DROUGHT MODELING and FORECASTING

–from Local to Global Scales–

Luis Samaniego

DROUGHT MODELING AND FORECASTING

DROUGHT MODELING AND FORECASTING
from Local to Global Scales

Kumulative Habilitationsschrift
zur Erlangung des akademischen Grades
doctor rerum naturalium habilitatus (Dr. rer. nat. habil.)
venia legendi: Hydrologie

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von

Luis Samaniego

Gutachter:
1. Prof. Dr. rer. nat. habil. Sabine Attinger
2. Prof. Dr. Kerstin Stahl
3. Prof.dr.ir. Remko Uijlenhoet

Verteidigt am 23. Februar 2021, veroffentlicht im Juni 2021

Drought Modeling and Forecasting / Luis Samaniego
Helmholtz Centre for Environmental Research - UFZ
Cover artwork: "The Bull of Heaven" / Sofía Samaniego

Printed in Germany.

To
María Eugenia,
Eduardo and Sofía

CONTENTS

1 Introduction **1**

1.1 Defining drought 1
1.2 From mythology to science 2
1.3 Socio-economic relevance 3
1.4 The grand challenge 5
1.5 Necessary elements for drought modeling 6
1.6 The parameterization and scale problems 8
1.7 Subject and aim of the thesis 10
1.8 Structure of the thesis 11
1.9 Research Statements 11

PART I DROUGHT MODELING

2 Relating macroclimatic circulation patterns **19**
L. Samaniego and A. Bárdossy

2.1 Introduction 19
2.2 Method 20
2.3 Application 25
2.4 Results and Discussion 27
2.5 Conclusions 32

3 Multiscale parameter regionalization **33**
L. Samaniego, R. Kumar, and S. Attinger

3.1 Introduction 33
3.2 Conceptualization of the System 36
3.3 Multiscale Parameter Regionalization 38

3.4 Application 40
3.5 Numerical Experiments and Evaluation Criteria 45
3.6 Analysis of Results 46
3.7 Conclusions 53

4 Uncertainty on drought analysis 57
L. Samaniego, R. Kumar, and M. Zink

4.1 Introduction 57
4.2 Soil Moisture Data 59
4.3 The mHM Soil Moisture Index 62
4.4 Results and Discussion 64
4.5 Summary and Conclusions 75

5 Towards seamless predictions 77
L. Samaniego et al.

5.1 Abstract 77
5.2 Introduction 77
5.3 Current parameterization techniques 79
5.4 Seamless parameterization framework 83
5.5 Experiments to reveal non-seamless parameterizations 87
5.6 Implementation of the parameterization protocol in PCR-GLOBWB 93
5.7 Conclusions 94

6 Multiscale and multivariate evaluation of water fluxes 97
O. Rakovec, ..., L. Samaniego

6.1 Abstract 97
6.2 Introduction 97
6.3 Data and Methods 99
6.4 Results and discussion 104
6.5 Conclusions 112

7 Conditioning mhm with patterns of land surface temperature 115
M. Zink, ..., and L. Samaniego

7.1 Abstract 115
7.2 Introduction 115
7.3 Study Domain and Data 117
7.4 Methodology 118
7.5 Results and Discussion 125
7.6 Summary and Conclusions 133

PART II DROUGHT FORECASTING & PREDICTION

8 Seasonal drought forecasting system over europe 139
S. Thober, ..., and L. Samaniego

8.1 Abstract 139
8.2 Introduction 139
8.3 Methods and Datasets 141
8.4 Results and Discussion 143

	8.5	Summary and Conclusions	151
9	**Propagation of uncertainties on to drought characteristics**		**153**
	L. Samaniego et al.		
	9.1	Abstract	153
	9.2	Introduction	154
	9.3	Method	155
	9.4	Results and discussion	156
	9.5	Conclusions	161
10	**Hydrological forecasts and projections in europe**		**165**
	L. Samaniego et al.		
	10.1	Abstract	165
	10.2	The rationale behind EDgE	166
	10.3	Climate data processing	167
	10.4	Hydrological modeling	168
	10.5	SCII: stakeholder co-design and estimation	169
	10.6	Uncertainties in Climate Projections	172
	10.7	Skill of Seasonal Forecasts	175
	10.8	Conclusions and Outlook	177
	10.9	Acknowledgments	177
		PART III DROUGHT IMPACTS	
11	**The german drought monitor**		**185**
	M. Zink, L. Samaniego et al.		
	11.1	Abstract	185
	11.2	Introduction	185
	11.3	Drought Monitoring	186
	11.4	Operational Drought Monitoring Framework	187
	11.5	Benchmark for the Recent 2015 Drought Event	190
	11.6	Conclusion and Outlook	192
12	**The effect of soil moisture anomalies on maize yield**		**195**
	M. Peichl, ..., and L. Samaniego		
	12.1	Abstract	195
	12.2	Introduction	195
	12.3	Data	197
	12.4	Regression Analysis	199
	12.5	Results and Discussion	201
	12.6	Conclusions	209
13	**Anthropogenic warming exacerbates european soil moisture droughts**		**211**
	L. Samaniego, S. Thober, et al.		
	13.1	Abstract	211
	13.2	Introduction	211
	13.3	Methods	212
	13.4	Results	212

13.5 Discussion and conclusions 215

PART IV LESSONS AND OUTLOOK

14 Lessons and outlook 229

14.1 Subject and aims of the thesis 229
14.2 General challenges and scientific relevance 229
14.3 Conclusions 231
14.4 Outlook 233

LIST OF FIGURES

1.1	Drought mythologized as the “Bull of Heaven”	2
1.2	Kircher’s system of springs, rivers and seas	3
1.3	Annual deaths by natural disaster	4
1.4	Economic losses due to Natural Hazards	4
1.5	Web of Science publications on floods and droughts	5
1.6	Typical drought modeling chain	6
1.7	The Richardson’s forecast factory	7
1.8	GCM resolution over time	8
2.1	Location of the Study Area	26
2.2	Time series and trends of selected variables.	28
2.3	Performance of three models to predict total drought duration in summer.	28
2.4	Spatial distribution of r of the total duration of high flows.	30
2.5	Time series for the total drought duration in summer.	31
2.6	Variation of the dispersion of the observed and calculated total drought duration in summer.	31
3.1	Schematic comparison between MPR and the standard simultaneous regionalization approaches.	35
3.2	General model structure.	36
3.3	Estimation of effective regionalized parameters at level-1.	37
3.4	Evaluation scheme for the evaluation of the continuity principle between two modeling scales.	40
3.5	Location of the Upper Neckar river.	41
3.6	Mean and quantile range of the spatial correlation of selected varioables.	47
3.7	NSE obtained between observed and simulated daily streamflow simulations using MPR and SR.	48
3.8	Sensitivity of the water balance across scales for the MPR and SR schemes).	49
3.9	Spatial variability of the porosity at three different modeling scales.	50
3.10	Conservation of mass for water fluxes across scales based on MPR and SR.	51
3.11	Discrepancy between fluxes simulated at two different modeling scales.	51

3.12 Land Surface Temperature from MODIS and simulated volumetric water content. 52
3.13 Performance of MPR and SR for daily discharge simulations. 53
4.1 Map of Germany indicating the main river basins 61
4.2 Ensemble monthly mean soil moisture fraction 65
4.3 Seasonality of the long-term soil moisture fraction 66
4.4 Parameter uncertainty of soil moisture fraction 66
4.5 Drought indices estimated with precipitation, runoff and soil moisture 67
4.6 Comparisin berainfall products DWD1 and DWD2 67
4.7 Parameter uncertainty of SMI 68
4.8 Time series of SMI and area under drought over Germany and major river basins 69
4.9 Probability of being at a drought severity class 69
4.10 Area under drought, duration, and magnitude of largest events 70
4.11 Sensitivity of the false positive rate to ensemble size 70
4.12 Severity at the peak of the largest drought events 71
4.13 Drought intensity evolution of the largest drought events 72
4.14 Bivariate density functions between drought area and total drought magnitude 73
4.15 Ensemble averaged Severity-Area-Duration curves of major drought events 74
4.16 Persistence map of the SMI 74
5.1 Porosity of typical LSM/HM over Pan-EU at various resolutions 81
5.2 Proposed seamless prediction framework 84
5.3 Seamless soil porosity fields at three spatial resolutions 86
5.4 Effect basin-wise calibration on soil moisture fraction 88
5.5 Porosity fields obtained using the majority upscale operator 89
5.6 Phase diagrams of monthly soil moisture fraction 91
5.7 Annual ET for the Rhine River with mHM, WaterGAP and PCR-GLOBWB 92
5.8 Porosity fields of PCR-GLOBWB before and after implementing MPR 93
5.9 Annual ET fields of PCR-GLOBWB for two spatial resolutions 94
6.1 Spatial map of the modelling domain 99
6.2 Schematic presentation of the experimental setup 103
6.3 KGE values between observed and simulated discharge for 400 EU basins 105
6.4 Factors influencing model predictability of discharge in terms of KGE 106
6.5 Model performance of mHM against observed TWS anomaly, actual ET, and SM 107
6.6 Model performance of discharge, actual ET and TWS 108
6.7 Monthly model performance for discharge, actual ET, TWS and SM 109
6.8 Monthly discharge, actual ET, TWS anomaly and SM standardized anomaly for 3 basins 110
6.9 Analysis of water balance closure error 110
6.10 Daily actual ET between mHM and local-scale eddy stations 111
7.1 German river basins used for parameter inference 118
7.2 Pattern similarity criterion 123
7.3 Simulated daily streamflow for mHM 125
7.4 Model performance variability at crossvalidation basins 126
7.5 Budyko plots and performance maps over Germany 127
7.6 Parameter range reduction based on streamflow and LST 128

7.7	Comparison of satellite retrieved and simulated LST	130
7.8	Simulated evapotranspiration obtained with two calibration strategies	130
7.9	Signal-to-noise ratio of evapotranspiration fields	131
7.10	Comparison of the optimization strategies using Q or Q and T_s	132
7.11	Evaluation of the ET at eddy flux towers	132
8.1	Skill of the NMME ensemble to reproduce reference SMI	144
8.2	NMME forecasting skill for precipitation and air temperatur	145
8.3	Exemplary time series of SM and SMI	146
8.4	Overall Pearson correlation and ETS for the NMME subensembles	147
8.5	Spatial distribution of ETS at one, three, and six month lead times	148
8.6	Area under drought based on the reference SMI dataset of ETS for the ESP	149
8.7	Probability density function for drought severity and drought area	150
9.1	Ensemble of the 30-year running mean RIs for scenarios RCP2.6 and RCP8.5	157
9.2	Mean range of RI of the GCMs and HMs variability	158
9.3	Ensemble mean drought magnitude and duration for RCP2.6 and RCP8.5.	159
9.4	CDFs of drought magnitude for historical and future periods RCP2.6 and RCP8.5	161
9-A.1	Graphical representation of the algorithm to estimate the GCM and HM uncertainty	163
9-A.2	Visualization of the GCM and HM uncertainties.	164
9-A.3	Mean range of drought magnitude and duration of the HMs and GCMs	164
10.1	EDgE modeling chain for seasonal forecast and climate prediction	167
10.2	Simulated mean daily streamflow from 1950 to 2011 for the EDgE domain	168
10.3	Workflow of the Climate Projection and Seasonal Forecast model chains	169
10.4	Projected changes in a high streamflow indicator under the RCP8.5	170
10.5	Verification of reference high flow soil moisture drought indicators	171
10.6	Uncertainty of the ensemble of the relative change of high streamflow indicator	172
10.7	Relative change of high streamflow indicator under the RCP8.5	173
10.8	Number of drought months per year expressed as signal to noise ratio	174
10.9	Effects of the sub-ensemble selection on the Brier Score	175
10.10	Brier skill score for dynamical seasonal forecasts over Europe	176
10-A.1	Evaluation of the hydrological models over 357 European basins	179
10-A.2	Effects of the sub-ensemble selection on the Brier Score skill	180
11.1	Framework of the German Drought Monitor	188
11.2	Soil water conditions from April to August in 2014 and 2015	190
11.3	Area affected in Germany during the drought event 2015 and hits on the GMD webpage	191
11.4	Ranking of the recently ongoing drought event in 2015	192
12.1	SMI data of May 2003 over German counties	199
12.2	BIC distribution for various polinomial models from May to October	202
12.3	Partial dose-response functions of the meteorological variables	206
12.4	Sensitivity of the functional form of temperature partial effects	207
12.5	Change of silage maize yield caused by significant SM Anomalies	208
13.1	Distribution functions of drought areas and durations for different levels of global warming	213
13.2	Spatial distribution of changes in drought area, duration and frequency	215
13.3	Changes in aridity for various warming levels	216

13.4	Changes of European population within large drought events	217
13-A.1	CDF of drought statistics using threshold τ of 0.1	218
13-B.1	Uncertainty in precipitation projections	219
13-B.2	Drought frequency for every hydrological model for various global warming levels	220
13-B.3	Projected changes in precipitation and temperatures for different levels of global warming	221
13-B.4	Verification of simulated streamflow	221
13-B.5	Projected global mean temperatures	222
13-B.6	Comparison between SMI and PDSI for RCP 2.6	224
13-B.7	Comparison between SMI and PDSI for RCP 6.0	225
13-B.8	Comparison between SMI and PDSI for RCP 8.5	225
14.1	Relationship of the theses with the elements of the modeling chain	230
14.2	Seamless SMI across spatial scales	232
14.3	Uncalibrated monthly NSE performance over 5500 GRDC stations	234
14.4	Pearson correlation of TWS anomalies of GRACE and mHM	235
14.5	Prototype of the South Asian drought monitor	235

LIST OF TABLES

2.1	Circulation Patterns according to Hess and Brezowsky	21
2.2	Explained variables, selected predictors, and data sources	26
2.3	Probability distributions and relevant characteristics.	27
2.4	Link functions used in this study.	27
2.5	Classification of circulation patterns.	27
2.6	Composition of the calibrated hydrological models.	29
2.7	Parameter estimates of the calibrated hydrological models.	29
2.8	r, BIAS, RMSE, and AIC for the period from 1961 to 1993.	30
3.1	Types of upscaling operators used in MPR	39
3.2	Description of basin predictors used in MPR.	41
3.3	Predictors used in the regionalization functions.	42
3.4	Regionalization (transfer) functions and upscaling operators used in mHM.	43
3.5	Effect of the sub-grid variability on three model efficiency statistics.	46
3.6	RMSE and NSE obtained with MPR and SR schemes at various modeling scales.	48
4.1	Uncertainty of characteristics of major drought events	70
4.2	Probability of finding a drought event in any of the top eight ranks	70
5.1	Data sources and parameterization methods	82
5.2	Efficiency of mHM, PCR-GLOBWB and WaterGAP	91
6.1	Description of input and evaluation datasets	100
6.2	Overview of the eddy covariance stations	113
7.1	Ranges for normalizing the error measures	124
7.2	Spearman rank correlations for streamflow and LST	129
8.1	Climate Forecasting models used in this study	141
8.2	The skill to reproduce reference SMI	145

10.1	Sectoral Climate Impact Indicators (SCIIs) in EDgE	169
10-A.1	Physiographic information used for the EDgE Project	178
11.1	Drought classes in the GDM	189
12.1	Mean and standard deviation of the meteorological variables over Germany	199
12.2	Pearson Correlation Coefficients of the Exogenous Variables	200
12.3	Adjusted Coefficient of Determination	203
12.4	Results of Regression Models employing precipitation and temperature	205
13.1	Multi-model ensemble statistics for different levels of global warming and IPCC regions	214

FOREWORD

I first became familiar with the work of Luis Samaniego in the fall of 2010, when he received the 2010 WRR Editor's Choice Award for his development of the Multiscale Parameter Regionalization (MPR) technique. Luis' work was based on the concept of the Representative Elementary Area (REA) which I had introduced a decade earlier, but which had been rarely referenced for parameterization of land surface models. In the MPR approach developed by Luis, the REA concept is one of the key concepts used to estimate effective parameters at multiple mesoscale resolutions based on the subgrid variability of geophysical predictors. At that time, the MPR approach was only built into the mHM model originally developed by Luis and his group at the UFZ and successfully tested in several basins in Germany. Although these concepts were originally applied at a regional scale, I could clearly see their potential for large scale modeling. In 2013, Luis visited me at Princeton to share the results of the first implementation of mHM in Germany and to brainstorm on potential cooperation opportunities. This mHM implementation would evolve to become the basis of the German Drought Monitoring system in place today at the UFZ. Later, during the Hyper-resolution Global Hydrological Modeling workshop held in Utrecht in 2014, I was pleased to see the first mHM implementation at the European level; with excellent cross-validation results. In 2016, Luis was invited to Princeton to deliver a talk entitled, "Towards Seamless Hydrologic Predictions Across Scales." Hearing this seminal presentation cemented my trust that Luis' approach has an extraordinary potential for seasonal forecasts and climate projections on a continental scale. I encouraged Luis to pursue funding to further develop his model into a fully operational seasonal forecasting system. With this goal in mind, we jointly work on several proposals and eventually secured funding from the C3S-Programme (ECMWF) to execute the project "End to End Demonstrator for Improved Decision Making in Europe (EDgE)". Based on EDgE results, Luis and I collaborated on five published papers covering topics related to seasonal forecasting and climate projections. Luis took the lead on two of them which are re-published in this book. These papers deal with a topic that has been at the center of my own research career – monitoring, modeling and prediction of droughts at continental and global scales. Luis' bottom up approach is at the forefront of this field and I am optimistic that his work will mark a striking change for operational systems and will provide a key contribution to address the immense challenge of delivering a global, and locally relevant, hyper-resolution system for drought (and flood) prediction and forecasting.

Eric F. Wood, NAE, FRSC, ATSE
Susan Dod Brown Professor (Emeritus) of Civil and Environmental Engineering

Princeton
27th of January of 2020

PREFACE

My fascination with Nature, science and water goes back to my childhood at the time when my father took me to the countryside to collect stones, watch big ferns and explore torrential rivers. From him I first learned the latin names of plants and fungi, and through his microscope, learned that there is much more stuff around us that we cannot see with the naked eye. At that time, the question that intrigued me most was: where does the water from a river come from? Years later, severe droughts and floods made me also wonder about the amazing power of water for sustaining and destroying life. The conscious aspiration to estimate the amount of water flowing at every moment in a river came much later during my master studies. From this first imagined question to later learning to calculate the answer and then actually estimating drought impacts, many things had to happen, from meeting my PhD supervisor (A. Bárdossy) to finding a postdoc job that allowed me develop the necessary tools (i.e., mathematical algorithms that constitute a model) to answer this question. To cope with this challenge, I needed to "stand on the shoulders of giants" in the discipline of land surface modeling (S. Manabe, R.A. Freeze, R.L.Harlan, J.C.I. Dooge, P.S. Eagelson, E.F. Wood) to be able to figure out, for example, how to include the sub-grid variability in a hydrological model and then to make it applicable everywhere. Furthermore, I needed to write, from scratch, the first lines of Fortran code for what was to become mHM and then to find a number of bright and extremely efficient colleagues to help me with this gargantuan enterprise. Later, it was necessary to understand how the leading experts in continental drought modeling (D.P. Lettenmaier, E.F. Wood, J. Sheffield, among others) are using statistical concepts to model droughts and then propose new methods so that the model and the drought indices are transferable across river basins and spatial scales. These were non-trivial tasks to master, but then, I was fortunate enough to work with a department head (S. Attinger) that believed that my work was promising and innovative, and hence, provided the resources along this wonderful intellectual adventure. Finally, I should recognize that this quest would not have been possible without the support and inspiring atmosphere I found at the Helmholtz Centre for Environmental Research - UFZ, and the infinite love and comprehension of my beloved family.

LUIS SAMANIEGO

Leipzig
March, 2020

ACKNOWLEDGMENTS

This work would not have been possible without the continuous support, inspiring discussions, and valuable feedback of all my co-authors over these years. Firstly, my utmost appreciation to Sabine Attinger for her confidence in my work. Looking in retrospect, I can state that the MPR technique is the synthesis between two modeling perspectives: one based on physical principles and the other one based on empirical hydrologic concepts. Her rigorous theoretical approach were the catalyst that inspired me to find a way to formulate a scaling procedure that embraced both standpoints. Many crucial papers in this work would not have been possible without the tenacious and inquisitive work of Rohini Kumar. He was my first PhD student and had to endure an uncountable number of setbacks until mHM with MPR finally worked. My deepest gratitude goes to all my colleagues that believed in the mHM project and contributed to its development: Matthias Cuntz, Maren Kaluza, Rohini Kumar, Juliane Mai, Oldrich Rakovec, Martin Schrön, Robert Schweppe, Pallav Shrestha, Stephan Thober, and Matthias Zink. My deep appreciation also goes to Mathias Zink for initiating the preparation of the multiple datasets that constitute the basis of the German Drought Monitor, for coding its operationalization scripts and to Andreas Marx and his team for keeping it running smoothly since 2014. My thanks to the EDgE modeling team –my dream team–: Rohini Kumar, Oldrich Rakovec, Stephan Thober, Niko Wanders, Ming Pan, Justin Sheffield, and Eric F. Wood. Without them, we could not have achieved the unthinkable in less than two years with so little resources. To Stephan Thober for his commitment and trust in all projects we have started and are now ongoing. His modeling and coding skills have been fundamental to accomplishing the results presented in this work, and as a co-author, one of most diligent and analytical persons with whom I have worked. My great appreciation and thanks to the UFZ scientific computing support team for keeping our Linux Cluster running and being there when we have troubles. I am deeply indebted to my dear friend Pablo Beltran for helping me to improve the English language in this thesis. Pablo has always provided constructive feedbacks that has helped me, over the years, to improve my writing skills. The remaining errors are mine. To Sofía Samaniego for the interpretation of the ancient Mesopotamian mythology, the "Bull of Heaven", and the beautiful artwork shown on the cover of this book. Finally to all institutions that provided funds to carry out many of the papers that constitute this thesis, among them: to the projects REKLIM, EDA, and ESM funded by the Helmholtz Association of German Research Centres, the Copernicus Climate Change Service and the European Centre for Medium-Range Weather Forecasts (ECMWF) who funded and coordinated EDgE, and the Federal Ministry of Education and Research (BMBF) who supported the project HOKLIM. Many thanks to the UFZ for supporting my position and many of my colleagues over the years.

L. S.

ACRONYMS

AGU	American Geophysical Union
BCE	Before current era
CE	Current era
CEDIM	Center for Disaster Management and Risk Reduction Technology at KIT
CHIME	Copernicus hyperspectral imaging mission
COPA	Committee of Professional Agricultural Organisations
COGECA	General Confederation of Agricultural Cooperatives
CONUS	Contiguous United States
DLR	Deutsches Zentrum für Luft- und Raumfahrt, the German Aerospace Center
DWD	Deutscher Wetterdienst, the German meteorological service
ECMWF	European Centre for Medium-Range Weather Forecasts
EDgE	End to End Demonstrator for Improved Decision Making in Europe
EM-DAT	The international disasters database
ENIAC	Electronic Numerical Integrator and Computer, first, Turing-complete, electronic general-purpose digital computer
EOS	Earth and space science news and analysis from AGU
E-OBS	Gridded meterological forcing dataset generated by the European Climate Assessment & Dataset project.
ESP	Ensemble streamflow prediction approach
ESM	Earth system model
ERA-5	Fifth major global ECMWF ReAnalysis
ET	Evapotranspiration
FLUXNET	Global network of micrometeorological eddy covariance tower sites

GCM	Global climate model
GDM	German Drought Monitor
GHM	Global hydrological model
GRACE	Gravity recovery and climate experiment from NASA and DLR
GRACE-FO	GRACE following mission
GRDC	Global Runoff Data Centre (WMO)
HRU	Hydrological response unit
HM	Hydrological model
H-SAF	EUMETSAT satellite application facility on support to operational hydrology and water management
ICON	(Icosahedral nonhydrostatic GCM operated by the DWD
ICOS	A European Research Infrastructure for Integrated Carbon Observation System
IFS-v5	Integrated forecasting system version five operated by the ECMWF
IPCC	Intergovernmental Panel on Climate Change
ISI-MIP2	The Inter-sectoral Impact Model Intercomparison Project from PIK
KIT	Karlsruhe Institute of Technology
LSM	Land surface model
LST	Land surface temperature
LSTM	Copernicus land surface temperature monitoring satellite mission
mRM	Multiscale routing model
mHM	Mesoscale hydrological model
ML	Machine learning algorithms
MODIS	Moderate resolution imaging spectroradiometer from NASA
MPR	Multiscale parameter regionalization
MOPEX	Model parameter experiment from USA
NASA	National Aeronautics and Space Administration
NMME	North American multi-model ensemble
NSE	Nash-Sutcliffe model efficiency
NWP	Numerical weather prediction
OFDA/CRED	International disaster data - our world in data
PIK	Potsdam Institute for Climate Impact Research
RCP	Representative concentration pathway for greenhouse gas concentrations
REA	Representative elementary area
RS	Remotely sensed or remote sensing
RI	Runoff index
SCII	Sectoral climate impact indicators
SM	Soil moisture
SMI	Soil moisture index
SPI	Standardized precipitation index
SR-15	IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels,
SWOT	Surface water and ocean topography from NASA and the National Centre for Space Studies (France)

tECV	Terrestrial essential climate variables from WMO
TWS	Terrestrial water storage anomaly from GRACE
WMO	World Meteorological Organization
WRR	Water Resources Research journal from AGU

INTRODUCTION

CHAPTER 1

INTRODUCTION

> For most of the history of our species we were helpless to understand how nature works. We took every storm, drought, illness and comet personally. We created myths and spirits in an attempt to explain the patterns of nature.
>
> —Ann Druyan

1.1 Defining drought

Drought is an Old English word originated from the ancient Germanic root "dreug", which means a continuous dry weather injurious to vegetation, leading to a shortage of water in water bodies such as creeks, rivers or lakes. Because this definition is quite general, there have been many interpretations of this word during the passing of the centuries which have led to confusion with other terms such as aridity or water scarcity. For these reasons, the World Meteorological Organization defined drought as "an insidious natural hazard characterized by lower than expected or lower than normal precipitation that, when extended over a season or longer period of time, is insufficient to meet the demands of human activities and the environment". This means that this natural hazard "is a temporary aberration, unlike aridity, which is a permanent feature of climate" (*WMO*, 2006). In other words, it denotes a transient state of the atmosphere or the hydrosphere.

The WMO also defined four types of droughts: meteorological, agricultural, hydrological and socioeconomic. The first three types define drought as a deficiency below a threshold defined over a predetermined period of time. The target variables in this case are: precipitation, soil moisture and runoff (*Samaniego et al.*, 2013). The socio-economic drought refer to the "relationship between the supply and demand for some commodity or economic good, such as water, livestock forage or hydroelectric power, that is dependent on precipitation" (*WMO*, 2006).

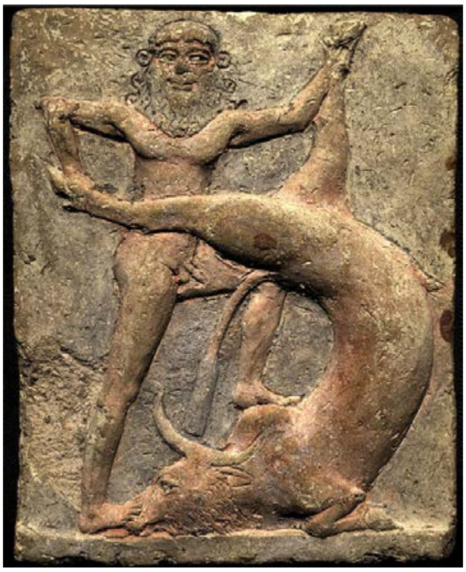

Figure 1.1 Ancient Mesopotamian terra-cotta relief (c. 2250 | 1900 BCE) showing Gilgamesh slaying the "Bull of Heaven". This mythological beast represents the first conceptualization of the drought (source Royal Museums of Art and History, Brussels).

Over millennia, this natural hazard could not be understood or defined but rather was associated with mystical connotations. As a result, drought occurrences became the origin of mythology, curses, and folklore. To my knowledge, the first personification of the drought hazard appears in the Epic of King Gilgamesh of Uruk in Mesopotamia (c. 2000 BCE). In this ancient heroic legend, Gilgamesh braved and defeated drought in the form of the mythical beast the "Bull of Heaven" (Fig. 1.1). This legend describes how the bull's voracious appetite caused the drought and wreaked havoc on the people in ancient Mesopotamia (*Heathcote*, 2016). Another well know Biblical passage was that of the Pharaoh's Dream (Genesis 41), in which seven years of famine will follow seven years of abundance.

Current research has shown that past mega-droughts changed the history of mankind by stimulating the expansion and migrations of early modern human populations (*Scholz et al.*, 2007). Over the course of centuries, several civilizations declined or collapsed due to the onset of long and severe droughts. The main consequences of these catastrophic events are now well documented by archeologists, for example, the Mycenaean and Hittite empires c. 1200 BCE (*Bryson and Murray*, 1977), the Neo-Assyrian Empire c. 609 BCE (*Sinha et al.*, 2011), the Mayan Civilization c. 1000 CE (*Evans et al.*, 2018; *Heathcote*, 2016), the Aztec Civilization 1454 CE (*Therrell et al.*, 2004), the Khmer Kingdom c. 1300 CE (*Stone*, 2009).

1.2 From mythology to science

Droughts have always been present during human history. References to the occurrence of this natural hazard abound in ancient religious books such as the Bible (e.g., in 2 Chronicles 7:13-14) and the Quran (e.g., in chapter 44), and in most cases, as one of the ultimate divine punishments for disobeying gods' commandments.

In history books, conversely, it is used to describe socio-economic situations characterized by crop failure; undersupply, shortage or scarcity of agricultural products; famine and mass-migration and as a cause of civilization collapse.

Despite all the human suffering, the socio-economic losses and the ecological consequences inflicted by the onset of droughts over the centuries, mankind was hopeless in their attempts to predict and ameliorate the consequences of droughts. In old times, it appeared that the only possible actions to mitigate the pain and loss caused by this natural event were religious and folkloric rituals. Ancient artists (*Tan*, 2015; *Therrell et al.*, 2004) and painters, impressed by the power of the human drama, immortalized such moments on canvas with extraordinary expressive and melancholic images. Representative examples are "Prayer in Time of Drought" by G. G. Myasoyedov (1881) or "Gorta" by L. L. Davidson (1846). Common people created maledictions such as "the curse of One Rabbit" by the Aztecs (*Therrell et al.*, 2004), and priests composed *(*pro-pluvia) rogations or prayers songs like the "Key to the Rain, a New Song for a Time of Drought", published in Prague in 1679 (cited in *Brázdil et al.*, 2018).

The Empiricism, which is the core of the scientific method (*Popper*, 1935) states that the first logical action to understand a natural phenomenon should be to obtain quantitative observations of the phenomenon of interest. This information will, in turn, help to discover patterns and make causal inferences and eventually discover natural laws describing the phenomenon. This mental setting was –perhaps– the rationale for ancient erudite scholars or industrious people to keep records of past hydrological events. The first written records of the occurrence of a drought event date from 206 BCE (*Tan*, 2015). Later, the Egyptians created sophisticated hydrological metering systems such as the Nilometer (c. 715), while in Central Europe, landmarks in buildings and "hunger stones" at the riverbeds became common starting in the 15th century (*Benito et al.*, 2015). The oldest mark appearing in a hunger stone, that is still readable, dates from the year 1417 at Děčí-Podmokly in the river Elbe (*Brázdil et al.*, 2018). Scholars realized that recording drought events was not enough to understand the causes, the extension, and the severity of droughts hitting a region of interest. They noticed that the key to understanding this phenomenon requires a realistic conceptualization of the global hydrological cycle.

The conceptualization of this fundamental cycle, alone, was one of the grand challenges for ancient and Middle Age philosophers and scientists alike. Understanding the main processes took millennia. Thales of Miletus (600 BCE) was the first scientific philosopher that asserted that one of the "ultimate stuff of the world" must be water because it is clearly vital for all known forms of life, it appears in all states of matter, and it covers most of the Earth's surface. Anaximander (c. 570 BCE) was the first to describe evaporation as the cause of the movement from water to the sky, one of the fundamental fluxes of this natural cycle. Xenophobes of Colophon (c. 530 BCE) contributed with concepts for the role of clouds as responsible for the transport and production of rain, which in turn feed springs and rivers. Anaxagoras of Clazomenaiz (c. 460 BCE) realized that these various water processes constitute a closed cycle involving the movement and storage of water (*Dooge*, 2001).

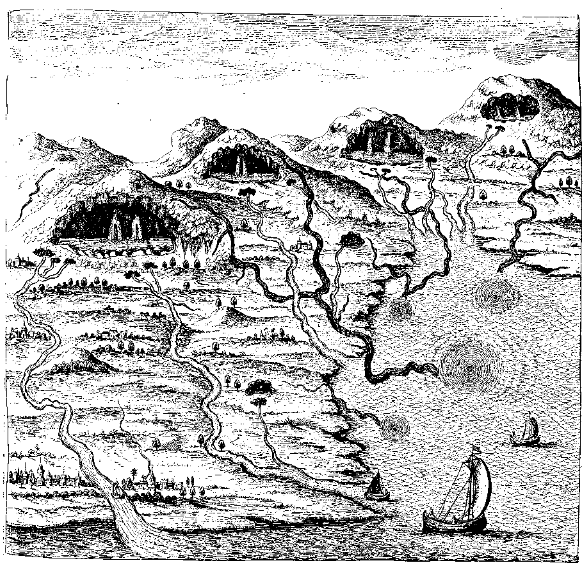

Figure 1.2 Reverse and erroneous conceptualization of the water cycle during the Middle Ages. Mundus Subterraneus – Kircher's system of springs, rivers and seas (1665 edn. vol. 1, p. 233)– is an example of a hydrological hypothesis formulated without observational evidence.

One of the first complete speculations about of the functioning of the water cycle is attributed to Aristotle (350 BCE), who considered that the water cycle was an endless, cyclical, and never changing system, i.e. in perpetual steady state, with the Sun's solar radiation as the main driving force behind the hydrological cycle. This idea prevailed until Vitruvious (1 BCE) stated that groundwater is the result of precipitation falling in the mountains, which after being infiltrating the Earth's surface, would appear later in streams and springs in the lowlands. Almost no progress was achieved until the Enlightenment. In this period there were also plenty of misconceptions about the water cycle. An example of one of them was provided by Athanasius Kircher (1665), who hypothesized an explicit explanation of the reverse hydrologic cycle in which springs on top of the mountains where fed by underground channels that link them with enormous whirlpools at the bottom of the seas (see Fig. 1.2). Kircher, like many of his contemporaries (e.g., Herbinius), could not explain the role of precipitation in the hydrological cycle.

The first correct inrterpretation of the water cycle "based on observations" is attributed to Leonardo da Vinci (c. 1500). Da Vinci was the first scholar to realize that the throughput of the main rivers surpasses by countless times, the volumes contained in the world's oceans. Da Vinci, as one of the luminaries of the Renaissance put forward a paradigm shift: "away from a dominant religion-centered paradigm of the Middle Ages to the science-centered paradigm, based on empiricism and deduction" (*Pfister et al.*, 2009). Another of Da Vinci's great innovations was the introduction of the water balance, relating inputs and outputs of the system. Probably, the anonymous book "Origin of Fountains", published in Paris in 1674 constitutes the first quantification of the main water cycle components: precipitation and runoff (*Dooge*, 1959). Although these contributions were great scientific advances in understanding the mechanisms of the hydrological cycle, the goal to understanding the evolution of droughts was still centuries away.

1.3 Socio-economic relevance

As briefly described above and according to numerous in-depth treatises about this topic (e.g., *Brázdil et al.*, 2018; *Bryson and Murray*, 1977; *Dai*, 2011; *Diamond*, 2011; *Heathcote*, 2016; *Scholz et al.*, 2007), droughts have been impactful during human history and for the evolution of ecosystems (*Godfree et al.*, 2019). Estimating the real economic losses and death tolls attributable to this hazard is quite difficult due to the lack of reliable sources and a good metric to estimate material cost, not to mention the need to adjust for economic inflation.

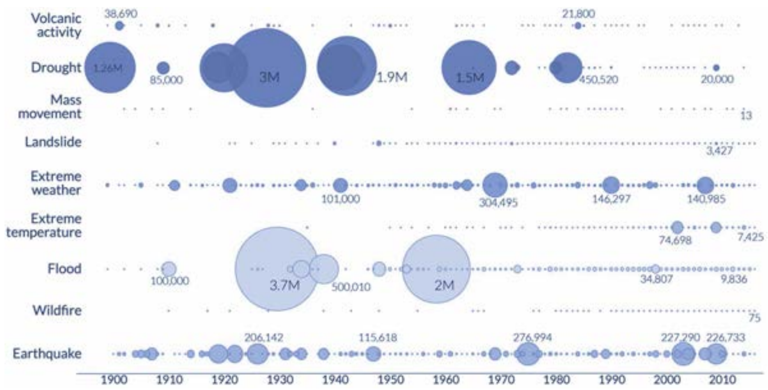

Figure 1.3 Global deaths from natural disasters (1900-2016) (*Ritchie and Roser*, 2019) (based on the OFDA/CRED International Disaster Database urlwww.emdat.be). The size of a bubble represents the total death count per year by type of disaster. Graph source: OurWorldlnData.org under CC-BY license.

As a result, reliable statistics exist only from the beginning of the 19th century. According to *Ritchie and Roser* (2019) (see Fig. 1.3), the deadliest hazards since 1900 are floods and droughts. This figure shows clearly that better planning and infrastructure, as well as opportune international relief assistance, have greatly contributed to minimizing the death toll from the 1970s onwards. Direct weather related hazards (e.g., droughts, extreme weather, floods, extreme temperature and wildfires) represent about 88% of the death toll since 1900, and specifically, droughts and floods about 51% and 30%, respectively. The drought death toll in these statistics includes the long term effects of drought-induced famines. According to the EM-DAT database (*Guha-Sapir et al.*, 2015), droughts affected 2.2 billion people worldwide between 1950 and 2014, thus making droughts the second most impactful natural disaster after floods (3.6 billion people affected). In the same period, the death toll has been 2.21 million people.

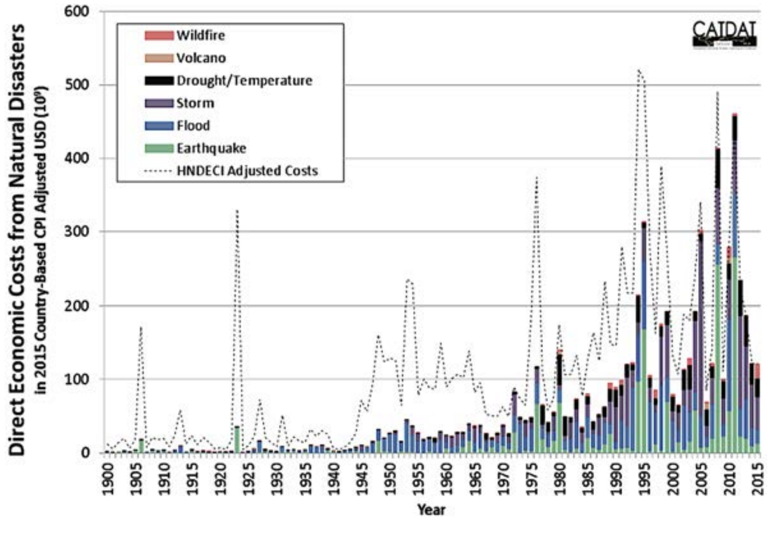

Figure 1.4 Economic losses in US$ due to natural hazards based on *Daniell et al.* (2016). Data source: Center for Disaster Management and Risk Reduction Technology CEDIM, KIT.

According to *Daniell et al.* (2016), natural disasters have caused a US $7 trillion loss since 1900. Figure 1.4 depicts the growth of economic losses based on a sophisticated global analysis of 35 000 natural disaster events since 1900. The losses have been estimated on a country-by-country basis including a GDP-deflator based price index to convert historical costs to 2015 US dollars. In absolute terms, economic losses due to natural hazards have significantly increased since the 1960s. Around 50% of economic losses between 1900 and 2015 have been caused by floods, droughts and related wildfires. This figure also indicates that drought related losses have significantly increased after the 1980s.

In Europe, in particular, *Zink et al.* (2016) indicated that "the costs per drought event during the period from 1950-2014 are estimated to be 621 Mio. EUR, the costliest amongst all natural disasters that occurred in this region (*Guha-Sapir et al.*, 2015)."

According to the European Commission (as reported in *Zink et al.* (2016)), "the frequency of droughts has increased since 1980 and will, very likely, further increase (*EEA*, 2012). To date, 11% of the European population and 17% of the area of the EU have been affected by water scarcity (*European Commission*, 2007, 2010). For example, the 2003 drought event, which covered major parts of Europe, caused 7,000 fatalities in Germany alone (*European Commission*, 2012) and had an agro-economical impact of 1.5 billion EUR. On the European level, the death toll was estimated to exceed 70,000 (*Robine et al.*, 2008), and the agro-economical impact was estimated to be 15 billion EUR (*COPA-COGECA*, 2003). This severe drought impacted many components of societal life. It disrupted irrigation, inland navigation, and power plant cooling (*Fink et al.*, 2004; *Parry et al.*, 2007)."

1.4 The grand challenge

For these reasons, droughts have been identified by the IPCC (*IPCC*, 2007) as the trigger of a web of impacts across many sectors leading to land degradation, migration (*Wilbanks et al.*, 2007) and substantial socio-economic costs. Reports cited above, along with the extensive literature available on this subject, clearly indicate that a better understanding of the evolution of droughts and its implications is of crucial importance for planning activities related to water resources, land use, infrastructure, power generation, wildfire mitigation, human health and welfare; and to food security intrinsically related to agricultural production (*Samaniego and Bárdossy*, 2007). Consequently, having the ability to globally monitor, model and forecast and/or predict the occurrence of droughts seamlessly, along several time scales going from weeks to seasons and to decades, constitutes one of the great challenges in hydro-meteorological sciences (*Wood et al.*, 2011).

The dificulty of this challenge is further complicated by the unprecedented anthropogenically induced climate change. Not providing satisfactory answers to this challenge would compromise the survival of mankind and quite likely contribute to a shift in the natural ecosystems towards unknown tipping points. Put differently, reaching those loci at which a natural ecosystem, driven by massive disturbances such as drought or wildfires, cannot recover its initial states (*Steffen et al.*, 2018).

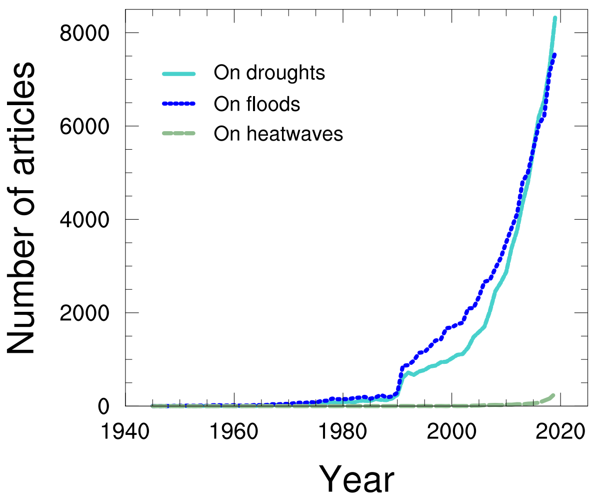

Figure 1.5 Number of ISI listed publications listed in the Web of Science `wcs.webofknowledge.com` under the keywords "drought", "flood" and "heatwaves" since 1945 until 2019. No restriction on disciplines. Date of access 2020/01/19.

Drought events during recent decades such as the Millennium Drought (2002–10) in Australia (*van Dijk et al.*, 2013), the 2003 European drought (*Fink et al.*, 2004; *Parry et al.*, 2007) or the California drought (2011-17) (*Swain et al.*, 2014), and the exorbitant increase in economic losses (see Fig. 1.4), have motivated extraordinary scientific productivity during the last two decades. Figure 1.5 shows the evolution of peer-review scientific articles published in all scientific disciplines covering every aspect related to three natural hazards directly linked with the water cycle over the land surface. It can be noted that, although the total number of written articles on floods are more numerous than those on droughts, the number of annual publications on the latter has surpassed that of the former after 2016. Research on heatwaves and their impacts is emerging only recently and it is not extensive yet. Research on these three subjects has significantly increased during the last two decades.

This figure also provides a clear starting point for a short recount on the evolution of the state-of-the-art on modeling the global water cycle. In some sense, the modest progress before the 1980s is a consequence of severe misconceptions undermining the progress in hydrology until as recent as 1965, among them: 1) ignoring the complexity of the spatio-temporal variability of dominant-hydrological processes which led to "downgrading hydrology from a natural science to an appendage of hydraulic engineering" (*Yevjevich*, 1968), which mainly deals with "classic problems of water supply and natural hazards reduction" (*Eagleson*, 1986); 2) the limited scale of interest, which was primarily that of the catchment (i.e., no more than 100 km^2), and with the assumption that "the atmosphere [is] an independent driver of the hydrological processes" (*Eagleson*, 1986); 3) the lack of coherent advancement in the hydrological sciences (*Dooge*, 1982) whose separate disciplines (*Rajaram et al.*, 2015) dedicated entire decades to performing research on specific theoretical problems, which on themselves are interesting, but missed addressing problems such as scalability, transferability, and parameterization (*Clark et al.*, 2016; *Samaniego* *et al.*, 2017); 4) the lack of global spatial data at high resolution (mostly remotely sensed) such as terrain elevation, soil texture, land cover, and forcing datasets; 5) the poor technological readiness levels of the software and hardware needed to address the challenge mentioned above (*Clark et al.*, 2017); and finally, 6) the lack of a Popperian approach for scientific discovery leading to hypothesis driven frameworks that address the scaling and similarity problems postulated by *Dooge* (1982) and summarized in *Peters-Lidard et al.* (2017) and *Samaniego et al.* (2017).

1.5 Necessary elements for drought modeling

To address this grand challenge, we need to know the past, the current, and the future states of energy and water fluxes of the atmosphere at given points in time and space as well as to be able to describe the evolution of the dominant hydrological processes that take place on the Earth's surface. These are non-trivial tasks, and the complexity to represent them into models has hindered progress until recently. The frustration of leading scientists during the 1980s regarding the slow progress in addressing this challenge is clearly expressed by Eamonn Nash, who said " I find it difficult to believe that we should enter the third millennium after Christ measuring rainfall in little buckets and guessing evaporation" (*Schultz*, 1988).

The advent of the "Electronic Age" (after the 1970s) offered unprecedented possibilities for acquiring global coverage of key hydrological variables via remote sensors alongside an exponential growth in computing power. As a result of these positive developments and the advancement in state-of-the-art meteorological and hydrological theories in the 1990s, the feasibility of addressing this challenge became real for the first time in human history.

Modeling is a complex human activity because of the crucial trade-offs that have to be made to reach a final objective, in this case modeling the occurrence of hydrological and agricultural droughts. According to *Popper* (1935), modeling is an interactive research process that starts with the observation of a natural system (e.g., the water cycle) aiming for a "mental" abstraction of the main elements that are necessary to faithfully describe the evolution of the system over time. Abstraction implies a reduction of the system's complexity, which is often formalized as a set of equations, which we call a model. Consequently, a model constitutes an elaborate hypothesis of the dynamics of the system that should be subject to falsification. In other words, its predictions should be contrasted with new data to establish the ability of the model (i.e., its skill) to reproduce them.

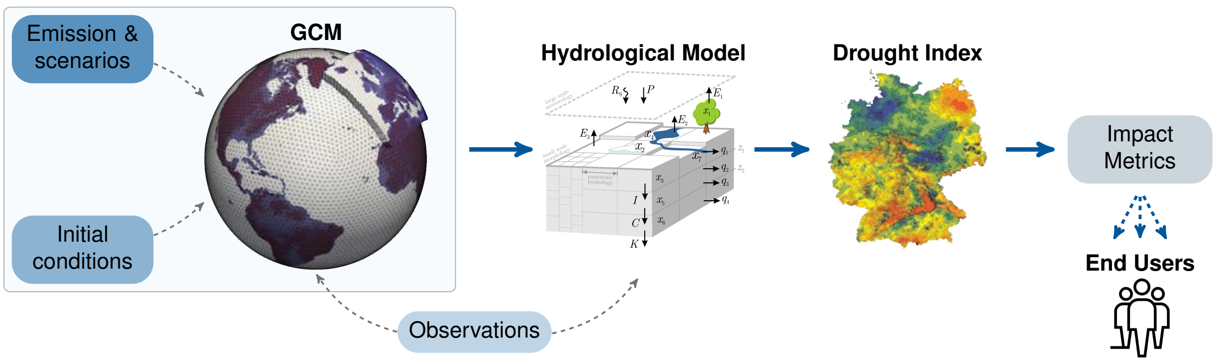

Figure 1.6 Typical drought modeling chain composed of a GCM, a HM, a drought index and an impact metric targeting a specific need of a group of stakeholders (i.e., end-users such as farmers, water planers, or dam operators). Graphic sources, GCM grid correspond to the ICON model (`www.dwd.de`); The hydrological model logo correspond to mHM v5.10 (`www.ufz.de/mhm`); and drought index (SMI) is based on *Samaniego et al.* (2013).

Under the current perspective, the simplest modeling chain required to reconstruct past droughts or to forecast them at a given location across the globe is depicted in Figure 1.6. It consists of a global climate model (GCM) (e.g., IFS-v5 ECMWF) to estimate the state of the atmosphere, a hydrological or land surface model (HM/LSM) (e.g., `www.ufz.de/mhm`) to simulate the fate of water over the landscape, a drought indicator (e.g., the soil moisture index (SMI) described in *Samaniego et al.* (2013)) to quantify the occurrence, extension and intensity of a drought event, and finally an impact model to estimate its consequences for a specific user, for example, changes in crop yield (e.g., *Peichl et al.*, 2018), which are necessary for economic planning or food security planning. This modeling chain requires, of course, additional information such as initial conditions, emission scenarios, and large amounts of observational data.

Realizing the input data, theories, numerical techniques, and computational capabilities required to develop this holistic drought modeling chain took five centuries of hard work since the visionary concept introduced by Leonardo da Vinci. First, all atmospheric physics and chemistry describing the hydrodynamics of a parcel of moist air needed to be discovered, including Newton's Laws of motion, the laws of thermodynamics and the fundamental principles of conservation of mass and energy (Lomonosov, Lavoisier, Noether, Helmholtz).

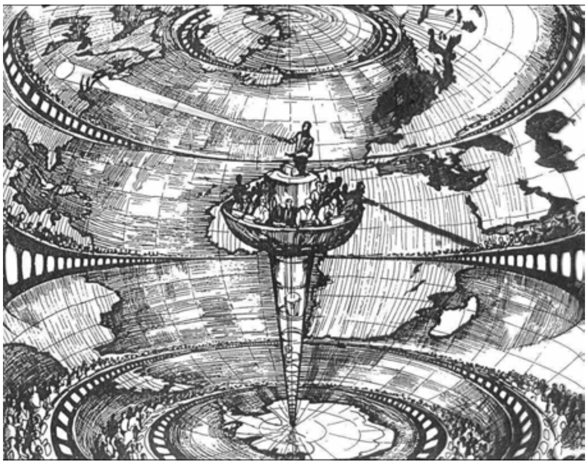

Figure 1.7 Artistic representation of the Richardson's central forecast-factory (A. Lannerback). Dagens Nyheter, Stockholm. Reproduced from L. Bengtsson, ECMWF, 1984.

The goal of predicting future states of the atmosphere based on its present state was expressed for the first time in the Bjerknes' Manifesto (*Bjerknes*, 1904). Bjerknes stated that if physical laws control the states of the atmosphere, then the "necessary and sufficient" conditions for a rational solution of the problem of meteorological prediction are:

1. One has to know with sufficient accuracy the state of the atmosphere at a given time.
2. One has to know with sufficient accuracy the laws according to which one state of the atmosphere develops from another.

The knowledge required to achieve this goal (and in the future to develop a GCM) was synthesized by Lewis F. Richardson in 1922 seminal book entitled "Weather Prediction by Numerical Process" (*Richardson*, 1922). This visionary meteorologist attempted, for the first time, to make a weather forecast by manually integrating the dynamic differential equations using finite differences! Its results were not encouraging but he was certain that "[p]erhaps some day in the dim future it will be possible to advance the computations faster than the weather advances and at a cost less than the savings to mankind due to the information gained. But that is a dream" (*Richardson*, 1922). He even imagined that to model the weather of the whole globe, a central forecast-factory would be needed (see Fig.1.7). In other words, he conceived the first massively parallel processor made up of 64 000 computers (i.e., persons performing predetermined computations) (*Richardson*, 1922, p.219).

The reasons for Richardson's failure were the poor initial conditions used for the atmosphere and the Earth's surface, the poor parameterization used for the hydrologic processes describing the evolution of key state variables such as soil moisture, and the error propagation in the numerical scheme he applied. The idea was a breakthrough, but premature. The first applications of Richardson's blueprint for a GCM were only possible after 1945 with the advent of the first electronic computers (e.g., the ENIAC), which dramatically increased the speed at which numerical algorithms could be solved.

The various kinds of numerical weather prediction models, global climate models, land surface schemes, and hydrologic models that exist today are the result of a continuous and elaborate formalization process that has led to a set of equations based on fundamental physical principles whose numerical solution is possible today by means of sophisticated algorithms. This modus operandi has not changed much since 1922, when Richardson wrote his seminal book in which the foundations for numerical weather forecasting were set down.

Furthermore, advances in hydrology towards a distributed and process-based description of dominant land surface hydrological processes are lagging behind for various reasons, some of which were mentioned above. The renaissance in this realm of science coincide with the emergence of new journals like Water Resources Research (AGU) (*Rajaram et al.*, 2015) in 1965. As a result, a few years later, an approach similar to that used by Richardson was put forward by *Freeze and Harlan* (1969) as the first blueprint for a physically-based, digitally-simulated hydrologic response model. Their conceptualization was based on physical principles and several parameterizations introduced by Darcy, Horton, De Wiest, Saint-Venant, Liggett, Woolhiser, Mannings among others. Although, *Freeze and Harlan* (1969) did not test their ideas with a numerical model — they conceded that their paper is "more an *artist' s conception* than a true *Blueprint*" —, it provided fresh new ideas to inspire land surface modeling pioneers to develop the first distributed hydrological models. In 1975, K. Beven was one of the first who attempted to apply the Freeze and Harlan blueprint to a real basin. The computer power and spatial information available at that time, however, only allowed him to simulate a tiny basin with an area of barely 21 ha. The simulation results, according to him, were not worth publishing (*Beven*, 2001).

Data and computation limitations were overwhelming at the time when Freeze and Harlan wrote their ground-breaking paper. Contemporary researchers, Amorocho and Hart, even wrote that (cited in *Freeze and Harlan*, 1969, p.238):

> Prohibitive amounts of input data would be required, far beyond practical limitations even for small experimental plots.

During the next decades, further improvements to the original blueprint (i.e., *Freeze*, 1974) were made and a number of models were developed following the original or modified conceptualizations (e.g., *Abbott et al.*, 1986; *Bronstert et al.*, 1998; *Kirkby*, 1988). These models, however, could not be run at scales larger than that of the watershed scale (10-50 ha) or in hillslopes. Recently, *Beven* (2002) proposed a "new" blueprint but until now there is no implementation based on his proposal.

1.6 The parameterization and scale problems

After the 1970s, once electronic computers became indispensable research tools, the skill and efficiency of the GCMs and LSM/HMs were mainly improved by: 1) increasing the complexity of the dynamical conceptualizations, e.g., from models omitting ocean dynamics to current GCMs fully resolving it, having now aerosols (SA), carbon cycle (TAR), dynamic vegetation (TAR) (*Bonan*, 1995), atmospheric chemistry (AR4), land ice(AR4) (*IPCC*, 2007); 2) improving the numerical algorithms used for solving the system of partial differential equations (*Brunet et al.*, 2015), 3) improving the quality of the initial conditions by developing data assimilation methods (*Brunet et al.*, 2015), and 4) by doubling resolution once the storage capacity and computational power allowed. These facts are clearly reflected in the change of the spatial resolution of the GCMs employed as the scientific basis for the IPCC assessment reports since 1990.

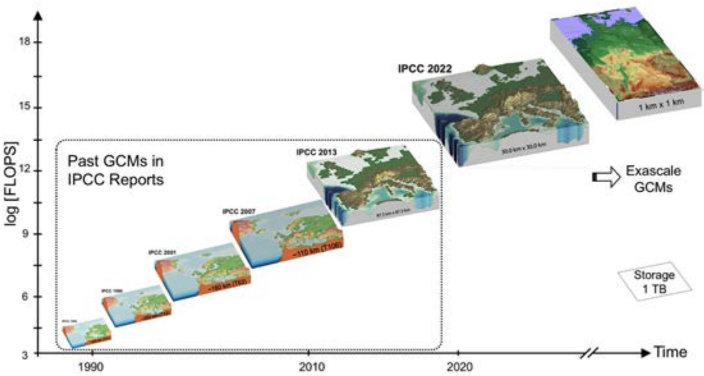

Figure 1.8 Evolution of the spatial resolution, computational demand, and storage of the GCMs employed in the IPCC reports since 1990. FAR (IPCC, 1990), SAR (IPCC, 1996), TAR (IPCC, 2001a), and AR4 (2007). Note: The scale of both axis is nonlinear and approximate. Adapted from *IPCC* (2007, WG1).

According to (*Lynch*, 2008), the GCM spatial resolution has doubled every five years since 1990 (Fig. 1.8). It should be noted that increasing the model resolution by a factor of two implies about ten times as much computing power and storage. Recently, the hyper-resolution initiative in land surface modeling has also opted along this pathway (*Bierkens et al.*, 2014; *Wood et al.*, 2011).

One of the major challenges to deploying the modeling chain depicted in Fig.1.6 is to distinguish and to recognize that "there are scales and physical processes that can not be represented by a numerical model, regardless of the resolution" (*Stensrud*, 2007). Parameterization is the "process by which these important processes that can not resolved directly by a numerical model are represented" (*Stensrud*, 2007). Put differently, it is a simplified and idealized representation of the physical phenomenon at a given scale in the form of a simplified equation that requires existing variables and numerical constants, often called parameters.

Model parameterizations on the land surface have changed little during the past decades. It should be noted that *Richardson* (1922, p.9) already recognized that the theory and "constants" (what we now call parameters) "must be appropriate to the size" of the grid element. He also suggested that these parameters should be found experimentally (e.g., *Richardson*, 1922, p.108), if possible. At present, many of these "constants" are still confined to their respective source codes or as lookup-tables as noted by *Mendoza et al.* (2015) and *Cuntz et al.* (2016) in the NOAH-MP LSM. Writing models and source codes with written down constants is an old practice, see e.g., *Crawford and Linsley* (1966)'s source code for the Stanford Watershed Model IV, in 1966. This poor practice has negative effects on results because it hinders the modeler's capability to explore the sensitivity of these parameters on outputs and the possibility to infer them using observations.

In a recent assessment towards seamless prediction of the Earth System across time scales, *Dirmeyer et al.* (2015) stated very clearly that improvements in understanding of hydrological land processes and their parameterization are crucial for increasing the predictability of GCMs at sub-seasonal (days to weeks) and seasonal forecasting (months) time scales. The major drawback that these authors see at the moment is the coarse resolution of the LSMs, which are a fundamental part of GCMs, and their lack of scalability.

In addition to that, *Bauer et al.* (2015) remarks that "grid-scale invariance" and improvements in physical process description have remarkably increased the skills of numerical weather prediction (NWP) models in the recent past. These authors, however, point out that for improving skill, increasing resolution is not the only answer; there is still a need for improving parameterizations of the land surface components of NWP models.

Parameterization of LSMs/HMs is therefore an old, ubiquitous, and recurring problem. Moreover, until as recently as 1982, more than a decade after the famous Freeze and Harlan blueprint was published, the progress on this fundamental component of the modeling chain was insignificant. For these reasons, *Dooge* (1982, p.269) concluded that:

> [T]he parameterization of hydrologic processes to the grid scale of general circulation models is a problem that has not been tackled, let alone solved.

Lack of effective progress did not mean that there were not brilliant ideas around to address the problem. In fact, the keys to the solution were hidden in the vast literature on the subject. For example, *Crawford and Linsley* (1966, p.9) had already expressed serious concerns with the problem of over-parameterization of HMs and the lack of parameters for the "ungaged[sic] areas". Their solution was to pursue parsimonious parameterizations (i.e., having a "minimum number of independent parameters") so that they can be extended to ungauged locations. *Freeze and Harlan* (1969, p.240,256) also put forward ideas for parameter regionalization when they indicated that it will be "necessary to extrapolate results of representative measurements of physical parameters to other points in the basin", and that the ideal blueprint may lead to over-parameterization and, hence, a "simplification of the model is needed" to achieve "workable dimensions". *Dooge* (1982, p.245) also realized that "the process of parametrization at the macro-scale of the micro-scale processes may be based either on the nature of these fine scale processes or else determined empirically at the macro-scale". In other words, he provided potential pathways that can be followed to derive a macro-scale model, either by process-based upscaling techniques or by a simplified effective model using effective parameters that represent the micro-scale processes.

Attempts to use inverse modeling (without regularization functions) to estimate parameters for HMs/LSMs would simply not work for making a model transferable across locations or scales. As a result of these experiences *Leavesley et al.* (1983, p.50) concluded that finding parameters for "distributed" hydrological models via optimization constitute an ill-posed problem due to the large number of degrees of freedom (i.e., unknown) and the few constraints that can be derived from integral observations such as streamflow. It should be noted that the word "distributed" was introduced to differentiate HMs from those that do not consider the spatial variability of the parameters and process within a basin. The latter we called "lumped" models. Leavesley's solution consisted of reducing the degrees of freedom by introducing a concept called hydrologic response units (HRUs), which represent areas of the landscape that have quasi-similar hydrological response and hence share the same hydrological parameters. This solution opened up huge expectations, but until now, there is no process-based method to derive the HRUs. Currently, only empirical approaches exist (*Samaniego et al.*, 2017).

An insight into the "parameterization problem" was, however, stated almost four decades ago. *Dooge* (1982) pointed out that this crucial issue is intimately related with the "scaling" problem (*Grayson and Blöschl*, 2000; *Sivapalan et al.*, 2004), which, in his opinion, was also a crucial "unresolved problem" in hydrology. In fact, two decades later, *Blöschl* (2001) pointed out that scaling is "the cornerstone for a unifying theory in hydrology". By the beginning of the 1990s, we still did not have a solution for these two problems, but at least we knew that they were related. The next milestone was achieved in a key workshop held at Princeton University (1990) aimed at exploring the status of land surface parameterizations within climate models. As a synthesis of this workshop, *Wood* (1990) concluded that among the reasons hindering the progress of improving the representation of land-atmosphere interactions in GCMs is the poor experimental settings to address the problem of "scale". With this aim in mind, he concluded that we should ask instead:

> What modeling experiments need to be performed to resolve the "scale" question and what is the tradeoff' among model complexity, the physical basis for land parameterizations and observational data for estimating model parameters?

This was, and still is, a fundamental question in hydrology, one among many others (*Blöschl et al.*, 2019; *Clark et al.*, 2016; *Peters-Lidard et al.*, 2017) that have not been addressed until recently (see e.g., *Samaniego et al.*, 2017). Another key insight of the Princeton workshop based on empirical results that, in my opinion, had serious implications for the further development of LSM/HMs, was formulated by *Wood* (1990) as follows: "[t]he inadequate representation [land-atmosphere interactions in GCMs] reflects the recognition that the well-known physical relationships, which are well described at small scales, result in different relationships when represented at the scales used in climate models."

An extended review on the evolution of process-based hydrologic models in recent decades along with the key theories and concepts was presented in *Clark et al.* (2017) and, thus, it is not necessary to repeat it here. The representative elementary area (REA) concept introduced by *Wood et al.* (1988) and later expanded by (*Wood et al.*, 1990; *Woods et al.*, 1995), however, is briefly introduced here because it is fundamental to the development of scale-independent land surface and hydrological models, to address the proposed grand challenge, and to develop a skillful modeling chain of the kind depicted in Fig. 1.6.

This REA concept and the empirical evidence supporting it (*Wood et al.*, 1990; *Woods et al.*, 1995) provided

> ... confidence that simpler macroscale models will perform well at large grid scales which have significant subgrid variability, thus questioning the wisdom of detailed land parameterization which ignores process heterogeneity.

Moreover, *Wood et al.* (1988) and many others, confirmed that the length of the REA (say ℓ_1) is process dependent and strongly influenced by topography. This implied that it would be possible to develop meso- and macro-scale hydrological models in which "the variability [of the parameters] can be explicitly represented only at scales larger than the element size (ℓ_1) while variability at the sub-element scale must be represented in a lumped way" (*Grayson and Blöschl*, 2000).

This fundamental conclusion was the inspiration to develop the mesoscale Hydrological Model (mHM, `www.ufz.de/mhm`) exhibiting, for the first time, a sophisticated regularization scheme called the "Multiscale Parameters Regionalization" (MPR) (*Samaniego et al.*, 2010a). In this model, MPR simultaneously addressed the parameterization and scaling problem mentioned previously. The scaling problem in MPR "is addressed by using process-specific representative elementary areas (REAs) that determine the minimum computational grid size ℓ_1 at which the continuum assumptions can be used without explicit knowledge of the actual patterns of the topography, soil, or rainfall fields" (*Samaniego et al.*, 2017).

Currently the mHM has been proven to be a plausible and parsimonious hypothesis to address the grand challenge, subject of this treatise. Up to now has been thoroughly tested and evaluated in Germany (*Kumar et al.*, 2010, 2013; *Samaniego et al.*, 2010a,b, 2013; *Wöhling et al.*, 2013; *Zink et al.*, 2017), in the conterminous USA (*Kumar et al.*, 2013b; *Rakovec et al.*, 2019), across Europe (*Kumar et al.*, 2015; *Rakovec et al.*, 2016a,b; *Samaniego et al.*, 2019a), Asia, Africa and South America (*Dembélé et al.*, 2020; *Huang et al.*, 2016; *Samaniego et al.*, 2011, 2016), and currently on over 5000 GRDC stations across the globe (*Samaniego et al.*, 2019b). Although, these results are quite encouraging, many operational land surface models used in GCMs supporting the IPCC reports still exhibit poor scalability and transferability, as shown in *Samaniego et al.* (2017).

1.7 Subject and aim of the thesis

This habilitation work focuses on the work carried out over the last decade aimed at putting together a modeling chain able to perform modeling and forecasting of hydrological and agricultural droughts from local to continental scales. In other words, I attempt to provide an answer to the trillion-dollar question implicit in Fig.1.4.

This thesis does not cover the immense work necessary to code, setup, parameterize and run a GCM, which is the first component of the modeling chain depicted in Fig. 1.6. This fundamental work has been subject of innumerable IPCC reports, books, research articles and dissertations, and hence out for scope of this work. This thesis, however, contributes fundamental insights on how to improve the parameterization of existing operational land surface models based on the criticism reported in *Samaniego et al.* (2017). Examples of such endeavors are ongoing work, having been reported at international geoscience conferences (e.g., *Thober et al.*, 2019a, 2020).

This work will also show how to use and preprocess the outputs of GCMs so that they can be used as drivers for hydrological and land surface models in offline mode (i.e., uncoupled from GCMs and used for drought impact assessment). A shortcoming of this approach is that it does not account for feedback from the highly resolved land surface water fluxes into the GCM. It is expected, however, that the new generation of Earth System Models, exhibiting improved land surface hydrology and parameterization schemes such as MPR will do so. In this aspect there is additional work being carried out.

Using GCM-simulated or observed forcing data, it will be shown how to estimate drought indices and their uncertainties, quantify probabilities of occurrence of drought events and their area, duration and magnitude, and, finally, to show potential pathways on how to use this modeling chain outputs for estimating drought impacts such as changes on wheat yield over a specific region and time.

In summary, key elements of this work are: 1) the propagation and quantification of uncertainties along the various elements of the modeling chain; 2) the strict verification criteria to evaluate the performance of the models across scales and locations; and 3) the development of high standard, reusable and portable open-source code that can be used for hydrological modeling of drought forecasting and projections.

1.8 Structure of the thesis

This thesis is subdivided into three methodological parts and one for discussion, synthesis and outlook to achieve the goal mentioned above, namely:

Part I Towards Drought Modeling across Scales, which covers six papers on the macro-drivers of droughts, the MPR development and mHM, parameter uncertainty, seamless predictions, multiscale evaluation of fluxes and states, and remotely sensed conditioning of mHM.

Part II Forecasting and Predicting Droughts, which covers three papers on the development and verification of a seasonal forecasting and prediction modeling chain, and the propagation of uncertainties along its various elements.

Part III Estimating Drought Impacts, which covers three papers on the development of the German Drought Monitor, the prototype of a data-based impact model for crop yield in Germany and the estimation of soil moisture droughts in Europe until 2100.

Part IV Lessons Learnt and Outlook, which synthesizes the main findings and potential pathways to improve the proposed modeling chain.

It should be noted that the selected published manuscripts (Chapters) used in each of these Parts are not sequential in time. This is the consequence of the order in which the research funds were procured and the length of the peer-review process rather than on methodological aspects. The selection of manuscripts is based on the relevance of the theses which were the subject of the individual papers that constitute this treatise on drought modeling and forecasting.

A summary of each Chapter and their main research statements are presented next. In each individual paper presented here, each thesis was treated as a research hypothesis subjected to strict statistical falsification.

1.9 Research Statements

The papers constituting the three parts of this thesis were selected according to their relevance to investigating specific components of the proposed model chain, which is assumed would help to address the challenge stated above. The research statements, or theses, that are put forward in each paper are denoted hereafter with the letter $\mathcal{T}_{p.i}$). The index p denotes a methodological part and i a running number.

1.9.1 Towards Drought Modeling across Scales

Chapter 2 Samaniego, L., and A. Bárdossy (2007), Relating macroclimatic circulation patterns with characteristics of floods and droughts at the mesoscale, *J. Hydrol.*, *335*, 109–123

The main objective of this paper is to "let the data speak for themselves" a la Gould. This means, use the law of parsimony (i.e., Occam's razor) in machine learning techniques and no ad-hoc assumptions to explore in a large data set which explanatory variables (e.g., many physiographic, land cover, and climatic characteristics) are able to predict seasonal extreme runoff characteristics like the total [hydrologic] drought duration at the mesoscale basins.

The main theses put forward in this study are:

$\mathcal{T}_{1.1}$: Annual land cover change fractions (i.e., forest area) and the dryness index tallying the total number of "dry spells" with decreasing antecedent precipitation during a season are statistically significant predictors of drought duration and specific streamflow deficit (i.e., streamflow below the 10% percentile threshold) in mesoscale basins during a hydrological summer season (MJJASO). Drainage area, floodplain's slope and drainage density are also significant predictors.

$\mathcal{T}_{1.2}$: Drought duration and specific streamflow deficit during summer are highly correlated variables in mesoscale basins.

Chapter 3 Samaniego, L., R. Kumar, and S. Attinger (2010a), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, *46*(5), w05523

This paper presents, for the first time, a series of cross-validation experiments covering several scales and locations using the process-based spatial-explicit model called mHM. This model exhibits the multiscale parameter regionalization technique as a way forward to addressing the *Dooge* (1982) scaling and parameterization problems, the issues of overparameterization and equifinality described by *Beven* (2001) and to address the transferability tests suggested by *Klemeš* (1986). Technically, MPR is a sophisticated regularization technique, commonly used in mathematics and statistics to solve inverse problems. This term refers to a process of introducing additional information to solve an ill-posed problem and/or to prevent overfitting and to improve the generalizability of the learnt model. In MPR, the additional information is given by regularized (or regionalized) model parameters at the sub-grid resolution of the model. The subgrid representation of the model parameters require additional geo-physiographic properties of the land surface, parameter specific regularization (or transfer) functions, and several scale-invariant coefficients called global parameters. For the novelty and success of this method, the authors of this article received the WRR Editors' Choice Award 2010. The source codes of mHM, mRM (routing) and the stand alone MPR can be found at `https://git.ufz.de/mhm`, `https://git.ufz.de/mhm/mrm`, and `https://git.ufz.de/chs/mpr`, respectively.

The main theses put forward in this study are:

$\mathcal{T}_{1.3}$: The REA concept is a fundamental notion to estimate effective parameters in a mesoscale hydrological model. The overall REA-scale can be inferred with the MPR technique.

$\mathcal{T}_{1.4}$: The finer the resolution of the subgrid variability of the model parameters, the better the model predictions at the REA scale.

$\mathcal{T}_{1.5}$: MPR global parameters enable the transferability of a model across scales and locations with a minimum performance loss.

Chapter 4 Samaniego, L., R. Kumar, and M. Zink (2013), Implications of Parameter Uncertainty on Soil Moisture Drought Analysis in Germany, *Journal of Hydrometeorology*, *14*(1), 47–68

The aim of this study was to analyze the effects of the MPR parameterization technique on simulated soil moisture over Germany. Using mHM the daily soil moisture fields at $4{\times}4\,\mathrm{km}^2$ were reconstructed using observed meteorological forcings interpolated with External Drift Krigging. In this study, a non-parametric method was used to estimate the soil moisture index (SMI). Major soil moisture drought events for Germany were also reconstructed for the first time. The implications of parameter uncertainty on drought identification were discussed and assessed. The code used for this study can be found at `https://git.ufz.de/chs/progs/edk_nc` and `https://git.ufz.de/chs/progs/SMI`.

The main theses put forward in this study are:

$\mathcal{T}_{1.6}$: A single parameter set for a given LSM or HM is inadequate to estimate water fluxes and related state variables at high spatiotemporal resolutions, considering that both inputs and model parameters over large modeling domains are subject to considerable uncertainties.

$\mathcal{T}_{1.7}$: Any drought characteristic (e.g., severity and duration) based on simulated soil moisture is prone to large variability due to parametric uncertainty.

$\mathcal{T}_{1.8}$: Ignoring parameter uncertainty in drought identification will led to large false positives.

Chapter 5 Samaniego, L., R. Kumar, S. Thober, O. Rakovec, M. Zink, N. Wanders, S. Eisner, H. Müller Schmied, E. H. Sutanudjaja, K. Warrach-Sagi, and S. Attinger (2017), Toward seamless hydrologic predictions across spatial scales, *Hydrology and Earth System Sciences*, *21*(9), 4323–4346

In this paper, we analyze the state-of-the-art LSMs and HMs to reveal that most of them do not have consistent hydrologic parameter fields across scales. We perform multiple experiments with the mHM, Noah-MP, PCR-GLOBWB, and WaterGAP models to demonstrate the pitfalls of deficient parameterization practices currently used in most operational models. We propose a general model protocol to describe how MPR can be applied to any LSM/HM.

The main theses put forward in this study are:

$\mathcal{T}_{1.9}$: Ad-hoc parameterization schemes, “brute” force optimization or inadequate upscaling operators lead to discontinuous and scale dependent parameter fields.

$\mathcal{T}_{1.10}$: The flux-matching test of water fluxes across scales is a necessary condition to obtain quasi scale-invariant MPR global model parameters.

$\mathcal{T}_{1.11}$: The multiscale parameter regionalization (MPR) technique provides a practical and robust method to estimate seamless parameter and flux fields across scales, in any LSM or HM.

Chapter 6 Rakovec, O., R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego (2016c), Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, *J. Hydrometeorol.*, *17*(1), 287–307

In this study, the mHM model parameterized with the MPR is tested across 400 European river basins. The model fluxes and states, constrained using the observed streamflow, are evaluated against gridded evapotranspiration, soil moisture, and total water storage anomalies, as well as local-scale eddy covariance observations. This multiscale verification is carried out in a seamless manner at the native resolutions of available datasets, varying from 0.5 to 100 km.

The main theses put forward in this study are:

$\mathcal{T}_{1.12}$: Performing parameter estimation on a LSM/HM based only on streamflow-related metrics is a necessary but not sufficient condition to warrant the proper partitioning of incoming precipitation into various spatially distributed water storage components and fluxes.

$\mathcal{T}_{1.13}$: Multivariate parameter estimation or assimilation scheme is necessary for improving the ability to predict regional water fluxes and states over large domains.

$\mathcal{T}_{1.14}$: MPR parameterization allows estimation of fluxes from local- (eddy covariance footprint, 10^2 m) to regional-scales (satellite footprint, 10^5 m), in a seamless manner.

Chapter 7 Zink, M., J. Mai, M. Cuntz, and L. Samaniego (2018), Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature, *Water Resources Research*, *54*, 2976–2998

In this study, we developed a bias-insensitive pattern-matching criterion to guide the parameter optimization on spatial patterns remotely-observed state variables (e.g., satellite-based land surface temperature). The proposed method is extensively tested in six distinct large German river basins and cross-validated in 222 additional basins in Germany. Additionally, a simple but efficient diagnostic algorithm (derived from the energy balance equation) to estimate land surface temperature within mHM. This paper was selected as Editors' Highlights and presented in EOS (AGU) (*Bierkens*, 2018).

The main theses put forward in this study are:

$\mathcal{T}_{1.15}$: The uncertainty in the global model parameters of a hydrological model will decrease when streamflow and land surface temperature are considered simultaneously.

$\mathcal{T}_{1.16}$: Model parameters constrained with remotely sensed land surface temperature significantly improve the estimation of evapotranspiration at basin and plot levels.

$\mathcal{T}_{1.17}$: The bias-insensitive pattern-matching criterion improves the identifiability of parameters of a hydrologic model.

1.9.2 Forecasting and Predicting Droughts

Chapter 8 Thober, S., R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego (2015), Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME), *Journal of Hydrometeorology*, *16*(6), 2329–2344

The main objective of this study was to evaluate the skill of the "North American Multi-Model Ensemble" (NMME) for soil moisture drought forecasting over Europe. Downscaled forcings of this ensemble were used as forcings of the mHM model. The skill of the NMME-based forecasts was compared against those based on the ensemble streamflow prediction (ESP) approach for the hindcast period of 1983–2009. Subensembles combinations of the NMME members were also investigated.

The main theses put forward in this study are:

$\mathcal{T}_{2.1}$: Dynamic drought forecasts (i.e., driven by GCMs) are consistently higher than that of ESP-based ones over the entire European domain and lead times.

$\mathcal{T}_{2.2}$: Subensembles selection is a promising alternative to the full ensemble and hence are useful for operational seasonal soil moisture (SM) drought forecasting because of marginal performance losses and low computational demand.

$\mathcal{T}_{2.3}$: Seasonal SM forecasts, specially over extreme conditions (i.e., droughts) exhibit strong and ubiquitous influence of the initial hydrologic conditions, highlight the need for robust parameterization schemes such as MPR.

Chapter 9 Samaniego, L., R. Kumar, L. Breuer, A. Chamorro, M. Flörke, I. G. Pechlivanidis, D. Schäfer, H. Shah, T. Vetter, M. Wortmann, and X. Zeng (2016), Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins, *Climatic Change*, *141*(3), 435–449

In this study, based on the ISI-MIP2 project simulations, we attempt to advance our understanding on the propagation of forcing and model uncertainties on to century-long time series of drought characteristics using an ensemble of GCMs and HMs under a broad range of climate scenarios and regions. In this study, a sequential sampling algorithm is proposed to address this issue and to disentangle the uncertainty originated by GCMs and HMs. Here a runoff index (RI), similar in concept to SMI is used to be able to compare runoff estimated by several HMs.

The main theses put forward in this study are:

$\mathcal{T}_{2.4}$: The uncertainty contribution of the GCMs on RI and derived drought characteristics outweighs that from the HMs regardless of the hydrological regime represented by the selected large-scale river basins.

$\mathcal{T}_{2.5}$: Given a GCM forcing, the drift in the RI time series of a given HM is practically indistinguishable from the ensemble RI. Therefore, the drift mainly arises from the uncertainty in the GCM forcings.

$\mathcal{T}_{2.6}$: The uncertainty in drought characteristics is dependent of the greenhouse gas concentration scenario (RCP).

Chapter 10 Samaniego, L., S. Thober, N. Wanders, M. Pan, O. Rakovec, J. Sheffield, E. F. Wood, C. Prudhomme, G. Rees, H. Houghton-Carr, M. Fry, K. Smith, G. Watts, H. Histal, T. Estrella, C. Buontempo, A. Marx, and R. Kumar (2019a), Hydrological forecasts and projections for improved decision-making in the water sector in Europe, *Bull. Am. Meteorol. Soc.*, *100*, 2451–2472

In this study, we developed a high-resolution multi-model ensemble of state-of-the-art climate and hydrological models to deliver, for the first time, an ensemble of 36 hydrometeorological change metrics co-designed with key water sector stakeholders in Europe. The model chain comprises two modes: one for seasonal forecasting (2 GCMs) and another for climate projections (5 GCMs). Both modeling chains share the same settings for the four LSM/HMs used in this project. This study constitutes the culmination of more than a ten-year quest for research needed to deploy, for the first time, a prototype of the modeling chain depicted in Fig. 1.6.

The main theses put forward in this study are:

$\mathcal{T}_{2.7}$: Due to parameter and structural uncertainties, an ensemble of multi-hydrological models is better suited to capture the uncertainty propagation in simulated fluxes and state variables, and thus, drought indices, than a single hydrological model.

$\mathcal{T}_{2.8}$: The combined GCM and HM uncertainties are not equally distributed over time or space.

$\mathcal{T}_{2.9}$: The uncertainty of derived SCIIs is not equally distributed between atmospheric (GCM) and land surface models (HM), with some regions in which the former is greater than that of the latter, and vise-versa.

$\mathcal{T}_{2.10}$: A multi-model ensemble (GCMs/HMs) exhibits higher forecasting skills (for droughts and floods) than Ensemble Streamflow Prediction (ESP).

1.9.3 Estimating Droughts Impacts

Chapter 11 Zink, M., L. Samaniego, R. Kumar, S. Thober, J. Mai, D. Schäfer, and A. Marx (2016), The German drought monitor, *Environmental Research Letters*, *11*(7)

In this study, the operationalization of the modeling chain used to deploy the German drought monitor (GDM `www.ufz.de/duerremonitor`) is presented and tested. It also provides an updated version of the largest soil moisture drought events in Germany since 1950, which were originally presented in (*Samaniego et al.*, 2013). The GDM produce daily, quasi-real time (latency of 4 days) SMI for Germany at a spatial resolution of $4{\times}4\,\mathrm{km}^2$. The GDM reached over 1 088 000 website hits on February 2019, which represent 25% of the UFZ visits until this date. The GDM system was the first implementation of the modeling chain shown in Fig.1.6 that used observed data instead of GCM-based simulations as forcing data, and focused on current soil moisture states on a soil column up to 1.8 m in depth.

The main theses put forward in this study are:

$\mathcal{T}_{3.1}$: The GDM system resolution and latency parameters allow detection of the emergence, the probability of occurrence, and the potential severity of ongoing drought events.

$\mathcal{T}_{3.2}$: The GDM system delivers timely information about the onset, extent, and intensity of drought events.

$\mathcal{T}_{3.3}$: The general public, the press, and many stakeholders are demanding and expecting accurate and timely information regarding the evolution of agricultural droughts in Germany.

Chapter 12 Peichl, M., S. Thober, V. Meyer, and L. Samaniego (2018), The effect of soil moisture anomalies on maize yield in Germany, *Natural Hazards and Earth System Science*, *18*(3), 889–906

In this study, a complete drought modeling chain as shown in Fig.1.6 is tested. Here parametric, reduced-form fixed-effect panel models are employed to investigate the intra-seasonal predictability of soil moisture (transformed to SMI) to estimate silage maize yield in Germany. In this study, we validate and compare results to similar state-of-the art approaches that use only meteorological variables.

The main theses put forward in this study are:

$\mathcal{T}_{3.4}$: Soil moisture improves the capability of statistical models aimed at predicting silage maize yield compared to standard methods that neglect it.

$\mathcal{T}_{3.5}$: Temporal patterns of seasonal soil moisture and its persistence contributed significantly to the crop yield model predictability.

$\mathcal{T}_{3.6}$: The SMI, as any anomaly based index, is advantageous for climate econometric impact models because they are less prone to systematic errors.

Chapter 13 Samaniego, L., S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. Wood, and A. Marx (2018), Anthropogenic warming exacerbates European soil moisture droughts, *Nat. Clim. Change*, *4*

In this study, using the ensemble of hydrological and land-surface models, forced with bias-corrected downscaled general circulation model output presented in *Samaniego et al.* (2019a), we estimated, for the first time in Europe, the impacts of 1–3 K global mean temperature increases on soil moisture droughts during the period from 2010 until 2100. Here, we also estimate the potential number of people in Europe that may be potentially affected by extreme droughts and related changes in aridity. These assessments focused on the major ecological regions in Europe. Currently, this paper is listed as a “Highly Cited Paper”, “Hot Paper” in the Web of Science.

The main theses put forward in this study are:

$\mathcal{T}_{3.7}$: Anthropogenic induced warming will cause unprecedented increases in the area affected by the largest soil moisture drought and its duration. Drought magnitude, consequently, will also increase.

$\mathcal{T}_{3.8}$: The largest historical droughts observed during the control period will become more frequent and thus, due to their increased occurrence, events of this magnitude will no longer be classified as extreme in the future.

$\mathcal{T}_{3.9}$: Changes in aridity, drought area, duration and frequency will be region specific. The highest changes will be expected in regions with higher air temperatures during the summer season.

PART I

TOWARDS DROUGHT MODELING ACROSS SCALES

CHAPTER 2

RELATING MACROCLIMATIC CIRCULATION PATTERNS WITH CHARACTERISTICS OF FLOODS AND DROUGHTS AT THE MESOSCALE

L. Samaniego[1] and A. Bárdossy[2]

[1] Helmholtz-Centre for Environmental Research-UFZ
[2] Institute of Hydraulic Engineering, University of Stuttgart

Originally published at:
Samaniego, L., and A. Bárdossy (2007), Relating macroclimatic circulation patterns with characteristics of floods and droughts at the mesoscale, J. Hydrol., 335, 109–123

Abstract

The prediction of extreme hydrological events in mesoscale catchments has been a main concern in hydrology because of their considerable societal impacts and because of the compelling evidence that anthropogenic activities significantly modify their occurrence likelihood. In this paper, nonlinear generalized models were used to predict extreme runoff characteristics like the specific volume, the frequency of high flows, and the total drought duration. Explanatory variables included many physiographic, land cover, and climatic characteristics such as mean slope, aspect, elevation, type of geological formations, shares of a given land cover type, and many composed indicators relating antecedent precipitation index and atmospheric circulation patterns. All time-dependent variables were estimated semiannually for each subcatchment. The proposed method was tested in 46 sub-catchments belonging to the Upper Neckar River basin covering an area of approximately 4000 km^2 during the period from 1961 to 1993. The results of this study indicated that macro circulation patters derived from either subjective or operational classifications combined with other explanatory variables can be effectively used to predict seasonal extreme runoff characteristics at the mesoscale. Moreover, the results indicated that most runoff characteristics exhibited a distributional element other than normal and that the selection of nonlinear generalized models was an appropriate choice to deal with the heteroscedasticity of model errors.

2.1 Introduction

The quantification of the magnitude and duration of meteorological extremes, as well as their probability of occurrence in a mesoscale catchment, has long been a main concern in hydrology and related disciplines because these anomalies might induce significant changes on several water-cycle-related state variables (e.g. soil moisture). These changes, in turn, might lead to the occurrence of either flood or drought spells (high- and low-flow regimes). The magnitude of the environmental degradation and the subsequent demographic and economic impacts at regional level are strongly dependent on the magnitude and duration of these events. Typical examples of the environmental degradation induced by the shortage of water are forest fires, stress on the supply side of the food chain, and soil erosion caused by the combined effects of vegetative cover reduction and wind action. An excess of water, on the other hand, also creates environmental problems such as the destruction of the endemic flora and fauna, insect infestations, and soil erosion caused by rill and gully erosion. Both extreme events entail substantial socioeconomic consequences such as heavy reduction of agricultural production and power generation, damages to the basic infrastructure and the manufacturing sector, as well as famines and people's migration.

Moreover, there is compelling evidence (e.g. see *Houghton et al.* (2001)) that the anthropogenic disruptions of the environment are significantly modifying the likelihood of occurrence of these events in a given period. Consequently, it would be crucial to investigate how the magnitude and the probability of occurrence of high and low flows might be affected by land use/cover and climatic changes.

Hydrological extremes occurring at the mesoscale are largely dependent on the occurrence of particular macroclimatic processes in the atmosphere, which are currently modeled by complex General Circulation Models (GCMs). GCMs, however, are not able to predict mesoscale hydrologic extremes mainly because these models are conceived to represent the main features of the atmospheric circulation rather than the regional climatic details. Moreover, GCM outputs depend on a given forcing scenario (e.g. CO_2 emissions) and, therefore, are highly uncertain.

In spite of that, it is possible to use their outputs, among them upper level winds, geopotential heights, and sea level pressure, as predictors of mesoscale observations such as precipitation. This procedure, which is generally termed empirical downscaling, is carried out using methods like multiple linear regressions, neural networks (*Cannon and Whitfield*, 2002), fuzzy rules (*Bárdossy et al.*, 2002), or regression tree approaches (*Li and Sailor*, 2000).

Another downscaling alternative is to use proxies for the state of the atmosphere at a given location and time as predictors of mesoscale variables. These proxies include: macroclimatic indexes such as the Pacific North America Index (PNA), the El Niño Southern Oscillation (ENSO), or the European atmospheric Circulation Patterns (CP) (*Hess and Brezowsky*, 1969). This type of information is particulary relevant to assess the *effects of macroclimatic changes* in mesoscale catchments, especially in ungauged basins or in those with very low density of meteorological stations. For instance, *Reddmont and Koch* (1991), *Bárdossy and Plate* (1992), and *Shorthouse and Arnell* (1997) used this procedure to link macroclimatic indices with the magnitude and the spatial distribution of meteorological variables (e.g. precipitation, temperature). Alternatively, *Dracup and Kahya* (1994), *Piechota and Dracup* (1996), and *Stahl and Demuth* (1999) linked them with stream flow characteristics such as partial duration of floods or drought spells. To our knowledge, however, there is no reference in the literature (excluding *Samaniego* (2003)) in which this approach has been used to assess simultaneously the influences of land use/cover and climate changes on hydrological extremes.

Two main types of CP classification techniques, subjective and objective, can be distinguished in the literature. The main advantage of the subjective classification is that the meteorologist's experience is applied on a daily basis in the classification. This advantage, however, entails a shortcoming for some practical applications such as medium range forecasting; namely, that the results of the classification can not be reproduced automatically (*Yarnal*, 1993). Objective classifications, on the contrary, operate on a given data set and derive day-by-day classified CPs using automated algorithms. The results of this kind of classification are ideal for the integration with climate-change simulations because they can be reproduced rapidly without human intervention (*Bárdossy and Filiz*, 2005; *Bárdossy et al.*, 2002).

In this study, low- and high-flow characteristics were related to existing CP classifications. The stochastic downscaling method was subsequently tested in the Upper Neckar Catchment in Southern Germany.

2.2 Method

2.2.1 Problem Formulation

The purpose of this paper is to find robust cause-effect relationships between a given runoff characteristic (i.e. one that accounts for a high-, or low-flow regime in a catchment) and a selected set of variables in given spatial and temporal domains. These relationships relate a given runoff characteristic with a set of explanatory variables such as: 1) physiographical factors, 2) shares of land cover types, and 3) climatic or meteorological factors. These variables are related to known physical processes that take place in a given catchment but are not the result of a physically-based analysis. The proposed method is consequently database driven and searches for significant signals within the available information. To apply this modeling approach, the following definitions are necessary.

Let Y_{il}^t denote the l runoff characteristics, related with low- or high-flows, observed in basin i during year t, and $f_l(\cdot)$ a nonlinear, differentiable, and monotonic function relating it to a set of explanatory variables $\{x_{i1}^t, \ldots, x_{ij}^t, \ldots, x_{in}^t\} \equiv \{\langle\mathbf{M}\rangle_i^t, \langle\mathbf{U}\rangle_i^t, \langle\mathbf{G}\rangle_i^t\}$. Here j denotes the variable index and n the total number of explanatory variables.

Table 2.1 European Circulation Patterns according to Hess and Brezowsky (1969).

Major Type	Sub-type	Index k	Description	Abbreviation
Zonal circulation	W	1	West, anticyclonic	Wa
		2	West, cyclonic	Wz
		3	Southern, West	WS
		4	Angleformed West	WW
Mixed circulation	SW	5	Southwest, anticyclonic	SWa
		6	Southwest, cyclonic	SWz
	NW	7	Northwest, anticyclonic	NWa
		8	Northwest, cyclonic	NWz
	HM	9	Central European high	HM
		10	Central European ridge	BM
	TM	11	Central European low	TM
Meridional circulation	N	12	North, anticyclonic	Na
		13	North, cyclonic	Nz
		14	North, Iceland high, anticyclonic	HNa
		15	North, Iceland high, cyclonic	HNz
		16	British Isles high	HB
		17	Central European trough	TRM
	NE	18	Northeast, anticyclonic	NEa
		19	Northeast, cyclonic	NEz
	E	20	Fennoscandian high, anticyclonic	HFa
		21	Fennoscandian high, cyclonic	HFz
		22	Norwegian Sea-Fennoscandian high, anticyclonic	HNFa
		23	Norwegian Sea-Fennoscandian high, cyclonic	HNFz
		24	Southeast, anticyclonic	SEa
		25	Southeast, cyclonic	SEz
	S	26	South, anticyclonic	Sa
		27	South, cyclonic	Sz
		28	British Isles low	TB
		29	Western Europe trough	TRW
Unclassified	U	30	Classification not possible	U

Each vector $\langle\mathbf{M}\rangle_i^t$, $\langle\mathbf{U}\rangle_i^t$, and $\langle\mathbf{G}\rangle_i^t$ is composed of variables that are evaluated at the mesoscale, which denote: 1) climatic or meteorological, 2) land cover, and 3) physiographical characteristics, respectively. The operator $\langle\cdot\rangle$ denotes a vector composed by either the integral or a spatial statistic obtained from a set of variables defined on a given spatial and temporal domain. Moreover, let $\boldsymbol{\beta}$ be a vector of parameters to be calibrated and validated based on historical records, and ε_{il}^t be an additive error term with zero mean but otherwise of undefined distribution. Based on these definitions, the general form of the relationship can be written as:

$$Y_{il}^t = f_l\big(\langle\mathbf{M}\rangle_i^t, \langle\mathbf{U}\rangle_i^t, \langle\mathbf{G}\rangle_i^t, \boldsymbol{\beta}\big) + \varepsilon_{il}^t \qquad \forall\, i = 1, \ldots, I \quad t = 1, \ldots, T \tag{2.1}$$

where I and T denote the total number of basins, and the total number of years of the calibration period respectively.

The variables required for eq. (2.1) are based on the best possible information describing the relevant basin's characteristics at the mesoscale. For example, one could use land cover classifications obtained from hyperspectral remote sensing data, physiographic characteristics of the basin derived from digital elevation models, soil types and their respective physical parameters together with geologic formations obtained from field campaigns and existing maps at the mesoscale, or time series for discharge and climatic variables. To account for macroclimatic changes one could use stochastic downscaling procedures to generate wetness and dryness indices that could be linked with high- and low flow regimes at the mesoscale. The derivation of such indices is described next.

2.2.2 Downscaling Circulation Patterns

In the present study, the subjective CP classification proposed by *Hess and Brezowsky* (1969) and two objective CP classifications developed by *Bárdossy and Filiz* (2005) were employed to downscale the macroclimatic state of the atmosphere as predictors for several mesoscale runoff characteristics.

The subjective classification is a synoptic meteorological classification defined for a large spatial domain (40° W, 30° N and 60° E, 80° N). This index is usually referred to as the European atmospheric circulation patterns and is currently used by the German Weather Service. This classification is based on mean air pressure distribution over Europe and the northern Atlantic Ocean and differentiates among three major circulation types called zonal, mixed, and meridional. These major types are, in turn, further subdivided according to the direction of movement of frontal zones, location of high and low pressure areas, and cyclonic and anti-cyclonic rotation. As a result, there are 29 plus one unclassified CPs as shown in Table 2.1.

The two objective classifications were optimized for a specific spatial domain and are specific for low- and high-flows. The fuzzy rule-based optimization was carried out with the method proposed by *Bárdossy et al.* (2002). The technique used for the classification of the high-flow related CPs was already reported in *Bárdossy and Filiz* (2005). A similar technique, with the exception of the objective function, was used to derive low-flow related CPs.

The purpose of the classification was to identify a selected number of classes so the members of these classes:

1. Behaved similarly with respect to rainfall and/or discharges. This means that the classes should be as homogeneous as possible.

2. Behaved differently from the mean. This means that the classification should enable a distinction with respect to the selected variable. A wet/dry situation has to be identified.

The classification was based on the large scale atmospheric variables such as the Sea Level Pressure (SLP) or the 500 hPa geopotential height anomaly obtained from the National Meteorological Center (NMC) gridpoint data set for different windows over Europe with a grid resolution of $5° \times 5°$. In this study the classification was performed by optimization using an objective function that evaluated the performance of the classification and the SLP data. In this study, two kinds of objective functions were considered.

For floods, positive increments of the discharges were used to define the objective function:

$$O_F(\mathbf{R}) = \frac{1}{TD} \sum_{t=1}^{T} \sum_{d=1}^{D} \left| \frac{\overline{z(\mathrm{CP}^{t(d)} = k)}}{\bar{z}} - 1 \right| \tag{2.2}$$

where $\bar{z}$ is the mean increase of the discharge on an arbitrary day at a given location i with $(z > 0)$. $\overline{z(\mathrm{CP}^{t(d)} = k)}$ is the mean increase of the discharge on days $t(d)$ with CP k ($\mathrm{CP}^{t(d)} = k$), and D the number of days of the year t. The rule system describing K CPs is represented by the matrix $\mathbf{R}$—for more details please refer to *Bárdossy and Filiz* (2005). This objective measures the relative performance of the classification compared to no-classification. The value of O_F is large if there are CP types which lead regularly to high increases of discharge and others which do not. This measure is thus related to flood peaks and flood volumes.

For droughts, this approach is not optimal, as small or no discharge changes are not necessarily related to low flows or droughts. Therefore rainfall data measured at different stations were used to define the objective function as:

$$O_D(\mathbf{R}) = \frac{1}{TDM} \sum_{t=1}^{T} \sum_{d=1}^{D} \sum_{m=1}^{M} \left| \ln \left(\frac{\overline{z_m(\mathrm{CP}^{t(d)} = k)}}{\bar{z_m}} \right) \right| \tag{2.3}$$

In this case $\bar{z_m}$ is the mean precipitation at location m and $\overline{z_m(\mathrm{CP}^{t(d)} = k)}$ is the mean precipitation at location m on days $t(d)$ with CP k ($\mathrm{CP}^{t(d)} = k$), and M is the number of measurement locations considered. The logarithmic transformation ensures that CPs with lower than normal precipitation are considered stronger than in the case of the linear type evaluation used in eq. 2.2.

Both classifications were calibrated and validated for a specific region and are composed of twelve categories ($K = 12$) ranging from CP1 to CP12 and one unclassified one called CP13. The procedure used in this study to downscale the proposed CPs is described next.

The *first step* was to cluster the CPs into two or three groups denoted as wet, normal and dry periods, using for this purpose a seasonal wetness index W_k (*Bárdossy*, 1993) estimated as follows:

$$W_k = \frac{\frac{1}{P} \sum_{t} \sum_{d \in S^t} p_k^{t(d)}}{\frac{1}{T} \sum_{t} \sum_{d \in S^t} \varphi_k^{t(d)}} \tag{2.4}$$

with

$$p_k^{t(d)} = \begin{cases} \langle p \rangle^{t(d)} & \text{if} \quad \mathrm{CP}^{t(d)} = k \ \wedge \ d \in S^t \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

$$\varphi_k^{t(d)} = \begin{cases} 1 & \text{if} \quad \mathrm{CP}^{t(d)} = k \,\wedge\, d \in S^t \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

$$P = \sum_t \sum_{d \in S^t} \langle p \rangle^{t(d)} \tag{2.7}$$

and where

- $\langle p \rangle^{t(d)}$ is the expected daily precipitation at the study area Ω. The operator $\langle \cdot \rangle$ denotes the integral of the daily precipitation p over the domain Ω occurring during the day d of the water year $t \in \{1, \ldots, T\}$.
- $\mathrm{CP}^{t(d)}$ is the atmospheric circulation pattern index according to Hess and Brezowsky for a given day d of the water year t.
- S^t is a set indicating whether a the day d belongs to summer or winter. S_t takes the values $\{1, \ldots, d_w\}$ and $\{d_w + 1, \ldots, D\}$ for winter and summer respectively. If t denotes a leap year, then $d_w = 182$ and $D = 366$; otherwise $d_w = 181$ and $D = 365$.
- P is the total summer or winter precipitation within Ω during the calibration period in millimeter.
- k is a CP-type index whose equivalence is either given in Table 2.1 or defined automatically by the classifier.

The wetness index defined in eq. (2.4) represents the ratio between the relative amount of precipitation in summer or winter occurring during those days with the CP-type k and the relative frequency of such CP. Using W_k, the CPs can be grouped into three categories by applying the following rules:

$$\text{if} \begin{cases} W_k \leq \tau_1 & \Rightarrow k \in \{\mathrm{Dry/Season}\} \\ \tau_1 < W_k \leq \tau_2 & \Rightarrow k \in \{\mathrm{Normal/Season}\} \\ W_k > \tau_2 & \Rightarrow k \in \{\mathrm{Wet/Season}\} \end{cases} \tag{2.8}$$

If only dry and wet periods—in any season—are required then the thresholds τ_1 and τ_2 must be equal.

The *second step* consisted of defining the following indices based on the previous classification, namely:

$$x_{ij}^t = \sum_{d=d_w+1}^{D} \vartheta_{ij}^{t(d)} \qquad j = 1, 2 \tag{2.9}$$

$$x_{ij}^t = \sum_{d=1}^{d_w} \vartheta_{ij}^{t(d)} \qquad j = 3 \tag{2.10}$$

with

$$\vartheta_{ij}^{t(d)} = \begin{cases} 1 & \text{if} \quad j = 1 \quad \wedge \begin{cases} x_{i4}^{t(d)} - x_{i4}^{t(d-1)} < 0 \\ \mathrm{CP}^{t(d)} \in \{\mathrm{Dry/Summer}\} \end{cases} \\ 1 & \text{if} \quad j = 2 \quad \wedge \begin{cases} x_{i4}^{t(d)} \geq \mathrm{F}_\gamma(x_4) \\ \mathrm{CP}^{t(d)} \in \{\mathrm{Wet/Summer}\} \end{cases} \\ 1 & \text{if} \quad j = 3 \quad \mathrm{CP}^{t(d)} \in \{\mathrm{Wet/Winter}\} \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

and

$$x_{i4}^{t(d)} = \sum_{c=0}^{C} (\upsilon)^c \langle p \rangle^{t(d-c)} \tag{2.12}$$

where

υ is the recession constant, commonly ranging within the interval $0.85 < \upsilon < 0.98$ (*Chow*, 1964).

$\mathrm{F}_\gamma(x)$ is a threshold value representing the γ-th percentile of variable x.

c is a time index denoting the precipitation occurred c days before the event $t(d)$. A common range for c is from 15 to 120 days.

The variable x_1 obtained with eq. (2.9) is aimed at finding a relationship between the occurrence of "dry" circulation patterns and low flows—i.e. $Y_i^{t(d)} \leq F_{10}(Y_i)$—occurring during summer. Hence, it tallies the total number of occurrences of CPs clustered as "dry periods" for a given catchment i during summer of a year t which have simultaneously a decreasing API (i.e. days where $\frac{d}{dt}x_4 < 0$).

Variables x_2 and x_3 in eq. (2.10), on the contrary, are intended to analyze relationships for peak flows—i.e. $Y_i^{t(d)} \geq F_{95}(Y_i)$. The former counts the number of days in summer on which both the occurrence of "wet" circulation periods and an antecedent precipitation index greater than a given threshold occur simultaneously. In other words, this index assumes that a flood may be expected if a certain climatic condition and a given amount of precipitation during a continuous period have already occurred. The latter tallies simply the number of occurrences of wet circulation periods during winter.

2.2.3 Modeling High- and Low-Flow Characteristics

Modeling extreme runoff characteristics—here denoted by the variable Y—is, in general, a challenging task because of their highly skewed probability density functions (PDF). Empirical studies carried out with extreme runoff characteristics have shown that the error term of multivariate models—represented by ε in eq. (2.1)—has a PDF other than the normal, or in other words, it is heteroscedastic (*Samaniego*, 2003). Consequently, standard methods such as multivariate linear or nonlinear regression generally fail to produce unbiased and acceptable fits if the common Gaussian assumption is not relaxed (i.e. the error term ε should be normally distributed with zero mean and a constant variance).

There are currently many possibilities to address this problem: 1) To introduce variable transformations as in *Montgomery and Peck* (1982) or to weight the residuals in the objective function according to their reliability as proposed in *Gentleman* (1974) and *Draper and Smith* (1981); 2) To use multivariate copulas to establish the dependence among the marginals of many random variables (*Favre et al.*, 2004; *Schweizer and Wolff*, 1981); or 3) To use a generalized model to deal with the variance of a given random variable in a proper manner (*Clarke*, 1994; *Lindsey*, 1999). In this paper, the latter approach was adopted.

In general, the structure of a generalized model is composed of three elements:

1. The deterministic element, also called the predictor, η_i^t, which is a suitable function of the explanatory variables $\mathbf{x}_i^t$, thus
$$\eta_{il}^t = f_l(\mathbf{x}_i^t, \boldsymbol{\beta}) \tag{2.13}$$

2. The distributional element, which indicates that the variance of the response Y_{ij}^t is an explicit function of the expectation at each observation denoted by μ_{il}^t, hence
$$\mathrm{E}[Y_{il}^t] = \mu_{il}^t \tag{2.14}$$
$$\mathrm{var}[Y_{il}^t] = \kappa V(\mu_{il}^t) \tag{2.15}$$
and that Y_{il}^t is h distributed with distributional parameters $\boldsymbol{\alpha}$.

3. The link function $g(\cdot)$, which is a nonlinear, monotone, and differentiable function that establishes a *link* between the deterministic and the stochastic part of the model, so that
$$g(\mu_{il}^t) = \eta_{il}^t \tag{2.16}$$

where $\mathbf{x}_i^t$ is a vector of observed variables at the basin i and point in time t, κ denotes a dispersion parameter, $\boldsymbol{\alpha}$ are the fitted parameters of the probability density function $h(\cdot)$, and $V(\cdot)$ is a variance function.

This formulation differs from the standard Generalized Linear Model (GLM) in one respect only, namely: that the function $f_l(\cdot)$ can be assumed either linear or nonlinear. By doing so, there is more flexibility to handle the intertwined relationships between the predictors and the predictand.

2.2.4 Model Selection and Parameter Estimation

The model selection was done in two steps: 1) a priori selection of a functional form for the eq. (2.1) according to a parsimonious criterion and/or available knowledge on the relationships between some of the related variables, and 2) the selection of the most robust model once a functional relationship is given.

In the present study, once a nonlinear relationship was selected, a constrained multi-objective optimization problem—whose solution space comprised all feasible combinations of given explanatory variables—was then solved. By definition, a feasible combination consisted only of statistically significative variables in which each subcategory must have at least one variable, i.e. $\langle \mathbf{M} \rangle_i^t$, $\langle \mathbf{U} \rangle_i^t$, $\langle \mathbf{G} \rangle_i^t \neq \emptyset$. Here, the significance of each variable was assessed by a nonparametric test (*Efron*, 1982).

Among all competing models, the most robust model is that which is least sensitive to the selection of the estimator. In the present study two estimators were used for the selection of the most robust model structure (i.e. variable composition), namely: L_1 and L_2. The former denotes the sum of the absolute residuals whereas the latter denotes the sum of the squares of residuals. These two estimators were selected a priori. Other possibilities are conceivable too, e.g. two different likelihood functions. Consequently, the best model should simultaneously surpass the others in two predefined objectives:

$$\Phi_o = \sum_{i=1}^{I} \sum_{t=1}^{T} \left(Y_{il}^t - \widetilde{Y}_{il}^t(L_o) \right)^2 \quad o = 1, 2 \quad \forall\, l \tag{2.17}$$

where Φ_o is a jackknife statistic (*Quenouille*, 1949) found by minimizing the estimator L_o with $o = 1, 2$. For the estimation of each $\widetilde{Y}_{il}^t$, the vector $\mathbf{x}_i^t$ was sequentially excluded from the data set. Subsequently, model parameters were calibrated with the rest of the sample and finally $\widetilde{Y}_{il}^t$ was evaluated for $\mathbf{x}_i^t$. This procedure was repeated $\forall\ i, t$. To improve the performance of the optimization algorithm and ease the comparison between the objective functions all variables (i.e. both the input and output ones) were transformed (scaled) to the interval $(0, 1]$. The jackknife statistics were not normalized as shown in eq. 2.17.

Once both the functional form and the model structure were selected, the estimation of model parameters $\boldsymbol{\beta}$ was carried out by maximizing the log-likelihood function $\ell(\cdot)$ whose general form for the explained variable Y_{il}^t is

$$\max_{\hat{\boldsymbol{\beta}}}\ \ell(\boldsymbol{\beta}) = \sum_{i,t} \ln h(Y_{il}^t | \mathbf{x}_i^t, \boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{2.18}$$

and, its corresponding goodness of fit was assessed by the Akaike's Information Criterion AIC (*Akaike*, 1973a) as follows

$$\mathrm{AIC} = -2\, \ell(\hat{\boldsymbol{\beta}}) + 2p^* \tag{2.19}$$

where $h(Y_{il}^t | \mathbf{x}_i^t, \boldsymbol{\alpha}, \boldsymbol{\beta})$ denotes the probability density function (PDF) of Y_{il}^t assuming that all observations are independent, and p^* is the number of parameters used to define the model $f_l(\cdot)$. It should be noted that minimizing the estimator L_2 implies that the explained variable and the residuals are assumed normal distributed. The maximum likelihood method was used here to introduce a given distributional element. For more details on this approach please refer to *Samaniego* (2003) and *Samaniego and Bárdossy* (2005).

2.3 Application

2.3.1 The Study Area and Information Available

The proposed method was tested in the upper catchment of the Neckar River upstream of the Plochingen gauging station covering an area of approximately 4000 km^2 (Fig. 2.1). The data concerning this Study Area were obtained from several sources as indicated in Table 2.2. All variables described in eq. 2.1 were integrated both in space, at subcatchment level, and in time, at semiannual intervals. The spatial domain of each subcatchment comprised the drainage area of 46 gauging stations ($I = 46$) located in the Study Area, whose area ranged from 4 km^2 to 4000 km^2. With respect to the time domain, the gauged discharge was aggregated from event scale (day) during the period from 1960.11.01 to 1993.10.31, into semiannual intervals termed summer or winter to minimize the existing autocorrelation, hence $T = 33$. To validate these data, an annual water balance for each subcatchment was calculated and all those observations exceeding a threshold value were excluded as outliers (*Samaniego*, 2003).

Every predictor and explained variable listed in Table 2.2 were estimated for each subcatchment and time interval based on data provided by several German State Agencies, as is indicated in the same Table.

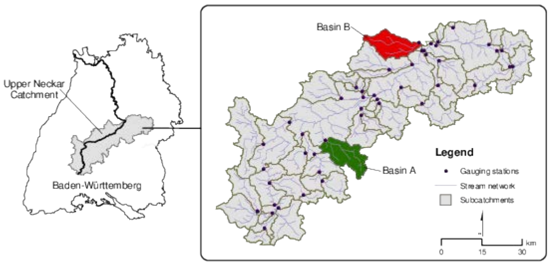

Figure 2.1 Location of the Study Area within the State of Baden-Württemberg, Germany. The location of subcatchments A and B are also displayed.

2.3.2 A Priori Structures of the Model

In the present study, four types of functional relationships $f_l(\cdot)$ were selected to represent presumed relationships between the explanatory variables and the predictor, therefore, they should be seen as *hypotheses* rather than *models*. The configuration of the predictor was based on criteria such as simplicity of model structure and parsimony of parameters. They, however, should help to reveal the kind of linkage that may exist among the system's processes represented here by the observables listed in Table 2.2. Consequently, four combinations of variables mixing linear and/or nonlinear functions were investigated, namely: 1) a multilinear (ML), 2) a potential (POT), and 3) two multilinear-potential (MLP1, MLP2) relationships. Formally, these relationships can be written as

$$\eta_{il}^t = \beta_0 + \sum_j \beta_j x_{ij}^t \tag{2.20}$$

$$\eta_{il}^t = \beta_0 \prod_j \left(x_{ij}^t\right)^{\beta_j} \tag{2.21}$$

$$\eta_{il}^t = \beta_0 + \sum_{j \in \mathbf{G} \cup \mathbf{U}} \beta_j x_{ij}^t + \beta_{J^*} \prod_{j \in \mathbf{M}} \left(x_{ij}^t\right)^{\beta_j} \tag{2.22}$$

and

$$\eta_{il}^t = \beta_0 + \sum_{j \in \mathbf{U}} \beta_j x_{ij}^t + \beta_{J^*} \prod_{j \notin \mathbf{U}} \left(x_{ij}^t\right)^{\beta_j} \tag{2.23}$$

Furthermore, to fulfil the model requirements described in section 2.2.3, these predictors were combined with several variance and link function alternatives shown in Tables 2.3 and 2.4 respectively.

Table 2.2 Explained variables, selected predictors, and data sources used in this study.

Source	Variable	Description
Time series of mean daily flows from 1961-1993 (Institute for Environmental Protection Baden-Württemberg, LfU)	Y_1	Specific volume of high flows in winter (mm)
	Y_2	Specific volume of high flows in summer (mm)
	Y_3	Total duration of high flows in winter (day)
	Y_4	Total duration of high flows in summer (day)
	Y_5	Frequency of high flows in winter (year^{-1})
	Y_6	Frequency of high flows in summer (year^{-1})
	Y_7	Total drought duration in summer (day)
	Y_8	Cumulative specific deficit in summer (mm)
European Circulation Patterns according to Hess and Brezowsky, and precipitation time series (German Meteorological Service, DWD)	x_1	A dryness index tallying the total No. of "dry periods" with decreasing API in summer (day)
	x_2	Total number of "wet periods" occurring, simultaneously with an API greater than a given threshold in summer (day)
	x_3	Total number of "wet periods" in winter (day)
Time series of daily precipitation and temperature for 288 meteorological stations in Baden-Württemberg from 1961 to 1993 (LfU and DWD). Each day was interpolated with External Drift Kriging to a spatial resolution of 300 m × 300 m	x_4	Antecedent precipitation index (API) (mm)
	x_5	Cumulative winter precipitation (mm)
	x_6	Cumulative summer precipitation (mm)
	x_7	Mean winter precipitation (mm)
	x_8	Mean summer precipitation (mm)
	x_9	Mean temperature in January (K)
	x_{10}	Mean temperature in July (K)
	x_{11}	Maximum temperature in January (K)
Topographic map 1:25 000 for 1961 (State Surveying Agency Baden-Württemberg, LVA) and LANDSAT scenes from 1975, 1984, and 1993	x_{12}	Fraction of forest cover (-)
	x_{13}	Fraction of impervious cover (-)
	x_{14}	Fraction of permeable cover (-)
DEM 30 m × 30 m (LVA)	x_{15}	Area of a given catchment (km^2)
	x_{16}	Trimmed mean slope $F_{(15)}$-$F_{(85)}$ $(^\circ)$
	x_{17}	Mean slope in floodplains $(^\circ)$
	x_{18}	Drainage density $(1/km)$
	x_{19}	Fraction of north-facing slopes (-)
	x_{20}	Mean elevation of the catchment (m)
	x_{21}	Difference between max. and min. elevation (m)
	x_{22}	Fraction of saturated areas (-)
Soil map 1:200 000 and Geological map 1:600 000 (LfU)	x_{23}	Mean field capacity (mm)
	x_{24}	Fraction of karstic formations (-)

Table 2.3 Probability distributions employed in this study and their relevant characteristics.

Distribution $Y_{il}^t \sim h(\boldsymbol{\alpha})$	Expectation $E[Y_{il}^t]$	Variance Function $V(\mu_{il}^t)$	Dispersion Parameter κ
Normal $\mathcal{N}(\mu_{il}^t, \sigma)$	μ_{il}^t	1	σ^2
Poisson $\mathcal{P}(\mu_{il}^t)$	μ_{il}^t	μ_{il}^t	1
Gamma $\mathcal{G}(a, b_{il}^t)$	$\mu_{il}^t = ab_{il}^t$	$(\mu_{il}^t)^2$	a^{-1}
Weibull $\mathcal{W}(a, b_{il}^t)$	$\mu_{il}^t = b_{il}^t \Gamma(1 + a^{-1})$	$(\mu_{il}^t)^2$	$\frac{\Gamma(1+2a^{-1})}{\Gamma^2(1+a^{-1})} - 1$

Note: a and b_{il}^t denote the shape and the scale parameters of the PDF, respectively, and $\Gamma(\cdot)$ is the gamma function.

Table 2.4 Link functions used in this study.

Name	Function $\mu_{il}^t = g^{-1}(\cdot)$
Identity	η_{il}^t
Logit	$\frac{1}{1+\exp(\eta_{il}^t)}$
Log	$\exp(\eta_{il}^t)$
Reciprocal	$(\eta_{il}^t)^{-1}$

2.4 Results and Discussion

2.4.1 CP Classifications

Based on the hydrological data available for the Study Area, the subjective and objective atmospheric circulation patterns were classified into two or three categories using parameters τ_1 and τ_2 as shown in Table 2.5. The adopted values for the recession constant υ were 0.95 and 0.85 for winter and summer respectively. Additional parameters required for this classification were the percentile level $\gamma = 0.80$, and the maximum range of c, whose adopted values for winter and summer were $c = 30$ and $c = 120$ days respectively. It is worth noting that according to the classification rules proposed for the subjective classification, most of the dry-CPs in this region are anticyclonic. This result confirms the adequateness of the adopted classification parameters since warmer and drier air masses in an anticyclone tend to suppress convective precipitation, which, in turn, leads to a reduction of the relative humidity. Most of the wet-CPs are, on the contrary, cyclonic.

Table 2.5 Classification of circulation patterns (CPs) for winter and summer seasons according to the wetness index W_j in the Study Area during the period from 1961 to 1993. The number of CPs and their total frequency of occurrence in percentage (ρ) for each category are also indicated.

Classification Type	τ_1	τ_2	Category	CP Winter	No.	ρ [%]	CP Summer	No.	ρ [%]
Subjective	0.6	1.0	Dry	BM, HB, HFa, HM, HNa, HNFa, NEa, NWa, Sa, SEa, SEz, SWa, Wa	13	41	BM, HB, HFa, HM, HNa, NEa, NWa, Sa, SEa, SWa, Wa	11	44
			Normal	HNFz, NEz, Sz, TB	4	6	HNFa, Na, Sz, U	4	4
			Wet	HFz, HNz, Na, NWz, Nz, SWz, TM, TRM, TRW, U, Ws, WW, Wz	13	53	HFz, HNFz, HNz, NEz, NWz, Nz, SEz, SWz, TB, TM, TRM, TRW, WS, WW, Wz	15	52
Objective high-flows	1.1	1.1	Dry	CP1, CP3, CP4, CP7, CP9, CP12,	6	61	CP1, CP3, CP4, CP6, CP7, CP9, CP12	7	69
			Wet	CP2, CP5, CP6, CP8, CP10, CP11, CP13	7	39	CP2, CP5, CP8, CP10, CP11, CP13	6	31
Objective low-flows	0.8	0.8	Dry	-			CP3, CP5, CP8, CP11, CP13	5	35
			Wet	-			CP1, CP2, CP4, CP6, CP7, CP9, CP10, CP12	8	65

2.4.2 Visualizing Possible Effects of Land Cover Change on Hydrological Extremes

Two subcatchments labeled A and B (Fig. 2.1) were selected to illustrate the evolution of some explanatory variables over time and the possible effects of land use/cover change. These subbasins have approximately the same extent (123 km^2 and 126 km^2 respectively) but differ largely on their main land cover type. Their average elevation above sea level is 630 m and 385 m respectively.

During the period from 1961 to 1993, subcatchment A exhibited slightly growing shares of forest and impervious cover whereas subcatchment B endured a steady land use transition from grassland (permeable land cover) to settlement (impervious land cover) and a steady decline of forest since the mid 70s. As a result of that, in 1993, the former was mainly covered by forest (48%) whereas the latter was extensively urbanized (56%) and hence mainly covered by impervious surfaces.

The behavior of these two subcatchments is depicted on the left and right panel of Fig. 2.2, respectively. Total drought duration in summer (Y_7), for instance, exhibited a remarkable decrement in the largely urbanized catchment although the total precipitation in the same half year as well as the mean air temperature in July did not reveal any significant trend. Moreover, the index x_1 showed a completely different behavior depending on the shares of land cover. This fact suggests that the combination of a rapid growth of impervious cover accompanied by a decline of forest may have led to a rapid shrink of the total drought durations. In contrast, slightly growing shares of forest and impervious areas may have led to an increase of total drought durations. Moreover, for a given meteorological drought, the effect on the runoff characteristic varied considerably. These facts are illustrated in Fig. 2.2 by the rectangles drawn with a dashed line. The frequency of high flows in winter (Y_5) and the index x_3 also revealed different behavior depending on the main land cover type. Subcatchment B, for instance, exhibited a clear positive trend whereas subcatchment A did not. A more detailed analysis of probable causal relationships is presented in section 2.4.3.

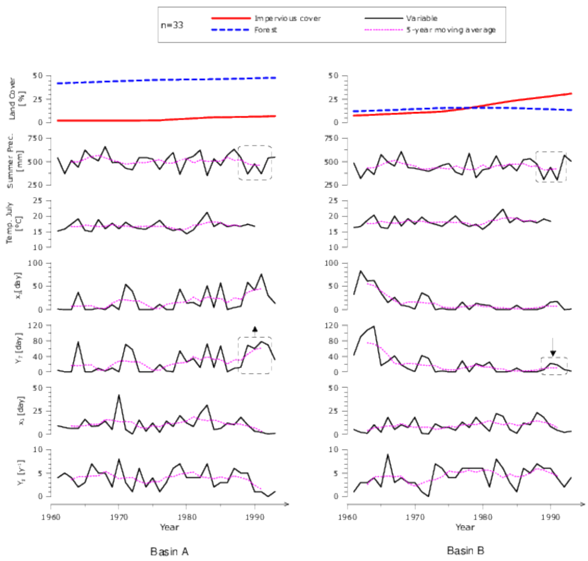

Figure 2.2 Time series and trends (by means of a 5-year running average) of selected variables, namely: the frequency of high flows in winter (Y_5), the total drought duration in summer (Y_7), indices x_1 and x_3 based on the subjective CP classification, summer precipitation (x_6), mean temperature in July (x_{10}), and fractions of land cover (x_{12} and x_{13}) for two subcatchments (A and B) depicted in Fig. 2.1. Additionally, it is also shown how the same climatic phenomenon (i.e. a sequence of dry years highlighted by a dashed line) have produced different outcomes depending on the land cover and morphological situation within the catchment.

Consequently, these observations seemed to indicate that although the explained variables were mainly governed by macroclimatic conditions, they might be significantly attenuated or enhanced by the land cover conditions as well as by morphological characteristics of the catchment.

2.4.3 Model Parametrization and Performance

The most robust models were selected by the method described in section 2.2.4. The composition of these models and their corresponding calibrated parameters are shown in Tables 2.6 and 2.7 respectively. Models for the runoff characteristics Y_1, Y_2, and Y_8 were not estimated because they are highly correlated with characteristics Y_3, Y_4, and Y_7, whose pairwise Pearson's correlation coefficients were 0.87, 0.89, and 0.71 respectively. For the remaining runoff characteristics, the best results were obtained with the identity link function, but, in any case, the predictor function appears to be multilinear.

To select among the several functional relationships considered in this study (MLP1, MLP2 and POT), a scatterplot (Fig. 2.3) of the performance indices Φ_1 and Φ_2 was used.

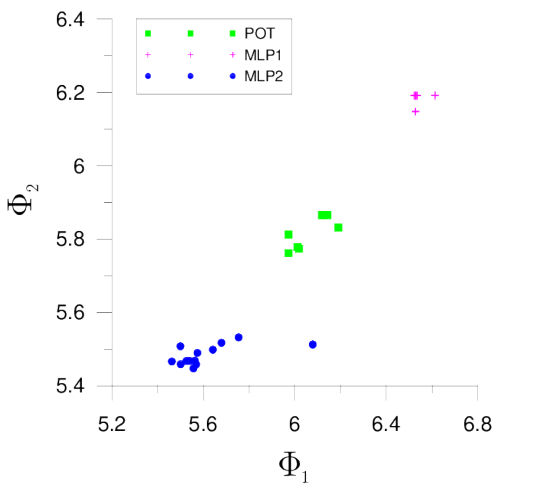

Figure 2.3 Performance of three models types (MLP1, MLP2 and POT) used to predict total drought duration in summer (Y_7). Φ_1 and Φ_2 are Jackknife statistics calculated by minimizing the L_1 and L_2 estimators respectively.

Since these performance indices denote Jackknife statistics-errors, the smaller the performance index, the higher the robustness of the model fit. In the example shown in Fig. 2.3, for instance, model type MLP2 was better than POT, and this, in turn, was better than model type MLP1. The performance indices for the ML type model do not appear in this graph because their corresponding values were greater than the maximum values shown in this graph. Similar results were obtained for the other characteristics. This finding indicated that the nonlinearity of the described system stemmed mainly from the meteorological and/or physiographic variables.

Table 2.6 Summary of the composition of the calibrated hydrological models (1 denotes that a variable is included in the model, otherwise it is omitted). Additionally, the sample size N as well as the type of predictor and distribution functions employed in each model are shown.

Runoff Characteristic	$\eta(\cdot)$	$h(\cdot)$	Variable x_i (i): 1	2	3	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	N
Y_3	POT	$\mathcal{N}$			1			1			1													1312
Y_4	MLP2	$\mathcal{N}$		1					1			1												1318
Y_5	POT	$\mathcal{P}$			1	1				1	1		1		1			1				1	1	1247
Y_6	MLP2	$\mathcal{W}$		1			1					1				1	1		1		1		1	977
Y_7	MLP2	$\mathcal{W}$	1								1			1		1	1			1				1263

Table 2.7 Parameter estimates and their respective p-values for each of the calibrated hydrological models (the climatic indices are based on the subjective classification).

Runoff Char.	Parameter β_i (i) [p-value]: 0	J^*	1	2	3	5	8	9	10	11	12	13	14
Y_3	1.419				0.951 $[\approx 0.]$			-0.116 $[\approx 0.]$			0.059 $[\approx 0.]$		
Y_4	0.889	3.365		1.143 $[\approx 0.]$					-0.502 [0.024]			0.105 $[\approx 0.]$	
Y_5	0.024				0.160 $[\approx 0.]$	0.761 $[\approx 0.]$				0.058 [0.018]	-0.131 $[\approx 0.]$		-0.166 $[\approx 0.]$
Y_6	0.100	4.923		0.454 $[\approx 0.]$			1.098 $[\approx 0.]$					0.012 [0.032]	
Y_7	-0.269	14.658	0.869 $[\approx 0.]$								0.017 [0.045]		

Runoff Char.	Parameter β_i (i) [p-value]: 15	16	17	18	19	20	21	22	23	24
Y_3										
Y_4										
Y_5		-0.637 $[\approx 0.]$			0.556 $[\approx 0.]$				0.225 [0.015]	-0.005 $[\approx 0.]$
Y_6			-0.825 [0.010]	-1.136 [0.004]		-0.289 [0.002]		0.486 [0.064]		-0.006 [0.024]
Y_7	-0.075 $[\approx 0.]$		-0.711 $[\approx 0.]$	-2.126 $[\approx 0.]$			0.236 [0.016]			

The model summary presented in Table 2.6 shows that with the exception of Y_3 and Y_4, the remaining runoff characteristics exhibited a distributional element other than normal. This means that the runoff characteristics Y_5, Y_6, and Y_7 as well as their associated additive error term (eq. 2.1) are heteroscedastic, or in other words, that they exhibit a non-stationarity variance either along time, or along some of the predictors, or both.

The goodness of the fit (r), bias, and the root of mean square error obtained for each calibrated model using both the subjective and objective CP classifications are summarized in Table 2.8. In general, these results indicated that the objective classification was as efficient as the synoptic one proposed by *Hess and Brezowsky* (1969). In the case of the low-flow characteristic (Y_7), the results obtained for both the r and the RMSE were significantly better than those obtained with the subjective classification. This important result showed an additional advantage of the objective classification, namely: that the learning (optimization) algorithm employed to find the fuzzy-rule classification system (see *Bárdossy et al.*, 2002) can be fine-tuned for a specific purpose, say predicting drought events at the mesoscale.

From all modeled characteristics, total drought duration in summer Y_7 exhibited the highest RMSE (13 days and 9 days depending of the classification used), which can be seen as an overall indicator of uncertainty associated with the estimation of this variable. In this case, the use of a "fine-tuned" objective CP classification (i.e for low-flows) led to a 30% reduction in the uncertainty of the predicted variable.

By inspection of Table 2.7 some hydrological interpretation of the empirical models is presented. Results obtained for total duration of high flows in winter and summer (Y_3 and Y_4 respectively) indicated that these runoff characteristics have a very strong correlation with the macroclimatic situation represented by the variables x_3 and x_9 in winter and x_2 and x_{10} in summer respectively. Morphological variables were irrelevant in these cases but land cover variables (x_{12} and x_9) were found to be statistically dependent and significant, although their contribution to the total explained variance was quite small. In both cases, mean temperature was inversely related to the total duration of high flows, which is plausible since the higher the temperature, the larger the evapotranspiration, and thus the lower the discharge.

Table 2.8 Correlation coefficient (r), bias (BIAS), and root of mean square error (RMSE) and the Akaike Information Criterium (AIC) obtained for the calibration period from 1961 to 1993 with yearly time steps.

Runoff	Subjective CP Classification				Objective CP classifications			
Characteristic	BIAS	RMSE	r	AIC	BIAS	RMSE	r	AIC
Y_3	-0.1	3.3	0.94	2919.	-0.2	3.9	0.91	3266.
Y_4	$+0.0$	2.3	0.94	1538.	$+0.0$	2.8	0.88	2403.
Y_5	$+0.0$	1.5	0.77	4327.	-0.0	1.5	0.76	4330.
Y_6	$+0.2$	1.1	0.87	1372.	-0.1	1.3	0.79	1693.
Y_7	$+0.0$	13.3	0.86	8751.	-0.3	9.3	0.92	7021.

The frequency of high flows (Y_5) in winter appeared as directly dependent on the meteorological conditions, especially the total precipitation (x_5), the maximum temperature in January (x_{11}) and the composed indicator of wet circulation patterns (x_3). Thus, the wetter a given year was, the higher the likelihood that a flood event would arise. Moreover, if the winter temperature in January was high, then the frequency of high flows tended to increase since the thawing process would be faster, which, in turn, may have increased discharge. The same direct relationship applied to the fraction of north-facing slopes (x_{11}) and the average field capacity (x_{23}). The former could be explained due to a reduced evapotranspiration and hence higher runoff (i.e. the conservation of mass principle that is implicit in the water balance of a basin). The shares of forest and permeable areas (x_{12} and x_{14} respectively) are inversely related variables, all perfectly plausible relationships from the hydrological point of view because land cover types with larger roughness coefficients tended to increase the time of concentration within a basin and hence reduce the likelihood of peak events. Causal relationships for the other variables were not quite clear, which indicates that they may represent artifacts found in the sample. During summer, the frequency of high flows (Y_6) also exhibited a direct relationship with the mean precipitation (x_8) and the composed index for wet circulation patterns (x_1), the share of saturated areas (x_5), and the share of impervious areas within a catchment (x_{13}).

Based on the data available, the total drought duration in summer (Y_6) appeared to have a complex nonlinear relationship with the predictors. For instance, it exhibited a nonlinear relationship with the macroclimatic conditions represented by the variable (x_1), which not only depends on the circulation patterns but also on the antecedent precipitation index. The latter, which is a proxy of the average soil moisture, was, in turn, directly related to the share of forest within a catchment (x_{12}) and inversely related to the share of impervious areas (x_{13}). Furthermore, x_{12} appeared as a linear predictor of Y_6 too. These relationships are also consistent with the catchment's water balance, as during a summer dry spell, the evapotranspiration from forest reaches the higher annual values, which, implies no recharge of the subsurface storage occurs, and, therefore, there should be less discharge into streams. The area of the basin (x_{15}) appeared to be inversely related to Y_6. In other words, the larger the basin's area, the larger its storage capacity, and hence, the longer the basin's baseflow can be kept above a given threshold. This, in turn, implies fewer days per year accounting for water deficits.

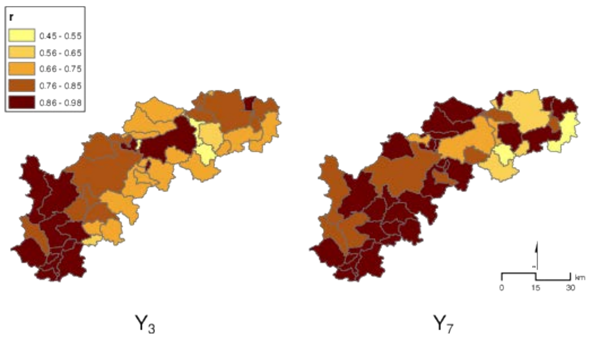

Figure 2.4 Spatial distribution of the correlation coefficient r of the total duration of high flows in winter (Y_3, left panel), and the total drought duration in summer (Y_7, right panel). In the particular case of basins A and B, r were 0.71 and 0.73 for Y_3; and 0.98 and 0.92 for the Y_7, respectively.

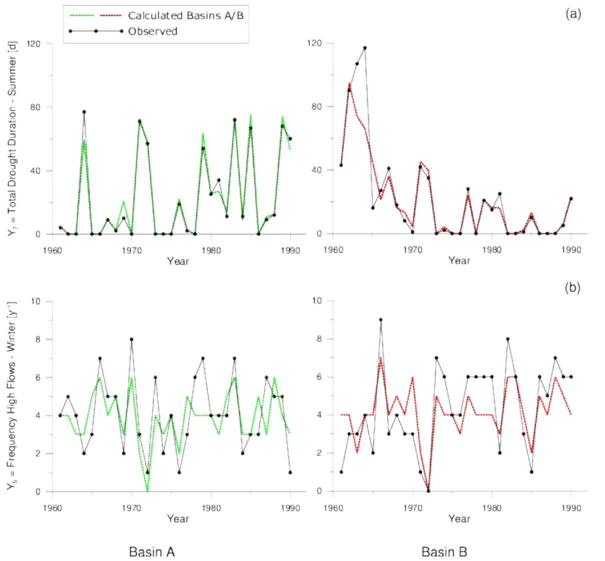

Figure 2.5 Observed and calculated time series for the total drought duration in summer [(Y_7), panel (a)] and the frequency of high flows in winter [Y_5, panel (b)] for basins A and B, respectively.

Based on the results shown in Table 2.8 and Fig. 2.5, it can also be inferred that the selected models perform relatively well because they were able to explain between 59% and 88% of the observed variance. This efficiency of the models, however, was not constant over space but exhibited considerable spatial-temporal variability within the studied area. For example, variability in Y_3 and Y_7 (Fig. 2.4), depended not only on the season (i.e. winter and summer) but also on the kind of hydrological regime (i.e. high- and low-flows) that was analyzed. In the present case, the lack of predictability seems to be associated with morphological characteristics such as the fraction of karstic formations underneath a given basin and its status as a headwater basin. In general, headwater subbasins performed better than those which are not. This is probably related with the number regulation structures (e.g. weirs and barrages) located along the courses of non-headwater basins. Moreover, subcatchments with lower fraction of karstic formations performed better than those with larger fractions of karstic fractions. The reason for that could be the unaccounted sources of water caused by size differences between surface and underground catchments, which, in turn, might alter the water budget accounting (particularly with respect to the baseflow) at a given gauging location.

2.4.4 Effects of the Distributional Element on the Simulated Variance

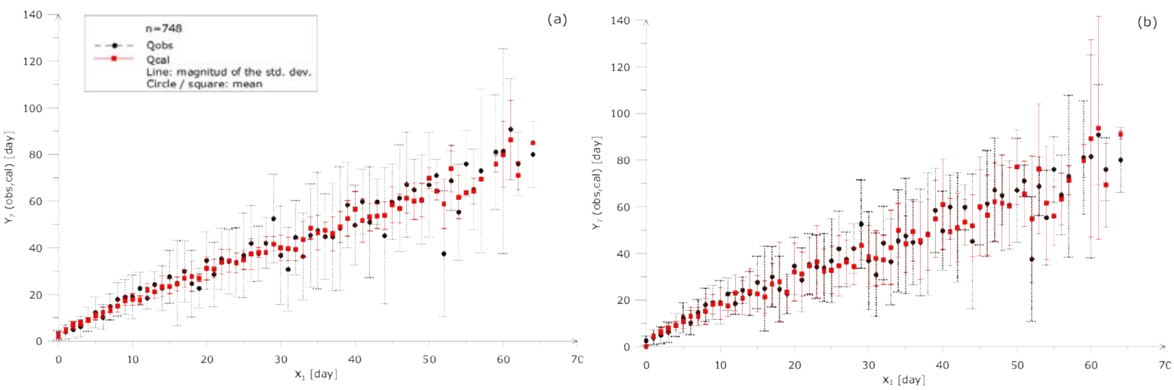

Figure 2.6 Variation of the dispersion of the observed and calculated total drought duration in summer (Y_7) as a function of the index x_1. The only difference between panels (a) and (b) is the kind of distribution function assumed for the explained variable, and thus, the error term: On the panel (a), Y_7 is assumed normal distributed, whereas on the panel (b) the variable is assumed Weibull distributed. Both continued and dashed lines represent the magnitude of the standard deviation whereas dots and rectangles represent the mean values at each level of predictor.

The correct selection of the distributional element when modeling highly heteroscedastic time series is a deciding factor on the model's quality, especially regarding the variance of the explained variable. This is clearly illustrated in Fig. 2.6, where the the total drought duration was modeled using two distributional elements as null hypothesis namely: 1) the normal distribution which is symmetrical and has a constant mean and variance; and 2) a skewed distribution like the Weibull which has a variance that depends on the value of the mean. For the sake of comparison, in both cases the a priori adopted function of the predictor, the sample size, and the explanatory variables were kept equal. Based on the results depicted in Fig. 2.6, it can be concluded that, under these assumptions, both models were able to predict relatively well the expected value of Y_7; but that the model based on the normal distribution completely underestimates the observed variability of Y_7 which largely depends on the index x_1. The model based on the Weibull distribution, on the contrary, exhibited a significant increase of the estimated variance along the abscissa, and, in this respect, performed much better than the former.

2.5 Conclusions

Based on the results of this study, some general conclusions can be drawn:

- The indices derived from either the objective or the subjective CP classifications were equally effective at explaining a large proportion of the variance of the proposed runoff characteristics. Consequently, objective classifications could become an efficient alternative in those cases where automated operational forecasts are required,
- The proposed dryness index (dry circulation patterns occurring with a decreasing antecedent precipitation index) exhibits a completely different behavior depending on the fraction of land cover within a catchment while almost constant seasonal-mean precipitation and temperature have been observed. Moreover, the dryness index is highly correlated with total drought duration,
- The proposed compounded wetness indexes also explain a large fraction of the observed variance of the total duration and the frequency of high-flows in winter and summer,
- The fraction of forest and/or impervious areas seem to have played—in the Study Area—a significative role in the hydrological extremes at mesoscale level,
- The large fraction of the observed variance of all extreme runoff characteristics is, however, explained by indices related to macroclimatic indices, and,
- The heteroscedasticity of the observed variables could only be acceptably simulated using nonlinear generalized models that assume non-gaussian distributions.

Acknowledgement

The authors appreciated the helpful comments from Prof. E. Zehe, Dr. J. Calabrese, and one anonymous reviewer.

CHAPTER 3

MULTISCALE PARAMETER REGIONALIZATION OF A GRID-BASED HYDROLOGIC MODEL AT THE MESOSCALE

L. Samaniego[1], R. Kumar[1], and S. Attinger[1]

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Samaniego, L., R. Kumar, and S. Attinger (2010a), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46(5), w05523

Abstract

The requirements for hydrological models have increased considerably during the previous decades to cope with the resolution of extensive remotely sensed datasets and a number of demanding applications. Existing models exhibit deficiencies such as overparametrization, the lack of an effective technique to integrate the spatial heterogeneity of physiographic characteristics, and the non-transferability of parameters across scales and locations. A Multiscale Parameter Regionalization (MPR) technique is proposed as a way to address these issues simultaneously. Using this technique, parameters at a coarser scale, in which the dominant hydrological processes are represented, are linked with their corresponding ones at a finer resolution in which input datasets are available. The linkage is done with upscaling operators such as the harmonic mean, among others. Parameters at the finer scale are regionalized through nonlinear transfer functions which link basin predictors with global parameters to be determined through calibration. MPR was compared with a standard regionalization (SR) method in which basin predictors instead of model parameters are firstly aggregated. Both methods were tested in a basin located in Germany using a distributed hydrologic model. Results indicate that MPR is superior to SR in many respects, especially if global parameters are transferred from coarser to finer scales. Furthermore, MPR, as opposed to SR, preserves the spatial variability of state variables and conserves the mass balance with respect to a control scale. Cross-validation tests indicate that the transferability of the global parameters to ungauged locations is possible.

3.1 Introduction

Hydrologic models have evolved in the previous decades to cope with the extensive data sets derived from Geographic Information Systems (GIS) and a plethora of remote sensing acquisition techniques. Rapidly increasing computational power has also contributed to this development. As a result, hydrologic models evolved from simple conceptual hydrologic models (e.g. SAC model (*Burnash et al.*, 1973a), HBV model (*Bergström*, 1995)) to spatially distributed models of varying complexities such as MIKE-SHE (*Refsgaard and Storm*, 1995), VIC-3L (*Wood et al.*, 1997), HL-RMS (*Koren et al.*, 2004), WASIM-ETH (*Schulla and Jasper*, 2007).

This development process has not always contributed to finding the "right" answers to the main problems of contemporary hydrology, particularly at the mesoscale, i.e. basins whose area are within the range of $[10^2, 10^4]\mathrm{km}^2$ (*Dooge*, 1986). These problems, according to *Beven* (2001) are: nonlinearity, scale, uniqueness, equifinality, and uncertainty.

Recent developments, especially from the so-called "physically based" hydrologic models have not properly addressed these problems, in most practical cases, mainly due to the following reasons. Firstly, media properties (both vegetation and soil) are essentially unknown or at least poorly known (*Blöschl et al.*, 2008), this implies that any system's characteristic will always exhibit some spatial variability regardless of the grid cell resolution chosen for modeling purposes. Hence, trying to use point scale physics at the basin scale implies that both the media and the boundary conditions should be known spatially at the scale of the equations. This, in turn, requires prohibitive amounts of input data, which in most cases goes far beyond practical limitations even for small experimental plots (*Zehe et al.*, 2006). Secondly, modelers are forced to find "effective parameters" via calibration, since the required information seldom exists. As a result, these types of models are transmuted into "overparameterized conceptual models" (*Beven*, 2001; *Blöschl*, 2001; *Kirchner*, 2006).

"Conceptual models", on the contrary, are parsimonious and computationally efficient than the former. They have been commonly applied in operational hydrology, due to these reasons. These models have been mainly focused on discharge generation and its predictive uncertainty, but have paid little attention to important issues like the spatio-temporal distribution of inputs, state variables, and water fluxes, which are fundamental for coupling them with regional climate models (*Seneviratne and Stöckli*, 2008). The application of these models usually requires parameter calibration.

In general, parameters of any mesoscale hydrologic model are by definition effective quantities that can not be measured but need to be inferred by an indirect procedure usually called calibration (*Beven*, 2001; *Gupta et al.*, 2002; *Kirchner*, 2006).

Model overparameterization further complicates this inference process due to the equifinality (*Beven*, 2001; *von Bertalanffy*, 1968) of feasible solutions. In most cases, finding them constitutes a difficult optimization problem. Moreover, solutions determined through calibration are not transferable to ungauged locations or to different scales (*Liang et al.*, 2004; *Troy et al.*, 2008) other than that used during calibration.

For these reasons regionalization techniques have been pursued in hydrological modeling. They have been aimed 1) to reduce model overparametrization (*Pokhrel et al.*, 2008), 2) to confine the parameter search to realistic values (*Gotzinger and Bárdossy*, 2007; *Hundecha and Bárdossy*, 2004), and 3) to allow the transfer of information (e.g. model parameters) from gauged to ungauged locations (*Abdulla and Lettenmaier*, 1997; *Mosley*, 1981).

Parameter regionalization techniques reported in literature can be categorized into two main groups: Firstly, parameter regionalization carried out after model calibration, or simply *post-regionalization*; and secondly, parameter regionalization carried out through simultaneous calibration of transfer-function parameters by assuming prior relationships between basin predictors (e.g. elevation, slope, soil texture, vegetation characteristics, etc.) and model parameters, or simply *simultaneous regionalization*. Parameters required for establishing these relationships are called "transfer-function parameters" (*Hundecha and Bárdossy*, 2004), global- or super-parameters (*Pokhrel et al.*, 2008).

In general, the **post-regionalization** technique consists of the following steps (*Abdulla and Lettenmaier*, 1997; *Parajka et al.*, 2005; *Seibert*, 1999): 1) select a set of gauged locations or basins, 2) calibrate a hydrologic model in each basin independently, 3) perform a multivariate input-output analysis (e.g regression, neural networks) to link model parameters obtained for each basin with a set of basin predictors, and 4) cross-validate the results in a gauged basin that was not taken into account during the calibration phase. This technique is quite simple to implement but quite disadvantageous due to the following reasons. Firstly, a set of calibrated model parameters can be a good solution to minimize a given error function but it might be a bad one to perform regionalization analysis because it does not conform with the physical range expected for a given parameter. Or in other words, these calibrated parameters might be only an artifact of the optimization algorithm. And secondly, because interactions among regionalization functions are not considered during the model calibration (*Boughton and Chiew*, 2007; *Heuvelmans et al.*, 2006; *Parajka et al.*, 2005; *Wagener and Wheater*, 2006), which implies that the transfer function parameters may turn out to be weak or even wrong estimations (*Kim and Kaluarachchi*, 2008).

The **simultaneous-regionalization** technique has been proposed to address the shortcomings of the previous approach as well as to take into account the spatial variability of the model parameters. The basic procedure is as follows: 1) select a group of gauged basins, 2) establish a priori functional relationships (i.e. transfer-functions) between model parameters and basin predictors, 3) calibrate the transfer function parameters coupled with the hydrologic model, and 4) cross-validate these parameters in a gauged basin that was not used during the calibration.

This approach has been used in several studies. For instance, *Hundecha and Bárdossy* (2004) employed the simultaneous-regionalization method in a semidistributed conceptual model (HBV-IWS) with five regionalized parameters. In this case, the spatial variability of the basin predictors was grouped into zones, which were defined within each sub-basins based on land cover classes, soil types, and elevation. A follow up of this study was proposed by *Gotzinger and Bárdossy* (2007), in which only the top-soil reservoir of the HBV-IWS model was conceived as spatially distributed. In this study, the authors introduced monotony and Lipschitz conditions into the optimization problem to ensure the continuity of the model parameters in neighboring cells which share similar properties. In both studies, models were able to reproduce quite well the discharge hydrograph. However, reasonable soil moisture patterns are unlikely to be obtained in both cases, since this regionalization technique employs discrete classes as basin predictors. Most recently, *Pokhrel et al.* (2008) followed a similar approach to regionalize HL-DHL model parameters based on a priori estimates derived by *Koren et al.* (2004).

A variant of this procedure was proposed by *Fernandez et al.* (2000) and *Kim and Kaluarachchi* (2008). In this case, the authors simultaneously calibrate both the model parameters and the transfer function parameters, which – in our opinion – is redundant because the model parameters are already coupled with the transfer function parameters.

Troy et al. (2008) proposed an alternative regionalization procedure aiming to reduce the computational time needed for the calibration of the VIC land surface model. To attain this goal, model parameters were obtained through calibration for a subset of the grid cells. Subsequently, parameters for the uncalibrated grid cells were found by linear interpolation. Their results indicated that calibrating the model at different temporal resolutions caused minimal changes in modeled runoff while transferring parameter sets across spatial resolutions did induce significant changes in model performance. The main shortcoming of this technique – in our opinion – is that neither the optimized parameter values nor the regionalization function (i.e. the linear interpolation) have a functional relationship with the physiographic characteristics of the uncalibrated grid cells whose spatial variability is not necessarily linear. Large bias in simulated runoff at some modeled locations, as reported by the authors, may be a consequence of this strong assumption.

The common feature among regionalization techniques reported in recent literature is that the sub-grid variability of the basin predictors is not taken explicitly into account. Hence, basin predictors are defined at the same scale as the modeling units (e.g. grid cells). Hereafter, these types of techniques will be denoted as **Standard Regionalization** (SR) to distinguish them from the technique proposed in this study.

In this study, a Multiscale Parameter Regionalization (MPR) technique is proposed to overcome the issues mentioned above. This approach is a type of simultaneous regionalization but it differs in many important aspects from those found in the reviewed literature (Fig. 3.1). Foremost, the regionalization is performed at a finer resolution (i.e. data input level) to account for the sub-grid variability of basin predictors. Subsequently, effective parameter values required for hydrologic modeling at a coarser grid are obtained with appropriate upscaling operators.

Regionalization at the finer scale is carried out with linear or nonlinear transfer functions which are based on process understanding and empirical evidence (e.g. pedo-transfer functions). These functions aim to establish a quasi-continuous link between model parameters and basin predictors through transfer function parameters (or global parameters). The purpose of the subsequent step is to estimate an effective value of a parameter that captures the emerging structure to suitably describe the dominant hydrological processes at a coarser grid.

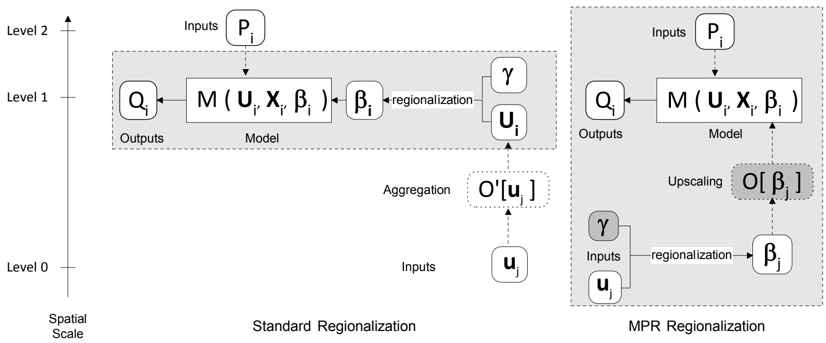

Figure 3.1 Schematic comparison between MPR and the standard simultaneous regionalization approaches. In both cases, the set of global parameters γ is to be obtained through calibration.

One of the main advantage of distributed models over lumped ones is their ability to reproduce streamflow at ungauged locations within a basin (*Reed et al.*, 2004). However, the performance of regionalization techniques within the context of simulating streamflow at internal ungauged locations is not satisfactory yet (*Pokhrel and Gupta*, 2010). There is also extensive literature dealing with the issue of how to transfer model parameters to ungauged basins based on dissimilarity measures, but this is out of the scope of this paper. Interested readers may refer to *Samaniego et al.* (2010b); *Wagener et al.* (2007) and sources therein.

MPR implemented within a spatially distributed mesoscale hydrologic model (mHM) (*Samaniego et al.*, 2010b) has successfully contributed to ameliorating the current shortcomings of existing distributed hydrologic models, namely overparametrization and the lack of an effective technique to integrate the spatial heterogeneity of soils, vegetation and topography into the model. Extensive numerical experiments carried out in this study supported the research hypothesis that explicitly accounting for the sub-grid variability in a regionalization technique (i.e. MPR) is essential to facilitate the transferability of global parameters to ungauged locations or to other modeling scales not considered during calibration. Furthermore, it contributed to the conservation of total water fluxes on a given control volume as well as retaining a reasonable level of accuracy for streamflow prediction.

Improvements in these areas are urgently needed for the efficient application of existing distributed hydrologic models in operational streamflow forecast at gauged and ungauged locations (PUB initiative) and for coupling these models with regional climate models.

3.2 Conceptualization of the System

3.2.1 General Problem Formulation

A mesoscale river basin is an open system that can be defined mathematically in various ways depending on how the spatio-temporal variability of the basin characteristics is described. If the basin characteristics can be assumed to be continuous in space, and its media characteristics and boundary conditions would be known at the point scale, and the governing process would be fully scalable, then a system of partial differential equations (PDE) would be suitable to describe the evolution of the dominant processes at this scale (*Freeze and Harlan*, 1969). Conversely, a system of ordinary differential equations (ODE) may be appropriate to describe the temporal evolution of these processes at a given location i (*Blöschl et al.*, 2008), if these characteristics appeared to be discrete with unknown scaling laws.

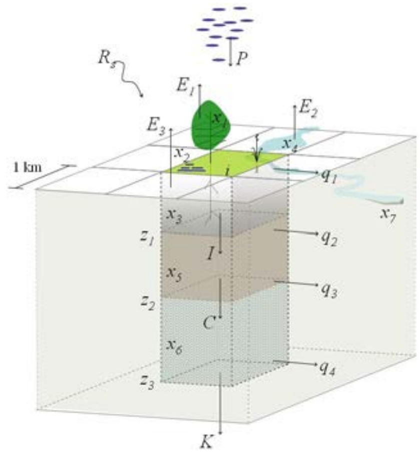

Figure 3.2 General model structure for a cell i at time point t draining to a stream section within this cell (graphic is not to scale).

Since the continuity and the scalability assumptions are quite difficult to justify at the mesoscale, most conceptual hydrologic models (e.g. HBV, SAC, VIC-3L) have adopted an ODE formulation, which may also include stochastic terms representing the uncertainty of the system. In general, let $\mathcal{M}\{\mathbf{f}, \mathbf{g}\}$ be a distributed mesoscale water balance hydrologic model that relates the state variables $\mathbf{x}$ with some observables categorized as inputs $\mathbf{u}$ and outputs $\mathbf{y}$. Here $\mathbf{f}$ and $\mathbf{g}$ denote a set of functional relationships that describe the evolution of the system and the quantification of model outputs, respectively. $\mathbf{u}$ is a set of fields (grids) representing the land cover, the physiographical and the meteorological variables. Based on this model, the rate of change of the state variables at a given cell i (Fig. 3.2) and point in time t are

$$\dot{\mathbf{x}}_i(t) = \mathbf{f}\big(\mathbf{x}_i(t), \mathbf{u}_i(t), \boldsymbol{\beta}_i(t)\big) + \boldsymbol{\eta}_i(t) \quad \forall i \in \Omega \tag{3.1}$$

where $\boldsymbol{\beta}$ is a vector of location specific parameters. Some of these parameters may vary in time to take into account changes in land cover (for more details; see Section 3.2.2). $\boldsymbol{\eta}$ is a vector of unmeasurable stochastic inputs, which can be interpreted as the degree of uncertainty originated due to the lack of knowledge about the dominant processes during the formulation of $\mathcal{M}$. Ω denotes the spatial domain of a river basin, and $i \in \Omega$.

Observed outputs such as streamflow or ground water levels $\mathbf{y}$ at given locations $m \in \Omega$ in time t are defined by

$$\mathbf{y}_m(t) = \mathbf{g}\big(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\beta}(t)\big) + \boldsymbol{\epsilon}_m(t) \tag{3.2}$$

where $\boldsymbol{\epsilon}$ is a vector denoting the uncertainty of the system originated by defects on measurements of both the inputs and outputs. A necessary condition of the lattice covering this domain Ω is, that there should be a unique point having the highest flow accumulation. This point is denoted hereafter as the basin's outlet.

It is worth noting that inputs fields $\mathbf{u}$ and outputs $\mathbf{y}$ can be measured at predetermined time intervals. Conversely, the state variables and their rate of change (represented as $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt}$) can only be inferred indirectly. To solve this system of ODEs, we assume for the sake of simplicity, that the stochastic term $\boldsymbol{\eta} = 0$ [Eq. (3.1)]. There are other available alternatives to solve explicitly this system of stochastic ODE, for example through Bayesian analysis (*Kavetski et al.*, 2006) or data assimilation (*Vrugt et al.*, 2005). The application of any of these techniques within the framework of the multiscale regionalization technique is possible, but out of the scope of this study. The assumption stated above, although disadvantageous in other respects, would allow us to test the efficiency of the proposed regionalization method.

3.2.2 Model Parametrization

As a result of the proposed solution scheme, the uncertainty originated by the combination of various causes, such as: defects in measurements, deficiencies of the model structure, inaccurate process representation, and degree of spatio-temporal discretization, is embedded in both model parameters $\boldsymbol{\beta}$ and measurements' error $\boldsymbol{\epsilon}$. This uncertainty is, in turn, translated into the model predictive uncertainty, which is also greatly magnified by over-parametrization (*Beven*, 2001).

The challenge is to formulate a parametrization method able to cope with data availability (i.e. resolution) but still robust enough to make reliable predictions under a changing environment.

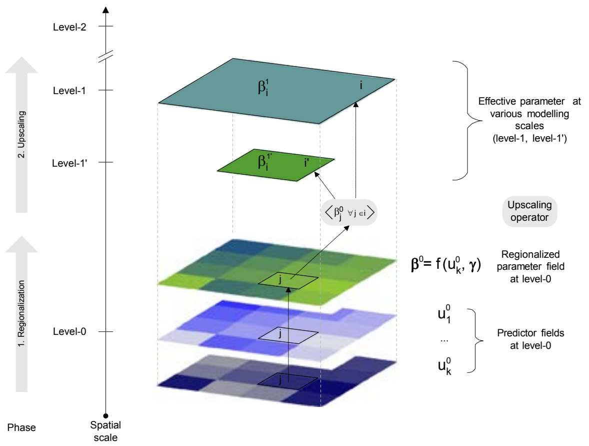

Figure 3.3 Estimation of effective regionalized parameters $\boldsymbol{\beta}^1$ at level-1 based on regionalized model parameters $\boldsymbol{\beta}^0$ at level-0. Note that global parameters $\boldsymbol{\gamma}$ are common for both the effective parameters at level-1 and level-1'. Given the level-0 information and a modeling scale, say at level-1, $\boldsymbol{\gamma}$ can be determined via calibration.

In this study, a Multiscale Parameter Regionalization (MPR) technique is introduced to overcome the over-parameterization problem, to explicitly account for the sub-grid variability within the parameter regionalization framework, and also to ease the transferability of global parameters to other scales (*Wagener and Kollat*, 2007) and locations, which are different from those used during calibration; all of these without inducing statistical significant bias in either streamflow predictions and/or simulated water fluxes. The basis of this method is that we are not interested in estimating *aggregated basin characteristics* having little or no information regarding the spatial variability of the natural factors that regulate the hydrological process at the sub-grid scale, but rather, in estimating *effective model parameters* that capture the emergent properties of these processes. To achieve this objective, the following hierarchy of spatial scales is considered.

3.2.3 Hierarchy of Spatial Scales

The spatial dimension of the dominant hydrological processes occurring at the mesoscale span over several orders of magnitude. In this study, three levels (Fig. 3.3) are differentiated to better incorporate and represent the spatial variability of input and state variables:

1. **Level-0**: Spatial discretization used to describe the sub-grid variability of relevant basin characteristics such as terrain elevation, slope and aspect, the main soil characteristics and number of horizons of pedotops, main geological formations of the basin as well as the land cover. The cell size and cell index at this level is denoted by ℓ_0 and j, respectively.

2. **Level-1**: Spatial discretization used to describe dominant hydrological processes (*Blöschl*, 2001) at the mesoscale, referred hereafter as the *modeling scale*. The cell size and cell index at this level is denoted by ℓ_1 (with $\ell_1 \gg \ell_0$) and i, respectively.

3. **Level-2**: Spatial discretization used to describe the variability of the meteorological forcings at the mesoscale, for example the formation of convective precipitation cells. The cell size at this level is denoted by ℓ_2, with $\ell_2 \geq \ell_1$.

3.3 Multiscale Parameter Regionalization

The MPR technique requires two phases to estimate the effective values of model parameters $\boldsymbol{\beta}$ at the modeling scale, namely: 1) regionalization and 2) upscaling. These phases are depicted in Fig. 3.3.

3.3.1 Parameter Regionalization

The first phase of MPR consists on establishing *a priori* relationships between fields of model parameters $\boldsymbol{\beta}^0$ and distributed basin predictors $\mathbf{u}^0$ at level-0 (Fig. 3.3). These spatio-temporal fields are merged together through linear or nonlinear transfer functions $f(\bullet)$ and scalar values denoted hereafter as *global parameters* $\boldsymbol{\gamma}$.

Global parameters $\boldsymbol{\gamma}$ are hypothesized as quasi-scale-independent scalar values that remain fixed across the whole modeling domain Ω. These global parameters along with the transfer functions and different basin predictors $\mathbf{u}^0$ determine the fields of model parameters $\boldsymbol{\beta}^0$ at level-0. These functions are based on process understanding and/or empirical evidence (e.g. pedo-transfer functions).

Commonly, static morphological variables such as soil texture properties; terrain elevation, slope, and aspect; average conductivity of main geological formations as well as time series of land cover classes available at level-0 scale can be used as basin predictors (*Abdulla and Lettenmaier*, 1997; *Hundecha and Bárdossy*, 2004). This is possible partly because these variables are intimately related with the generating process at the ℓ_0 scale such as the runoff generation at hillslopes (*Becker and McDonnell*, 1998).

The general formulation of a regionalization or transfer function $f(\bullet)$ is

$$\beta^0_{lj}(t) = f_l\Big(\boldsymbol{u}^0_j(t), \boldsymbol{\gamma}\Big) \tag{3.3}$$

where $\beta^0_{lj}(t)$ denotes the l-th model parameter defined at the cell j of level-0 in time point t, $l = 1, \ldots, p$, with p denoting the total number of model parameters. $\boldsymbol{u}^0_j$ refers to a v-dimensional predictor vector for cell j and $\boldsymbol{\gamma}$ is a s-dimensional vector of global parameters. Vector dimensions v and s denote the total number of basin predictors and the total number of free parameters to be calibrated, respectively. Some model parameters are time-dependent due to changes in basin predictors such as land cover.

The main objective of this first step is to reduce model overparameterization, to ease the transferability of global parameter sets from gauged to ungauged catchments (*Samaniego et al.*, 2010b), and, if possible, to increase overall model performance. Moreover, the parameter regionalization performed at level-0 scale (smallest scale supported by the input data), also ensures that MPR explicitly account for the sub-grid variability within the modeling framework (i.e. preprocessing of input data is not required as it is commonly done in the SR method).

3.3.2 Estimation of Effective Parameters

The second phase of MPR consists in upscaling of the l^{th} regionalized model parameter $\beta^0_l(t)$ from level-0 to the modeling level-1 (cell i) in a way that the resulting parameter $\beta^1_{li}(t)$ becomes an effective parameter that encapsulates the emerging features of a given process at this scale (Fig. 3.3). The main challenge in this phase is therefore to find the best, often non-linear, aggregation or upscaling rules, hereafter denoted as upscaling operators.

Moreover, these operators should also take into account the characteristics of the sub-grid variability of a given parameter (i.e. second and higher moments) and its propagation via non-linear equations describing the hydrological system. If this is not done properly, significant biases in predicted variables would be introduced. The reason for that stems from the fact that predicting the evolution of an aggregated variable x^1 at a larger scale through a non-linear process $\mathcal{P}(x^1)$ may be quite different from predicting the evolution of the subgrid-scale variability of variable x^0, since

$$\overline{\mathcal{P}(x^0_j)} \neq \mathcal{P}(x^1_i), \quad \forall j \in i \quad \forall i \in \Omega \tag{3.4}$$

where

$$x^1_i = \overline{x^0_j}. \tag{3.5}$$

with the overbar denoting the arithmetic mean. The magnitude of the difference between the $\overline{\mathcal{P}(x^0_j)}$ and $\mathcal{P}(x^1_i)$ would depend mainly on the temporal gradient of the function $\mathcal{P}$ and the variance of x (*Nykanen et al.*, 2001).

Each process and its related effective parameters, however, would have to be analyzed on a case by case approach, since no generally agreed upon upscaling theory exists for dominant hydrologic processes at the mesoscale

(*Binley et al.* (1989) and sources therein). Consequently, the upscaling operators can be considered an approximation to account for the influence of sub-grid variability on the model parameters. The schematic representation of this phase is shown in Fig. 3.3.

The general form of an upscaling operator O applied to the l-th model parameter at level-1 ($\beta^1_{li}(t), \quad l = 1, \ldots, p \quad \forall i \in \Omega$) is given by:

$$\beta^1_{li}(t) = \mathrm{O}_l \Big\langle \beta^0_{lj}(t) \quad \forall j \in i \Big\rangle \tag{3.6}$$

where p denotes the total number of model parameters, and i, j are indices related to grid cells at level-1 and level-0, respectively. $\mathrm{O}_l\langle \bullet \rangle$ is the upscaling operator applied to the model parameter l. There are a number of possible operators or upscaling functions that can be used within the MPR method. Their selection and type, however, should be based on conceptual and/or process understanding, and subject to evaluation. Additionally, parsimonious relationships should be preferred to complex ones. In this study, five kinds of upscaling operators were tested, namely: the majority operator $\mathcal{M}$, the arithmetic mean $\mathcal{A}$, the maximum difference $\mathcal{D}$, the geometric mean $\mathcal{G}$, and the harmonic mean $\mathcal{H}$ (Table 3.1).

Table 3.1 Types of upscaling operators used in MPR to derive an effective parameter at level-1 based on regionalized parameters at level-0. i and j denotes the cell index at these levels, respectively. b is value of the field with the highest frequency of occurrence. $|\cdot|$ denotes the cardinality of the set.

Name	Notation	Estimation	Condition
Arithmetic mean	$\mathcal{A}$	$\beta^1_i(t) = \mathcal{A}\langle\, \beta^0_j(t)\,\rangle = \frac{1}{n}\sum_j \beta^0_j(t)$	$\forall j \in i$
Maximum difference	$\mathcal{D}$	$\beta^1_i(t) = \mathcal{D}\langle\, \beta^0_j(t)\,\rangle = \max\big(\beta^0_j(t)\big) - \min\big(\beta^0_j(t)\big)$	$\forall j \in i$
Geometric mean	$\mathcal{G}$	$\beta^1_i(t) = \mathcal{G}\langle\, \beta^0_j(t)\,\rangle = \big(\prod_j \beta^0_j(t)\big)^{\frac{1}{n}}$	$\forall j \in i$
Harmonic mean	$\mathcal{H}$	$\beta^1_i(t) = \mathcal{H}\langle\, \beta^0_j(t)\,\rangle = \frac{n}{\sum_j \frac{1}{\beta^0_j(t)}}$	$\beta^0_j(t) > 0, \forall j \in i$
Majority	$\mathcal{M}$	$\beta^1_i(t) = \mathcal{M}\langle\, \beta^0_j(t)\,\rangle = b$	$\lvert\{\beta^0_j(t) = b\} \quad \forall j \in i\rvert \rightarrow \max \quad b \in \mathbb{N}$

3.3.3 Remarks on the MPR Technique

Fisrt, the regionalization in MPR is carried out at the lowest spatial resolution supported by the data (i.e. level-0), which contributes to preserving the spatial variability of both the predictors and the regionalized parameters. This characteristic not only distinguishes this approach from standard regionalization methods (Fig. 3.1) but also minimizes the bias introduced by simple aggregation of predictors. In contrast with MPR, basin predictors in standard regionalization techniques are firstly aggregated from level-0 to level-1, and afterwards parameter regionalization is performed at modeling scale (e.g. *Hundecha and Bárdossy* (2004), *Pokhrel et al.* (2008)).

Second, MPR greatly reduces the level of model complexity as denoted by the following inequality

$$p \times n_\Omega \gg s \tag{3.7}$$

where n_Ω denotes the number of cells contained within the basin Ω. For example, if a hydrologic model requires $p = 28$ parameters per cell (e.g. mHM, see Section 3.4.1) and would be calibrated without regionalization in a basin covering an area 1000 km^2 with a resolution of 1 km^2 (i.e. $n_\Omega = 1000$ cells), then the optimization algorithm would have to search for a good solution to an optimization problem with 28 $\times$ 1000 = 28 000 degrees of freedom, a daunting computational task! Conversely, if MPR would be applied within mHM, then merely $s = 62$ global parameters $\boldsymbol{\gamma}$ would have to be estimated. Consequently, MPR becomes quite advantageous during model calibration because sampling in a lower dimensional space improves dramatically the convergence speed of any optimization algorithm (*Pokhrel et al.*, 2008).

Third, MPR not only considerably reduces the parameter uncertainty by minimizing the number of the free parameters to be calibrated (see above) but also allows to ascertain effective parameters $\boldsymbol{\beta^1}$ at various modeling scales, using the same upscaling operators, and without re-calibrating the global parameters $\boldsymbol{\gamma}$ (Fig. 3.3). Global parameters are expected to be time-invariant and quasi scale-independent.

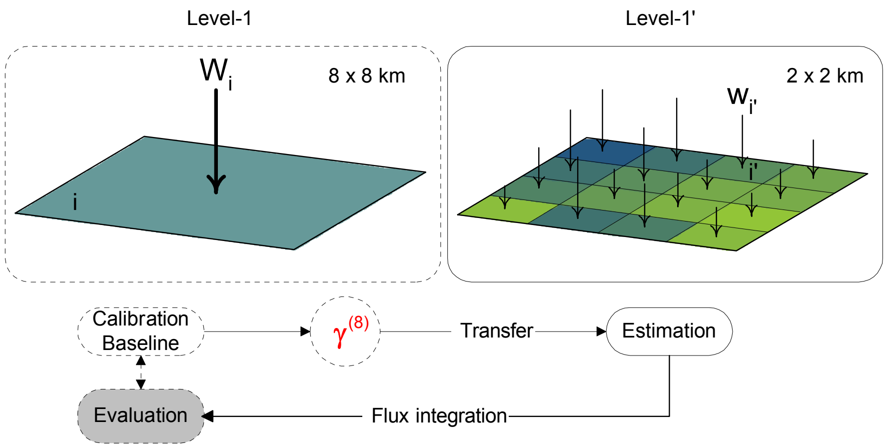

Figure 3.4 Evaluation scheme for the evaluation of the continuity principle between two modeling scales. In this example, level-1 is the control scale at 8×8 km whereas the level-1' is the simulations scale at 2×2 km. Here $\boldsymbol{\gamma}^{(8)}$ denote global parameters (scalars) calibrated at level-1. $W_i(t)$ and $w_{i'}(t)$ is the estimated water fluxes at each scale, respectively.

Finally, MPR also ensures that the continuity principle (i.e. mass conservation on a given control volume at level-1) is satisfied to a larger extent if global parameters calibrated for a coarser modeling scale (e.g. $\ell_1 = 8$ km) would be used for a finer one (e.g. $\ell_1 = 2$ km) as indicated in Fig. 3.4. It should be noted that the calibration of global parameters at a coarser modeling scale is advantageous because this implies a significant reduction in the computational time required for model calibration (*Troy et al.*, 2008). Large deviations in the distribution of fluxes are a clear indication of a biased and poor regionalization technique.

3.4 Application

3.4.1 The Model

A mesoscale grid-based conceptual hydrologic model (mHM) was employed in this study to test MPR. mHM includes a number of new features and improvements that ease the implementation of the proposed regionalization technique. This model is based on numerical approximations applied in known hydrologic models such as HBV (*Blöschl et al.*, 2008; *Hundecha and Bárdossy*, 2004) and VIC-3L (*Liang et al.*, 1994). mHM has been applied in 38 mesoscale basins in Germany ranging in size from 70 km^2 to 4000 km^2 (*Samaniego et al.*, 2010b).

In general, this model simulates the following processes (Fig. 3.2): canopy interception, snow accumulation and melting, soil moisture dynamics, infiltration and surface runoff, evapotranspiration, subsurface storage and discharge generation, deep percolation and baseflow, and discharge attenuation and flood routing. An extended description of the model, however, is out of the scope of this paper. The main equations of mHM are briefly presented in Appendix 3.7. The model requires $p = 28$ parameters per modeling cell (level-1). Using MPR, however, only $s = 62$ global parameters need to be calibrated. mHM is entirely written in Fortran 2003.

3.4.2 Implementation of MPR and SR in mHM

A short description of all basin predictors employed in both regionalization methods (MPR and SR) is shown in Table 3.2. In this table, $\mathbf{u}$ denotes the spatial variability of a given variable either at level-0 or level-1 scale. These predictors, excluding land cover, leaf area index (LAI), and fraction of impervious cover on floodplains, are time-independent variables.

Not all model parameters of mHM are required to be regionalized because they do not exhibit spatial variability at the mesoscale level and thus can be assumed as global parameters. Among these parameters are: β_2, β_4, β_9, β_{11}, β_{12}, β_{14} (see definition in the Appendix). Others were regionalized as indicated in Table 3.3.

The performance of several types of upscaling operators in MPR were studied during the calibration and evaluation phases. Most of these operators, as well as the relationships between catchment characteristics and parameter fields at level-0, are based on process understanding and/or empirical evidence. For the soil-related characteristics, however, only the harmonic mean was used as suggested by *Zhu and Mohanty* (2002). A summary of the transfer functions used in mHM is shown in Table 3.4.

mHM was also parameterized with the standard regionalization method (SR) to assess the efficiency of both techniques. In SR, basin predictors were first upscaled from level-0 to the required modeling scale at level-1 and then the regionalization of parameters was performed. To make a fair comparison between both methods, model parameters in SR were regionalized with the same functional relationships used in MPR (Table 3.3). As a result, the total number of global parameters $\boldsymbol{\gamma}$ in both methods is the same. The major difference between SR and MPR method is the level at which the parameter regionalization was performed (i.e. the input data scale, or level-0, in MPR and the modeling scale, or level-1 in SR).

Table 3.2 Description of basin predictors used in MPR. $u_\bullet$ denotes a field describing the spatial variability of a given basin characteristic at level-0, if not specified otherwise.

Variable	Description
u_1	Land cover class (time dependent).
u_2	Leaf area index (LAI) (time dependent).
u_3	Fraction of impervious cover on the floodplains (time dependent).
u_4	Sink free DEM.
u_5	Terrain slope.
u_6	Aspect.
u_7	Flow directions based on the DEM.
u_8	Flow accumulation.
u_9	Length of the reach segment in cell i (level-1).
u_{10}	Slope of the reach segment in cell i (level-1).
u_{11}	Length of flow path based on flow direction (u_7).
u_{12}	Mean clay percentage in the root zone.
u_{13}	Mean sand percentage in the root zone.
u_{14}	Mineral bulk density in the root zone (*Rawls*, 1983).
u_{15}	Mean clay percentage in the vadose zone.
u_{16}	Mean sand percentage in the vadose zone.
u_{17}	Mineral bulk density in the vadose zone (*Rawls*, 1983).
u_{18}	Hydraulic conductivity of major geological formation.
u_{19}	Fraction of karstic formations within a cell i.

3.4.3 Study Area: The Upper Neckar Basin

mHM with both regionalization schemes (i.e. SR and MPR) was applied to the Upper Neckar river basin, near Stuttgart in Germany. It covers an area of approximately 4000 km^2 (Fig. 3.5) and is bounded by the north-western edge of the Swabian-Jura on the right bank of the Neckar river and by the Black Forest on its left bank. The elevation ranges from 240 m to 1014 m a.s.l. with a mean elevation of 546 m. 90 % of the area has mild slopes ranging from 0° to 15°. The annual precipitation is approximately 900 mm per year. The geology of the catchment is composed mainly of altered keuper, claystone-jura, claystone-keuper, limestone-jura, loess, sandstone and shelly limestone. Approximately 35 % of the basin contains karstic formations. The climate of this basin is moist with mild winter according to the Köppen notation. The daily mean air temperature in the coldest and warmest months (i.e. January and July) is -0.8 °C and 17.0 °C. Soil freezing may occur during the winter at higher altitudes (e.g. Black Forest).

3.4.4 Data Availability

Data availability should be carefully considered before a modeling attempt is carried out because this, in turn, constrains the model structure. Typical information available for the Upper Neckar basin is detailed next.

1. Meteorological information is available at hourly or daily intervals from a network of 288 stations (German Weather Service, DWD).

2. Land cover data was derived from Landsat TM5 scenes (30 m × 30 m) classified for 1975, 1983, 1989, 1993, and 2004 (*Samaniego et al.*, 2008). For this study, land cover classes were aggregated as follows: class 1, composed of permeable areas covered by coniferous, deciduous, and mixed forest; class 2, mainly composed of impervious areas with land usage such as settlements, industrial parks, highways, airport runways, and railway tracks; and class 3, mainly composed of permeable areas covered fallow lands, or those surfaces covered by crops, grass, and orchards. Wetlands and water bodies were included into this class because they are insignificant in this region.

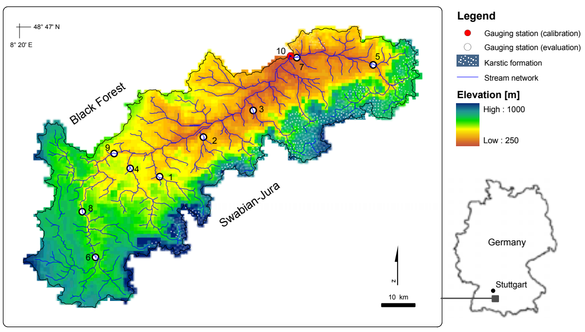

Figure 3.5 Location of the Upper Neckar river basin within Germany. The basin's outlet corresponds to gauge Nr. 10. at Plochingen

3. Weekly leaf area index (LAI) and daily land surface temperature (LST) were obtained from Moderate Resolution Imaging Spectroradiometer (MODIS), NASA, with a spatial resolution of 1000 m × 1000 m for the period from 2001 to 2007. These data are available freely from https://wist.echo.nasa.gov/api/.

4. Soil texture at different horizons as well as the geological formations were obtained from digital soil maps at the scale of 1 : 25 000 (LfU, Environmental Agency Baden-Württemberg).

Table 3.3 Predictors used in the regionalization functions. All model parameters β_l^0 are regionalized at level-0, $l = 1, \ldots, 28$. Superscript index 0 is not shown to ease notation.

Process	Model Parameter	Predictor variables	Reference
Interception	β_1	LAI	(*Dickinson*, 1984; *Fenicia et al.*, 2008)
Snow accum. & Melting	β_2,β_4	–	–
	β_3,β_5	Land cover	(*Gotzinger and Bárdossy*, 2007; *Hundecha and Bárdossy*, 2004)
Infiltration root zone	β_6	Soil texture, land cover	(*Zacharias and Wessolek*, 2007)
	β_7	Soil texture, land cover	(*Brooks and Corey*, 1964)
	β_8	Soil texture, land cover	(*Koren et al.*, 1999)
	β_9,β_{11}, β_{13}	–	–
	β_{10}, β_{12}	Soil texture	*Patterson and Smith* (1981)
Surface Runoff	β_{14}	–	–
EVT root zone	β_{15}, β_{16}	Soil texture, land cover	(*Kutilek and Nielsen*, 1994)
	β_{17}	Land cover	(*Kutilek and Nielsen*, 1994)
Fast interflow	β_{18}	Soil texture, land cover	(*Booij*, 2005)
	β_{19}	Slope	(*Booij*, 2005)
Slow interflow	β_{20}	Soil texture	(*Booij*, 2005)
	β_{21}	Soil texture, elevation	(*Booij*, 2005)
Baseflow	β_{22}	Soil saturated hydraulic conductivity	(*Liang et al.*, 1994)
	β_{23}, β_{24}	Geological formations	(*Le Moine et al.*, 2007)
Routing	β_{25}	Length, slope and land cover of drainage path within the cell	–
	β_{26}	Length, slope and fraction of impervious area of floodplains of the reach segment	(*Tewolde and Smithers*, 2006)
	β_{27}	Slope of the reach segment	(*Tewolde and Smithers*, 2006)
PET	β_{28}	Slope, aspect	(*Shevenell*, 1999)

5. Terrain elevation was obtained from the SRTM sensor (NASA) with a spatial resolution of 90 m $\times$ 90 m. This data was obtained freely from http://srtm.csi.cgiar.org/.

6. The hourly or daily basin's streamflow at various locations shown in Fig. 3.5 was obtained from LfU (Institute for Environmental Protection Baden-Württemberg, Germany) and the DWD.

3.4.5 Discretization and Data Processing

Several spatial resolutions were used in this study for the various levels defined in Section 3.2.3:

Level-0: Physiographic variables $\mathbf{u}$ such as terrain elevation, slope, aspect, soil texture properties, and land cover as well as LAI were defined at four spatial resolutions to test the influence of the sub-grid variability on the parametrization scheme. The selected spatial resolutions were: $\ell_0 =$ (100, 500, 1000, 2000) m.

Level-1: Several modeling resolutions were considered to test the performance of mHM with both regionalization techniques (i.e. MPR and SR). This test includes also the sensitivity of global parameters obtained for a particular modeling resolution, and then used in a different one. The selected modeling resolutions were: $\ell_1 =$ (2, 4, 8, 16, 32) km.

Level-2: In this study, the meteorological data was derived at the same resolution as the modeling scale (i.e. $\ell_2 \equiv \ell_1$).

The vertical discretization of the soil layer was carried out based on soil horizon depth obtained from the soil map. In general, its total depth varies between 30 cm to 90 cm in the study area.

Daily interpolated fields of meteorologic forcings at level-2 (e.g. precipitation, maximum and minimum temperature) were obtained with external drift Kriging (EDK) using terrain elevation as a drift variable. Additionally, long term daytime/nighttime fluctuations of these forcings were also used to better represent their intra-day variability. Potential evapotranspiration (PET) was derived with the procedure proposed by *Hargreaves and Samani* (1985). This variable was additionally corrected to take into account the influence of elevation and aspect (*Shevenell*, 1999).

Table 3.4 Regionalization (transfer) functions and upscaling operators used in mHM. For simplicity three land cover classes, two soil layers, and two geological formations are employed. Forest $\equiv 1$, impervious cover $\equiv 2$, and permeable cover $\equiv 3$. Time index t only used for time dependent parameters. Superscript indexes of $\boldsymbol{\beta}$ (0,1) are not shown to ease notation. Subindexes i and j refer to cells at level-1 and level-0, respectively. $\Vert\beta_j\Vert_i$ denotes a locally normalized field values located within the cell i, i.e. $\Vert\beta_j\Vert_i = \frac{\beta_j}{\max \beta_j}\ \forall j \in i.\Vert \bullet \Vert$ is a globally normalized field.

Process	Parameter l	Operator	Regionalization Function
Interception	1	$\beta_{1i}(t) = \mathcal{A}\langle\, \gamma_1 u_{2j}(t)\, \rangle_i$	
Snow accum. & Melting	2	$\beta_{2i} = \gamma_2$	
	3	$\beta_{3i}(t) = \mathcal{A}\langle\, \beta_{3j}(t)\, \rangle_i$	$\beta_{3j}(t) = \begin{cases} \gamma_3 & u_{1j}(t) = 1 \\ \gamma_4 & u_{1j}(t) = 2 \\ \gamma_5 & u_{1j}(t) = 3 \end{cases}$
	4	$\beta_{4i} = \gamma_6$	
	5	$\beta_{5i}(t) = \mathcal{A}\langle\, \beta_{5j}(t)\, \rangle_i$	$\beta_{5j}(t) = \begin{cases} \gamma_7 & u_{1j}(t) = 1 \\ \gamma_8 & u_{1j}(t) = 2 \\ \gamma_9 & u_{1j}(t) = 3 \end{cases}$
Infiltration root zone	6	$\beta^k_{6i}(t) = \mathcal{H}\langle\, \beta^k_{6j}(t)\, \rangle_i$	$o_j(t) = \begin{cases} \gamma_{10} & u_{1j}(t) = 1 \\ \gamma_{11} & u_{1j}(t) = 2 \\ \gamma_{12} & u_{1j}(t) = 3 \end{cases}$ $\varrho^k_j(t) = \begin{cases} \frac{1}{\frac{o_j(t)}{\varrho_o} + \frac{1-o_j(t)}{u_{14j}}} & k = 1 \\ u_{14j} & k = 2 \end{cases}$ $\beta^k_{6j}(t) = \begin{cases} \gamma_{13} + \gamma_{14} u_{12j} + \gamma_{15} \varrho^k_j(t) & u_{13j} < f_s \\ \gamma_{16} + \gamma_{17} u_{12j} + \gamma_{18} \varrho^k_j(t) & \text{otherwise} \end{cases}$
	7	$\beta^k_{7i}(t) = \mathcal{H}\langle\, \gamma_{19}\, \varrho^k_j(t)\, \rangle_i$	
	8	$\beta_{8i}(t) = \begin{cases} \mathcal{H}\langle\, \gamma_{20} + \gamma_{21} \frac{u_{13j}}{100}\, \rangle_i & \mathcal{M}\langle\, u_{1j}(t)\, \rangle_i = 1 \\ \mathcal{H}\langle\, \gamma_{20} + \gamma_{22} \frac{u_{13j}}{100}\, \rangle_i & \mathcal{M}\langle\, u_{1j}(t)\, \rangle_i = 2 \\ \mathcal{H}\langle\, \gamma_{20} + \gamma_{23} \frac{u_{13j}}{100}\, \rangle_i & \mathcal{M}\langle\, u_{1j}(t)\, \rangle_i = 3 \end{cases}$	
	9	$\beta_{9i} = \gamma_{24}$	
	10	$\beta_{10i} = \mathcal{H}\langle\, \gamma_{25} - \gamma_{26} \frac{u_{12j}}{100}\, \rangle_i$	
	11	$\beta_{11i} = \gamma_{27}$	
	12	$\beta_{12i} = \mathcal{H}\langle\, \gamma_{28} + \gamma_{29} \frac{u_{12j}}{100}\, \rangle_i$	
	13	$\beta_{13i} = \gamma_{30}$	
Surface Runoff	14	$\beta_{14i} = \gamma_{31}$	
EVT root zone	15	$\beta^k_{15i}(t) = \mathcal{H}\langle\, \gamma_{32}\, \beta^k_{6j}(t)\, \rangle_i$	
	16	$\beta^k_{16i}(t) = \mathcal{H}\langle\, \gamma_{33}\, \beta^k_{6j}(t)\, \rangle_i$	
	17	$\beta_{17i}(t) = \mathcal{A}\langle\, \beta_{17j}(t)\, \rangle_i$	$\beta_{17j}(t) = \begin{cases} \gamma_{34} & u_{1j}(t) = 1 \\ \gamma_{35} & u_{1j}(t) = 2 \\ \gamma_{36} & u_{1j}(t) = 3 \end{cases}$
Fast interflow	18	$\beta_{18i} = \mathcal{H}\langle\, \gamma_{37}(1 + \Vert\, \Theta_j \Vert)\, \rangle_i$	$\Theta_j = \begin{cases} \gamma_{13} + \gamma_{14} u_{6j} + \gamma_{15} u_{17j} & u_{16j} < f_s \\ \gamma_{16} + \gamma_{17} u_{6j} + \gamma_{18} u_{17j} & \text{otherwise} \end{cases}$
	19	$\beta_{19i} = \mathcal{A}\langle\, \gamma_{38}(1 + \Vert u_{5j} \Vert)\, \rangle_i$	
Slow interflow	20	$\beta_{20i} = \mathcal{A}\langle\, \gamma_{39} + \gamma_{40}\, (1 + \Vert\kappa_j\Vert) + \gamma_{41}\, (1 + \Vert\mathcal{D}\langle\, u_{4j}\, \rangle\Vert)\, \rangle_i$	$\kappa_j = \gamma_{42} e^{\gamma_{43}\, u_{16j} - \gamma_{44}\, u_{15j}}$
	21	$\beta_{21i} = \mathcal{A}\langle\, \gamma_{45}\, (1 + \Vert\mathcal{D}\langle\, u_{4j}\, \rangle\Vert)\, \rangle_i$	
Baseflow	22	$\beta_{22i} = \mathcal{A}\langle\, \gamma_{46}\, (1 + \Vert\kappa_j\Vert_i)\, \rangle_i$	
	23	$\beta_{23i}(t) = \begin{cases} \gamma_{47} & \mathcal{M}\langle\, u_{18j}(t)\, \rangle_i = 1 \\ \gamma_{48} & \mathcal{M}\langle\, u_{18j}(t)\, \rangle_i = 2 \end{cases}$	
	24	$\beta_{24i}(t) = 1 + (-1)^{\gamma_{49}}\, \gamma_{50}\, \mathcal{M}\langle\, u_{19j}\, \rangle_i$	
Routing	25	$\beta_{25i}(t) = \mathcal{G}\langle\, \Upsilon_j(t)\, \rangle_i$	$\psi_{j'} = \begin{cases} \gamma_{51} & u_{1j}(t) = 1 \\ \gamma_{52} & u_{1j}(t) = 2 \\ \gamma_{53} & u_{1j}(t) = 3 \end{cases}$ $\Upsilon_j(t) = \sum_{j' \in \{j \to i_o\}} \psi_{j'}\, \frac{(u_{11j'})^{\gamma_{54}}}{(u_{5j'})^{\gamma_{55}}}$
	26	$\beta_{26i}(t) = \gamma_{56} \Vert u_{9i} \Vert\, (\, 1 - \Vert u_{3i} \Vert\,)^{\gamma_{57}}\, \Vert u_{10i} \Vert^{\gamma_{58}}$	
	27	$\beta_{27i}(t) = \gamma_{59} \Vert u_{10i} \Vert$	
PET	28	$\beta_{28i}(t) = \begin{cases} \mathcal{A}\langle\, \gamma_{60} + \frac{\gamma_{61} - \gamma_{60}}{\gamma_{62}}\, u_{6j}(t)\, \rangle_i & u_{6j}(t) < \gamma_{62} \\ \mathcal{A}\langle\, \gamma_{60} + \frac{\gamma_{61} - \gamma_{60}}{360 - \gamma_{62}}\, (360 - u_{6j}(t))\, \rangle_i & \text{otherwise} \end{cases}$	

PET was afterwards disaggregated according to its long-term day-night variability to better describe the daily soil moisture dynamics. In areas, where surfaces are covered by impounded water, the daily PET was adjusted with the pan-evaporation coefficient (f_p) to estimate the maximum free-water surface evaporation rate.

The spatio-temporal variability of LAI was approximated by combining the land cover data at level-0 (based on LANDSAT TM5) with the long-term weekly LAI for each land cover class.

3.4.6 Parameter Identification

Good sets of global parameters $\boldsymbol{\gamma}$ were identified with a split-sampling technique using a constrained optimization algorithm based on simulated annealing (SA) (*Aarts and Korst*, 1990). The following estimation procedure was used:

Algorithm 1:

1. Randomly select an initial set of global parameters $\boldsymbol{\gamma}$ within their predefined ranges and constraints.
2. Estimate model parameters $\boldsymbol{\beta^0}$ at level-0, using [Eq. (3.3)].
3. Estimate effective model parameters $\boldsymbol{\beta^1}$ at level-1, using [Eq. (3.6)].
4. Set an initial annealing temperature τ (a priori estimate).
5. Calculate the current objective function Φ [Eq. (3.8)].
6. Randomly select an index (ι) with $1 \leq \iota \leq s$.
7. Randomly modify the element γ_ι of the vector $\boldsymbol{\gamma}$ and formulate a new vector $\boldsymbol{\gamma}^*$.
8. Estimate model parameters $\boldsymbol{\beta^{0*}}$ and effective model parameters $\boldsymbol{\beta^{1*}}$ using [Eq. (3.3)] and [Eq. (3.6)], respectively.
9. Calculate the new objective function Φ^*.
10. If $\Phi^* \leq \Phi$ then replace $\boldsymbol{\gamma}$ by $\boldsymbol{\gamma}*$. Else calculate $\pi = \exp(\frac{\Phi - \Phi^*}{\tau})$. With the probability π, replace $\boldsymbol{\gamma}$ by $\boldsymbol{\gamma}^*$.
11. Repeat steps (6)-(10) M times (with M being the length of the Markov chain of SA (*Aarts and Korst*, 1990)).
12 .Reduce the annealing temperature τ and repeat steps (6)-(11) until the objective function Φ achieves a minimum.

The overall model efficiency Φ was estimated as a weighted combination of four estimators based on the Nash-Sutcliffe efficiency (NSE) between observed and calculated streamflows using three different time scales (daily, monthly and yearly) as well as the logarithm of the daily streamflow to de-emphasize the effects of the peak flows over the low flows. These objective functions are denoted by ϕ_n, $n = 1, 4$. The overall objective function to be minimized is then

$$\Phi = \Big(\sum_n \mathrm{w}_n^p (1 - \phi_n)^p \Big)^{\frac{1}{p}} \tag{3.8}$$

where $p > 1$, and $\sum_{n=1}^{4} \mathrm{w}_n = 1$. Here p is an exponent according to the compromise programming technique (*Duckstein and Opricovic*, 1980) and w_n denote the degree of importance of each objective. High values of p, say $p = 6$, should be chosen to avoid substitution of objective function values at low levels. In this study, the estimators related with daily streamflows were twice as important as the long-term ones, thus $\{\mathrm{w}_n\} = \{\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}\}$. The NSE for a given time scale t' (say day, month or year) is given by

$$\phi_n = 1 - \frac{\sum\limits_{t'} \left(y_n(t') - \hat{y}_n(t')\right)^2}{\sum\limits_{t'} \left(y_n(t') - \overline{y}_n(t')\right)^2} \tag{3.9}$$

where $\overline{y}_n(t')$ is the mean value of the observations time series over the calibration period. The index n denotes here the daily, monthly, yearly, and the transformed $\ln\big(y(t')\big)$ streamflow discharges. y and $\hat{y}$ are the observed and simulated streamflows at a given time scale.

The calibration and evaluation periods selected for all simulations described in this study ranged between 1979-11-01 to 1988-10-31 and between 1988-11-01 to 2001-10-31, respectively. For the estimation of fluxes and state variables both periods were used. Six months spin up time was used to establish reliable initial conditions for the state variables. This period, however, was not accounted for in the overall objective function.

3.5 Numerical Experiments and Evaluation Criteria

Various numerical experiments were carried out to evaluate the performance of a parametrization technique with respect to the following criteria:

1. Sensitivity of model efficiency measures to the spatial resolution of basin predictors at level-0. 2. Sensitivity of model efficiency measures to global parameters $\boldsymbol{\gamma}$ calibrated at modeling scales and/or locations different from that currently used. 3. Degree of disruption of the mass balance in a control volume at level-1 caused by the transfer of global parameters $\boldsymbol{\gamma}$ from modeling scales and/or locations different from that currently used. 4. Preservation of spatial patterns of the state variables at various modeling scales (i.e. level-1).

3.5.1 Effect of the Sub-grid Variability

The purpose of this numerical experiment was to assess the effect of the sub-grid variability of predictors (i.e. level-0) on model efficiency, given a predefined modeling level-1. Aggregated statistics based on streamflow only [e.g. the root mean squared error (RMSE) or the Nash Sutcliffe Estimator (NSE)] may not be sufficient to identify these effects mainly due to the nonlinearity of the system. For this reason, the spatial variability statistic $\bar{r}$ based on water fluxes or state variables was also estimated in the following algorithm:

Algorithm 2:

1. Set the spatial resolution of level-1, e.g. $\ell_1 = 2$ km
2. Set a spatial resolution for level-0 data, ℓ_0.
3. Calibrate the model at level-1 based on level-0 information using Algorithm 1.
4. Estimate model efficiency at level-1 (e.g. bias, RMSE, NSE).
5. Estimate the spatial variability statistic $\bar{r}$ [Eq. (3.10)].
6. Repeat (2) to (5) for various spatial resolutions of input data or basin predictors, e.g. $\ell_0 =$ (100, 500, 1000, 2000) m.

Here $\bar{r}$ denotes the expectation of r estimated as

$$\bar{r} = \mathrm{E}\Big[r\big(\mathbf{x}_i^{(100)}(t), \mathbf{x}_i^{(J)}(t)\big)_t \;\forall i \in \Omega\Big],\; \forall t \tag{3.10}$$

where r is the spatial correlation of two fields in time t. It should be noted that, the correlation coefficient r is estimated over the space for each point in time t. $\mathbf{x}_i^{(J)}(t)$ denotes the value of a state variable or a water flux at cell i (level-1) in time point t estimated with effective parameters obtained with level-0 information at a spatial resolution $\ell_0 = J$. J is the level-0 discretization with $J =$ (100, 500, 1000, 2000) m. The simulation with $\ell_0 = 100$ m was used as a baseline for the estimation of $\bar{r}$.

3.5.2 Transferability of Global Parameters across Modeling Scales

The transferability of global parameters may introduce bias either because of the unaccounted spatial heterogeneity of basin predictors or because of the assumptions required to define the regionalization functions and the upscaling rules. In this study, the procedure depicted in Fig. 3.4 was employed to test the mass conservation on a given control volume:

Algorithm 3:

1. Set the spatial resolution of level-0 ($\ell_0 \times \ell_0$) and level-1 ($\ell_1 \times \ell_1$). The latter is denoted as the **control** scale. Level-2 scale is set equal to that of level-1. E.g. ℓ_0 = 100 m and ℓ_1 = 8 km.
2. Find global parameters $\boldsymbol{\gamma}^{(\ell_1)}$ at the control scale (use Algorithm 1), e.g. if ℓ_1 = 8 km then find $\boldsymbol{\gamma}^{(8)}$ as shown in Fig. 3.4.
3. Set a new modeling level-1' for evaluation such that $\ell_1' < \ell_1$, e.g. ℓ_1' = 2 km.
4. Simulate fluxes at level-1' using the set of global parameters obtained for the control scale in (2) (e.g. $\boldsymbol{\gamma}^{(8)}$).
5. Estimate global efficiency measures of the model at level-1' (e.g. RMSE, NSE).
6. Integrate water fluxes obtained in (4) [Eq. (3.11)] from level-1' to level-1 and estimate the statistic r_f [Eq. (3.12)].
7. Repeat (3) to (7) for various spatial resolutions of level-1.

It is worth noting that in MPR, model parameters β^1 are not transferred from one modeling scale ℓ_1 to a different one at level-1, but the global parameters $\boldsymbol{\gamma}^{(\ell_1)}$. Based on them, regionalized fields of parameters β^0 at level-0 can be upscaled to any modeling scale (level-1) as indicated in step (4) of Algorithm 3 (Fig. 3.3).

This algorithm basically consists in finding the absolute deviations between fluxes of every cell i at a control scale defined a priori (e.g. ℓ_1 = 8 km) and the corresponding integral of water fluxes obtained at the finer scale (e.g. ℓ_1 = 2 km). In this case, both simulations (i.e. at coarser and finer modeling scales) have to be forced by the same meteorological drivers and employ the same global parameters $\boldsymbol{\gamma}$, which can be obtained at the coarser resolution via calibration. This implies that model parameters at level-0 ($\boldsymbol{\beta^0}$) are common for both modeling scales. Formally, the absolute deviations at point in time t can be calculated by

$$\left|W_i(t) - \int_i w_{i'}(t)\right| \to 0 \quad i' \in i \quad \forall i \in \Omega \tag{3.11}$$

where $W_i(t)$ and $w_{i'}(t)$ denote the fluxes at time t estimated at the coarser and finer cells i and i' respectively. Ω represents the domain of a simulation.

A possible estimator for the conservation of mass at every grid i over a time interval is the NSE between the water fluxes obtained at the coarser scale $W_i(t)$ (used as baseline values since no observations exist) and the aggregated flux at grid i denoted as $\int_i w_{i'}(t)$. In this case, a negative NSE value indicates a complete mismatch between these two variables. If the principle of continuity is fully satisfied, then the expectation of the NSE statistic denoted by r_f would tend to one, formally,

$$r_f = E\Big[\mathrm{NSE}\big(W_i(t), \langle w_{i'}(t)\rangle\big)_i\Big] \to 1 \tag{3.12}$$

where E is the expectation of the NSE evaluated at every cell i. Larger deviations of r_f from its ideal value (i.e. 1) would indicate that the global parameters obtained with a given regionalization technique can not be transferred to other modeling scales.

3.5.3 Preservation of Spatial Patterns

The spatial similarity of two fields at two different scales can be estimated with the averaged spatial correlation coefficient r_s given by

$$r_s \equiv r_s\big(\mathbf{x}_j^{(\ell_0)}, \mathbf{x}_i^{(\ell_1)}\big) \quad \forall j \in i \quad \forall i \in \Omega \tag{3.13}$$

where $\mathbf{x}_j^{(\ell_0)}$ and $\mathbf{x}_i^{(\ell_1)}$ two fields at scale ℓ_0 and ℓ_1, respectively.

3.6 Analysis of Results

3.6.1 Sensitivity of Model Efficiency to the Sub-grid Variability

The results of the numerical experiment obtained with Algorithm 2 support the hypothesis that the sub-grid variability of the basin predictors played an important role both in the prediction of daily streamflow and the spatial distribution of water fluxes. For instance, the RMSE between observed and simulated streamflow increased by 12 %, if the ℓ_0 resolution varied from 100 m to 2000 m, with a fixed modeling scale of ℓ_1 = 2 km (Table 3.5). Correspondingly, the NSE and the coefficient of correlation r showed a decrement of 2 %.
The variability of the spatial distribution of soil moisture in the top soil layer, as well as the actual evapotranspiration and runoff, exhibited a reduction up to 40 % of $\bar{r}$ (Eq. (3.10)) with respect to their corresponding baseline simulation (Fig. 3.6). It is worth mentioning that MPR simulations in which $\ell_0 = \ell_1 = 2000$ m are equivalent to those determined with the SR method. As a result, input data with a resolution $\ell_0 = 100$ m was used for the remaining numerical experiments employing the MPR technique.

Table 3.5 Effect of the sub-grid variability ℓ_0 = (100, 500, 1000, 2000) m on three model efficiency statistics (RMSE, NSE, and r) obtained for the daily discharge simulation for a given modeling scale (ℓ_1 = 2000 m). These statistics were evaluated during the period from 1979.11.1 to 2001-10-31.

ℓ_0	ℓ_1	RMSE	NSE	r
m	m	mm/d	-	-
100	2000	0.38	0.88	0.94
500	2000	0.41	0.87	0.92
1000	2000	0.41	0.86	0.91
2000	2000	0.42	0.86	0.91

3.6.2 Sensitivity of Streamflow Simulations to Global Parameters

Results obtained with Algorithm 3 are presented in Table 3.6. Both statistics, RMSE and NSE, shown in this table were estimated with the daily streamflow simulations at the outlet of the basin (Gauge Nr. 10 in Fig. 3.5) for a given scale. Both parametrization methods were not significantly different from each other, with respect to the RMSE and the NSE, when the model was calibrated and evaluated at a given modeling scale as shown in Table 3.6 (values in boldface on the diagonal). In this case, MPR performed marginally better than SR (approximately 2 % for both statistics). This implies that there exist various levels of discretization and thus parameterizations that provide equally acceptable solutions for modeling streamflow. These results agreed with those obtained by the Distributed Model Intercomparison Project (DMIP) (*Reed et al.*, 2004) with respect to the performance of lumped versus distributed hydrologic models. The performance regarding other state variables, however, may be unacceptable as will be shown afterwards.

Deficiencies of the regionalization methods, nevertheless, became apparent when global parameters were shifted across modeling scales as can be appreciated in the off-diagonal values in Table 3.6. The performance of SR, as compared to MPR, showed a significant deterioration when the global parameters were calibrated for a coarser modeling scale (say ℓ_1 =8 km) and subsequently applied in a finer one ($\ell_1 = 2$ km) as shown in the upper triangular matrices of Table 3.6. The RMSE obtained with MPR and SR was, on average, 0.46 mm/d and 0.62 mm/d, respectively. This means that the error of the daily streamflow simulations obtained with SR was, on average, 34 % higher than that estimated with MPR. The NSE, as expected, exhibited the opposite relationship. Mean NSE, in this case, was 0.80 and 0.64 for MPR and SR, respectively. Moreover, the average reduction of NSE with respect to the baseline simulations (i.e. no shift of global parameters) was 53 % for SR, whereas, this statistic only exhibited 12 % reduction for MPR. It is worth noting that the regionalization technique proposed by *Troy et al.* (2008) also exhibited a significant deterioration in the VIC model performance when model parameters calibrated at coarser resolutions were applied to finer resolutions.

Shifting global parameters calibrated for a given modeling scale to its immediate lower one (e.g. from 8 km to 4 km), however, does not induce a significant decrease in the performance of MPR as compared to that of SR. The increment of the RMSE was, on average, 1.1% and 30.7% for MPR and SR, respectively (Table 3.6). The decrement of NSE for MPR was at most 3%, whereas, SR exhibited reductions in NSE up to 11%.

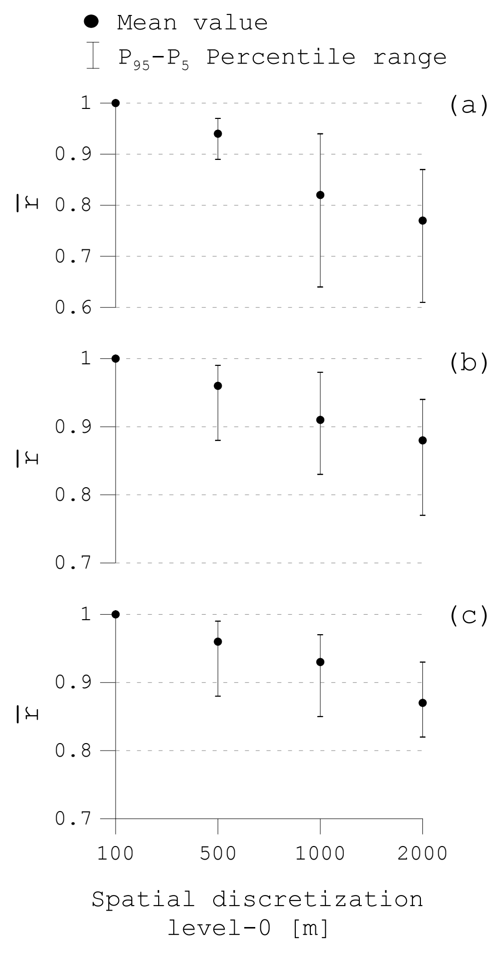

Figure 3.6 Mean and $P_{95} - P_5$ quantile range of the spatial correlation ($\bar{r}$) of (a) daily soil moisture, (b) daily total evapotranspiration, and (c) runoff between simulations obtained with a fixed modeling scale $\ell_1 = 2$ km but varying input data resolution $\ell_0 = (100, 500, 1000, 2000)$ m. The baseline values correspond to the simulations obtained with the smallest resolution. These statistics were estimated for the period from 1979-11-01 to 2001-10-31

Transferring global parameters calibrated for a finer modeling scale (say $\ell_1 = 2$ km) to a coarser one ($\ell_1 = 8$ km) was not so significant with both regionalization techniques as it was described above for the opposite case. In those cases, the RMSE obtained with MPR were only 1% lower than those obtained by SR. NSE estimated with both methods was, on average, 0.83 and 0.82 for MPR and SR, respectively (based on the lower triangular matrices of Table 3.6). The same tendency for both methods was observed when global parameters were

Table 3.6 RMSE and NSE obtained with both the multiscale (MPR) and the standard (SR) parameter regionalization techniques at various modeling scales obtained with daily streamflow values. Values on the diagonal (bold) refer to the statistics obtained for a model calibrated and evaluated at a given scale, whereas off-diagonal values denote statistics obtained with global parameters γ calibrated at other modeling scales. Results were obtained at Plochingen gauging station (basin outlet) during the period from 1979-11-01 to 2001-10-31.

Simulation Scale	Calibration Scale									
	2 km	4 km	8 km	16 km	32 km	2 km	4 km	8 km	16 km	32 km
	RMSE [mm/d]									
	MPR					**SR**				
2 km	**0.38**	0.38	0.45	0.53	0.72	**0.42**	0.46	0.64	0.78	0.90
4 km	0.35	**0.33**	0.35	0.41	0.57	0.38	**0.34**	0.46	0.63	0.77
8 km	0.39	0.39	**0.33**	0.36	0.49	0.41	0.37	**0.34**	0.46	0.63
16 km	0.46	0.47	0.39	**0.35**	0.38	0.47	0.46	0.40	**0.35**	0.45
32 km	0.54	0.55	0.50	0.42	**0.39**	0.53	0.56	0.50	0.42	**0.38**
	NSE [-]									
	MPR					**SR**				
2 km	**0.88**	0.87	0.82	0.76	0.56	**0.86**	0.81	0.64	0.46	0.29
4 km	0.89	**0.90**	0.89	0.85	0.73	0.87	**0.90**	0.81	0.65	0.48
8 km	0.86	0.87	**0.90**	0.89	0.80	0.85	0.88	**0.90**	0.81	0.66
16 km	0.82	0.82	0.87	**0.89**	0.87	0.81	0.81	0.86	**0.89**	0.82
32 km	0.75	0.75	0.78	0.84	**0.88**	0.75	0.72	0.78	0.84	**0.88**

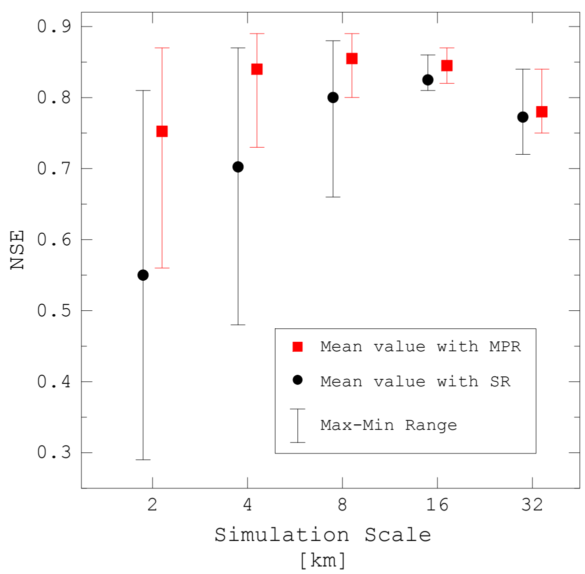

Figure 3.7 Mean and range of NSE obtained between observed and simulated daily streamflow simulations using MPR and SR at various modeling scales. Both statistics were determined with global parameters shifted from both coarser and finer scales to a given simulation scale. Simulations were carried out during the period from 1979-11-01 to 2001-10-31.

shifted to its immediately higher scale (e.g. from 4 km to 8 km.

The interpretation for this behavior of both methods is related with the amount of information used for regionalization. As a consequence, shifting global parameters from finer to coarse resolutions, for both methods, provide higher stability rather than the other way around. For the same reason, MPR performed significantly better than SR, when global parameters were shifted from coarser to finer modeling scales. Model parameters estimated with MPR, at any modeling scale, are intrinsically linked with their sub-grid variability (Fig. 3.3).

The average performance for both regionalization schemes at a given simulation scale obtained only with transferred global parameters is shown in Fig. 3.7. The average and the range of NSE depicted in this figure for a given scale were estimated based on their respective values of Table 3.6 (i.e. along rows). In general, model performance tends to increase as the simulation scale increases, regardless of the regionalization scheme (Fig. 3.7). However, there seems to be an upper limit, which for this study basin is around $\ell_1 = 8$ km. After this threshold scale is reached, further spatial discretization tends to decrease model performance. This behavior has also been noticed in the VIC model (*Liang et al.*, 2004).

The range of NSE for both regionalization schemes tends to decrease towards the threshold scale. SR exhibited larger ranges than MPR for all simulation scales, though. Furthermore, the mean of NSE obtained with MPR was, on average, 12% higher than that of SR (Fig. 3.7).

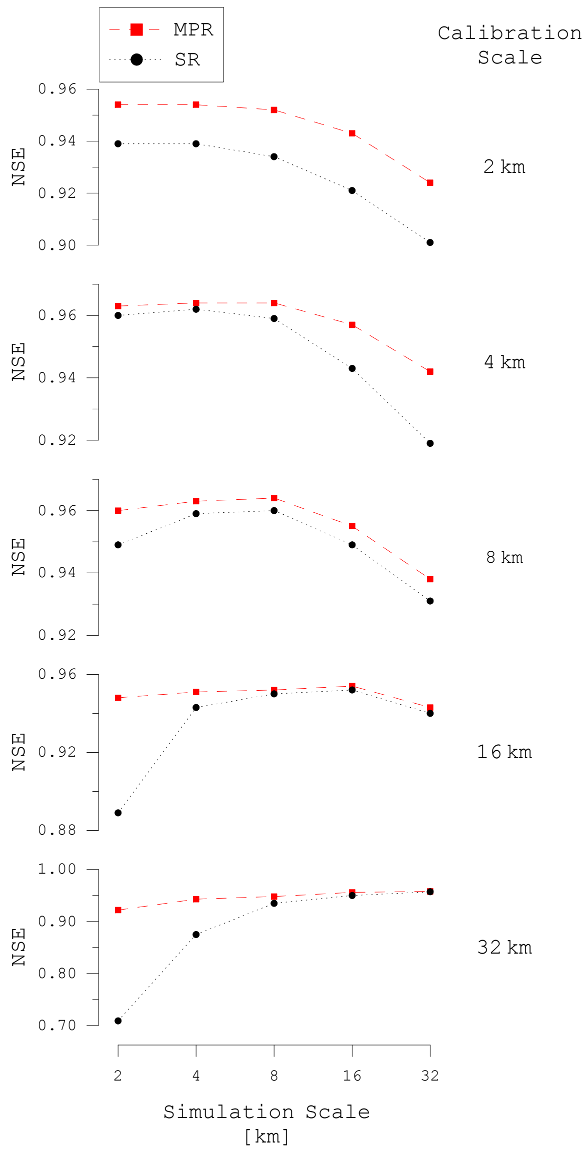

Figure 3.8 Sensitivity of the monthly water balance obtained by shifting global parameters from the calibration to the simulation scales for both regionalization schemes (MPR and SR). Baseline for the NSE are the monthly streamflow observations from 1979-11-01 to 2001-10-31.

The monthly water balance for both regionalization schemes exhibited similar behavior with respect to shifting global parameters as mentioned above (Fig. 3.8). The deterioration in performance, nevertheless, was not as higher as that estimated with daily streamflow simulations. The NSE between monthly observed and simulated streamflow during the period from 1979-11-01 to 2001-10-31 was at least 0.92 and 0.70 for MPR and SR, respectively. The specific long-term mean annual discharge at the outlet of the basin obtained with MPR and SR was quite close to the observed long-term mean of 407 mm/y for all spatial resolutions employed. The bias between the observed and simulated long-term annual discharge for both regionalization techniques, with or without shifting global parameters, was smaller than 7 mm/y.

3.6.3 Effects of the Sub-grid Variability on Model Parameters

Model parameters ($\boldsymbol{\beta}^1$) varied considerably depending on the regionalization method employed (Fig. 3.1), which denote the large degree of equifinality characterizing the model parameter space. A good example of this can be appreciated in Fig. 3.9, which depicts the spatial distribution of the porosity of the top ($k = 1$) soil layer β_6^1, obtained at various modeling scales [i.e. $\ell_1 = (2, 4, 8)$ km] with both parametrization methods. Comparison of the sub-grid distribution of β_6^0 ($\ell_0 = 100$ m) with the corresponding effective parameters β_6^1 obtained with MPR and SR respectively, showed that the MPR method preserved the spatial pattern significantly better than the SR method (Fig. 3.9). The spatial correlation coefficient r_s [Eq. (3.13)] between β_6 at level-0 and the corresponding field obtained with MPR at level-1 was, on average, 25 % greater than that obtained with SR.

The larger deviations observed in the SR technique can be mainly attributed to the upscaling mechanism (Fig. 3.1) described in Section 3.4.2. These, in turn, led to the emergence of significantly different spatial patterns. MPR, on the contrary, did not exhibit such large deviations because of the two step regionalization procedure which inherently accounts for sub-grid variability (Fig. 3.9).

3.6.4 Sensitivity of the Mass Balance to Global Parameters Calibrated at Various Modeling Scales

Two important water fluxes and one state variable were selected to carry out the continuity test: (a) actual evapotranspiration, (b) total discharge, and (c) the soil moisture of the top soil layer (depth 5 cm). Since no observations are available for any of these variables, simulated values obtained with global parameters calibrated at each control scale were used as a baseline for the estimation of NSE. Moreover, to ensure comparability, all simulations were driven by the same meteorological factors, which were estimated at the smallest modeling scale ($\ell_1 = 2$ km) and then aggregated to the required one. Deviations from the optimal value (i.e. no difference between fluxes or state variables) were quantified as indicated in Algorithm 3 with the statistic r_f [Eq. (3.12)] for every simulation and for both regionalization approaches independently.

The $P_{95} - P_5$ quantile range and the mean of NSE (r_f), between a simulated flux from a given control scale (ℓ_1) and the corresponding areal aggregated flux from a selected simulation scale (ℓ_1') are shown in Fig. 3.10 as an error bar with continuous line and solid circle, respectively. Each simulation, in this case, employed independently calibrated global parameters. For instance, if $\ell_1 = 4$ km and $\ell_1' = 2$ km are selected as control and simulation scales, respectively, then global parameters $\gamma^{(4)}$ and $\gamma^{(2)}$ need to be determined via calibration. Based on them, r_f between fluxes $W_i(t)$ and $\langle w_{i'}(t)\rangle$ (Eq.3.12) was estimated taking the former as baseline (Fig. 3.4). In the same Figure, dashed lines and empty circles depict also the $P_{95} - P_5$ quantile range and the mean of NSE, but in this case, fluxes at the finer scale ($\langle w_{i'}(t)\rangle$) were estimated with global parameters obtained at the control scale. For the previous example, this would imply that both fluxes $W_i(t)$ and $\langle w_{i'}(t)\rangle$ have to be estimated with $\gamma^{(4)}$.
It was determined based on these three variables that the soil moisture of the top soil layer exhibited the highest sensitivity to the spatial resolution ℓ_1, whereas the actual evapotranspiration (AET) was the less sensitive variable. These results also corroborated a previous study of *Liang et al.* (2004), in which the AET obtained with VIC-3L model was found to be less sensitive as compared to soil moisture and runoff based on simulations with transferred parameters.

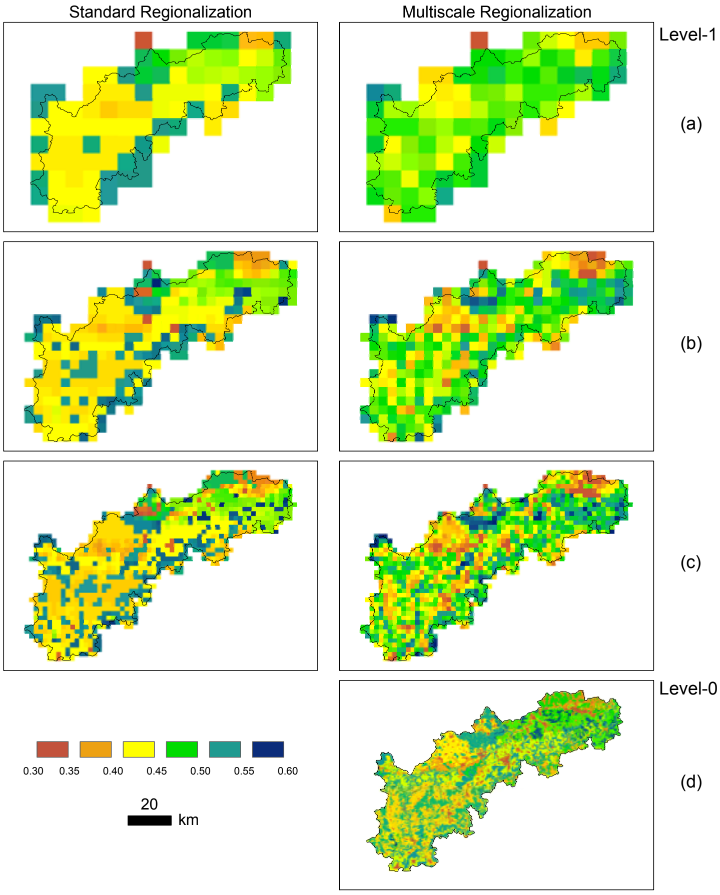

Figure 3.9 Spatial variability of the porosity (mm mm^{-1} of the top-soil layer (i.e. $\beta_6^{(1)}$) estimated at three different modeling scales, $\ell_1 = (2, 4, 8)$ km, and for both regionalization techniques (SR and MPR) is depicted in panels (a),(b), and (c) respectively. The porosity of the top soil layer at $\ell_0 = 100$ m (i.e. $\beta_6^{(0)}$) is provided in panel (d) as a reference.

SR exhibited systematic deficiencies as compared with MPR (Fig. 3.10). For instance, the NSE obtained with SR was, on average (r_f), not only much less than that obtained with MPR, but also exhibited a considerably larger range of variability as compared with MPR. Furthermore, MPR was less sensitive to the modeling scale than SR, specially when global parameters were calibrated at a given control scale (level-1) and then applied in other simulation scales (level-1').

It was also determined that r_f between soil moisture fields of the top soil layer obtained with the SR method was in most cases less than zero when the global parameters were calibrated at the control and simulation scales independently [panel (c) of Fig. 3.10]. For the same simulations, however, r_f was greater than 0.85 [panel (b) of Fig. 3.10], which indicates that the total discharge flux $Q(t)$ was quite insensitive to both the scale and the regionalization method employed.

The spatial variability of the NSE [Eq. (3.12)] depicted in Fig. 3.11 based on a simulation performed at $\ell_1' = 2$ km with global parameters obtained at the control scale $\ell_1 = 4$ km clearly shows the degree of influence that a regionalization technique may have on the dynamics and the mass balance of relevant water fluxes and state variables. This simulation indicated that global parameters obtained with SR are scale specific, thus not transferable, since almost 53 % of the grid cells did not conserve the mass balance (say NSE < 0.95) as compared with the 2.5 % of the grid cells with MPR. The locations at which the NSE is less than this threshold are mostly karstic formations (Fig. 3.5).

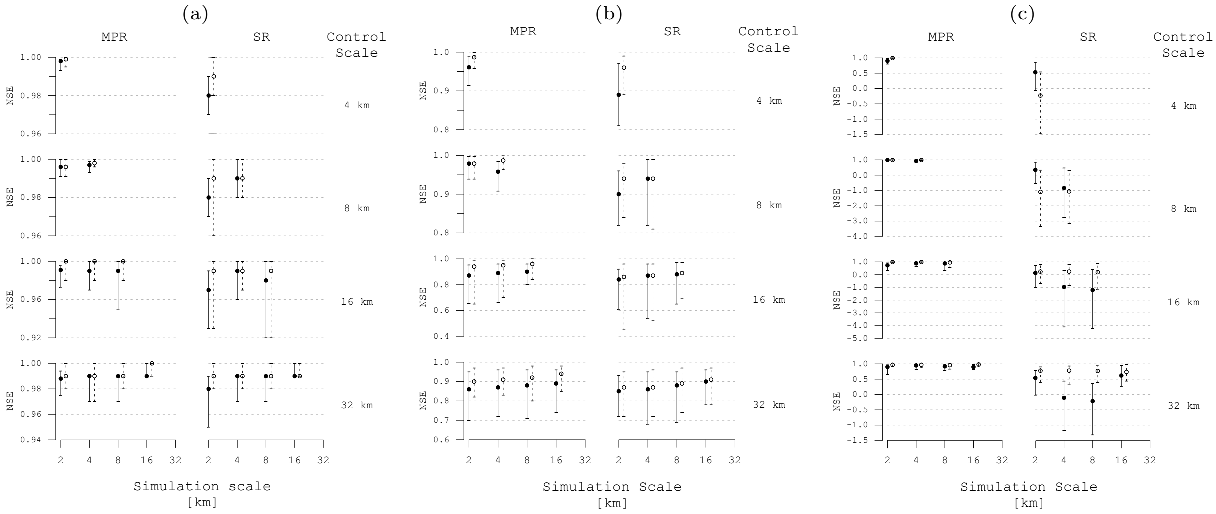

Figure 3.10 Evaluation of the conservation of mass for three water fluxes at various control scales based on MPR and SR parameterizations. Filled circles and continuous lines denote the mean and $P_{95} - P_5$ quantile range of the NSE [Eq. (3.12)] between the fluxes obtained at a given simulation scale compared with those obtained at the respective control scale (assumed as baseline for NSE). Empty circles and dotted lines indicate the same statistics but using global parameters obtained at the given control scale (i.e. no shift). From top to down: (a) actual evapotranspiration, (b) total discharge, and (c) the soil moisture of the top soil layer.

3.6.5 Preservation of Spatial Patterns

Daily time series of LST during the year 2000 were used as a proxy for the spatio-temporal variability of the top-layer soil moisture fields at the scale of $\ell_1 = 2$ km. A strong negative correlation between the volumetric water content of this layer and the LST (MODIS) is expected (*Chauhan et al.*, 2003; *Wang et al.*, 2007). As an example, a short sequence of LST and top layer soil moisture fields obtained with both regionalization methods is depicted in Fig. 3.12 to visualize the dynamics of these variables.
Based on these results, it was noticed that the soil moisture patterns calculated with MPR and the LST were closely related to each other. The dynamics of the moisture pattern obtained with SR, on the contrary, did not exhibit a strong dependence. As a result, the Spearman's rank correlation coefficient between the LST and the soil moisture of the top-layer obtained with the MPR and the SR technique varied between [-0.82,-0.57] and [-0.69,-0.07], respectively. The maximum difference observed between the Spearman's rho estimated with MPR and SR was as high as 41 %.

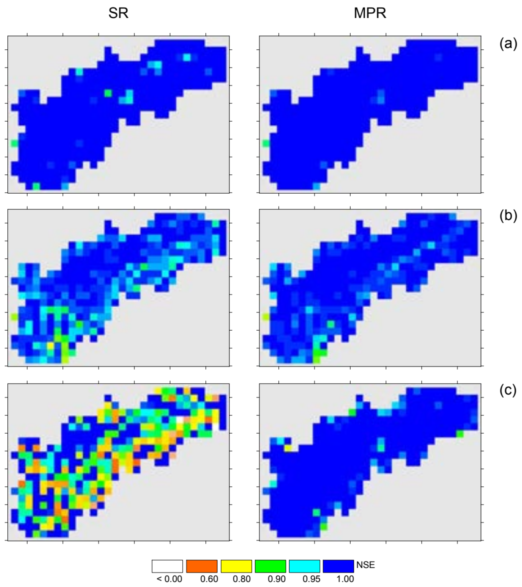

Figure 3.11 Discrepancy between fluxes simulated at two different modeling scales (the control scale $\ell_1 = 4$ km and the finer scale $\ell_1' = 2$ km) during the period from 1979-11-01 to 2001-10-31. The NSE was used as a measure of correspondence between fluxes simulated at the control scale – assumed as baseline – and the areal aggregation of fluxes obtained from the finer scale, as depicted in Fig. 3.4. Global parameters were estimated at the control scale. The spatial distribution of the NSE for daily evapotranspiration, total discharge, and soil moisture of the top soil layer is depicted in panels (a), (b), and (c), respectively.

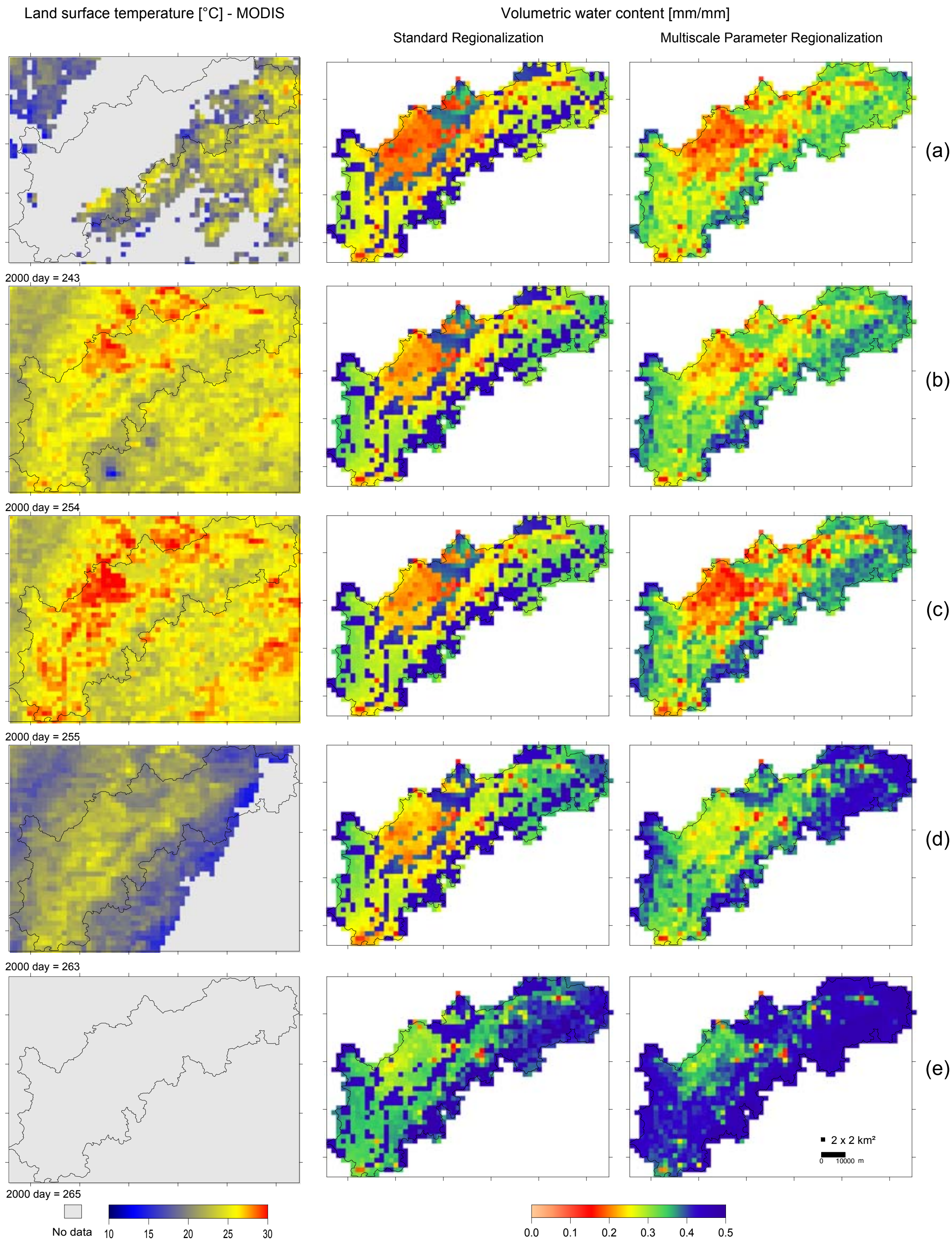

Figure 3.12 Land Surface Temperature (°C) from MODIS and simulated volumetric water content (mm mm^{-1}) in the top-soil layer estimated with SR and MPR. Both variables are depicted for various days during year 2000 at the modeling scale $\ell_1 = 2$ km.

3.6.6 Transferability of Global Parameters to Ungauged Locations

Ten gauging stations within the Upper Neckar basin were selected as cross-validation locations (Fig. 3.5) to test the efficiency of mHM to reproduce streamflow at internal locations. In these simulations, the modeling scale was set to $\ell_1 = 4$ km because this discretization covers all internal stations. Sets of global parameters for MPR and SR were obtained at this discretization as well as three coarser scales, namely: $\ell_1 = (8, 16, 32)$ km.

The NSE obtained for daily streamflow simulations during the evaluation period using MPR was, on average, 6 % greater that those obtained with SR (Fig. 3.13). This result corresponds to simulations obtained with global parameters estimated at the modeling scale of $\ell_1 = 4$ km. The median reduction of NSE for internal gauging stations with respect to the performance obtained at the outlet (gauge Nr. 10) was 15 % and 20 % for MPR and SR, respectively. Simulations obtained with global parameters calibrated at $\ell_1 = 4$ km were used as reference.
In the case that global parameters were estimated at the other modeling scales, then the NSE for SR was, on average, 18 % lower than that obtained with MPR (Fig. 3.13). Here, the median reduction was 16 % and 28 % for MPR and SR, respectively. This implies that MPR is more robust than SR for streamflow predictions at internal locations.

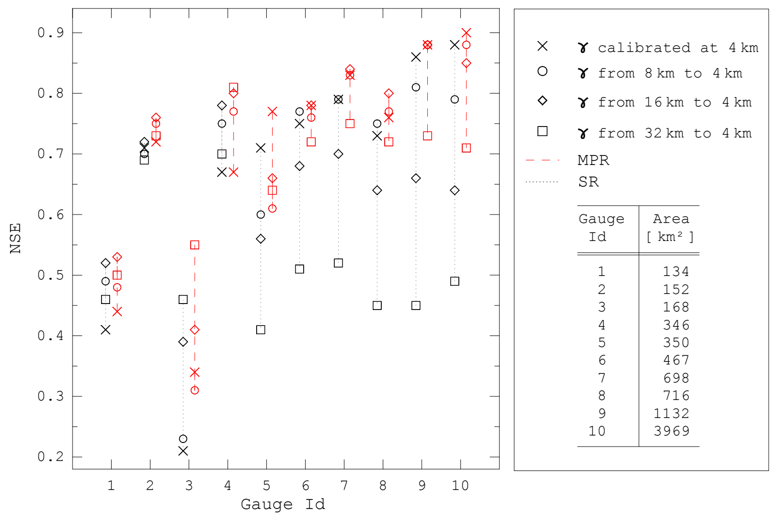

Figure 3.13 Performance of MPR and SR for daily discharge simulations on several internal locations within the upper Neckar basin, during the period from 1979-11-01 to 2001-10-31. Simulations for both methods were carried out at $\ell_1 = 4$ km using global parameters $\boldsymbol{\gamma}$ obtained from different scales: $\ell_1 = (4, 8, 16, 32)$ km.

Both methods, however, showed relatively poor performance for internal stations located within karstic formations (e.g. gauge Nr. 3, Fig. 3.5). For those locations, the NSE obtained with MPR was 62 % greater that that obtained with SR. These results, however poor, are still better than those reported in previous studies (*Gotzinger and Bárdossy*, 2007).

3.7 Conclusions

We have introduced in this paper a multiscale parameter regionalization technique (MPR) implemented within a fully spatially distributed conceptual hydrologic model (mHM) suited for research and operational purposes at the mesoscale. The framework of MPR is, however, not limited to mHM. It can be applied to any distributed hydrologic model.

The efficiency of the MPR technique was compared with the standard parameter regionalization (SR) procedure often used in recent literature. The hydrologic model and both regionalization techniques were applied in a river basin covering an area of 4000 km^2.

Results of this study indicated that both regionalization techniques do not exhibit significant differences in global efficiency measures (e.g. NSE of streamflow simulation) as long as the model is calibrated and evaluated at a given modeling scale (level-1). These results also pointed out the extent of the equifinality of global parameter sets and a substantial shortcoming of the calibration procedure when the objective function does not consider components other than observed and simulated streamflow.

Substantial differences in efficiency between MPR and SR, however, became apparent when the global parameters were calibrated at a coarser modeling scale and then transferred to a finer one. In such a case, MPR exhibited a clear superiority with respect to SR. As a consequence of this analysis, MPR can substantially reduce the amount of computational time required for model calibration starting from the premise that one can calibrate the global parameters at a coarser scale and then transfering them to a finer scale without having to recalibrate.

The dynamics of state variables such as soil moisture and water fluxes varied significantly depending on the regionalization method employed as well as the modeling scale used for the calibration of global parameters. Therefore, it is crucial for the evaluation of any regionalization technique, to assess the error induced into the mass balance, at a given control volume, when global parameters are shifted across scales.

Compelling evidence is presented in this study with respect to the effect of accounting for the sub-grid variability in the regionalization method, as well as the importance of the upscaling sequence (i.e. either predictors or parameters) to satisfy the continuity principle. Moreover, cross-scale experiments bring us to the conclusion that upscaling regionalized model parameters (i.e. MPR), instead of performing parameter regionalization with upscaled basin predictors (i.e. SR), lead to significantly different spatio-temporal distributions for both state variables and fluxes, due to the nonlinearity of the system.

Another advantage of MPR over SR is its capability to produce better streamflow simulations in cross-validated locations (assumed ungauged for testing), which is of great importance for the prediction in ungauged basins (PUB). Experiences with upscaling techniques lead us to the conclusion that there are no explicit simple averaging rules for the various model parameters. Further investigation regarding upscaling operators and their fundamental properties, required to describe dominant processes in a mesoscale control volume, is still needed. Finally, additional research is also required to address the effects of the various sources of errors within the MPR framework.

Acknowledgement

The authors would like to thank the Editor Scott W. Tyler, the Associate Editor Dennis P. Lettenmaier and three anonymous reviewers for their valuable suggestions and critique that substantially improved the quality of the original manuscript. We would like also to thank Steffen Zacharias for his advise on soil science issues.

Appendix: Main Equations of the Grid-based Model (mHM)

mHM simulates the spatio-temporal evolution of dominant hydrologic processes (Fig. 3.2) such as: canopy interception, snow accumulation and snowmelt, evapotranspiration, soil infiltration and freezing, surface and subsurface runoff generation, percolation, subsurface storage, base flow and a flow routing process. This model is driven by precipitation, temperature, potential evapotranspiration, and land cover. Its state equations $\mathbf{f}$ for a given cell i in time point t are **State equations:**

$$\left.\begin{aligned}
\dot{x}_{1i} &= P_i(t) - F_i(t) - E_{1i}(t)\\
\dot{x}_{2i} &= S_i(t) - M_i(t)\\
\dot{x}^k_{3i} &= (1-\rho^k) I_i^{k-1}(t) - E^k_{3i}(t) - I^k_i(t)\\
\dot{x}_{4i} &= \rho^1\big(R_i(t) + M_i(t)\big) - E_{2i}(t) - q_{1i}(t)\\
\dot{x}_{5i} &= I_i^2(t) - q_{2i}(t) - q_{3i}(t) - C_i(t)\\
\dot{x}_{6i} &= C_i(t) - q_{4i}(t)\\
\dot{x}_{7i} &= \beta_{26i} Q_i^1(t) + \beta_{26i}\beta_{27i}\left(Q_i^0(t) - Q_i^1(t)\right)
\end{aligned}\right\} \quad \forall i \in \Omega \tag{3.14}$$

The main governing equations of mHM are provided next (subindexes are omitted whenever possible to ease reading):

Response fluxes:

$$F(t) = \max\left\{P(t) + x_1(t-1) - \beta_1(t)\ ,\ 0\right\} \tag{3.15}$$

$$S(t) = \begin{cases} 0 & T(t) > \beta_2 \\ P(t) & \text{otherwise} \end{cases} \tag{3.16}$$

$$R(t) = \begin{cases} P(t) & T(t) > \beta_2 \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

$$M(t) = \begin{cases} \min\left\{x_2(t-1)\ ,\ \varphi(t)\big(T(t) - \beta_2\big)\right\} & T(t) > \beta_2 \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

$$I^k(t) = \big(1 - \rho^k(t)\big) I^{k-1}(t) \left(\frac{x_3^k(t-1)/d^k - \theta_r^k}{\beta_6^k - \theta_r^k}\right)^{\beta_7^k(t)} \tag{3.19}$$

$$E_1(t) = \left(\frac{x_1(t)}{x_1^{\max}(t)}\right)^{2/3} E_p \tag{3.20}$$

$$E_2(t) = \frac{x_4(t)}{\beta_{14}}\left(\frac{E_p(t)}{f_p(t)} - E_1(t)\right) \tag{3.21}$$

$$E_3^k(t) = \begin{cases} \alpha\, \beta_{17}^k \left(E_p'(t)\right) & k = 1 \\ \alpha\, \beta_{17}^k \left(E_p'(t) - \sum_{l=1}^{k-1} E_3^{k-1}(t)\right) & \text{otherwise} \end{cases} \tag{3.22}$$

$$q_1(t) = \max\{\rho(t)\big(R(t)+M(t)\big) + x_4(t-1) - \beta_{14}, 0\} \quad (3.23)$$

$$q_2(t) = \max\{I^2(t) + x_5(t-1) - \beta_{18}(z_2 - z_1)\,,\,0\}\beta_{19} \quad (3.24)$$

$$q_3(t) = \beta_{20}\,(x_5(t-1))^{\beta_{21}} \quad (3.25)$$

$$q_4(t) = \beta_{23}x_6(t-1) \quad (3.26)$$

$$C(t) = \beta_{22}x_5(t-1) \quad (3.27)$$

$$K(t) = \beta_{24}C(t) \quad (3.28)$$

$$Q(t) = \sum_{k=1}^{\delta} q(t-\delta+k)\mu(k) \quad (3.29)$$

$$Q_i^0(t) = Q_{i'}(t) + Q_{i'}^1(t) \quad (3.30)$$

$$Q_i^1(t) = Q_i^1(t-1) + \nu_1\left(Q_i^0(t-1) - Q_i^1(t-1)\right) + \nu_2\left(Q_i^0(t) - Q_i^0(t-1)\right) \quad (3.31)$$

$$\varphi(t) = \begin{cases} \beta_3(t) + \beta_4 R(t) & R(t) < \frac{\beta_5(t)-\beta_3(t)}{\beta_4} \\ \beta_5(t) & \text{otherwise} \end{cases} \quad (3.32)$$

$$I_i^0(t) = R(t) + M(t) \quad k=1 \quad (3.33)$$

$$d^k = z_1^k - z_1^{k-1} \quad (3.34)$$

$$\vartheta(t) = x_3^1(t)\big(1 - f_l(t)\big) \quad (3.35)$$

$$\rho_T(t) = P[\vartheta(t) > \beta_8(t)] = 1 - \int_0^{\beta_8(t)} f(\vartheta)d\vartheta = 1 - \frac{1}{\Gamma(\beta_9)}\int_0^{\upsilon} \xi^{\beta_9-1}e^{-\xi}d\xi \quad (3.36)$$

Auxiliary Variables / Constraints :

$$\upsilon = \frac{\beta_8(t)}{\theta} = \frac{\beta_8(t)}{\frac{E(\vartheta)}{\beta_9}} \approx \frac{\beta_8(t)\beta_9}{\vartheta(t-1)} \quad (3.37)$$

$$f_l(t) = \begin{cases} \beta_{12} & a_T \le \beta_{10} \\ \beta_{12} + \frac{1-\beta_{12}}{\beta_{11}-\beta_{10}}(a_T(t) - \beta_{10}) & a_T < \beta_{11} \\ 1 & \text{otherwise} \end{cases}$$

$$\alpha = \begin{cases} 0 & x_3^k(t-1) \le \beta_{15}^k \\ \frac{x_3^k(t-1)-\beta_{15}^k}{\beta_{16}^k - \beta_{15}^k} & \beta_{15}^k < x_3^k(t-1) \le \beta_{16}^k \\ 1 & \text{otherwise} \end{cases}$$

$$\rho^k(t) = 1 - \begin{cases} \big(1-\rho_T(t)\big)\big(1-\rho_U(t)\big) & k=1 \\ 1 & \text{otherwise} \end{cases}$$

$$a_T(t) = a_T(t-1) + \beta_{13}\big(T(t) - a_T(t-1)\big) \quad (3.38)$$

$$E_p(t) = \beta_{28}(t)E_p^* \quad (3.39)$$

$$E_p'(t) = E_p(t) - E_1(t) - E_2(t) \quad (3.40)$$

$$E_p \ge E_1 + E_2 + E_3 \quad (3.41)$$

$$1 = \sum_k \beta_{17}^k \quad (3.42)$$

$$1 = \sum_{k=1}^{\delta} \mu(k) \quad (3.43)$$

$$\delta = \frac{\beta_{25}}{\Delta t} - 1 \quad (3.44)$$

$$\nu_1 = \frac{\Delta t}{\beta_{26i}(1-\beta_{27i}) + \frac{\Delta t}{2}} \quad (3.45)$$

$$\nu_2 = \frac{\frac{\Delta t}{2} - \beta_{26i}\beta_{27i}}{\beta_{26i}(1-\beta_{27i}) + \frac{\Delta t}{2}} \quad (3.46)$$

$$\frac{1}{2(1-\beta_{27})} \le \frac{\beta_{26}}{\Delta t} \le \frac{1}{2\beta_{27}} \quad (3.47)$$

Notation

Inputs

P	Daily precipitation depth, mm d^{-1}.
E_p	Daily potential evapotranspiration (PET), mm d^{-1}.
T	Daily mean air temperature, °C.
R_s	Solar Radiation, W/m^2.
ATI	Antecedent temperature index, K.

States

x_1	Depth of the canopy storage, mm.
x_2	Depth of the snowpack, mm.
x_3	Depth of soil moisture content in the root zone, mm.
x_4	Depth of impounded water in reservoirs, water bodies, or sealed areas, mm.
x_5	Depth of the water storage in the subsurface reservoir, mm.
x_6	Depth of the water storage in the groundwater reservoir, mm.
x_7	Depth of the water storage in the channel reservoir, mm.

Fluxes

S	Snow precipitation depth, mm.
R	Rain precipitation depth, mm.
M	Melting snow depth, mm d^{-1}.
F	Throughfall, mm d^{-1}.
E_1	Actual evaporation intensity from the canopy, mm d^{-1}.
E_2	Actual evapotranspiration intensity, mm d^{-1}.
E_3	Actual evaporation from free-water bodies, mm d^{-1}.
E_p	Potential evapotranspiration, mm d^{-1}.
I	Recharge, infiltration capacity or effective precipitation, mm d^{-1}.
C	Percolation, mm d^{-1}.
q_1	Surface runoff from impervious areas, mm d^{-1}.
q_2	Fast interflow, mm d^{-1}.
q_3	Slow interflow, mm d^{-1}.
q_4	Baseflow, mm d^{-1}.
K	Gain/loss flux in a leaking linear reservoir, mm d^{-1}.

Output

Q_i^0	Simulated discharge entering the river stretch at cell i, m^{3}s^{-1}.
Q_i^1	Simulated discharge leaving the river stretch at cell i, m^{3}s^{-1}.
$Q_{i'}$	Contribution from the upstream cell i.
Q	Hydrograph at the outlet of a grid cell.

Other variables

Δt	Simulation time interval.
z_i	Depth of sub-surface layer i, $i = 1, \ldots, 3$.
z_1^k	Depth of root-zone horizon k, $k = 1, \ldots, \lambda$.
x_3^k/d^k	Average water content in the k root-zone horizon, m^3 m^{-3}.
ρ^k	Overall influx fraction accounting for the impervious cover within a cell.
$\rho_T(t)$	Portion of the permeable areas.
$\rho_U(t)$	Fraction of impervious surfaces in a grid cell.
$\varphi(t)$	modified degree-day factor, mm d^{-1} $^\circ$C^{-1}.
θ_r^k	Residual soil moisture content.
$\vartheta(t)$	Soil ice content (i.e. water equivalent) of the first root zone horizon.
θ	Scale factor of the distribution $f(\vartheta) \sim \Gamma(\xi, \beta_9)$, with $\xi = \frac{\vartheta}{\theta}$.
$E(\cdot)$	Expectation of the soil ice content on a given cell.
f_l	Fraction of unfrozen water in the first root zone layer.
a_T	Antecedent temperature index (ATI).
$\mu(\beta_{25}, \Delta t)$	Triangular unit hydrograph.
i_o	Outlet of every cell i.
κ_j	Saturated hydraulic conductivity based on *Campbell and Shiozawa* (1994) pedotransfer function.
o_j	Fraction of organic matter.
ϱ_o	Average organic matter bulk density (= 0.224 g/cm^3) (*Rawls*, 1983).
ϱ	Bulk density according to Adams (*Rawls*, 1983).
f_s	Sand fraction threshold according to *Zacharias and Wessolek* (2007) (= 66.5 %).
ψ	Adjustment factor (*Chow et al.*, 1988).
Υ	Time of concentration (similar to Kirpich's Eq.) of each cell $j \in i$ along the drainage path $\{j \to i_o\}$.
i_o	Outlet of cell i at level-0.
E_p^*	PET estimated by a standard method for a horizontal surface e.g. *Hargreaves and Samani* (1985).

Parameters

β_1	Effective maximum canopy storage, mm.
β_2	Threshold temperature for phase transition snow/rain, $^\circ$C.
β_3	Degree-day factor during rainless days.
β_4	Rate of increase of the degree-day factor per unit of precipitation, d^{-1} $^\circ$C^{-1}.
β_5	Maximum degree-day factor reached during rainy days.
β_6^k	Maximum soil moisture content.
β_7	Parameter that determines the relative contribution of rain or snowmelt to runoff.
β_8	Critical value of soil ice content above which the soil is practically impermeable.
β_9	Shape factor of the distribution $f(\vartheta) \sim \Gamma(\xi, \beta_9)$.
β_{10}	ATI threshold below which unfrozen water content reaches its minimum.
β_{11}	ATI threshold above which no frozen water exist.
β_{12}	Minimum fraction of unfrozen water content.
β_{13}	Weighting multiplier ranging from 0.1 to 1.
β_{14}	Maximum ponding retention in impervious areas.
β_{15}	Permanent wilting point.
β_{16}	Soil moisture limit above which the actual transpiration is equated with the PET.
β_{17}^k	fraction of roots in the k^{th} horizon.
β_{18}	Maximum holding capacity of the second reservoir (unsaturated zone).
β_{19}	Fast-recession constant.
β_{20}	Slow-recession constant.
β_{21}	Exponent that quantifies the degree of nonlinearity of the cell response.
β_{22}	Effective percolation rate.
β_{23}	Baseflow recession rate.
β_{24}	Fraction of the groundwater recharge that might be gained or lost either as deep percolation or as intercatchment groundwater flow in nonconservative catchments.
β_{25}	Duration of the TUH.
β_{26}	Muskingum travel time parameter.
β_{27}	Muskingum attenuation parameter.
β_{28}	Aspect correction factor of the PET.

Indices

k	Index denoting the root zone layer, $k = 1, 2$.
t	Time index for each Δt interval.

CHAPTER 4

IMPLICATIONS OF PARAMETER UNCERTAINTY ON SOIL MOISTURE DROUGHT ANALYSIS IN GERMANY

L. Samaniego[1], R. Kumar[1], and M. Zink[1]

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Samaniego, L., R. Kumar, and M. Zink (2013), Implications of Parameter Uncertainty on Soil Moisture Drought Analysis in Germany, Journal of Hydrometeorology, 14(1), 47–68

Abstract

Simulated soil moisture is increasingly used to characterize agricultural droughts but its parametric uncertainty, which essentially affects all hydrological fluxes and state variables, is rarely considered for identifying major drought events. In this study, a high-resolution, 200-member ensemble of land surface hydrology simulations obtained with the mesoscale Hydrologic Model is used to investigate the effects of the parametric uncertainty on drought statistics such as duration, extension, and severity. Simulated daily soil moisture fields over Germany at the spatial resolution of 4×4/,km^2 from 1950 to 2010 are used to derive a hydrologically consistent soil moisture index (SMI) representing the monthly soil water quantile at every grid cell. This index allows a quantification of major drought events in Germany. Results of this study indicated that the large parametric uncertainty inherent to the model, did not allow discriminating major drought events without a significant classification error. The parametric uncertainty of simulated soil moisture exhibited a strong spatio-temporal variability, which significantly affects all derived drought statistics. Drought statistics of events occurring in summer with at most six months duration were found to be more uncertain than those occurring in winter. Based on the ensemble drought statistics, the event from 1971 to 1974 appeared to have 67% probability of being the longest and most severe drought event since 1950. Results of this study emphasize the importance of accounting for the parametric uncertainty for identifying benchmark drought events as well as the fact that using a single model simulation would very likely lead to inconclusive results.

4.1 Introduction

Drought is a recurrent and extensive climatic phenomenon characterized by below-average water availability whose duration might last for several years. It is considered as one of the most costly natural disasters because it often induces huge socio-economic losses (*Wilhite*, 2000) as well as environmental degradation. During the summer of 2003, for instance, several parts of Europe endured the highest temperatures of the last 500 years (*Fink et al.*, 2004; *Luterbacher et al.*, 2004) and one of the most extensive and severe drought in records. In Germany alone, the estimated loss in the agricultural sector was 1.5 billion Euros (*COPA-COGECA*, 2003). In extreme cases, prolonged drought spells might lead to unprecedented environmental disasters often associated with the decline of human societies (*Haug et al.*, 2003; *Hodell et al.*, 1995) or the trigger for mass migrations and famine (*Field*, 2000). Droughts occur indifferently in high and low rainfall areas and in virtually all climatic zones (*Dracup*, 1991; *Mishra and Singh*, 2010), although the most severe human consequences happen in arid regions.

Currently, hydro-meteorologic mechanisms originating droughts are relatively well understood. In general, droughts are driven by extreme macroclimatic variability originated by atmospheric interactions and feedback between the atmosphere, the oceans, and the land surface (e.g. *McCabe and Palecki*, 2006; *Nicholson*, 2000). This variability is, in turn, related to the solar activity as well as atmospheric composition, and strongly affected by anthropogenic activities (*Sheffield et al.*, 2009).

Our ability to make reliable drought predictions, however, is not satisfactory (*Wilhite*, 2000) although there is vast scientific literature on this topic. One of the main reasons is related to the insufficient knowledge regarding the processes controlling drought development and persistence, as well as, its spatio-temporal variability (*Sheffield et al.*, 2009). Another reason stems from the fact that there is no clear definition of this phenomenon (*Wilhite and Glantz*, 1985) since it depends upon the variable that is used for its quantification.

Droughts have been mainly classified into three types: (1) meteorological drought, usually defined as an extreme anomaly of precipitation; (2) hydrological drought, which is related to a deficit in the supply of surface and subsurface water, and (3) agricultural drought, being a combination of meteorological and hydrological droughts leading to deficits in root zone soil moisture available to vegetation (*Wilhite and Glantz*, 1985). Since precipitation and discharge data are widely available, a plethora of drought indices have been proposed in the scientific literature to quantify meteorological and hydrological droughts, for instance: the Palmer Drought Severity Index (*Palmer*, 1965), the Standard Precipitation Index (*McKee et al.*, 1993), the Regional Deficiency Index (*Stahl and Demuth*, 1999), among others.

It is widely accepted, however, that these empirical indices are not adequate to represent extreme water stress conditions that would lead to a significant reduction of biomass and crop yield (*Keyantash and Dracup*, 2002; *Mishra and Singh*, 2010). In Germany, for example, *Döring et al.* (2011) have shown that empirical drought indices based only on available data such as precipitation, temperature do not constitute adequate measures to describe agricultural drought stress because they do not explicitly account for the available water stored in the root zone, which is ultimately the plant's life supporting substance.

Direct soil moisture observations, on the other hand, are not available at regional level because measuring this variable at large scales is not logistically and economically feasible (*Vereecken et al.*, 2008). This implies that hydrologic or land surface models would have to be employed for the estimation of the soil water content. Soil moisture, in contrast to precipitation or discharge, constitutes a good index for quantifying agricultural drought because it controls the proportion of the rainfall that percolates, runs off or evaporates from the earth surface (i.e. root zone). Concisely, it integrates precipitation and evapotranspiration as well as the delays introduced by interception, snow accumulation, and melting over periods of days to weeks. In other words, soil moisture in the root zone is a governing factor sustaining vegetative growth and thus it is a direct indicator of agricultural drought (*Keyantash and Dracup*, 2002). Land surface models such as VIC-3L (*Liang et al.*, 1996a) and SIM (*Soubeyroux et al.*, 2008), for example, have been used recently to assess agricultural drought characteristics in the USA and France, respectively (*Andreadis et al.*, 2005; *Sheffield et al.*, 2004; *Vidal et al.*, 2010). There are, however, several key issues that should be considered, if simulated soil moisture is chosen for quantifying agricultural droughts.

Modeling soil moisture dynamics at large-scales (e.g. grid cells greater than 500 m) is difficult and uncertain as was demonstrated by the PILPS project (*Chen and Coauthors*, 1997). In this project, 23 land surface models (LSMs) exhibited significant differences between modeled and measured soil moisture (among other variables) although all models were based on fundamental principles of mass and energy conservation and forced with identical atmospheric conditions. This experiment also indicated the existing interplay between this state variable and other fluxes such as latent heat as well as the substantial parameter uncertainty that is related with these physical processes. At larger scales, the sub-grid variability of the variables involved and the nonlinearity of the processes make the modeling of soil moisture even more complicated because parametrization schemes might become scale dependent (*Nykanen and Foufoula-Georgiou*, 2001). It should be noted that effective model parameters (e.g. saturated soil water content or porosity) at large scales can only be estimated but not measured. This, in turn, constitutes a new source of uncertainty that should be taken into account when modeling soil moisture dynamics. Consequently, a drought monitoring and early warning system based on a soil moisture index, which does not fully take into account the predictive uncertainty of the simulation model, might be inadequate for real applications and/or for impact assessment.

Most of the soil moisture drought studies (*Andreadis and Lettenmaier*, 2006; *Shukla et al.*, 2011; *Vidal et al.*, 2010) found in the literature have not addressed the epistemic uncertainty related to parametrization, model structure, and input data. More recently, *Wang et al.* (2011) argued that state variables, such as soil moisture, are strongly dependent on the parametrization of the LSMs and the quality of the meteorological forcing data. Similar results have been found by *Mo et al.* (2012a), who concluded that the primary source of uncertainty between two drought monitoring systems operated in the USA is originated from precipitation data, and in a minor degree from air temperature, shortwave and longwave radiation, and wind speed. As a result, substantial discrepancies with in-situ measurements have been found (*Entin et al.*, 2000), which are mainly attributed to the variability of topography, soil, vegetation, and root structure, but could also stem from uncertainty sources mentioned above. Specifically, finding a robust parametrization scheme for a LSM or a hydrological model, which is able to produce reliable estimates of water fluxes at high spatial resolution over large domains, is one of the grand challenges of contemporary hydrology (*Beven and Cloke*, 2011).

It has been noted, however, that multi-model ensembles are able to describe the anomalies and seasonal variability of soil moisture. *Wang et al.* (2009, 2011) successfully applied this technique to reproduce agricultural drought characteristics in the continental United States and China. In both studies, six LSMs were used to generate soil moisture fields for a period of almost 100 years in the USA and 56 years in China. However, in those studies, only a single simulation for each LSM was used.

In this study, we argue that a unique parameter set for a given LSM is inadequate to estimate water fluxes and related state variables at high spatio-temporal resolutions, considering that both inputs and model parameters over large modeling domains are subject to considerable uncertainties due to the reasons mentioned above (see also *Rosero et al.*, 2011). Thus, we hypothesize that any drought characteristic (e.g. severity, duration) based on simulated soil moisture is prone to large variability due to parametric uncertainty, which, if it is not taken into account, will lead to incorrect estimates of drought characteristics.

The main objectives of this study and the rationale behind them is summarized below. 1) To obtain a consistent ensemble of daily soil moisture fields for Germany since 1950 at a spatial resolution of 4×4 km. Such reconstruction is fundamental to characterize historical drought events and their related characteristics. To the best of our knowledge, this is the first study to perform nationwide agricultural drought reconstruction for Germany. Long-term soil moisture simulations are also fundamental for initializing hydrologic or regional climate models and the basis to fulfil the remaining objectives. 2) To develop a reliable soil moisture drought index (SMI) for Germany at a high spatial resolution. Such SMI is key for implementing a monitoring system and adaptation strategies at regional scale. Available global soil moisture analyses have a spatial resolution 0.5° or larger, which is too coarse for a regional drought analysis. 3) To identify benchmark agricultural drought events occurring in summer and winter in Germany during the last 60 years and the uncertainty of their main statistical characteristics. These exceptional events are necessary to identify potential climate change effects on the hydrological cycle. The uncertainty associated with drought characteristics such as coverage area, duration, and severity, will be quantified by means of a Monte Carlo method. Ensemble model simulations would allow us to assess the reliability of the predictions which, in turn, will lead to minimize the number of false positive drought events (i.e. cases in which the SMI indicates that a given event is below a certain threshold for a given characteristic when in fact it is not). Additionally, the effect of the ensemble size on the false positive rate will be investigated. 4) To identify regions in Germany prone to strong drought persistence as well as areas exhibiting significant trends in monthly soil moisture fields. These insights would provide hints for mitigation and adaptation measures at regional scale.

4.2 Soil Moisture Data

Soil water availability in the root zone is a direct indicator of agricultural drought because it constitutes a governing factor of the state of vegetative growth through the availability of water for transpiration (*Keyantash and Dracup*, 2002). Measuring soil moisture content over the entire domain of Germany at a spatial resolution of 4 × 4 km, for example, is logistically and economically infeasible (*Vereecken et al.*, 2008). LSMs or hydrologic models are therefore often employed to estimate this key variable over large spatial domains and longer periods (*Andreadis and Lettenmaier*, 2006; *Mishra et al.*, 2010; *Sheffield and Wood*, 2007; *Wang et al.*, 2009, 2011).

In this study, the mesoscale Hydrologic Model, mHM (*Samaniego et al.*, 2010a) was used to generate a large ensemble of daily soil moisture fields for the period from 1950 to 2010. A three layer soil scheme was used to model the soil moisture dynamics over the entire root zone depth (i.e. approximately up to 2 m below ground). The depth of the first two layers was fixed to 5 cm and 25 cm, whereas the depth of the last one was variable according to soil characteristics provided by the soil texture map. The spatial resolution of each grid was 4×4 km (level-1). A short description of mHM and the generation of ensemble soil moisture fields are given below.

4.2.1 The mHM

The mesoscale Hydrologic Model is a process-based water balance model (*Samaniego et al.*, 2010a) that has been developed over the last five years at the Helmholtz Centre for Environmental Research - UFZ. This spatially explicit model does not differ significantly from existing large scale hydrologic models (e.g. the HBV and the VIC-3L model) on how dominant hydrologic processes at the meso- and macro-scales are conceptualized, but on how the effective parameters of the model are quantified at a selected modeling scale and on how the sub-grid variability of physiographic characteristics provided at level-0 is taken into account for the estimation of these effective parameters. These two fundamental differences constitute the core of the multiscale parameter regionalization technique (*Samaniego et al.*, 2010a) that is embedded into mHM. Extensive numerical experiments have shown that this technique is capable of coping with the large spatio-temporal variability of the input data and as a result, mHM is able to produce quite good performance at multiple spatial resolutions and locations other than those used during calibration (i.e. proxy basin and flux-matching tests).

Currently, mHM has been evaluated in more than one hundred basins in Germany ranging from 4 km^2 to 47 000 km^2 (*Kumar et al.*, 2010, 2013; *Samaniego et al.*, 2010a). This model is driven by disaggregated fields of daily forcings such as precipitation, temperature, and potential evapotranspiration. It accounts for the following hydrological processes: canopy interception, snow accumulation and melting, evapotranspiration, infiltration, soil moisture dynamics in three layers, surface runoff, subsurface storage, discharge generation, percolation, baseflow, and flood routing within the river reaches. Readers may refer to *Samaniego et al.* (2010a) for a detailed model description as well as its parametrization.

The morphological and physiographic data required for setting up mHM include a digital elevation model (50×50 m) acquired from the Federal Agency for Cartography and Geodesy, a vector soil map containing information on soil textural properties such as sand and clay contents of different soil horizons, and a vector map of hydro-geologic formations containing properties such as saturated hydraulic conductivity. Both vector maps at a scale of 1:1 000 000 were obtained from the Federal Institute for Geosciences and Natural Resources of Germany. Three Corine land cover seamless vector data (http://www.eea.europa.eu) for the years 1990, 2000, and 2006 were employed to account for the changes in states of land cover over the simulation time period (1950-2010). Land cover states, prior to the year 1990, were inferred from the Corine 1990 map. Monthly variability of the leaf area index was estimated for each land cover class with MODIS scenes from 2001 to 2009. These data are freely available from https://lpdaac.usgs.gov/get_data. For a detailed description on data processing and setting up mHM in several river basins, interested readers may refer to *Kumar et al.* (2010); *Samaniego et al.* (2010a). Previous data sets were re-sampled on a common spatial resolution of 100×100 m denoted as level-0. This level of information provides the sub-grid variability of all morphological and physiographic variables required to run the model at any coarser resolution denoted as level-1 (e.g. 4 km). The time series of discharge data across several gauging stations were acquired from the EURO-FRIEND program (http://ne-friend.bafg.de) and the Global Runoff Data Centre (http://www.bafg.de).

Gridded fields of daily average precipitation as well as maximum, minimum, and average air temperatures at 4×4 km spatial resolution (level-2) were estimated from their respective point measurement data from about 5600 rain gauges and 1120 meteorological stations, operated by the German Meteorological Service (DWD). Two interpolation techniques were used to derive the daily fields of precipitation, which are detailed in section 4.2.2. Gridded estimates for temperature fields were obtained with external drift kriging, wherein the terrain elevation was used as a drift variable. The daily fields of potential evapotranspiration were estimated with the Hargreaves and Samani method (*Hargreaves and Samani*, 1985) and were subsequently corrected to account for the spatial variability of the terrain aspect.

4.2.2 Ensemble description and experimental design

Two major sources of parametric uncertainty were identified through sensitivity analysis. The most important one is related with the variability of the global calibration parameters of mHM (i.e. space and time independent), and the second one is related with the parameters required for the rainfall interpolation method. Consequently, the uncertainty tree was divided into two main branches, each one driven by two independent interpolation methods but both based on the same rainfall measurements. These two branches were denoted as DWD1 and DWD2. Other meteorological variables such as daily, minimum, and maximum temperature required in both branches were kept the same. This assumption was taken considering 1) that precipitation interpolation is one the most important source of error in the input data (*Mo et al.*, 2012a), and 2) that the areal coverage of snow-dominated areas in Germany is geographically limited.

The DWD1 branch was created with external drift kriging using terrain elevation as a drift and a combined variogram that comprised a nugget and an exponential part. The resolution of this product was 4×4 km, with daily time steps from 1950 to 2010. The best fit parameters (i.e. nugget, range, and sill) were found through a cross-validation procedure.

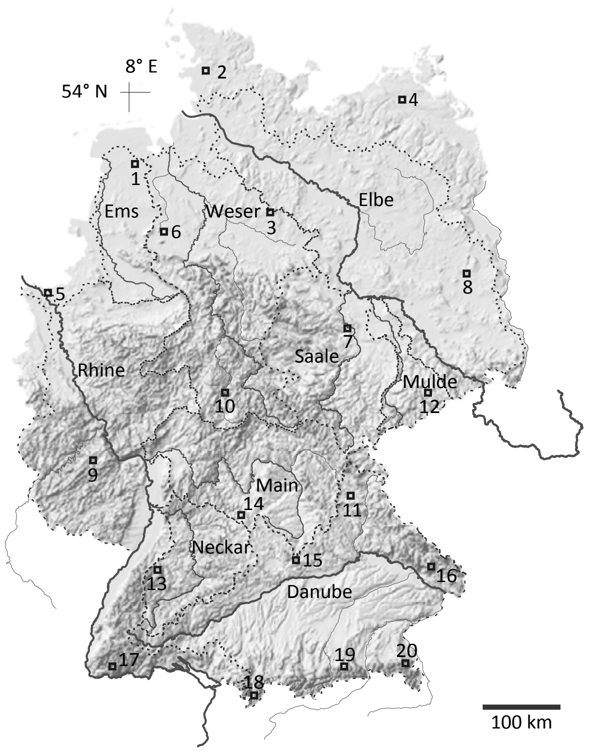

Figure 4.1 Map of Germany indicating the main river basins used for this study. Selected locations for uncertainty analysis of the soil moisture climatology are depicted with a dot.

The DWD2 branch was obtained by re-sampling the original daily REGNIE (www.dwd.de) product available at 1×1 km into a regular grid similar to that of the DWD1 data set. The k-nearest's neighbor technique and a standard geo-referencing algorithm were employed for this purpose. The DWD1 data was used to complete this set with daily fields from 1950 to 1959 since the REGNIE data set is only available from 1960 to 2010. The REGNIE data is based on multiple linear regression having elevation, geographic location, and aspect as predictors.

Within each branch, the propagation of the parameter uncertainty into the soil moisture simulations was evaluated by an ensemble of one hundred best parameter sets of mHM. The following procedure was implemented for their selection. First, in every major river basin depicted in Fig. 4.1, the dynamically dimensioned search algorithm (*Tolson and Shoemaker*, 2008) was employed to find good sets of global parameters which exhibit an acceptable model efficiency [e.g. Nash-Sutclife-Efficiency of at least 0.75] during the evaluation period (for details refer to *Kumar et al.*, 2010, 2013). In the next step, all parameter sets found for a given basin were transferred to the remaining ones. Finally, only those sets exhibiting a model efficiency greater than or equal to 0.65 at recipient locations were retained as members of the best global parameter sets. This implies that these super sets of global parameters are able to reproduce water fluxes in all major river basins in Germany with an efficiency of at least 0.65. It may be noted that a single set of VIC-3L model parameters for a large domain in the midwestern United States was used in a study by *Mishra et al.* (2010) for assessing historical drought events. In contrast to that, in this study the ensemble of 200 model realizations was used for the subsequent analysis of historical drought events in Germany including both uncertainty branches.

In general, mHM requires at least five years of spin-up time to equilibrate. To minimize the influence of initial conditions, all state variables (e.g. water content at a given soil layer) in each ensemble member were initialized with their climatological averages corresponding to the precise time of year at the initialization (*Rodell et al.*, 2005). The climatological average was estimated as the long term mean of a given state variable within a seven-day window around the first of January. The DWD1 precipitation estimate was employed to estimate the long term mean. This procedure allowed to reduce the spin-up time to one year without inducing large bias due to inappropriate initial conditions. Thus, model simulations during the starting year 1950 were discarded from the following analysis.

4.3 The mHM Soil Moisture Index

The absolute values of the soil moisture states estimated with mHM do not allow a direct comparison of derived drought indices across the study domain because anomalies in absolute terms reflect climatological and morphological characteristics (*Andreadis et al.*, 2005), rather than strong deviations from the respective normal conditions, which is the main characteristic that defines a drought event. Instead of absolute values, agricultural droughts can be quantified as "*deficit of soil moisture relative to its seasonal climatology at a location*" (*Sheffield et al.*, 2004). The main idea behind this definition is to develop an index that varies between 0 and 1, which indicates drier to wetter conditions, respectively. The apparent selection for such an index is the conditional cumulative distribution function of the soil water content in the root zone at a given location i and time of the year m. This kind of normalization is inspired by the Standardized Precipitation Index (*McKee et al.*, 1993). The procedure to estimate a Soil Moisture Index (SMI) based on mHM soil moisture simulations is described next.

4.3.1 Aggregation and normalization

Daily mHM soil moisture from three soil layers was averaged for every grid cell to obtain monthly states. These monthly values were, in turn, normalized with respect to the corresponding total root zone saturated water content (i.e. porosity times the total depth of the soil layers) to estimate the monthly soil moisture fraction (x) of the total soil column, namely:

$$x = \frac{\sum_l x^l}{\sum_l x_S^l} \tag{4.1}$$

where, x^l is the monthly soil moisture at root zone layer l [mm], x_S^l is porosity or the saturated water content of root zone layer l [mm]. In the present study $l = 3$. In this case, the indexes i and m are omitted to ease the notation.

4.3.2 Estimation of the SMI

The monthly soil moisture fraction (Eq. 4.1) may exhibit heavily skewed, non-gaussian distributions (*Koster et al.*, 2009) whose shape varies depending on climate and soil characteristics. The distribution of this random variable can also be multi-modal (*Vidal et al.*, 2010), which is an indication of preferential states of seasonal soil moisture (*Laio et al.*, 2002; *Rodriguez-Iturbe et al.*, 1991). Consequently, describing this random variable with unimodal theoretical distribution [e.g. the beta distribution (*Sheffield et al.*, 2004)] is not appropriate. Instead of making assumptions regarding the theoretical distribution of this variable, which would induce an additional source of uncertainty, a non parametric technique was adopted to estimate the probability density function of the monthly soil moisture fraction at every cell within the domain, denoted hereafter as $\hat{f}(x)$. The estimation procedure is as follows.

Given a set of data from one of the ensemble members $x_1, x_2, \ldots, x_n$ that corresponds to the monthly soil moisture fractions of a given cell within the domain during month m (e.g. January), the kernel density estimate at a given value x can be obtained by

$$\hat{f}(x) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{x - x_k}{h}\right) \tag{4.2}$$

where $K(x)$ is the smoothing kernel, n the sampling size, and h the bandwidth. The sampling size in this case is equal to 60. There are various possibilities to select $K(x)$ (*Wilks*, 2011), however the Gaussian kernel is appealing in this case because of its unlimited support. The optimal selection of the bandwidth $\hat{h}$ can be obtained by minimizing the unbiased cross-validation criterium (*Scott and Sain*, 2005) given by

$$\min_{\hat{h}} \left[\int \hat{f}(x|h)^2 dx - \frac{2}{n} \sum_{k=1}^{n} \hat{f}_{-k}(x|h) \right] \tag{4.3}$$

where, $\hat{f}_{-k}(x|h)$ is the leave-one-out density estimate at x when observation x_k is not taken into account. This optimization was performed with a generalized reduced gradient algorithm. Once the optimal bandwidth is found, the best fit of the empirical distribution function can be estimated $\hat{f}$.

Finally, the mHM soil moisture index for a given cell and month, which denotes the quantile at the soil moisture fraction value x, can be obtained by numerically integrating the expression

$$\mathrm{SMI} = \int_0^x \hat{f}(u)du \tag{4.4}$$

4.3.3 Identification of drought events

Droughts are regional phenomena covering large contiguous areas over long periods. Understanding the spatial-temporal patterns and their relationships with other variables is therefore a fundamental step for drought prediction. Previous drought studies carried out in Germany, however, have been focused on statistical analysis of readily available point observations such as river discharge or precipitation data (*Demuth and Heinrich*, 1997; *Franke et al.*, 2004; *Schindler and Mayer*, 2007; *Schindler et al.*, 2007; *Stahl and Demuth*, 1999), and in general, they are limited to a regional scale rather than to the national scale. To the best of our knowledge, studies investigating the spatial-temporal drought variability over the whole German territory are not available in the scientific literature.

The retrospective reconstruction of soil moisture analysis in Germany provides a unique data set to estimate fundamental characteristics (e.g. severity and areal extent) of the major agricultural droughts occurred in Germany since 1950 at a high spatial resolution. Drought events were identified in this continuous spatio-temporal data set with the method proposed by *Andreadis et al.* (2005).

First of all, regions under drought stress were identified with the threshold method (*Dracup et al.*, 1980). This implies that cells fulfilling $\mathrm{SMI}_t < \tau$ were selected as potential regions under drought at the monthly time step t. The selection of the truncation level τ is fundamental for this method. A common value adopted in the literature is $\tau = 0.2$ (*Andreadis et al.*, 2005; *Vidal et al.*, 2010). This threshold indicates that a given cell is enduring a soil water deficit occurring less than 20% of the time.

In the second step, drought clusters at every monthly time step have to be consolidated in space. This means that all clusters whose area is less than a minimum threshold area will be excluded from further analysis. This step is necessary to eliminate small isolated areas that are suffering a drought but are too small to be considered as a regional event. In this study the minimum cluster area was set to 640 km^2 (i.e. 40 cells).

The final step of the drought event identification consists consolidating independent spatial clusters over successive time steps into a regional, multi-temporal cluster. This kind of clustering is necessary because the spatial variability of a drought event is vast, composed of many branches that can either merge together or split over time. The only condition to join clusters over time is that the overlapping area should be larger than 6400 km^2 (i.e. 400 cells). Overlapping areas less than this threshold area was considered as independent drought events.

Both threshold areas (i.e. the minimum cluster area and the overlapping area) were determined though sensitivity analysis but primarily based on rules of thumb often followed in the literature (e.g. *Andreadis et al.*, 2005; *Vidal et al.*, 2010). The main criteria for the selection of these parameters was the stability of drought characteristics described in the following section. It should be noted that the selection of smaller areas, enduring drought conditions, leads to the proliferation of smaller clusters that are not contiguous over time and hence can not be considered as part of a regional phenomenon.

4.3.4 Quantification of drought characteristics

Drought characteristics such as mean duration, mean areal extent, total magnitude, intensity, and severity-area-duration curves were quantified for every drought event and every ensemble member. The *mean duration* (D) of a spatio-temporal drought event is defined as the average of the drought duration of every cell within a drought event. This statistic is given in months. The *mean areal extent* (A) is defined as the average of a region under drought from the onset until the end of the drought event, expressed as percentage of the total German surface area. The *total magnitude* (M) is defined as the spatio-temporal integral of the SMI below the threshold value τ (i.e. the deficit) over those areas which are affected by the drought event, or explicitly

$$M = \sum_{t=t_0}^{t_1} \int_{A_t} \big(\tau - \mathrm{SMI}_i(t)\big)_+ \tag{4.5}$$

where, t_0 and t_1 denote the onset and the ending months of a given drought event. A_t is the area under drought at a given time step t, expressed as the percentage of total German surface area. i denotes a given location within the domain A_t, and $(\cdot)_+$ the positive part function. Thus, M is expressed in months times percentage of total German surface area.

Above described three statistics are useful to rank drought events based on the overall impact but they do not allow to estimate the impact of the drought after some months from the onset. This could be better quantified with the *drought intensity* (I_d) at a given duration d from the onset of the event. This statistic can be estimated as

$$I_d = \frac{1}{d} \sum_{t=t_0+1}^{t_0+d} \int_{A_t} \big(\tau - \mathrm{SMI}_i(t)\big)_+ \tag{4.6}$$

This statistic would also allow to estimate the impact of various events during summer and winter, by discriminating the time step $t_0 + d$ to a corresponding season.

Another commonly used method to benchmark drought events is based on the severity-area-duration curves (SAD) proposed by *Andreadis et al.* (2005). The *severity* (S_d) for every cell for a given duration d in months can be estimated as

$$S_d = 1 - \frac{1}{d} \sum_{t \in d} \mathrm{SMI}_t \tag{4.7}$$

The SAD curves for durations of 3, 6, 9, and 12 months for a given ensemble realization were constructed as follows. Firstly, the grid cells were ranked according their severity. The procedure starts with those cells having the maximum severity. Then, the severities of the adjacent cells were summed up progressively until a threshold area is reached. Afterwards, the average severity is estimated for those selected cells. The cumulative area and the average severity constitute the abscissas and ordinates of the SAD curves for a given duration. In this study, regular area intervals equivalent to the area of 20 grid cells were selected (i.e. every 320 km^2). This procedure is repeated until the whole area of a given drought event is covered.

The monthly evolution of these statistics was estimated for every member of the ensemble. Based on the ensemble simulations, the uncertainty of the four selected statistics was analyzed.

4.4 Results and Discussion

4.4.1 mHM evaluation

The performance of mHM was evaluated against observations of daily streamflow, latent heat and soil moisture measured at various eddy covariance (EC) stations acquired from www.fluxdata.org, as well as, with soil moisture observations obtained with a cosmic ray neutron probe (*Rivera Villarreyes et al.*, 2011). Seven large river basins in Germany were selected to cross-validate mHM performance with respect to observed daily streamflow. In this proxy basin test, global calibration parameters of mHM obtained at every river basin were transferred to the remaining test basins. For instance, from Neckar to Danube, Main, Ems, Saale, Mulde, and Weser basins (Fig. 4.1). The procedure to find the best hundred global parameter sets is described in section 2.4.2.2.

High efficiency in this kind of test is a good indication of model performance in ungauged locations. The ensemble mean Nash-Sutcliffe Efficiency (NSE) obtained with mHM using the best hundred global parameter sets at proxy basins during the validation period from 1965 to 1999 varied from 0.50 to 0.88, which is quite acceptable considering that these basins have significantly different hydrologic regimes. Model evaluation on those basins with at-site calibrated parameter sets during the same period exhibited on average a NSE value ranging from 0.74 to 0.93. During the calibration period (2000-2004), the NSE varied from 0.84 to 0.96. These tests indicated that mHM can be used for hydrological predictions within Germany.

The coefficient of determination between the simulated latent heat fluxes against observations across several eddy covariance (EC) stations varied between 0.50 and 0.74 during the period 2000-2002. The model domain in this case was reduced to a cell size of 100 $\times$ 100 m. Considering the various factors that influence the EC measurements and the fact that mHM is driven by disaggregated hourly values of precipitation and temperature as well as known scaling issues with EC measurements, these results can be regarded as satisfactory. The soil moisture anomalies estimated with mHM were able to explain up to 75% of the variance of their observed anomalies at various EC stations during the same period. Soil moisture estimates were obtained with standard TDR probes.

The model at the EC sites was forced with observed hourly precipitation and hourly temperature instead of the interpolated data as used for running the model over the whole domain.

The cosmic ray neutron probe, on the other hand, is a promising alternative because it allows an estimate of the soil water content over a control volume with a diameter of approximately 600 m and a depth of 0.3 m, which in this case, corresponds to the tillage depth setup in mHM. The coefficient of determination (r^2) between the mHM soil moisture anomaly and the cosmic ray neutron probe, reported by *Rivera Villarreyes et al.* (2011), was 0.57 for the period from August to September of 2011. Correspondingly, the r^2 between the simulated and the mean of soil moisture anomalies measured with 16 frequency domain reflectometry probes located within the same control volume was 0.79.

4.4.2 Retrospective reconstruction of soil moisture fields

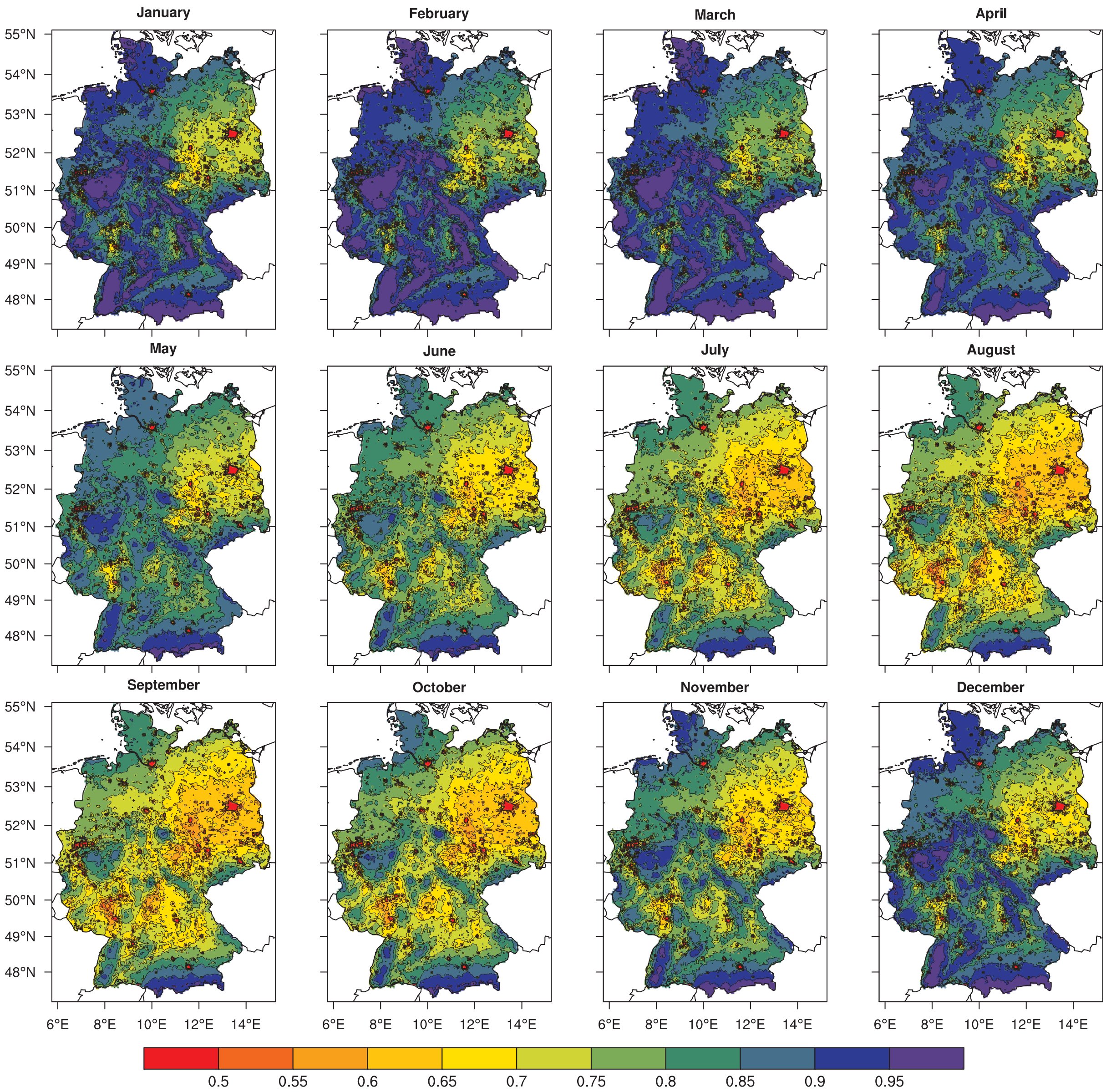

Figure 4.2 Ensemble monthly mean soil moisture fraction over Germany for the period 1950 to 2010.

The basis for the analysis of agricultural drought analysis in Germany was the reconstruction of daily soil moisture fields since 1950. Two hundred realizations of these fields were estimated for the whole of Germany at an hourly basis based on the premise that a single simulation is not sufficient for such analysis because of parameter uncertainty.

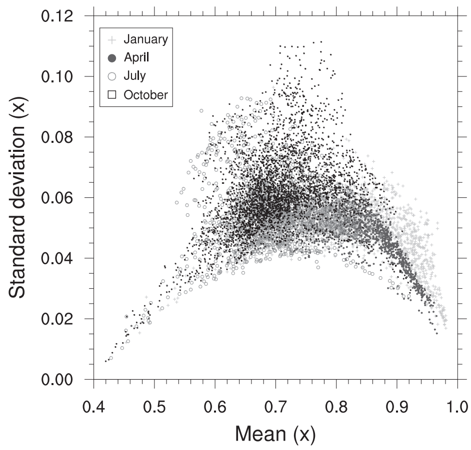

Figure 4.3 Seasonality of the long-term soil moisture fraction x in the Rhine basin. Each point denotes the mean and the standard deviation of x at a given grid cell within this basin.

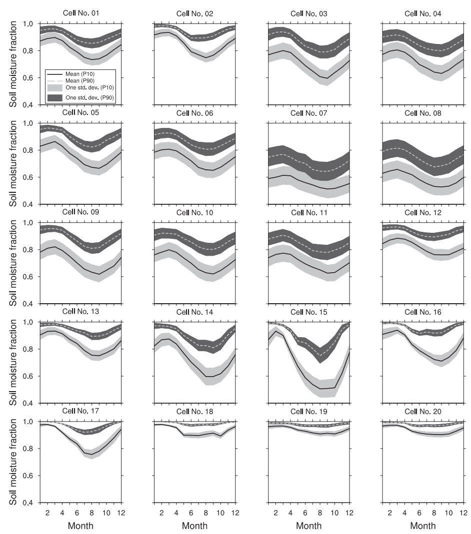

Figure 4.4 Parameter uncertainty of percentiles P_{10} and P_{90} of x at selected cells in Germany. The ensemble mean of these statistics is shown with a continuous line whereas their standard deviation is depicted in solid color. The location of the grid cells are shown in Fig. 4.1. The position of the panels approximately resembles the geographic location of these grid cells.

For the subsequent analysis, simulated hourly fields were aggregated to daily and monthly time steps. Monthly soil moisture values were then normalized as indicated in Eq. 4.1 to ease comparison across locations. The ensemble long-term mean for each month (Fig. 4.2) is the most evident statistic to evaluate these results and to verify whether the annual variability of soil moisture corresponds to the known climatology of major geographic regions in Germany. The variability of the spatial patterns shown in Fig. 4.2 indicate almost saturated conditions the whole year round in mountainous areas such as the Black Forest, the Harz mountains, and the Bavarian Alpine Foreland. Quasi-permanent dryer conditions have been observed on the North German Plain. The variability of the long term mean of the soil moisture fraction x with respect to its standard deviation indicates a clear seasonality describing wetter and less variable conditions in winter opposed to less wet but highly variable conditions in summer (Fig. 4.3).

Results indicated that not only the ensemble monthly climatology of the soil moisture fraction x, depicted in Fig. 4.2, but also other statistics such as the 10th and 90th percentiles of x (P_{10}, P_{90}) exhibits seasonality and strong dependency to geographic location. The annual variability of these two percentiles for selected cells within Germany is depicted in Fig. 4.4. The geographic location of the selected cells is shown in Fig. 4.1. The variability and the value of both percentiles indicate marked hydro-climatic regimes in Germany, for instance, humid regions with moderate seasonality on the North Sea (cells 1 and 2), very humid regions with very little seasonality in the alpine regions (cells 18-20), very humid regions with marked seasonality on the Black Forest (cells 13 and 17), moderately dry regions with marked seasonality in the North German Plain (cells 7 and 8), and regions with large seasonality in the pre-alpine regions (cells 14 and 15). In general, the standard deviation of the 90th percentile of x is less than that of the 10th percentile based on the 200-member ensemble. This corroborates the findings of *Meng and Quiring* (2008); *Schaake et al.* (2004) which point out that the parametric uncertainty in drier regions (cells 7, 8, 11, 15) is much higher than in humid regions (cells 17-20). The standard deviation of both percentiles exhibits not only seasonal variability, clearly depicted in cell no. 15 shown in Fig. 4.4, but also strong geographic dependency. This indicates that there is a complex interplay between climatic conditions and parametrization of the soil moisture processes.

4.4.3 Comparison with other indices

The same method proposed in section 13.5 to estimate the SMI can be used to estimate drought indices based on precipitation and surface runoff generated at each cell before routing (*Shukla et al.*, 2011). The results of these three drought indices are shown in Fig. 4.5 for one of the ensemble realizations obtained with DWD1. The three upper panels of this figure indicate how different the spatial distribution of the drought index might become depending on the variable used to describe a drought event. Among the three variables, the drought index based on precipitation exhibits the largest spatiotemporal variability because of the lack of memory of the precipitation process, which is one of the main reasons for considering it not appropriate for describing water stress in vegetation (*Döring et al.*, 2011). The drought index based on surface runoff is correlated to the SMI but still quite weak due to fast runoff generation processes. The SMI, as compared with the other two indices, exhibits the largest persistence.

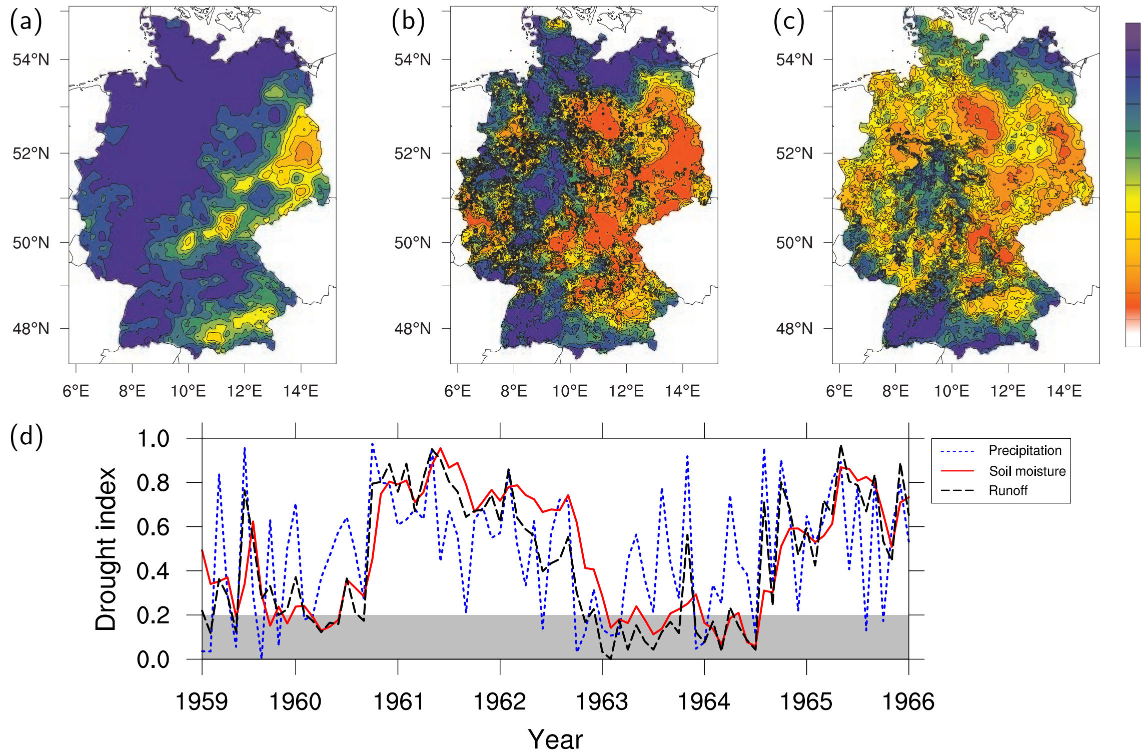

Figure 4.5 Drought indices estimated with precipitation (a), runoff (b) and soil moisture (c) at 1960-08. Panel (d) depicts the time series of the averaged values over Germany from 1959 to 1969. The solid grey area indicates the drought occurrence.

4.4.4 Sensitivity of the parameter uncertainty related to precipitation interpolation

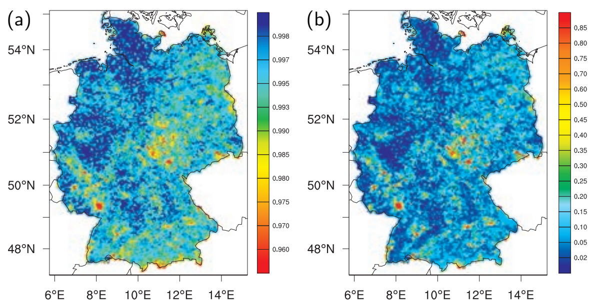

Figure 4.6 Ensemble mean of the Pearson correlation coefficient (a) and mean coefficient of variation (b) between monthly soil moisture fraction estimated with rainfall products DWD1 and DWD2 but same model parameters.

Among the two sources of parametric uncertainty investigated in this study, the first one was related to the interpolation methods used to regionalize rainfall point data. For this purpose, two methods were employed to estimate the gridded fields of precipitation data, as denoted by DWD1 and DWD2 (see section 2.4.2.2 for details). Since both methods use the same input data, any possible variation in soil moisture simulations — ceteris paribus— could be attributed to the kriging weights and the variogram parametrization used in DWD1, or the linear weights of the multi-linear regression method employed in DWD2. In this respect, two question were pursued in this study. (1) How important is this source of uncertainty for the estimation of soil moisture? And, (2) how is this uncertainty distributed over space? To answer these questions, the Pearson correlation coefficient (r) of the monthly soil moisture fractions at every grid cell obtained with both precipitation products (i.e. DWD1 and DWD2) were estimated separately for all 100 global parameter sets. From these r values, the ensemble mean ($\hat{r}$) and the coefficient of variation of r were calculated for every cell within the domain. These statistics are depicted in panels (a) and (b) of Fig. 4.6, respectively.

In general, most of the grid cells within Germany exhibit a $\hat{r}$ value greater than 0.98, which indicates a high degree of agreement between any pair of simulations driven by DWD1 and DWD2 forcings but having the same global model parameters. There are very few places where this statistic is less than 0.98, but in every case greater than 0.95. This finding along with the very low coefficient of variation indicated a quite low sensitivity of the monthly soil moisture fraction to the precipitation interpolation parameters. The lower values of $\hat{r}$ were obtained mainly in cells located in and around mountainous regions such as the Harz, the Alps, and the Swabian Jura (Fig. 4.6).

4.4.5 Overall parameter uncertainty of the soil moisture index SMI

The two major sources of parametric uncertainty described above induced considerable variability into the SMI as shown in Fig. 4.7, which depicts the areal average of the SMI over major German river basins, denoted hereafter as $\langle \mathrm{SMI} \rangle$. It can be noticed from this figure that the overall parameter uncertainty of $\langle \mathrm{SMI} \rangle$ is neither constant in space nor over time. The $\langle \mathrm{SMI} \rangle$ obtained with each ensemble member exhibited a large variability within the interquartile range of SMI but a relatively small one at its extreme quartiles (Fig. 4.7). This behavior is closely related with the high variability of the standard deviation of the soil moisture fraction around the middle ranges of its mean value (e.g. between 0.6-0.8 as depicted in Fig.4.3).

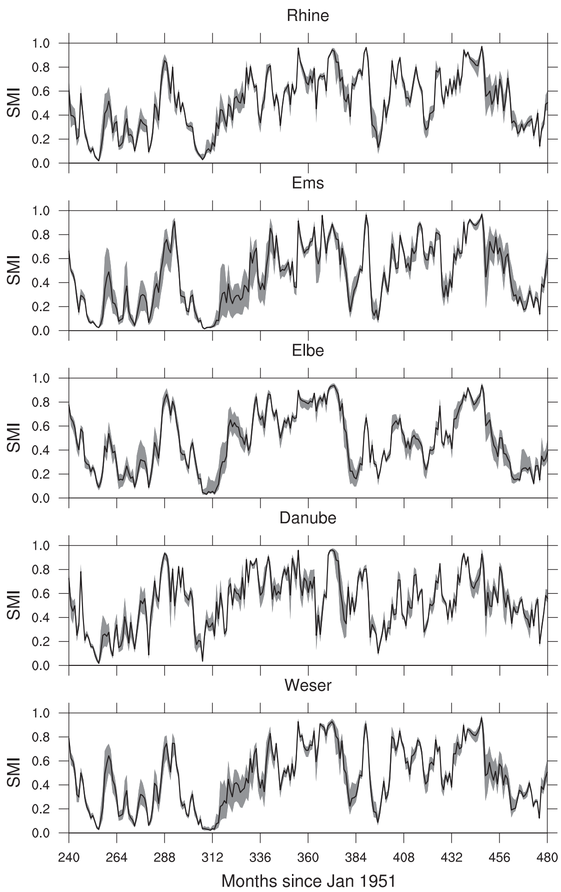

Figure 4.7 Parameter uncertainty of SMI averaged over six major basins in Germany from 1971-01-01 to 1991-12-31. The light grey depicts the variability of the ensemble $\langle \mathrm{SMI} \rangle$ and the black line represents the ensemble mean $\overline{\langle \mathrm{SMI} \rangle}$.

For further analysis, the temporal variability of $\langle \mathrm{SMI} \rangle$ within the ensemble simulations is estimated by its range $R(t) = \langle \mathrm{SMI(t)} \rangle_{\max} - \langle \mathrm{SMI(t)} \rangle_{\min}$, at every point in time t. $R(t)$ denotes the ensemble uncertainty of the soil moisture index over a given domain at time t. The long-term average of $R(t)$ is approximately 0.124 with a standard deviation of 0.014. The correlation coefficient estimated between the range of time series $R(t)$ for every pair of major basins, depicted in Fig. 4.1, varied from 0.25 to 0.88. This implied that the uncertainty of the SMI is not only the result of independent errors arising from model parametrization, but also the result of systematic interdependencies between soil moisture and climatic variables such as precipitation (P) and potential evapotranspiration (E_p). Based on these results, it was determined that the standard deviation of $R(t)$ tends to decrease as the ratio E_p/P increases. Moreover, given the data provided for each major basin (Fig. 4.7), the null hypothesis that the time series of the ensemble uncertainty $R(t)$ constitutes white noise can be safely rejected provided that the p-value of the Fisher's Kappa statistic was less than 0.001.

The 12-month moving average of $\langle \mathrm{SMI} \rangle$ depicted in panel (a) of Fig. 4.8 over the reconstruction period (1951-2010) showed a considerable reduction in uncertainty compared with the monthly values of $\langle \mathrm{SMI} \rangle$, but still not small enough to be considered negligible. The 12-month moving average of the percentage of area under drought (with respect to the surface area of Germany) exhibited a considerable uncertainty at the peaks of the events (panel (b) of Fig. 4.8). This result, however, allows preliminary identification of major drought events covering at least 50% of the German territory, namely those in the periods 1953-1954, 1959-1960, 1964-1965, 1972-1973, 1976-1977, 1992-1993, 2003-2004.

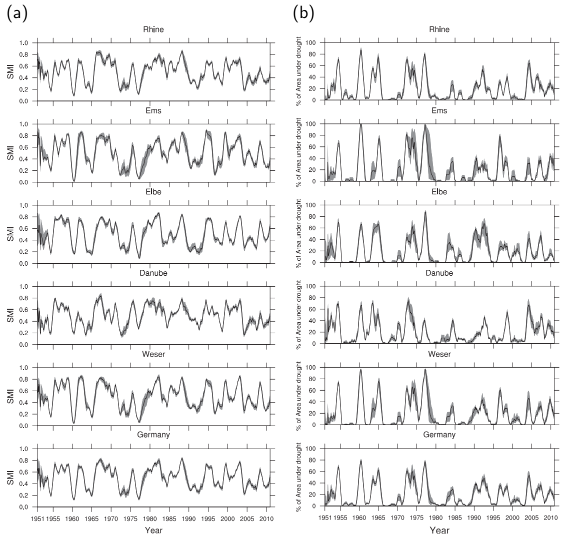

Figure 4.8 Panel (a): 12-month moving average of $\langle \text{SMI} \rangle$ over Germany and major river basins including uncertainty during the period from 1951-01-01 to 2010-12-31. Panel (b): Area under drought. The light grey line depicts the variability of the ensemble $\langle \text{SMI} \rangle$ and the black line represents the ensemble mean $\overline{\langle \text{SMI} \rangle}$.

The parametric uncertainty of the SMI also has a strong influence on drought severity classes commonly used for monitoring purposes. Panel (a) of Fig. 4.9 depicts the probability of finding a cell, at a given point in time, under one of the five drought severity classes used by the United States Drought Monitor (http://droughtmonitor.unl.edu). These classes denote abnormal (DO), moderate (D1), severe (D2), extreme (D3), and exceptional (D4) dry conditions, which correspond to: $0.2 < \text{SMI} \leq 0.3$, $0.1 < \text{SMI} \leq 0.2$, $0.05 < \text{SMI} \leq 0.1$, $0.02 < \text{SMI} \leq 0.05$, and $\text{SMI} \leq 0.02$, respectively. This figure shows also that there are areas, in which, no unique drought class can be assigned due to parametric uncertainty. A possibility to assign a unique class to a cell is to choose a class with the largest probability, as shown in the panel (b) of Fig. 4.9 for May 1976.

4.4.6 Identification of major drought events based on mean duration, mean areal extent and total magnitude

Major drought events were found in this study using the technique described in section 3.4.3.3. These benchmark events are required for the future analysis of possible consequences of climate change on agricultural droughts. The drought clustering algorithm was applied to every ensemble realization to find the spatio-temporal evolution of all drought events during the reconstruction period from 1951-2010. For every event, drought characteristics such as mean duration (D), total magnitude (M), and mean areal extent (A), among others, were evaluated using the procedure illustrated in section 3.4.3.4. The ensemble average of these characteristics, i.e. $\hat{D}$, $\hat{M}$, and $\hat{A}$ are depicted in Fig. 4.10. The corresponding uncertainty of these characteristics is presented in Table 4.1.

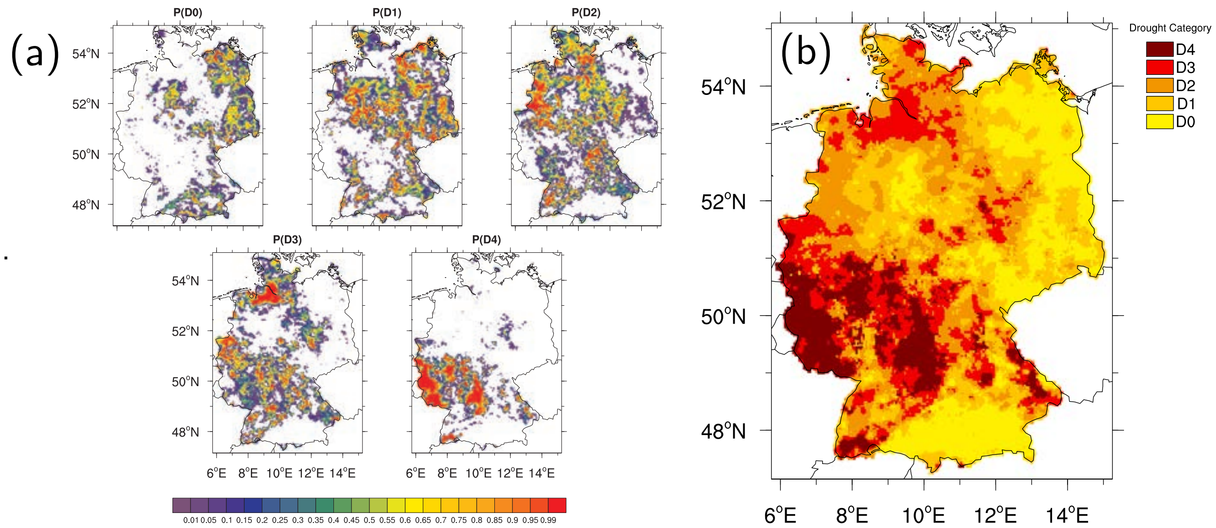

Figure 4.9 Panel (a): Probability of being at a drought severity class D0,...,D4 for May 1976. Panel (b): Most likely drought severity class based on the ensemble simulations. Classification according to the US Drought Monitor (http://droughtmonitor.unl.edu).

The eight largest drought events identified during the last 60 years in Germany are the following periods: 1962-1965, 1971-1974, 1975-1978, 1959-1960, 1953-1954, 1991-1993, 2003-2005, and 1995-1997. It is worth noting

Table 4.1 Uncertainty of characteristics of major drought events in Germany since 1950. Uncertainty of the characteristics and mean ± standard deviation.

Period	Duration [month]	Area [%]	Magnitude [% area × month] × 10^3
1953-1954	8.0 ± 0.2	70.8 ± 3.0	24.6 ± 1.0
1959-1960	12.0 ± 0.2	59.2 ± 2.3	36.3 ± 0.7
1962-1965	14.5 ± 0.9	41.5 ± 1.5	36.8 ± 2.0
1971-1974	14.8 ± 4.6	43.1 ± 5.0	36.7 ± 12.9
1975-1978	12.4 ± 0.8	43.5 ± 4.9	36.5 ± 1.9
1988-1991	5.9 ± 0.2	22.7 ± 2.0	11.1 ± 1.1
1991-1993	9.3 ± 1.5	29.3 ± 4.2	20.7 ± 3.6
1995-1997	8.5 ± 2.3	24.7 ± 6.7	11.8 ± 3.2
2003-2005	7.6 ± 0.5	32.1 ± 4.5	17.1 ± 1.6
2005-2007	5.6 ± 1.0	24.7 ± 3.4	11.5 ± 2.2

Table 4.2 Probability of finding a drought event in any of the top eight ranks. Here, only the eight largest events in Germany since 1950 were selected. The sum of the likelihood is not necessarily one due to the truncation of the table up to only the eighth rank. Values in bold represent the largest likelihood based on the ensemble simulations.

Event	Ranking likelihood							
	1	2	3	4	5	6	7	8
1953-1954		0.04	0.31	**0.56**	0.09			
1959-1960	0.34	**0.51**	0.15					
1962-1965		0.43	**0.48**	0.08	0.01			
1971-1974	**0.67**				0.03	0.19	0.10	0.01
1975-1978		0.02	0.06	0.36	**0.56**			
1991-1993						**0.59**	0.09	0.08
1995-1997						0.07	**0.53**	0.29
2003-2005					0.03	0.10	0.23	**0.59**

that the event from 2003-2005, appears in this overall ranking in the 7th position. *Vidal et al.* (2010) also noticed this fact and concluded that 2003 hardly appears as a benchmark event in France. This is a rather controversial conclusion because in this year the highest temperatures during the last 500 years were recorded (*Luterbacher et al.*, 2004). In Germany alone, great losses in the agricultural sector (*COPA-COGECA*, 2003) were reported. A likely explanation for this paradox is provided in section 4.4.7.

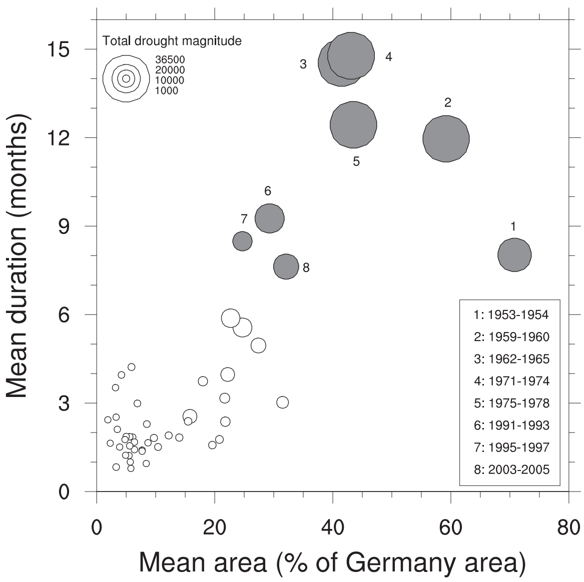

Figure 4.10 Area under drought, duration, and magnitude of the eight largest events in Germany since 1950 based on the ensemble ⟨SMI⟩.

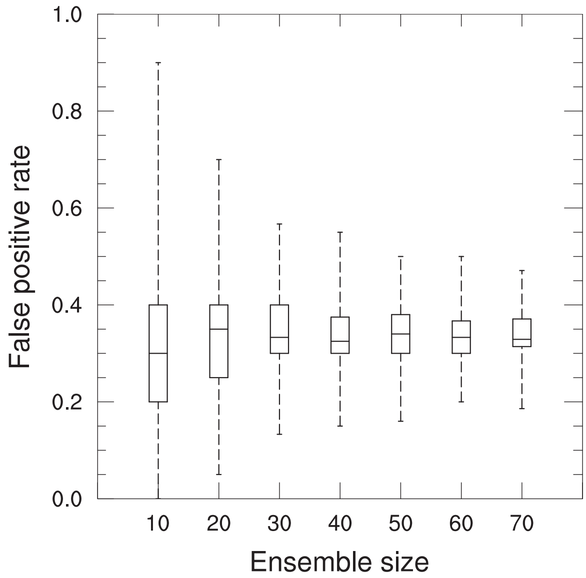

Figure 4.11 Sensitivity of the false positive rate (α) to ensemble size. In this example, α denotes the probability of rejecting the null hypothesis that the event from 1971-1974 ranks 1st among all drought events from 1950 to 2010. The size of the bootstrapping realizations was 1000.

The three drought characteristics D, M, and A depicted in Fig. 4.10, are highly correlated with each other. The Pearson correlation coefficient between D and M, is the highest, and equal to 0.97, whereas those between (D and A) and (M and A) are 0.80 and 0.87, respectively. This indicates that this triplet has low dimensionality. In fact, the first eigenvector of the correlation matrix of this triplet alone explains 92% of the total variance.

Using the k-means cluster analysis, three main groups of drought events were distinguished, 1) events with a large areal extent and duration, i.e. events 1962-1965, 1971-1974, 1975-1978, and 1959-1960; 2) events with the largest areal extent and moderate duration, i.e. 1953-1954; and 3) events with moderate areal extent and duration, i.e. 1991-1993, 2003-2005, and 1995-1997. Based on the ensemble SMI mean ($\overline{\text{SMI}}$), the event from 1971-1974 exhibited the longest duration, and the event from 1953-1954 covered the largest area. The events from 1962-1965 and 1971-1974 reached the two largest magnitudes.

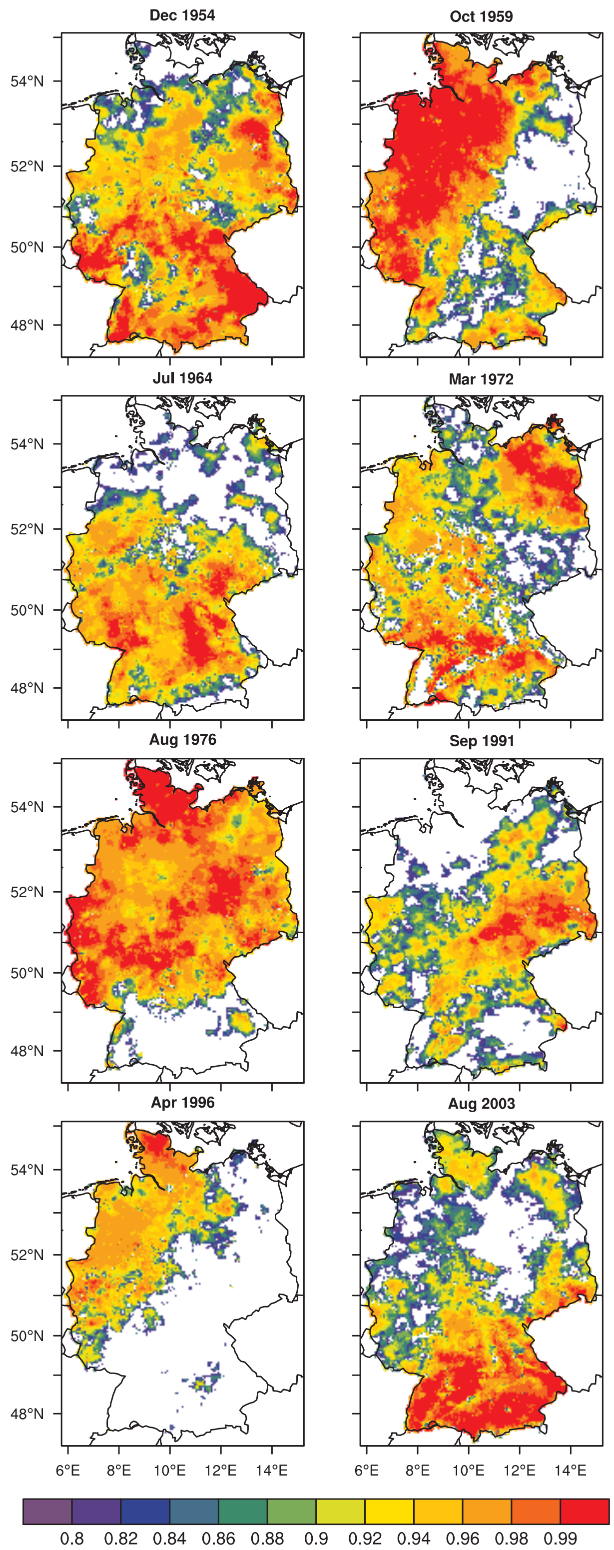

Figure 4.12 Severity at the peak of the eight largest drought events from 1951-01-01 to 2010-12-31 based on the ensemble mean SMI.

The absolute ranking of these extreme drought events is rather difficult due to the parameter uncertainty as illustrated in Table 4.2. This table presents an estimate of the probability to order every event into the eight top ranks using a linear, equal-weighted, normalized indicator composed of D and A, as an example. The results presented in this table indicate that the maximum probability of finding an event in one of the top ranks is not greater than 0.67. The ranking of a given event spans at least over three categories. Low ranking events tend to have a much larger ranking spread than the top ones, though.

The size of the ensemble also played a very important role to estimate the probability of finding an event in a given rank $(1 - \alpha)$, where α denotes the false positive rate. Fig. 4.11, for example, shows the probability of not identifying the event from 1971-1974 as the largest since 1951. This figure clearly shows that the variance of the false positive rate is strongly dependent on the ensemble size. These results were obtaining by bootstrapping the 200 ensemble simulations without replacement and limiting the number of realizations to 1000 for a given sample size. This figure showed also that the first two moments of α tend to stabilize with ensemble sizes larger than 50. Consequently, it is safe to conclude that small ensemble sizes would lead to misleading results. An ensemble with 200 members, as realized in this study would lead to safer results. These Monte Carlo realizations clearly highlighted the role of parametric uncertainty in identifying the benchmark drought events which should be handled carefully.

The spatial distribution of severity (S_d) based on $\overline{\text{SMI}}$ at the peak of the eight largest drought events is shown in Fig. 4.12. It can be observed from this figure that each event has its own peculiarities with respect to the spatial distribution of the affected areas. The drought event during December 1954 has the largest areal coverage, with 93.5% of the German territory under water stress, whereas the event during April 1996 had the lowest coverage with 46.5%. The latter drought event at its peak was particularly concentrated on the northwest part of Germany. The event of 1976, with its peak in August, had spread over whole Germany with an exception of the Alpine Foreland. The latter areas endured the highest severity during August 2003.

4.4.7 Uncertainty of large events occurring in summer and winter

As mentioned before, the ranking of drought events based on ensemble characteristics (D, M, and A) does not allow the identification of their impact at a given point in time from their onset, nor to differentiate them according to their level of incidence in a particular season. The drought intensity proposed in Eq. 4.6 enables estimating the transient evolution of a drought event from its onset, and by so doing, it allows quantifying how fast a drought event covered a given area and by what magnitude. Panel (a) in Fig. 4.13 shows the results of plotting drought intensity versus duration from the onset (d) of a given event for the ten largest events since 1950. Panel (b) in the same figure depicts the results obtained by ranking the drought intensities of all events at various durations from their onsets (e.g. 3, 6, ... months). The classification of an event into summer or winter was estimated with the procedure illustrated in section 3.4.3.3 (Eq. 4.6). The ensemble SMI mean (i.e. $\overline{\mathrm{SMI}}$) was used instead of individual realizations for both analyses because the former is an unbiased estimate of the SMI, and thus leads to a robust estimate of the evolution of the drought intensity.

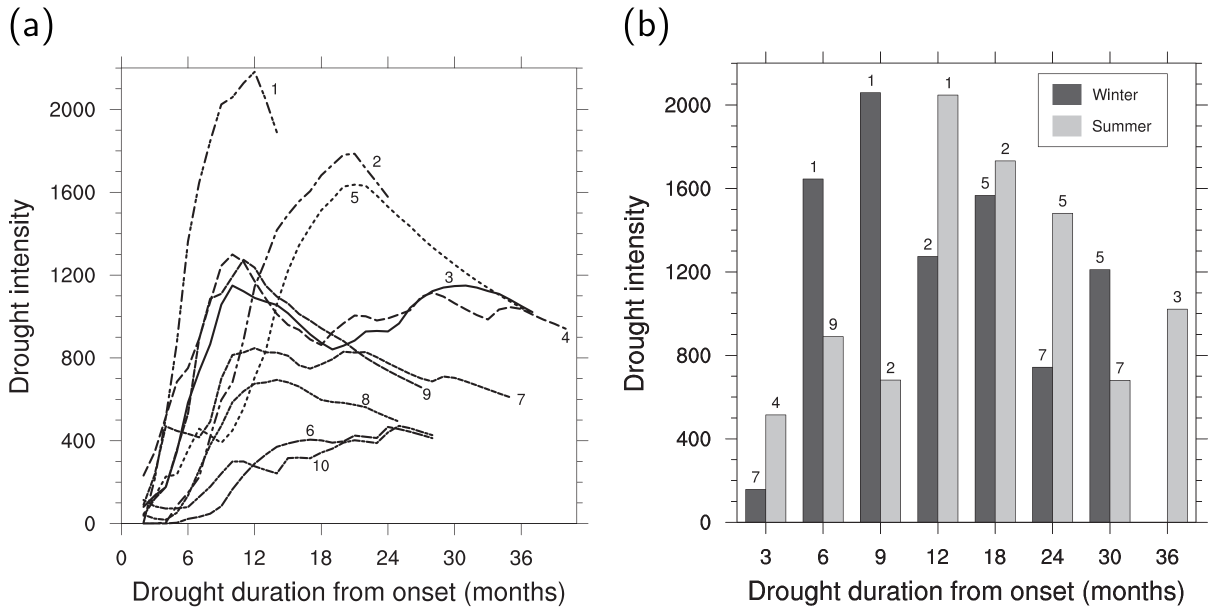

Figure 4.13 Panel (a): Drought intensity evolution for the 10 largest drought events since 1950. Panel (b): Major drought events for a given duration and season of occurrence. The numbers denote the following events: 1: 1953-1954, 2: 1959-1960, 3: 1962-1965, 4: 1971-1974, 5: 1975-1978, 6: 1988-1991, 7: 1991-1993, 8: 1995-1997, 9: 2003-2005, 10: 2005-2007.

Based on the results described above and shown in Fig. 4.13, it was found that at 3 month duration, summer events have much larger drought intensity than the corresponding ones in winter. At 6 and 9 months duration, the opposite happens. The events with more than a 9 month duration mostly reach their higher intensities during summer as compared to winter ones. However, droughts having a duration of 30 months or more are more intense during winter months. The event 1953-1954 not only exhibits the largest intensities at 6 and 9 month durations during winter months (Nov-Apr), but also the largest intensity in summer at 12 months duration. The event 2003-2005 is, according to these results, the summer event with the largest intensity at 6 months duration. Among the 10 largest drought events in Germany during last 60 years, the 1953-1954 event had the largest intensity peaking within a relatively short period of time (less than 12 months). This event, however, lasted for only one and a half years. Four drought events, namely, 1962-1965, 1971-1974, 1975-1978, and 1991-1993, spanned over the period of more than 30 months (i.e. two and a half years). According to this analysis, the decade of 1970 could be regarded as the most severe drought period in Germany. The drought events 1962-1965 and 1971-1974 clearly exhibited more than one peak over their whole life span. The analysis also indicated that most of the historical drought events in Germany have their peaks during 6 to 12 months of duration.

The empirical bivariate density function between the average drought area (A) and the total magnitude (M) was constructed to analyze the uncertainty in overall drought characteristics (D, M, and A) based on the ensemble realizations. The large number of model runs also allowed to assess the uncertainty in time evolution of these characteristics. The four most intense drought events with 6 months and at least 30 months duration after its onset were selected to illustrate this procedure, namely: the events 1953-1954 and 2003-2005 for shorter duration, and the events 1975-1978 and 1962-1965 for longer duration, respectively (Fig. 4.14). It is worth noting that the events 1953-1954 and 2003-2005 are classified as winter and summer events, respectively, at 6 months duration (Fig. 4.13 (b)). Likewise, the events 1975-1978 and 1962-1965 peaked in winter and summer, respectively. Droughts that are peaking within a relatively short time (up to 6 months) from their onset are quite relevant because they have large repercussion on socio-economic activities.

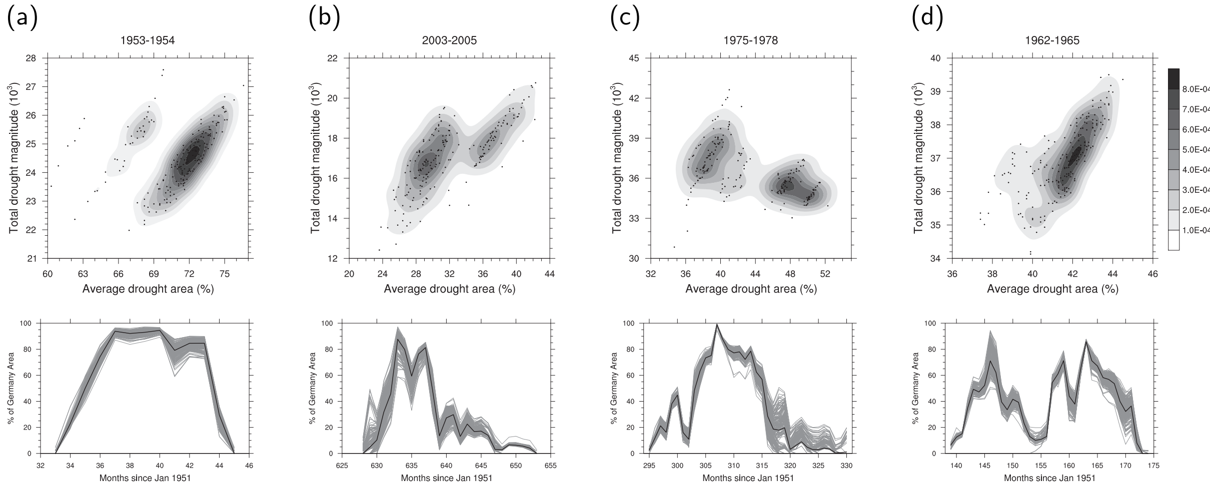

Figure 4.14 Top: Bivariate density functions between drought area and total drought magnitude of four major events. Panels (a) and (b) depict the most intense drought events with 6 months duration after its onset in winter and summer, respectively. Panels (c) and (d) correspond to the most intense drought events having a drought duration of at least 30 months, in winter and summer, respectively. Bottom: Predictive uncertainty and evolution of the area under drought for the selected events.

Based on the ensemble results, the density function for each event was estimated independently with a bivariate Gaussian kernel smother algorithm. The estimation of the bandwidths in both directions was carried out in a similar way as presented in section 3.4.3.2. The results of this analysis are depicted in the top panels (a) to (d) of Fig. 4.14, which clearly supports the research hypothesis that the parametric uncertainty of soil moisture has a strong implication for drought characterization. Most events exhibit multimodal behavior which is the combined result of the uncertainty of the model parametrization and drought identification (e.g. clustering, threshold).

Events having shorter durations and peaking in winter (1953-1954) appear to be more certain than those peaking in summer (2003-2005) as can be noted by the larger spread of the respective distribution (Fig. 4.14 (a) and (b)). Consequently, the probability density values for the summer event are lower than those of the winter event. However, at longer durations no conclusive comparison could be drawn from this analysis because longer events experience various seasons over many years. The time evolution of the area under drought $A(t)$ for each events, as depicted in bottom panels of Fig. 4.14, also supports the assertion that a single model realization would lead, very likely to a high rate of false alarms for drought monitoring.

4.4.8 Uncertainty of the Severity-Area-Duration curves

SAD curves obtained with the ensemble SMI mean ($\overline{\mathrm{SMI}}$) for the eight largest drought events in Germany at duration 3, 6, 9, and 12 months are depicted in panel (a-d) of Fig. 4.15. From this analysis, the event from 1975-1978 appears to be the most severe and extensive event at durations ranging from 3 to 9 months. Based on this measure, the 2003-2005 event, however, hardly appears as a benchmark event at longer durations and area coverage. The event from 1953-1954 is quite severe at 3 and 6 months, but not at longer durations. The apparent contradiction of these results, can be clarified with the individual evolution graphs presented in Fig. 4.13.

SAD curves have often been used to rank drought events (*Andreadis et al.*, 2005; *Sheffield et al.*, 2009). Due to parametric uncertainty, however, they exhibit large variability as shown in panel (e) of Fig. 4.15. This, again, corroborate our hypothesis that a single model run would lead to unsatisfactory conclusions and event ranking. These results indicate that the SAD variability increases as the area under drought and duration increase. The variability of the SAD curve with a 12-month duration is almost twice as much as that for 3 months. The variability of SAD curves for summer events is higher than that estimated for winter at any duration.

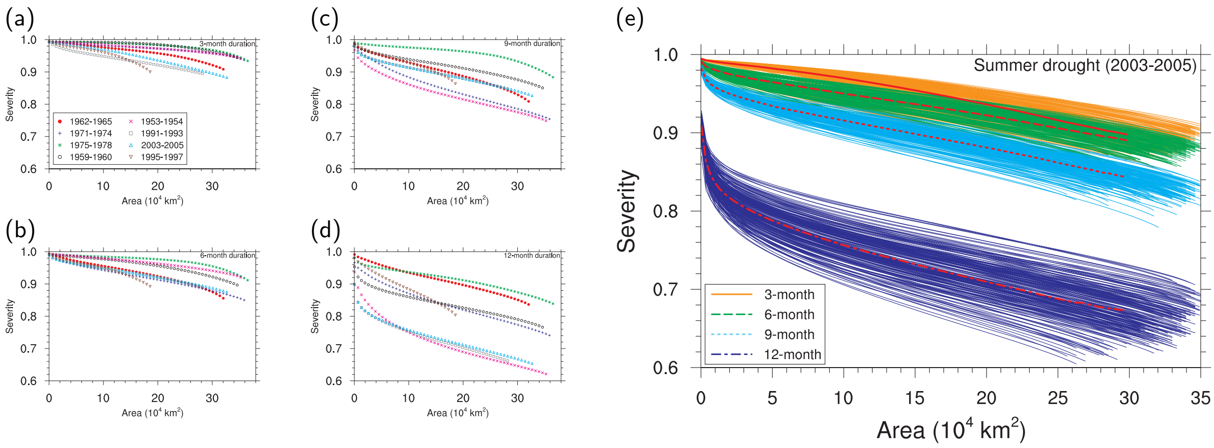

Figure 4.15 Panels (a) to (d): Ensemble averaged Severity-Area-Duration (SAD) curves of eight major drought events for 3, 6, 9, and 12 months duration since 1950 over Germany. Panel (e) depicts the predictive uncertainty of the SAD curves obtained for the event 2003-2005. In this panel, lines in red denote the ensemble mean.

4.4.9 Drought persistence and SMI trends

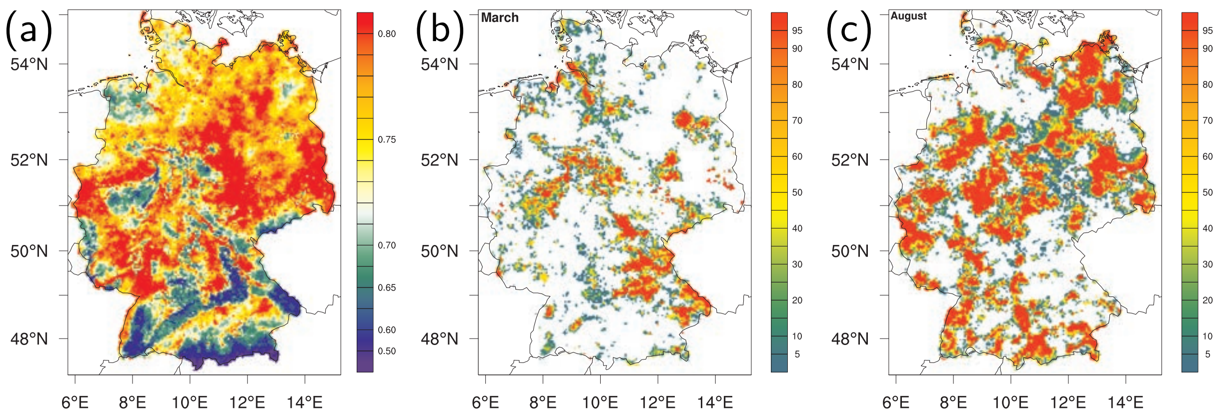

Figure 4.16 Persistence map of the SMI (a), and regions with positive (b) and negative(c) SMI trends (5% significance). Panels (b) and (c) depict the percentage of ensemble members indicating a significant trend.

Characterizing areas prone to remain under severe drought conditions when they are already suffering one constitute a relevant piece of information for water resources planning. The level of persistence of the severe drought events can be quantified with a two-state Markov chain with two states: $\mathrm{SMI} \leq 0.2$ and $0.2 < \mathrm{SMI} \leq 1$.

The persistence of severe drought can be estimated for each ensemble member as the probability $\pi_{00} = \Pr\left(\mathrm{SMI}(t+1) \leq 0.2 \mid \mathrm{SMI}(t) \leq 0.2\right),\ \forall t$. The ensemble mean of π_{00} is depicted in panel (a) of Fig. 4.16 for the whole of Germany. This figure indicates that most of the Northeast German Plain comprising the area of the Elbe, Saale, and Mulde river basins, as well as large extensions along the Main and Rhine rivers, exhibit drought persistence greater than 0.8. The Northwest German Plain, comprising the Ems and Weser river basins, tend to have lower drought persistence than the eastern part of Germany, with an average value of π_{00} less than 0.7. The Alpine

Foreland located within the Danube basin and areas in and around the Black Forest, on the contrary, exhibit the largest variability in drought persistence within Germany ranging from less than 0.4 to 0.8. It is worth nothing that those areas exhibiting large drought persistence have been also classified as areas with medium to high agricultural suitability according to a recent study conducted by UBA-PIK (www.pik.de). These regions comprise large plains within the Saale river basin around the cities of Halle and Magdeburg, and flood plains of the Rhine river on the western side of the Black Forest.

Mann-Kendall tests on monthly $\overline{\mathrm{SMI}}$ indicate that there are large extensions of the German territory showing positive trends (i.e. getting wetter) during winter months but negative trends in summer months, at 5% significance level. The largest areas exhibiting significant trends were detected in March and August as depicted in Fig. 4.16, panels (b) and (c), respectively. It is worth noting that positive SMI trends tend to occur in areas with low persistence and negative trends in areas with high persistence. These trends are, in turn, related, with observed trends in temperature and precipitation. Further details on this aspect are beyond the scope of this paper.

4.5 Summary and Conclusions

In this study we have presented a method to derive a soil moisture index based on a process based hydrological model. This model uses a multiscale parametrization method that goes beyond standard calibration approaches. Great emphasis has been put on testing this model in all major river basins in Germany, especially with respect to the transferability of global parameters across locations and scales. Ongoing tests with Fluxnet and cosmic ray neutron probe data have also been presented. Using this model a consistent ensemble of high resolution daily soil moisture fields for Germany since 1950 at a spatial resolution of 4×4 km were obtained.

Based on this soil moisture reconstruction, a soil moisture index (SMI) representing the corresponding monthly quantile was estimated with the kernel density approach. The derived SMI exhibits high correspondence with total grain yield of Germany and allows to identify major drought events in Germany, that have also been identified using other techniques (e.g. tree rings) and reported in the literature (*Büntgen et al.*, 2010). This approach has advantages over standard empirical approaches or those obtained from satellite derived products, which are too coarse to account for soil moisture at high spatio-temporal resolutions and quite uncertain because the algorithms used to infer soil moisture do not take into account the water balance of large river basins. Consequently, the proposed technique has a large potential to be used as a monitoring tool in the future. More research is, however, needed to evaluate the SMI against times series of annual crop yield at regional scale. Further research is also required to identify potential driving mechanisms, the feedback effects, and the spatio-temporal correlations of soil moisture with other hydrological state variables such a snow depth, and climatic variables.

The effects of other sources of uncertainty stemming from model structure and quality of meteorological data on the soil moisture index should be further investigated. Potential benefits of using ensembles of multi-model, multi-parameter soil moisture simulations should be also carried out. Both issues, however are out of the scope of this study.

Based on the results of this study, the following conclusions were drawn. 1) The main source of parametric uncertainty of the soil moisture index is related with global model parameters. This uncertainty is seasonally and regionally varying. This corroborates, findings of other researchers who have advocated for multi-model ensembles to account for model uncertainty. In summary, one single model run is not enough for estimating benchmark events. 2) The uncertainty of overall statistics used for estimating drought events are highly sensitive to this kind of uncertainty. This sensitivity is the result of non-linear relations and branching effects caused by the clustering method. 3) Events peaking during summer with at most 6 months duration tend to exhibit a much large uncertainty than those peaking during winter. 4) The SMI is not a stationary variable. Many regions in Germany exhibited significant trends during the study period. Potential triggering mechanisms and drivers behind these trends might be the observed changes of precipitation and temperature, as well as, other feedback mechanisms. A detailed trend attribution, however, is out of the scope of this study. 5) The identification of benchmark drought events should be based on combined criteria such as SAD or intensity duration curves. Robust estimates can only be made with an ensemble SMI due to the uncertainty mentioned before.

Acknowledgment

This study was carried out within the context of the REKLIM project. The Authors would like to acknowledge the collaboration of the Federal Institute of Hydrology (BfG) supporting the European Water Archive (EWA) and the Global Runoff Data Centre (GRDC), the Federal Institute for Geosciences and Natural Resources (BGR), the Federal Agency for Cartography and Geodesy (BGK), the German Weather Service (DWD), the European Environment Agency (EEA), FLUXNET, and the NASA Land Processes Distributed Active Archive Center (LP DAAC) for providing the data employed in this study. We are also very grateful with Mr. Rivera Villarreyes for kindly providing cosmic ray neutron probe data for validating our model. We are also like to thank Ms. Elisabeth Krüger, Mr. Pablo Beltran, and two anonymous reviewers for their valuable comments.

CHAPTER 5

TOWARD SEAMLESS HYDROLOGIC PREDICTIONS ACROSS SPATIAL SCALES

L. Samaniego[1] et al.

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Samaniego, L., R. Kumar, S. Thober, O. Rakovec, M. Zink, N. Wanders, S. Eisner, H. Müller Schmied, E. H. Sutanudjaja, K. Warrach-Sagi, and S. Attinger (2017), Toward seamless hydrologic predictions across spatial scales, Hydrology and Earth System Sciences, 21(9), 4323–4346

5.1 Abstract

Land surface and hydrologic models (LSMs/HMs) are used at diverse spatial resolutions ranging from catchment-scale (1–10 km) to global-scale (over 50 km) applications. Applying the same model structure at different spatial scales requires that the model estimates similar fluxes independent of the chosen resolution, i.e., fulfills a flux-matching condition across scales. An analysis of state-of-the-art LSMs and HMs reveals that most do not have consistent hydrologic parameter fields. Multiple experiments with the mHM, Noah-MP, PCR-GLOBWB, and WaterGAP models demonstrate the pitfalls of deficient parameterization practices currently used in most operational models, which are insufficient to satisfy the flux-matching condition. These examples demonstrate that J. Dooge's 1982 statement on the unsolved problem of parameterization in these models remains true. Based on a review of existing parameter regionalization techniques, we postulate that the multiscale parameter regionalization (MPR) technique offers a practical and robust method that provides consistent (seamless) parameter and flux fields across scales. Herein, we develop a general model protocol to describe how MPR can be applied to a particular model and present an example application using the PCR-GLOBWB model. Finally, we discuss potential advantages and limitations of MPR in obtaining the seamless prediction of hydrological fluxes and states across spatial scales.

5.2 Introduction

... "If it disagrees with experiment, it's wrong". Richard P. Feynman

Land surface and hydrologic models (LSMs/HMs) are currently used at diverse spatial resolutions ranging from 1 to 10 km in catchment-scale impact analysis and forecasting (*Addor et al.*, 2014; *Christensen and Lettenmaier*, 2007) to over 50 km in global-scale climate change simulations to estimate land surface boundary conditions of key state variables (*Bierkens*, 2015; *Haddeland et al.*, 2011; *Wanders and Wada*, 2015). The fundamental conditions behind the applicability of the same LSM/HM model structure at different spatial scales requires that the model parameterizations are scale invariant and that the model estimates similar fluxes across a range of spatial resolutions. In other words, it must fulfill the flux-matching condition across scales so that the mass conservation principle can be ensured (*Wood*, 1997).

A parameterization is a simplified and idealized representation of subgrid physical phenomenon that is either "too small, too brief, too complex, or too poorly understood" to be explicitly represented by a model at a given resolution (*Edwards*, 2010). Parameterizations require variables called predictors, effective parameters and constants also called transfer, global, or super parameters (*Pokhrel and Gupta*, 2010). Super parameters are often parameters in empirical relationships that have been found with measurements in the field or in the laboratory, e.g., regression parameters in pedotransfer functions (*Cosby et al.*, 1984). They are often tuned to represent observed variables and often have no physical meaning. These parameters constitute simplified surrogates to compensate for the missing subgrid processes that are not accounted for within a modeling system (*Brynjarsdottir and O'Hagan*, 2014).

Effective parameters of LSMs/HMs are usually obtained by ad hoc procedures (e.g., automatic calibration) at a given spatial resolution for a given modeling domain. As a consequence of this standard practice, parameter fields of LSMs/HMs often exhibit artificial spatial "discontinuities" such as calibration imprints circumscribing river basin boundaries, and consequently they are not seamless (*Li et al.*, 2012b; *Merz and Blöschl*, 2004). Inconsistent patterns of effective parameter fields for land surface geophysical properties across spatial scales constitute a clear indication that their parameterizations are not scale invariant. There are several reasons explaining this parameterization deficiency. With the advent of electronic computers, the performance of general circulation models (GCMs), numerical weather prediction (NWP) models (*Pielke Sr*, 2013), land surface models (*Liang et al.*, 1994; *Niu et al.*, 2011; *Sellers et al.*, 1997), and hydrologic models (*Batjes*, 1996; *Lindstrom et al.*, 1997; *Samaniego et al.*, 2010a; *van Beek et al.*, 2011) has been increased mainly by improving model conceptualization (i.e., the number of process descriptions) and/or spatial resolution since the storage capacity and computational power allowed for it (*Bierkens et al.*, 2014; *Le Treut et al.*, 2007; *Wood et al.*, 2011). As a result, parameterizations in LSMs have also increased in their complexity during the past decades (*Fisher et al.*, 2014; *Sellers et al.*, 1997). The procedures to estimate effective parameters required for the parameterizations, however, remained unchanged. For example, LSMs evolved from simple aerodynamic bulk transfer schemes with uniform description of surface parameters during the 1970s to detailed LSMs having a consistent description of the exchange of energy and matter between the atmosphere, the vegetation, and the land surface (*Sellers et al.*, 1997). State-of-the-art LSMs, such as the Community Land Model version 4 (*Bonan et al.*, 2011) and Noah-MP (*Niu et al.*, 2011), however, still use quite simple pedotransfer functions based on work of *Clapp and Hornberger* (1978) and *Cosby et al.* (1984) to estimate fundamental soil properties such as porosity (*Oleson et al.*, 2013).

Further reasons that have prevented the improvement of parameterization techniques are

- the lack of procedures and theories for linking physical properties (e.g., soil porosity) that can be measured at the field scale with "effective" parameter values that represent the aggregate behavior of the land characteristics at the scale of a grid cell required in LSMs or HMs,
- poor understanding of the scaling of parameters (*Dooge*, 1982) and its influence on the hydrological response of the system (*Wood*, 1997; *Wood et al.*, 1988),
- limited inclusion of subgrid heterogeneity in hydrological parameterizations and multiscale modeling of hydrologically relevant variables as suggested by *Famiglietti and Wood* (1994, 1995); *Liang et al.* (1996b),
- lack of significant progress on the applicability of seminal upscaling theories (*Dagan*, 1989; *Gelhar*, 1993; *Kitanidis and Vomvoris*, 2010; *Miller and Miller*, 1956; *Neuman*, 2010) developed for subsurface hydrologic problems into LSMs/HMs, and
- lack of transparency in most of the existing LSM/HM source codes with respect to the meaning, origin, and uncertainty associated with the hard-coded numerical values (i.e., parameters) either in the code or in the look-up tables (*Cuntz et al.*, 2016; *Mendoza et al.*, 2015).

Consequently, it is possible to assert that model parameterization is an old, ubiquitous, and recurring problem in land surface and hydrologic modeling. Considering this lack of coherent development during the past decades, we can still concur with *Dooge* (1982, p. 269) and say that the "parameterization of hydrologic processes to the grid scale of general circulation models is a problem that has not been approached, let alone solved."

There are potential methods available in the literature that may lead toward coherent parameterizations and prediction of water and energy fluxes in LSMs/HMs. For example, (1) sidestepping the scaling problem of key model parameters by assuming scale-independent distribution functions with regionalized distribution parameters (*Intsiful and Kunstmann*, 2008), (2) finding strong links between model parameters to mapped geophysical attributes via regularization procedures (*Pokhrel and Gupta*, 2010), and (3) finding strong links between of observed functional responses of hydrological systems and geophysical characteristics (*Yadav et al.*, 2007). These methods, however, alone may not satisfy the flux-matching criteria.

In contrast to these existing methods, we argue that the multiscale parameter parameterization (MPR) technique (*Samaniego et al.*, 2010a) offers a framework to link the field scale (observations) with the catchment scale (*Dooge*, 1982). MPR also accounts for the effect of the spatial variability and non-linearity of geophysical characteristics in the parameterization of hydrologic processes that operate at a range of spatial resolutions (*Dooge*, 1982; *Wood et al.*, 1988). Depending on the conditions imposed on the parameter estimation technique, MPR can lead to parameterizations that satisfy the flux-matching criteria and hence contributes to obtaining seamless parameter and water flux fields. Because MPR relies on empirical transfer functions and upscaling operators to link geophysical properties with model parameters, it provides a very effective procedure to transfer "global parameters" to scales and locations other than those used in calibration (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a,b). This dependency on several transferable coefficients also contributes to minimizing a serious drawback of spatially explicit models called "overparameterization" (*Beven*, 1995).

In this study, we analyze to which extent existing LSM/HM parameterizations are limited to obtain seamless predictions of water fluxes and states across multiple spatial resolutions. Through several modeling experiments addressing *Wood* (1990)'s query (i.e., "What modeling experiments need to be performed to resolve the scale question . . . "), we demonstrate that a large portion of the predictive uncertainty in existing LSMs/HMs originates from the deficient estimation of effective parameters, which leads to a lack of scale invariance and thus to their poor transferability across scales and locations. These experiments also aim to help the modeler to reveal poor-performing parameterizations, i.e., those that exhibit non-seamless fields. Finally, based on our past experiences and aiming to address the challenges stated above, we develop a protocol that systematizes the application of the MPR technique for any LSM/HM and demonstrate its effectiveness by implementing it into the PCR-GLOBWB model.

5.3 Current parameterization techniques

5.3.1 The state-of-the-art

The most common parameterization techniques found in the literature are (1) look-up tables (LUTs), (2) manual or automatic calibration, (3) hydrologic response units (HRUs), (4) representative elementary watersheds (REWs), (5) a priori regularization functions, (6) simultaneous regionalization/regularization functions, and (7) dissimilarity-based metrics to transfer model parameters.

The simplest technique to assign a parameter value to a modeling unit (e.g., grid cell, HRU, or subcatchment) is based on a LUT. In this case, a categorical index associated with a modeling unit links it with information taken from an external reference file (i.e., the LUT) which maps this index with parameter values that are usually taken from the literature. This technique is commonly used in most of the (operational) LSMs such as CABLE, CHTESSEL, CLM, JULES, and Noah-MP (*Best et al.*, 2011; *ECMWF*, 2016; *Kowalczyk et al.*, 2006; *Niu*, 2011; *Oleson et al.*, 2013; *Viterbo and Beljaars*, 1995). A disadvantage of this method is the difficulty to perform sensitivity analysis (*Cuntz et al.*, 2016). Moreover, the number of classes defined in LUT is often limited to a few (e.g., 13 soil classes in Noah-MP) resulting in non-seamless parameter fields that are not continuous.

Manual or automatic calibration is a commonly used technique to parameterize spatially lumped hydrologic models (e.g., *Andréassian et al.*, 2014; *Burnash et al.*, 1973b; *Crawford and Linsley*, 1966; *Edijatno et al.*, 1999; *Fenicia et al.*, 2011; *Lindstrom et al.*, 1997; *Martina et al.*, 2011; *Singh et al.*, 2014) and semi-distributed hydrologic models (e.g., *Hundecha and Bárdossy*, 2004; *Hundecha et al.*, 2016; *Kavetski et al.*, 2003; *Leavesley et al.*, 1983; *Lindström et al.*, 2010; *Merz and Blöschl*, 2004). The aim is to minimize the disagreement between model simulations and observations. In the majority of the cases, the target variable is streamflow. The main drawback of this parameterization technique is that the parameter fields, which are obtained by colocating lumped model parameters from sub-basins, are doubtful because they exhibit sharp discontinuities along individually calibrated sub-basin boundaries despite having spatial continuity in basin physical attributes like soil, vegetation, and geological properties that govern spatial dynamics of hydrological processes (*Blöschl et al.*, 2013; *Li et al.*, 2012b; *Merz and Blöschl*, 2004). In addition, the "patchwork quilt" parameter fields shown in these references exhibit significant sensitivity to the calibration conditions as demonstrated by *Merz and Blöschl* (2004). Thus, models that are parameterized with this technique may exhibit (1) poor predictability of state variables and fluxes at locations and periods not considered in calibration and (2) sharp discontinuities along sub-basin boundaries in state, flux, and parameter fields (e.g., *Lindström et al.*, 2010; *Merz and Blöschl*, 2004). Parameter fields derived from basin-wise "calibrated" lumped models lack spatial seamlessness and thus are "inadequate representations of real-world systems" (*Savenije and Hrachowitz*, 2017). Moreover, excessive reliance on parameter calibration leads to deficient performance at interior points of the basin or at other locations at which the model was not calibrated (*Brynjarsdottir and O'Hagan*, 2014; *Lerat et al.*, 2012; *Pokhrel and Gupta*, 2010).

There have been many attempts to improve the parameterization of lumped and semi-distributed models by further discretizing the sub-basins into a given number of regions that exhibit nearly similar hydrologic behavior, i.e., the so-called HRU concept initially proposed by *Leavesley et al.* (1983) and further developed by others (e.g., *Beldring et al.*, 2003; *Blöschl et al.*, 2008; *Flügel*, 1995; *Viviroli et al.*, 2009; *Zehe et al.*, 2014). Unfortunately, results obtained in these parameterization attempts have not been very successful in realistically representing the spatial variability of model parameters, states, and fluxes because of the lack of regionalized parameters and the unabridged reliance on parameter calibration to improve model performance (*Kumar et al.*, 2010). Commonly, the effective parameters estimated for the HRUs are found by automatic calibration. Efforts have been made to enforce continuity on parameter fields (*Gotzinger and Bárdossy*, 2007; *Singh et al.*, 2012) but with somewhat limited success during the transferability of parameters across scales and locations. In addition, models parameterized using HRUs do not lead to mass conservation of water fluxes (i.e., flux-matching) when applied to scales other than those used for calibration (*Kumar et al.*, 2010, 2013). Recent attempts have been made to improve the HRU concept to increase the seamless representation of parameters, states, and fluxes (*Chaney et al.*, 2016a). However, this concept has not been tested for scalability and seamlessness of the estimated fields at coarse resolutions. Lately, a thermodynamic reinterpretation of the HRU concept was proposed by *Zehe et al.* (2014), but to date, the implementation of this approach has not found its way into meso-scale to macro-scale LSMs/HMs.

The representative elementary watershed approach (*Reggiani et al.*, 1998) is an interesting theoretical concept, which scales mass and momentum balance equations. Unfortunately, to the best of our knowledge, it has not been used to estimate effective parameters at meso- and regional scales.

A priori regularization functions (e.g., pedotransfer functions) were introduced by *Koren et al.* (2013) to ensure the "inappropriate randomness in the spatial patterns of model parameters", i.e., the lack of seamlessness. Unfortunately, in this case, the parameters (or coefficients) of regularization functions were not subject to parameter estimation or to the verification of their ability to predict fluxes and states across various scales. The use of empirical point-scale-based relationships to link geophysical characteristics with LSM/HM parameters and the assumption that their coefficients are universally applicable with certainty (e.g., the coefficients in the *Clapp and Hornberger* (1978) pedotransfer functions) are the major reasons for the proliferation of hidden parameters in LSM/HM code (*Cuntz et al.*, 2016; *Mendoza et al.*, 2015). It is of pivotal importance to understand that these point-scale relationships should not be applied beyond the scale at which they were derived.

Many types of regionalization (or regularization) approaches have been tested for semi-distributed and distributed models. According to *Samaniego et al.* (2010a), these approaches can be broadly classified into post-regionalization and simultaneous regionalization approaches, depending on if the regionalization function parameters (or global parameters) are estimated after (*Abdulla and Lettenmaier*, 1997; *Livneh and Lettenmaier*, 2013; *Seibert*, 1999; *Wagener and Wheater*, 2006) or during the model calibration (*Fernandez et al.*, 2000; *Gotzinger and Bárdossy*, 2007; *Hundecha and Bárdossy*, 2004; *Pokhrel and Gupta*, 2010). None of these procedures consider the subgrid variability of the model parameters or geophysical characteristics. *Livneh and Lettenmaier* (2013) noted that most of these regionalization procedures exhibit limited transferability because of the use of discrete soil texture classes as predictors, and very likely discontinuous parameter fields.

Recently, a dissimilarity-based regionalization technique was used by *Beck et al.* (2016) to generate an ensemble of global parameters of the Hydrologiska Byråns Vattenbalansavdelning (HBV) model at a 0.5° resolution for global-scale hydrological modeling. A shortcoming of this approach is the use of ad hoc nearest-neighbor interpolation of parameter fields to fill gaps where no donor basins are available in (geographically) surrounding regions. Following a similar concept of that of *Beck et al.* (2016), the parameterization method proposed by *Bock et al.* (2016) for the contiguous United States (CONUS) will likely lead to discontinuous parameter fields for reasons similar to those mentioned above.

Many attempts have been made in the land surface modeling community to address Dooge's challenges, especially with respect to the transferability of model parameters across locations and scales, and to obtain seamless parameter fields. One of the earliest prominent experiments was conducted in the Project for Intercomparison of Land-surface Parameterizations (PILPS) (*Wood et al.*, 1998). In this project, calibrated LSM parameters were transferred from small catchments to their nearest computational grid cells. The results indicated that LSMs exhibited poor transferability across space, leading to significant differences in the partitioning of water and energy fluxes. For instance, *Troy et al.* (2008) used calibrated variable infiltration capacity (VIC) model parameters from small basins to generate parameter fields for continental-scale land surface modeling by "linearly interpolating to fill in those grid cell not calibrated" on a sparse grid. As noted by *Samaniego et al.* (2010a), this type of regionalization is inadequate because of the nonlinearity of soil and geological formations. The spatial patterns of model parameters that would be obtained by ad hoc extrapolations based on calibrated parameters from small basins or grid cells would most likely lead to unrealistic parameter fields with spatial discontinuities circumscribing river basins, as shown in recent studies by *Wood and Mizukami* (2014) and *Mizukami et al.* (2017) for the VIC model parameters.

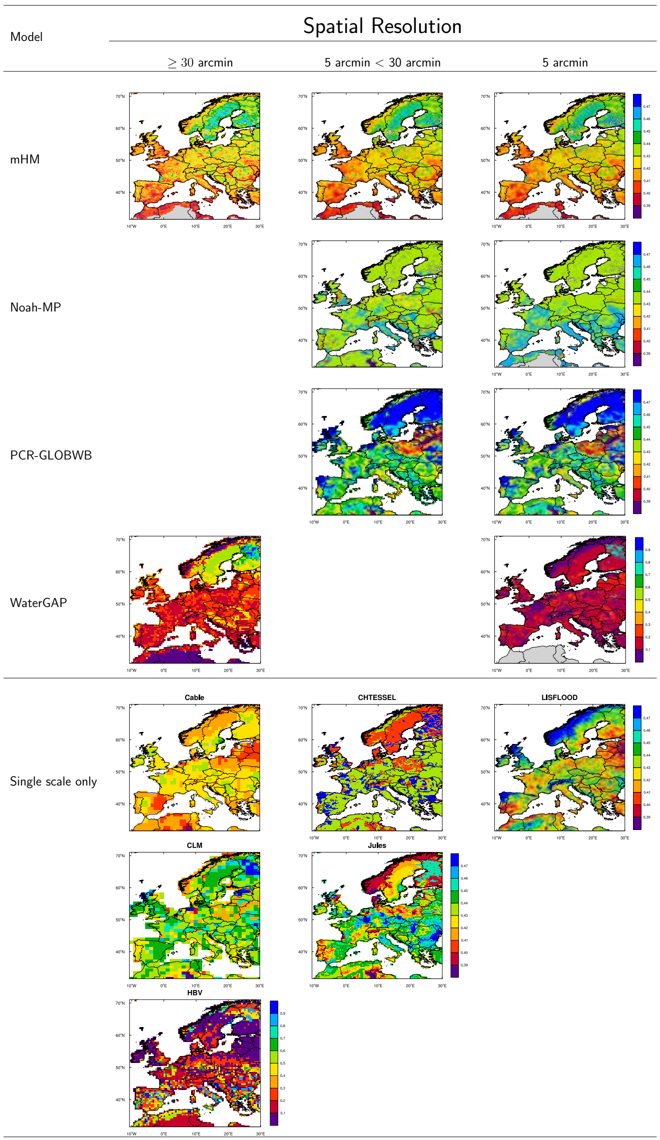

Figure 5.1 Porosity fields (top 2 m) of typical LSM/HM over Pan-EU at various resolutions: CABLE (1°), CLM (1°), CHTESSEL (0.11°), JULES (35 km), LISFLOOD (EFAS, 5 km), mHM (EDgE-C3S, 5 km), Noah-MP (CORDEX-EU, 0.11°), and PCR-GLOBWB (EDgE-C3S, 5 km). Normalized available water capacity of WaterGAP2 (HyperHydro, 30 arcmin), [3, 536] mm, WaterGAP3 (HyperHydro, 5 arcmin), [1, 960] mm, and HBV [50, 698] mm. In brackets, the normalization values, denoted as [min, max], are provided only for HBV and WaterGAP.

Recent community-driven efforts, such as the Protocol for the Analysis of Land Surface Models (PALS) and the Land Surface Model Benchmarking Evaluation Project (PLUMBER) (*Haughton et al.*, 2016), indicate that the hurdles noted in PILPS have not been overcome. Thus, it is required to gain understanding on whether the inferior predictability of many LSMs evaluated with empirical benchmarks in the PLUMBER project (e.g., CABLE, CHTESSEL, JULES, Noah) may be the result of deficient parameterizations, among other factors.

5.3.2 Parameterization of soil porosity and available water capacity in selected LSMs/HMs

The above-mentioned challenges that we face in estimating key physical parameters in LSMs/HMs have been intensively discussed in many studies (*Bierkens*, 2015; *Bierkens et al.*, 2014; *Clark et al.*, 2016, 2017; *Gupta et al.*, 2014; *Mizukami et al.*, 2017; *Peters-Lidard et al.*, 2017). To further visualize the problems and to understand the deficiencies of current parameterization techniques, we selected a representative sample of LSMs/HMs used for research and/or operational purposes, namely CABLE, CLM, JULES, LISFLOOD, Noah-MP, mHM, PCR-GLOBWB, WaterGAP2 (30 arcmin), WaterGAP3 (5 arcmin), CHTESSEL, and HBV. These models vary in process complexity and spatial resolution.

We selected soil porosity as an example to visualize existing shortcomings because it is one of the most common parameters in many LSMs/HMs. This parameter controls the dynamic of several state variables and fluxes such as soil moisture, latent heat, and soil temperature, and its sensitivity has been demonstrated in various studies (*Cuntz et al.*, 2015, 2016; *Goehler et al.*, 2013; *Mendoza et al.*, 2015).

A representation of the porosity of the top 2 m soil column in these models over the Pan-European domain (Pan-EU) is shown in Fig. 5.1. The Pan-EU domain was selected for depiction, but we note that the problem is general and persistent across other domains (*Mizukami et al.*, 2017). For cases in which a HM does not use this parameter, the "available water capacity" (WaterGAP) or the "field capacity" (HBV) were selected as a surrogate due to their similarity with porosity. Both surrogate fields are normalized (in space) to ease their comparison with the porosity fields. Soil porosity is expressed in $m^3\,m^{-3}$ to ease the comparison among different models.

The following lessons can be learned from Figure 5.1: 1) There is a large variability in the parameterization of this key physical parameter because none of the analyzed models have comparable spatial patterns or comparable estimates at a given location. It should be noted that the definition of the selected parameter is rather simple: it represents the ratio of the volume of voids to the total volume in the soil column. One can now wonder how large the uncertainty of other parameters would be (e.g., hydraulic conductivity) whose relationship with soil properties is very nonlinear. 2) The degree of seamlessness strongly depends on the level of aggregation and the upscaling of underlying soil texture fields. For example, the proxy of porosity for WaterGAP is substantially different in spatial pattern and magnitude for 30 arcmin and 5 arcmin simulations. On the contrary, the spatial pattern and magnitude for porosity used in mHM remain almost unchanged for application at 30 and 5 arcmin resolution. 3) A parameter field becomes highly discontinuous and patchy when, for a given model, the parameter is calibrated in a limited domain (or basins) and then extrapolated to other regions (e.g., as shown in the panel corresponding to the HBV). 4) These experimental results confirm the postulation of *Dooge* (1982) that the parameterization of the existing state-of-the-art LSMs/HMs at large and continental scales is still an unsolved problem.

The analysis of current parameterization techniques allow us to put forward the following questions: 1) Why are there such large differences between models in estimating a parameter that has a physical meaning? 2) What are the consequences of poor parameterizations on the spatiotemporal dynamics of state variables and fluxes? 3) What are the consequences of model calibration on parameter fields? 4) Are current model parameterizations scale invariant? 5) Do the fluxes estimated with these models at various scales satisfy the fundamental mass conservation criterion (hereafter denoted as the flux-matching test)?

Table 5.1 Data sources and parameterization method used by models used in this study

Model	Parameterization Method	References	Source code & Projects
CABLE	Pedo-transfer functions, look-up table, dominant soil type	*Kowalczyk et al.* (2006)	`www.cawcr.gov.au/publications/technicalreports/CTR_057.pdf`
CLM	Pedo-transfer functions, look-up table, mosaic approach	*Oleson et al.* (2013)	`www.cesm.ucar.edu/models/cesm1.2/clm/`
CHTESSEL	Look-up table, dominant soil type	*ECMWF* (2016); *Viterbo and Beljaars* (1995)	`www.ecmwf.int/search/elibrary`
HBV	k-NN interpolation, calibrated parameter	*Beck et al.* (2016)	`www.gloh2o.org/hbv-simreg/`
JULES	Look-up table, dominant soil type	*Best et al.* (2011)	`jules.jchmr.org`
LISFLOOD	Pedo-transfer functions, mosaic approach, arithmetic mean	*De Roo and Wesseling* (2000)	`ec.europa.eu/jrc/en/publication/`
mHM	MPR	*Samaniego et al.* (2010a)	`edge.climate.copernicus.eu www.ufz.de/mhm`
Noah-MP	Look-up table, dominant soil type	*Niu* (2011)	`www.jsg.utexas.edu/noah-mp www.meteo.unican.es/wiki/cordexwrf`
PCR-GLOBWB	(Original) pedo-transfer functions with averaged predictors (New) MPR	*van Beek et al.* (2011); *Wada and Bierkens* (2014) *Samaniego et al.* (2010a)	`pcraster.geo.uu.nl/projects/applications/pcrglobwb/`
WaterGAP (2,3)	Look-up tables	*Batjes* (1996); *Müller Schmied et al.* (2014)	`www.uni-kassel.de/einrichtungen/en/cesr/research/projects/active/watergap.html www.uni-frankfurt.de/45218063/WaterGAP`

5.4 Seamless parameterization framework

5.4.1 The flux-matching postulation

The key postulation aiming at obtaining scalable (global) parameters that are transferable across locations and scales was proposed by *Samaniego et al.* (2010a) and further tested in *Kumar et al.* (2013,b) and *Rakovec et al.* (2016a). We hypothesize that flux matching across scales leads to quasi-scale-invariant global parameters $\hat{\gamma}$; thus,

$$\sum_i \sum_t \left| W_i(\hat{\gamma}, t) a_i - \sum_{k \in i} w_k(\hat{\gamma}, t) a_k \right| \to 0, \quad \forall i \in \Omega. \tag{5.1}$$

Here, k denotes the subgrid elements constituting a given modeling cell i with area a_k. i denotes a modeling grid cell i with area a_i. W_i and w_k denote fluxes at two modeling scales ℓ_1 and ℓ_1', respectively, with $\left(\frac{\ell_1}{\ell_1'}\right)^2 = \frac{a_i}{a_k}$. Ω denotes the modeling domain, e.g., a river basin, and t a point in time. It should be noted that the topology of the cells at either level is not specified. Normally, rectangular grid cells are used for convenience, but this is not a necessary condition. This strong flux-matching condition can be used as a penalty function or as an additional test to discriminate parameter sets obtained with conventional parameter estimation approaches.

5.4.2 The MPR approach

MPR, proposed by *Samaniego et al.* (2010a), aims to estimate model parameters that are seamless across scales, satisfy the flux-matching conditions (see Sect. 5.4.1), and enable the transferability of global or transfer-function parameters across scales and locations (*Kumar et al.*, 2013b; *Livneh et al.*, 2015; *Rakovec et al.*, 2016b; *Samaniego et al.*, 2010a,b; *Wöhling et al.*, 2013). The development of MPR is ongoing. Regionalization functions used in MPR for the mHM model (`www.ufz.de/mhm`) by *Samaniego et al.* (2010b) were further improved by *Kumar et al.* (2013). More recently, a model-agnostic implementation of MPR has been proposed by *Mizukami et al.* (2017) and tested in the VIC model in over 500+ basins in the CONUS. The study of *Mizukami et al.* (2017), in contrast to the present study, does not include flux-matching tests nor the evaluation of model skill across different spatial scales.

The scaling problem in MPR is addressed by using process-specific representative elementary areas (REAs) that determine the minimum computational grid size ℓ_1 at which the continuum assumptions can be used without explicit knowledge of the actual patterns of the topography, soil, or rainfall fields (*Wood et al.*, 1988). The REA of a specific process, such as streamflow, can be determined by conducting a careful sensitivity analysis as shown by *Samaniego et al.* (2010a). To estimate an "effective" model parameter (e.g., total soil porosity) at the selected modeling scale, it is first necessary to estimate its variability at a much finer scale $\ell_0 \ll \ell_1$ such that the effects of its spatial heterogeneity can be adequately represented. In other words, the parameter at the fine scale ℓ_0 represents the minimum support at which the proposed equations are still valid. *Barrios and Francés* (2011) indicated that a suitable estimate of ℓ_0 for a given parameter could be near its correlation length. The subgrid variability of a parameter $\boldsymbol{\beta}_0$ depends, in turn, on the spatial heterogeneity of geophysical and biophysical characteristics ($\mathbf{u}_0$), such as terrain elevation, slope and aspect, soil texture, geological formation, and land cover, which are now available at hyper-resolution for the entire globe. The mathematical relationships that link model parameters with these characteristics at the finer resolution are called pedotransfer, regionalization, or regularization functions f (*Clapp and Hornberger*, 1978; *Cosby et al.*, 1984; *Wösten et al.*, 2001). The constants required in these functions are usually denoted as global parameters $\hat{\gamma}$; thus, $\boldsymbol{\beta}_0 = f(\mathbf{u}_0, \hat{\gamma})$. Note that the fields $\boldsymbol{\beta}_0$ and $\mathbf{u}_0$ are dependent on space and time, but the vector $\hat{\gamma}$ is not.

Regularization functions are commonly used in mathematics and statistics to solve ill-posed problems (which is the case when the parameters of a distributed LSM/HM are determined by calibration) and/or to prevent overfitting. The direct consequence of the regularization is the substantial decrease in degrees of freedom of the optimization problem because the cardinality of the gridded parameter fields $\#\{\boldsymbol{\beta}_0\}$ is orders of magnitude larger than that of the vector of the global parameters $\#\{\hat{\gamma}\}$. Hence, MPR is a parsimonious parameterization technique that offers spatially continuous parameter fields and removes spatial discontinuities in water fluxes and states, as observed by *Gotzinger and Bárdossy* (2007) and discussed by *Mizukami et al.* (2017). From the Bayesian point of view, the regularization functions impose a prior distribution on the model parameters. Consequently, greater care should be taken in their selection.

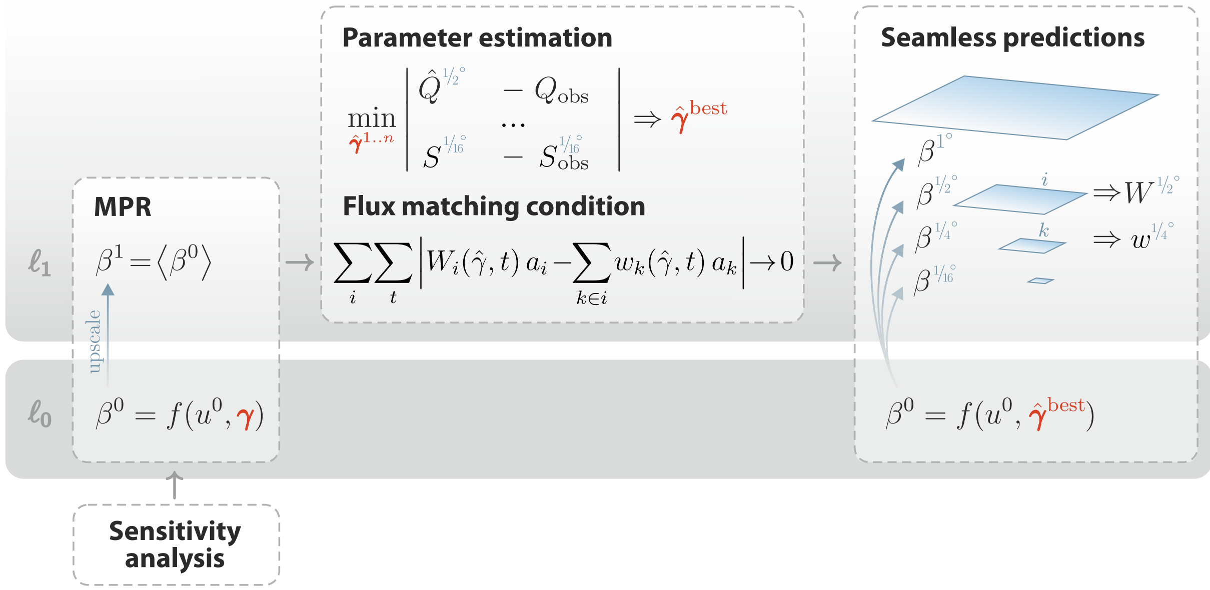

Figure 5.2 Schematic representation of the proposed seamless prediction framework based on *Rakovec et al.* (2016a). It includes a preliminary sensitivity analysis, MPR estimation, global-parameter estimation, a flux-matching test, and multiscale seamless prediction. W_i and w_k are the fluxes at the i and k cells of the $1/2°$ and $1/4°$ resolutions, respectively (as an example). Q_{obs} and S_{obs} are the observed time series of streamflow and soil moisture, respectively. The operator $|\cdot|$ is a compromise dissimilarity metric composed of many independent observations at various scales.

The second step of the MPR approach consists of upscaling the subgrid distribution of a regionalized parameter to the modeling scale. In other words, $\boldsymbol{\beta}_1 = \langle \boldsymbol{\beta}_0 \rangle$. Here, the symbol $\langle \cdot \rangle$ represents an averaging or scaling operator that is parameter specific, and thus $\boldsymbol{\beta}_1$ denotes the upscaled effective parameter field. It is important to note that this scaling operator is not necessarily the arithmetic mean.

A schematic representation of the MPR procedure can be seen in Fig. 5.2. In short, the motto of MPR is “estimate first, then average”, whereas other existing regionalization methods follow the opposite approach of “average first, then estimate.” Because the processes in LSMs/HMs are highly nonlinear, this sequence of operations does not commute. The consequences can be dramatic (to be shown in the results section). The latter, which is the standard approach, does not preserve fluxes/states across scales, whereas MPR does to a considerable extent. The key question here is in finding the right scaling rule for the model parameters such that the fluxes/states are preserved across a range of spatial scales.

Model parameters at the ℓ_1 scale (i.e., 1 to 100 km) are called “effective” parameters because they cannot be measured by physical means at this resolution and can only be inferred by heuristic relationships $f(\cdot)$. Thus, it is essential that the inequality $\ell_0 \ll \ell_1$ is fulfilled so that the law of large numbers leads to stable estimates of the effective parameter $\boldsymbol{\beta}_1$ having low uncertainty. Since every LSM/HM (e.g., those mentioned in Sect. 5.3) contains “effective” model parameters, depending on heuristic relationships (that are hidden in the source code in many cases; *Cuntz et al.*, 2016; *Mendoza et al.*, 2015), it is logical that existing LSMs/HMs are subject to parameter uncertainty. These models can be treated as stochastic models, even though their governing equations are deterministic in nature and based on physical principles such as the conservation of mass and energy (*Clark et al.*, 2015; *Nearing et al.*, 2016). Effective parameters should not be the pure result of a blind calibration algorithm. MPR varies from other regionalization approaches in that the introduced relationships may lead to seamless parameter fields and model simulations fulfilling the flux-matching condition.

Currently, MPR is the only method that consistently and simultaneously addresses the scale, nonlinearity, and overparameterization issues if global parameters are estimated simultaneously at multiple locations (i.e., basins). The MPR approach also addresses the principle of scale-dependent subgrid parameterization (i.e., “net fluxes must satisfy the conservation of mass” proposed by *Beven*, 1995) but does not adhere to Beven’s other principles, such as that subgrid parameterizations may be data and scale dependent (principles 3 and 4 in *Beven*, 1995), because exhaustive tests reported in the above-mentioned references carried out over hundreds of river basins do not appear to support them. We find MPR to be a robust technique that has the ability to provide "effective parameters" and is capable of addressing the scaling problem; in this sense, it diverges from the Beven's view (*Beven*, 1995, p. 507) that these "effective parameters" are an "inadequate approach to the scale problem". Furthermore, MPR differs on the regionalization and aggregation scheme (i.e., patch model areal weighting) proposed by *Beven* (1995, p. 520).

The selection of regionalization functions and scaling operators is fundamental to ensuring the transferability of global parameters across scales and to guarantee the seamlessness of parameter fields across scales, e.g., from ℓ_1 to $2\ell_1$ and so on. *Samaniego et al.* (2010a) proposed that the key to determining them is the flux-matching condition mentioned above. A seamless parameter field $\boldsymbol{\beta}_1$ can be interpreted as the corollary of the flux-matching condition. Moreover, MPR employs geophysical properties at ℓ_0 that allow for a representative sample at the hyper-resolution promoted by *Wood et al.* (2011) and *Bierkens et al.* (2014).

5.4.3 Protocol for implementing the MPR approach

The development of LSMs/HMs and their parameterizations should be guided by a strict hypothesis-driven framework (*Nearing et al.*, 2016) that aims at finding parsimonious and robust parameter sets that fulfill the flux-matching condition and a number of efficiency metrics that are not used during the parameter estimation phase. A multivariate, multiscale evaluation assessing the reliability of model simulations should follow the scheme presented in *Rakovec et al.* (2016b). Based on our previous experiences, we synthesize a formalized scheme (i.e., protocol) for systematically implementing the MPR technique in other LSMs/HMs with the aim to obtain a robust and seamless parameterization. A graphical depiction of the estimation procedure at multiple scales is shown in Fig. 5.2.

1. Retrofit the source code of an LSM/HM so that all model parameters are exposed to analysis algorithms. Parameters are the values of a model that can be considered random variables, i.e., those that are subject to various outcomes and can be fully defined by a probability density function. Parameters should not be confused with numerical or physical constants.

2. Determine a set of the most sensitive model parameters through a sensitivity analysis (SA). For computationally expensive LSMs such as CLM or Noah-MP, computationally frugal methods such as the elementary effects method (*Morris*, 1991), its enhanced version such as that proposed by *Cuntz et al.* (2015), or the distributed evaluation of local sensitivity analysis (DELSA; *Mendoza et al.*, 2015; *Rakovec et al.*, 2014) are of particular interest because use of the popular standard Sobol' method (*Sobol'*, 2001) can be computationally expensive although still possible (*Cuntz et al.*, 2016).

3. Regionalize sensitive model parameters that exhibit marked spatial variabilities. The selection of the regionalization function $f(\cdot)$ can be guided by existing literature or by step-wise methods (e.g., *Samaniego and Bárdossy*, 2005). This regularization step should be conducted at the highest available spatial resolution for all predictor fields. This resolution is denoted as level ℓ_0. The output of the regularization is the parameter field $\boldsymbol{\beta}_0$.

4. Estimate effective parameter fields $\boldsymbol{\beta}_1$ using upscaling operators based on the underlying subgrid variability $\boldsymbol{\beta}_0$. The scale ℓ_1 is determined by synthetic experiments aimed at finding the optimal REA for processes related to the parameter in question (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a).

5. Estimate the global parameters $\hat{\gamma}$ using standard optimization algorithms (simulated annealing, shuffled complex evolution (SCE), dynamically dimensioned search (DDS)) by minimizing a compromise metric that includes observations at multiple scales and locations (*Duckstein and Opricovic*, 1980; *Rakovec et al.*, 2016b). The compromise metric could also include hydrologic signatures to extract as much information from a time series as possible (*Nijzink et al.*, 2016).

6. Perform multi-basin, multiscale, multivariate cross-validation tests to evaluate the robustness of the regionalization functions, scaling operators, and global parameters (*Rakovec et al.*, 2016b).

7. Evaluate the parameter seamlessness and the preservation of the statistical moments of fluxes and states across scales (seamless prediction step in Fig. 5.2).

8. If the cross-validation tests provide satisfactory results (e.g., Kling–Gupta efficiency (KGE) of the compromise solution > 0.6), then evaluate the flux-matching condition given by Eq. (5.1). If the total error is too large to be tolerated, repeat steps 3 to 8.

It should be noted that any of the steps above can be tested within a sequential hypothesis-testing framework (*Clark et al.*, 2016). A substantial difference from a standard model optimization exercise is that the transfer function $f(\cdot)$ (step 3) and the upscaling operator (step 4) can also be modified in the modeling protocol.

Failure to satisfy the imposed condition, such as the flux-matching test, after exhaustively testing the options in steps 3 to 6 may indicate deficits in process understanding and/or poor data. Consequently, the evaluation step should also provide guidance on detecting and separating the errors stemming from process conceptualization (modeling) and input data.

5.4.4 Seamless parameter fields across multiple scales using MPR

In Sect. 5.4.2, it was postulated that the MPR technique aims at estimating seamless parameter fields across scales which minimize the occurrence of artificial discontinuities and ease the transferability of model parameters across scales and locations. The latter has been tested and reported in many studies in Europe, USA, and other basins worldwide (*Kumar et al.*, 2013,b; *Rakovec et al.*, 2016a,b; *Samaniego et al.*, 2011). In this study, we provide evidence in favor of the former postulation.

To achieve this goal, the mHM model is parameterized using MPR (*Samaniego et al.*, 2010a) with hyper-resolution fields of geophysical characteristics at $\ell_0 = 500$ m resolution as input. Among them, the land cover data were obtained from the Corine datasets (`http://land.copernicus.eu/pan-european/corine-land-cover`), and the soil texture information was derived from SoilGrids (`soilgrids.org`). These very detailed and homogenized soil texture fields provide the fractions of clay and sand, mineral bulk density, and fraction of organic matter for six soil horizons up to 2 m deep. A hyper-resolution digital elevation model (DEM) over Europe (approximately 30 m) from the GMES RDA project (EU-DEM; `www.eea.europa.eu/data-and-maps/data/eu-dem`) was used to derive terrain characteristics such as slope, aspect, and flow direction. The underlying hydrogeological characteristics are based on the International Hydrogeological Map of Europe (IHME; `www.bgr.bund.de/ihme1500`), available at a 1 : 1 500 000 scale. Details on the pedotransfer function used for these simulations can be found in *Livneh et al.* (2015). mHM global parameters were obtained by closing the water balance over selected river basins in Europe (*Rakovec et al.*, 2016b).

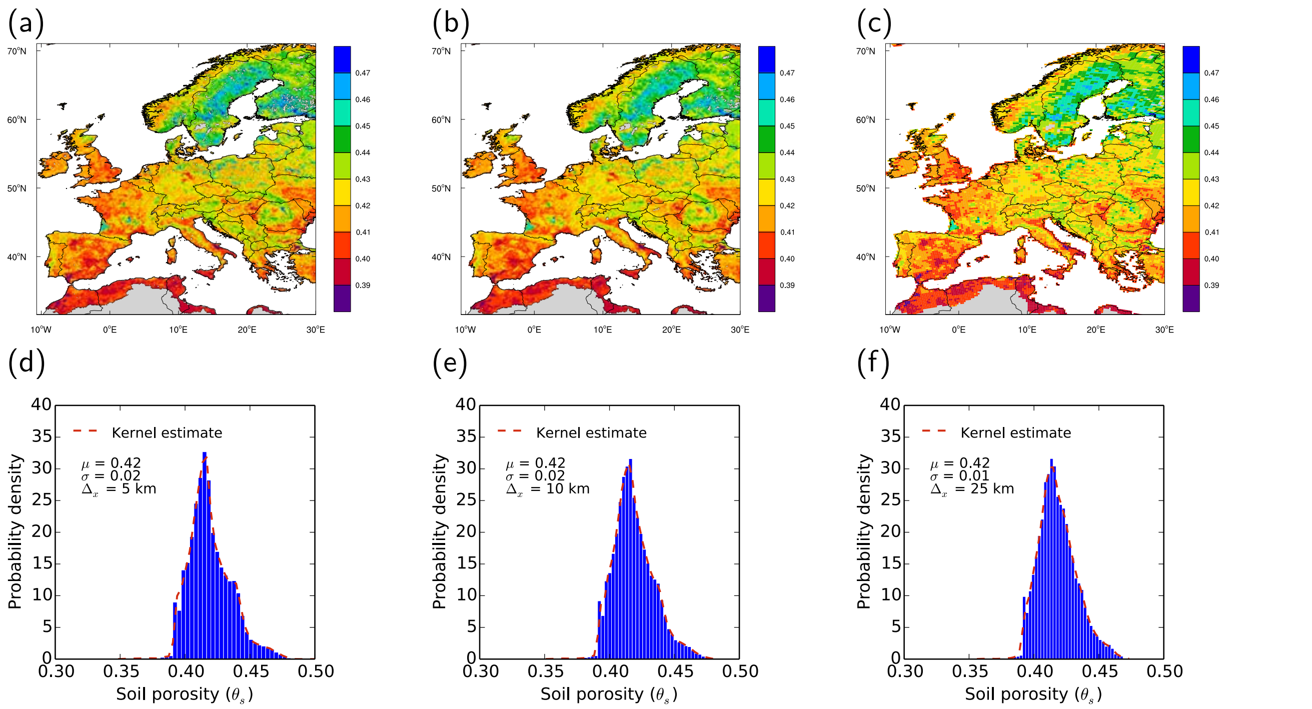

Figure 5.3 Seamless soil porosity (top 2 m) fields obtained using MPR at three spatial resolutions ℓ_1: **(a)** 5 km, **(b)** 10 km, and **(c)** 25 km, respectively. Lower panels **(d)**–**(f)** show the empirical distribution function of porosity at the respective resolution and method.

Based on these settings, which constitute the basis for the EDgE project (edge.climate.copernicus.eu), we estimated porosity fields at three modeling resolutions of ℓ_1= 5, 10, and 25 km, based on the same ℓ_0 support information. Following the MPR procedure depicted in Fig. 5.2, the parameter fields for the mHM model at these three resolutions can be estimated. Results are shown in Fig. 5.3.

The results illustrate that the MPR approach can preserve the spatial pattern of the porosity fields (see Fig. 5.3a, b, and c) and the first and second moments of its probability density function shown in Fig. 5.3e–g. Two-sample Kolmogorov–Smirnov tests indicate that there is insufficient evidence to reject the null hypothesis that any of the three possible pairs of empirical distributions were drawn from the same unknown distribution. This highlights that the MPR approach leads to consistent parameter fields across scales. In this case, the mean porosity is estimated to be 0.42 $m^3\,m^{-3}$ independent of the scale.

5.4.5 Limitations of the MPR approach

The MPR approach, as any method, has some limitations. One of the crucial aspects of MPR is the selection of transfer functions and upscaling operators. Existing theories could be the first guess, but in the event that nothing is available, the protocol proposed in Sect. 5.4.3 could be used to guide the search of robust transfer functions. Testing the model parameterization for flux-matching conditions across a range of basin and spatial scales may help to identify adequate upscaling operators. This procedure, although tedious, is the only solution for the moment.

In the event that some state variables change over time (e.g., land cover/use), or during parameter estimation, the MPR algorithm has to be linked to the model because every time a global parameter ($\hat{\gamma}$) is re-estimated, all related model parameters ($\boldsymbol{\beta}_1$) have to be updated as illustrated in Fig. 5.2. The computational cost of performing MPR is therefore larger than other parameterization method discussed before.

Another limitation of the applicability of the MPR technique until recently was its availability only as an intrinsic module of the mHM model (`www.ufz.de/mhm`). This implies that tailored algorithms (i.e., source code) to perform the regionalization and upscaling of parameters for a target LSM/HM have to be developed from scratch, as it is demonstrated here as a case study for the PCR-GLOBWB model. This activity is of course time-consuming and not pleasing due to its complexity. For this reason, *Mizukami et al.* (2017) have started a community effort to develop a model-agnostic MPR implementation (MPR-flex), which has been so far evaluated for the VIC model.

The availability of high-resolution biophysical characteristics at the spatial scale ℓ_0 constitutes another limitation of the applicability of MPR. Since the subgrid variability is fundamental to estimating robust effective parameter values at coarser scales, the minimum scale at which a model can be applied (ℓ_1) is strongly determined by the data availability. For example, if the soil data are available for the Pan-EU domain at $\ell_0 = 250$ m, the ℓ_1 should not be lower than 1000 m, so that each modeling cell (ℓ_1) has a representative number of underlying subgrid cells (ℓ_0).

MPR has been mainly developed for a hydrologic model representing the water cycle. However, land surface models also include the energy and carbon cycles and thus have greater complexity. In particular, they have more detailed representation of vegetation. It is a topic for future research to develop a MPR approach (i.e., transfer functions and upscaling operators) for plant functional-type-specific parameters such as carboxylation rate and the slope of the Ball–Berry equation for stomatal conductance (*Ball et al.*, 1987), which are required for a successful implementation of MPR in LSMs.

Finally, the computational effort for MPR is also considerably larger in comparison with other methods, because of its requirement to estimate model parameters ($\boldsymbol{\beta}_0$) at the highest resolution at which the biophysical characteristics are available. The computational time, however, could be substantially reduced by using a restart file (i.e., a dataset containing a copy of all parameters, state variables, and fluxes of a model at a given point in time). If this capability is available, the MPR estimation can be greatly reduced for operational simulations because the effective parameter fields and past modeled states do not need to be estimated often.

5.5 Experiments to reveal non-seamless parameterizations

In this section we perform four modeling experiments, inspired on *Wood* (1990)'s recommendation, to investigate: 1) the effects of the over-calibration of global parameters on the spatial patterns of modeled state variables. 2) The effects of a parameterization technique on the spatial pattern of effective parameters. 3) The effects of a parameterization technique on the dynamics of a state variable. And, 4) the effects of not satisfying the flux-matching condition on simulated flux across different spatial scales. In these experiments four models are employed: mHM, Noah-MP, PCR-GLOBWB, and WaterGAP.

5.5.1 Effects of on-site model calibration

As noted in the introduction, on-site (basin-specific) parameter estimation based on HRU or similar techniques (such as clustering grid cells or sub-basins into regions that exhibit quasi-similar hydrological behavior) leads to non-seamless parameter fields such as those reported in *Merz and Blöschl* (2004). Here, we go one step further to show the consequences of this common practice on state variables such as soil moisture. Our postulation is that an on-site calibration of global parameters $\hat{\gamma}$ leads to biased state variables even with regularization techniques such as MPR. To falsify this postulation, we performed two model simulations denoted "on-site" and "multisite" calibration schemes. In both cases, we used the mHM setup described in *Rakovec et al.* (2016a) over the Pan-EU domain at a 0.25° resolution.

In the first simulation, we perform on-site calibrations at 400 river basins in the Pan-European domain. Subsequently, the respective optimized parameter sets are used in each corresponding basin to generate the target variable, in this case, the daily soil moisture of the top 1 m soil column. Lastly, daily soil moisture fields are assembled using the independent basin simulations for the entire Pan-EU domain. The results of this experiment are shown in Fig. 5.4a for a day in August 2005. In the second simulation, the global parameters $\hat{\gamma}$ are estimated simultaneously for a set of 13 basins covering various hydroclimatic regimes in the Pan-EU domain. The corresponding soil moisture field for the same point in time is depicted in Fig. 5.4b.

The first simulation shows clear evidence of strong spatial imprint in the soil moisture fields that is easily identifiable because the shapes of the constituent river basins (Fig. 5.4a) are apparent. Another interesting feature is a strong wet bias in a basin located in center of the Iberian Peninsula compared to its neighboring regions. Wet soils during this period are very unlikely because the entire region was enduring a prolonged and extreme drought. Moderate dry bias is apparent in basins in southwest Germany, and a strong dry bias was detected in basins in west Croatia, south Lithuania, south Hungary, and north Bosnia and Herzegovina. Conversely, the soil moisture field obtained with the multi-basin parameter estimation does not exhibit these nuisances and thus can be regarded as a spatially seamless field. In this case, parameter estimation with a large sample of geophysical characteristics and many streamflow time series to estimate efficiency measures leads to a well-posed parameter estimation problem.

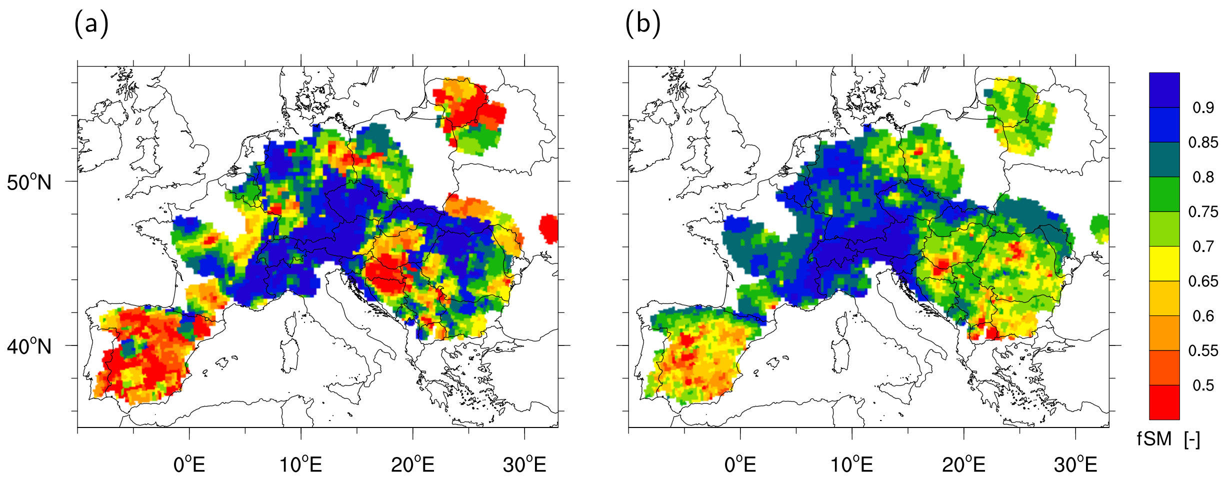

Figure 5.4 mHM simulations of soil moisture as the fraction from saturation $\frac{\theta}{\theta_s}$ for a day in August 2005 conducted with **(a)** basin-wise parameter estimation and **(b)** seamless parameter estimation. Panel **(b)** shows a seamless soil moisture field.

Based on these results, it can be concluded that parameter sets obtained using the on-site parameter estimation technique do not lead to seamless parameter fields or state variables. Moreover, automatic optimization algorithms, such as SCE or DDS, tend to overlearn from time series with large observational errors, which in turn leads to poor identifiability of parameters (*Brynjarsdottir and O'Hagan*, 2014) and biased simulations, as demonstrated above. Consequently, parameter estimation should be performed with a representative sample of basins that adequately cover the variability of hydrological regimes and geophysical properties (e.g., soil types) (*Kumar et al.*, 2015). It is worth noting that if the parameters of a model are estimated in a small basin with very few soil types, a single geological formation, or very flat terrain, then it is very likely that some parameters cannot be constrained during calibration. The obtained parameter set is biased to the specific basin in which it has been estimated, and hence it is not skillful for seamless and continental-scale simulations.

5.5.2 Effects of a parameterization technique on spatial patterns of effective parameters

The effects of the commonly used parameterization techniques to generate the porosity fields of LSMs (such as CHTESSEL and Noah-MP depicted in Fig. 5.1) are important to investigate. These fields are obtained by combining the majority (or dominant) upscaling operator and a look-up table containing categorical values of model parameters tabulated for a limited set of dominant soil types (e.g., *Niu*, 2011, p. 20., *ECMWF*, 2016, p. 137). The majority-based operator is mostly used for estimating grid-specific vegetation classes in LSMs (*Li et al.*, 2013a).

The porosity field, based on a majority upscaling for the Noah-MP model used in EURO-CORDEX (`www.euro-cordex.net`) at an approximately 12 km resolution, is depicted in Fig. 5.1. Compared with the other model-derived porosity fields, the Noah-MP field appears to be most homogeneously distributed in space. It is very likely that the spatial heterogeneity is underrepresented in this case as the default soil LUT contains only 13 soil classes. It should be noted that a model such as CABLE that uses a porosity field with an approximately 100 km resolution has a larger variability than that of Noah-MP at 12 km.

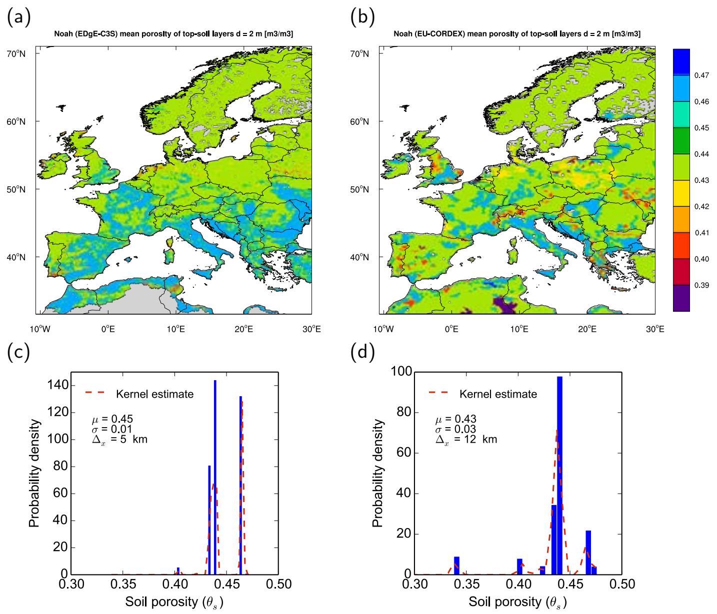

Figure 5.5 Porosity fields obtained using the majority upscale operator for spatial resolutions of **(a)** 5 km and **(b)** 12 km with the Noah-MP model used in the EDgE and EURO-CORDEX projects, respectively. Lower panels **(c)**–**(d)** show the empirical distribution function of porosity at the respective resolution and method.

The following experiment is carried out to evaluate whether the variability of the soil map or the upscaling operator has a larger effect on the derived porosity field. The highest resolution soil map available for Europe is used and applied in the same manner to derive porosity fields as described above. The texture field is provided by the SoilGrids dataset (`http://soilgrids.org`) at 1000 m resolution (level-0). The upscaled porosity field is generated at 5 km for the EDgE project. The soil characteristics for Noah-MP are estimated using the same look-up table as in the EURO-CORDEX–Noah-MP case. The comparison of both parameter fields (i.e., EDgE–Noah-MP and EURO-CORDEX–Noah-MP) and the main statistical moments describing the spatial variability of the porosity fields are shown in Fig. 5.5. The results clearly indicate the inappropriateness of the majority-based upscaling operator for this parameter in both cases. It leads to reduction of the variance of the porosity field and thus can be considered the least sensitive operator. This means that the informational content of the hyper-resolution soil maps, commonly available globally, is almost lost.

Notably, although the overall mean of the porosity estimated using MPR over the Pan-EU domain for mHM (Fig. 5.3a) is only 6.6 % lower than that calculated using the majority-based approach for Noah-MP (Fig. 5.5a), the spatial patterns obtained by both models are very different. The evidence of this remarkable dissimilarity can also be visualized by comparing the empirical density functions shown in Figs. 5.3d and 5.5c, both corresponding to a field at $\ell_1 = 5$ km and with the same input data. A detailed evaluation conducted by *Samaniego et al.* (2012) in Germany showed that large porosity values estimated with the majority-based approach could overestimate those obtained with MPR by up to 40 %, whereas in other locations, underestimation up to 15 % from those estimated by MPR can be found.

Other upscaling operators, such as the weighted arithmetic mean, are commonly used in LSMs in combination with the mosaic approach. For example, in CLM (*Oleson et al.*, 2013, see p. 160), the texture class of the subunits of the cell, called tiles, are provided in a look-up table. The upscaled porosity field obtained using this approach is shown in Fig. 5.1 at a 1° (100 km) resolution. Methods based on the majority and weighted arithmetic mean operators exhibit some similarity and lack spatial variability. In both cases, the spatial mean is approximately $0.43\,\mathrm{m}^3\,\mathrm{m}^{-3}$.

Hydrologic models that do not use soil porosity tend to use a similar conceptualization and values denoted as the total available water capacity (TAWC; WaterGAP versions 2 and 3) and field capacity (FC; HBV). For these types of conceptual models, normalized values of these parameters are used as surrogates for soil porosity. The consistency of the spatial patterns of TAWC and FC are compared here instead of their actual values. A distinctive difference in the patterns can be observed. For example, WaterGAP3 exhibits lower values than WaterGAP2, whereas the pattern of the normalized FC in HBV is the opposite in many locations (e.g., Spain, Germany, and Scandinavia).

Details of the parameterization schemes used to estimate TAWC and FC are beyond the scope of this study. Interested readers may refer to *Müller Schmied et al.* (2014) or *Beck et al.* (2016), respectively. However, the TAWC in WaterGAP is obtained by linking the soil type provided by the FAO soil map with available water capacity values estimated by *Batjes* (1996). Thus, no scaling rule or form of regularization is used in this case. The field capacity parameters used in HBV were determined using an ad hoc nearest-neighbor interpolation technique that relies on calibrated parameters from nearby similar donor basins that might exhibit very different geophysical characteristics. The parameter fields obtained for two versions of WaterGAP (30 and 5 arcmin) and HBV are depicted in Fig. 5.1. It can be concluded that the parameterization technique employed is not scale invariant as revealed by distinct parameter sets from WaterGAP model versions, which are operated at different resolutions. The regionalization proposed by *Beck et al.* (2016) leads to a patchwork-quilt field that does not resemble to any other field presented. Evident from Fig. 5.1, the HBV field lacks seamlessness that may result in non-seamless fields of water fluxes and states.

5.5.3 Effects of a parameterization technique on the dynamics of a state variable

There is a complex interplay between soil moisture (SM) and latent heat (LH) in LSMs/HMs. Improving our understanding of soil–land–atmosphere feedback is fundamental for making reliable predictions of water and energy fluxes. In this context, we carry out a sensitivity experiment to investigate the effects of soil-related parameterizations (e.g., soil porosity) on latent heat and soil moisture. Two contrasting modeling paradigms (Noah-MP and mHM) are employed.

The WRF/Noah-MP system is forced with ERA-Interim at the boundaries of the rotated CORDEX grid (`www.meteo.unican.es/wiki/cordexwrf`) at a spatial resolution of 0.11° covering Europe from 1989 to 2009. To ease the comparison, the process-based hydrological model mHM (`www.ufz.de/mhm`) is driven with daily precipitation and temperature fields generated by the WRF/Noah-MP system during the same period. The spatial resolution of mHM is fixed at $5 \times 5\,\mathrm{km}^2$. The main geophysical characteristics in WRF/Noah-MP of land cover and soil texture are represented with a $1 \times 1\,\mathrm{km}^2$ MODIS and a single-horizon, coarse-resolution FAO soil map with 16 soil texture classes, respectively. The porosity field of Noah-MP is estimated by applying a majority-based operator to values for different soil classes, as shown in Fig. 5.5b.

The settings of the mHM model used in this experiment are described in Sect. 5.4.4. In contrast to those of Noah-MP, the global parameters of mHM estimated using the MPR technique are obtained by closing the water balance over selected river basins in Europe (*Rakovec et al.*, 2016b). The porosity fields obtained for mHM over the Pan-EU are depicted in Fig. 5.3.

The phase diagrams of the monthly fraction of soil water saturation $\mathrm{fSM} = \frac{\theta}{\theta_s}$ (i.e., plots of monthly $\mathrm{fSM}(t)$ vs. $\mathrm{fSM}(t+1)$) are subsequently investigated to understand the effect of differences in porosity estimates of the top 2 m soil column on the soil moisture dynamics (Fig. 5.6). Two locations in Germany are selected in which Noah-MP systematically over- or underestimated the latent heat fluxes with respect to mHM (the latitude and longitude coordinates of the center of the selected Noah-MP grids are A (54° N, 10° E) and B (51° N, 7° E), respectively). At location A, the majority-based approach underestimates the MPR soil porosity by $-10\,\%$, whereas in location B, it overestimates it by $40\,\%$. This experiment unambiguously shows that, at locations where Noah-MP overestimates latent heat with respect to mHM, the temporal variance (i.e., dynamic) of the monthly SM time series simulated by Noah-MP is almost doubled compared to that of mHM, leading to much lower soil moisture values (Fig. 5.6a). Conversely, underestimation of latent heat greatly reduces the variance of the soil moisture dynamics (Fig. 5.6b).

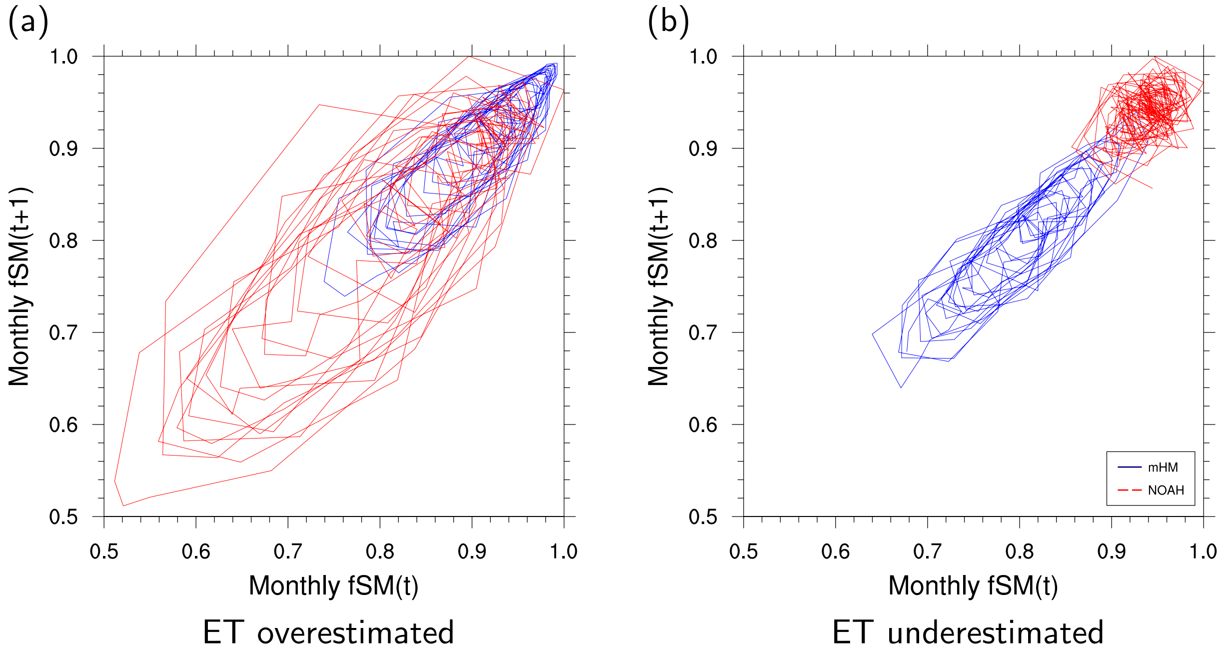

Figure 5.6 Phase diagrams of monthly soil moisture fraction for two locations in Germany, **(a)** 54° N, 10° E and **(b)** 51° N, 7° E, in which the latent heat estimated by Noah-MP is over- or underestimated with respect to corresponding estimates of mHM. The models have identical forcings.

5.5.4 Effects of not satisfying the flux matching condition

In Sect. 5.3, we postulated that ad hoc parameterization schemes do not necessarily fulfill the flux-matching test performed with a flux simulated by a given model at two modeling resolutions ($\ell_1 = 5$ and 30 arcmin). A detailed description of how to perform this test is provided in *Samaniego et al.* (2010a). The following experiment is conducted with three models (mHM, PCR-GLOBWB, and WaterGAP) in an attempt to falsify the above postulation. All models use the same forcings and geophysical information. The simulations are conducted in the Rhine River upstream of the Lobith gauging station. All three models are driven by daily forcing with a spatial resolution of 5 km, which was kindly provided by the EFAS team at JRC (`www.eea.europa.eu`). Additional details of the modeling settings of this experiment are provided in *Sutanudjaja et al.* (2015) and at `www.hyperhydro.org/`.

Table 5.2 Efficiency of mHM, PCR-GLOBWB and WaterGAP obtained for the Rhine basin at Lobith station during 2003 for spatial resolutions of 5 and 30 arcmin.

Model	5 arcmin		30 arcmin	
	KGE	Bias [m^3s^{-1}]	KGE	Bias [m^3s^{-1}]
mHM	0.96	61.19	0.96	21.74
PCR-GLOBWB	0.93	-20.61	0.86	248.09
WaterGAP (3,2)	0.83	143.02	0.90	-41.99

The KGE and bias values of these three models obtained for both scales at the Lobith station during 2003 are reported in Table 5.2. The daily streamflow time series during this year is selected for evaluation because it exhibits strong temporal dynamics, with wet conditions in the beginning of the year followed by a drought during the summer and fall seasons.

The performances obtained for the three models are satisfactorily, but the results shown in Table 5.2 indicate that mHM is the only model that can have higher KGE values regardless of the spatial modeling resolution.

The flux-matching test presented in Sect. 5.4.1 is performed with simulated evapotranspiration (ET) because it is the largest flux in the water cycle besides precipitation, and is prone to the largest predictive uncertainties (*Mueller et al.*, 2013). To ease the comparison, collocated grids are employed for every model such that every coarser scale grid cell has exactly the same number of underlying cells at finer resolution (5 arcmin). The results of this test are shown in Fig. 5.7a, b. They reveal that mHM exhibits the best flux-matching between these two scales.

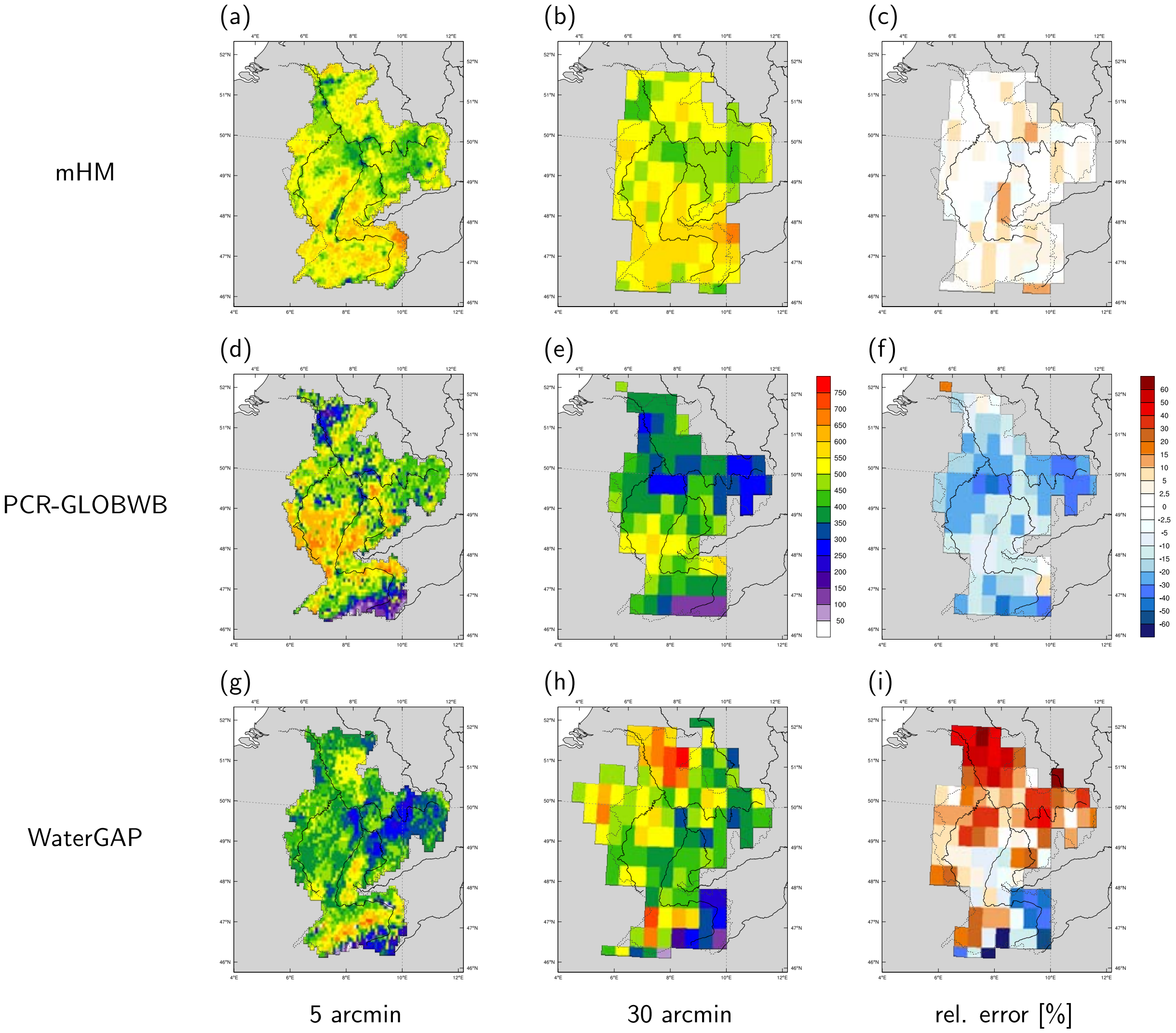

Figure 5.7 Multiscale simulation of annual ET for the Rhine River in 2003 with mHM, PCR-GLOBWB, and WaterGAP (versions 3 and 2) at spatial resolutions ℓ_1 of 5 and 30 arcmin, respectively. The relative errors in percentage of the coarse field estimates with respect to the finer ones (aggregated to the coarser level) for mHM, PCR-GLOBWB, and WaterGAP are shown in panels **(c)**, **(f)**, and **(i)**, respectively.

This experiment also shows that the MPR technique implemented in mHM leads to ET fields that are of similar magnitude at both scales, indicating a close conservation of mass leading to the lowest relative errors (Fig. 5.7c) among the three models.

The PCR-GLOBWB and WaterGAP models reveal large inconsistencies in preserving the spatial pattern of annual ET across two modeling scales, although the streamflow performance at the outlet is good (greater than 0.83 in both cases). PCR-GLOBWB at coarse resolution tends to underestimate ET (up to 50 %) compared with those at finer resolution (Fig. 5.7f). Conversely, the coarser version of WaterGAP tends to overestimate ET (up to 60 %) compared with those at the finer resolution (Fig. 5.7i). Interestingly, it can be observed that changes in model resolution affect the dynamic of water fluxes in those models that do not use any consistent scaling rules for model parameterization. These results also confirm the postulation that "streamflow-related metrics are a necessary but not sufficient condition to warrant the proper partitioning of incoming precipitation P into various spatially distributed water storage components (e.g., SM) and fluxes (e.g., ET)" (*Rakovec et al.*, 2016a). Because all models are forced with the same forcings, share the same geophysical information, and have almost similar hydrological process descriptions, it can be safely concluded that the parameterization method used in the models caused the ET mismatch. To falsify this postulation, the MPR parameterization protocol proposed in Sect. 5.4.3 is next applied to PCR-GLOBWB.

5.6 Implementation of the parameterization protocol in PCR-GLOBWB

To evaluate the consistency of land surface fluxes before and after MPR implementation, we analyze the impact of MPR on evaporative fluxes and soil moisture content in PCR-GLOBWB (*Sutanudjaja et al.*, 2016; *van Beek et al.*, 2011; *Wada and Bierkens*, 2014) over the Rhine River basin during 2003. The model is used to simulate the hydrological states at two different spatial resolutions ($\ell_1 = 5$ and 30 arcmin), and the sensitivity to MPR implementation is evaluated using a field difference method (in line with Eq. 5.1):

$$\Delta = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(100\frac{W(t)-w(t)}{w(t)}\right)^2}, \tag{5.2}$$

where W and w are the coarse and fine resolution simulations of variable W, respectively, and T is the total time series length.

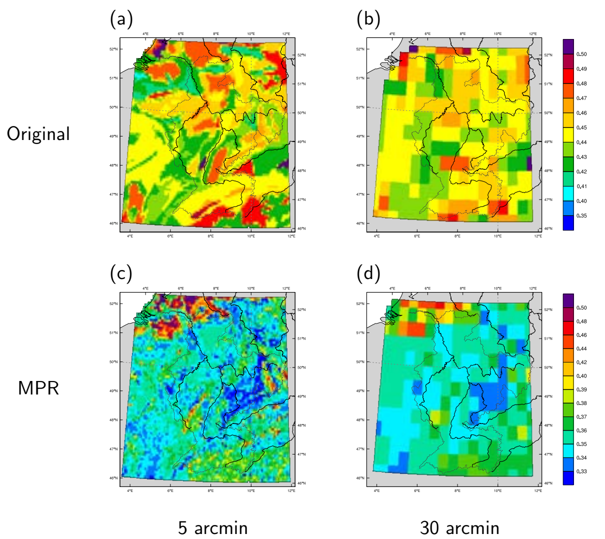

Figure 5.8 Porosity fields of PCR-GLOBWB before **(a, b)** and after implementing MPR **(c, d)** for two spatial resolutions of 5 and 30 arcmin. Dotted lines denote the Rhine Basin and the continuous line is the main EU river basin network.

The original PCR-GLOBWB parameterization does not include consistency in upscaling as enforced by MPR, leading to a larger difference in soil properties. Figure 5.8 depicts the porosity fields of this model before and after the implementation of MPR. Figure 5.8a and b clearly show the problems mentioned in Sect. 5.3, for example, lack of coherence in spatial patterns and the existence of spatial discontinuities of parameter fields at two scales. The porosity fields obtained with the MPR technique shown in Fig. 5.8c and d, on the contrary, exhibit a typical seamless spatial structure in which the main features of the field can be distinguished across scales. It is worth noting that differences seen between Fig. 5.8a and c are not only due to the improved upscaling procedure, but also due to a modified pedotransfer function. The parameters of the pedotransfer function have also been included in the calibration within the MPR approach. These differences in soil hydraulic properties influence the derived hydrological properties, leading to changes in saturated conductivity and storage capacity in the unsaturated zone. The considerable differences in ET fluxes are shown in Fig. 5.9a and b, and are the result of these changes. When MPR is employed, we observe that the difference in actual average Rhine Basin evapotranspiration between the two scales Δ drops from 29 to 9.4 % (Fig. 5.9d, e). For the total column soil moisture, we find a stronger decrease in Δ from 25 to 6.9 %, clearly indicating the benefits of MPR implementation. The error fields in Fig. 5.9c and f show a clear benefit of implementing MPR in PCR-GLOBWB. It should be noted, however, that the improvements are not as high as those obtained for mHM as shown in Fig. 5.7c. This is related to the fact that all effective parameters related to the evaporation and soil dynamic processes have been scaled with MPR in mHM, whereas in PCR-GLOBWB, only soil porosity has been scaled with this technique. Nevertheless, it is remarkable to see the improvements in flux matching (Fig. 5.9f) by scaling a single parameter of PCR-GLOBWB using MPR.

We also observe a slight increase in the discharge performance (KGE) at Lobith. The original KGEs are 0.86 ($\ell_1 = 5$ arcmin) and 0.93 ($\ell_1 = 30$ arcmin), whereas the KGEs with MPR implementation are 0.91 and 0.93, respectively. Another advantage is that PCR-GLOBWB is calibrated at a coarser resolution, whereas this model is calibrated for each spatial resolution individually in the original setup and with lower consistency in the discharge simulation.

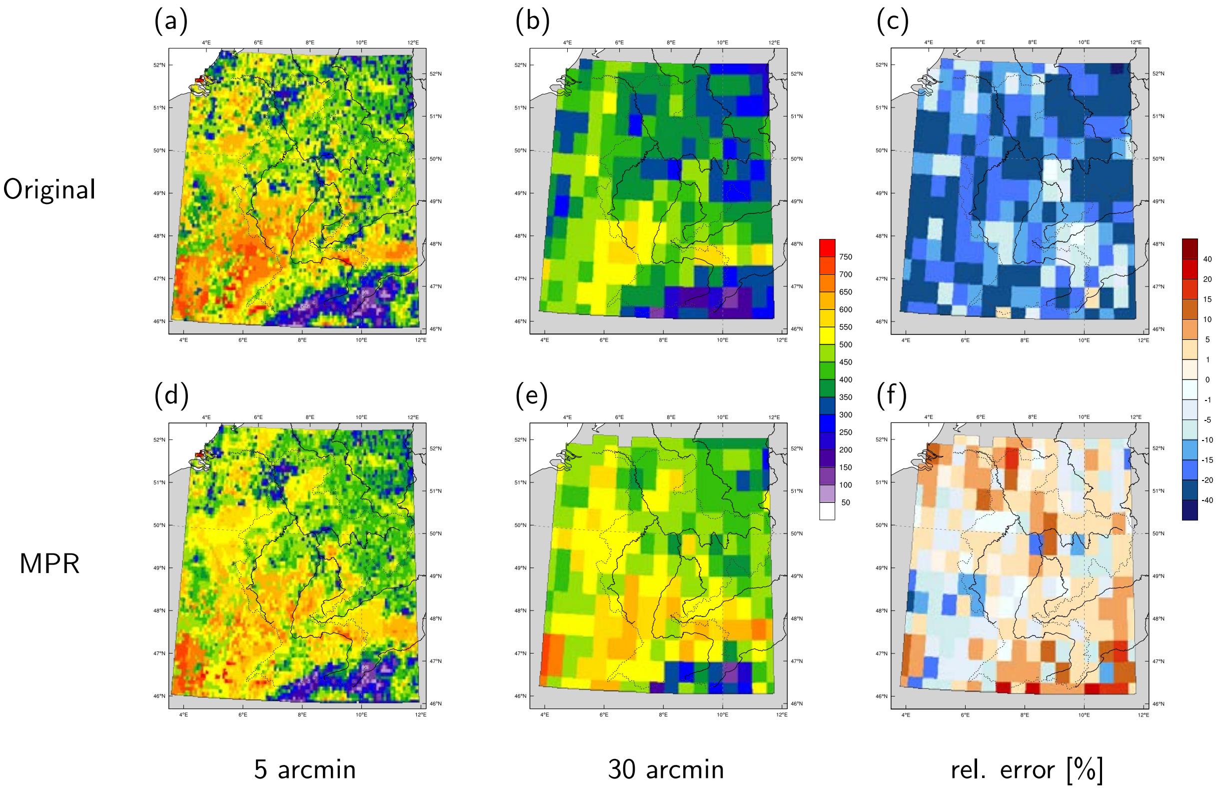

Figure 5.9 Annual ET fields in 2003 of PCR-GLOBWB before **(a, b)** and after implementing MPR **(c, d)** for two spatial resolutions of 5 and 30 arcmin. Dotted lines denote the Rhine Basin and the continuous line is the main EU river basin network. The relative errors in percentage of the coarse field estimates with respect to the finer ones are shown in panels **(c)** and **(f)**, respectively.

From these evaluations, we conclude that MPR implementation leads to significant improvement in the flux-matching and discharge simulations across scales, allowing for more consistency across scales for hydrological model simulations. Notably, additional parameters in PCR-GLOBWB still need to be regionalized within the MPR framework, which could potentially lead to better performance and transferability.

5.7 Conclusions

Hyper-resolution modeling initiatives (*Bierkens et al.*, 2014; *Wood et al.*, 2011) challenge the hydrological community to intensify efforts to make water (quantity and quality) and energy flux predictions "everywhere" and for these predictions to be "locally relevant." The predictions should have small uncertainties to be useful for the end users. These grand challenges also imply that the next generation of land surface and hydrologic models must incorporate probabilistic descriptions of the subgrid variability of geophysical land surface properties – such as POLARIS (*Chaney et al.*, 2016b) and SoilGrids (*Hengl et al.*, 2017) – to cope with the large uncertainties that characterize the related process below the representative elementary area (REA) scale. Consequently, great efforts should be made in hyper-resolution monitoring at the global scale in improving the computational efficiency of LSMs/HMs and in the development of scale-invariant parameterizations for these models. In this study, we have shown that the state-of-the-art parameterizations need to be improved to address this grand challenge, especially with respect to better fulfill the flux-matching condition.

We revisited a technique called MPR (*Samaniego et al.*, 2010a), originally available only in mHM but recently implemented in PCR-GLOBWB as a part of this study. Moreover, we proposed a "parameterization protocol" as a guideline to apply MPR and to retrofit existing LSMs/HMs to ease the implementation of MPR in the latter. We also discuss the advantages and limitations of MPR which should be considered while applying this concept to other LSMs/HMs.

This study has shown that two models that use ad hoc parameterizations can have reasonable efficiency with respect to simulated streamflow but poor performance with respect to distributed fluxes such as evapotranspiration. The implementation of this protocol in PCR-GLOBWB in this study increased the model efficiency by almost 6 % and improved the consistency of simulated ET fields across scales. For example, the estimation of evapotranspiration without MPR at 5 and 30 arcmin spatial resolutions for the Rhine River basin resulted in a difference of approximately 29 %. Applying MPR reduced this difference to 9 %. For total soil water, the differences without and with MPR are 25 and 7 %, respectively. We have also shown that the PCR-GLOBWB global parameters can be transferred across scales with consistent ET patterns and model efficiency.

In general, it can be concluded that the estimation of global parameters is feasible with MPR and that these scalars are transferable across scales and locations. The successful application of MPR implies that the averaging procedure of geophysical properties matters and that having the right physics with incorrect "effective" parameters leads to inconsistent fluxes and states. Consequently, MPR is a step forward to quasi-scale-invariant parameterizations and is feasible to implement in existing LSMs/HMs whose goal should be seamless parameter fields across scales that do not exhibit artificial spatial "discontinuities" such as calibration imprints, and that lead to consistent predictions across scales. We consider that this feature is the key for the next generation of LSM and NWP models such as the model for prediction across scales (MPAS) (`www.mmm.ucar.edu`) and the nested-domain ICON (`www.earthsystemcog.org/projects/dcmip-2012/icon-mpi-dwd`). Furthermore, a proper implementation of MPR in process-based (conceptual) models may contribute to recent efforts towards identifying their "effective" parameters through observational datasets at the scale of interest (*Savenije and Hrachowitz*, 2017).

Finally, we would like to reiterate that a flux obtained from a land surface/hydrologic model should always be evaluated with local observations when available and across scales. If "it disagrees with the experiment, it's wrong."

Acknowledgements

We kindly acknowledge our data providers: Noah-MP: Kirsten Warrach-Sagi (University of Hohenheim), PCR-GLOBWB: Niko Wanders (Princeton University), WaterGAP: Hannes Mueller-Schmied (University of Frankfurt), JULES: Anne Verhoef (The University of Reading), LISFLOOD: Peter Salamon (JRC), CABLE: Matthias Cuntz (formerly UFZ, now INRA), CLM: David Lawrence (UCAR) and E. Sutanudjaja and M. Bierkens, et al. for providing results from the HyperHydro WG1 Workshop 9-12 June 2015, Utrecht. This study was carried out within the Helmholtz-Association climate initiative REKLIM (`www.reklim.de`). This work has been partly funded by Helmholtz Alliance EDA - Remote Sensing and Earth System Dynamics, through the Initiative and Networking Fund of the Helmholtz Association, Germany. This study was partially performed under a contract for the Copernicus Climate Change Service (`edge.climate.copernicus.eu`). ECMWF implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. We thank Martin Schrön for kindly contributing to the artwork of Figure 5.2. Rens van Beek is acknowledged for providing support with the MPR implementation in PCR-GLOBWB. This article is a result of the Symposium in Honor of Eric F. Wood: Observations and Modeling across Scales, Princeton, New Jersey, USA, 2–3 June 2016.

CHAPTER 6

MULTISCALE AND MULTIVARIATE EVALUATION OF WATER FLUXES AND STATES OVER EUROPEAN RIVER BASINS

O. Rakovec[1], ..., L. Samaniego[1]

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Rakovec, O., R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego (2016c), Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, J. Hydrometeorol., 17(1), 287–307

6.1 Abstract

Accurately predicting regional scale water fluxes and states remains a challenging task in contemporary hydrology. Coping with this grand challenge requires among other things a model that makes reliable predictions across scales, locations, and variables other than those used for parameter estimation. In this study, the mesoscale hydrologic model (mHM) parameterized with the multiscale regionalization technique is comprehensively tested across 400 European river basins. The model fluxes and states, constrained using the observed streamflow, are evaluated against gridded evapotranspiration data, soil moisture and total water storage anomalies, as well as local-scale eddy covariance observations. This multiscale verification is carried out in seamless manner at the native resolutions of available dataset varying from 0.5 km to 100 km. Results of cross-validation tests show that mHM is able to capture the streamflow dynamics adequately well across a wide range of climate and physiographical characteristics. The model yields generally better results (with lower spread of model statistics) in basins with higher rain gauge density. Model performance for other fluxes and states is strongly driven by the degree of seasonality that each variable exhibits with the best match being observed for evapotranspiration, followed by total water storage anomaly, and the least for soil moisture. Results show that constraining the model against streamflow only may be necessary but not sufficient to warrant the model fidelity for other complementary variables. The study emphasizes the need to account for other complementary datasets besides streamflow during parameter estimation to improve model skill with respect to "hidden" variables.

6.2 Introduction

Since the pioneering work of *Crawford and Linsley* (1966), the efficiency of computational hydrologic models has been evaluated against streamflow observations that are available at determined locations within a river basin (e.g., *Bergström*, 1995; *Dawdy and Lichty*, 1968; *Duan et al.*, 1992; *Hundecha and Bárdossy*, 2004; *Kumar et al.*, 2013b;

Samaniego et al., 2010a; *Seibert*, 2000; *Sorooshian and Dracup*, 1980; *Troy et al.*, 2008; *Yilmaz et al.*, 2008). This kind of continuous *in situ* measurement is essential for understanding the governing relationships between rainfall and runoff in a particular drainage basin. The information content of this time series fundamentally differs from other point measurements such as soil moisture and latent heat in the sense that it represents the integral basin response to a sequence of hydrometeorologic events under particular physiographic and climatic conditions that uniquely characterizes a river basin. Because of this fundamental characteristic, streamflow gauging has been and will be part of the core of national hydro-meteorologic monitoring programs and the basis for sound water resources management. It is therefore not surprising that streamflow time series has been the focus for seminal hydrologic work in the past (*Horton*, 1935; *Hurst*, 1951; *Kuichling*, 1889; *Nash*, 1957; *Rodriguez-Iturbe and Valdes*, 1979; *SCS*, 1973; *Sherman*, 1932).

In the last years, however, a tendency towards a more comprehensive assessment of model structural adequacy has taken shape with an overall aim to improve the representation of different hydrological processes incorporated within a model (e.g., *Clark et al.*, 2011; *Gupta et al.*, 2012; *Shuttleworth*, 2012). The rational behind this assessment is the need to get the right answers for the right reasons (*Blöschl*, 2001; *Kirchner*, 2006) which goes beyond just assessing the model performance against observed streamflow or associated signature measures (e.g., *Euser et al.*, 2013; *Kumar et al.*, 2010; *Pokhrel et al.*, 2012; *Samaniego and Bárdossy*, 2007; *Yilmaz et al.*, 2008). Additional motivation for such assessment is driven by the growing need to simulate spatially distributed land surface fluxes controlled by local soil moisture availability and land surface hydrology. Consequently, complementary datasets representing internal hydrologic states and fluxes, such as soil moisture and evapotranspiration are required to achieve this goal. New kinds of observations and/or proxy data obtained from remote sensing and/or *in situ* measurement are being increasingly available, although at different spatial and temporal resolutions, e.g., monthly total water storage anomaly from GRACE at $1^\circ \times 1^\circ$, near surface soil moisture from ESA-CCI at $0.25^\circ \times 0.25^\circ$, and 30-minute eddy flux measurements of latent heat with a foot print of hundreds of hectares.

Several recent studies have evaluated the capability of hydrologic and/or land surface models to represent internal model fluxes and/or states (e.g., *Cai et al.*, 2014; *Li et al.*, 2012a; *Livneh and Lettenmaier*, 2012; *Sutanudjaja et al.*, 2014; *Xia et al.*, 2014, 2015). A common shortcoming in these studies has been the incompatibility of the scales at which simulated state variables and fluxes are compared with the observations (i.e., data are measured at different spatial scales from those at which models usually operate). The scaling issue poses a major obstacle in performing a comprehensive model evaluation (e.g., *Blöschl*, 2001; *Gentine et al.*, 2012; *Samaniego et al.*, 2010a; *Tetzlaff et al.*, 2010). Often, the *in situ* measurements of soil moisture or the evapotranspiration inferred at eddy covariance sites are compared with much coarser gridded model outputs (e.g., *Xia et al.*, 2014, 2015). *In situ* measurements often exhibit much finer support than the smallest representative elemental volume of hydrologic models (*Blöschl*, 1999; *Blöschl et al.*, 1995; *Wood*, 1995).

This scale discrepancy problem is exaggerated when a model is evaluated simultaneously against multiple datasets available at different spatial resolutions. In such a case, different upscaling/downscaling rules have to be employed to enable comparison between simulations and observations. Alternatively, a quasi-scale independent model parameterization scheme that allows to reliably represent processes at different spatial resolutions is required to tackle this scaling problem. The latter has the advantage that a processes-based modeling approach can be used to estimate hydrologic fluxes/states across multiple scales (*Gentine et al.*, 2012; *Kumar et al.*, 2013; *Samaniego et al.*, 2010a). Most of the existing modeling approaches, however, exhibit scale dependent performance, which means that the model parameterization obtained at a given spatial resolution induces large bias in hydrologic fluxes and states when applied to other resolutions (e.g., *Boone et al.*, 2004; *Haddeland et al.*, 2002; *Kumar et al.*, 2013; *Samaniego et al.*, 2010a; *Stöckli et al.*, 2007; *Troy et al.*, 2008).

Recently, *Samaniego et al.* (2010a) proposed a multiscale parameter regionalization (MPR) method that allows to make hydrologic predictions at different scales using a same set of model (transfer) parameters but without losing much of the model performance. The method explicitly accounts for the sub-grid variability of the essential aspects of the physical processes that are embedded within model parameters (e.g., soil porosity) and ensures that water fluxes simulated at different scales are comparable. The MPR method incorporated within the mesoscale hydrologic model (mHM; *Samaniego et al.*, 2010a) has been tested across a variety of climate and land surface conditions at different spatial resolutions ranging from 4 km to 100 km (*Kumar et al.*, 2010, 2013,b; *Samaniego et al.*, 2013). To date, these scaling studies have mainly focused on evaluating model performance against streamflow and conducting flux matching experiments using modeled variables at multiple scales and locations.

In this study we specifically evaluate the ability of the MPR method to reproduce the spatio-temporal dynamics of various water fluxes and states observed at multiple resolutions. The model parameterization constrained using streamflow observations across 400 European river basins is evaluated against complementary datasets that include gridded upscaled *in-situ* evapotranspiration (ET) data, satellite-based soil moisture (SM) and total water storage (TWS) anomalies, as well as local-scale eddy covariance data and their native resolutions. Alternative data fusion possibilities (such as data assimilation) to mitigate the limitations of models are beyond the scope of the present study and require future investigation. The multiscale evaluation approach followed here differs from previous hydrological model assessment studies that have covered the European domain using for example, the LISFLOOD model (e.g., *Wanders et al.*, 2014), the PCR-GLOBWB model (e.g., *Sutanudjaja et al.*, 2014; *Wada et al.*, 2010) or the WaterGAP model (e.g., *Werth and Güntner*, 2010). Although these studies focused on evaluating model skill on multi-variables, they have been operated on limited number of basins and/or with little consideration to the scaling discrepancy problem while verifying model outputs against observations.

We hypothesize that parameter estimation based only on streamflow related metrics is a necessary but not a sufficient condition to warrant the proper partitioning of incoming precipitation (P) into various spatially distributed water storage components (e.g., SM) and fluxes (e.g., ET). In the presented study, the multiscale and multivariate verification of water fluxes and states is carried out by executing mHM in a "seamless manner" (i.e., multiscale model simulation, in which each scale realization can be run simultaneously using a same set of model transfer parameters) at the native resolutions of available datasets varying from 0.5 km to 100 km.

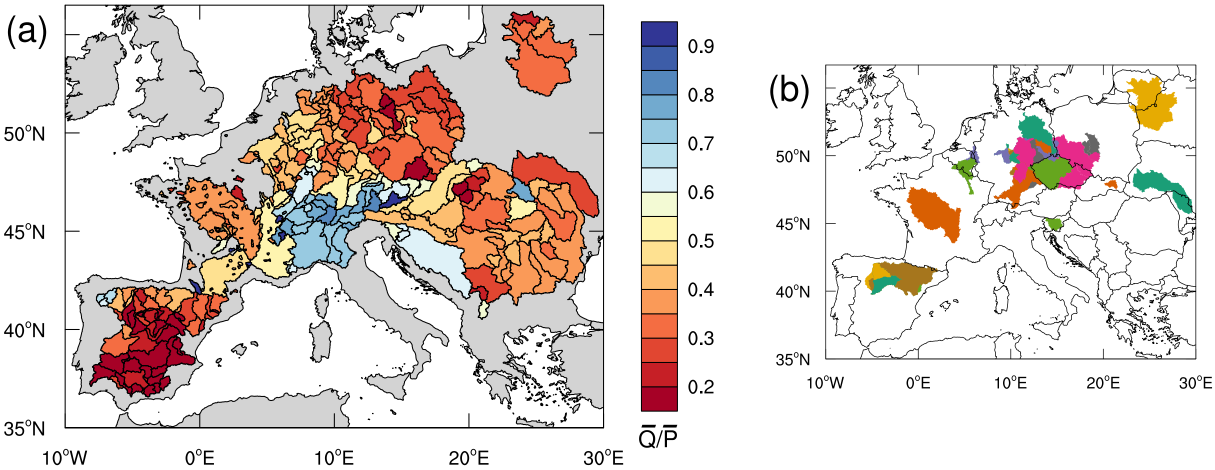

Figure 6.1 (a) Spatial map of the modelling domain showing the runoff ratio ($\bar{Q}/\bar{P}$) for 400 European basins used in this study. The smaller basins are overlaid on larger ones. (b) 36 donor basins provide an ensemble of plausible parameter sets (γ) constrained using the observed streamflow (different colors are used to distinguish between individual basins).

6.3 Data and Methods

6.3.1 Study area and datasets

The study is carried out in 400 European river basins (Fig. 6.1a) with drainage area varying from 10^2 km^2 to 10^6 km^2. These basins span over distinct climate conditions ranging from the dry-summer subtropical (Mediterranean, Southern Europe) to maritime temperate (Western Europe) and warm summer continental (Eastern Europe) climate types according to Köppen-Geiger classification (*Rubel and Kottek*, 2010). Figure 6.1a shows the span of runoff ratio ($\bar{Q}/\bar{P}$) which represents the long-term average partitioning of the precipitation ($\bar{P}$) into runoff ($\bar{Q}$) and actual ET ($\overline{ET}$). The runoff ratio is a comprehensive measure of physiographic basin and regional climate descriptors (e.g., *Berger and Entekhabi*, 2001; *Sankarasubramanian and Vogel*, 2002) that ranges between 0 and 1. Basins with smaller $\bar{Q}/\bar{P}$ values represents relatively drier conditions with higher evaporative rates (e.g., Southern Spain), while larger $\bar{Q}/\bar{P}$ represents humid or mountainous basins with lower evaporation rates (e.g., alpine regions).

The physiographical datasets used to setup the model mainly include digital elevation model, soil textural properties, and land cover states. An overview of these datasets is provided in Table 6.1. Since these datasets are available at different spatial resolutions, they are mapped on a common spatial resolution of 500 m × 500 m. This fine level datasets then allow to account for sub-grid variability of basin physical characteristics in parameter regionalization as described further in section 6.36.3.2.

Table 6.1 Description of input and evaluation datasets. ECAD, European Climate Assessment & Dataset. SRTM, Shuttle Radar Topography Mission. CGIAR-CSI, Consultative Group on International Agricultural Research: Consortium for Spatial Information. CORINE, coordination of information on the environment. EEA, European Environment Agency. ESD, European Soil Database. HWSD, The Harmonized World Soil Database. GRDC, Global Runoff Data Centre. GRACE, Gravity Recovery and Climate Experiment. FLUXNET, Flux Network. ESA-CCI, European Space Agency-Climate Change Initiative

	Variable	**Description**	**Reference**
Model setup	Meteorological forcing inputs (precipitation, air temperature, potential evapotranspiration)	Daily E-OBS product of 0.25° × 0.25° resolution	ECAD[a], http://www.ecad.eu (*Haylock et al.*, 2008)
	Terrain characteristics (e.g., elevation, slope, aspect, flow direction and flow accumulation)	SRTM[b] Digital Elevation Model data of 90 m × 90 m resolution	CGIAR-CSI[c], http://srtm.csi.cgiar.org (*Jarvis et al.*, 2008)
	Land cover (e.g., major class: forest, permeable, impervious cover)	CORINE[d] land cover dataset of 100 m × 100 m	EEA[e], http://www.eea.europa.eu
	Soil textural properties (sand and clay content, bulk density, horizon depth)	30 arc-second raster based on ESD[f]	HWSD[g] (*FAO/IIASA/ISRIC/ISSCAS/JRC*, 2012)
Model evaluation	Streamflow data	Daily observed streamflow	GRDC[h], http://www.bafg.de/GRDC; French basins (*Giuntoli et al.*, 2013)
	TWS anomaly	Gridded product of 1° × 1° resolution	GRACE[i] (*Landerer and Swenson*, 2012; *Swenson and Wahr*, 2006)
	Actual ET	Gridded product of 0.5° × 0.5° resolution	FLUXNET[j] (*Jung et al.*, 2011)
		In situ observations	Eddy covariance sites http://gaia.agraria.unitus.it
	SM	Gridded product of 0.25° × 0.25° resolution	ESA-CCI[k] http://www.esa-soilmoisture-cci.org (*Dorigo et al.*, 2014; *Liu et al.*, 2011)

The meteorological forcing data for the mHM consist of the daily gridded fields of precipitation, and average, maximum and minimum air temperatures at 0.25° × 0.25° resolution for the period 1950–2010. These datasets are acquired from the European Climate Assessment and & Dataset project (E-OBS, v8.0, *Haylock et al.*, 2008). These fields were created using the external drift Kriging interpolation technique based on ground-based observation networks. The potential evapotranspiration is derived using the temperature-based method of *Hargreaves and Samani* (1982) at the same spatial resolution (0.25° × 0.25°).

Streamflow records are commonly used to constrain the model parameterization and to evaluate its performance. Daily streamflow data between 1950 and 2010 were obtained from the Global Runoff Data Centre for this purpose. The data availability varies from station to station with the median record length of 43 years. All basins used in this study have undertaken a first-order data quality check, so that they do not violate the physical constraints imposed by the Budyko relationship (*Budyko*, 1974) and do not exhibit any obvious unnatural behavior in the discharge time series. More detailed analysis, particularly on the degree of regulation of European river basins, is deemed beyond the scope of the study because of the lack of support information. Besides streamflow, model performance is evaluated against complementary datasets, namely the total water storage anomaly, actual evapotranspiration and soil moisture. A brief overview of these datasets is given below.

TWS anomaly The TWS anomaly represents an important measure on seasonal and inter-annual variability of the terrestrial water storage, and is of critical interest for water resource management. The state of TWS affects infiltration rates, subsurface flows, groundwater recharge and runoff generation (e.g., *Li et al.*, 2012a). The remotely sensed anomalies of the Earth's gravity field retrieved by the Gravity Recovery and Climate Experiment (GRACE release 05, *Landerer and Swenson*, 2012) are used in this study to evaluate the simulated TWS of mHM. The global GRACE gridded dataset has 1° × 1° spatial and monthly temporal resolution. Although the GRACE product is available at coarse spatial and temporal resolutions, its application in hydrologic studies is increasing (e.g., *Andersen et al.*, 2005; *Cai et al.*, 2014; *Forman et al.*, 2012; *Li et al.*, 2012a; *Livneh and Lettenmaier*, 2012; *Orth and Seneviratne*, 2015; *Su et al.*, 2010; *Zaitchik et al.*, 2008). The TWS anomaly is analyzed using a combined product composed of different solutions obtained from three processing centers: GFZ (Geoforschungs Zentrum Potsdam, Germany), CSR (Center for Space Research at University of Texas, USE) and JPL (Jet Propulsion Laboratory, USA). The TWS anomaly is calculated via removing their corresponding long term mean estimates, which cover the baseline period from January 2004 to December 2009 (*NASA*, 2015). The arithmetic mean of these three products used here is the most effective way to reduce noise in the gravity field within the available scatter of the three solutions (*Sakumura et al.*, 2014). The evaluation period for TWS anomaly ranges between 2004 and 2012.

Actual ET Actual ET (latent heat flux) includes evaporation of water from soil, surface water bodies, canopy interception, and transpiration from plants leaves. It represents the second largest flux of the hydrologic cycle; on average 60% of terrestrial precipitation is returned back to the atmosphere via ET (e.g., *Oki and Kanae*, 2006). In this study, the modelled ET is evaluated against data at two distinct resolutions from (a) fine scale eddy covariance observations at 27 CarboEurope sites (*Göckede et al.*, 2008; *Mauder et al.*, 2008), and (b) the 0.5° gridded ET dataset derived from the FLUXNET observations (*Jung et al.*, 2011).

Basic information for the eddy covariance stations is provided in Table 6.2 (Appendix). The foot print of the observations covers approximately several hundred meters. Only stations with an almost complete record for the years 2004-2007 are chosen from the CarboEurope database. Data are processed after *Papale et al.* (2006) and unit imputation (gap-filling) is done by marginal distribution sampling (*Reichstein et al.*, 2005). Observed latent heat fluxes were corrected by authors for missing energy balance closures with a Bowen ratio approach similar to *Kessomkiat et al.* (2013).

The gridded FLUXNET ET product is acquired from the Department Biogeochemical Integration at the Max Planck Institute for Biogeochemistry, Jena, Germany. The FLUXNET ET product is obtained by upscaling observations of biosphere-atmosphere fluxes of carbon and energy from eddy covariance flux tower sites using model tree ensembles (MTE) (*Jung et al.*, 2011). The global monthly ET product is available at 0.5° × 0.5° for the period 1982–2011. We refer to *Jung et al.* (2011) for detailed description of the processing algorithm used to generate this dataset.

ESA-CCI surface SM SM acts as a switch and integrator of various energy and water fluxes between the land surface and the atmosphere and is the life-giving substance for vegetation. Estimating correctly the degree of soil saturation is the key point in hydrological modeling because it influences the partitioning of precipitation into ET and runoff. It also has a direct effect on society in terms of agriculture management as well as flood and drought predictions. Moreover, it integrates precipitation and evaporation over periods of days to weeks, thus introducing memory in the hydrological cycle.

The ESA Climate Change Initiative (CCI) provides a global SM product based on the retrievals from four passive (SMMR, SSM/I, TMI, and AMSR-E) and two active (ERS AMI and ASCAT) coarse resolution microwave sensors. The interested reader may refer to *Liu et al.* (2011) and *Dorigo et al.* (2014) for detailed description of this dataset. The ESA-CCI dataset represents near surface SM (0.5–2 cm) at 0.25° × 0.25° spatial resolution for the period 1978–2013. The recent study by *Dorigo et al.* (2014) shows that the skill of the merged product compared to the skill of the individual input products of the passive/active sensors with respect to *in situ* observation has a comparable and/or better performance than the individual input products in terms of the Spearman rank correlations. We emphasize that the ESA-CCI SM product is rescaled to the dynamic range of the GLDAS-Noah surface soil moisture fields, and therefore it could not be considered as an independent dataset representing absolute true soil moisture (*ESA*, 2015).

6.3.2 The mHM and the MPR

The mesoscale hydrologic model (mHM) used in this study is a grid based distributed model that is based on numerical approximations of dominant hydrologic processes applied in known HMs such as the Hydrologiska Byråns Vattenbalansavdelning (HBV; *Bergström*, 1995) and the Variable Infiltration Capacity (VIC; *Liang et al.*, 1994) models. Specifically the model accounts for the following processes: canopy interception, snow accumulation and melting, SM dynamics, infiltration and surface runoff, ET, subsurface storage and discharge generation, deep percolation and baseflow, and flood routing. The snow accumulation and melting processes are modeled using a modified degree-day method which accounts for the enhanced snowmelt during the intense precipitation events (*Hundecha and Bárdossy*, 2004). The incoming precipitation and snowmelt is partitioned into a root zone soil moisture and runoff components depending on the degree of soil saturation using a power function similar to the HBV model. The model uses three soil layers to describe the root zone soil moisture dynamics: the depth of the first soil layer is 5 cm, the second one is 25 cm, and the third layer up to 100 cm. The soil moisture processes in the first two soil layers account for the variation in soil organic matter over time with changes according to land cover type. Actual ET from soil layers is estimated as a fraction of potential evapotranspiration depending on the soil moisture stress and fraction of roots in each soil layer. The runoff generation process in mHM accounts for surface, fast- and slow-interflow and baseflow components. The interflow component represents the fast reaction to weather signals while the baseflow represents the slow and permanent groundwater flow. Finally, the total runoff produced at each grid cell is routed to the neighboring downstream grid cell via the Muskingum-Cunge flow routing algorithm. For a complete model description, interested readers may refer to *Samaniego et al.* (2010a); and the model code can be downloaded from `www.ufz.de/mhm`. To date, mHM has been applied over a large number of river basins across Germany and the U.S. (*Kumar et al.*, 2010, 2013,b; *Livneh et al.*, 2015; *Samaniego et al.*, 2010a, 2013).

The model uses three distinct levels of information to better account for spatial heterogeneities of input data, hydrological processes, and meteorological forcings. The lowest level (ℓ_0) describes information on input data related to physiographical and morphological characteristics of a basin. The intermediate level (ℓ_1) is used to model the governing hydrological processes, while the highest level (ℓ_2) contains information on meteorological data sets. Typically, the spatial resolution of ℓ_1 and ℓ_2 is the same, in order of kilometers depending on the availability of the forcing data set (24 km $\approx 0.25°$ in this study). The resolution of ℓ_0 input data is much finer than the other two, in the order of hectometers (500 m $\approx 0.004°$ in this study).

The Multiscale Parameter Regionalization (MPR) technique (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a) is used to efficiently incorporate the sub-grid ℓ_0 information within the modeling level ℓ_1 using a two-step parameterization technique (see Figure 6.2). In the first step, model parameters (β^0, e.g., porosity) are linked to available basin physical characteristics (e.g., terrain slope, sand and clay contents) using a set of pedo-transfer functions (f) and global parameters (γ). This linkage is established at the ℓ_0 spatial resolution to account for the sub-grid variability of input data and β^0 parameters. In the subsequent step, the ℓ_0 fields of model parameters (β^0) are aggregated to generate the effective regional parameter fields (β_1) at the modeling ℓ_1 level. The aggregation is performed using upscaling operators such as the harmonic, geometric, or arithmetic means, which satisfy flux matching conditions i.e., minimal discrepancy between aggregated water fluxes simulated across multiple resolutions. A set of global parameters (γ) is usually inferred via a suitable parameter estimation technique. This two step parameter regionalization technique allows the model to run efficiently in a seamless manner at multiple resolutions using the same set of global parameters γ (see Figure 6.2).

According to *Gupta et al.* (2014) the benefit of a regionalization method such as MPR stems from the fact that it "regularizes the optimization problem, providing constraints that greatly reduce the degrees of freedom (number of unknowns to be inferred) to a relatively small number of regional transfer function coefficients". The MPR technique, in addition of regularizing the optimization problem, takes into account the sub-grid variability of the essential aspects of the physical process that represent a given model parameter (e.g., soil porosity, wilting point, or hydraulic conductivity). Previous studies have demonstrated the effectiveness of the MPR approach over other existing parameterization techniques based on hydrological response units, lumped parameterizations and standard regionalization that do not account for the sub-grid variability of model parameters (*Kumar et al.*, 2010, 2013; *Samaniego et al.*, 2010a, 2011).

6.3.3 Experimental design and model setup

The goal here is to comprehensively evaluate the skill of mHM to represent the spatio-temporal variability of modeled fluxes and states at multiple scales. The experimental design of model parameter estimation and seamless verification is schematically shown in Fig. 6.2. In the initial phase, the model parameters are constrained against observed streamflow to obtain an ensemble of plausible model parameters (γ) with following procedure.

1. For each basin:

   (a) Prior to the parameter estimation of the model parameters, a subset of "informative" parameters (γ^*) is identified using the sequential screening method developed by *Cuntz et al.* (2015). This screening method is an adaption of the Morris method (*Morris*, 1991). In the first iteration, the model is evaluated at several points along trajectories of the parameter space and Elementary Effects are determined. Parameters with an Elementary Effect above a certain threshold are considered to be "informative". The next iteration of the method only takes into account "non-informative" parameters to test whether they are sensitive at other regions of the parameter space. This iterative procedure is repeated until no additional parameters are marked to be informative.

   (b) Find an optimal set of $\hat{\gamma}$ identified in step 1(a) by maximizing the Kling-Gupta efficiency (KGE; *Gupta et al.*, 2009). The Shuffled Complex Evolution (SCE) algorithm (*Duan et al.*, 1993) is used to maximize:

$$\max_{\hat{\gamma}} \mathrm{KGE}(r, \alpha, \beta), \tag{6.1}$$

with

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \tag{6.2}$$

where r is the Pearson correlation coefficient between observed (Q_{obs}) and simulated $Q(\hat{\gamma})$ streamflow; α denotes the measure of relative variability in the simulated and observed values (ratio of the standard deviations); and β is the ratio between the mean simulations and mean observations, i.e., bias. The parameter estimation period varies from basin to basin and ranges between 4 and 16 years of data depending on the availability of the observed streamflow. Prior to the model parameter estimation, a default run in the period from 1951 to 2010 is conducted to ensure appropriate initializations of internal model states and fluxes. Additionally, 5 years of data prior to the parameter estimation period are used to spin-up the model. The model is executed at daily time step and spatial resolution of $0.25° \times 0.25°$. The optimal set of $\hat{\gamma}$ inferred by this procedure corresponds to the complete basin and not to individual grid cells. This step is repeated ten times with random initialization of the parameter space to partially account for the uncertainty of the SCE algorithm and best performing parameter set is selected.

(c) Transfer $\hat{\gamma}$ to all remaining basins for streamflow cross-validation. Estimate the KGE($\hat{\gamma}$) as a measure of transfer efficiency.

2. Select best parameter sets ($\hat{\gamma}^{\mathrm{best}}$) from the pool of $\hat{\gamma}$ based on a cross-validation median KGE value larger than a threshold κ.

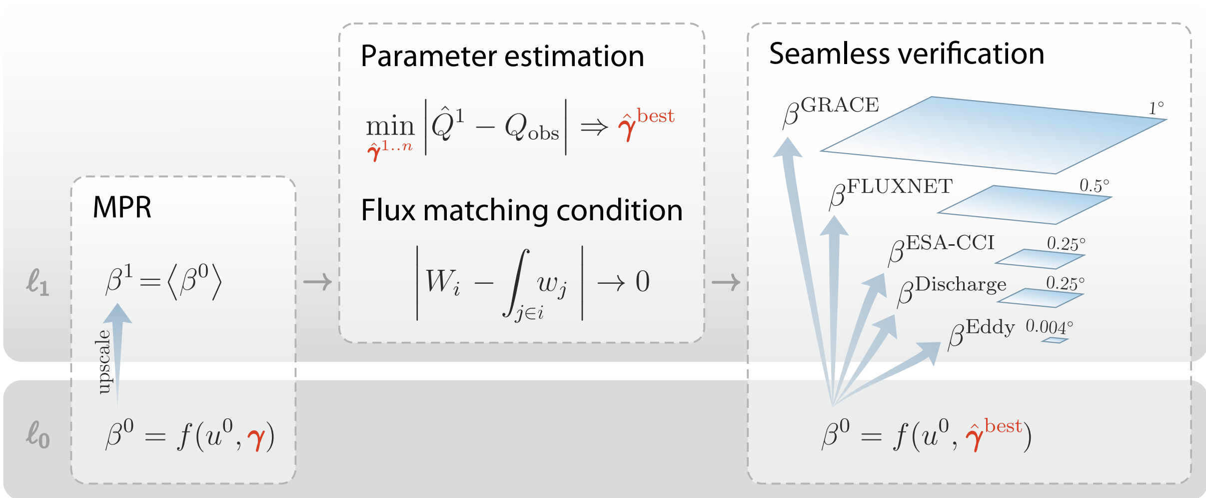

Figure 6.2 Schematic presentation of the experimental setup of the multiscale and multivariable model evaluation. Left panel depicts the two-step MPR scheme to incorporate the sub-grid variability of basin physical characteristics (u^0) available at the ℓ_0 level. Effective parameters β^1 at the modelling level ℓ_1 are estimated using a set of transfer functions f, global (parameter estimation) parameters γ, and upscaling operators $\langle . \rangle$. Middle panel shows the subsequent procedure to estimate a best set of global parameters ($\hat{\gamma}^{\mathrm{best}}$) calibrated against observed discharge (Q); and upscaling operators using the water fluxes matching conditions across multiple ℓ_1 resolutions (W_i and w_j denote the fluxes estimated at the coarser and finer cells i and j, respectively; for more details see *Samaniego et al.*, 2010a). Right panel illustrates the seamless (multiscale) verification approach for water fluxes and states at multiple modelling scales $\ell_1 = 0.004°, \ldots, 1°$ using the same sub-grid level information (β^0) and $\hat{\gamma}^{\mathrm{best}}$.

Following this procedure, an ensemble of best 36 parameter sets ($\hat{\gamma}^{\mathrm{best}}$) satisfying the threshold κ of 0.55 is selected to represent the "cross-validation uncertainty" of model output. The κ criterion of 0.55 is not directly related to performance in an individual basin, but rather it represents the median KGE value in a cross-validation over 400 basins. The location of the respective 36 basins spans over the entire study domain (Fig. 6.1b), which indicates the representativeness of the donor sample.

In the second phase (Fig. 6.2, right panel), the ensemble of 36 parameter sets is used to conduct model simulations at the native scale of the complementary datasets using the following procedure.

For each parameter set $\hat{\gamma}^{\text{best}}$:

1. Estimate effective model parameters $\boldsymbol{\beta}^0$ at level $\boldsymbol{\ell}_0$ using transfer functions f applied to the basin's physical characteristics (u^0) (Fig. 6.2, left panel).

2. Aggregate $\boldsymbol{\beta}^0$ to $\boldsymbol{\beta}^1$ at multiple modeling levels $\boldsymbol{\ell}_1$ using upscaling operators $\langle . \rangle$. The spatial resolution of $\boldsymbol{\ell}_1$ varies from 0.004° to 1° depending on the variable of interest (Fig. 6.2, right panel).

3. Run mHM using the $\boldsymbol{\beta}^1$ parameter fields at multiple scales and evaluate its performance for selected water fluxes and states.

Following the aforementioned algorithm, the model is executed 36 times at the spatial resolutions of $1° \times 1°$, $0.5° \times 0.5°$, $0.25° \times 0.25°$ over the entire domain, and at $0.004° \times 0.004°$ across 27 eddy grid points (see Fig. 6.2).

The streamflow evaluation is conducted on the entire record of available streamflow observations within the simulation period 1951–2010. The model skill for complementary gridded data sets is evaluated during the period of 2004–2010 for the ESA-CCI SM product, 1982–2011 for the gridded LANDFLUX ET data, 2004–2012 for the GRACE TWS anomaly. The ET evaluation at eddy flux stations is limited to the period 2004–2007 due to data availability issues. Three CORINE land cover scenes corresponding to the years 1990, 2000 and 2006 are taken into consideration. Model simulations prior to 1990 use the CORINE 1990 land cover map.

The model performance is evaluated using multiple statistical criteria that includes r, α, and β (Eq. 7.12), which quantifies mismatch between model and observations with respect to the temporal dynamics, variability and biases, respectively. Additionally, the model evaluation results are presented using a Taylor diagram (*Taylor*, 2001). This two-dimensional diagram quantifies concisely how well model simulations match observations in terms of r, α, and the root-mean-square difference.

Finally, a two dimensional histogram of the marginal distributions of the observed and simulated values (also called empirical copula) is used to describe the statistical dependency between the marginal distributions of two random variables (*Nelsen*, 2006). The copula can be generally written as

$$P(x \leq X, y \leq Y) = C(F(x), G(y)), \tag{6.3}$$

where x is the observed quantity with distribution function F, y is the simulated value with distribution function G, the left hand side of the equation is the joint probability P of x and y, and C is the copula between F and G. The copula-based model evaluation allows to quantify the stochastic dependence of simulated variables with respect to observations along the entire range of the variable.

6.4 Results and discussion

6.4.1 Model evaluation using observed streamflow

The mHM performance for the median KGE values between observed and simulated Q is depicted in Fig. 6.3a. The results indicate that mHM is capable of simulating daily discharge well over the Pan-EU domain considering that around 70% of the total area exhibits a median KGE value exceeding 0.5. It should be noted that the KGE values at a given basin are not obtained by means of on-site parameter estimation, but rather by transferring global parameters ($\hat{\gamma}$) from other basins. This guarantees consistent representation of hydrological processes at different locations, because the goal here is to obtain parameter sets that represent hydrological fluxes and states across the entire domain applicable for making predictions in ungauged basins. On-site parameter estimation yield even better performances, but increase the dimensionality of the parameter space for the entire domain since each basin in this case requires a set of global parameters. Additionally, Fig. 6.3b shows the ensemble of the cumulative frequency distributions of model performance based on 36 sets of global parameters ($\hat{\gamma}^{\text{best}}$). This figure illustrates a considerably narrow variation of model performance due to equally good performing parameter sets with KGE values higher than 0.55 for 50% of the basins.

The model performance in terms of KGE tends to be homogeneously distributed over space (Fig. 6.3a). Hot spots of poorer model performance occur in notably human influenced river basins such as those in Southern Spain, where the pressure on water resources is high and observed Q is far from natural conditions due to irrigation diversions, hydroelectric power generation, and flood control, (e.g., *Batalla et al.*, 2004; *Lorenzo-Lacruz et al.*, 2012). The mHM does not include human influenced processes, as the majority of other rainfall-runoff models, thus below normal performance is expected in those areas. We note that the basins which violate the physical constraints of the Budyko curve are removed prior to the analysis, as discussed earlier, however, this first-order quality check may not be sufficient to filter basins with significant anthropogenic activities. The mHM performance to simulate naturalized streamflow dynamics in other heavily human influenced U.S. basins is adequate and comparable to other existing models (*Kumar et al.*, 2013b; *Livneh et al.*, 2015). However, the lack of naturalized streamflow dataset in the present study domain limits such type of model evaluation. Another hot spot can be found in Eastern Europe (Romanian Carpathian Mountains) where the model systematically underestimates snow melt driven floods (in spring). The same behavior is observed by on-site parameter estimation (not shown). Additionally to the model conceptual error, the poor performance in these areas can be related to observation errors, such as the precipitation undercatch discussed in the following section.

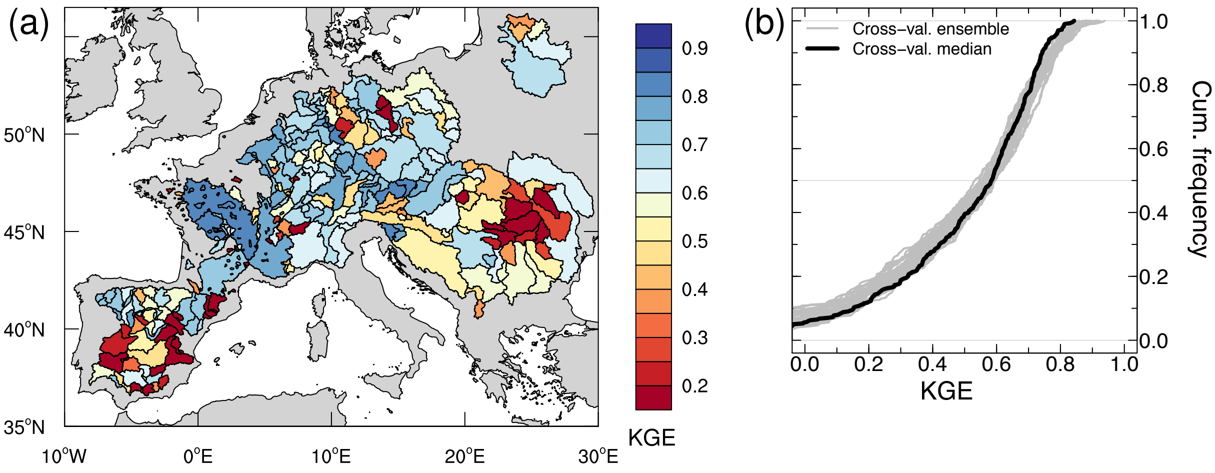

Figure 6.3 (a) Spatial maps of the modelling domain showing median KGE values between observed and simulated discharge for 400 European basins based on the cross-validation analysis (spatial model resolution of $0.25^\circ \times 0.25^\circ$, daily time step). (b) Cumulative frequency of the KGE values for the cross-validation uncertainty based on 36 parameter sets (grey), and the median KGE value (black) shown in (a).

6.4.2 Factors influencing Q predictability

The model performance of KGE and its three components (see r, α, β in Eq. 7.12) is further evaluated in Fig. 6.4 for basic basin characteristics such as area, rain gauge density and runoff ratio ($\bar{Q}/\bar{P}$). In general, the spread in uncertainty decreases with increasing basin area and also the model performance tends to be improved: KGE and r increase, while α and β converge towards their ideal value of one. This type of model performance dependency indicates that smaller basins are more susceptible to errors in model inputs than the larger ones, which also stems from averaging and the central limit theorem. The closer to the representative elementary area (REA), the more difficult it is to model due to the increased effect of small processes considered neither in the model nor in the data. Such kind of model dependency is also reported in previous studies (e.g., *Kumar et al.*, 2013; *Merz et al.*, 2009; *Reed et al.*, 2004).

Additionally, Fig. 6.4 depicts the relation between model performance and rain gauge density (number of rain gauges per 1000 km^2) to investigate the effect of forcing uncertainty. The model exhibits systematically better performance in regions with relatively higher rain gauge density, particularly in terms of variability. This promotes the importance of having a dense observation network for meaningful hydrological simulations, which is in particular important for capturing small-scale features such as convective cells (*Alfieri et al.*, 2014). The median rain gauge density is 0.4 gauges per 1000 km^2 and does not meet the standard of the World Meteorological Organization (WMO) in which the tolerable rain gauge density in flat regions is around 1–2 gauges per 1000 km^2, while it increases to 4–10 gauges per 1000 km^2 for mountainous regions (e.g., *Dingman*, 2004). Therefore, a precipitation product based on sparsely distributed rain gauge data can lead to higher modelling errors arising from imperfect precipitation estimates. However, we would like to emphasize that the E-OBS dataset used here is the best possible freely available dataset that exists at the moment with a relatively long temporal coverage, large spatial extent and fine spatial resolution (*Hofstra et al.*, 2009).

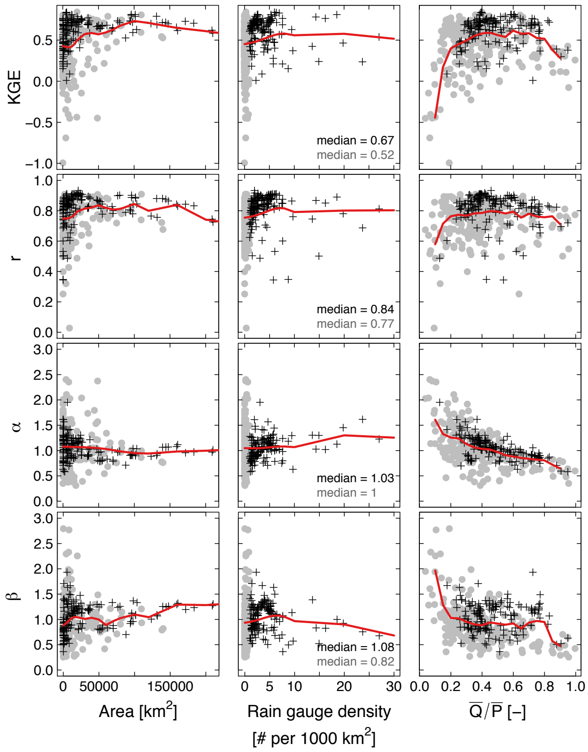

Figure 6.4 Basin area, rain gauge density, and runoff ratio as factors influencing model predictability of discharge in terms of KGE (see also Fig. 6.3a), and decompositions into the three components (r, α, and β in Eq. 7.12). Black pluses show basins which have at least 1 rain gauge per 1000 km^2, grey filled circles otherwise, and median values of the y-axis are provided in corresponding colors. Kernel regression is used to produce a smooth red line for the whole sample (black and grey).

Moreover, Fig. 6.4 illustrates that the model performance with respect to $\bar{Q}/\bar{P}$ is usually superior in intermediate physiographic and climatic regimes ($0.3 < \bar{Q}/\bar{P} < 0.7$), whereas the performance deteriorates towards both extremes. Generally, the model tends to overestimate the observed mean and variability in relatively moisture-limited (dry) basins. Note that these basins contain areas with human influenced activities where the model performs poorly (as discussed before). On the other hand, the extreme energy-limited basins exhibit a large bias and a systematically underestimated variance. These shortcomings can be related to several factors: precipitation underestimation due to lower rain gauge density, insufficient evaporation rates and/or model deficiency in capturing sub-grid snow processes. In general, the spread of model statistics (r, α, β) is considerably lower and model yields better results in basins with higher rain gauge density, which is observed regardless of the selected basin characteristics (compare grey filled circles with black crosses in Fig. 6.4 and their median values).

6.4.3 Spatial evaluation using complementary data

The model is further evaluated against the following complementary data (not being used to constrain the model) at monthly temporal resolution and native spatial resolution, namely: fields of the total water storage (TWS) anomaly from GRACE ($1° \times 1°$), FLUXNET ET ($0.5° \times 0.5°$), and SM from ESA-CCI ($0.25° \times 0.25°$). Figure 6.5 shows the model performance in terms of median r of the original data (top), median r of standardized anomalies (middle), and corresponding copulas of the latter one (bottom).

Overall, the model represents the TWS anomaly and the actual ET adequately well, while the performance for SM is not satisfactory in terms of correlation for the original time series (Fig. 6.5a–c). Presented hydrological variables exhibit strong seasonality and the performance criteria based on, for example, correlation coefficient for such variables is not adequate to show the actual model skill. Therefore, standardized anomalies of both observations and simulations are estimated by removing their respective monthly means and standard deviations. The correlation for the standardized anomalies shows in general deterioration for TWS and ET, however, slight improvement in SM when compared to the data with retained seasonality (Fig. 6.5d–f). The mHM results are consistent in findings of the recent study by *Orth and Seneviratne* (2015).

A reasonably good agreement is achieved for TWS anomaly in large part of the study domain. A relatively larger error can be observed in the Alps and coastal areas. This can be attributed to the fact that mHM lacks the capability to explicitly represent the glacial and tidal processes in these areas. Also, the GRACE data are not suitable to accurately quantify ice mass changes in glaciers (e.g., *Jacob et al.*, 2012) and there are relatively higher measurement and leakage errors provided in GRACE dataset along the coastal line. The leakage error of GRACE stands for "the residual errors after filtering and rescaling" from the raw original product to estimate the TWS anomaly (*NASA*, 2015).

The temporal dynamics of modeled ET resembles quite well the FLUXNET derived ET product with the majority (75%) of cells exceeding correlation coefficients (r) larger than 0.9 for the original time series, and larger than 0.59 for the standardized anomalies. Poor model performance is noticed on the Iberian peninsula. Notably, in these areas the model performance for observed Q is also poor (Fig. 6.3).

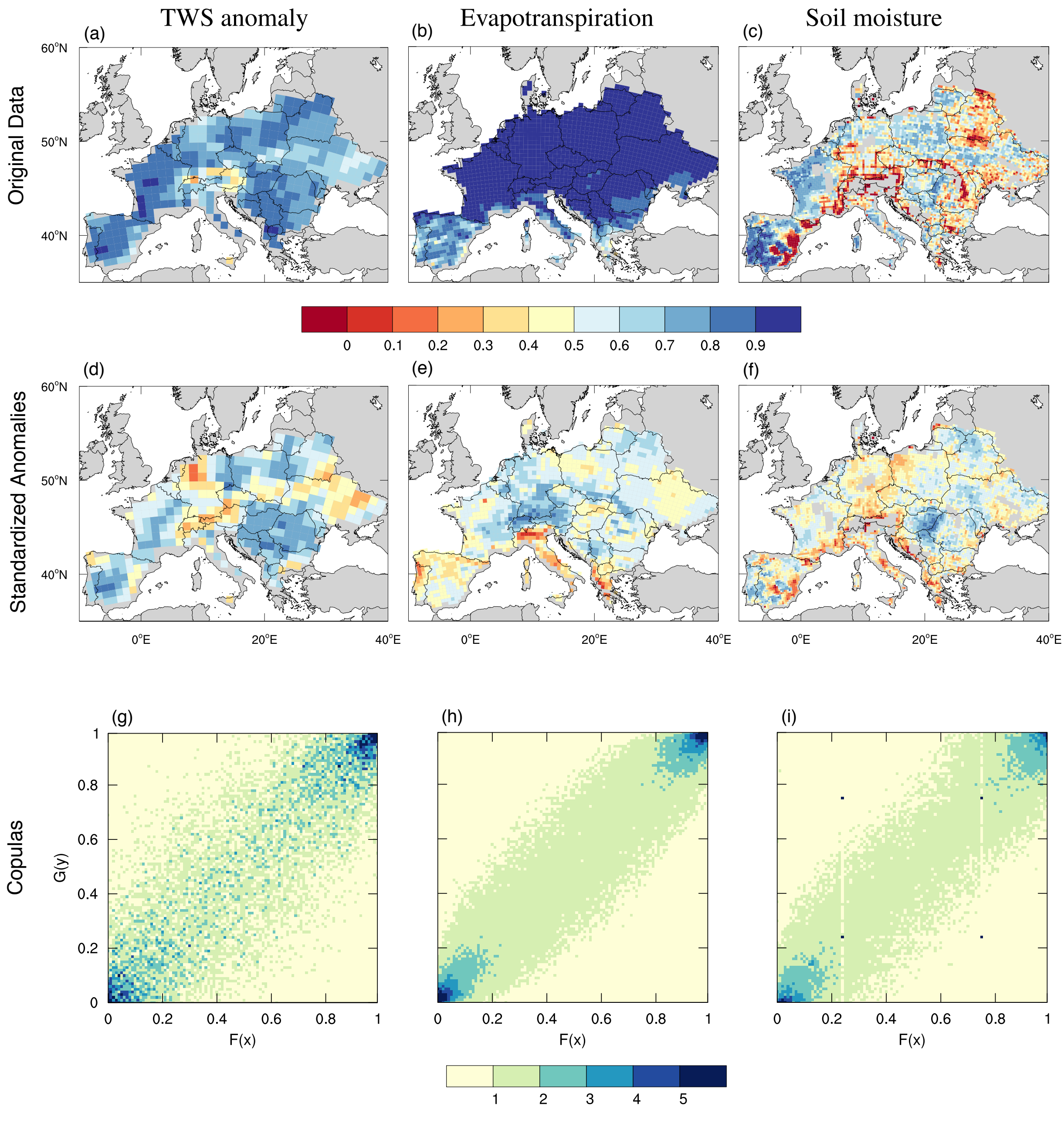

Figure 6.5 Model performance of mHM simulations and total water storage anomaly (GRACE, left column), actual evapotranspiration (gridded FLUXNET, middle column), and soil moisture (ESA-CCI, right column) observations in terms of medians of the Pearson correlation coefficient r of the original time series (a, b, c), r of the standardized anomalies (d, e, f), and empirical copula densities of the standardized anomalies (g, h, i). F(x) is a distribution function of an observed variable and G(y) is a distribution function of a simulated variable (see Eq. 6.3). The Spearman rank correlations are 0.61 for TWS anomaly, 0.55 for evapotranspiration, and 0.49 for soil moisture.

In comparison with the two aforementioned variables, r for soil moisture exhibits the poorest performance, however, overall correlation is quite comparable to those obtained in other recent studies (e.g., *Dorigo et al.*, 2014; *Lievens et al.*, 2015). Stripes in the model performance may correspond to an artifact of the retrieval algorithm, which is a typical characteristic in the observation through satellite microwave instruments (*ESA*, 2015). Notably, the comparison between satellite derived soil moisture and distributed models is challenging because of data and modeling reasons. Firstly, there is limited information on the exact depth of the soil layer that is used for the ESA-CCI satellite product besides other potential retrieval problems such as vegetation coverage, snow and ice content. Secondly, the biases across the statistical moments are very typical for surface soil moisture data derived from satellite retrievals, ground measurements, and models so that they need to be quantified and corrected (*Reichle and Koster*, 2004). Thirdly, the top thin first soil layer of most distributed models is not a good representation of actual soil moisture. Water evaporates as vapour from the soil surface. Soil water models that calculate only liquid water flow have to compensate for the missing process of vapour transport. Most of the models therefore include a very thin upper layer that counteracts the liquid water flow. Soil moisture products such as from cosmic ray sensors (*Zreda et al.*, 2012) represent larger soil volumes (*Köhli et al.*, 2015) and could become more feasible for model-data comparison in future studies.

The overall model evaluation using empirical copula densities of the standardized anomalies shows a strong statistical dependency between observations and simulations in particular for high and low quantiles (Fig. 6.5g–i). This has strong implications for drought and flood monitoring using satellite products alone. The statistical dependency for values in-between the extremes is close to the diagonal but with considerable spread for the three variables analyzed. This is related to grid cells exhibiting low correlation coefficients between their simulated and observed anomalies (panels d–f). Among the three variables, the relationship between the observed and simulated TWS anomaly exhibits larger scatter because of the reduced sample size due to a coarser spatial resolution and limited temporal availability of datasets (panel g). Overall, the copula densities indicate that the matching between observations and simulations needs to be improved for normal conditions as compared to the extremes. The Spearman rank correlations estimated based on these copulas varies from 0.49 to 0.61.

6.4.4 Basin scale evaluation of modelled fluxes and states

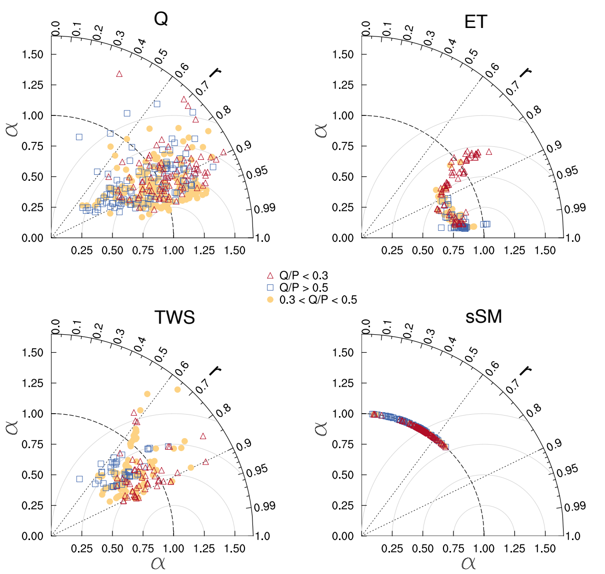

Figure 6.6 Evaluating model performance in terms of r and α for discharge (Q), actual evapotranspiration (ET), total water storage anomaly (TWS) and standardized soil moisture (sSM) observations at monthly time step using Taylor diagrams. Data are normalized by the standard deviation of the observations and classified according to the runoff ratio coefficient ($\bar{Q}/\bar{P}$).

The quantitative evaluation of model performance at basin scale is presented with Taylor diagrams (Fig. 6.6) for monthly estimates of Q, TWS anomaly, ET, and SM. The basins are further classified into three categories based on runoff coefficients ($\bar{Q}/\bar{P}$) representing wet/mountainous, intermediate and dry climatic regimes. In general, the model is able to represent the temporal dynamics of observed Q adequately well with correlations varying between 0.75 and 0.95 in the majority of the analyzed basins. The observed variability is also well captured by the model regardless of the $\bar{Q}/\bar{P}$ characteristics with a median α value of around 1. On the other hand, the variability in ET is systematically underestimated ($\alpha < 1$), while temporal dynamics is well represented with r exceeding 0.8 in the majority of basins. Furthermore, the performance of the model is relatively low for the total water storage anomaly and SM. This is observed independently of the runoff ratio. The correlation for the TWS anomaly ranges mostly between 0.6 and 0.9 with higher values being noticed for basins lying in the water limited regime, which is also seen in Fig. 6.5. The poorest performance among all analyzed variables is observed for simulating the soil moisture dynamics with r less than 0.6 in the majority of basins.

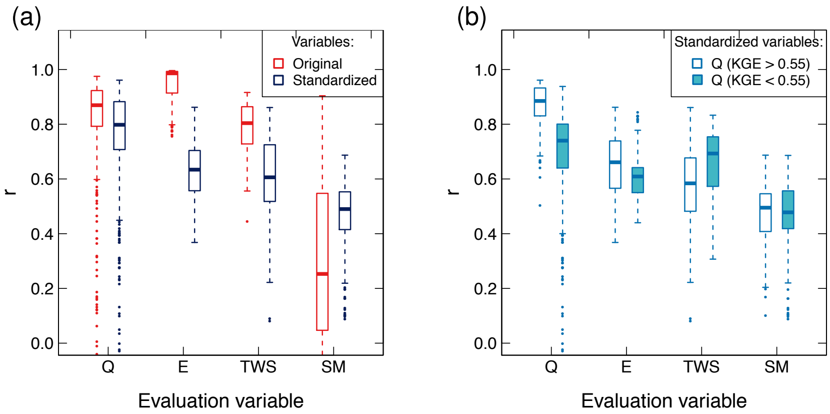

Figure 6.7 Evaluating model performance for discharge (Q), actual evapotranspiration (ET), total water storage anomaly (TWS) and soil moisture (SM) at monthly time step using the correlation coefficient. (a) Correlation is derived for the original time series (red; identical to values shown in Fig. 6.6), and the time series with removed annual cycle normalized by long term standard deviation from the monthly data (blue). (b) Results of the standardized anomalies are differentiated into two equal-size groups based on the KGE values yielding better performing basins (above median KGE) and worse performing basins (below median KGE).

The model performance in terms of correlation for the original time series (shown in Fig. 6.6) is contrasted against their corresponding standardized values for different variables in Fig. 6.7a. All variables exhibit lower performance for their standardized estimates with the exception of soil moisture. The largest deterioration is noticed for the ET followed by the TWS anomaly and the least for Q. The sequence of this deterioration corresponds to the degree of seasonality among the analyzed variables. The best model performance for Q can be partly attributed to the fact that the model is constrained against this variable. Despite the lower model performance for standardized variables, the majority of basins has r values above 0.4 which is well beyond the threshold limit of 0.2 to be statistically significant at the 95% confidence interval.

Based on the results shown in Fig. 6.7a, the test of differences in mean skill scores between the standardized distribution of Q and other modeled variables have p-values lower than 10^{-5}. This indicates that the null hypothesis that Q alone can sufficiently constrain model components responsible for internal fluxes and states in a cross-validation mode can be safely rejected. To further support the aforementioned hypothesis, Figure 6.7b differentiates the results of the standardized anomalies into two equally-sized groups based on the median KGE cut-off value of 0.55 (as discussed in section 6.46.4.1). On average a significant deterioration in model skill score (p-value < 0.01) is observed for other complementary variables in comparison to discharge with the exception of TWS anomaly for which no conclusive deterioration in model skill can be noticed for the worse performing basins. The deterioration is more pronounced for the group of basins yielding on average better model performance in terms of discharge, which to some degree reflect the over-fitting of model parameters during the parameter estimation against the observed discharge. This also indicates that other complementary data are required to appropriately constrain the model in well performing basins. Such datasets are also of great usage in data scarce regions, where streamflow observations are not available to constrain the model.

Finally, Figure 6.8 shows the monthly dynamics of observed and simulated fluxes and states for three randomly selected basins in dry, intermediate and wet climatic conditions based on the runoff coefficients. The magnitude and timing of Q are well matched and observations are mostly covered within the uncertainty bounds. Contrary to Q, ET and TWS anomaly exhibit very regular inter-annual variability. The model is able to follow this behavior quite well, although it tends to underestimate the gridded FLUXNET ET in the intermediate and wet basins. The largest discrepancy between the model and observation occurs for SM consistent with our previously discussed results.

6.4.5 Model errors in relation to water balance closure

The aforementioned results of basin scale model performance for different variables illustrate the existence of potential errors between observations and simulations (Figs. 6.5–6.8). They can be attributed to a number of factors, which can be mainly related to model and input data errors. The former constitutes error due to the improper model structure and/or parameterizations, whereas the latter represents errors due to imperfect forcings and/or response variables. An analysis is carried out to understand the relationship between the errors in input data to errors in individual variables. The residuals in water balance closure ($\bar{P} - \bar{Q} - \overline{ET}$) based on 20 water years (1989–2008) of observed P, Q, and the gridded FLUXNET ET, are taken as a proxy for the input data error. This error is analyzed across basins with different climatic characteristics and contrasted against errors in simulated Q and ET.

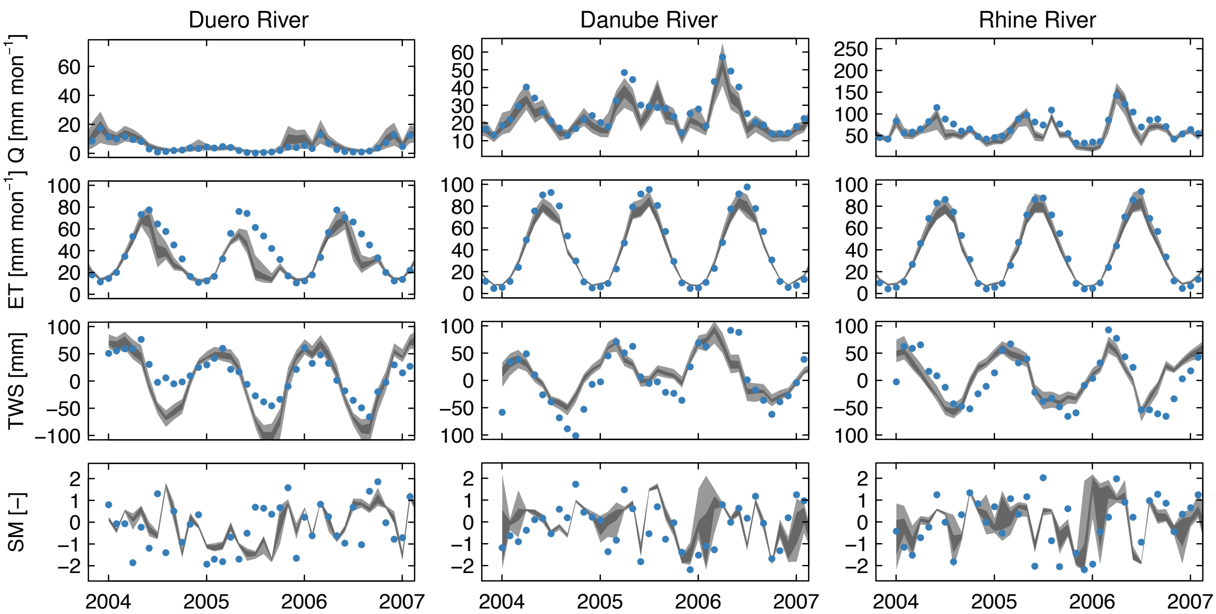

Figure 6.8 Time series of monthly discharge (Q), actual evapotranspiration (ET), total water storage (TWS) anomaly and soil moisture (SM) standardized anomaly for 3 randomly selected basins: (left column) the Duero River at Toro (basin area ≈ 42 000 km^2, $\bar{Q}/\bar{P} = 0.18$); (middle column) the Danube River at Zimnicea (basin area ≈ 660 000 km^2, $\bar{Q}/\bar{P} = 0.41$); and (right column) the Rhine River at Basel-Rheinhalle (basin area ≈ 36 000 km^2, $\bar{Q}/\bar{P} = 0.72$). Observations are shown in blue, mHM simulations are shown given the cross-validation uncertainty with its 95% confidence bounds (light grey) and inter-quartile range (dark grey). Note the different scales for Q (upper row).

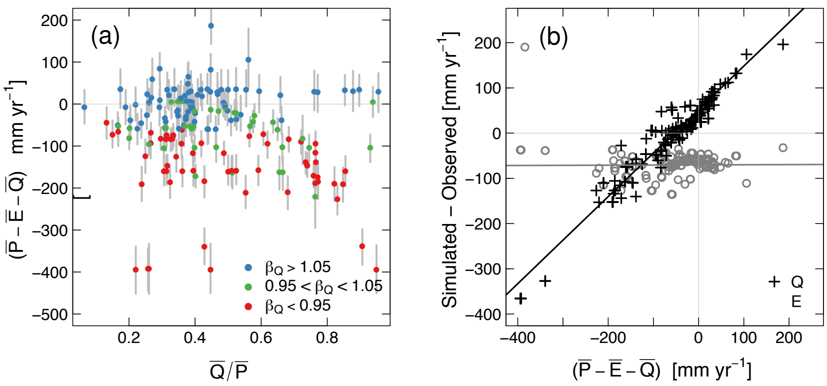

Figure 6.9 Analysis of water balance closure error for the 179 basins with full coverage of observed data for the water years of 1989–2008: (a) Scatter plot between the runoff ratio and median annual water balance closure error with 95% confidence bounds estimated using the bootstrapping method using 1000 bootstrap samples (grey lines). Color indicates discharge model performance in terms of bias β. (b) Relation between median annual water balance closure error and bias between model and observation for streamflow (black) and evapotranspiration (gray).

Overall, the errors in water balance closure are rather independent from the physiographical characteristics with majority of the basins having average annual values between -200 mm yr^{-1} and 100 mm yr^{-1} (Fig. 6.9a). The negative water balance errors are caused by either an underestimated source term (P) or an overestimated sink term (ET+Q). Note that ET is not directly measured but is rather estimated by upscaling observations of biosphere-atmosphere fluxes of carbon and energy from eddy covariance flux tower sites with its own uncertainties (*Jung et al.*, 2011). As noted by *Velpuri et al.* (2013), the errors in ET can yield up to 50% of the mean annual ET values in certain regions. The underestimation of $\bar{P}$ is likely due to inadequate representation of rain gauge coverage failing to capture the small-scale convective events.

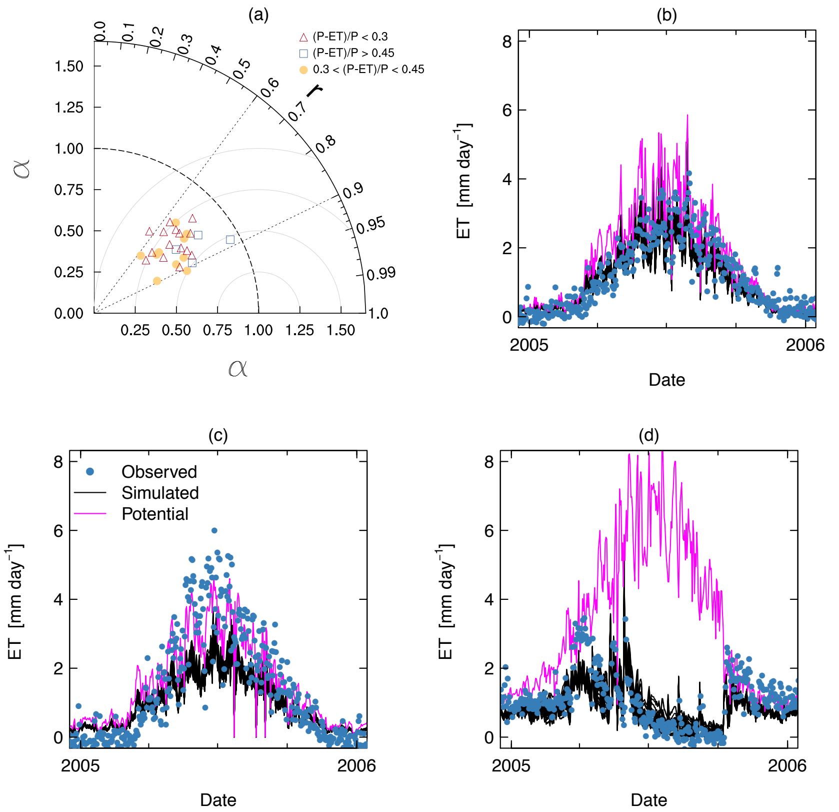

Figure 6.10 (a) Comparison of the daily actual ET between mHM and local-scale estimated derived at 27 eddy stations (see Table 6.2) between 2004 and 2007 using Taylor diagrams. Three daily time series are shown for (b) evergreen needle leaf forest (Tharandt-Anchor in Germany); (c) grassland (Monte Bondone in Italy); and (d) savanna (Mitra II in Portugal). mHM simulations are shown for 36 best parameter sets (black).

The model overestimates the observed discharge in basins where the positive water balance closure errors occur ($\beta_Q > 1.05$, Fig. 6.9a), while underestimations are observed in basins with negative closure errors ($\beta_Q < 0.95$, Fig. 6.9a). This is further supported in Fig. 6.9b which indicates that the water balance closure error follows a close relation with the errors between observed and simulated discharge, with a correlation coefficient of around 0.96. The ET errors do not exhibit any dependency to observed water balance errors, since the correlation coefficient is 0.02. The simulated ET estimates averaged across the investigated basins are consistently underestimated by approximately 70 mm yr^{-1} with respect to observations. Furthermore, the slope of the best fitted line between water balance closure error and Q error is nearly one meaning that on average 100 mm water balance closure error would translate to around 100 mm of simulated streamflow error. The slope can also be interpreted as the elasticity of the fitted line here illustrating the sensitivity of proportional changes in modeling error (in case Q) to the changes in water balance closure error. Results of this analysis indicate that a substantial part of the error in modeled variables can be safely attributed to the erroneous observational datasets. These results highlight the need for better quantification of model errors together with erroneous observational data.

6.4.6 Local-scale evaluation of ET

The multiscale evaluation of mHM is further carried out against daily ET estimated at eddy covariance stations with distinct vegetation cover. Results of this analysis, summarized in Figure 7.11a, indicate that the model is able to capture the temporal dynamics of ET with correlations ranging between 0.6 and 0.9 across 27 eddy covariance stations. In analogy to the gridded scale ET simulation results (see Fig. 6.6), the model systematically underestimates the observed variability indicating the lack of the model to represent the observed range of ET dynamics (i.e., from dry to wet phases). Figures 7.11b–7.11d show the time series of observed and modeled ET at three distinct locations. The temporal dynamics of observations is well represented by the model at the forest and grassland sites with correlations of more than 0.88. Relatively poor performance is observed at the semiarid savanna site ($r = 0.74$). However, the model is able to capture the observed variability including the sudden jumps in ET values observed at this savanna site. Furthermore, the observed magnitude and variability of ET at the grassland site is strongly underestimated by the model particularly during summer. This is due to the fact that the potential ET that is used to force the model at a local scale is generally lower than the observed actual ET.

Results of this analysis provide a first-order confidence that the parameter estimates obtained at much coarser scale can be transferred to finer ones. There is, however, a number of factors that influence the modeling results at local scale, mainly related to the representation of hydrological processes as well as input data. For example, constraining the model against the local forcings instead of large scale E-OBS meteorological forcings may improve its performance. Another limiting factor could be due to the estimation of temperature based potential ET estimates (*Hargreaves and Samani*, 1982), which do not account for other environmental factors, such as wind speed or humidity (*Cristea et al.*, 2012). Finally, the current model version does not account for lateral flows particularly relevant at the small scales.

6.5 Conclusions

The performance of the mesoscale hydrologic model (mHM) parameterized with the multiscale parameter regionalization (MPR) technique is comprehensively evaluated against various *in situ* and satellite-based observations over 400 European river basins. The multiscale evaluation of internal model fluxes and states is carried out at the native resolution of available data varying from 0.5 km to 100 km using an ensemble of cross-validated model parameters constrained only against observed streamflow. Results show that the model is able to perform well for simulating daily discharge over a wide range of climatic and physiographic conditions with KGE greater than 0.55 in more than 50% of the basins. The streamflow predictability deteriorates in basins with a poor rainfall gauge network and in heavily regulated river basins (e.g., Southern Spain). Besides the improvement needed in observational networks, further efforts are needed to incorporate large scale reservoirs operations, irrigation and other human induced water withdrawal and storage activities.

The multiscale evaluation for the complementary datasets generally shows reasonable but lower performance in comparison to streamflow, which is used to constrain the model parameters. The model shows the best agreement with the gridded FLUXNET evapotranspiration ($r > 0.8$), followed by the GRACE total water storage anomaly ($0.6 < r < 0.9$) and the least for the ESA-CCI merged soil moisture ($r < 0.6$). This performance is strongly related to the degree of seasonality that the selected variable exhibits. The skill of the model deteriorates when the annual cycle is removed from each variable except for the soil moisture with majority of the basins exhibiting $r > 0.4$ for the de-seasonalized complementary datasets. The analysis of water balance closure errors indicates that a part of the error in modeled variables is due to erroneous observational datasets. While the error between the observed and simulated discharge is closely related to the errors in the water balance closure estimates, modeled ET is consistently underestimated with respect to observations on average by a constant error of 70 mm yr^{-1}.

The local-scale evaluation of evapotranspiration at several eddy covariance sites further supports the functionality of multiscale parameterization of mHM. While the model is able to capture the temporal dynamics of observed evapotranspiration at most of the sites, it consistently underestimates the observed variability regardless of the locations. Besides improvement in the model parameterization to account for the local scale processes in detail (e.g., sub-grid variability of snow and runoff generation processes), future studies may focus on further enhancement in model performance by constraining the model with site-specific information.

This study provides first order confidence on the ability of the mHM to simulate fluxes and states across a range of spatial scales and varying climatic and physiographic conditions. Due to the implemented MPR technique, it has been possible to run the model at disparate scales native to the observational data, without re-calibrating the model. Although the model yields good performance while conditioned on observed discharge, further improvements are expected by optimally exploiting other reliable complementary datasets together with the streamflow. Results of this study indicate that the null hypothesis that streamflow alone can sufficiently constrain model components responsible for internal fluxes and states in a cross-validation mode can be safely rejected. Therefore, further research should focus on multivariate parameter estimation or assimilation schemes for improving the ability to predict the regional water fluxes and states over large domains.

Acknowledgments

This work is partially funded by the Helmholtz Alliance EDA – Remote Sensing and Earth System Dynamics, through the Initiative and Networking Fund of the Helmholtz Association, Germany. It is also part of the Helmholtz Alliance Climate Initiative REKLIM (www.reklim.de). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (`http://ensembles-eu.metoffice.com`) and the data providers in the ECA&D project (`http://www.ecad.eu`), the U.S. Geological Survey's Earth Resources Observation and Science Center and the NASA's Land Processes Distributed Active Archive Center, the European Environment Agency, the Harmonized World Soil Database, the Global Runoff Data Centre, the Gravity Recovery and Climate Experiment (available at http://grace.jpl.nasa.gov with processing algorithms provided by Sean Swenson, and supported by the NASA MEaSUREs Program), Department Biogeochemical Integration at the Max Planck Institute for Biogeochemistry, Jena, Germany, and the European Space Agency-Climate Change Initiative. The French River runoff data were kindly provided by Jean-Philippe Vidal. Additionally, we acknowledge the PIs at the CarboEurope Eddy covariance sites and the corresponding projects: CarboEuropeIP (EU-FP6), CarboItaly (IT-FISR), CarboMont (EU-FP5), NitroEuropeIP (EU-FP7), GHG-Europe (EU-FP7), and GreenGrass (EU-FP5). These data were downloaded from the European Fluxes Database Cluster at http://gaia.agraria.unitus.it We thank the three anonymous reviewers and the Editor for their comments, which improved the quality of the manuscript.

Appendix: Eddy Covariance Stations

Details on the Eddy Covariance Stations Table6.2 provides detailed information on the eddy covariance stations.

Table 6.2 Overview of the eddy covariance stations including landcover, geographic coordinates and period used in this study.

Number	Site Name (Code[a])	Landcover[b]	Latitude (°N)	Longitude (°E)	Available Period
1	Neustift (AT-Neu)	GRA	47.11667	11.3175	2004–2007
2	Brasschaat (BE-Bra)	MF	51.3092	4.52056	2004–2007
3	Lonzee (BE-Lon)	CRO	50.5522	4.74494	2004–2007
4	Vielsalm (BE-Vie)	MF	50.3055	5.99683	2004–2007
5	Oensingen (CH-Oe1)	GRA	47.2856	7.73214	2004–2007
6	Oensingen (CH-Oe2)	CRO	47.2863	7.73433	2004–2007
7	Grillenburg (DE-Gri)	GRA	50.9495	13.5125	2004–2007
8	Hainich (DE-Hai)	DBF	51.0793	10.452	2004–2007
9	Mehrstedt (DE-Meh)	GRA	51.2753	10.6555	2004–2006
10	Tharandt (DE-Tha)	ENF	50.9636	13.5669	2004–2007
11	Wetzstein (DE-Wet)	ENF	50.4535	11.4575	2004–2007
12	Soroe (DK-Sor)	DBF	55.4869	11.6458	2004–2007
13	Las Majadas (ES-LMa)	SAV	39.9415	-5.7734	2004–2007
14	Vall d'Alinya (ES-VDA)	GRA	42.1522	1.4485	2004–2007
15	Grignon (FR-Gri)	CRO	48.844	1.95243	2004–2007
16	Hesse (FR-Hes)	DBF	48.6742	7.06462	2004–2007
17	Le Bray (FR-LBr)	ENF	44.7171	-0.7693	2004–2007
18	Bugac (HU-Bug)	GRA	46.6911	19.6013	2004–2007
19	Matra (HU-Mat)	GRA	47.8469	19.726	2004–2006
20	Lavarone (IT-Lav)	ENF	45.9553	11.2812	2004–2006
21	La Mandria (IT-LMa)	GRA	45.5813	7.15463	2004–2006
22	Monte Bondone (IT-MBo)	GRA	46.0156	11.0467	2004–2007
23	Renon (IT-Ren)	ENF	46.5878	11.4347	2004–2007
24	Roccarespampani 2 (IT-Ro2)	DBF	42.3903	11.9209	2004–2007
25	San Rossore (IT-SRo)	ENF	43.7279	10.2844	2004–2007
26	Loobos (NL-Loo)	ENF	52.1679	5.74396	2004–2007
27	Mitra IV (PT-Mi2)	GRA	38.4765	-8.0246	2004–2007

[a] First two capital letters of the code stand for the country code.
[b] GRA: grassland; MF: mixed forest; CRO: cropland; DBF: deciduous broadleaf forest; ENF: evergreen needles forest; SAV: savanna.

CHAPTER 7

CONDITIONING A HYDROLOGIC MODEL USING PATTERNS OF REMOTELY SENSED LAND SURFACE TEMPERATURE

M. Zink[1], ..., and L. Samaniego[1]

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Zink, M., J. Mai, M. Cuntz, and L. Samaniego (2018), Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature, Water Resources Research, 54, 2976–2998

7.1 Abstract

Hydrologic models are usually calibrated using observed river runoff at catchment outlets. Streamflow, however, represents an integral response of the entire catchment and is observed at a few locations worldwide. Parameter estimation based on streamflow has the disadvantage that it does not consider the spatiotemporal variability of hydrologic states and fluxes such as evapotranspiration. Remotely sensed data, in contrast, include these variabilities and are broadly available. In this study, we assess the predictive skill of satellite-derived land surface temperature (T_s) with respect to river runoff (Q). We developed a bias insensitive pattern-matching criterion to focus the parameter optimization on spatial patterns of T_s. The proposed method is extensively tested in six distinct large German river basins and cross-validated in 222 additional basins in Germany. We conclude that land surface temperature calibration outperforms random drawn parameter sets, which could be meaningful for calibrating hydrologic models in ungauged locations. A combined calibration with Q and T_s reduces the root mean squared error in the predicted evapotranspiration by 8% compared to flux tower observations but reduces the NSEs of the streamflow predictions by 6% on average for the six large basins. Our results show that patterns of T_s better constrain model parameters when considered in a calibration next to Q, which finally reduces parametric uncertainty.

7.2 Introduction

Hydrologic models (HM) are usually calibrated against streamflow at outlets of basins and thus only consider integrated signals to model the response of the entire basin. That procedure ensures the fulfillment of the mass balance but has no control over the spatial distribution of hydrologic fluxes and states such as evapotranspiration and soil moisture. Drought monitoring and forecasting, however, rely on spatially representative simulations of evapotranspiration and soil moisture. Model parameterizations aiming on streamflow lead to adequate estimations of discharge, but in general induce discontinuities (i.e., a lack of seamlessness) in parameter fields (*Merz and*

Blöschl, 2004; *Samaniego et al.*, 2017), which in turn, leads to diminished accuracies for other fluxes such as evapotranspiration (*Rakovec et al.*, 2016a; *Zink et al.*, 2017). To overcome these deficiencies, parameterization of a hydrologic model assisted by spatially distributed satellite observations is investigated in this study.

Spatially distributed ground observations of land surface fluxes and states do not yet exist for regional or larger scales and likely will never exist (*Vereecken et al.*, 2008). Thus, satellite data remain the only resource for spatially explicit observations of the Earth surface. From the perspective of a hydrological modeler, satellite soil moisture or evapotranspiration observations are preferable for constraining hydrologic models because those variables are model inherent. Those data, however, have several disadvantages. First, the estimation of satellite soil moisture and evapotranspiration is based on inverse modeling techniques (e.g., *Bastiaanssen et al.*, 1998; *Mu et al.*, 2007; *Wagner et al.*, 2007), which convert the satellite signal into hydrologic state variables and fluxes and rely on own parameterization schemes for the soil and vegetation. Using the parameterization of the hydrological model for this conversion ensures consistency in the parameterization between remote sensing product and hydrological model and should be preferred. Second, the satellite retrievals still underlie large uncertainties and inaccuracies (*Sheffield and Wood*, 2011). Third, the spatiotemporal resolutions of freely available satellite products with regional to global coverage such as soil moisture and evapotranspiration are coarse ($\geq 25\ km$ and $\geq 30\ d$) compared to the spatial resolutions of mesoscale hydrologic models, which range from $1\ km$ to $10\ km$ and from hourly to daily temporal scales, respectively.

An alternative source of data is land surface temperature (T_s), which is based on satellite-based thermal-infrared (TIR) observations. TIR is directly interlinked with T_s through the radiative temperature equation (*Li et al.*, 2013b), which depends only on corrections for atmospheric and emissivity effects (*Li et al.*, 2013b), not on soil or vegetation characteristics. For those reasons, satellite-derived T_s are regarded as a more robust source of information compared to soil moisture or evapotranspiration retrievals and consequently is preferred in this study. Formally, T_s is defined as the temperature of the interface between the Earth's surface and atmosphere (*Niclòs et al.*, 2011) and is directly connected to latent heat via the energy balance equation at the land surface. *Lakshmi* (2000) showed that a strong relationship between T_s and soil moisture exists, and that the calibration of a land surface model (LSM) with that variable was able to improve its soil moisture estimation.

As a consequence, land surface temperature was considered to be a promising variable for enhancing the spatial representation of evapotranspiration and/or soil moisture in hydrological models. The land surface temperature is the connection between the water and the energy balances and depends on the estimated evapotranspiration within the hydrological model. The state-of-the-art, however, does not indicate conclusive results. *McCabe et al.* (2005), for example, observed changes in the spatial distribution of evapotranspiration when calibrating a land surface model (LSM) with T_s. *Boni et al.* (2001) and *Reichle et al.* (2010) assimilated T_s using a variational assimilation scheme and an Ensemble Kalman Filtering technique, respectively. Both studies employed LSMs which implicitly solve the energy balance and thus estimate T_s. *Boni et al.* (2001) concluded that the control of surface temperature on evaporation is feasible, whereas *Reichle et al.* (2010) did not observe any effect on surface energy fluxes.

The calibration of a hydrological model with T_s was originally proposed by *Crow et al.* (2003). In that study they found that the consideration of spatially averaged T_s, besides streamflow, improved monthly evapotranspiration predictions by up to 20%. Similar efforts were undertaken by *Corbari et al.* (2010, 2015) and *Silvestro et al.* (2013, 2015).

All those studies have in common that only selected model parameters were considered during the calibration process, whereas the remaining parameters were estimated by prior knowledge, i.e., by transferring parameters from remote locations or by setting them using expert knowledge. It should be noted that the land surface models employed in those studies explicitly solved the energy balance and thus inherently estimated the land surface temperature. Their authors, however, did not specifically focus on the spatial distribution of T_s. Consequently, models were either calibrated using basin averaged T_s (*Silvestro et al.*, 2013, 2015) or compared observations and simulations using standard error metrics such as bias or root mean squared error (*Corbari et al.*, 2010, 2015). *Reichle et al.* (2010); *Stisen et al.* (2011), and *Koch et al.* (2015) on the other hand suggested using bias insensitive metrics, which only consider the spatial patterns of land surface temperature. Such measures are preferred because satellite-derived T_s is known to be biased when compared to ground observations (*Li et al.*, 2013b; *Niclòs et al.*, 2011; *Reichle et al.*, 2010; *Trigo et al.*, 2008). The novelty of our study is to take advantage of pattern matching criteria for constraining hydrological models.

Spatially distributed data can hence improve model calibration, improving hydrologic states and fluxes. It also impacts the estimation of model parameter uncertainty by improving parameter identifiability and by reducing equifinality (*Beven*, 1993, 2001).

In this study, we postulate that the simultaneous calibration of streamflow Q and T_s will affect the spatial distribution of evapotranspiration (E) and improve estimates of E when compared to eddy covariance observations. Furthermore, we hypothesize that the uncertainty in the global model parameters of a hydrological model will decrease when Q and T_s are considered simultaneously because this approach would better constrain parameters related to land surface processes such as evapotranspiration. The aim of the third analysis is to assess the predictive skill of the hydrological model regarding streamflow if the model is calibrated with T_s alone to evaluate if T_s could be a useful variable for model calibration if streamflow data are not available.

Finally, to test those hypotheses, we developed a diagnostic algorithm to estimate land surface temperature within the mesoscale Hydrologic Model (mHM). For the parameter estimation we propose a non-parametric, bias-insensitive pattern-matching criterion.

7.3 Study Domain and Data

7.3.1 Meteorological Data

The forcings needed for mHM were provided by the German Meteorological Service (DWD). The approximately 2000 precipitation and 1100 temperature stations covering Germany were interpolated using external drift Kriging (*Ahmed and De Marsily*, 1987; *Zink et al.*, 2017). A digital elevation model was used as an external drift. The potential evapotranspiration was estimated based on the Hargreaves-Samani equation (*Hargreaves and Samani*, 1985) and using the interpolated fields of minimum, maximum and average daily air temperature. Precipitation, average temperature and potential evapotranspiration are the main forcings for solving the water balance within the hydrologic model mHM. The spatial resolution is 4×4 km^2 since we consider that to be the lowest spatial resolution supported by the station input data.

The proposed land surface temperature module requires net radiation as an additional input. Observed T_s is also needed as an input for the evaluation of the modeled T_s. In general, observations of land surface temperature and radiation are unfortunately not available or are very sparse. In Germany, for example there are at most 60 radiation measurement stations. Alternative data sources relevant for this study are reanalysis or remote sensing data. These data sources are required to be of similar spatiotemporal resolutions as the above-mentioned meteorological forcings. Reanalysis data are typically coarser than 0.25° and are thus not appropriate for this study. Satellite data are derived from either polar orbiting or geostationary satellites. Although, the spatial resolutions (e.g., 1 km) of polar-orbiting satellites such as TERRA are large and equidistant, their temporal resolutions are coarse (one to two overpasses per day). Geostationary satellites (e.g., Meteosat), however, have high temporal (≥15 minutes) but lower spatial resolutions. Their spatial resolution decreases with increasing latitude.

Herein, we use data from Meteosat Second Generation (MSG) that have been processed by the Land Surface Analysis - Satellite Application Facility (LSA-SAF, *EUMETSAT* (2016)). The data have an average spatial resolution of 3.5 × 6.5 km^2 for Germany and a temporal resolution of 15 or 30 minutes depending on the data product. LSA-SAF provides land surface temperature and downwelling radiation data. The required net radiation is estimated based on downwelling shortwave and longwave radiation (30 minute resolution), land surface temperature (15 minute resolution), emissivity (15 minute resolution), and albedo (1 day resolution) products from LSA-SAF (see section 7.4.2, Equations 7.4a-7.4c). The inter-daily data have been aggregated to daily values by filtering only valid data points derived from quality flags provided by LSA SAF. Consequently, data points affected by high cloud coverage are discarded. Days with less than 50% valid data for a particular pixel are also discarded for the calculation of the daily mean value. Finally, pattern analysis is performed only for pixels having valid data for the respective day. Data gaps are unproblematic in this study because they are neglected by the land surface temperature module and the error measures presented in section 7.4.

The downwelling shortwave radiation (*EUMETSAT*, 2016) is evaluated using available observations from 28 stations in the year 2009 (*Deutscher Wetterdienst (DWD)*, 2011). The evaluation showed that both datasets are comparable, with an average Pearson correlation coefficient of 0.93 (standard deviation=0.04) and a relative bias of 5% (standard deviation=2%), and were thus applicable for this study. In addition, we validated the LSA-SAF's land surface temperatures at two eddy covariance stations, i.e., stations E2 and E4 (Figure 7.1), where radiometric temperatures are measured (*Kutsch et al.*, 2008; *Rebmann et al.*, 2010). The comparison revealed a bias in the satellite retrievals of approximately 2.7 K. Similar biases of 2 K to 3 K have been reported in the literature (*Li et al.*, 2013b; *Niclòs et al.*, 2011; *Reichle et al.*, 2010; *Trigo et al.*, 2008). These biases have to be taken into account when comparing modeled and satellite-retrieved T_s.

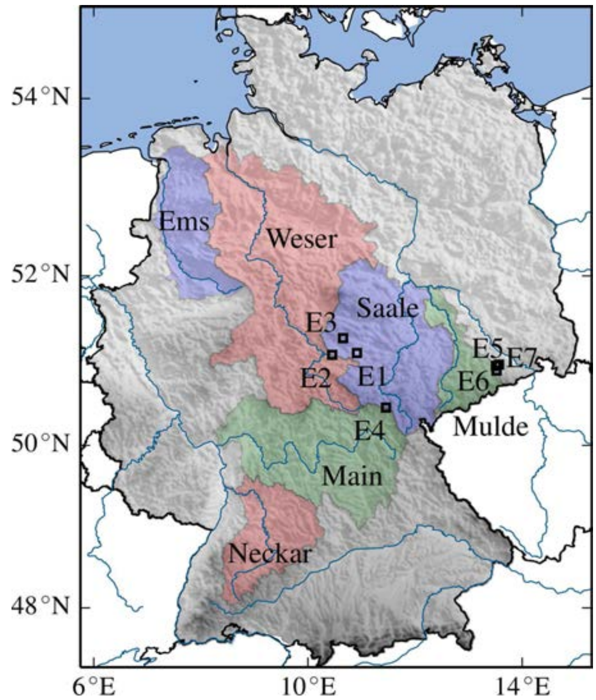

Figure 7.1 The main basins used for parameter inference and numerical experiments. The six major inner German river basins span over a climate gradient ranging from maritime influence in the Ems to continental climate in Main and Neckar. The points E1 to E7 depict the location of eddy flux tower observations, which have been used for evaluating the simulated evapotranspiration.

The study period has been restricted to the year 2009 because this was the only period in which meteorological, LSA SAF and discharge data for each of the 6 main study sites have been available. Discharge data past 2009 were not available for all six catchments. On the other hand broadband emissivity data from LSA SAF have just been available from 2009 onwards. The restriction to one year of data constitutes a clear limitation of this study and can be overcome in future studies due to longer available discharge data. The year 2009 is, however, within the 30th and 70th percentile of the climatology (1951-2010) in terms of annual catchment water availability in all six catchments and can therefore be seen as a valid choice. The catchment water availability is estimated as precipitation minus potential evapotranspiration ($P - E_p$). The climatology of the annual discharge is for all catchments within the 25th and 75th percentile with exception of the Ems basins which was slightly drier in 2009 showing a percentile of 24 (see supplemental material).

7.3.2 Study Domain and Land Surface Properties

The study domain covers a large fraction of Germany. Intensive analyses will be presented for the six inner German river basins presented in Figure 7.1. These are the largest inner German basins and differ in size, hydrologic behavior and climatic conditions. They range from a flat, agriculturally dominated, maritime influenced basin in northern Germany (Ems) to a snow influenced and more continental basin with distinct slopes in the south (Neckar). A detailed description of the basins can be found in *Zink et al.* (2017).

The land surface is characterized by a digital elevation model provided by the *Federal Agency for Cartography and Geodesy (BKG)* (2010), a soil and hydrogeological map offered by the *Federal Institute for Geosciences and Natural Resources (BGR)* (1998, 2009), and land cover information from the *European Environmental Agency (EEA)* (2009). These data are discretized to a spatial resolution of 100×100 m^2. The *Global Runoff Data Centre* (2017) and the *European Water Archive (EWA)* (2011) provided the daily streamflow data.

Next to the six main study sites we use 222 uncalibrated basins for evaluating parameters regarding discharge and 7 eddy flux towers for evaluating evapotranspiration simulations. A detailed description and metadata of these sites can the found in *Zink et al.* (2017).

7.4 Methodology

7.4.1 The mesoscale Hydrologic Model mHM

This study's computational experiments were conducted employing the mesoscale Hydrologic Model mHM (`www.ufz.de/mhm`) (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a). It is a process-based and spatially distributed model that was developed for the estimation of hydrologic fluxes and state variables on the land surface. These states and fluxes are derived by closing the water balance on every grid cell. Within a grid cell the governing processes are conceptualized as discrete reservoir models for the different compartments of the hydrologic cycle, e.g., interception, snow accumulation, soil moisture, and evapotranspiration. The estimation of hydrologic fluxes and states is highly dependent on the quality of the evapotranspiration estimate since it is the second most important flux in the water balance, following precipitation.

The evapotranspiration within mHM is based on potential evapotranspiration (PET). The evapotranspiration flux in mHM is mainly estimated by reducing the potential evapotranspiration according to the available amount of soil water within the root zone (*Feddes et al.*, 1976). Minor contributions to evapotranspiration originate from the interception storage and evaporation from surface water retention. Evapotranspiration estimation is highly dependent on the representation of the soil water retention in land surface models and is therefore described in the following.

Within the root zone the amount of evaporative water is determined within the different soil layers. The number and depth of the soil layers can be defined by the user whereas the depth of the deepest soil layer is derived by mHM from the soil map and thus varies among grid cells. In this study, the soil was discretized into three layers. The first layer ends 5 cm below the surface, the second at 25 cm and the third is the soil map dependent layer. The evapotranspiration of the single soil layers is consecutively estimated with depth as a function of (a) potential evapotranspiration, (b) soil water content, (c) soil hydraulic properties (permanent wilting point, field capacity, and saturated soil moisture content) and (d) the fraction of roots. This functional relationship contains several model parameters.

The model parameters are derived by employing the Multiscale Parameter Regionalization (MPR) technique (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a). This methodology is based on space and time invariant parameters, the so called *global parameters* (*Pokhrel et al.*, 2008). These *global parameters* parameters are subject to calibration and are used for the estimation of spatially distributed parameter fields. These fields result from the functional relationships (transfer functions) between the *global parameters* and physiographical characteristics of the catchment.

In MPR, the transfer functions (e.g., the pedotransfer functions for the estimation of soil parameters) are connected to the morphological input (e.g., soil textural properties) and thus lead to model parameters (e.g., porosity and soil hydraulic conductivity). In the example, the *global parameters* are the coefficients of the pedotransfer functions. The model parameter estimation is performed on the resolution of the morphological input (e.g., 100×100 m^2). They must be upscaled to determine the model parameters at the hydrologic model resolution (e.g., 4×4 km^2). The applied upscaling rules are different for the various model parameters (e.g., the geometric mean for the porosity and soil hydraulic conductivity). Detailed information about model parameters and upscaling rules are provided in *Kumar et al.* (2010, 2013); *Samaniego et al.* (2010a); *Zink et al.* (2017).

Compared to other parameter estimation approaches such as hydrologic response units (*Flügel*, 1995), the advantages of MPR are (1) the ability to choose flexible model resolutions without the necessity to rescale inputs, (2) transferability of the *global parameters* across locations (*Kumar et al.*, 2010; *Rakovec et al.*, 2016a; *Zink et al.*, 2017), and (3) transferability across scales (*Kumar et al.*, 2013,b; *Samaniego et al.*, 2010a) without recalibrating the model. mHM also showed its capability in impact assessment studies such as those of *Samaniego et al.* (2013) and *Thober et al.* (2015) and for operational purposes (*Zink et al.*, 2016).

7.4.2 Development of a Land Surface Temperature Module

A goal of this study is to incorporate spatially distributed information into the hydrologic model mHM to improve the spatial representativeness of the hydrologic fluxes and states. Herein, we aim on evapotranspiration because it has a large impact on water balance. Evapotranspiration is linked with land surface temperature by the energy balance. We employ satellite-derived land surface temperature fields within the hydrologic model mHM. The spatiotemporal distribution of land surface temperature (T_s) was used to constrain mHM in addition to streamflow (Q).

Because the purpose of mHM is to solve the water balance equation, land surface temperature was not yet calculated. The energy balance is used to simulate land surface temperature. In consequence, the evapotranspiration acts on the energy and the water balances.

The following section will introduce a parsimonious module for estimating land surface temperature based on modeled evapotranspiration (E) and given short- and longwave radiation inputs. The module is called the land surface temperature module in the following. It can be coupled to any hydrologic model and was interfaced with mHM in this study.

On the one hand, the evapotranspiration E $[mm\ d^{-1}]$ is determined by closing the water balance

$$E = P - Q - \Delta S \tag{7.1}$$

with mHM . Where P is precipitation $[mm\ d^{-1}]$, Q is streamflow $[mm\ d^{-1}]$, and ΔS is the change in the storages $[mm\ d^{-1}]$, e.g., soil moisture. On the other hand the energy balance of the land surface can be written as:

$$R_n = \lambda E + H + G + S \tag{7.2}$$

with net radiation R_n $[W\ m^{-2}]$, latent heat flux λE $[W\ m^{-2}]$, sensible heat flux H $[W\ m^{-2}]$, soil heat flux G $[W\ m^{-2}]$ and any storages S $[W\ m^{-2}]$, for example photosynthetic or biomass heat storage. The latent heat flux λE is determined by converting the mass flux of the evapotranspiration E estimated by mHM (Equation 7.1) to an energy flux. For that reason, the evapotranspiration E $[mm\ d^{-1}]$ is multiplied by the latent heat of vaporization λ $[kJ\ kg^{-1}]$. The density of water is taken as $\varrho = 1000\ kg\ m^{-3}$. The latent heat of vaporization λ is approximated by $\lambda = 2501 - 2.37T_a$ using the air temperature T_a $[°C]$ (*Dyck and Peschke*, 1995).

The land surface temperature is estimated using the temporal resolution of one day because this is the temporal resolution of the meteorological input. For daily time steps it is assumed that the soil heat flux G and the storage terms S are negligible (*Haverd et al.*, 2007), such that Equation 7.2 simplifies to

$$H = R_n - \lambda E. \tag{7.3}$$

To solve Equation 7.3 the net radiation R_n has to be provided as an input to mHM. Because spatially comprehensive measurements of the net radiation are not available (see section 7.3.1), it is estimated from incoming radiation components from satellites. We therefore use

$$R_n = Q_S^{\mathrm{in}} - Q_S^{\mathrm{out}} + Q_L^{\mathrm{in}} - Q_L^{\mathrm{out}} \tag{7.4a}$$

$$Q_S^{\mathrm{out}} = \alpha Q_S^{\mathrm{in}} \tag{7.4b}$$

$$Q_L^{\mathrm{out}} = \epsilon\sigma\widehat{T_s}^4 \tag{7.4c}$$

Q_S^{in} and Q_S^{out} are the incoming and outgoing shortwave radiation $[W\ m^{-2}]$, respectively, and Q_L^{in} and Q_L^{out} are the incoming and outgoing longwave radiation $[W\ m^{-2}]$, respectively. The outgoing shortwave radiation Q_S^{out} is estimated using Equation 7.4b, in which α is the albedo of the land surface $[-]$. The outgoing longwave radiation Q_L^{out} is approximated as the emission of a gray body emissivity ϵ which can be calculated using the Stefan-Boltzmann law (Equation 7.4c) and the and the Stefan-Boltzmann constant $\sigma = 5.67 \cdot 10^{-8}\ W\ m^{-2}\ K^{-4}$.

Equation 7.3 for sensible heat flux H $[W\ m^{-2}]$ modifies to

$$H = (1-\alpha)Q_S^{\mathrm{in}} + Q_L^{\mathrm{in}} - \epsilon\sigma\widehat{T_s^4} - \lambda E. \tag{7.5}$$

The thermodynamic formulation of the sensible heat H can be written as

$$H = \varrho_a c_p \frac{\widehat{T_s} - T_a}{r_a} \tag{7.6}$$

where T_a is the air temperature $[K]$, $\widehat{T_s}$ is the model derived land surface temperature $[K]$, r_a is the aerodynamic resistance $[s\ m^{-1}]$, ϱ_a is the density of air ($\varrho_a = 1.29\ kg\ m^{-3}$) and c_p is the specific heat capacity of air, which is assumed to be constant ($c_p = 1004\ J\ kg^{-1}\ K$). Combining Equation 7.5 and 7.6 leads to a fourth degree polynomial in $\widehat{T_s}$:

$$(1-\alpha)Q_S^{\mathrm{in}} + Q_L^{\mathrm{in}} - \lambda E + \frac{\varrho_a c_p}{r_a}T_a - \frac{\varrho_a c_p}{r_a}\widehat{T_s} - \epsilon\sigma\widehat{T_s}^4 = 0. \tag{7.7}$$

In summary, $\widehat{T_s}$ is the modeled variable of interest, Q_S^{in}, Q_L^{in}, α, and ϵ are satellite-retrieved variables, ϱ_a, σ, and c_p are constants, T_a is measured air temperature, which is an input for mHM, λE is derived by closing the water balance with mHM (Equation 7.1), and r_a is the aerodynamic resistance which is still unknown but will be explained in the following.

Solving the Equation 7 for $\widehat{T_s}$ leads to four possible solutions which are the roots of the quartic equation. The root that falls within the interval $[0\ K, 500\ K]$ is the feasible result for $\widehat{T_s}$. During all the experiments, only one of the four roots fulfilled this requirement. The calculation of that root is presented in Appendix 7.6.

The aerodynamic resistance r_a $[s\ m^{-1}]$ is calculated following *Allen et al.* (1998):

$$r_a = \frac{\ln\left(\frac{z_m - d}{z_{0m}}\right)\ln\left(\frac{z_h - d}{z_{0h}}\right)}{k^2 u_z} \tag{7.8}$$

where z_h is the height of the humidity measurement $[m]$, d is the zero plane displacement height $[m]$, z_{0m} is the roughness length for momentum transfer $[m]$, z_{0h} is the roughness length for heat transfer $[m]$, k is the von Karman constant ($k = 0.41$), and u_z is the wind speed $[m\ s^{-1}]$ at the wind speed measurement height z_m in $[m]$. It is assumed that the measurement heights of wind speed and humidity are equal, i.e., $z = z_m$=z_h.

The approximations of the three variables $d = \frac{2}{3}h_c$, $z_{0m} = 0.123h_c$, and $z_{0h} = 0.1z_{0m}$ are taken from *Allen et al.* (1998). The constant coefficients for d, z_{0m} and z_{0h} have been implemented as the *global parameters* p_{48}, p_{49}, and p_{50} in the land surface temperature module, respectively. These parameters need to be calibrated, and their ranges were chosen to be between $\pm10\%$ of the values reported by *Allen et al.* (1998). Thus Equation 7.8 therefore becomes

$$r_a = \frac{\ln\left(\frac{z - p_{48}h_c}{p_{49}h_c}\right)\ln\left(\frac{z - p_{48}h_c}{p_{49}p_{50}h_c}\right)}{k^2 u_z}. \tag{7.9}$$

The equation shows that besides the given height z and the measured wind speed u_z, r_a is dependent on the estimation of the parameters p_{48}, p_{49}, and p_{50} and the canopy height h_c. The Multiscale Parameter Regionalization (MPR) technique was employed to estimate canopy height h_c because no spatially comprehensive information regarding h_c was available. Hence, h_c becomes a calibration parameter that is dependent on the land cover information.

The canopy height for the mixed land cover class takes the monthly evolution of the leaf area index (LAI) into account. The mixed land cover class is a generalized class consisting of grasslands, agricultural areas, and pastures. The relationship is assumed to be

$$h_{c,\mathrm{mix}}(m) = p_{47}\frac{\mathrm{LAI}(m)}{\max\limits_{m}\mathrm{LAI}(m)}, \quad m = 1, \ldots, 12\ . \tag{7.10}$$

in which $h_{c,\mathrm{mix}}(m)$ is the canopy height $[m]$ of the mixed land cover class (mix) for month m, $\mathrm{LAI}(i)$ is the leaf area index $[m^2\ m^{-2}]$ for month m, and p_{47} is a calibration parameter $[m]$.

Both of the other land cover classes (forest (for) and sealed (seal)) are assumed to be constant in canopy height over the course of a year and do not depend on LAI:

$$h_{c,\mathrm{for}} = p_{45} \qquad \text{and} \qquad h_{c,seal} = p_{46}\ . \tag{7.11}$$

The canopy height is estimated at the resolution of the physiographic input data, i.e., 100×100 m^2. Various upscaling operators were tested for the upscaling to the model resolution, i.e., 4×4 km^2, and the arithmetic mean was proven to perform best.

In summary, we have presented the development of a land surface temperature module that can be coupled to any environmental model. Satellite-derived radiation components ($Q_S^{(in)}$ and $Q_L^{(in)}$), air temperature (T_a), wind speed (u_z), and modeled evapotranspiration (E) are used as inputs for the land surface temperature module. The necessary steps for estimating land surface temperature ($\widehat{T_s}$, Equation 7.7) are

1. the estimation of E as residual of the water balance (Equation 7.1) and

2. the calculation of $\widehat{T_s}$ (Equation 7.7) based on the aerodynamic resistance (Equation 7.9).

To approximate the aerodynamic resistance r_a (Equation 7.9) the three *global parameters* p_{48} to p_{50} connected to the displacement height and the roughness lengths as well as the three *global parameters* p_{45} to p_{47} connected to the canopy height (Equation 7.10 and Equation 7.11) are necessary. An additional parameter p_{51} was introduced into Equation 7.7 to account for the biases that have been observed in the satellite-retrieved T_s. This parameter could be neglected because a bias insensitive error measure was designed for calibrating mHM with T_s (see section 7.4.3).

These seven *global parameters*, $p_{45} - p_{51}$, are estimated through the automated calibration of the mHM model. The difference between the satellite-derived T_s and simulated land surface temperature $\widehat{T_s}$ is minimized during model calibration. The calibration procedure will be explained in the following.

The land surface temperature model driven by satellite observations has, however, certain limitations. First, the satellite data of land surface temperature, short and longwave radiation, albedo and emissivity on its own come with uncertainties. These uncertainties arise from assumptions in the underlying models for translating satellite reflectances to derived variables. Another source of uncertainty are meteorological model data, which are used to correct for atmospheric influences on the observed satellite signal, e.g. atmospheric humidity. Another source of uncertainty are model parameters of the land surface temperature module itself which are further discussed in section 7.5.2.

7.4.3 Optimization of the Coupled mHM-Land Surface Temperature Model

The aforementioned T_s module is coupled to mHM. For simplicity, the coupled mHM-land surface temperature model is denoted as mHM in the following. All of the parameters, including the mHM *global parameters* (44 parameters) and the seven additional parameters of the T_s module are herein referred to as the mHM parameters. The coupled model, therefore, has 51 *global parameters* for the purpose of calibration.

The coupled model will be calibrated against land surface temperature or streamflow or a combination of both. The performance regarding the two model outputs, i.e., streamflow and land surface temperature, is estimated using a weighted objective function. In general the objective function Φ is estimated by

$$\Phi = \left(\sum_{i=1}^{n} (w_i)^p (\phi_i)^p \right)^{\frac{1}{p}} \tag{7.12}$$

where w_i is the weight ($\sum_{i=1}^{n} w_i = 1$) of the error measure ϕ_i of the n objectives. Different error measures ϕ_i are considered because streamflow Q only depends on time, whereas T_s is a spatiotemporal variable. Following *Duckstein* (1984), the exponent p was set equal to 6 to assure numerical stability and assure a compromise solution. The exponent p ensures that the progress in competing objectives does not compensate each others effects, e.g., if ϕ_1 improves and ϕ_2 declines the overall objective function could improve because the improvement of ϕ_1 outperforms ϕ_2 if a compromise solution is not applied. The power law ensures that the improvement of ϕ_1 gets less weight and thus the optimizer will put more emphasis on improving ϕ_2. The different error measures ϕ for streamflow and land surface temperature are described in the following.

Error Measure for Streamflow Q The Nash-Sutcliffe Efficiency (NSE) (*Nash and Sutcliffe*, 1970) is applied to assess the model performance regarding streamflow. To obtain satisfactory estimates of high flows as well as of low flows, the NSE is determined for the daily streamflow (ϕ_1) and the logarithm of the daily streamflow (ϕ_2), respectively. For the optimization against streamflow alone, ϕ_1 and ϕ_2 are considered in the objective function. The weights are chosen to be equal for both criteria ($w_1 = w_2 = 0.5$).

Error Measure for Land Surface Temperature T_s An error measure for quantifying the differences between modeled and satellite-retrieved T_s is developed in this section. The satellite retrievals of T_s had an inherent bias of approximately 2 K to 3 K at the temporal resolution of one day in comparison to ground measurements (see section 7.3.1). It is assumed that the patterns delivered by the satellite measurements are trustworthy. Thus, an objective that compares patterns of the satellite-retrieved and model estimated land surface temperature qualitatively was targeted.

The application of error measures that are sensitive to a biases such as the mean squared error was therefore not considered. A bias resistant, local and non-parametric measure denoted as pattern similarity (P) was developed. Mathematically, the new pattern similarity criterion can be expressed as

$$\phi_3 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i,j \in \Omega}^{N} P_{ij}(t) \tag{7.13a}$$

$$P_{ij}(t) = \frac{1}{2M} \sum_{k=1}^{M} \left[\tilde{\text{sgn}} \left(\widehat{T}_{s,ij}^{(k)}(t) - \widehat{T}_{s,ij}(t) \right) \tilde{\text{sgn}} \left(T_{s,ij}^{(k)}(t) - T_{s,ij}(t) \right) + 1 \right] \tag{7.13b}$$

where i and j are the elements of the spatial domain Ω, which in total consists of N cells, T is the number of time steps, $P_{ij}(t)$ is the pattern similarity criteria at cell (i,j) at a particular time step t, $T_{s,ij}^{(k)}$ is the land surface temperature of the k^{th} of M neighbors of the center cell (i,j), and $T_{s,ij}$ is the land surface temperature of the center cell itself. The pattern similarity criterion is normalized with M, the number of neighbors of the center cell (i,j). M typically equals eight but can vary at the basin boundaries. The notation without a hat (T_s) is used for the satellite-derived land surface temperature, whereas the model simulated temperature is denoted with a hat ($\widehat{T_s}$). The $\tilde{\mathrm{sgn}}$ operation determines the sign of the argument a as follows:

$$\tilde{\mathrm{sgn}}(a) = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{if } a \leq 0 \end{cases}. \tag{7.14}$$

An example for the pattern similarity criterion is depicted in Fig. 7.2.

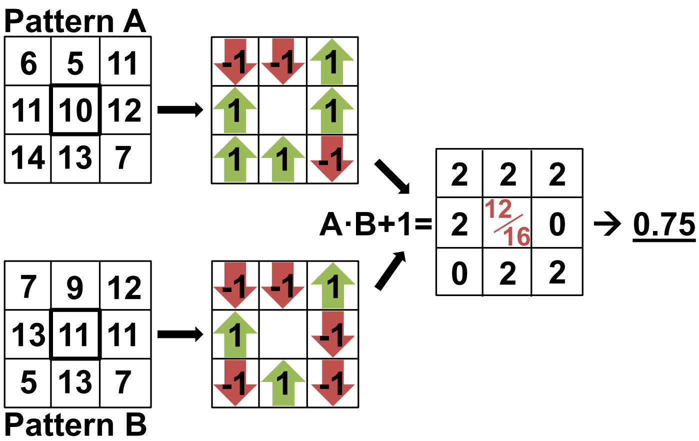

Figure 7.2 Schematic description of the pattern similarity criterion according to Equation 7.13b. In the upper left row, an example pattern A with the center pixel having a value of 10 is illustrated (e.g., satellite-retrieved T_s in $°C$). The sign of the comparison between the center pixel with its neighboring pixels is shown on its right. If the respective neighboring pixel is larger than the center pixel (green arrow) the value 1 is assigned to this pixel (e.g., 5 pixels in pattern A), otherwise (red arrow) the value -1 is assigned to them (e.g., 3 pixels in pattern A). This analysis is repeated for a pattern B (e.g., simulation of $\widehat{T_s}$), as depicted in the lower row. The results of both comparisons are multiplied and increased by 1. Thus, the dissimilar pixels between patterns A and B are assigned values of 0, whereas pixels with the same tendency are assigned values of 2. The elements of the resulting matrix are summed and divided by twice of the number of neighbors (e.g., 16). For the given example, the pattern similarity criterion is 0.75, meaning that three-quarters of the neighbors showed the same relation to its center value in both patterns A and B.

The criterion is based on a 3×3 pixel search raster. Its center cell is subtracted from the eight neighboring cells. The difference becomes negative, and the $\tilde{\mathrm{sgn}} = -1$ if the value of the center cell is greater than the neighbor. In the opposite case the sign becomes positive ($\tilde{\mathrm{sgn}} = 1$). This procedure is applied to both fields under comparison, i.e., the satellite-retrieved T_s and the modeled land surface temperature $\widehat{T_s}$. The two resulting 3×3 signum matrices are multiplied together. The resulting matrix has a negative entry (-1) where the elements of both factors had different signs and a positive entry (+1) where the factors had the same sign.

Thus, a negative entry appears when the modeled grid cell shows a different tendency compared to the measured land surface temperature. The entry is positive when the grid cell tendencies are in correspondence. To avoid the results canceling out when summed, the eight single results are increased by one. Hence, for full correspondence the sum of the elements of the search raster yields 16 but it is zero for full disagreement. Finally, the sum is scaled between zero and one. A P of 1, therefore, indicates full agreement of the patterns, i.e., no dissimilarity.

The scaling assures comparability with other error measures such as the Nash-Sutcliffe Efficiency or correlation coefficient. A pattern similarity of 0 not only corresponds to full dissimilarity but means that the two patterns are inverse to each other. A P of 0.5 indicates randomly diverging patterns.

The 3×3 local search window is applied to every cell (i,j) within the domain Ω and all time steps t of the patterns under comparison. The overall pattern similarity is then calculated as the mean of the single values (see Equation (7.13a)).

Numerical tests showed that a combination of the pattern similarity criterion with another bias resistant criterion, i.e., the Pearson correlation coefficient, results in the best model performances regarding streamflow and land surface temperature. The Pearson correlation coefficient is the fourth error measure ϕ_4. It is calculated as the correlation between the vectorized T_s and $\widehat{T}_s$ fields over all time steps within a catchment. It can be interpreted as the temporal mean of the spatial correlation of T_s and $\widehat{T}_s$ fields at every time step.

The criteria for pattern similarity ϕ_3 and ϕ_4 are equally weighted ($w_3 = w_4 = 0.5$) in the objective function if applied to calibration against the land surface temperature (only ϕ_3 and ϕ_4 are considered in the objective function).

The calibration with respect to a combination of land surface temperature and streamflow data were conducted using all four error measures ϕ_1, ϕ_2, ϕ_3, and ϕ_4 as objectives. The weights are defined as $w_1 = w_2 = \frac{1}{3}$ and $w_3 = w_4 = \frac{1}{6}$. The higher weighting of the streamflow error measures was chosen to ensure a correct partitioning of water in the hydrological system. In comparison with other weighting schemes this setup has proven to perform best.

Some of the objectives, e.g., the correlation, were varying within a very small range when compared to their maximal ranges (e.g., [-1,..,1]) if the model is calibrated. The objectives ϕ_i $(i = 1, ..4)$ were normalized by their potential ranges to avoid the dominance of any objective:

$$\phi_i = \frac{\phi_i - \phi_i^{\min}}{\phi_i^{\max} - \phi_i^{\min}} \tag{7.15}$$

Table 7.1 The applied ranges for normalizing the error measures.

	$\phi_i^{\min}$	$\phi_i^{\max}$
ϕ_1	-0.99	0.9
ϕ_2	-0.99	0.9
ϕ_3	0.64	0.66
ϕ_4	0.93	0.97

where $\min$ and $\max$ denote the upper and the lower bounds of the particular objective i, respectively. $\phi_i^{\min}$ and $\phi_i^{\max}$ were determined based on 55 000 simulations in two of the basins under investigation, i.e., Ems and Neckar, using random parameters. To ensure sampling over the entire parameter domain a stratified sampling strategy was applied to generate 55 000 parameter sets (*Morris*, 1991). Empirical ranges were determined for the maxima and minima of the objective functions. We reviewed 20 calibrations in each of the six catchments under investigation and assessed the minimum and maximum ranges of each objective. The adopted ranges of each error measure are given in Table 7.1.

7.4.4 Experimental Design

To address three different hypotheses, several numerical experiments were conducted, which will be explained within this section. In common to all of the experiments were the six basins under investigation depicted in Figure 7.1. All the experiments were designed as ensemble simulations based on 20 independent parameter optimization runs to address the issue of parameter estimation uncertainty. The standard calibration of hydrologic models with streamflow served as a reference or baseline scenario. All the model calibrations were conducted using the Dynamically Dimensioned Search algorithm (*Tolson and Shoemaker*, 2007) and employing a budget of 1000 iterations.

We calibrated the hydrologic model mHM using land surface temperature alone to assess its streamflow prediction performance. The resulting model performances were compared to simulations based on 1000 random parameter samples that were derived using Latin hypercube sampling (*McKay et al.*, 1979). The determined parameter sets were transferred to 222 uncalibrated locations to assess their validity and stability.

The aim of the second experiment was to assess the impact of a combined calibration of streamflow Q and land surface temperature T_s. It is expected that this approach had a high impact on the modeled evapotranspiration because the land surface temperature characterizes the near-surface atmospheric conditions and is directly connected to E via Equation 7.7.

We hypothesize that calibrating the hydrologic model mHM using land surface temperature and streamflow will lead to a better constraint of parameters without deteriorating the model performance regarding streamflow significantly. The degrees of the parameter constraints were determined by analyzing the final parameter values of the 20 independent model calibrations. To avoid influential effects of outliers, the spread in the ensemble parameters was determined using the difference between the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles, r_5 and r_{95}:

$$R_r^i = 1 - \frac{r_{95}^i - r_5^i}{r_{max}^i - r_{min}^i}, i = 1, \ldots, n \tag{7.16}$$

where r_{min} and r_{max} denote the lower and upper limits of the initial parameter range of each parameter i. If the parameter range reduction R_r equals 1, the parameter range converges to a single value in the independent runs. If R_r, equals 0 the optimized parameter is spread over the entire initial range and therefore is not constrained during calibration.

In addition, we hypothesize that the spatial variability of the resulting evapotranspiration fields will decline. The spatial variability of the modeled E was estimated using the signal-to-noise ratio (S), which is defined as

$$S_E(t) = \frac{\mu_E(t)}{\sigma_E(t)} \tag{7.17}$$

where μ denotes the mean and σ the standard deviation of an evapotranspiration field E at a particular time step t. We will present an evaluation of the evapotranspiration estimates at the local scale based on eddy flux data. This comparisons at a different model resolution, i.e., 100×100 m^2, follows the methodology presented in *Zink et al.* (2017).

The calibration period was limited to the year 2009 due to the availability of streamflow and land surface temperature observations. All the simulations had a model spin up period of 5 years. The majority of the analyses focused on the year 2009, but the simulations within the 222 basins were based on the entire observational time series. The separation of the time series into calibration and validation period was unnecessary because this study focused on the benefit of using T_s for parameter inference compared to classical Q calibration.

7.5 Results and Discussion

7.5.1 The Performance of Land Surface Temperature Calibrations Regarding Streamflow

In this section, we will present results obtained by calibrating the hydrologic model mHM with patterns of land surface temperature T_s alone using the objective functions ϕ_3 and ϕ_4 as described in section 7.4.3 in order to assess the performance of land surface temperature calibrations regarding streamflow.

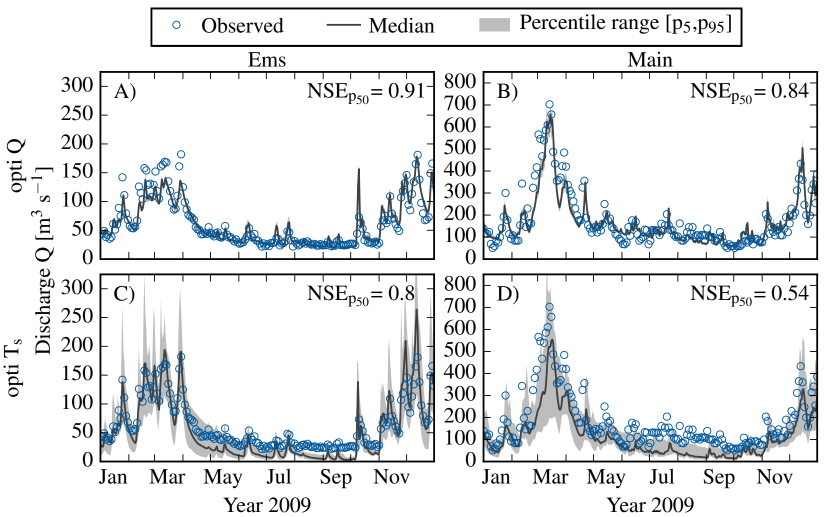

Figure 7.3 Simulated daily streamflow when calibrating the hydrologic model mHM with streamflow Q (panels A and B) and with land surface temperature T_s (panels C and D) for the basins Ems (panels A and C) and Main (panels B and D). The gray bands depict the uncertainty in the 20 ensemble model simulations as assessed by the range of the 5$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ percentiles of the estimated streamflow. The black line is the median of the ensemble streamflow simulations. Its performance $\mathrm{NSE}_{\mathrm{P50}}$ is given in the top right corner of each panel.

An ensemble of 20 parameter sets, that are calibrated individually in each basin is used for a forward run to predict streamflow. Panels C and D of Figure 7.3 exemplarily show the observed and simulated streamflow time series of the two basins Ems and Main out of the six basins under investigation (Figure 7.1). For comparison, panels A and B of Figure 7.3 depict streamflow predictions obtained by 20 independent classical calibrations with streamflow. The median streamflow estimated from the T_s calibrations shows an unexpected good mapping of the observed streamflow, revealing NSEs of 0.8 and 0.54 for the Ems and Main basins, respectively. As seen, the performance of the median NSE decreases when mHM is calibrated with T_s (Figure 7.3 panels C and D). Especially low flow periods are usually underestimated (July to September).

This underestimation results from insufficient estimated slow interflow and baseflow. These hydrologic processes are insufficiently modeled because T_s is non-informative regarding them. This means that parameters that are connected to slow interflow and baseflow are insensitive to a calibrations using land surface temperature. Similar studies that used T_s for model calibration limited the number of calibrated parameters to those connected to soil water storage and evapotranspiration (*Corbari and Mancini*, 2014; *Corbari et al.*, 2015; *Crow et al.*, 2003; *Gutmann and Small*, 2010; *Silvestro et al.*, 2013, 2015). In those studies, all other parameters are determined by prior knowledge, e.g., transfers from remote locations or expert knowledge.

The uncertainties arising from the parameter estimation process are depicted as gray bands in Figure 7.3. The streamflow uncertainty increases for the T_s calibration compared to the classical calibration with streamflow. In the case of the Ems river basin, the highflows of the flood event in spring 2009 are within the uncertainty bands for the T_s calibration, which was not the case for the Q calibration. In contrast, some of the parameter sets from T_s calibration performed very poorly when estimating flood events, e.g., in spring in the Main basin (Figure 7.3 panel D). The high uncertainty in the streamflow simulations is reasoned in the weak estimation of interflow and routing parameters when the model is calibrated with land surface temperature (see section 7.5.2). The T_s calibration approach shows stronger pronounced flood peaks as compared to Q calibration. This indicates that the direct runoff and fast interflow component are enabled more rapidly when compared to the calibration with streamflow. The uncertainty ranges are, however, reasonable given that streamflow itself was not involved in the model calibrations.

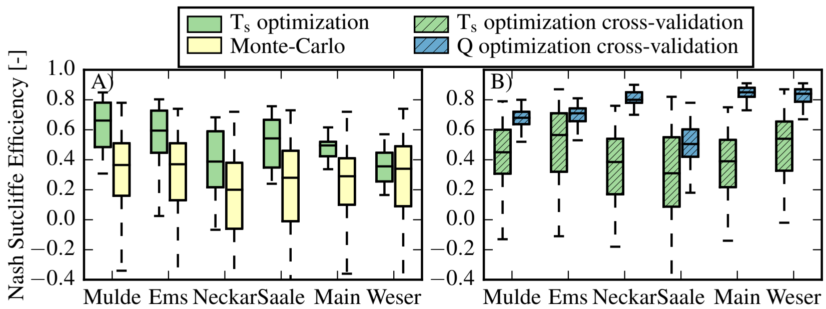

Figure 7.4 A) shows the model performance regarding streamflow when the model is either calibrated with land surface temperature (T_s) or driven by parameters using Latin hypercube sampling (Monte-Carlo). For the calibration with T_s, 20 independent calibration runs are performed, whereas 1000 sampled parameter sets are used for the Monto-Carlo simulations. B) depicts the model performance when calibrating mHM with either Q or T_s and transferring the parameters to the other basins. The variability, therefore, arises from the 100 parameter sets, which are derived at the remaining five different donor basins.

The median NSE of the 20 model calibrations with T_s varies between 0.36 and 0.66 for the six basins and is on average 0.51 (Figure 7.4 panel A). Note that these are the medians of the NSEs obtained from the 20 calibrated parameter sets, whereas the reported NSE in Figure 7.3 is calculated using the median streamflow time series. The median NSE of Monte-Carlo simulations using 1000 Latin hypercube sampled parameter sets is on average 0.20 lower. The lower median is caused by low model performances leading to insufficient NSEs falling below 0. These results indicate that the calibration based on T_s prevents poor model performances (NSEs $<$ -0.1) and results in streamflow performances which could be meaningful if no discharge data are available for calibration. The variabilities in the performance criteria obtained by T_s calibration are significantly lower (average standard deviation of 0.22) compared to those resulting from the Monte-Carlo simulations (average standard deviation of 0.39). Please note, that the streamflow simulation performance of both experiments, T_s calibration and Monto-Carlo simulations (Figure 7.3 A), are significantly below calibrations using streamflow (shown in Figure 7.11 panel B). The average median and standard deviation of these calibrations are 0.87 and 0.01, respectively.

The NSE uncertainties of the T_s calibrations increased (average standard deviation of 0.26) if the model parameters are transferred to remote locations (Figure 7.4 panel B). One reason is the five times higher number of ensemble simulations using 100 parameter sets, consisting of 20 parameter sets from each of the other five basins. Another reason is that some transferred parameters were not well adjusted for transfer to another location because different hydrologic processes are important in distinct basins. The Neckar basin, for example, has a significant groundwater contribution to the runoff process due to the karstic nature of the subsurface. Such processes will play a minor role in the Ems basin, for example, which is mainly located on a ground moraine. Hence, some subsurface parameters are not well constrained in the Ems basin and will lead to an insufficient representation of karstic processes in the Neckar basin. Nevertheless, the median NSEs are comparable to the on-site calibrations (Figure 7.4 panel A) which confirms the transferability and stability of the inferred parameters.

The comparison of transferred parameters obtained by Q calibration with those acquired by T_s calibration show an average deterioration of the latter by 39% (Figure 7.4 panel B). This behavior was expected because a cross-validation of land surface temperature inferred model parameters with streamflow cannot outperform a calibration employing Q. The best results of the T_s calibrations (upper edges of the box-plots) are, however, at least as good as the median model performances of the Q calibrations for most of the basins.

The 120 optimized parameter sets (20 from each basin) are transferred to 222 additional basins to assess their ability to reproduce streamflow observations (Figure 7.5). Some of those basins are sub-basins of the six donor basins. This cross-validation experiment assesses the stability and validity of the derived parameters since the basins were not involved in the parameter inference process (*Klemeš*, 1986). The average median model performance is 0.4 for the basins (Figure 7.5 panels A and B). The median NSE of 0.5 is exceeded by 45% of all basins when considering daily streamflow. On a monthly basis, the average median NSE increases to 0.61 and the number of catchments with NSE$\geq$0.5 increases to 77% (Figure 7.5 panels D and E). These results further support the hypothesis that land surface temperature can inform the model calibration if no streamflow data is at hand. The reported NSEs are significantly lower compared to NSEs originating from parameter sets obtained by streamflow calibrations as presented by *Zink et al.* (2017). Please note the remarkable uncertainty in Figure 7.5 panels D and E. These uncertainties differ substantially for daily and monthly streamflow and show significant higher values than those derived by calibrations using streamflow data (*Zink et al.*, 2017).

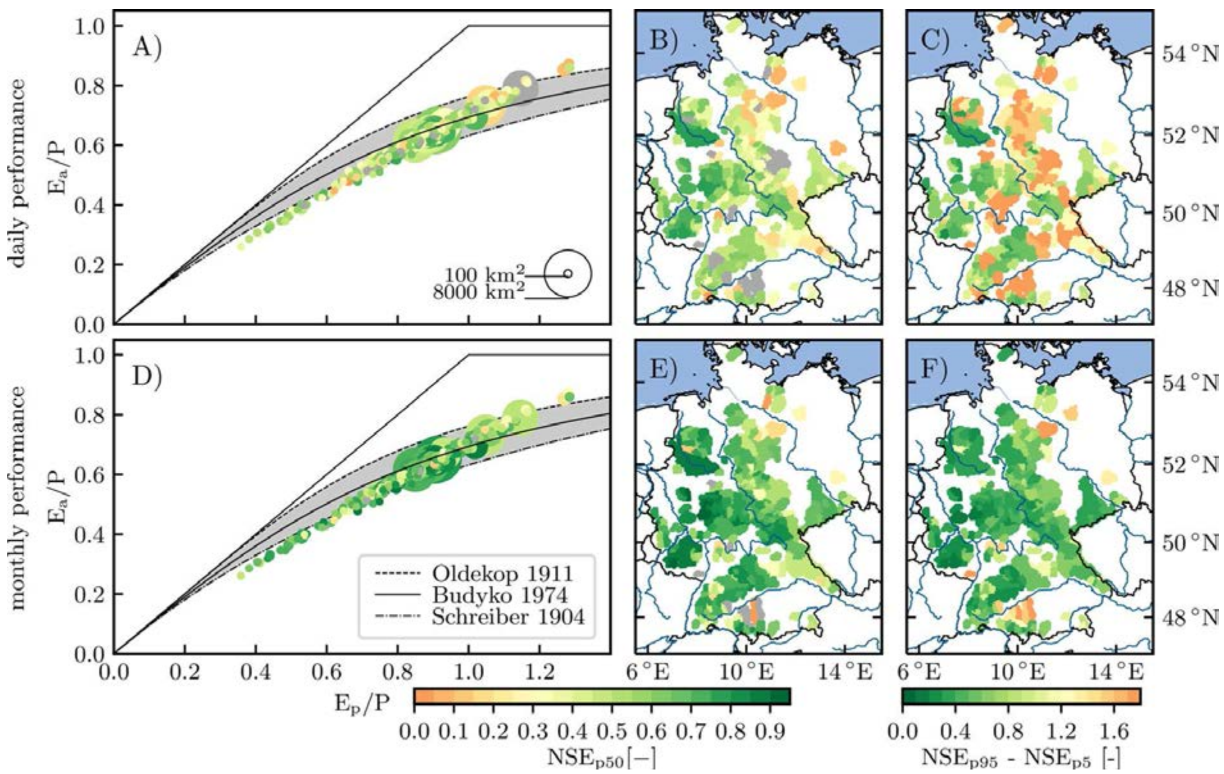

Figure 7.5 Budyko plot and performance maps for 120 parameter sets of the six donor basins (Figure 7.1) at 222 basins spread over Germany. The parameter sets are based on calibrations using land surface temperature alone. The upper row depicts evaluations based on daily values (panels A, B, and C), whereas the lower row depicts monthly streamflow evaluations (panels D, E, and F). In the first column the basins are presented as Budyko plots (panels A and D), which are color-coded based on the ensemble median NSE for daily (panel A) and monthly (panel D) streamflow values. The gray band envelops different estimations of the Budyko curve (*Budyko*, 1974; *Ol'dekop*, 1911; *Schreiber*, 1904). A separation to energy- ($E_p/P < 1$) and water-limited basins ($E_p/P > 1$) can be made based on the x-axis. The center column depicts the location of the 222 basins shown in the Bydyko plots using the same color code (panels B and E). The right column shows the range of the 5^{th} and 95^{th} ensemble percentiles for the NSE on daily (panel C) and monthly (panel F) basis. Panels A, B, D, and E share the left color bar, and panels C and F share the right color bar. The simulation period is adopted according to the available streamflow observations but is at least 10 years (average=42 years).

Using satellite-derived land surface temperature for calibrating hydrologic models is consistent with efforts to predict streamflow in ungauged basins (*Hrachowitz et al.*, 2013; *Sivapalan et al.*, 2003). In particular, the results of the six study basins (Figure 7.4) show that T_s could be worth considering if no discharge data are available.

Corbari and Mancini (2014) found similar results for the calibration of a distributed "Energy-Water Balance" model. In their study, the calibration with T_s did not outperform the streamflow estimation employing the standard parameterization of Energy-Water Balance model. *Silvestro et al.* (2015) also found that a land surface temperature calibration lead to performance losses if compared to streamflow calibrations. The deteriorations shown in those studies are lower than those mentioned above. However, that can be attributed to the calibration procedures employed. *Corbari and Mancini* (2014) and *Silvestro et al.* (2015) restricted the number of parameters to be calibrated only to those connected to soil moisture and evapotranspiration. The remaining parameters are estimated from prior knowledge under the assumption that they are insensitive to T_s. To restrict the calibration to a subset of parameters is a good idea in general but implies a risk. The parameters that remain for calibration may be insensitive or have low sensitivities with regard to T_s if compared to a full parameter calibration (see, e.g., *Cuntz et al.*, 2015). Hence, a proper sensitivity analysis or parameter screening should be performed before excluding parameters from calibration. In this study, all the parameters are therefore purposed for optimization, and parameters that are insensitive to T_s but are important for streamflow prediction may have been included.

7.5.2 Calibration of mHM with Streamflow and Land Surface Temperature

This section addresses the question wether a combined calibration of land surface temperature and river streamflow can improve the model parameter identifiability. We analyze the impact of such an approach on the spatial variability of evapotranspiration, estimation of streamflow, and estimation of evapotranspiration at eddy flux towers.

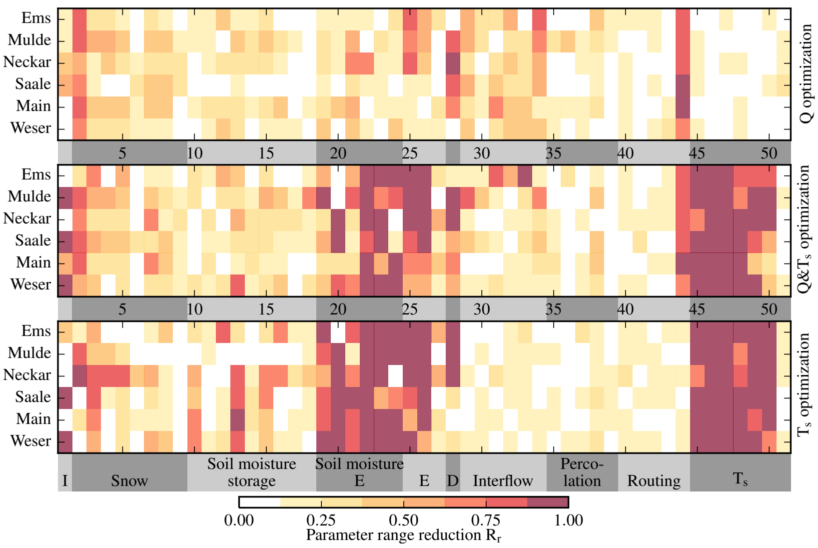

Figure 7.6 The parameter range reduction R_r (Equation ref) is shown. The panels show wether mHM was calibrated against streamflow (upper row), streamflow and land surface temperature (middle row), or land surface temperature (lower row). The parameter range reduction is assessed by scaling the range of a particular parameter resulting from 20 independent calibration runs with the initial parameter range (see Equation 7.16). A low value (light yellow) indicates a small range reduction, whereas a high value (dark red) indicates a well-constrained parameter. The parameters are grouped according to their appearance in the different model processes. Abbreviations: I - interception, D - direct runoff, E - evapotranspiration, T_s - land surface temperature.

Identifiability of Model Parameters: Parameter Range Reduction The parameter range reductions for the calibration of the models with a) streamflow alone, b) streamflow and land surface temperature, and c) land surface temperature alone are determined according to Equation 7.16. One hypothesis of this study is that adding a diagnostic land surface temperature model to an exiting hydrologic model helps to better constrain certain model parameters. As indicated in section 7.4.3, only the patterns of T_s were involved in model calibration of mHM through the pattern similarity criterion (Equation 7.13), which uses the nominal values only indirectly. The spreads of the ensemble parameters normalized using their initial ranges are shown in Figure 7.6. Dark red colors characterize well-constrained model parameters, whereas light yellow colors identify parameters that are almost randomly drawn from their initial ranges and are thus uninformed by the data. The figure gives some indications of the identifiability and hence the sensitivity of the parameters regarding the various variables used for calibration (Q, T_s, or both).

The interpretation of the sensitivity is analogous to the parameter range reduction - if the parameter range is reduced, the particular parameter is sensitive with respect to the individual variable.

The most obvious difference between the three optimization strategies can be observed in the group of soil moisture evapotranspiration parameters (p_{19} to p_{24}). These parameters primarily govern water extraction from the soil due to evapotranspiration. They are constrained best if calibrated with T_s (Figure 7.6 bottom panel). The ranges also narrow significantly when mutually calibrated with Q and T_s (Figure 7.6 center panel). Two out of the three evapotranspiration parameters (parameters 25 and 26) show a similar behaviors when T_s is involved in the calibration. This confirms that using patterns of satellite-derived land surface temperature for parameter optimization helps better constrain model parameters, especially those connected to evapotranspiration.

The results shown in Figure 7.6 also indicate that using T_s only for parameter optimization may not be sufficient because some parameters are not well constrained. The snow threshold temperature (parameter 2), for example, is not as well constrained when Q is not considered in the model calibration in most of the basins. The snow threshold temperature parameter defines the aggregate state of precipitation. If the air temperature is below that threshold, precipitation is treated as snow, and otherwise, it is considered as rain. T_s is a bad estimator for snow threshold temperature because that parameter is only important in winter. During the cold season, evapotranspiration is low, and as a consequence, the impact of T_s on the modeled water fluxes is also low. The Neckar basin, is an exception; for which the snow threshold temperature is well constrained if calibrated with T_s.

The fifth routing parameter, i.e., parameter 44, is almost insensitive to T_s (Figure 7.6 bottom panel). Parameter 44 is the dominating parameter for routing water through the model domain. Further, the interflow parameters (parameters 31-34) show lower range reductions when compared to parameter optimizations that included the Q observations. These insensitivities explain the mismatches in low flows observed in Figure 7.3. Moreover, the strongly pronounced peaks in Figure 7.6 are reasoned in the weak estimation of the threshold for activating/deactivating of the fast inflow process (parameter 29 in Figure 7.6).

The high parameter range reductions of the parameters 45-50 confirm the proper implementation and parameterization of the diagnostic land surface temperature module. This is an important aspect because increasing the number of model parameters due to the implementation of a new process should not lead to a distraction of the optimization algorithm caused by those parameters.

Parameter 51, which is one of the newly introduced T_s parameters, characterizes a bias correction parameter for T_s. That parameter was implemented during the investigation of different objective functions, in which we also tested bias sensitive error measures, e.g., NSE or SSE (results not shown). The fact that the parameter is not well constrained underpins the conclusion that the pattern similarity criterion is bias insensitive.

These results confirm the hypothesis that the consideration of spatially distributed, satellite-retrieved land surface temperature fields next to streamflow improves the identifiability of parameters of the hydrologic model mHM.

Spatial Patterns of Land Surface Temperature and Evapotranspiration Here, the aforementioned higher constraint in evapotranspiration related parameters among others using a simultaneous calibration with streamflow and land surface temperature is analyzed regarding its effect on the spatial variability of land surface temperature and evapotranspiration.

Table 7.2 Improvement of Spearman rank correlations for streamflow and land surface temperature calibration compared to streamflow only calibrations for every catchment and season season of the year 2009.

Improvement [%]	Winter	Spring	Summer	Autumn
Mulde	3.09	3.27	3.63	1.38
Ems	2.4	3.45	4.04	-1.63
Neckar	0.63	2.47	2.73	0.27
Saale	3.13	2.24	3.9	0.64
Main	3.11	3.64	6.29	1.75
Weser	1.42	5.24	5.61	1.8
average	2.14	3.41	4.51	0.57

A pattern analysis, independent from the Pearson correlation coefficient and the pattern similarity criterion, is based on the Spearman rank correlation coefficient. This metric is estimated between satellite retrieved and simulated land surface temperature. Table 7.2 reports the differences of the Spearman rank coefficients in terms of improvement using Q and T_s compared to streamflow only calibrations. The average improvement over the median of the six basins is approximately 3%. The main improvements of the spatial patterns of T_s are observed in spring and summer as shown in Table 7.2.

The maximum improvement of approximately 6% in the Main basin in summer can be understood from the structure of the hydrologic model mHM. Most of the model internal, effective parameters are based on physiographic input data, e.g., the soil textural properties. Hence, the optimization of the *global* parameters using land surface temperature additionally to streamflow can only improve spatial patterns in a limited range. Nevertheless, the impact on the spatial distribution of evapotranspiration can be significant, e.g., if evapotranspiration simulations improve in the same order of magnitude. A boxplot of the annual average Spearman rank correlations for both calibrations, i.e., Q only and $Q\&T_s$, can be found in the supplemental material.

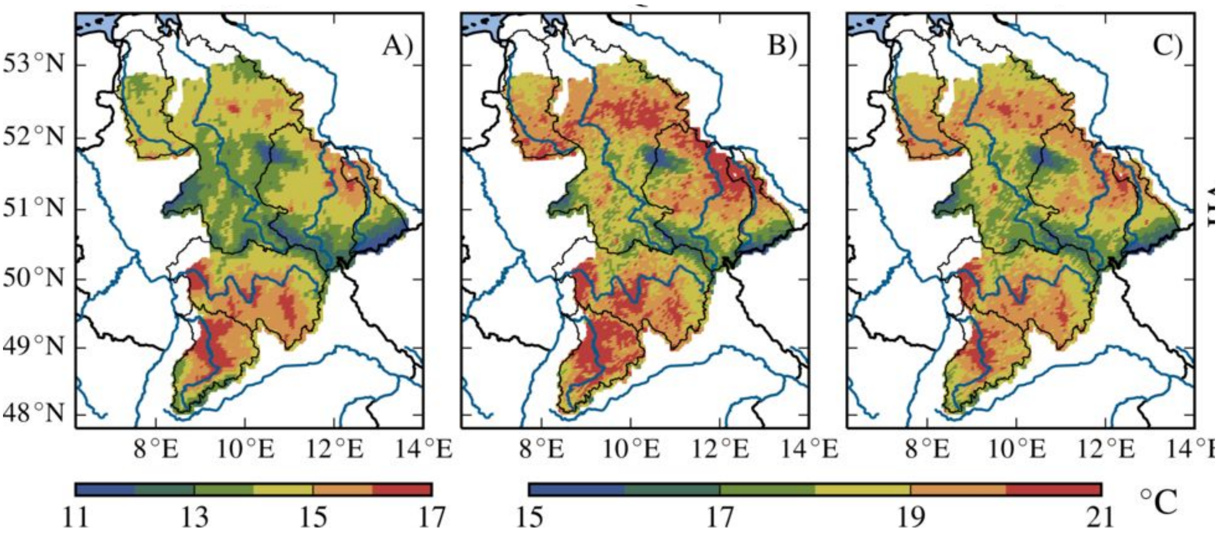

Figure 7.7 Comparison of A) the satellite retrieved land surface temperature (T_s) and simulated land surface temperatures obtained by B) streamflow only (Q) and C) simultaneous ($Q\&T_s$) calibrations. The plot shows the average land surface temperature in summer 2009 (June, July, August - JJA) for all six basins under investigation (see Figure 7.1). Observations and simulations refer to different color bars because of an inherent bias of the satellite retrieved land surface temperature (see section 7.3.1). Please note that the land surface temperature is derived with the best objective function value of each basin.

Figure 7.7 illustrates the patterns of satellite retrieved and simulated land surface temperatures for summer 2009. The visual inspection of this figure shows that the patterns obtained by the simultaneous calibration with land surface temperature and streamflow have a lower spatial variability than those using the Q only calibration. The resulting simulated T_s pattern is closer to the observed pattern but is not identical. The pattern matching, i.e., the Spearman rank coefficient, between panels A and C is approximately 5% larger than that between panels A and B. Plots of the remaining seasons can be found in the supplemental material.

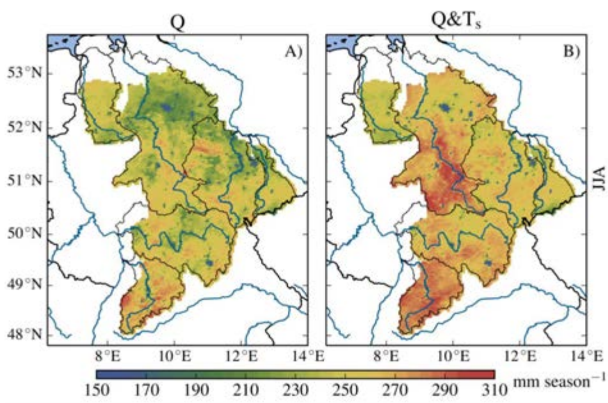

Figure 7.8 Simulated evapotranspiration for model calibrations based on A) streamflow (Q) or B) streamflow and land surface temperature simultaneously ($Q\&T_s$). The plot shows the sum of evapotranspiration in summer 2009 (June, July, August - JJA) for the six basins (see Figure 7.1). Please note that the shown evapotranspiration is derived with the best objective function value of each basin.

Figure 7.8 shows the evapotranspiration in the summer of 2009 in all basins under investigation. Panel A displays the results of an optimization with streamflow alone, and panel B shows the results of the calibration with streamflow and land surface temperature. The pattern of the evapotranspiration when calibrated with streamflow shows higher spatial variability in comparison to the simultaneous Q and T_s calibration. The mean E is approximately 6% higher if the model is calibrated with Q and T_s. The average evapotranspirations for the Q versus Q and T_S calibration are 245 mm season^{-1} and 259 mm season^{-1} in the summer of 2009, respectively. The locations with evapotranspiration values less than 160 mm season^{-1} are sealed areas. Plots of the remaining seasons can be found in the supplemental material. The visual comparison in Figure 7.8 supports the hypothesis that the spatial field of evapotranspiration has a higher spatial variability if the model is optimized with streamflow only.

The spatial variability of the evapotranspiration decreases if land surface temperature, which carries some information about the spatial distribution, is included in the calibration process. *McCabe et al.* (2005) also found that T_s had an effect on the spatial variability of evapotranspiration but did not quantify it.

In this study, we quantify the impact on the spatial fields of evapotranspiration using signal-to-noise ratio (S_E, see Equation 7.17). For the two example basins, Ems and Main, the smoothed signal-to-noise ratio over the course of 2009 is shown in Figure 7.9. The S_E is higher for the calibration with streamflow and land surface temperature. This means the fields of E are smoother and have lower spatial variability than those obtained by calibration with streamflow. This smoothing is not only caused by T_s but to a significant extent by the air temperature as well, which is a very sensitive variable in Equation 7.7. Air temperature has very high spatial covariance, i.e., low spatial variability, compared to, e.g., precipitation, which propagates to evapotranspiration in the proposed framework.

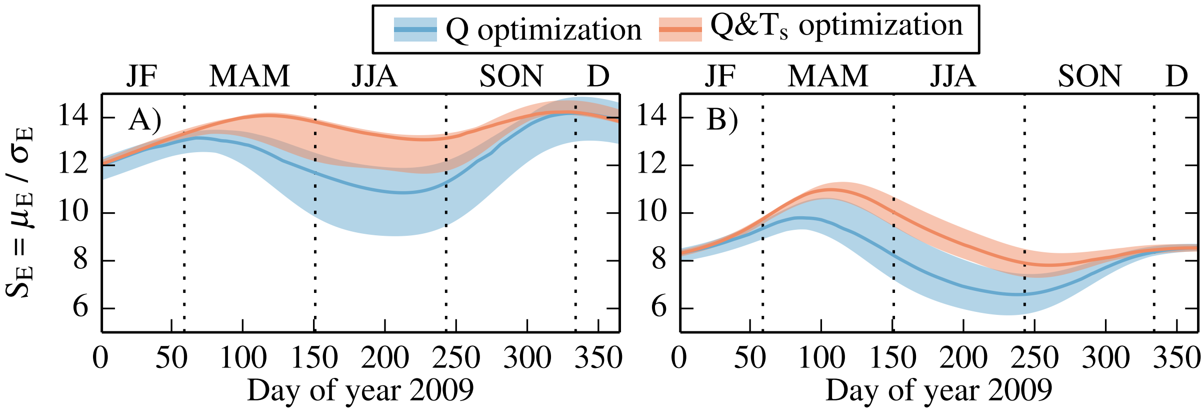

Figure 7.9 Kernel-smoothed signal-to-noise ratio (S, Equation 7.17) of evapotranspiration fields for the A) Ems and B) Main basins. Low values characterize noisy fields, whereas high values describe spatially smooth patterns. The uncertainty bands depict the difference between the 5^{th} and 95^{th} percentiles of the signal-to-noise ratio of the 20 on-site calibrated parameter sets. The solid line represents the 50^{th} percentile of the respective basin. The red band/line marks calibrations using land surface temperature (T_s) and streamflow (Q) simultaneously, whereas blue indicates calibrations using streamflow alone. Please note that the y-axes do not start at 0.

A significant impact of the calibration procedure on simulated evapotranspiration can only be observed between April and September (Figure 7.9). During winter, evapotranspiration is very low and thus uncertain model parameters do not have a significant effect on either the magnitude or the spatial variability of E. The uncertainty of the modeled E is, however, low during winter.

Figure 7.10 panel A presents the average signal-to-noise ratios for 2009 for the six basins and panel B presents the streamflow NSE. The signal to noise ratio S_E is higher for all of the Q-T_s calibrations, as panel A shows. Furthermore, the spreads are smaller compared to the Q calibration for all the basins, with exception of the Mulde. Figures 7.9 and 7.10 confirm that the spatial variability of evapotranspiration is reduced if land surface temperature is considered during model calibration.

Streamflow Simulations Finding a compromise solution for optimizing hydrologic models with Q and T_s should not deteriorate the streamflow simulation significantly. Figure 7.10 panel B shows that for four out of the six basins this condition is fulfilled. For the Neckar and Weser basins the streamflows deteriorate significantly by more than 5% for the simultaneous calibrations. A possible reason could be the weighting scheme between the two objectives, Q and T_s. For some basins, it may be necessary to increase the weighting of the objective function considering streamflow, i.e., ϕ_1 and ϕ_2 in Equation 7.12. *Crow et al.* (2003) studied the effects of weighting Q and T_s differently and found that the model performance differs based on the chosen weighting scheme. The herein proposed weighting was determined during precedent tests for the Ems and Neckar basin and showed good results for these test cases. Generally, the weighting of different objectives can be argued in one way or another. Ideally, a Pareto optimization would provide closer insight to the offset between both objectives. However, the decision regarding which objective should be preferred remains a subjective choice. The NSE does not improve by assimilating T_s in any of the basins, with exception of the Mulde (Figure 7.10 panel B). On average, the median streamflow performance deteriorates by 6% if T_s is considered in the calibration. A range of -11% to 14% in performance dif-

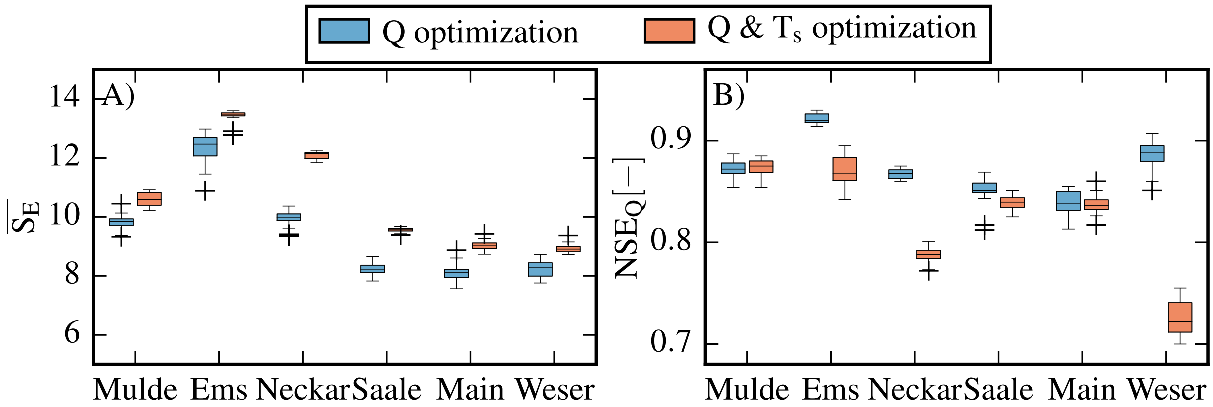

Figure 7.10 Comparison of the optimization strategies using Q only or Q and T_s regarding A) the average spatial variability of evapotranspiration (here S_E) and B) Nash-Sutcliffe performance criterion (NSE) of daily streamflow (panel B). The results shown in panel A are estimated by averaging the daily signal-to-noise ratios of 2009 (e.g., Figure 7.9). Panel B shows the performance of the simulated streamflow of 2009. The spread of the values (uncertainty) stems from the 20 independent parameter estimations. In both panels, large values indicate better performance. Please note that the y-axes in both panels do not start at 0.

ferences was found by *Corbari and Mancini* (2014) if the model is calibrated by T_s and Q simultaneously, which is comparable to the findings of this study.

Evapotranspiration at Eddy Flux Towers Figure 7.11 compares the performances of simulated evapotranspiration, determined on a $100 \times 100\ m^2$ spatial resolution, using the classical streamflow calibration with the combined Q-T_s calibration. The Pearson correlation coefficient between the observations and simulation is increasing when mHM is calibrated with both Q and T_s (Figure 7.11 panel A). The medians of the correlation and RMSE improve by 5% and 8%, respectively. Major improvements are achieved in summer when the evapotranspiration is highest; for example, the median correlation coefficient in summer improves from 0.36 to 0.67 at station E3.

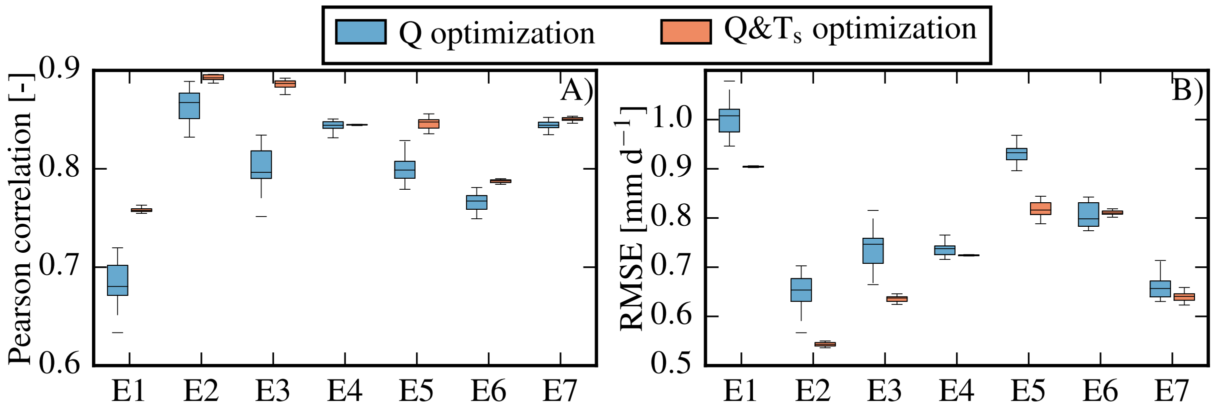

Figure 7.11 Evaluation of the evapotranspiration (E) estimates at 7 eddy flux towers (Figure 7.1). A) shows the Pearson correlation coefficient and B) shows the root mean square error (RMSE) between the flux tower observations and model simulations using 20 parameter sets inferred by either calibration with streamflow Q (blue) or streamflow and land surface temperature $Q\&T_s$ (red). The observational periods of the flux towers range from 3 to 10 years, and, on average there are 6 years of data. Note that high Pearson correlation coefficients are beneficial, whereas the opposite holds for the RMSE. Please note that the y-axes do not start at 0.

Another important effect is the reduction in uncertainty of the evapotranspiration simulations (Figure 7.11). At some stations, the uncertainty bands are hardly visible for the Q-T_s calibration. This behavior can be directly attributed to the parameter range reduction (see section 7.5.2). The uncertainties in the E estimates have to decrease because the parameter estimation uncertainties of the parameters related to evapotranspiration decreased (parameters 19 to 27 in Figure 7.6). Please note that the eddy flux towers are unevenly distributed among catchments and their analyses are biased towards the Mulde and the Saale catchment because each of them are hosting 3 out of the 7 stations under investigation. Comparing the improvements in E estimation and the deterioration in Q simulation, it is difficult to draw a conclusion. We consider that a combined calibration with Q and T_s is beneficial based on the tradeoff of performances of the two major water balance variables E and Q and, moreover, the improved parameter identifiability.

7.6 Summary and Conclusions

The results of the study confirm that accounting only for spatial variabilities in land surface temperature in the parameter inference process results in model simulations whose efficiency could be meaningful if no streamflow data are available for calibration and evaluation. It is, however, a step forward towards predictions in ungauged basins. Land surface temperature data are broadly and freely available over the entire globe and thus represent a valuable source of information for hydrologic modeling.

A second finding of this study is that calibrating the hydrologic model mHM with Q and T_s leads to better-constrained model parameters, even if the implementation of the diagnostic land surface temperature module requires additional model parameters. In particular, parameters connected to evapotranspiration were better constrained when compared to streamflow only calibrations. This finding indicates that the classical calibration of hydrologic models can be improved by incorporating spatial information originating in satellite data.

The herein presented methodology is a step forward for considering such spatially distributed observations, even if they are inherently biased. The developed pattern similarity criterion is a first attempt to assess the spatial structures of spatially distributed observations.

Some limitations of this methodology were also observed. Model performance with regard to streamflow decreased despite the fact that the model parameters were better constrained. At the same time, the model performance regarding evapotranspiration increased at the seven eddy flux measurement sites.

Parameters connected to interflow and routing could not be sufficiently constrained if only T_s was considered in the calibration process. Further research must be performed to explore other sources of satellite data, which may overcome this discrepancy. GRACE data, for example, seem to be a promising alternative for assessing subsurface model parameters.

Research must also be dedicated for investigating new measures to incorporate either spatial or temporal information of satellite data. *Cloke and Pappenberger* (2008) and *Koch et al.* (2015) have already made some efforts in that direction, but the studies concerning bias-insensitive pattern matching criteria in hydrology remain rare.

Two-step calibration is another approach that could make better use of satellite information. However, a sensitivity analysis must be used initially to identify model parameters that are sensitive to the respective model variables, e.g., streamflow or land surface temperature. Based on that knowledge, the hydrologic model would be calibrated first with land surface temperature and second with streamflow by only considering the sensitive parameters for the respective variables.

Acknowledgments

We kindly acknowledge our data providers the German Meteorological Service (DWD), the Joint Research Center of the European Commission, the European Environmental Agency, the Federal Institute for Geosciences and Natural Resources (BGR), the Federal Agency for Cartography and Geodesesy (BKG), the European Water Archive, the Global Runoff data Centre, and the Land Surface Analysis Satellite Applications Facility (LSA SAF). Further, we acknowledge the projects EuroFlux (EU-FP4), CarboEuroFlux (EU-FP5), and CarboEuropeIP (EU-FP6), IMECC (EU-FP6) as well as Christian Bernhofer, Axel Don, Mathias Herbst, Alexander Knohl, Olaf Kolle, and Corinna Rebmann for the provision of eddy covariance data. This work was funded by the Helmholtz Alliance EDA - Remote Sensing and Earth System Dynamics, through the Initiative and Networking Fund of the Helmholtz Association, Germany. We thank Stefan Rimkus and two anonymous reviewers and the editors, John D. Bolten and Marc Bierkens, for their valuable comments, which helped us to improve the quality of this manuscript.

Apendix: General Solution of a Depressed Quartic Equation

Based on the energy balance equation (see Section 3.2, Equation 7.7), the following functional relationship for $\widehat{T_s}$ was found

$$(1-\alpha)Q_S^{\text{in}} + Q_L^{\text{in}} - \lambda E + \frac{\varrho_a c_p}{r_a} T_a - \frac{\varrho_a c_p}{r_a}\widehat{T_s} - \epsilon\sigma\widehat{T_s}^4 = 0. \tag{7.18}$$

The aim of this appendix is to find the real roots of Equation 7.18. For simplicity, it would be easier to rearrange this equation as

$$\widehat{T_s}^4 + q\widehat{T_s} + r = 0, \tag{7.19}$$

with

$$q = \frac{\varrho_a c_p}{\epsilon\sigma r_a} \tag{7.20}$$

$$r = -\frac{1}{\epsilon\sigma}\left((1-\alpha)Q_S^{in} + Q_L^{in} - \lambda E + \frac{\varrho_a c_p}{r_a} T_a\right) \tag{7.21}$$

Because the cubic term of the $\widehat{T_s}$ formulation (Equation 7.7) is absent, the resulting equation is called a depressed quartic equation having this general form:

$$x^4 + px^2 + qx + r = 0. \tag{7.22}$$

A depressed quartic equation can be solved explicitly by a method discovered by Ferrari in 1545 (*Cardano*, 1993). The first step of his method consists on rewriting the depressed quartic equation into two parts as follows

$$\left(x^2 + \frac{p}{2}\right)^2 = -qx - r + \frac{p^2}{4} \tag{7.23}$$

In the second step, an arbitrary variable m is introduced into the left-hand side so that the right-hand side can be factorized. To keep the equality, the corresponding terms are added to the right-hand side. The resulting equation is

$$\left(x^2 + \frac{p}{2} + m\right)^2 = 2mx^2 - qx + m^2 + mp + \frac{p^2}{4} - r. \tag{7.24}$$

Since m is an arbitrary factor, it can be selected so that one get a perfect square on the right-hand side (i.e., a single positive solution). A perfect square solution for a quadratic equation can be obtained when the discriminant of x on the right-hand side is equal to zero, in other words, when

$$(-q)^2 - 4(2m)(m^2 + mp + \frac{p^2}{4} - r) = 0. \tag{7.25}$$

By rearranging the discriminant, we obtain the "resolvent cubic of a quartic equation" (R):

$$R = 8m^3 + 8pm^2 + (2p^2 - 8r)m - q^2 = 0 \tag{7.26}$$

whose real roots can be found using Cardano's formula (*Cardano*, 1993) (see section "General Real Roots of the Resolvent" below). Assuming that we have found a solution for the resolvent such that $m \neq 0$, the right-hand side of Equation 7.24 can be factorized as follows

$$\left(x^2 + \frac{p}{2} + m\right)^2 = \left(x - \frac{q}{4m}\right)^2 \tag{7.27}$$

$$= \left(\sqrt{2m}x - \frac{q}{2\sqrt{2m}}\right)^2 \tag{7.28}$$

which can be rearranged and decomposed into two factors to obtain the following expression

$$\left(x^2 + \sqrt{2m}x + m + \frac{p}{2} - \frac{q}{2\sqrt{2m}}\right)\left(x^2 - \sqrt{2m}x + m + \frac{p}{2} + \frac{q}{2\sqrt{2m}}\right) = 0. \tag{7.29}$$

Finally, Ferrari's method solves the depressed quartic equation by applying the quadratic formula to every factor independently. Note that the solution is dependent on the real root m of the resolvent R:

$$x = \pm_1 \sqrt{\frac{m}{2}} \pm_2 \sqrt{-\left(\frac{p}{2} + \frac{m}{2} \pm_1 \frac{q}{2\sqrt{2m}}\right)}, \tag{7.30}$$

where $\pm_1$ and $\pm_2$ denote the four corresponding occurrences of + and -.

General Real Roots of the Resolvent

The real roots of the resolvent R can be found with Cardano's solution for a general cubic equation (*Cardano*, 1993) of the form

$$m^3 + am^2 + bm + c = 0. \tag{7.31}$$

Using this notation, the real root of this equation is given by the expression:

$$m = \sqrt[3]{-\frac{1}{2}t + \sqrt{D}} + \sqrt[3]{-\frac{1}{2}t - \sqrt{D}} - \frac{1}{3}a \tag{7.32}$$

with

$$D = \left(\frac{s}{3}\right)^3 + \left(\frac{t}{2}\right)^2 \tag{7.33}$$

and

$$s = \frac{3b - a^2}{3} \quad t = \frac{2a^3}{27} - \frac{ab}{3} + c \tag{7.34}$$

Solution of $\widehat{T_s}$

Since the quadratic term of the Equation 7.19 is missing ($p = 0$), the equation of the resolvent (7.26) can be simplified to a depressed cubic equation

$$m^3 - rm - \frac{1}{8}q^2 = 0. \tag{7.35}$$

whose real root m_0, based on Equation 7.32 (with $a = 0$, $b = -r$ and $c = -\frac{1}{8}q^2$, $s = b$, $t = c$), is

$$m_0 = \sqrt[3]{\frac{q^2}{16}\left(1 + \sqrt{1 - \frac{256}{27}\frac{r^3}{q^4}}\right)} + \sqrt[3]{\frac{q^2}{16}\left(1 - \sqrt{1 - \frac{256}{27}\frac{r^3}{q^4}}\right)} \tag{7.36}$$

This finally leads to the real solution of the quartic Equation (7.30) in the interval [0 K, 500 K] (which is the only root of interest for this study):

$$\widehat{T_s} = \sqrt{\frac{1}{2}}\left(-\sqrt{m_0} + \sqrt{\frac{q}{\sqrt{2m_0}} - m_0}\right). \tag{7.37}$$

PART II

FORECASTING AND PREDICTING DROUGHTS

CHAPTER 8

SEASONAL SOIL MOISTURE DROUGHT PREDICTION OVER EUROPE USING THE NORTH AMERICAN MULTI-MODEL ENSEMBLE (NMME)

S. THOBER, ..., AND L. SAMANIEGO[1]

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Thober, S., R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego (2015), Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME), Journal of Hydrometeorology, 16(6), 2329–2344

8.1 Abstract

Droughts diminish crop yields and can lead to severe socio-economic damages and humanitarian crisis (e.g., famine). Hydrologic predictions of soil moisture droughts several months in advance are needed to mitigate the impact of these extreme events. In this study, the performance of a seasonal hydrologic prediction system for soil moisture drought forecasting over Europe is investigated. The prediction system is based on meteorological forecasts of the North American Multi-Model Ensemble (NMME) that are used to drive the mesoscale Hydrologic Model (mHM). The skill of the NMME based forecasts is compared against those based on the Ensemble Streamflow Prediction (ESP) approach for the hindcast period of 1983-2009. The NNME based forecasts exhibit an Equitable Threat Score that is on average 69% higher than the ESP based ones at a six month lead time. Among the NMME based forecasts, the full ensemble outperforms the single best performing model CFSv2, as well as all subensembles. Subensembles, however, could be useful for operational forecasting because they are showing only minor performance losses (less than 1%), but at substantially reduced computational costs (up to 60%). Regardless of the employed forecasting approach, there is considerable variability in the forecasting skill ranging up to 40% in space and time. High skill is observed when forecasts are mainly determined by initial hydrologic conditions. In general, the NMME based seasonal forecasting system is well suited for a seamless drought prediction system as it outperforms ESP based forecasts consistently over the entire study domain at all lead times.

8.2 Introduction

Droughts appear worldwide and belong to the most devastating natural catastrophes. Droughts are defined as dry anomalies and occur in all compartments of the hydrological cycle (*Sheffield and Wood*, 2011) such as the atmosphere (meteorological drought), streamflow and groundwater (hydrological drought), and root zone soil moisture (agricultural drought). We focus here on agricultural droughts because they are able to reduce crop yields leading to substantial socio-economic damages. For example, the 2003 European drought has caused losses in the order of 13 bn EUR (*COPA-COGECA*, 2003), whereas in the U.S. it is estimated that droughts lead to damages of 10 bn USD on average per event (mainly agricultural but also others such as livestock, *Smith and Katz*, 2013; *Smith and Matthews*, 2015). In developing countries, droughts even threaten the livelihood of societies. The 2010-2011 drought in the Horn of Africa, for example, led to a severe humanitarian crisis affecting around 12 million people (*Dutra et al.*, 2013; *Relief*, 2011). Drought early warnings can help to mitigate the impact of these disasters several months in advance, but only if they are based on skillful seasonal forecasting systems.

State-of-the-art seasonal forecasting systems employ either dynamical or statistical frameworks to generate a drought forecast. Statistical frameworks, for example, use conditional distribution functions of observed historical datasets for drought prediction (*Shahrbanou Madadgar and Hamid Moradkhani*, 2013). Dynamical prediction systems represent the physics of the Earth system and typically constitute of Coupled General Circulation Models (CGCMs), which provide climate forecasts (CFs) of meteorological variables (e.g., precipitation and air temperature). These forecasts are then used to force a hydrological model that can reliably simulate the land surface components of the hydrological cycle such as root zone soil moisture (SM). Previous studies have assessed the forecast skill of experimental prediction systems for specific drought events (*Dutra et al.*, 2013; *Luo and Wood*, 2007) as well as for multi-decadal hindcast periods (*Mo and Lettenmaier*, 2014; *Mo et al.*, 2012b; *Shukla and Lettenmaier*, 2011; *Shukla et al.*, 2014; *Wang et al.*, 2011; *Yuan et al.*, 2011, 2013a,b, 2015). In these studies, the Ensemble Streamflow Prediction (ESP) approach is frequently used as a benchmark for representing climatological skill (*Day*, 1985). ESP is a statistical method that resamples meteorological forcings from a historic dataset to represent the forcing uncertainty under unknown future conditions. It has been used to discriminate between the impact of initial hydrologic conditions (IHCs) and that of CFs on hydrologic predictions (*Shukla and Lettenmaier*, 2011; *Shukla et al.*, 2013; *Wood and Lettenmaier*, 2008).

Previous studies indicate that SM predictability depends strongly on the region considered. For example, ESP based SM forecasts in the Western United States are as skillful as CF based ones while the latter only add value at one month lead time (*Mo et al.*, 2012b; *Shukla and Lettenmaier*, 2011). In contrast, the National Center for Environmental Prediction Climate Forecasting System (CFS) version one and two provide more skillful SM drought forecasts than ESP in the Central and Eastern United States up to six months lead time (*Yuan et al.*, 2013a). This might be related to stronger correspondence of drought to the El Niño-Southern Oscillation (ENSO) in these regions and thus a higher atmospheric predictability (*Mo*, 2011; *Mo and Lyon*, 2015). A similar finding has been observed by *Dutra et al.* (2013) for a hindcast of the 2010-2011 Horn of Africa drought using the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting systems S3 and S4. They reported high predictability for periods associated with a La Niña event and less predictability otherwise. Although such ENSO teleconnections are weaker in Europe, *Yuan et al.* (2015) observed that CGCM based drought forecasts exhibit higher skill than ESP based ones up to five months lead times over the Danube river basin. In that study, the authors employed the recent North American Multi-Model Ensemble (NMME) which comprises 71 realizations of a multi-institutional, multi-model ensemble of climate forecast models up to lead times of 9-10 months (*Kirtman et al.*, 2014). The spatio-temporal distribution of SM drought forecasting skill using NMME over Europe has, however, not yet been fully evaluated. A high forecasting skill irrespective of the location and lead time is a fundamental requirement for a seamless prediction system.

Few studies focused on drought predictability during particular drought phases such as the development, onset, and recovery. In one of these, *Mo* (2011) reported that drought recovery is more difficult to predict in the United States as it evolves on a shorter time scale than the development. *Yuan and Wood* (2013) reported that NMME models add skill to forecasts of meteorological drought onsets in tropical regions, but not in extra-tropical ones. In contrast to precipitation, SM drought predictability depends strongly on the IHCs (*Wood and Lettenmaier*, 2008), which are substantially drier during the recovery than during the development phase. This characteristic has not been exploited when investigating the impact of IHCs on SM forecasts.

Multi-model forecasting ensembles such as CFSv2, ECMWF S4, and NMME have ever-increasing ensemble sizes to provide a better estimate of model uncertainty. This implies that they also offer more than one meteorological forcing time series for assessment studies. Nonetheless, most assessment studies focus only on the grand ensemble mean (*Dutra et al.*, 2013; *Mo et al.*, 2012b; *Yuan et al.*, 2013a, 2015, among others). Few studies related the performance of the grand ensemble to that of individual models (*Mo and Lettenmaier*, 2014; *Yuan and Wood*, 2013). *Thober and Samaniego* (2014) recently showed that investigating subensembles, which do not take all realizations into account, has the potential to increase ensemble performance for reproducing extreme precipitation and temperature indices. Considering the fact that SM predictability is highly dependent on the quality of precipitation forecasts, subensembles could help either to increase the forecasting skill, or to reduce computational load for operational forecasts without loosing predictability.

Given the current knowledge regarding NMME based SM drought forecasts over Europe, four research questions constitute the main goal of this study. 1) Are NMME based drought forecasts more skillful than ESP based ones over larger parts of the European domain? 2) How is the drought forecasting skill distributed in space and time? 3) How skillful are subensembles in forecasting European droughts in comparison to single NMME models and the full ensemble? 4) How do IHCs impact drought forecasting skill during drought development and recovery?

To address these research questions, the mesoscale Hydrologic Model (mHM, *Kumar et al.*, 2013; *Samaniego et al.*, 2010a) is used to simulate SM for monthly NMME based precipitation and air temperature forecasts for the hindcast period of 1983-2009. These NMME based forecasts are contrasted against those based on the ESP approach, which serve as a benchmark in this study. The mHM derived SM forecasts are then transformed to a quantile based soil moisture index (SMI). The SMI lies in the interval [0,1] and a threshold of 0.2 is used to classify droughts. This cutoff implies that the lower 20% of SM states occurring in a given period (e.g., a month) are considered as drought. Reference SMI fields are created using the observation based E-OBS dataset (*Haylock et al.*, 2008) to assess the skill of the different forecasting approaches employing the Pearson correlation coefficient and the Equitable Threat Score (ETS).

8.3 Methods and Datasets

8.3.1 Climate Forecasts

The forecasting dataset used in this study incorporates realizations of eight global climate models from the North American Multi-Model Ensemble (NMME) with ensemble members varying between 6 and 24 per model (Table 8.1, see also *Kirtman et al.* (2014)). Monthly CFs of precipitation and air temperature are provided globally at a $1^\circ \times 1^\circ$ spatial resolution for lead times up to eight months. In total 101 realizations are used in this study available from the International Research Institute for Climate and Society. The performance of these models for soil moisture drought forecasts is analyzed for the overlapping hindcast period of 1983-2009. It has to be mentioned that not all of these models are participating within the NMME phase two real-time dataset (*NMME*, 2014). The analysis of the hindcast dataset in this study, however, provides the opportunity to investigate the performance of a large ensemble of seasonal climate model predictions in comparison to that of a simple statistical approach. The analysis is conducted over the European domain covering an area between 10°W-45°E and 35°N-55°N.

Table 8.1 Climate Forecasting models used in this study, Institute they are developed at, and ensemble members available (see *Kirtman et al.* (2014) for details).

Acronym	Model	Institute	Ensemble members
CCSM3	Community Climate System Model, Version 3	University of Miami, Rosenstiel School of Marine and Atmospheric Science	6
CM2p1	Climate model version 2.1	Geophysical Fluid Dynamics Laboratory	10
ECHAMA	ECHAM version 4.5 anomaly coupled	International Research Institute for Climate and Society	12
ECHAMD	ECHAM version 4.5 direct coupled	International Research Institute for Climate and Society	12
GEOS5	Goddard Earth Observing System Model version 5	National Aeronautics and Space Administration	12
CFSv1	Climate Forecasting System version 1	National Center for Environmental Prediction	15
CFSv2	Climate Forecasting System version 2	National Center for Environmental Prediction	24
CanCM3	Canadian Coupled Global Climate Model version 3	Canadian Meteorological Center	10

8.3.2 Construction of Soil Moisture Forecasts

The well-constrained mesoscale Hydrologic Model (mHM, *Kumar et al.*, 2013; *Samaniego et al.*, 2010a) is used here to generate gridded estimates of soil moisture (SM) fields over the study domain. mHM is a spatially explicit distributed hydrologic model in which hydrological processes are conceptualized similar to these of other existing large-scale models like the VIC (*Liang et al.*, 1996a) and the WaterGAP model (*Döll et al.*, 2003). It is driven by daily gridded fields of precipitation, air temperature, and potential evapotranspiration to simulate different components of the terrestrial hydrological system such as canopy interception, snow accumulation and melt, soil moisture and infiltration, runoff generation and evapotranspiration, deep percolation and base flow, and flood routing between grid cells. The model is open source (www.ufz.de/mhm) and readers interested in more details may refer to *Samaniego et al.* (2010a). To date, mHM has been successfully applied to several river basins in Germany, North America, and Europe (*Kumar et al.*, 2013,b; *Samaniego et al.*, 2010a, 2013, 2014). In this study, a similar model setup with respect to terrain, soil, and land cover characteristics as used by *Rakovec et al.* (2016a), who demonstrated the ability of mHM to adequately represent the spatio-temporal dynamics of runoff, evapotranspiration, soil moisture, and total water storage anomaly over a wide range of European river basins.

The reference monthly SM field is obtained by forcing mHM with the observation based gridded E-OBS dataset (v8.0, *Haylock et al.*, 2008) during the period 1950-2010. The E-OBS dataset is aggregated to 1° grid resolution to be compatible with the resolution of the North American Multi-Model Ensemble (NMME) dataset. This reference SM field is then used to represent initial hydrologic conditions (IHCs) at the beginning of each month during the hindcast period (1983-2009).

Furthermore, the E-OBS dataset is used to set up the NMME and ESP based forecasts. The Ensemble Streamflow Prediction (ESP) forecast ensemble is created by resampling the meteorological dataset (i.e., E-OBS) of the hindcast period for a given target month excluding the year of that month, which is similar to the approach of previous studies (*Day*, 1985; *Shukla et al.*, 2013; *Twedt et al.*, 1977; *Wood and Lettenmaier*, 2008, among others). In total, the ESP forecasting ensemble consists of 26 members. The spatio-temporal variability of the E-OBS dataset is employed to disaggregate NMME based monthly precipitation forecasts to their corresponding daily values using a multiplicative cascade approach (*Thober et al.*, 2014). This approach preserves the observed spatial patterns at the daily time scale as well as the monthly amount of the forecasted precipitation. Each monthly NMME forecast is stochastically disaggregated to an ensemble of 25 daily realizations, thus increasing the overall ensemble size to 2525 (= 101 × 25). The daily weights for disaggregating the monthly temperature forecasts are derived from the E-OBS dataset for a given target month. This procedure is similar to the rescaling technique used by *Yuan et al.* (2015). The rescaled temperature estimates are then also used to adjust potential evapotranspiration, which is calculated using the Hargreaves-Samani approach (*Hargreaves and Samani*, 1985). The daily mHM derived SM fields for both forecasting systems are then averaged to their monthly estimates. A representative SM field for a given NMME model realization is created by averaging the corresponding estimates derived from the 25 disaggregated meteorological forecasts because there is no significant variability among the latter fields as they are all forced with the same monthly precipitation and air temperature.

8.3.3 Calculation of Soil Moisture Index

The monthly soil moisture (SM) fields are converted into their respective quantiles using a non-parametric kernel density estimation method for the drought analysis. The kernel density $\hat{f}(x)$ is estimated by

$$\hat{f}(x) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{x - x_k}{h}\right) \tag{8.1}$$

for a given sample of n SM fractions $x_1, \ldots, x_n$, bandwidth h, and kernel function K. A Gaussian kernel is used in this study and the bandwidth h is estimated by an optimization against a cross-validation error estimate (see *Samaniego et al.* (2013) for details). The respective quantiles, hereafter denoted as soil moisture index (SMI), and the corresponding distribution functions are estimated for each grid cell and calendar month independently. This procedure removes the seasonality of simulated SM and allows the comparability of SMI across locations. A SMI threshold value of 0.2 is used here to identify drought events following previous studies (*Andreadis et al.*, 2005; *Samaniego et al.*, 2013; *Sheffield et al.*, 2012; *Vidal et al.*, 2010, among others).

The monthly SM estimates are converted to their respective standardized anomalies prior to the conversion of SM to SMI to ensure their comparability across different realizations, climate models, and forecasting methods (*Koster et al.*, 2009). The standardized anomalies are obtained by removing the seasonal mean and standard deviation. In this approach, the distribution function $\hat{f}$ is estimated only once using the reference SM anomalies. The forecasted SM anomalies are converted to SMI using this unique distribution function. This procedure provides a fair comparison between NNME and ESP based forecasts. In this study, no bias correction is applied to the NMME forecasts because the SMI calculation and the standardization of SM forecasts accounts for biases, particularly in the mean and standard deviation, as long as these biases are small and do not lead to unrealistic model behavior. The standardization of SM has also been exploited in previous studies to ensure comparability among different SM products (*Dirmeyer et al.*, 2004; *Koster et al.*, 2009; *Wang et al.*, 2011). It is worth mentioning that bias correction is crucial for the correct quantification of hydrological fluxes in other applications where even small biases would modify the results substantially such as streamflow predictions (e.g., *Luo et al.*, 2007; *Mo and Lettenmaier*, 2014).

Three SMI forecasting ensembles are created in this study: two based on NMME forecasts and one based on ESP. The two NMME based approaches differ with respect to the employed averaging scheme. In the first approach, SMI forecasts are created for all 101 model realizations independently and these are then averaged to obtain a grand NMME ensemble mean for SMI. This approach is denoted as $\overline{\mathrm{SMI}}$. In the second approach, the SM fields are first averaged over all model realizations to create a grand NMME ensemble mean for SM. The latter is then transformed to its respective SMI. This approach is denoted as $\mathrm{SMI}(\overline{\mathrm{SM}})$. These two approaches will provide different results, because the SMI calculation is a highly non-linear transformation. Investigating these two averaging schemes will help to determine the best possible NMME drought forecasting skill.

8.3.4 Subensemble Selection

The North American Multi-Model Ensemble (NMME) based forecasts are further evaluated with respect to the performance of subensembles, as these might give a better performance as the full ensemble but with a reduced computational demand. There are several subensemble selection methods available to identify the best performing subensemble and the backward search algorithm is used in this study as suggested by *Thober and Samaniego* (2014). This algorithm is computationally efficient because it does not require the evaluation of all possible subensemble combinations. The algorithm is summarized here:

1. Select all NMME models as the first subensemble.
2. Sequentially remove a remaining model from the subensemble and evaluate the corresponding performance (e.g., Pearson correlation coefficient R).
3. Repeat step 2 for all remaining models contained in the subensemble.
4. Replace the subensemble with the combination exhibiting the highest performance found in steps 2 and 3.
5. Repeat steps 2 to 4 until the subensemble contains only a single model.
6. Select the combination with the highest performance as the best performing subensemble.

8.4 Results and Discussion

8.4.1 Representation of Spatio-Temporal SMI Dynamics

The overall skill of the NMME and ESP based forecasts to mimic the spatio-temporal dynamics of the reference soil moisture index (SMI) is analyzed for different lead times using the Pearson correlation coefficient R (Figure 8.1). Two different averaging schemes have been employed to create the NMME based forecasts (Section 28.3.3). All three methods have a comparably high skill at one month lead time (R ≈ 0.9), confirming the strong influence of initial hydrologic conditions (IHCs) on SM forecasts at a short lead time (*Shukla et al.*, 2013; *Wood and Lettenmaier*, 2008). Expectedly, the forecasting skill decreases with increasing lead time, but the rate of this decrement is method dependent. For instance, the spatially averaged R value for ESP based forecasts drops from 0.90 at one month lead time to 0.32 at six months lead time (around 65% loss; Figure 8.1, panels g-i). For NMME based forecasts, which have been created by the $\overline{\mathrm{SMI}}$ averaging approach, the skill decreases from 0.87 to 0.25 (around 71% loss; Figure 8.1, panels a-c). This is the strongest decrement among all considered methods, and also the lowest performance at any lead time. On the contrary, NMME based forecasts created by the $\mathrm{SMI}(\overline{\mathrm{SM}})$ averaging approach, have the highest performance and the lowest decrement among all considered methods (around 42% loss; Figure 8.1, panels d-f).

The outperformance of the $\mathrm{SMI}(\overline{\mathrm{SM}})$ approach is also present for all four seasons (Table 8.2). The SMI forecasting skill is highest in winter (DJF) for all considered methods which might be related to snow pack that has a high influence on soil moisture development in the following months. NMME based forecasts also benefit from a higher precipitation forecasting skill during these seasons, particularly at a one month lead time (see Figure 2 in *Mo and Lyon*, 2015). For one and three month lead times, low forecasting skills are observed during autumn (SON). Interestingly, these shift to summer (JJA) for six month lead time. This implies that the forecasting skill is small for forecasts ending at the beginning of winter. This might be related to the fact that higher evapotranspiration during summer and autumn reduce SM persistence during these seasons. Since the ordering of the different methods does not change with season (Table 8.2), the average forecasting skill over the whole year is investigated in the following analysis.

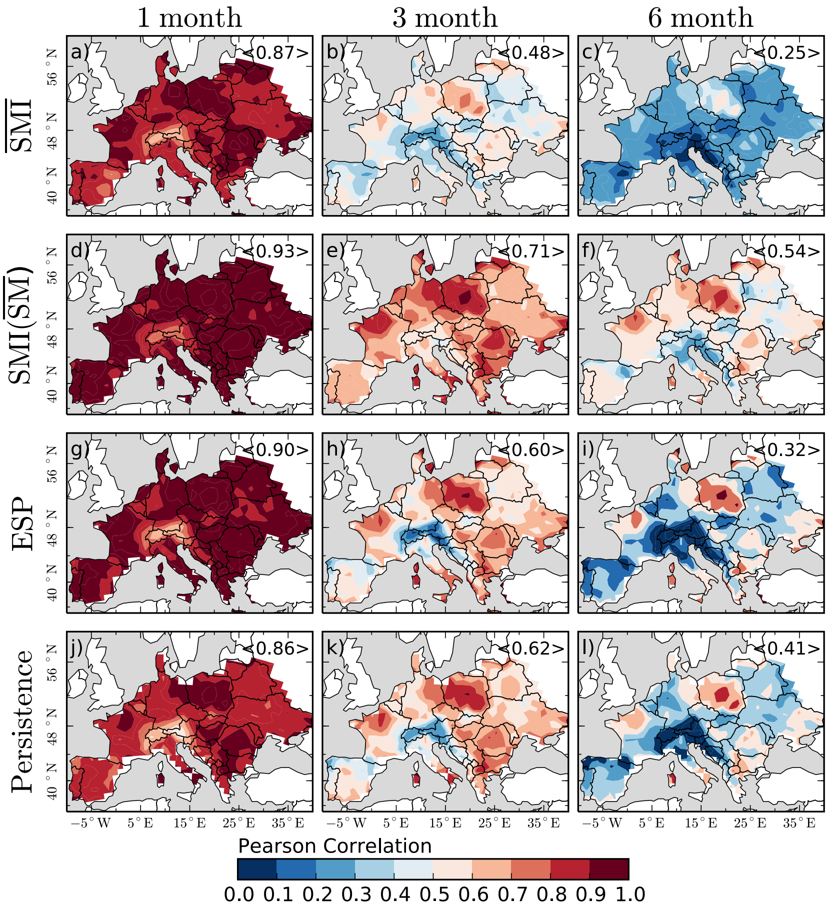

Figure 8.1 The skill to reproduce reference SMI is illustrated in terms of the Pearson correlation coefficient R between the forecasted and reference SMI for lead times of one, three, and six months. The skill of the NMME ensemble is depicted for two averaging schemes: $\overline{\text{SMI}}$ and $\text{SMI}(\overline{\text{SM}})$ in panels a-c and d-f, respectively. In the panels g-i, the skill of the ESP approach is shown. The persistence of reference SMI (estimated as Pearson auto-correlation) is displayed in the panels j-l. The spatial average of the corresponding R is depicted in the upper right corner of each panel.

Although the different forecasting methods yield distinctively different skill, the spatial patterns among the corresponding forecasts are very similar (Figure 8.1, panels a-i). This is observed for any lead time. Regions exhibiting consistently higher skill are located for all methods in Poland, Northern France, and Eastern Ukraine and relatively less skill in the Alps (i.e., Northern Italy, Switzerland, and Austria) and in the Pyrenees along the Spanish-French border. These patterns compare remarkably well with those of the persistence map of reference SMI (Figure 8.1, panels j-l). A high persistence (i.e., auto-correlation) of reference SMI indicates that SM states are exhibiting a long memory, which induces a high dependence of SMI forecasts on IHCs. In this study, perfect knowledge of IHCs is assumed (i.e., they are the same for all forecasts and the reference dataset), which leads to a high SMI forecasting skill (i.e., a high R) at locations exhibiting high SM persistence. On the contrary, SMI forecasts at locations having a short memory will be more dependent on CFs and the large uncertainty therein reduces the ability to represent reference SMI dynamics.

NMME precipitation forecasting skill is very low over Europe (Figure 8.2, panels a-c) as found in previous studies (*Mo and Lyon*, 2015; *Yuan and Wood*, 2012a; *Yuan et al.*, 2015). It is, however, significant for one month lead time. Temperature forecasting skill is comparatively high and does not decrease with increasing lead time (Figure 8.2, panels d-f). Notably, the cumulative precipitation and average air temperature are considered at three and six month lead times because droughts are creeping events that depend more on the integrated forecasting skill than at the forecasting skill of a particular month. It appears that the seasonality helps to achieve a skillful forecast for temperature at a long lead time. ESP based predictions do not exhibit any skill for temperature and precipitation forecasts because this method uses only climatological information. It is thus not surprising that the relatively high skill in temperature and the significant skill in precipitation for one month lead time induces a higher skill into the NMME based forecasts compared to those based on ESP.

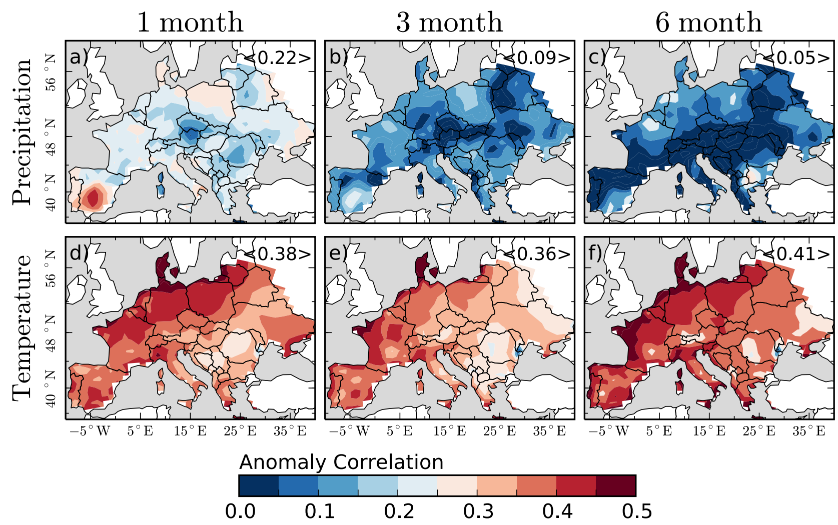

Figure 8.2 Meteorological forecasting skill is quantified for NMME based precipitation (panels a-c) and air temperature (panels d-f) predictions using the anomaly correlation (i.e., the Pearson correlation between forecasted and reference standardized anomalies). The anomalies of cumulative precipitation and average air temperature are considered at three and six month lead times. The spatial average of the corresponding anomaly correlation is depicted in the upper right corner of each panel.

The spatial patterns of SMI($\overline{\mathrm{SM}}$) forecasting skill show a higher agreement with the reference SMI persistence than with those of meteorological forecasting skill (compare Figure 8.1 and Figure 8.2). This highlights the fact that the IHCs have a higher impact on the spatial variability of SMI forecasting skill than the CFs. The latter, however, causes the outperformance of NMME based forecasts in comparison to ESP based ones. These results illustrate the complex interactions between IHCs, CFs, and SMI forecasting skill.

In general, the NMME based forecasts outperform the ESP based ones by 69% on average at a six month lead time (Figure 8.1; compare ETS in panel f and i).

Table 8.2 The skill to reproduce reference soil moisture index (SMI) is presented in terms of Pearson correlation coefficient R between the forecasted and reference SMI for lead times of one, three, and six months averaged over the four seasons DJF, MAM, JJA, and SON. Forecasting skill is depicted for the season the forecast is initialized.

Lead time	Method	DJF	MAM	JJA	SON
	$\overline{\mathrm{SMI}}$	0.93	0.89	0.86	0.83
1 month	SMI($\overline{\mathrm{SM}}$)	0.97	0.93	0.92	0.91
	ESP	0.96	0.90	0.88	0.88
	$\overline{\mathrm{SMI}}$	0.56	0.50	0.47	0.41
3 month	SMI($\overline{\mathrm{SM}}$)	0.77	0.73	0.69	0.63
	ESP	0.73	0.57	0.55	0.55
	$\overline{\mathrm{SMI}}$	0.29	0.29	0.18	0.24
6 month	SMI($\overline{\mathrm{SM}}$)	0.63	0.58	0.44	0.51
	ESP	0.39	0.28	0.21	0.37

A similar outperformance has also been reported by *Yuan et al.* (2015) using bias corrected CFs. No bias correction is applied to the CFs in the present study because the SMI calculation using standardized SM anomalies implicitly accounts for biases in SM as long as the obtained SM dynamics are not unrealistic (e.g., a constantly saturated soil). This illustrates that bias correction of state-of-the-art CFs might not be required to obtain a high forecasting skill for SM drought prediction. An analogous finding was reported by *Yuan and Wood* (2012b) for streamflow, who demonstrated that driving a hydrologic model with raw CFs and subsequently bias correcting the simulated streamflow results in a skillful prediction of the latter.

8.4.2 The Effect of Model Averaging

Additional to the initial land surface conditions, the averaging scheme employed to create the North American Multi-Model Ensemble (NMME) based forecast has a decisive impact on the skill of representing reference soil moisture index (SMI) dynamics (Figure 8.1, panels a-f). Notably, the ensembles created by the SMI($\overline{\mathrm{SM}}$) averaging scheme outperform Ensemble Streamflow Prediction (ESP) based forecasts, while the ensembles created with the $\overline{\mathrm{SMI}}$ approach do not. This implies that the kind of averaging applied can have large impacts on the conclusions drawn in previous studies investigating the capabilities of ensemble drought prediction systems (*Mo and Lettenmaier*, 2014; *Mo et al.*, 2012b; *Wang et al.*, 2011; *Yuan et al.*, 2013a, 2015). The SMI values of individual models are often recasted to the one of the ensemble in these studies and the skill of drought prediction systems might be further increased by using averaging schemes that preserve the frequency of SMI values and therefore capture extremes.

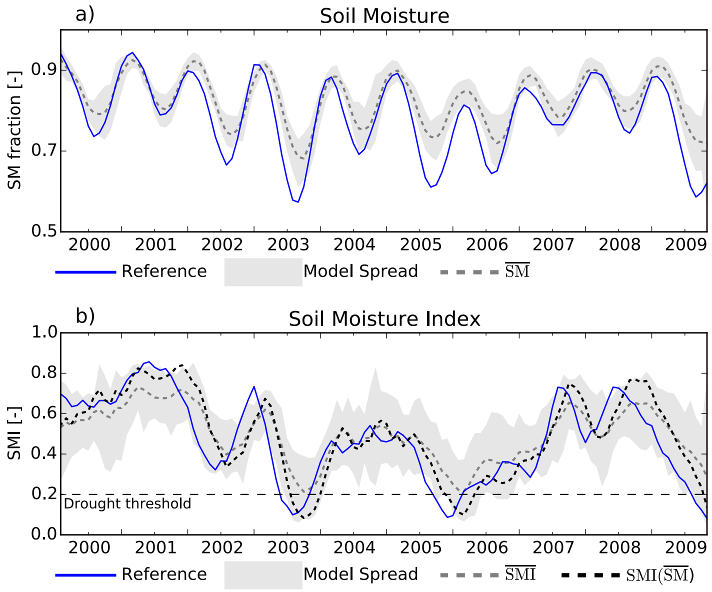

Figure 8.3 For a given grid cell (located in Central France at 47.19° N, 3.21° E), the exemplary time series of SM and SMI are depicted in panels a and b, respectively. In both panels, the blue line delineates the dynamics of the reference dataset and the gray band shows the uncertainty obtained from the 24 ensemble members of the CFSv2 forecasts at two months lead time. The gray dashed line in the top panel a denotes the average of the CFSv2 SM ensemble. The gray and black dashed lines in the bottom panel b denote the SMI ensemble derived by the $\overline{\mathrm{SMI}}$ and $\mathrm{SMI}(\overline{\mathrm{SM}})$ averaging scheme, respectively. The thin horizontal dashed line illustrating the drought threshold 0.2 is displayed for clarity.

The 24-member CFSv2 ensemble is used as one example to illustrate the impact of different averaging schemes on SMI dynamics (Figure 8.3). A strong annual cycle can be observed for both the forecasted and the reference soil moisture (SM) fractions. The mean SM forecast tends to overestimate the reference one, but the latter is mostly within the uncertainty bound of the forecast (Figure 8.3a). The SMI, however, does not exhibit an annual cycle because the climatology of SM is treated separately for each calendar month in the SMI estimation (Section 28.3.3). The ensemble SMI forecasts tend to show a similar temporal dynamic as the reference one, but at the expense of an increased model spread compared to their respective SM forecasts (Figure 8.3b). Due to the increased model spread for SMI, there is always a SMI forecast which is not under drought at a given forecast date. As a result, the $\overline{\mathrm{SMI}}$ averaging approach does not detect drought events given a 0.2 drought threshold (i.e., no time step is identified to be under drought). The reason is that the average of different SMI indices is not a quantile based index itself. For example, it does not fulfill the condition that 20% of the time steps exhibit a SMI less than 0.2. The $\mathrm{SMI}(\overline{\mathrm{SM}})$ scheme captures both the wet and dry extremes better than the $\overline{\mathrm{SMI}}$ scheme and also preserves the property that 20% of the SMI time steps are below 0.2, which is crucial for drought analysis. The same effect was noticed for the other NMME models. Hence, the averaging scheme based on the $\mathrm{SMI}(\overline{\mathrm{SM}})$ approach is used in the further analysis.

8.4.3 Subensemble and Single Model Performance for SMI and Drought Forecasts

Investigating the performance of subensembles is crucial to correctly determine the best possible performance of a given ensemble dataset. The backward selection algorithm proposed by *Thober and Samaniego* (2014) is used to identify subensembles of decreasing size based on Pearson correlation coefficient R and Equitable Threat Score (ETS), separately. The former criteria accounts for both wet and dry extremes, while the latter ETS is used to measure the skill of forecasts to capture drought events based on a 0.2 SMI threshold (see Appendix A for further details of the ETS). The selected subensemble should exhibit a high skill regardless of location and time step considered, which is a basic requirement for a seamless prediction system. Additionally, it is assumed that different subensembles distribute forecasting skill over seasons in a similar fashion, which has been also observed for the different forecasting methods (Table 8.2). For these reasons, the performance criteria are averaged over space, lead time, and forecasting time step.

The skill of any considered subensemble is higher than those of the single models for both criteria (Figure 8.4a). On the contrary, ESP has the lowest performance among all considered approaches for R and only marginally outperforms the worst performing model (CCSM3) for ETS. CFSv2 is the best performing model and the ordering of the single models is the same for R and ETS with the exception of the 2nd and 3rd best models which swap their places (CanCM3 and GEOS5). As a consequence, the models selected within the subensembles are quite similar for the two criteria (Figure 8.4b). Only the selected subensembles of size six are different by more than one model. For both criteria, the backward search algorithm correctly identifies CFSv2 as the single best performing model. It is worth noting that the algorithm would select a different model if the best performing model would have been deselected in a previous iteration. Such a result has been reported for the ENSEMBLES dataset (*Thober and Samaniego*, 2014).

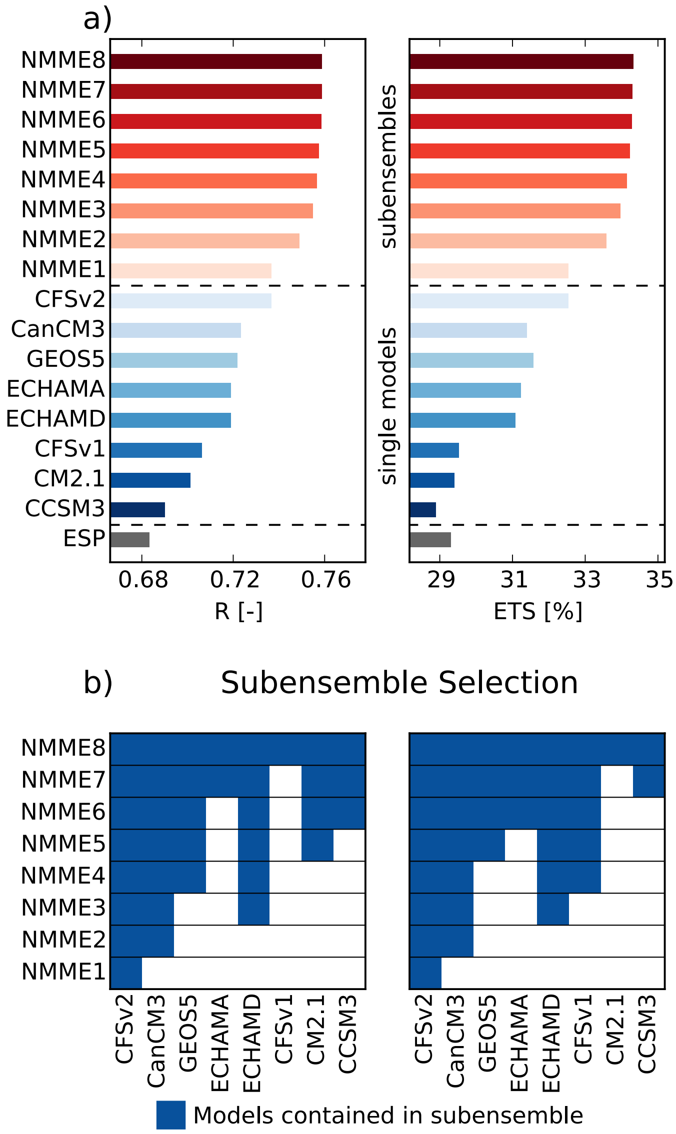

Figure 8.4 In the top panel a, the overall Pearson correlation and ETS estimates are shown for the NMME subensembles (red bars), single models (blue bars), and ESP (gray bar). These estimates are averaged over space and lead times to meet the requirements of a seamless prediction system. The SMI of NMME subensembles is obtained by the $\mathrm{SMI}(\overline{\mathrm{SM}})$ averaging scheme. In the bottom panel b, the single models contained within a selected subensemble for Pearson correlation and ETS are depicted by blue boxes.

The performance of the subensembles decreases monotonically with decreasing ensemble size for both criteria (Figure 8.4a). This justifies the approach pursued in previous studies to use the full ensemble as it exhibits the best possible performance (*Mo and Lettenmaier*, 2014; *Yuan and Wood*, 2013; *Yuan et al.*, 2015, among others). However, the selected subensembles containing four models require 60% of the computational costs of the full ensemble to achieve a skill, which is only 0.3% and 0.5% less than that of the full ensemble for R and ETS, respectively. This highlights that operational forecasting could benefit from using subensembles in favor of the full ensemble because of the reduced computational demand. The performance of the full North American Multi-Model Ensemble (NMME) ensemble (NMME8) is contrasted with that of a subensemble containing four models (NMME4) in the following analysis to further illustrate this aspect. Without loss of generality, NMME4 evaluated against ETS is chosen because it shows a similar performance as that evaluated against R (R value is only 1% less). The four models contained in NMME4 are CFSv2, CanCM3, ECHAMD, and CFSv1 (Figure 8.4b). Only two of these models (CFSv2 and CanCM3) are, however, currently operational in the NMME phase 2 (*NMME*, 2014).

Although subensembles consistently outperform single models and ESP, the spread of both criteria is relatively narrow. This is due to the fact that the initial hydrologic conditions are the same for all forecasting methods, which reduces the variability among the different soil moisture forecasts. In other words, the high variability in climatic forecasts is dampened while propagating through the hydrologic system exhibiting long memory. It is worth mentioning that substantially different subensemble performances have been observed for atmospheric variables like extreme precipitation indices (*Thober and Samaniego*, 2014).

8.4.4 Spatio-Temporal Distribution of Drought Forecasting Skill

It is desirable for a drought prediction system to be seamless with a high forecasting skill regardless of the location and the lead time. The forecasting skill of most prediction systems, however, varies in space and time (*Dutra et al.*, 2013; *Shukla et al.*, 2013; *Yuan and Wood*, 2013). The spatio-temporal distribution of Equitable Threat Score (ETS) is analyzed here to understand these variations as well as the factors that influence drought forecasting skill.

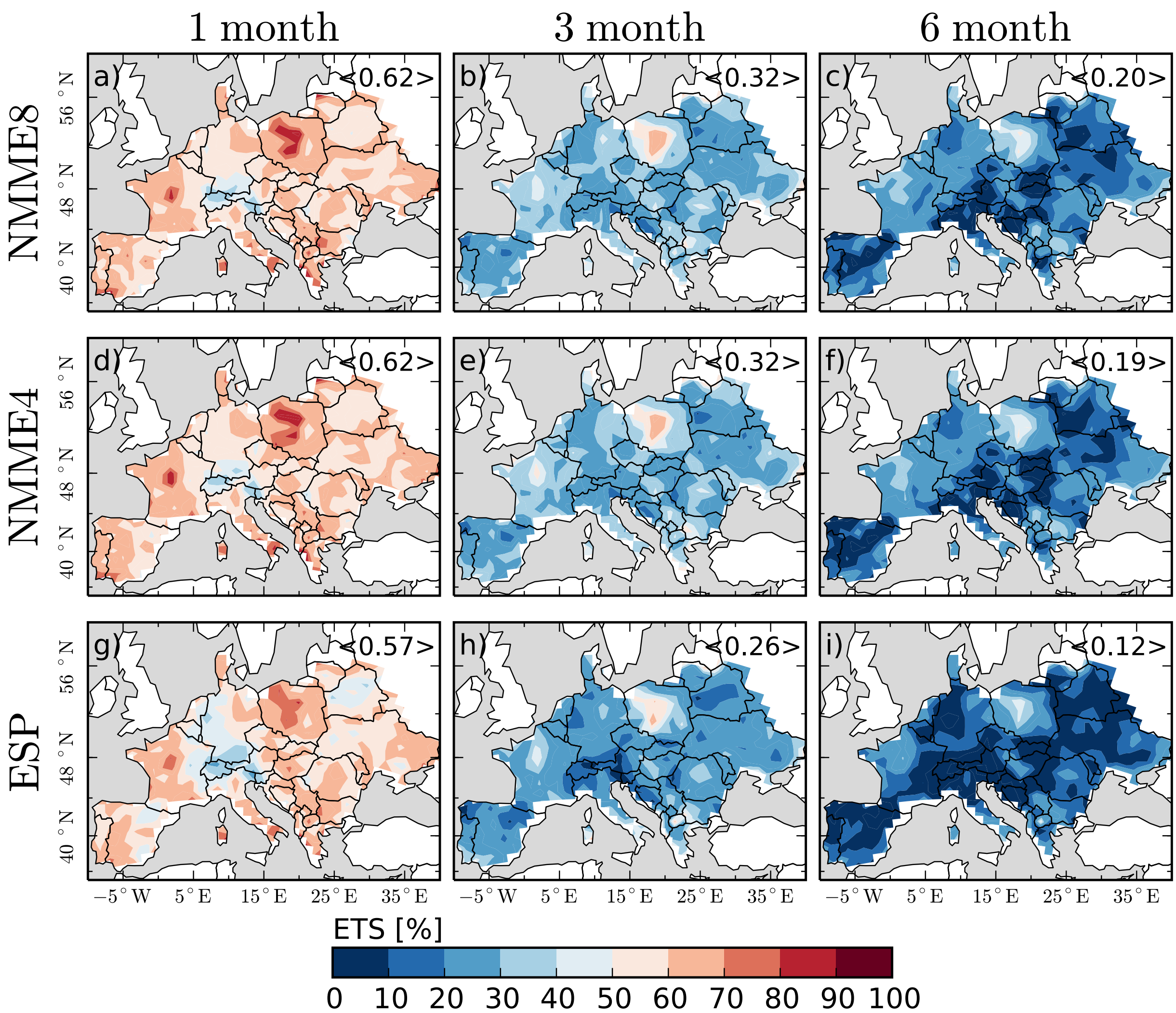

Figure 8.5 Spatial distribution of ETS at one, three, and six month lead time is displayed for the full NMME ensemble (NMME8) in panels a-c, NMME subensemble containing four models (NMME4) in panels d-f, and ESP in panels g-i. The NMME based forecasts are obtained by the $\mathrm{SMI}(\overline{\mathrm{SM}})$ averaging scheme. The corresponding spatial averages of ETS are denoted in the upper right corner of every panel.

Distinctive spatial patterns in ETS are observed for both NMME and ESP (Figure 8.5), which are similar to those of the Pearson correlation for the reference soil moisture index (SMI) dynamics (Figure 8.1, panels j-l). This illustrates that the impact of initial hydrologic conditions (IHCs) is also evident for extreme conditions. The differences in ETS between two locations across the study domain are as high as 40% (e.g., difference between Switzerland and Poland at one month lead time for NMME8; Figure 8.5a). These spatial differences are larger than the differences between the NMME8 and ESP forecasting approaches, which range up to 8% on average at six month lead time. It is worth noting that the spatial distribution between NMME8 and NMME4 is very similar (Figure 8.5). At 90% of the grid cells, the differences between these two ensemble based forecasts are smaller than 5% in terms of ETS irrespective of the lead time.

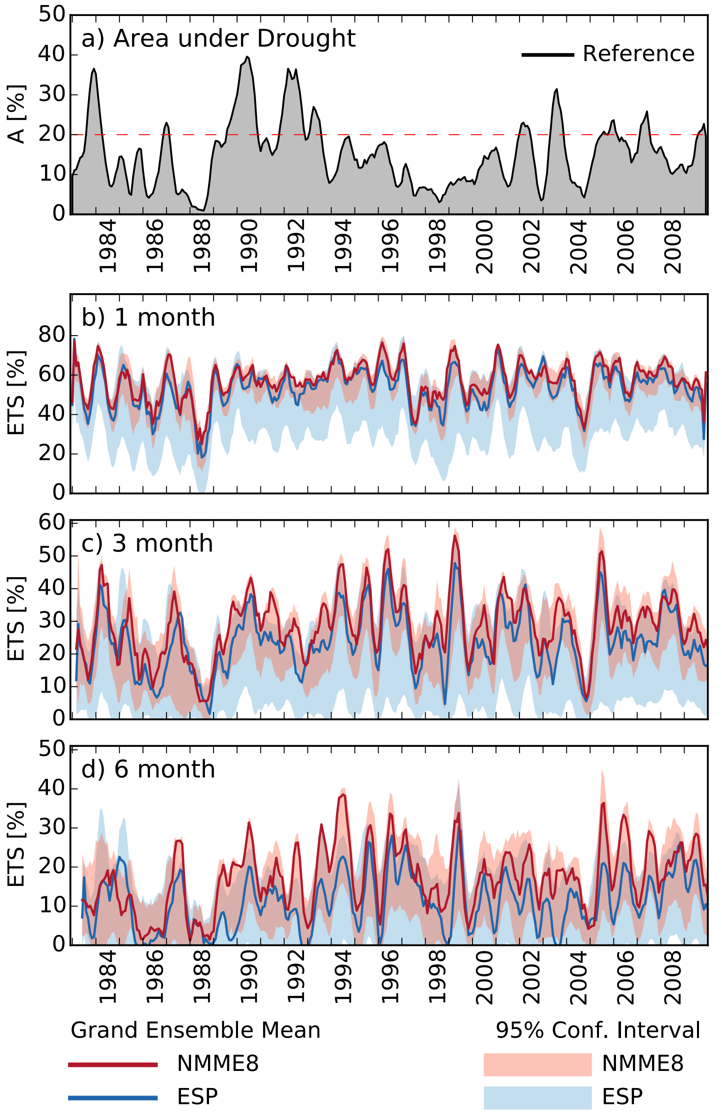

Figure 8.6 The top panel depicts the fraction of area under drought based on the reference SMI dataset. The thin horizontal dashed line is added for clarity displaying the threshold for droughts covering more than 20% of the European domain. Panels b-d illustrate the temporal variability of ETS for the ESP (blue lines) and the full NMME ensemble (red lines) based SM drought forecasts. The NMME based forecasts are obtained by the $\mathrm{SMI}(\overline{\mathrm{SM}})$ averaging scheme. Additionally, the 95% confidence interval for the single ESP and NMME ensemble members is depicted as light red and blue bands, respectively. Ticks mark the end of the respective year. The scale of y-axis are different for each panel for clarity.

This skill of both the NMME8 and ESP forecasting methods also depends on the forecast date (Figure 8.6, panels b-d). The differences between the smallest and highest ETS can be also as high as 40% for both forecasting methods, whereas the maximum difference between NMME8 and ESP forecasts at any given time step is at most 20%. Both forecasting methods, as expected, show lower ETS values at longer lead times, but the rate of decrement is less for NMME8 than for ESP. This leads to the relative outperformance of 69% on average at a six month lead time as discussed above (Section 38.4.1). These results illustrate the added value of an ensemble seasonal forecasting system at longer lead times (*Mo and Lettenmaier*, 2014). In general, NMME8 forecasts significantly outperform ESP ones at any location and lead time at a 5% significance level, which has also been reported by *Yuan et al.* (2015) using the VIC land surface model over the Danube basin in Europe. This result is obtained by applying a Student's t-test, which has been previously used in drought prediction studies (*Wilks*, 2011; *Yuan et al.*, 2015). A similar result is obtained for the NMME4 subensemble, which requires only 60% of the computational demand as compared to the NMME8 (not shown).

The spread of single model performance is significantly narrower for the full NMME ensemble (19% on average) as compared to that of ESP (29% on average) at a 5% significance level (Figure 8.6, panels b-d). A similar result is obtained when the same number of samples (forcing members) is evaluated for NMME8 and ESP. The higher uncertainty for the ESP based forecasts can be mostly attributed to poorly performing forecasts. The spread of ETS for the NMME8 based forecasts is often located within the upper tail of that estimated for the ESP based ones. The skill of the full NMME ensemble is comparable to that of the best performing model at a given forecast date (i.e., the upper limit of single model spread shown in Figure 8.6, panels b-d), which has also been reported for an NMME based prediction system over the CONUS (*Mo and Lettenmaier*, 2014). It is worth noting that there exists not a single model that outperforms all others at all forecasting dates. For example, CFSv2 only outperforms all other models at 20% of all forecasting dates, although it is the overall best performing model (as discussed above; Figure 8.4). This again highlights the advantage of using ensemble based forecasts over ones based on a single model.

The temporal dynamics of ETS for the full NMME ensemble and ESP are quite similar (Figure 8.6, panels b-d), which again signifies the role of IHCs for drought predictions. Low ETS values are generally observed during periods of drought recovery with less extensive droughts (e.g., 1988, during autumn 1998, and at the end of 2004; Figure 8.6a). Both forecasting methods overestimate the drought extent during these periods, which results in a high false alarm rate and thus reduces ETS. On the contrary, high ETS values are observed during drought development phases (e.g., during 1990, 1994, and summer of 2005). These results illustrate that the drought forecasting skill varies depending on the states of drought events (e.g., drought development and recovery). These are defined in the following section.

8.4.5 Forecasting Skill during Drought Development and Recovery

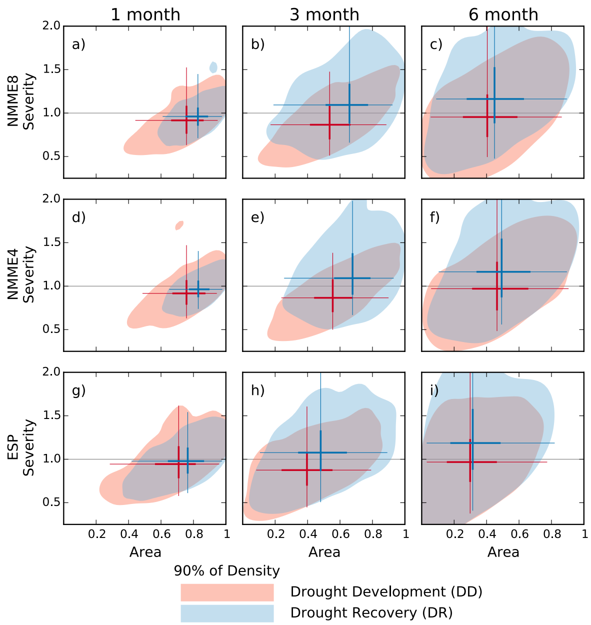

Figure 8.7 Probability density function for drought severity and drought area is illustrated for different lead times for forecasts obtained by single NMME and ESP ensemble members. The performance for all NMME models is shown in panels a-c and only for four NMME models in panels d-f. The performance for ESP based forecasts is displayed in panels g-i. The area containing 90% of the density for both characteristics is depicted in each panel as red and blue regions for drought development and recovery, respectively. Additionally, the spread for each characteristic is shown as box plots for the different drought phases (95% confidence interval as thin lines, the spread between the 25th and 75th quantiles as thick lines, and the median is located at the intersection).

To further investigate the forecasting skill during drought development and recovery phases, two drought characteristics are analyzed for major drought events that cover more than 20% of the European domain (e.g., the 1983, 1990, and 2003 drought; see also Figure 8.6a). A drought time step is defined as development (recovery) if it occurs before (after) the peak extent of the respective event. The two characteristics are the drought severity and the area under drought (see Appendix B for details). Both of these characteristics are normalized by their corresponding reference estimates (based on E-OBS) to make them comparable among different events. The perfect forecast would correspond to a value of one for both characteristics. The drought characteristics during both phases are calculated for all NMME and ESP ensemble members separately. Finally, a probability density function is estimated jointly for the two characteristics using a kernel estimation method (Equation A.1) to assess their associated spread, following the procedure used by *van Loon et al.* (2014).

In general, the forecasted drought severity matches the median reference one quite well, with deviations less than 20% irrespective of the lead time, drought phase, and forecasting method (horizontal lines in Figure 8.7). On the contrary, substantial underestimations in drought area are observed with increasing lead time up to 55% for NMME8, 51% for NMME4, and 68% for ESP (vertical lines in Figure 8.7). Additionally, these are more pronounced during drought development phases than during recovery phases. In summary, the drought forecasts exhibit a higher mismatch in correctly detecting reference drought location. If a drought has been correctly forecasted at a given location, then it is likely that the severity of this event would be comparable to that of the reference one.

The spread of drought severity and area increases with lead time for all forecasting methods (see regions containing 90% of the density in Figure 8.7). Expectedly, the relatively larger uncertainty in climatic forecasts at longer lead times causes a higher spread in drought characteristics (*Shukla and Lettenmaier*, 2011; *Wood and Lettenmaier*, 2008). This spread is larger during the drought recovery than during the development phases at a long lead time, which is in agreement with *Mo* (2011) who reported that drought development is more predictable than drought recovery.

The spread is also remarkably similar for the NMME8 and NMME4 based forecasts. For example, there is a comparable overlap of spread estimated during the drought development and the recovery phases at a three month lead time. This overlap is considerably different from that observed for Ensemble Streamflow Prediction (ESP) based forecasts (Figure 8.7, compare panels b, e, and h). These results illustrate that the NMME4 subensemble also has a similar performance as the full NMME ensemble during different drought phases, but only requiring 60% of the computational resources.

In general, all forecasting methods underestimate the reference drought severity during the drought development phases at all lead times (Figure 8.7). This results from too wet forecasts leading to higher soil moisture index conditions as compared to the relatively drier reference ones. On the contrary, drought severity is overestimated during the drought recovery phases at three and six months lead times. The forecasts are drier than the reference one in this case. In other words, they are not able to add sufficient SM to recover from the drought. These results illustrate the fundamental influence of initial hydrologic conditions (IHCs) that persist throughout the drought forecasts leading to a consistent lag of these with respect to the reference soil moisture index dynamics (see also Figure 8.3b). This is expected for ESP as it represents a climatological forecast and the skill is mainly derived from the correct representation of IHCs (*Koster et al.*, 2004; *Shukla et al.*, 2013). The skill of NMME based forecasts has a similar dependence on the IHCs as ESP despite that NMME models represent physical dynamics of the Earth system. They do, however, provide a substantially better forecast for drought area as compared to ESP (Figure 8.7).

8.5 Summary and Conclusions

In this study, the skill of a seasonal hydrologic prediction system for soil moisture (SM) drought forecasts is evaluated over Europe for a 27 year hindcast period (1983-2009). The prediction system is based on meteorological forecasts of the North American Multi-Model Ensemble (NMME) that are used to drive the mesoscale Hydrologic Model (mHM). The skill of NMME based forecasts is contrasted with that of the Ensemble Streamflow Prediction (ESP) approach. The obtained SM estimates from both forecasting approaches are transformed to a quantile based soil moisture index (SMI) to conduct a drought analysis using a 0.2 SMI threshold. Drought prediction skill is quantified in terms of the Equitable Threat Score (ETS) employing a reference SMI field. The latter has been created using the observation based E-OBS dataset.

NMME based forecasts significantly outperform ESP based ones particularly at a long lead time (i.e., up to 69% higher ETS at six month lead time). This is achieved only if the SMI has been calculated for the grand ensemble SM mean. In contrast, the grand ensemble SMI obtained by averaging single NMME model based SMIs does not outperform the ESP based one. Among the NMME based forecasts, the full ensemble outperforms the single models as well as all selected subensembles. There is a considerable variability in the skill of SMI forecasts over Europe (i.e., up to 40% in space and time), regardless of the forecasting approach. This variability is strongly related to the persistence of reference SM, illustrating the strong impact of initial hydrologic conditions (IHCs) on SM drought forecasts. The IHCs are respectively wetter during drought development phases than during drought recovery phases, which induces an underestimation of drought severity during the former and an overestimation during the latter phase.

The main conclusion of this study is that NMME based forecasts are useful for seasonal SM drought prediction over Europe, which is in accordance with recent studies for the CONUS and GEWEX river basins using the VIC land surface scheme (*Mo and Lettenmaier*, 2014; *Yuan et al.*, 2015). The NMME based forecasts are well suited for a seamless prediction system as their skill is consistently higher than that of ESP based ones over the entire study domain at all lead times.

The selected subensembles only show performance losses less than 1% on average in comparison to the full ensemble, but at 60% of the computational demand. Subensembles thus provide a promising alternative to the full ensemble and might be useful for operational seasonal SM drought forecasting. The subensemble skill has been averaged over space, lead time, and forecasting time step because the subensemble should exhibit a high skill regardless of the location and time step to be useful for a seamless prediction system. Alternative selection methods, however, could take the spatio-temporal variability of forecasting skill into account in the selection process. They should also test whether the skill of subensembles is stationary in time, which is crucial requirement for operational forecasting. Moreover, bias correction of raw meteorological data has little impact on SM drought forecasting skill because the calculation of the quantile based SMI already accounts for systematic biases, particularly in the mean and standard deviation, as long as these do not lead to unrealistic SM dynamics.

The results of this study illustrate the ubiquitous impact of IHCs on SM drought forecasting skill. The uncertainty associated with imperfect IHCs is, however, not considered here. Methods for further evaluating this aspect such as the reverse ESP approach have been investigated in previous studies using observational datasets (*Shukla and Lettenmaier*, 2011; *Shukla et al.*, 2013; *Wood and Lettenmaier*, 2008). With the increase of computational resources, these should also be considered in the evaluation of ensemble SM drought prediction systems such as those based on the NMME. Future studies could investigate the NMME phase two data containing real-time forecasts instead of the hindcast dataset explored in this study.

Acknowledgments

This study was carried out within the Helmholtz-Association climate initiative REKLIM (www.reklim.de) and was supported by the Helmholtz Interdisciplinary School for Environmental Research (HIGRADE). We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). We also would like to thank the International Research Institute for Climate and Society (IRI; http://iridl.ldeo.columbia.edu/ SOURCES/.Models/.NMME/) for making the NMME dataset available. Moreover, we would like to thank the editor and three anonymous reviewers for their constructive comments that helped to further improve this text.

Appendix A: Equitable Threat Score

Forecast verification for discrete events (e.g., a drought event) is commonly carried out using measures that are based on a 2×2 contingency table (*Wilks*, 2011). In this study, we use the Equitable Threat Score (ETS) as skill measure, which is defined as

$$ETS = 100\frac{a - a_{ref}}{a - a_{ref} + b + c}, \tag{8.2}$$

where a is the number of drought events that occur in both the forecast and the reference dataset (commonly called hits), b is the number of drought events that occur in the forecast but not in the reference dataset (commonly called false alarms), and c is the number of droughts that occur not in the forecast but in the reference dataset (commonly called misses). a_{ref} is defined as

$$a_{ref} = \frac{(a+b)(a+c)}{n}, \tag{8.3}$$

where n is the total number of time steps. ETS is used in this study because it condenses the hit rate ($a/(a+c)$) and the false alarm rate ($b/(a+b)$) into one metric. An ETS of 100% indicates a hit rate of 1 and a false alarm rate of 0, which means that all drought events are forecasted perfectly.

Appendix B: Drought severity and area

Two drought characteristics are evaluated during the drought development and recovery phase. These are the fraction of correctly forecasted drought area and the drought severity of this area. For a given time step t, the former is defined as

$$A(t) = \frac{a(t)}{a(t) + c(t)}, \tag{8.4}$$

where $a(t)$ is the number of grid cells under drought both in the forecast and the reference dataset at time step t and $c(t)$ is the number of grid cells under drought that occur not in the forecast but in the reference dataset at time step t. It is worth mentioning that this area is equivalent to the hit rate estimated over space.

The drought severity is calculated for the grid cells that exhibit a drought both in the forecast and the reference dataset. For a given time step t, the drought severity is defined as

$$S(t) = \sum_{i \in a(t)} [\tau - \mathrm{SMI}_i(t)]_+, \tag{8.5}$$

where τ is the SMI drought threshold (here 0.2), $(\cdot)_+$ is the positive part function, and $a(t)$ is defined as above. A large deviation from the drought threshold leads to higher severity indicating a more severe drought. The severity of the forecast is then normalized by that of the reference dataset (calculated over the same area $a(t)$ using equation 8.5) to make them comparable among different drought events.

CHAPTER 9

PROPAGATION OF FORCING AND MODEL UNCERTAINTIES ON TO HYDROLOGICAL DROUGHT CHARACTERISTICS IN A MULTI-MODEL CENTURY-LONG EXPERIMENT IN LARGE RIVER BASINS

L. Samaniego[1] et al.

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Samaniego, L., R. Kumar, L. Breuer, A. Chamorro, M. Flörke, I. G. Pechlivanidis, D. Schäfer, H. Shah, T. Vetter, M. Wortmann, and X. Zeng (2016), Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins, Climatic Change, 141(3), 435–449

9.1 Abstract

Recent climate change impact studies studies have presented conflicting results regarding the largest source of uncertainty in essential hydrological variables, especially streamflow and derived characteristics that describe the evolution of drought events. Part of the problem arises from the lack of a consistent framework to address compatible initial conditions for the impact models and a set of standardized historical and future forcings. The ISI-MIP2 project provides a good opportunity to advance our understanding of the propagation of forcing and model uncertainties on to century-long time series of drought characteristics using an ensemble of hydrological model (HM) projections across a broad range of climate scenarios and regions. To achieve this goal, we used six regional preconditioned hydrological models set up in seven large river basins: Upper-Amazon, Blue-Nile, Ganges, Upper-Niger, Upper-Mississippi, Rhine, and Upper-Yellow. These models were forced with bias-corrected outputs from five CMIP5 general circulation models (GCMs) under two extreme representative concentration pathway scenarios (i.e., RCP2.6 and RCP8.5) for the period 1971-2099. The simulated streamflow was transformed into a monthly runoff index (RI) to analyze the attributions of the GCM and HM uncertainties on to drought magnitudes and durations over time. The results indicated that GCM uncertainty mostly dominated over HM uncertainty for the projections of runoff drought characteristics, irrespective of the selected RCP and region. In general, the overall uncertainty increased with time. The uncertainty in the drought characteristics increased as the radiative forcing of the RCP increased, but the propagation of the GCM uncertainty on to a drought characteristic depended largely upon the hydro-climatic regime. Although our study emphasizes the need for multi-model ensembles for the assessment of future drought projections, the agreement between the GCM forcings was still too weak to draw conclusive recommendations.

9.2 Introduction

Droughts are creeping hydro-meteorological events that can bring societies and natural systems to their limits, and induce large famines, drinking and irrigation water shortfalls, natural fires, the degradation of soil and water quality, and, in many cases, considerable socio-economic losses. Many climate change impact studies have attributed increases in the affected areas of drought and changes in the severity and the duration of droughts due to global warming (*Briffa et al.*, 2009; *Dai*, 2013; *Held and Soden*, 2006; *Mueller and Zhang*, 2015; *Sheffield and Wood*, 2008a). The projections obtained for the twenty-first century indicate that it is likely that "severe and widespread" droughts may occur (*Dai*, 2013), and, if so, droughts will "set in quicker and be more intense" (*Trenberth et al.*, 2014). Contradictory results have been reported in recent years (e.g., *Dai*, 2013; *Sheffield et al.*, 2013) regarding this topic. For these reasons, the IPCC-AR5 summarized these plausible changes as follows: "There is medium confidence that droughts will intensify in the 21st century in some seasons and areas, due to reduced precipitation and/or increased evapotranspiration" (*Seneviratne et al.*, 2012).

It is recognized that the quantification of the predictive uncertainties in essential hydrological variables (e.g., streamflow) and their attribution to the main sources is of particular interest (*Pappenberger and Beven*, 2006) in climate change studies. Currently, however, the uncertainties intrinsic to historical meteorological observations and those related to the modeling chains used to estimate drought indices and/or derived characteristics have not yet led to conclusive results with respect to the effects of global warming on droughts (*Dai*, 2013; *Sheffield et al.*, 2013; *Trenberth et al.*, 2014).

There are a number of factors that contribute to the epistemic predictive uncertainty (*Beven et al.*, 2011) of hydrological variables, including observational errors in hydro-meteorological data, errors induced by the interpolation of meteorological data, hydrological model (HM) structures and their internal parameterizations (*Kumar et al.*, 2013; *Vetter et al.*, 2015) parametric uncertainties in hydrological models (*Samaniego et al.*, 2013), uncertainties related to mapping of subgrid physiographic information such as soil textures (*Livneh et al.*, 2015; *Samaniego et al.*, 2010a) and land use/cover changes (*Samaniego and Bárdossy*, 2006), and uncertainties related to water management practices. In the case of future hydrological projections, there are other important sources of uncertainty such as the initial conditions of general circulation models (GCMs) (*Buizza*, 2002), parameterizations of GCMs, errors caused by the numerical approximations used to solve the underlying GCM equations at a given resolution (*Buizza*, 2002), assumptions made to downscale (*Blöschl and Montanari*, 2010) and bias-correct GCM model outputs (*Ehret et al.*, 2012), and uncertainties stemming from emission scenarios.

Few studies have sought to assess the ability of regional and global HMs to reproduce extreme hydrological events (*Donnelly et al.*, 2015; *Gudmundsson et al.*, 2012) as well as disentangling the individual contributions of different uncertainty components in hydrological projections (*Pechlivanidis et al.*, 2016). In the recent literature, there are different (contradictory) views regarding the dominant sources of the uncertainties. *Prudhomme et al.* (2014), for example, concluded that the main source of uncertainty in projections of hydrological drought severity and the deficit index arises from the variability among the HMs, whereas a substantial but smaller share was attributed to GCM variability. Their analyses were based on outputs from several global hydrological and land surface models that were forced with various GCMs projections for the 21st century under different representative concentration pathways (RCPs). Similarly, *Haddeland et al.* (2011) concluded that HM structure is a major source of uncertainty that should be considered in climate change impact studies. Their study was, however, limited to a historical hydrological analysis based on WATCH forcing data to drive global (uncalibrated) HMs.

Other authors, however, concluded that GCMs can outweigh the contributions of the uncertainties in global HMs (*Giuntoli et al.*, 2015) and regional HMs (*Arnell*, 2011; *Bosshard et al.*, 2013; *Teng et al.*, 2012; *Vetter et al.*, 2015). *Giuntoli et al.* (2015) also observed that the dominant source of uncertainty in summer and autumn comes from GCMs, with the exception of snow-dominated regions. *Bosshard et al.* (2013), however, suggested that the uncertainties attributed to HMs and post-processing may gain importance in winter and spring or in regions affected by declining water resources. The results from those studies have been used to note that the individual contributions of these sources of uncertainty are not additive (*Bosshard et al.*, 2013) due to the non-linearities of the modeling chain. Along the same lines, *Arnell* (2011) concluded that the uncertainty in GCM outputs is considerably larger than that of the parameter uncertainty in HMs based on simulations in UK basins.

Most of these multi-model climate change assessments were carried out in regions that share similar hydro-climatic regimes, and in many cases, they lacked a consistent framework to address compatible initial conditions (e.g., similar forcings during model spin-up and a consistent protocol for ensemble model setup) and a set of standardized historical and future forcings for HMs. For these reasons, it is worth outlining frameworks that would allow the identification and quantification of the various factors that contribute to the predictive uncertainties in key hydrological variables and derived drought characteristics (*Bosshard et al.*, 2013; *Haddeland et al.*, 2011; *Schewe et al.*, 2014; *Vetter et al.*, 2015).

The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, www.isimip.org) provides a unique opportunity to analyze the propagation of uncertainties stemming from the emission scenarios, GCM-derived forcings, and HM structural differences on to hydrological drought characteristics using an ensemble of model projections across a range of climatic regions. In this study, we focused on these three dominant sources of uncertainty and their impacts on monthly runoff index (RI) using an ensemble of six HMs setup at seven large river basins around the globe. We acknowledge that other sources of uncertainty are also important, but they were not considered here because of lack of information.

The main hypotheses guiding this research were formulated as follows. 1) The uncertainty contribution of the GCMs on RI and derived drought characteristics outweighs that from the HMs regardless of the hydrological regime represented by the selected large-scale river basins. 2) Given a GCM forcing, the drift in the RI time series of a given HM is practically indistinguishable from the ensemble RI. Therefore, the drift mainly arises from the uncertainty in the GCM forcings. 3) The uncertainty in drought characteristics is RCP dependent.

The rationale behind these hypotheses is based on the fact that simulations of the dynamics of the Earth's atmosphere and oceans carried out using GCMs are affected by chaotic behaviors intrinsic to the atmosphere (*Cretat and Pohl*, 2012) and are also very sensitive to initial conditions (*Buizza*, 2002), subgrid scale process parameterizations and the simulation of various feedback mechanisms. These effects, in turn, induce large variabilities in GCM outputs (e.g., precipitation) that are later used as forcings for HMs. A HM, on the other hand, simulates a deterministic hydrodynamic system that has storage components characterized by a long-range memory (e.g., vadose zone), which induces a significant attenuation of the meteorological forcings. As a result, the internal variability of an HM, ceteris paribus, is expected to be much lower than that of the GCMs. In addition to both sources of uncertainty, the emission scenarios that guide future GCM projections would add considerable variability to forcing variables, especially at end of the 21st century. In summary, in this study, we aimed to disentangle the effects of those three sources of uncertainty on drought characteristics.

9.3 Method

9.3.1 Study area and data sets

This study is conducted in seven large river basins around the world that have an area greater than $10^5\ \mathrm{km}^2$ and that represent a wide range of hydro-climatic and physiographic conditions: 1) the Upper-Amazon at the gauging station Sao Paulo de Olivenca (equatorial fully humid), 2) the Blue-Nile at Kartoum (warm temperate with dry winters and warm summers), 3) the Ganges at Farakka (warm temperate with dry winters and hot summers), 4) the Upper-Mississippi at Alton (warm temperate or snow dominated, humid with hot summers), 5) the Upper-Niger at Dire (equatorial, partly monsoonal with dry winters), 6) the Rhine at Lobith (warm temperate, humid with warm summers), and 7) the Upper-Yellow at Tangnaihai (arid with dry winters and cold).

We use the daily streamflow simulations from the following six hydrological models: HBV, HYPE, mHM, SWIM, VIC, and WaterGAP3. With an exception of HBV and HYPE, the remaining four models are run consistently across all the study basins. HBV simulations are used where HYPE simulations are not available. Therefore, the effective number of HMs is five. These HMs vary in their complexities, spatial discretizations, and process representations. They are constrained against observed streamflows using the historical WATCH forcing (*Weedon et al.*, 2011). Interested readers may refer to *Krysanova and Hattermann* (2016) for more details on HM description, setup and the basins' characteristics.

Every HM is driven by five CMIP5-GCM climate projections for the period from 1971 to 2099 under two future emission scenarios that have radiative forcings by the end of century of 2.6 Wm^{-2} and 8.5 Wm^{-2} and are denoted hereafter as RCP2.6 and RCP8.5, respectively. These scenarios are chosen because they represent two extreme future conditions. The GCMs are GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, and NorESM1-M (*Taylor et al.*, 2012). The ranges of uncertainty in annual temperature and precipitation projections of the set of these five GCMs are comparable with those of all CMIP5 models (see protocol-report on www.isimip.org). The required forcings, including precipitation and air temperature, are bias corrected to match the corresponding long-term monthly means of the WATCH-forcing for the overlapping reference period 1960-1999 (*Hempel et al.*, 2013). The HM initializations and simulations corresponding to every GCM under the given RCP scenarios are carried out following the ISI-MIP2 protocol (www.isimip.org), which specifies standardized forcings and periods for calibration and verification, spin-up and projection periods. Note that the selection of the forcings, hydrological models and river basins is constrained by the data availability in the ISI-MIP2 project.

9.3.2 Runoff drought index and derived characteristics

Monthly discharge time series for every GCM under a given RCP scenario are estimated based on the simulated daily discharges for the period 1971-2099. To ease the comparisons among the streamflow simulations obtained with the different RCPs, GCMs, and HMs, the resulting monthly streamflow time series are converted into a quantile based runoff index (RI) using a non-parametric kernel density estimator (KDE) (*Wilks*, 2011). The KDEs are separately obtained for every calendar month, basin, and GCM/HM combination to allow comparability across the time, space and model combinations. The KDEs estimated for the period from 1971 to 2000 are used as a reference to recast the future streamflow time series (2001-2099) under both RCP scenarios. Details can be found in Supplement 9.5. It should be noted that RI is a probabilistic index that denotes the monthly quantile of simulated streamflow and therefore ranges between [0,1].

Two drought characteristics are considered in this study: total drought magnitude M and duration D. The total magnitude ofdrought events occurring within a period T (30 years) is estimated as the temporal integral of the monthly RIs below a threshold value τ. In this study, a threshold value τ of 0.2 is used as a reference. This value denotes a 20% probability of occurrence, which is normally used to denote the onset of a moderate drought (*Andreadis et al.*, 2005; *Kumar et al.*, 2016; *Samaniego et al.*, 2013; *Vidal et al.*, 2010). The duration of drought spells within the period T is therefore the total number of months in which the RIs is below the threshold value τ. Further details are provided in Supplement 9.5.

Both drought characteristics are estimated for rolling windows of T years. T should be large enough to capture climatological changes (e.g., T=30). The shift between two consecutive rolling windows is S years. For example, if T=30, and S =10, as used in this study, the drought characteristics are estimated for the following rolling windows: 1971-2000, 1981-2010, ... , and 2071-2099.

9.3.3 Uncertainty contributions of the GCMs and HMs

The goal here is to disentangle the GCMs' and HMs' uncertainty contributions on the simulated RIs and the derived drought characteristics. A set of 25 hydrological simulations corresponding to the five GCMs and five HMs for every study basin and RCP is used for this analysis. The individual contribution of a given GCM or HM is estimated using a sequential sampling procedure similar to that proposed by *Schewe et al.* (2014).Based on this procedure, the component of the GCM uncertainty $\overline{R}_{cG}(.)$ for a given RCP scenario c, is characterized by computing the range of a drought characteristic across all GCMs for each HM individually and then averaging it over all the HMs. The HM uncertainty component $\overline{R}_{cH}(.)$ is estimated in a similar fashion but by first computing the range across all HMs for each GCM individually and then averaging it over all GCMs. This procedure is applied to every basin and RCPs under investigation. Additional details on this procedure are provided in Supplement 9.5 and in Fig. 9-A.1.

The uncertainty components are quantified using a range statistic ($R_\bullet$) for both drought characteristics mentioned above. The range is preferred to other dispersion measures such as variance, inter-quantile range or the median absolute deviation because it is the most useful statistic to understand the full range of the dispersion given the small sample of GCMs and HMs available in the ISI-MIP2 project. Sampling with replacement can be used to generate confidence intervals for the range statistics. The number of realizations should be large enough to obtain reliable statistics (*Demuth and Heinrich*, 1997). In the present study, N=1000 is sufficient for convergence. The advantage of this method stems from the fact that it can be easily implemented within a bootstrapping framework by randomizing the selection of the HMs and GCMs with replacement (see Step 6 in Supplement 9.5). As a result, the confidence interval and significance level of $\overline{R}_\bullet$ can be estimated without any assumptions of normality that are necessary for parametric tests in standard procedures such as the Analysis of Variance (ANOVA). A non-parametric (bootstrapping) procedure is preferred to reduce the effects of the biased variance estimation.

9.4 Results and discussion

9.4.1 Propagation of the forcing and model uncertainties on to the RIs

The multi-model ensemble of the RIs for every river basin exhibits a large predictive uncertainty. Fig. 13.1 depicts the 30-year rolling averages of the ensemble means and ranges of the RIs based on all combinations of the GCMs and HMs. The moving average of RI is estimated to investigate potential hydro-climatological trends until the end of the 21st century. The two extreme scenarios with respect to the GCMs' radiative forcing (i.e., RCP2.6 and RCP8.5) are selected to show two extreme situations under future climates. In general, this analysis shows that in all cases the uncertainties in the RI ensembles increase with time. It can also be observed that in almost all cases, the uncertainties in the projected RIs for RCP 8.5 are higher compared to those for RCP2.6. This behavior can be directly attributed to the different GCM climate sensitivities of temperature and/or precipitation to increasing CO_2 (*Aich et al.*, 2014).

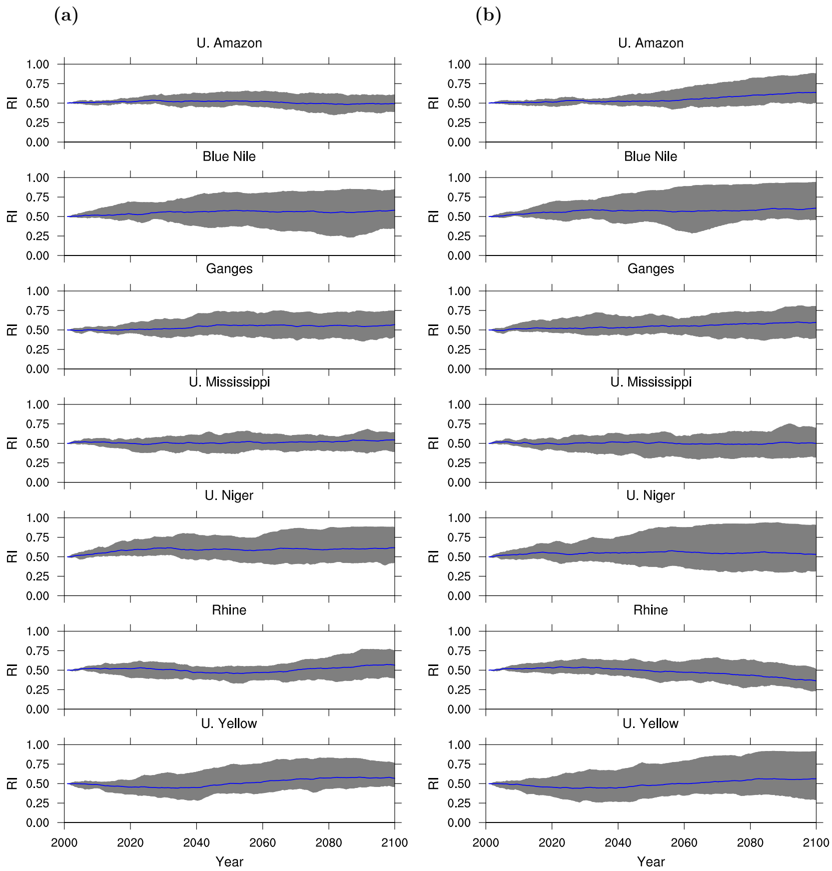

Figure 9.1 Multi-model ensemble of the 30-year running mean RIs for scenarios (a) RCP2.6 and (b) RCP8.5, for the period 1971-2099. Each running window is separated by a one-month interval (e.g., 1971-01 to 2000-12, 1971-02 to 2001-01, and so on). The blue line depicts the ensemble mean, and the gray bound denotes the ensemble uncertainty of the RI index.

The non-parametric Mann-Kendall test shows positive trends (at the 5% significance level) in the 30-year moving average RIs under RCP8.5 for more than 70% of the total model combinations (25) for the Upper-Amazon(20), Ganges(21), and Upper-Yellow (18) basins (Fig. 13.1). This implies that these basins will be, on average, wetter at the end of the century under RCP 8.5 than in the reference period. Three basins (the Blue-Nile, Upper-Mississippi, and Upper-Niger) show inconclusive results, which means that the number of model combinations that were positive or negative were comparable (the maximum difference did not exceed five). In contrast, for the Rhine basin, 24 out of the 25 model combinations agree that there is a consistent negative trend in RI (i.e., dryer than the reference period) for this scenario. The GCM projected precipitation over this basin does not exhibit a significant trend, but a strong increase in temperature is observed (*Mishra et al.*, 2016). This, in turn, affects the water balance by increasing evapotranspiration and reducing runoff. This drying trend in runoff is consistent with the recent findings in *Vetter et al.* (2015).

Under the RCP2.6 scenario, most model combinations show positive trends for the 30-year moving average RIs in five out of the seven basins: Upper-Amazon(19), Blue-Nile(16), Ganges(20), Upper-Mississippi(20), and Upper-Yellow (24) (Fig. 13.1). The remaining two basins show inconclusive results. In summary, under both RCP scenarios, the 30-year moving average RIs in the Ganges and Upper-Yellow basins exhibit positive trends. These trends are likely due to increases in precipitation. In both basins, all the GCMs project wetter than normal conditions under both scenarios (*Mishra et al.*, 2016).

The ensemble variability in the 30-year RIs corresponding to the RCP2.6 scenario is generally smaller than that exhibited for RCP8.5. This is a likely consequence of an increase in the uncertainties in the precipitation and temperature projections under RCP8.5 (*Mishra et al.*, 2016). Among the basins, a relatively lower spread in RI is observed in the Upper-Amazon, Ganges, and Rhine basins for both RCPs (Fig. 13.1). In contrast, comparatively large ensemble uncertainties are shown mainly for the Blue-Nile, Upper-Niger and Yellow river basins. These results show that the overall RI uncertainty arising from the combination of the GCM and HM uncertainties varies across the basins. For example, the Yellow River basin exhibits a small spread in the precipitation projections (*Mishra et al.*, 2016) among GCMs but a large uncertainty in the RI projections (Fig. 13.1). This example shows that the differences in the parameterizations of hydrological processes among HMs (e.g., evapotranspiration, snow melt) may lead to significant disagreements among HMs in some regions, as was also observed in previous studies (*Aich et al.*, 2014; *Pechlivanidis et al.*, 2016; *Prudhomme et al.*, 2014).

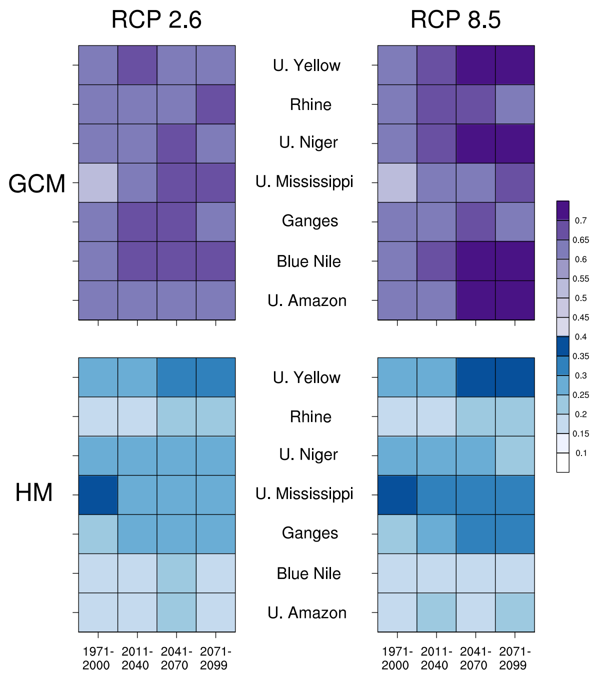

Figure 9.2 Mean range of RI ($\bar{R}$) depicting the contribution of the GCMs (top panels) and HMs (bottom panels) variability for four selected periods between 1971 and 2099 under scenarios RCP2.6 (a) and RCP8.5 (b). A lighter color implies small ensemble variability for a given group (e.g., HM or GCM). A darker color implies the opposite.

The summary of the uncertainty ranges of the RIs depicted in Fig. 13.2 shows that the long-term mean range of RI ($\bar{R}$) is time and RCP dependent. For a given RCP scenario, the mean range due to GCM variability for M HMs is estimated as $\bar{R} = \frac{1}{M n_T} \sum_t R_{mt}$, where R_{mt} denotes the range of RIs for the m HMs obtained by varying the GCMs (see Fig. 9-A.1). A 30-year interval is selected to capture climatological changes in the runoff index; therefore $n_T = 360$ months, and $M = 5$. Similarly, the mean range due to HM variability can be estimated. Four 30-year periods are selected for depiction: 1971-2000 (historical), 2011-2040 (near future), 2041-2070 (mid century), and 2071-2099 (end of the century). The GCM and HM contributions to the mean range of RI is shown in the upper and bottom panels of Fig. 13.2, respectively. A lighter color in the upper panel of Fig. 13.2, for example, indicates a small variability among the RIs obtained by fixing the HMs and varying the GCMs. Therefore, a good agreement among the GCM forcings should exist.

It can be concluded from Fig. 13.2 that the magnitude of $\bar{R}$ tends to increase with time. The proportion of the GCM uncertainty in RI over all the basins and time periods is at least 50% higher than that of the HMs under both RCPs. The GCM uncertainty contribution can be at least three times as much as those of the HMs. These results provide evidence that supports the working hypothesis that the uncertainty in RI due to GCM forcings largely dominates the HM-induced uncertainty. This result supports the findings of previous studies (*Bosshard et al.*, 2013; *Giuntoli et al.*, 2015; *Schewe et al.*, 2014; *Vetter et al.*, 2015).

It is worth noting that the contribution of the HM uncertainty, although small (0.15-0.40), cannot be neglected, especially in the context of drought analysis, considering that the threshold used to estimate drought characteristics is usually taken as 0.2 (*Samaniego et al.*, 2013; *Vidal et al.*, 2010). The variability in the ensemble RI ($\bar{R}$) is slightly higher in the RCP8.5 scenario compared to that of RCP2.6, for either the GCM or HM contributions (Fig. 13.2). On average, the estimated difference is smaller than 5% across all the basins and periods.

The 30-year moving average RI time series under the RCP8.5 scenario are used to depict the degree of coherence among the RIs estimated using the five HMs driven by a single GCM (GFDL-ESM2M) (Fig. 9-A.2a), and the lack of coherence between the RIs estimated using a given hydrological model (mHM) but forced with the five GCMs (Fig. 9-A.2b). Similar patterns are observed for all the HM and GCM combinations as well as those for RCP 2.6.

These results support the second postulation stated in the introduction that given a GCM forcing, the RI time series of every HM is practically indistinguishable or very close from their ensemble mean. The null hypothesis associated with this postulation can be safely rejected in 33% of all the model combinations (7 basins $\times$ 2 RCPs $\times$ 5 GCMs $\times$ 5 HMs) at the 5% significance level (based on the Studentized Bootstrap test). However, 90% of the model combinations are rejected in the case variability in the GCMs. Therefore, given this evidence, it is unlikely that the RI uncertainty due to HM variability would dominate over that of the GCM variability. The HM related uncertainty on RI is, however, not negligible.

Although the HMs have different conceptualizations for the dominant hydrological processes, their responses after transforming streamflow into RI for a given forcing are comparable. To a large extent, a quantile index such as RI can remove the systematic bias of every model simulation. This transformation leads to a large degree of coherence among RIs obtained from different HMs driven by the same climatic forcing (*Wang et al.*, 2009).

The variability in RI among the HMs is mainly due to the differences in the parameterizations of hydrological processes such as evapotranspiration, soil moisture accounting and runoff generation mechanisms. For example, models such as HBV and HYPE share many similarities with mHM with respect to runoff generation mechanisms, but the latter differs from the former ones by the number of soil layers, the soil moisture redistribution and root-water uptake process. WaterGAP, on the other hand, has a single soil layer and therefore has limits to account for the dynamics of soil moisture redistribution. It should be noted that HYPE and SWIM can handle water management activities but this capability was not implemented for the simulations used in this study. A detailed description of the major hydrological processes in the HMs is given in *Krysanova and Hattermann* (2016). Other factors that may lead to differences in the simulated RIs include spatial discretization, inadequate spin-up time, and deficient parameter estimations. Additionally, the method adopted to account for evapotranspiration, as discussed in *Sheffield et al.* (2013), plays a key role in estimating drought characteristics.

9.4.2 Propagation of forcing and model uncertainties on to drought magnitude and duration

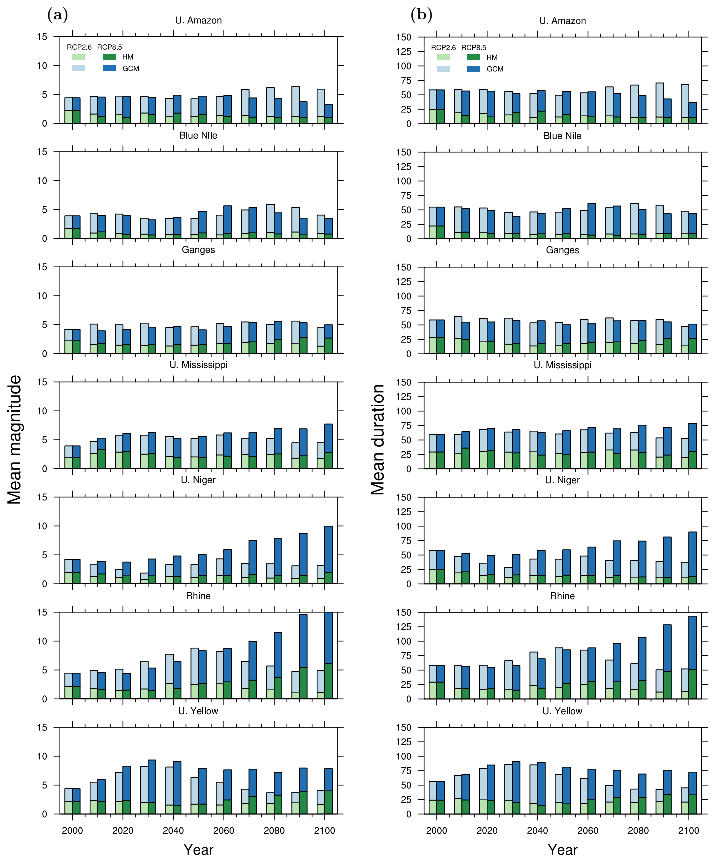

Figure 9.3 Ensemble mean drought magnitude (M) [% months] (a) and duration (D) [months] (b) corresponding to 25 model simulations (5 GCMs and 5 HMs) for a running window of 30-y for RCP2.6 and RCP8.5. The bar height correspond to the ensemble mean of the drought characteristic I (see Section 2.3). Each running window is separated by 10 year intervals (e.g., 1971-2000, 1981-2010, and so on). The contribution due to the HMs and GCMs variability is shown as stack bars. The HM and GCM contribution are estimated as $f_{HM} = \overline{R}_H/\left(\overline{R}_H+\overline{R}_G\right)$, and $f_{GCM} = 1 - f_{HM}$, respectively.

The 30-year ensemble means of the drought magnitudes and durations between 1971 and 2099 for both RCPs are shown in Fig. 13.3. The mean of each drought characteristics is estimated from the multi-model ensemble comprised of the 25 combinations of GCMs and HMs (Section 9.3.3). The uncertainties in these characteristics attributed to GCM and HM variability are depicted as stack bars. The results of this analysis clearly show that the uncertainty contribution of the GCMs on the drought characteristics outweigh those from the HMs, regardless of the hydro-climatic regimes and RCPs. Similar to the results obtained for the RI, the HM uncertainty contribution to the drought characteristics, although smaller compared to that of the GCMs, cannot be neglected. These results corroborate findings of previous studies (*Bosshard et al.*, 2013; *Giuntoli et al.*, 2015; *Vetter et al.*, 2015).

The share of the HM variability on the projected drought characteristics tends to remain constant and less important than that of the GCMs in the basins that exibhit strong streamflow annual cycles such as the Upper-Amazon, Blue-Nile, and Upper-Niger (stack bars in Fig. 13.3). This behavior is apparent under both RCP scenarios. In the midlatitude basins (e.g., the Upper-Yellow and Rhine), on the contrary, the uncertainty contribution of the HMs tends to increase by the end of the century, and reaches levels that are almost comparable to those of the GCMs.

Overall, among all the river basins, the Rhine exhibits the strongest increase in drought magnitude and duration under RCP8.5 (Fig. 13.3). Future climate projections from the 5 GCMs under the RCP8.5 scenario show a moderate increase in monthly precipitation with respect to the reference period 1981-2010 for the entire Rhine river basin. A change of 3.8% is expected for the period 2006-2035 and 1.0% for 2070-2099. Under scenario RCP2.6, those changes are approximately 2.3% and 6.6%, respectively. The monthly temperatures for those periods are expected to increase by 1.0 and 4.4°C under RCP8.5 compared to the reference period. However, under RCP2.6, the increases are 1.1 and 1.5°C (*Krysanova and Hattermann*, 2016; *Pechlivanidis et al.*, 2016). The changes due to increased temperature overcompensate the effect due to the increase in precipitation. These factors lead to a reduction in snow cover and an increase in evapotranspiration that would influence the hydrological regime of the Rhine causing a strong reduction in streamflow, and hence a higher increase in drought magnitude by the end of the century under RCP8.5. A contrasting effect (decreasing trend) is observed for RCP2.6 in this basin by the mid-century onwards, which is in line with the finding of *Pechlivanidis et al.* (2016). A second strongest increase in the drought magnitude under RCP8.5 is observed in the Upper-Niger basin due to the significant temperature increase by the end of the century (*Mishra et al.*, 2016).

On the contrary, the Upper-Yellow basins show declining trend in drought magnitude and duration by the end of the century, particularly under RCP8.5. This trend corresponds to the increasing trend in the RI time series for this basin, which may result from the projected increases in precipitation (*Mishra et al.*, 2016).

The results depicted in Fig. 9-A.3 further confirm the dominance of the GCM uncertainties over those of the HMs for drought characteristics. The darker blue lines that denote the GCM contributions are always above the lighter blue lines that denote the HM contributions for both RCPs. In this case, the mean range $\bar{R}$ of the 30-year drought magnitude and duration is used as a measure of the uncertainty contributions from the GCMs and HMs (see Section 9.3.3).

It is worth noting that the propagation of uncertainty from the RI to the mean drought magnitude and duration is nonlinear. In general, an increase in uncertainty over time is observed for both characteristics. To enable a comparison between the uncertainty estimates for RI and the drought characteristics, the ratios between their respective ranges and mean values are estimated. Based on this statistic, an incremental change of approximately 60% on average was estimated for the propagation of uncertainty from the RI to drought magnitude under RCP8.5 for the three projection periods shown in Fig. 13.2. No systematic pattern in this statistic is detected among those periods. The reason behind the increase is the non-linear behavior of the truncation level method used to estimate drought magnitude (and duration) based on the RI times series. Note that a drought can occur only when RI falls below a certain threshold (say 0.2). As a consequence, the uncertainty in the drought characteristics cannot be inferred based only on the uncertainty in RI.

The results shown in Fig. 9-A.3 also indicate that the uncertainty in the drought characteristics due to GCM and HM variability increases with time. Among the two scenarios, the uncertainty in the drought characteristics increases as the radiative forcing increases. This is a consequence of the relatively larger uncertainty in the projection of the GCM forcings under RCP8.5 (*Mishra et al.*, 2016). To some extent, the increase in the uncertainty due to the HM variability with time could be related to the departure from the reference calibration conditions.

The relative contribution of HM uncertainty on to drought characteristics depends on the hydro-climatic regime, which differ among the basins (Fig. 9-A.3). The HM uncertainty tends to be larger in snow-dominated regions and lower when the climatic conditions are wet, which is in agreement with the findings of *Giuntoli et al.* (2015); as an example, refer to the lighter blue lines in Fig. 9-A.3 for the Upper-Mississippi, Rhine, and Upper-Yellow basins in comparison to those of the Upper-Amazon and Ganges basins, which are mainly characterized by a humid climate.

9.4.3 Expected changes in the CDF of the drought magnitude

Fig. 13.4 presents the cumulative distribution function (CDF) of the estimated drought magnitude based on all 25 model combinations for the two 30-year windows corresponding to the historical (1971-2000) and the end of the century (2071-2099) periods. The results shown in this figure provide supporting evidence that the projected drought characteristics are RCP dependent. Based on the Kolmogorov-Smirnov test, the null hypothesis associated with this postulation can be safely rejected at the 1% level of significance for the Blue-Nile, Upper-Niger, Rhine, and Upper-Yellow river basins and at the 5% level in the Upper-Mississippi basin. For the Upper-Amazon and Ganges, the null hypothesis can be rejected at the 10% level of significance. The tests are conducted based on the CDFs from RCP2.6 and RCP8.5 for the end of the century. The analysis of the drought duration (not shown) shows similar tendency as that of the magnitude.

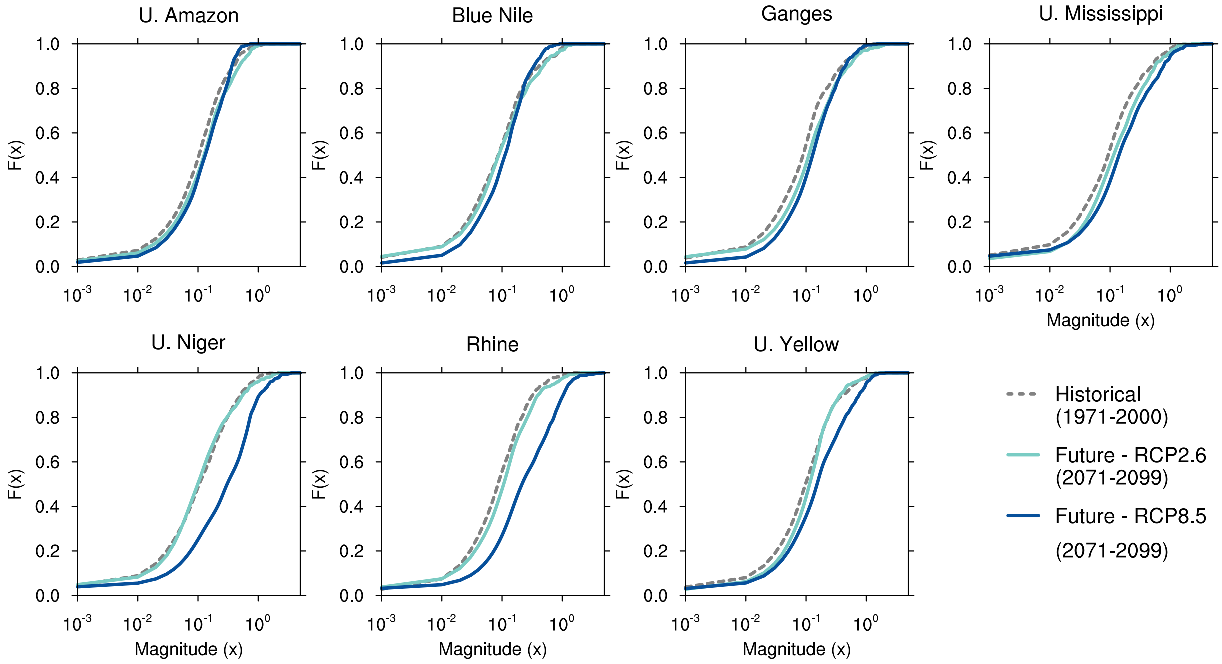

Figure 9.4 Cumulative distribution functions of drought magnitude for historical and future periods under scenarios (a) RCP2.6 and (b) RCP8.5 based on all 25 model combinations.

The propagation of the GCM related uncertainty for a given RCP scenario on to a drought characteristic depends on the hydro-climatic regime of the basin. The Upper-Niger and the Rhine basins under RCP8.5, for example, exhibit the largest increases in drought magnitude by the end of century compared to the historical estimates. The Upper-Amazon, on the contrary, does not show such a large deviation. Despite these deviations, the drought magnitude estimated for the 30-year period at the end of the century under RCP8.5 is on average higher than that of the historical period. This implies that, on average, more severe droughts can be expected in the study basins by the end of the century under the RCP8.5 scenario. For RCP2.6, not all the basins exhibit deviations in future drought magnitude compared to the historical ones.

9.5 Conclusions

In this study an attempt is made to understand the propagation of forcing and model uncertainties on to century-long time series of drought characteristics using an ensemble of five hydrological model (HM) projections across a range of climate scenarios and regions. The models in seven large scale river basins are driven by an ensemble of five GCMs under RCP2.6 and RCP8.5 for the period 1971-2099.

The following conclusions are derived based on the obtained results. 1) The GCM uncertainties dominated the HM uncertainties for the runoff index and derived drought magnitude and duration, irrespective of the RCPs and studied basins. The ranges of the drought characteristics due to the GCMs and HMs uncertainties were, however, basin (hydroclimatic regime) specific. 2) The uncertainties due to the HMs, although smaller than those of the GCMs, cannot be neglected for hydrological drought projections. Therefore, multi-model drought assessments should consider an ensemble of HMs and GCMs. 3) By the end of the century under RCP8.5, most of the study basins would probably endure, on average, higher magnitude droughts compared with those of the reference period (1971-2000). In particular, a comparatively larger propensity to hydrological droughts is observed in the Upper-Niger and Rhine basins.

We acknowledge that the selection of GCMs, HMs and basins may not fully encompass the entire variability needed to generalize our results. A comparison across regions that have a range of consensus in future projections of precipitation and temperature could be a subject of further studies. This study, however presents a method to analyze the effects of the dominant sources of uncertainty for hydrological drought projections. We encourage that future studies should consider more hydrologically diverse river basins, GCMs, HMs and other sources of uncertainty for climate change impact assessments, including HM parametric uncertainties, downscaling and bias corrections of GCM forcing. Other essential hydrological variables such as soil moisture and groundwater levels should be considered for regional drought assessments.

Acknowledgements

This study was carried out within Helmholtz-Association climate initiative REKLIM (www.reklim.de). This study was conducted under the Inter-Sectoral Impact Model Inter-comparison Project. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling for the CMIP5 simulations.

Appendix

Estimation of the runoff index RI

The runoff index RI is estimated as $\mathrm{RI_t} = \hat{\mathrm{F}}^{-1}(\mathrm{x_t})$, where, x_t is the simulated monthly streamflow at time t and $\hat{F}^{-1}$ is the inverse of the empirical distribution function associated with the kernel density estimator $\hat{f}(x)$ of the corresponding calendar month at the time t. $\hat{f}(x)$ is estimated as

$$\hat{f}(x) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{x - x_k}{h}\right) \tag{A.1}$$

where $x_1, \ldots, x_n$ denote the streamflow values corresponding to a given calendar month for the entire reference period, n is the sample size, and K denotes a Gaussian kernel function with a bandwidth of size h.

The bandwidth h is estimated by minimizing a cross-validation error estimate (see *Samaniego et al.* (2013) for details) for the historic period (1971-2000) separately for every calendar month, basin, HM, and GCM to allow comparability across the time, space and model.

It should be noted that the standardized runoff index (SRI) proposed by *Shukla and Wood* (2008) is equivalent to the RI because both indices denote percentiles associated with the monthly streamflow time series. The main difference between these two indices is that the SRI is obtained by fitting a gamma density function to the streamflow sample and then expressing the corresponding quantiles as standard-normal deviates. The RI proposed in this study, simplifies the procedure by estimating the distribution of streamflow with a non-parametric kernel density function (eq. A.1).

Estimation of the drought characteristics

Total drought magnitude M and duration D are derived based on the monthly RI time series. The total magnitude of drought events occurring within a period T is defined as the temporal integral of the RI below a threshold value τ. It is estimated by

$$M(T) = \sum_{t=t_0}^{t_1} \Delta t \big(\tau - \mathrm{RI}_t\big)_+ \tag{A.2}$$

where, t_0 and t_1 denote the onset and the ending months of the period T, respectively. Therefore, the length of the window T is $t_1 - t_o + 1$ months. The expression $(\cdot)_+$ is the positive part function and Δt denotes the time step in months. M is expressed in [% months] (*Sheffield and Wood*, 2008a).

The duration of drought spells within the period T is therefore the total number of months in which the RI is below a threshold value τ. It is estimated by

$$D(T) = \sum_{t=t_0}^{t_1} \mathbf{1}_{\mathrm{RI_t} \leq \tau} \tag{A.3}$$

where, $\mathbf{1}(\cdot)$ denotes the indicator function (i.e., one if the condition is true and zero otherwise) to estimate the total occurrence of drought affected months during the period T.

Sequential sampling algorithm

The sequential sampling algorithm to evaluate the contribution of the GCM variability for a given basin and RCP is listed below and visualized in Fig. 9-A.1.

1. Select a RCP scenario (c).
2. Select a HM (m), $m = 1, 2, \ldots, M$.
3. Select a GCM (g), $g = 1, 2, \ldots, G$.
4. Estimate the monthly streamflow for the entire simulation period (P_0 to P_1) and transform it to the RI time series.
5. Estimate a drought characteristic I_{cmg} (Section 9.3.2) over a rolling window of size T, with steps of size S.
6. Repeat steps 3-5 for all the GCMs or resample the GCMs N times with replacement (* see details below).
7. Estimate the range of I_{cmg} over the GCMs for every rolling window, denoted as $R_{cm}(T)$.
8. Repeat steps 2-7 for all the HMs.
9. Estimate $\overline{R}_{cG}(T) = \frac{1}{M}\sum_m R_{cm}(T)$ for every rolling window T to represent the uncertainty of the drought characteristic I due to the variation in the GCMs.

Note (*): Sampling with replacement can be used to generate confidence intervals for the range statistics (e.g., $\overline{R}_{cG}(T)$). The number of realizations should be large enough to obtain reliable statistics (*Davison and Hinkley*, 1997). In the present study, N=1000 is sufficient for convergence.

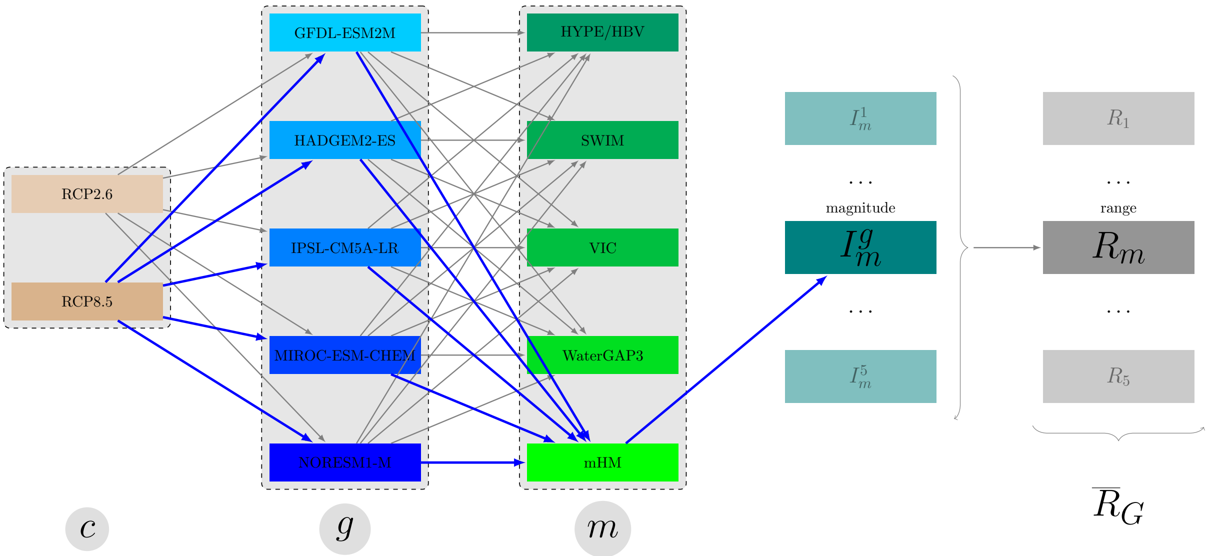

Figure 9-A.1 Graphical representation of the algorithm presented in Section 9.5 to estimate the GCM and HM uncertainty contributions from a given emission scenario c. Here, for example, it is visualized the estimation of the GCM uncertainty component $\overline{R}_{cG}(.)$ under RCP8.5 (c = 2). The subindex c and the rolling window T are dropped to improve readability. The GCM index g is written as superscript for the same reason.

Here, M and G denote the sample size for the HMs and GCMs, respectively (in this study, $M = G$ =5). P_0 and P_1 denote the start and end of the simulation period (i.e., 1971 and 2099, respectively). T is the length of a rolling window in years. It should be large enough to capture the climatological changes (e.g., T=30). S denotes the shift between two consecutive rolling windows in years. For example, if T=30 and S =10, as used in this study, $R_{cm}(T)$ is estimated for the following rolling windows: 1971-2000, 1981-2010, ... , and 2071-2099. A graphical representation of this algorithm is shown in Fig. 9-A.1.

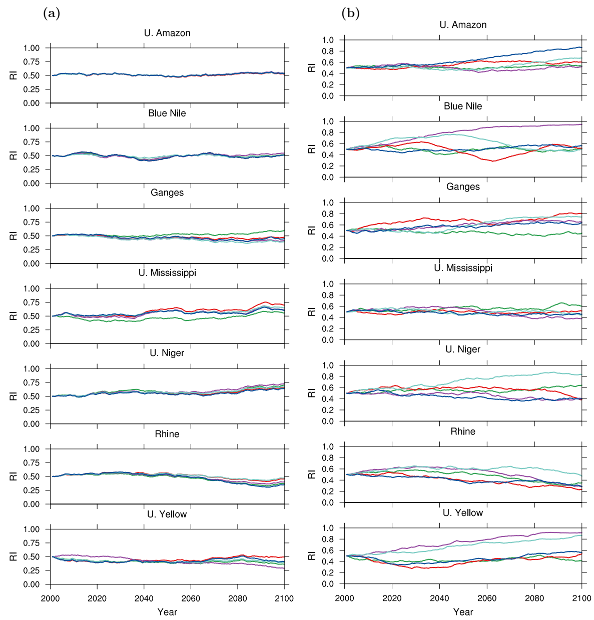

Figure 9-A.2 A 30-year running mean of monthly RI for (a) five HMs forced by the GFDL-ESM2M model and (b) the same index from the mHM model forced with five GCMs. Both cases consider RCP8.5 forcings for the period from 1971 to 2099. Each rolling window is separated by one month interval (e.g., 1971-01 to 2000-12, 1971-02 to 2001-01, and so on). Panel (a): green = HYPE/HBV, red = SWIM, violet = VIC, light blue = WaterGAP3, blue = mHM. Panel (b): green = GFDL-ESM2M, red = HADGEM2-ES, violet = IPSL-CM5A-LR, light blue = MIROC-ESM-CHEM, blue = NORESM1-M.

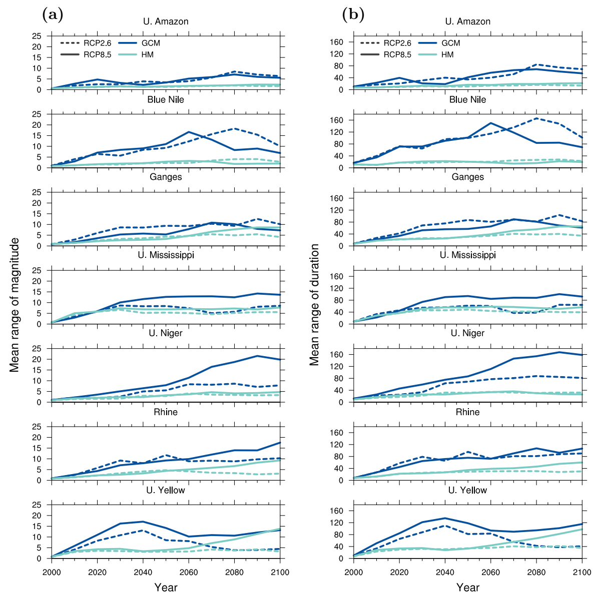

Figure 9-A.3 Mean range of drought magnitude [% months] (a) and duration [months] (b) depicting the contribution of the HMs ($\overline{R}_H$) and GCMs ($\overline{R}_G$) variability for a running window of 30-y for RCP2.6 and RCP8.5. Each running window is separated by 10 year intervals (e.g., 1971-2000, 1981-2010, and so on).

CHAPTER 10

HYDROLOGICAL FORECASTS AND PROJECTIONS FOR IMPROVED DECISION-MAKING IN THE WATER SECTOR IN EUROPE

L. Samaniego[1] et al.

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Samaniego, L., S. Thober, N. Wanders, M. Pan, O. Rakovec, J. Sheffield, E. F. Wood, C. Prudhomme, G. Rees, H. Houghton-Carr, M. Fry, K. Smith, G. Watts, H. Histal, T. Estrella, C. Buontempo, A. Marx, and R. Kumar (2019a), Hydrological forecasts and projections for improved decision-making in the water sector in Europe, Bull. Am. Meteorol. Soc., 100, 2451–2472

10.1 Abstract

Simulations of water fluxes at high spatial resolution that consistently cover historical observations, seasonal forecasts, and future climate projections are key to providing climate services aimed at supporting operational and strategic planning, and developing mitigation and adaptation policies. The EDgE is a proof-of-concept project funded by the Copernicus Climate Change Service programme that addresses these requirements by combining a multi-model ensemble of state-of-the-art climate model outputs and hydrological models to deliver Sectoral Climate Impact Indicators (SCIIs) co-designed with private and public water sector stakeholders from three contrasting European countries. The final product of EDgE is a water-oriented information system implemented through a web application. Here, we present the underlying structure of the EDgE modeling chain, which is composed of four phases: 1) climate data processing, 2) hydrological modeling, 3) stakeholder co-design and SCII estimation, and 4) uncertainty and skill assessments. Daily temperature and precipitation from observational data sets, four climate models for seasonal forecasts, and five climate models under two emission scenarios are consistently downscaled to 5 km spatial resolution to ensure locally relevant simulations based on four hydrological models. The consistency of the hydrological models is guaranteed by using identical input data for land surface parameterizations. The multi-model outputs are composed of 65 years of historical observations, a 19-year ensemble of seasonal hindcasts, and a century-long ensemble of climate impact projections. These unique, high-resolution hydro-climatic simulations and SCIIs provide an unprecedented information system for decision-making over Europe and can serve as a template for water-related climate services in other regions.

Capsule Summary

Development of a high resolution multi-model ensemble of state-of-the-art climate and hydrological models to deliver hydro-meteorological change metrics co-designed with key water sector stakeholders in Europe.

10.2 The rationale behind EDgE

Existing water-oriented decision support systems are either designed as early-warning systems to provide forecasts of hydrological floods and droughts, or as monitoring platforms aiming to provide information on the current state of variables of interest such as streamflow or soil moisture. Such systems normally target national-, continental-, and global-scales with examples of systems developed for Australia (*Emerton et al.*, 2016), Africa (*Sheffield et al.*, 2014), Europe (EFAS, *Thielen et al.*, 2009), and North America (*Demargne et al.*, 2014).

Whilst the multiplication of water-focused climate services undeniably helps downstream decision making, a number of areas of improvements can be identified. First, only few systems operate at different prediction horizons, most focusing on a single function targeting monitoring, short to medium-range forecasting, seasonal or climate time scales. This means that users need to refer to different systems and services depending on their planning scale, each generally associated with different and inconsistent types of information and delivered services, hence requiring users to develop different application tools for each independent one, and to take much care when interpreting their different outcomes, as they might not provide exactly the same information. Two notable exceptions are the European Flood Awareness System (*Arnal et al.*, 2018; *Thielen et al.*, 2009) and the Global Flood Awareness System (*Alfieri et al.*, 2013), both components of the Copernicus Emergency Management Service, which are currently the only operational forecasting systems providing ensemble streamflow forecasting and flood early warning at both medium range and seasonal time frames for Europe and the world, respectively; however, none currently offer predictions beyond a few months, limiting their use to short and medium term planning.

Second, existing continental or global-scale systems (e.g., for Africa, *Sheffield et al.* (2014); North America, *Lawrimore et al.* (2002); or Europe, *Horion et al.* (2012)) typically operate at resolutions of 0.25° or coarser (note the European Drought Observatory has a multi-scale approach) and do not quantify the uncertainty associated with the monitoring and forecasting hydro-meteorological chain. At national scale, monitoring systems are typically available at a high spatial resolution (e.g., the German Drought Monitor at 4 km scale, `www.ufz.de/droughtmonitor`, *Zink et al.* (2016); or the UK Drought portal at 5 km scale, `https://eip.ceh.ac.uk/apps/droughts/`) and provide timely information for decision-making and the general public, whilst high resolution national forecasting services are normally based on a single hydrological model, not capturing some of the important uncertainty. This limits robust local decision-making as single deterministic estimates are often given instead of probabilistic ones.

Third, most climate service portals, especially when designed for climate projections, only focus on a few specific climate-related variables. For example, the Royal Netherlands Meteorological Institute Climate Explorer (`climexp.knmi.nl/`) is a web application to visualize and analyze global climate data, but does not include hydrological-derived indices relevant for the water sector (see e.g., *National Research Council*, 2001). This can limit downstream applications because the information relevant for local decision making is absent and would require further processing (and associated resources) by users, thus, hampering potential uptake.

The project EDgE "End-to-end Demonstrator for improved decision-making in the water sector in Europe" (EDgE, `https://youtu.be/PqoRi6eSM2w`) was designed to address three specific gaps in existing climate services for a more user-focused delivery for the water sector: 1) development of a high-resolution (5 km), multi-model system (using two land surface models (LSM) and two hydrological models (HM) established with a common set of land surface properties across continental Europe) where both uncertainty in the atmospheric forcing and hydrological impact are accounted for and summarized to users; 2) delivery of a consistent and comparable multi-scale sets of seasonal forecasts and climate impact projections based on the same hydro-climate modeling chain; and 3) provision of 36 climate and water indicators (so-called sectoral climate impact indicators SCIIs) and a unique web information service co-designed with over 30 different public- and private-sector stakeholders from different hydro-climatic regions and water-related industries across Europe, to facilitate the uptake of the service. The rationale of EDgE is simple: better-informed operational and strategic planning decisions can be made only with a timely, coherent and co-designed water-oriented information system (*Lourenço et al.*, 2015). EDgE was one of two proof-of-concepts (PoC) for the water sector for the European Copernicus Climate Change Service Sectoral Information Systems programme.

In this paper, we present briefly the technical features of the EDgE modeling chain, discuss the continental scale application of the underlying model components, including the forcing data, and discuss the co-design and interpretation of impact indicators. Although the stakeholder feedback process and the development of visualization tools were fundamental parts of the PoC, they are not covered here. For completeness, note that EDgE was a proof-of-concept project and does not provide a real-time service; however, all indicators estimated under both seasonal forecast and climate impact prediction mode are accessible at `http://edge.climate.copernicus.eu`.

10.3 Climate data processing

Daily temperature and precipitation from the E-OBS 25-km gridded product (v12) and the underlying station data (*Haylock et al.*, 2008) were used as historical meteorological forcing data for the period 1950 to 2015. In addition, gridded daily wind speed was obtained from the European Flood Alert System forcing (*Thielen et al.*, 2009), made available by the European Centre for Medium-Range Weather Forecasts (ECMWF) for the period 1990–2014. Gridded historical observations were used to develop the hydrological model historical simulations, to bias correct the future climate prediction forcings and to generate the hydrological initial conditions for the seasonal hindcasts (i.e., forecasts of past periods).

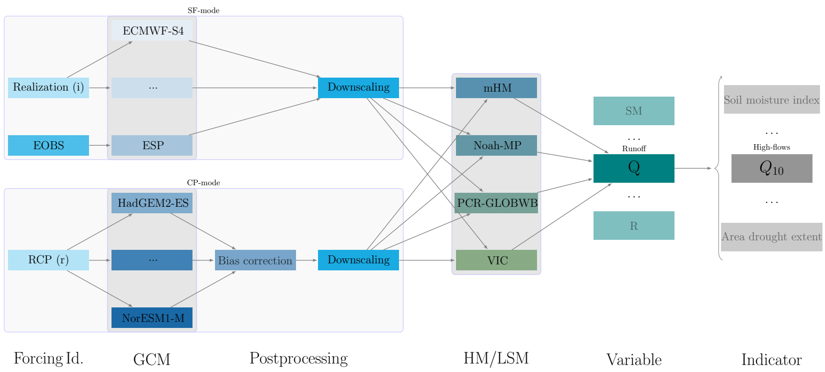

Figure 10.1 EDgE modeling chain for seasonal forecast (SF) and climate prediction (CP) modes, shown here for the high flow indicator Q_{10}. Both chains use four hydrological models to compute values for the terrestrial Essential Climate Variables (tECVs) which are the basis to estimate the sectorial climate impact indicators (SCIIs) requested by focus groups. For the SF-mode the number of climate realizations can vary across models ($i = 10$ to 15). For the CP-Mode, the representative climate pathway (RCP) are set to RCP2.6 an RCP8.5 ($r = 1{,}2$). ESP is included as benchmark for the dynamic SF models.

For the climate change simulations, daily temperature and precipitation from five bias-corrected Global Climate Models (GCMs) from the Coupled Model Intercomparison Project Phase 5 (CMIP5: HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M and NorESM1-M) were used to drive the HM/LSMs during the period from 1950 to 2099 under two Representative Concentration Pathways (RCPs; RCP2.6 and RCP8.5, see CP-mode in Figure 10.1). This data set was made available by the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, *Warszawski et al.*, 2014) at a spatial resolution of 0.5°, and was selected as it benefited already from a trend-preserving bias-correction ((*Hempel et al.*, 2013); see further detail in Appendix A).

For the seasonal forecast simulations, daily temperature and precipitation hindcasts from four GCMs run in seasonal forecast mode (SF-GCM) were used to drive the HM/LSMs (see SF-mode in Figure 10.1). These comprise two models from the North-American Multi-Model Ensemble (NMME: Canadian Climate Model version 4 CanCM4 and the Geophysical Fluid Dynamics Laboratory Forecast-oriented Low Ocean Resolution model (GFDL-FLOR) and two European models from the Copernicus Climate Change Service (C3S: ECMWF system 4 (ECMWF-S4) and the Météo-France modeling system version 5 (LFPW)). The number of NMME SF-GCMs were selected to counterweight those provided by the ECMWF. The number of realizations among SF-GCMs were chosen so that each member become a similar weight in the multi-model ensemble. In total, the EDgE SF-GCM multi-model ensemble contained 52 realizations comprised of 10 members from CanCM4, 12 members from GFDL-FLOR, and 15 members each from ECMWF-S4 and LFPW. Daily hindcasts starting on the first day of each month within the hindcast period 1993–2011 were used. The SF-GCMs forcings were downscaled from their native spatial resolution (of 1° for NMME models and 0.75° for C3S models) to the hydrological model resolution (see section below) without previous drift or bias correction, the relatively short hindcast period (1993–2011) being considered not sufficiently long to train a robust bias correction algorithm.

The spatial resolution of all forcing data was considered too coarse for deriving water sector indicators relevant for water managers and practitioners at local and regional levels in Europe. Instead, a spatial resolution of $5{\times}5$ km^2 was selected to derive all EDgE products as a trade-off between a spatial resolution that is informative for practitioners across Europe and a scale at which it is still feasible to estimate water-related variables using current computational facilities and geophysical forcing information. The forcing data were hence downscaled from their native resolution to the common $5{\times}5$ km^2 resolution prior to their use as input to the hydrological modeling chain. Details of the statistical downscaling technique is provided in Appendix A.

10.4 Hydrological modeling

Hydrological and land surface models (HM/LSMs) are the backbone of the EDgE modeling chain (Figure 10.1); they comprise: the mesoscale Hydrological Model (mHM); the PCRaster Global Water Balance model 2 (PCR-GLOBWB); the Variable Infiltration Capacity model (VIC); and the Noah Land Surface model with multi-parameterization option (Noah-MP). All four models are process-based, simulating canopy interception, snow accumulation and melting, infiltration, evapotranspiration, and runoff generation. They were selected based on the diversity of their underlying process representations and their wide-spread use in hydrological applications to capture as much as possible the structural uncertainty within the hydrological modeling component. A summary of the land surface data used to parametrize these models up can be found in the Appendix B.

mHM (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a) is a grid-based distributed hydrological model equipped with a multiscale parameter regionalization scheme, developed with a special focus on running seamlessly at multiple spatial resolutions ranging from 1 km to 50 km (*Kumar et al.*, 2013b; *Rakovec et al.*, 2016c; *Samaniego et al.*, 2017), and ready to be implemented in an operational setting (*Kauffeldt et al.*, 2016). PCR-GLOBWB (*Sutanudjaja et al.*, 2018; *van Beek et al.*, 2011; *Wanders and Wada*, 2015) is a grid-based hydrological and water resources model, developed to represent the terrestrial water cycle at global and continental scales, with a special emphasis on including human water uses. VIC (*Cherkauer et al.*, 2003; *Liang et al.*, 1994) is a macro-scale hydrological model that solves full water and energy balances to represent the land-surface hydrology and near-surface atmospheric fluxes. The VIC model has been implemented in catchment to global scale applications for understanding catchment behavior, extreme hydrological events, hydrological predictability, and climate change impacts (a.o., *Sheffield and Wood*, 2008b; *Sheffield et al.*, 2014; *Yuan et al.*, 2015). The Noah-MP model provides several upgrades of the Noah LSM (*Niu et al.*, 2011), which was originally developed as the land-surface scheme for numerical weather prediction (*Ek et al.*, 2003). In this study, we use the same process parametrizations as in *Cuntz et al.* (2016).

Within EDgE, all four HMs/LSMs were established using the same high resolution (500 m) morphologic, land cover, and soil databases (Appendix B). Differences between models originate only from different process representations. All HMs/LSMs are setup at a spatial resolution of 5 km, simulate daily water fluxes and states, and were calibrated with standard procedures described in Appendix C. These models were subsequently evaluated in hundreds of river basins across Europe that cover a wide range of hydro-climatological regimes. For more details, refer to Figure10-A.1.

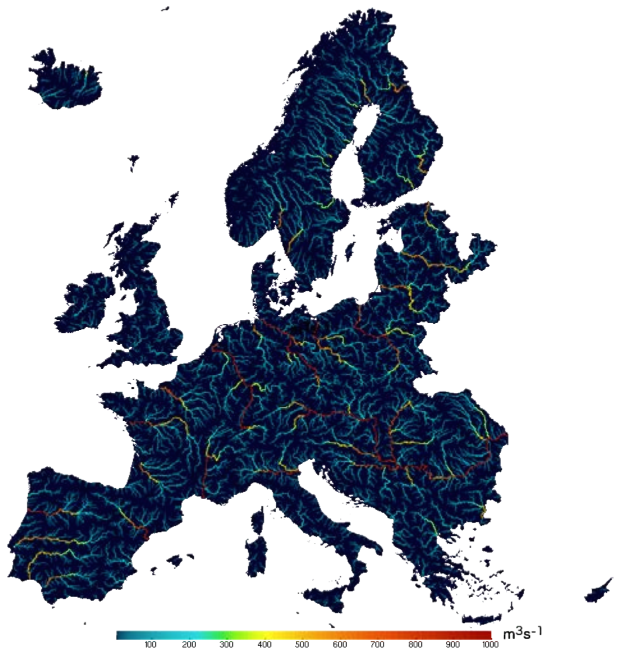

Figure 10.2 Simulated mean daily streamflow from 1950 to 2011 for the EDgE domain at 5 km spatial resolution. Simulated daily runoff was obtained with the mHM model forced with downscaled E-OBS forcing data. River routing was carried out with the mRM algorithm.

Another hallmark of the hydrological modeling chain is the use of a common river routing scheme to minimize predictive uncertainty from inconsistencies in the channel network. The gridded runoff fields generated by the four HMs/LSMs are routed through the same 5 km river network using the multi-scale routing model (mRM, *Thober et al.*, 2019b) that was originally developed for mHM (*Samaniego et al.*, 2010a). mRM has the ability to simultaneously route cell-generated runoff to multiple outlets, allowing streamflow to be generated over the entire domain simultaneously (Figure 10.2). The key characteristic of mRM is its capacity to estimate streamflow at various spatial resolutions without recalibration of river-routing parameters within the employed Muskingum scheme. This simplification of the Saint-Venant equations that only accounts for wave advection and attenuation is justified in all hydrological models used in this project because the river reaches within the domain are unlikely to exhibit abruptly changing hydrographs with supercritical flows over large distances (see details in *Thober et al.* (2019b) and references therein).

10.5 SCII: stakeholder co-design and estimation

The general graphical representation of the hydro-meteorological modeling chains used to generate terrestrial Essential Climate Variables (tECVs, *Sessa and Dolman*, 2008) for climate predictions and seasonal forecasts is shown in Figure 10.1. Four tECVs (Table 10.1) were stored from each simulation for the historical, climate prediction and seasonal forecast modes. These are: streamflow (Q), soil moisture in the top 2 m (SM), snow water equivalent (SWE), and groundwater recharge (R).

Table 10.1 Sectoral Climate Impact Indicators (SCIIs) derived from terrestrial Essential Climate Variables (tECVs): Streamflow (Q) ($m^3 s^{-1}$); top 2 m soil moisture as fraction of saturation (SM) ($m^1 m^{-1}$), groundwater recharge (R) (mm d^{-1}) and snow water equivalent (SWE) (m) and meteorological forcing data: potential evapotranspiration (PET) (mm d^{-1}), precipitation (P) (mm d^{-1}), daily average temperature (T) (°C). X denotes any of the tECVs. X_p denotes the value of X that is equaled or exceeded $p\%$ of the time over a time horizon. Lead time is denoted by ℓ.

Index type	Statistic	Time Horizon	tECV Variables	Notes
		Climate Predictions		
Relative change	Daily X_{10}	30 y	Q,R	Relative change in high values (X_{10}) w.r.t the reference period.
Relative change	Median annual X_{max}	30 y	Q	Peak values. w.r.t the reference period.
Relative change	Daily X_{90}	30 y	Q,R	Low values. w.r.t the reference period.
Relative change	Daily X_{95}	30 y	Q,R	w.r.t the reference period.
Relative change	Monthly mean X	1 month	Q,R, PET,P,SWE,T	w.r.t the reference period. For each calendar month.
Relative change	Seasonal mean X	3 month	Q,R,PET,P,SWE,T	w.r.t the reference period. Seasons considered DJF, MAM, JJA, SON.
Relative change	Annual mean X	1 y	Q,R,PET,P,SWE,T	w.r.t the reference period.
Percentile index	Monthly $F(X)$	Monthly	SM,R	$F(X)$ indicates the percentage of the time that X at a given location and point in time will take a value less than or equal to X
Duration	$F(X) < 0.2$	Monthly	SM	Number of consecutive months over a 30-year window in which the $F(X) < 0.2$, which indicates the onset of a moderate drought.
Relative change	Area $F(X) < 0.2$	Monthly	SM	Relative change of area of a basin $F(X) < 0.2$ w.r.t. a ref. period.
		Seasonal Forecast		
Probabilistic	Monthly quintiles, X_p	$1, \ldots, 6$	Q,SM,R	% of realizations of monthly forecasted X for every quintile category and lead-time. Cutoffs 20th, 40th, 60th, 80th percentiles.
Probabilistic	Above X_{10}	$1, \ldots, 6$	Q,SM,R	High values. % of realizations above the reference monthly X_{10} for each ℓ
Probabilistic	Below X_{90}	$1, \ldots, 6$	Q,SM,R	Low values. % of realizations above the reference monthly X_{90} for each ℓ

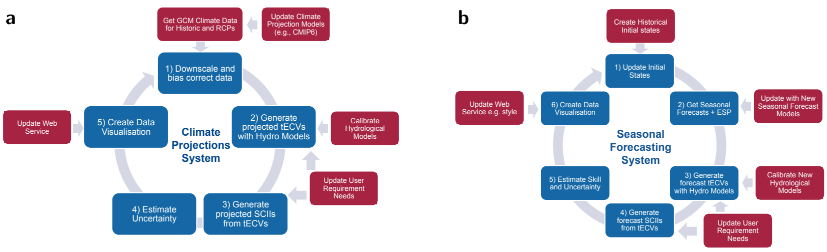

Figure 10.3 Workflow of the Climate Projection (a) and Seasonal Forecast (b) model chains. Both chains include occasional (red) and routine tasks (blue).

In the Climate Projections (CP) chain, all four HMs/LSMs were forced with the downscaled GCM data for the period from 1950 to 2099 under RCP2.6 and RCP8.5. The period 1971 to 2000 selected by the IPCC was adopted to represent present-day conditions (*Hoeg-Guldberg et al.*, 2019) based on the historical HMs/LSMs simulations driven by the GCM data sets. The workflow of the operationalization of the CP-chain includes the following steps (Figure 10.3(a): a) obtain GCM projections, b) perform bias correction (in this case, this step was not necessary because the ISI-MIP forcing data are already bias-corrected.) and downscaling, c) run the multi hydro-meteorological modelling chain to generate an ensemble of target variables (tECVs), d) generate an ensemble of indicators (SCIIs) based on the tECVs, e) estimate uncertainty in the ensemble, and f) export ensemble outputs/indicators for visualization. This workflow also includes occasional re-configurations, such as re-calibration of the HMs/LSMs, modification of the SCIIs according to user needs, and updates of the web service.

In the Seasonal forecast (SF) chain, all four HMs/LSMs were forced monthly with the downscaled SF-GCM 7-month hindcast data representative of the period 1993–2011. The SF-chain differs from the CP-chain because it requires initialization of state variables for each monthly simulation. The workflow of the SF-chain is depicted in Figure 10.3(b), with monthly operations (blue) carried out on a routine basis, and occasional operations (red). The main steps of the operational SF-chain are: a) update initial state variables through restart files obtained from the respective E-OBS reference historical run, b) obtain and downscale SF-GCM data, c) run the multi hydro-meteorological modelling chain to generate an ensemble of tECVs, d) estimate an ensemble of SCIIs based on the tECVs, e) estimate the ensemble forecast uncertainty and skill, and f) export ensemble outputs/indicators for visualization. Occasional tasks include the creation of historical initial states for the HMs/LSMs and the re-calibration of the HMs/LSMs.

The EDgE Sectoral Climate Impact Indicators (SCIIs) for the water sector were identified from an open survey conducted at the onset of the project (see Appendix E). The users —which included members from consultancy, academia, NGOs, water user associations, local, regional and national authorities— were asked for the information that they would require from a climate service delivering seasonal forecasts (SF) and long-term climate projections (CP). The end-users requested SF-based indices for lead times from 1 to 6 months, and CP-based indices for each decade up to 2100. To restrict the number of projections, two Representative Concentration Pathways (RCPs) were chosen, that define the lowest (RCP2.6) and highest (RCP8.5) emission scenarios. End-users were also interested in obtaining information regarding relative changes compared to the baseline 1971–2000.

The selected SCIIs for the CP and SF modeling chains are listed in Table 10.1. SCIIs based on forcing variables (i.e., precipitation, temperature and potential evapotranspiration) were derived directly from the bias-corrected and downscaled forcing data sets.

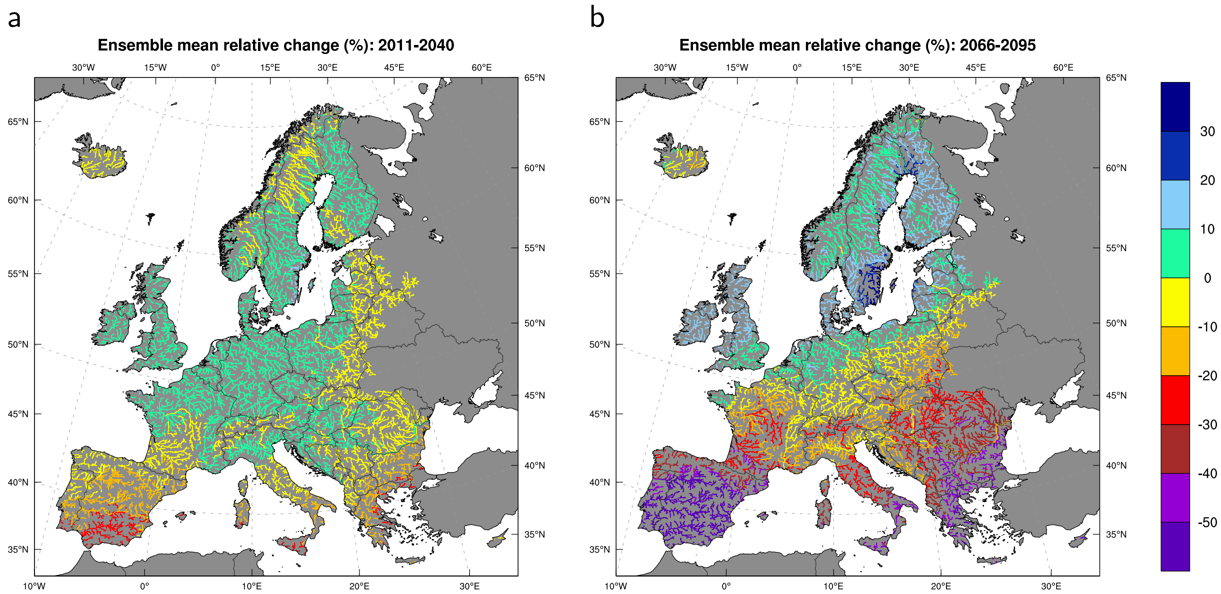

Figure 10.4 Multimodel ensemble mean of the projected changes in a high streamflow indicator (Q_{10}) for two future time periods: (a) 2011–2040 and (b) 2066-2099, both under the RCP8.5 scenario. Q_{10} is the daily streamflow equaled or exceeded 10% of the time over a 30-year window. The historical reference period is from 1971 to 2000.

The majority of CP-based SCIIs denote the changes in a given hydro-climatic variable (e.g., high flow as Q_{10}) over a future 30-year period with respect to their historical reference values. A reference value was established for every GCM/HM or GCM/LSM combination using the model simulations for 1971–2000. Subsequently, SCIIs were estimated for every 30-year period, starting from 2011 until 2095 with a gap of 5 years between each 30-year periods (i.e., 2011–2040, 2016–2045, . . ., 2066–2095). Some of the CP-based SCIIs related to soil moisture and groundwater recharge are expressed as quantile indices of monthly values for the entire period 1971–2099 (*Samaniego et al.*, 2013). Figure 10.4 shows the projected changes in one of the SCIIs (high flow indicator – Q_{10}) for different future periods using the multi-model ensemble consisting of 5 GCMs and 4 HMs/LSMs. This figure shows, for example, the level of detail made available by the EDgE simulations. Significantly drier hydrological conditions are expected across the Mediterranean region. Extended analyses of the implications of global warming on floods, low flows, and soil moisture droughts based on EDgE water projections are available in *Thober et al.* (2018), *Marx et al.* (2018), and *Samaniego et al.* (2018), respectively.

The SF-based SCIIs are expressed as the percentage of realizations that detect a reference-based indicator at a given month and lead-time (varying from 1 to 6 months) to assess the accuracy of the available seasonal forecast data. The reference is based on the E-OBS historical simulations conducted for every HMs/LSM separately. For example, the high flow indicator quantifies the percentage of ensemble realizations above the reference monthly river flow (Q_{10}) for each lead-time, with Q_{10} previously derived for each calendar month from the relevant reference model run. Similarly, the low flow indicator was based on assessment of the streamflow forecasts falling below the reference monthly river flow (Q_{90}). Another category of SF-SCIIs are the counts of monthly forecasts falling within each of the quantile levels ($\leq 20\%$, $20\% - 40\%$, $40\% - 60\%$, $60\% - 80\%$, $> 80\%$), with the five level limits derived for each calendar month from the relevant historic reference run.

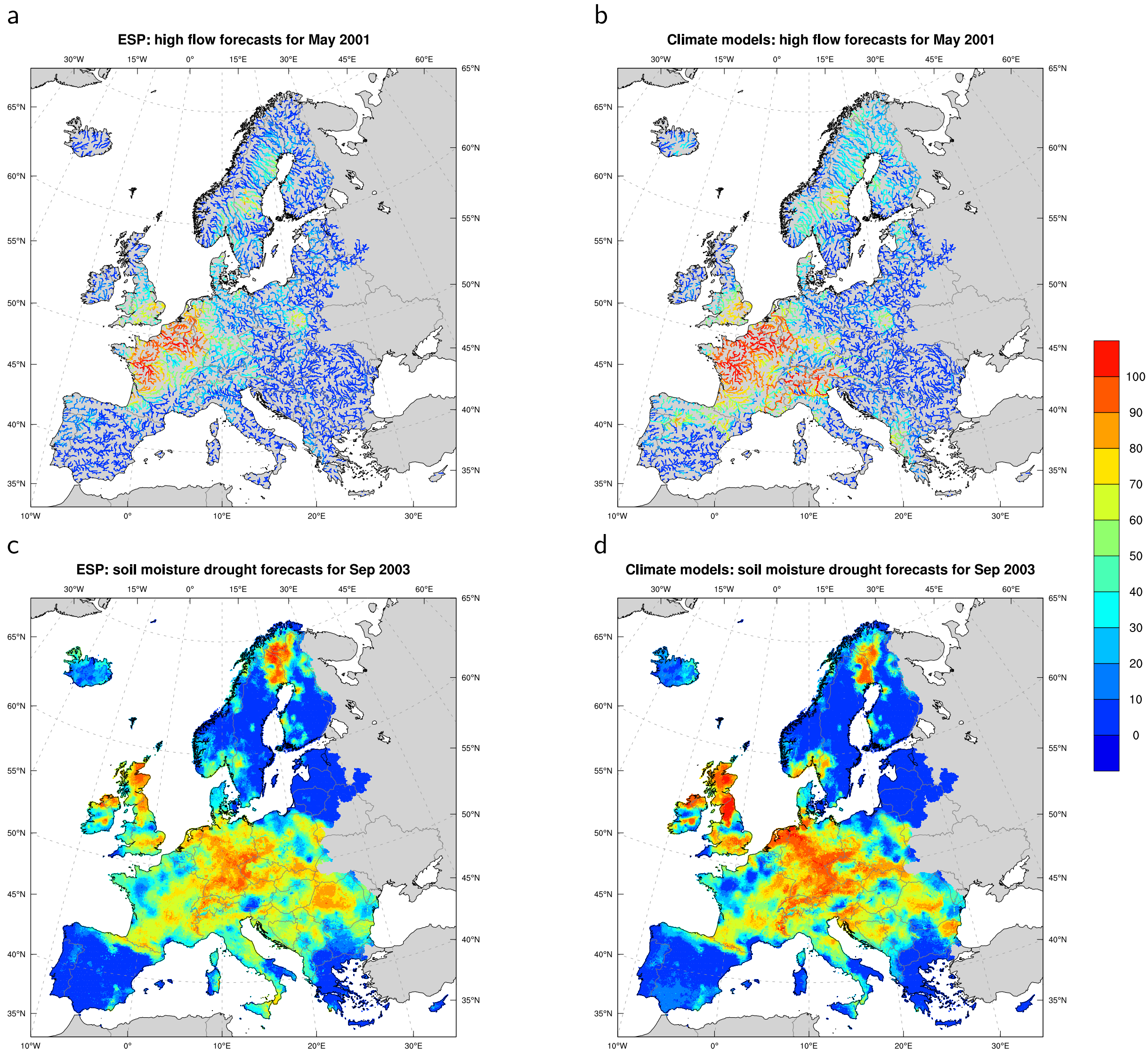

Figure 10.5 Multimodel ensemble mean percentage of forecast realizations detecting the reference run based high flow (Q_{10}) and soil moisture drought indicators at one month lead time for the ensemble streamflow predictions (ESP, (a) and (c)) and climate model (GCM, (b) and (d)) forced hydrological model runs. The reference runs are based on hydrological model simulations driven by the observed meteorological forcing data (here the E-OBS data set).

Figure 10.5 presents the results of two SF-based SCIIs related to high river flow and soil moisture drought (i.e., quantile level $\leq 20\%$) at one month lead time. It shows two extreme events, a flood occurring in May 2001 in France and the Alps regions (panels a, b); and the 2003 soil moisture drought in central Europe (panels c, d). The climatological-based Ensemble Streamflow Prediction (ESP) (*Day*, 1985) was used to provide a benchmark of the GCM-based seasonal forecasts. In general, the GCM-based seasonal forecast showed a larger agreement with observations than the ESP-based forecast. A detailed analysis of the seasonal forecasting skill and uncertainty associated with EDgE seasonal streamflow hindcasts is reported by *Wanders et al.* (2019).

It was clear from the consultation with users that information on the uncertainty of climate projections and the skill of seasonal forecasts is critical for outputs to be used to their full potential (*Taylor et al.*, 2015). This is especially true for large ensembles, like the EDgE seasonal forecast and climate projection chains, that contain many combinations of models, seasons, and geographical locations. Within this project, skill and uncertainty information were combined with expert knowledge to provide end-users with both quantitative and qualitative information, designed to facilitate interpretation. The simulation quality was determined for each individual hydro-meteorological model combination and initialization months, and was conducted for all parts of the geographical domain independently. In the following two sections we analyze the uncertainty and the skill of both modeling chains.

10.6 Uncertainties in Climate Projections

Due to project limitations, uncertainty analysis did not include that from parameter data, downscaling method, and geophysical data. Consequently, the term uncertainty here refers to the spread among the hydro-meteorological ensemble members, calculated overall meteorological and hydrological model combinations.

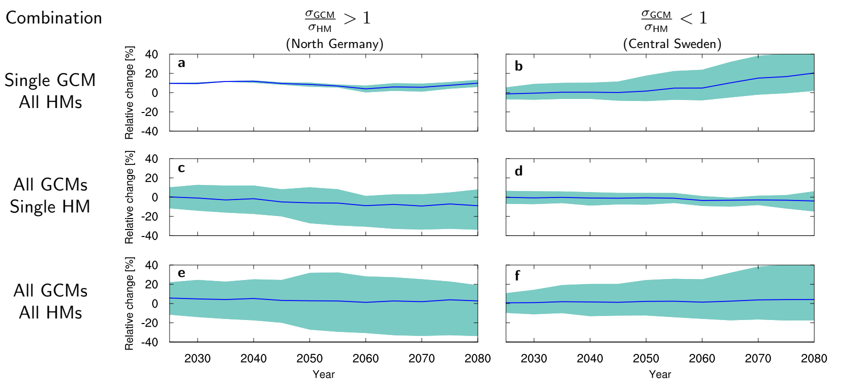

Figure 10.6 Uncertainty of the multi-model ensemble of the relative change of high streamflow indicator (Q_{10}) for consecutive 30-year periods with respect to the reference period for different combinations of GCMs and HMs/LSMs and for two distinctive locations in Central Europe ($\approx$ E9.38°,N52.12°) and Scandinavia ($\approx$ E16.72°,N62.98°) characterized by a larger and smaller ratios of GCM/HM uncertainty ($\frac{\sigma_{GCM}}{\sigma_{HM}}$), respectively. The reference runs are based on hydrological model simulations driven by the historical forcing data of the respective GCM.

The ability of GCMs and HMs/LSMs to capture the predictive uncertainty of key water fluxes and state variables has been extensively discussed in the literature (e.g., *Giuntoli et al.*, 2015; *Prudhomme et al.*, 2014, and sources therein). The propagation of forcing data and meteorological model uncertainties and its dependency on RCP and time horizon were shown by *Samaniego et al.* (2016) for streamflow for a few selected basins. In EDgE, the assessment was extended to all indicators shown in Table 10.1, and estimated over a large domain and at the high resolution of the EDgE modeling framework. Figure 10.6 provides clear evidence, when investigating the impact of climate change on water-related variables, of the importance of a multi-hydrological model ensemble over a single hydrological model to better capture the uncertainty propagation in water fluxes and state variables. There are regions where GCM forcing uncertainty clearly dominates the HM/LSM structural uncertainty for the high streamflow indicator (Q_{10}, see panels: a, c, and e in Figure 10.6), but there are other locations in which the opposite is true (panels b, d, and f in Figure 10.6). Notably, the spatial distribution of uncertainty varies between all indicators. Consequently, it is not possible to establish *a priori* the optimal size, or membership, of ensemble GCMs and ensemble HMs/LSMs. Ideally, the ensemble size should be as large as possible to be able to identify the uncertainties originating from either GCMs or HMs/LSMs. These results clearly indicate that a single model or a sub-ensemble may work well in a given location but perform poorly in another, and hence should not be recommended for an operational system.

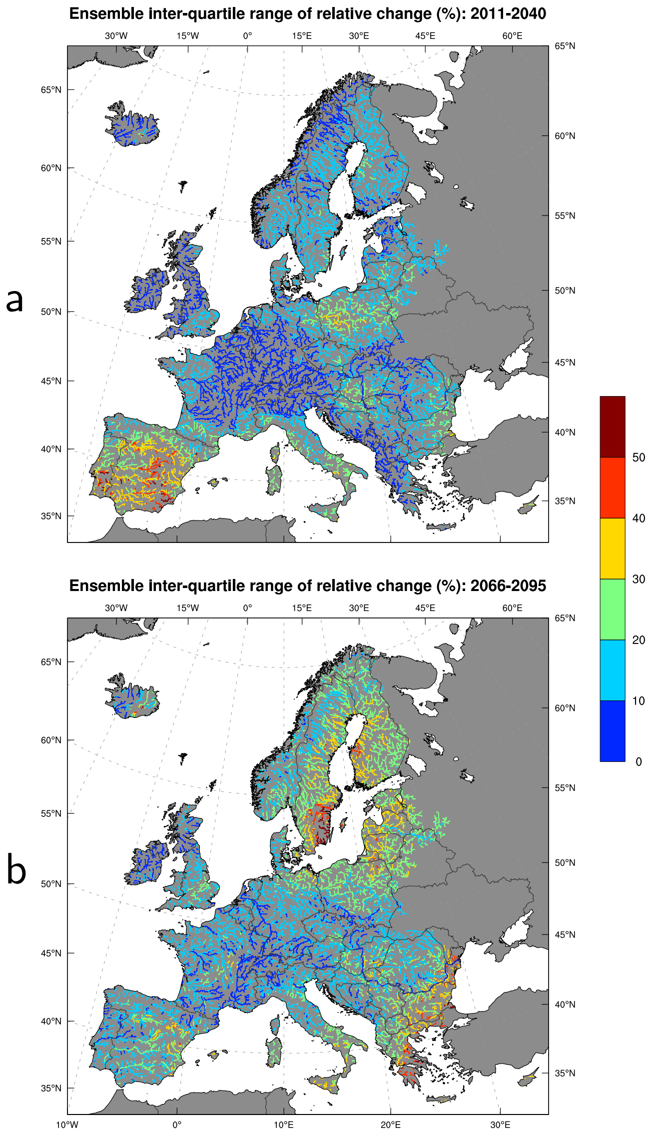

Figure 10.7 Ensemble inter-quartile range of the relative change of high streamflow indicator (Q_{10}) for two 30-year periods with respect to the reference period: (a) 2011-2040 and (b) 2066-2099, both under the RCP8.5 scenario. Southern Europe exhibits, for example, a significant decrease in uncertainty by the end of the century. The historical reference period is from 1971 to 2000.

The uncertainty of the climate projections under a given RCP were estimated as in *Samaniego et al.* (2016), which used the long-term mean of the inter-quartile range for a given SCII as the uncertainty metric, so that the contributions stemming from GCMs and HMs/LSMs could be disentangled. The 40-member ensemble inter-quartile range of the relative change of the high streamflow indicator (Q_{10}) shows significant regional changes within a given RCP scenario for two future periods (2011-2040 and 2066-2099, panel a and b in Figure 10.7, respectively). Southern Europe, in particular, exhibits a significant decrease in uncertainty by the end of the century. This suggests that GCM projections and HMs/LSMs tend to have a consistent estimate of the projected changes. The same method applied to sub-samples of the ensemble can be used to systematically quantify uncertainties stemming from different origins (i.e., uncertainties from hydrological model structure and input uncertainty — given that all HMs/LSMs use an identical set of underlying physiographical land surface characteristics), using, for example, the full ensemble inter-quartile range spread as the benchmark uncertainty. Users of the online platform can experiment with different combinations to assess the quality of a given combination of GCMs and LSMs/HMs. Note however, that a smaller ensemble might not capture the full uncertainty of the hydrological and meteorological combinations.

A detailed analysis of the contribution by GCMs and HMs/LSMs to the total uncertainty has been evaluated within the High-resolution Climate Indicators for 1.5 Degree Global Warming project (HOKLIM; `www.ufz.de/hoklim`), which employed EDgE simulation data. It was found that the total uncertainty in modeled variables is dominated by the choice of hydrological models in Alpine and semi-arid regions for both floods and low flows (*Marx et al.*, 2018; *Thober et al.*, 2018). In these regions, different representations of snow processes and soil water redistribution in the HMs/LSMs have an impact on the projected climate change signal comparable to the different meteorological forcing data from the GCMs.

A further analysis of the GCMs and HMs/LSMs uncertainty contribution to the soil moisture drought duration (duration in Table 10.1) under different global warming levels reveals that the ensemble spread is dominated by the GCMs in comparison to the HMs/LSMs (Figure 10.8, lower panels), with the ratio of GCM and HM/LSM uncertainties ($\frac{\sigma_{\mathrm{GCM}}}{\sigma_{\mathrm{HM}}}$) being particularly high in the Continental and Atlantic region. In these humid regions, the variability of precipitation among GCMs has the largest effect on soil moisture drought development. In Scandinavia, the Southern Alps and parts of the Mediterranean, the HMs/LSMs contribute at least as much to the uncertainty as GCMs. In cold regions, this might be related to the importance of snow processes, where the rate of snowmelt and accumulation varies substantially among the HMs/LSMs. In arid regions such as the Mediterranean, the soil water restriction on evapotranspiration also varies among HMs/LSMs. This process greatly influences how fast soils dry and thus drought development. These results provide substantial evidence against the assumption that the uncertainty of derived SCIIs is equally distributed between atmospheric (GCM) and land surface/hydrological models(HMs/LSMs), in accordance with the findings of *Samaniego et al.* (2016) for low-flow duration.

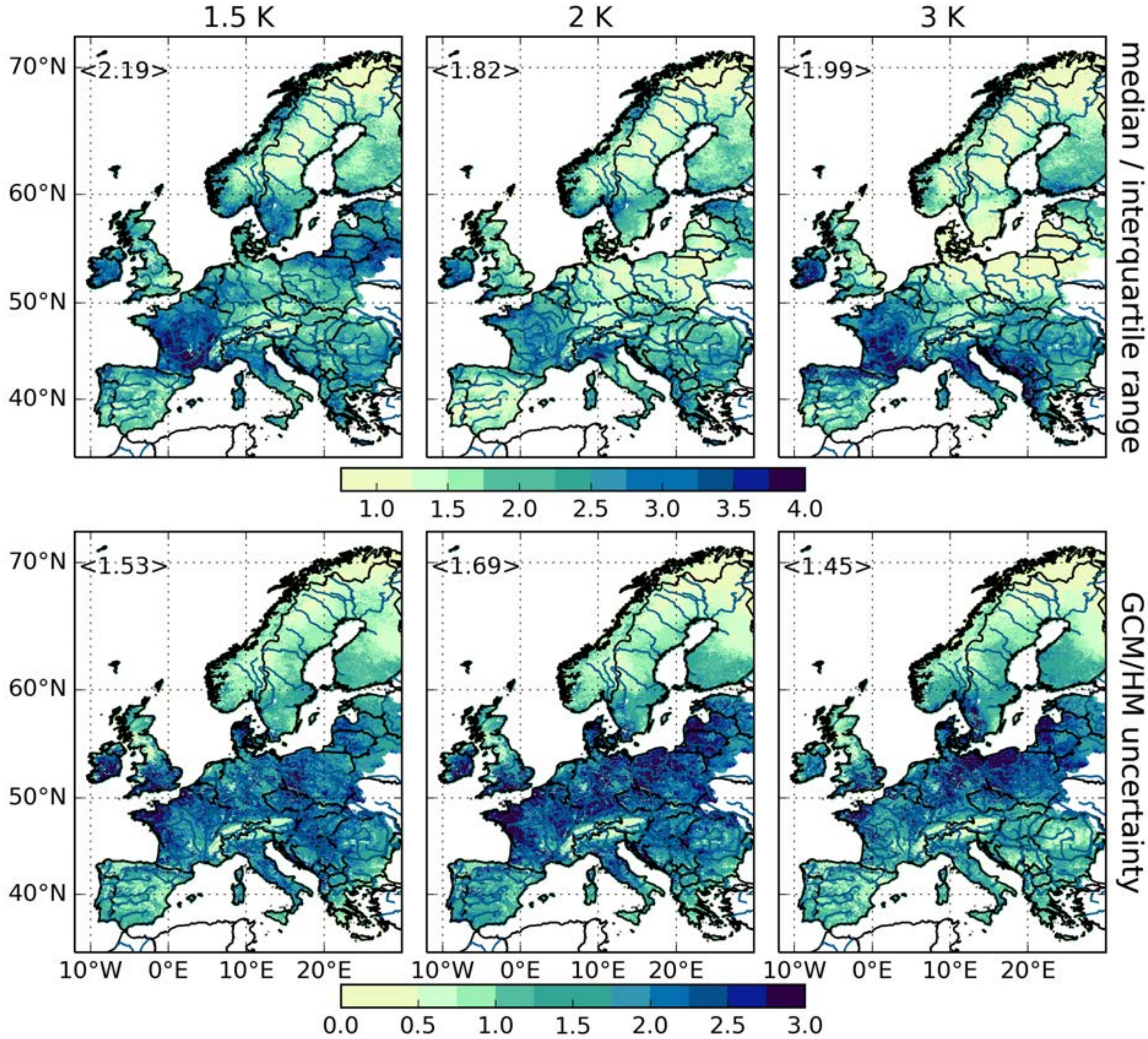

Figure 10.8 Uncertainty in the estimated number of drought months per year expressed as signal to noise ratio (upper panel), and as the ratio of the uncertainty contribution of global climate models with respect to hydrological models (lower panel, using the method described in *Samaniego et al.* (2016)). The columns from left to right correspond to a global warming level of 1.5 K, 2 K, and 3 K, respectively. The number in the brackets of each panel denotes the mean value over space.

The spatial distribution of the signal to noise ratio for drought duration calculated as the median divided by the inter-quartile range is shown in the upper panels in Figure 10.8. Low signal to noise ratios are generally found in Scandinavia, Germany, and Poland, irrespective of the amount of global warming. Note that the spatial distribution of GCMs/HMs/LSMs uncertainty contribution does not correlate with that of the signal to noise ratio (compare Figure 10.8, upper and lower panels). For example, for a global warming of 3 K (Figure 10.8, right column), the signal to noise ratio is low in Scandinavia, where uncertainty is dominated by HMs/LSMs, but it is also low in Poland, where uncertainty is dominated by GCMs. In other words, low confidence in future projections (i.e., low signal to noise ratio) can be created by both HMs/LSMs and GCMs, and be undifferentiated.

10.7 Skill of Seasonal Forecasts

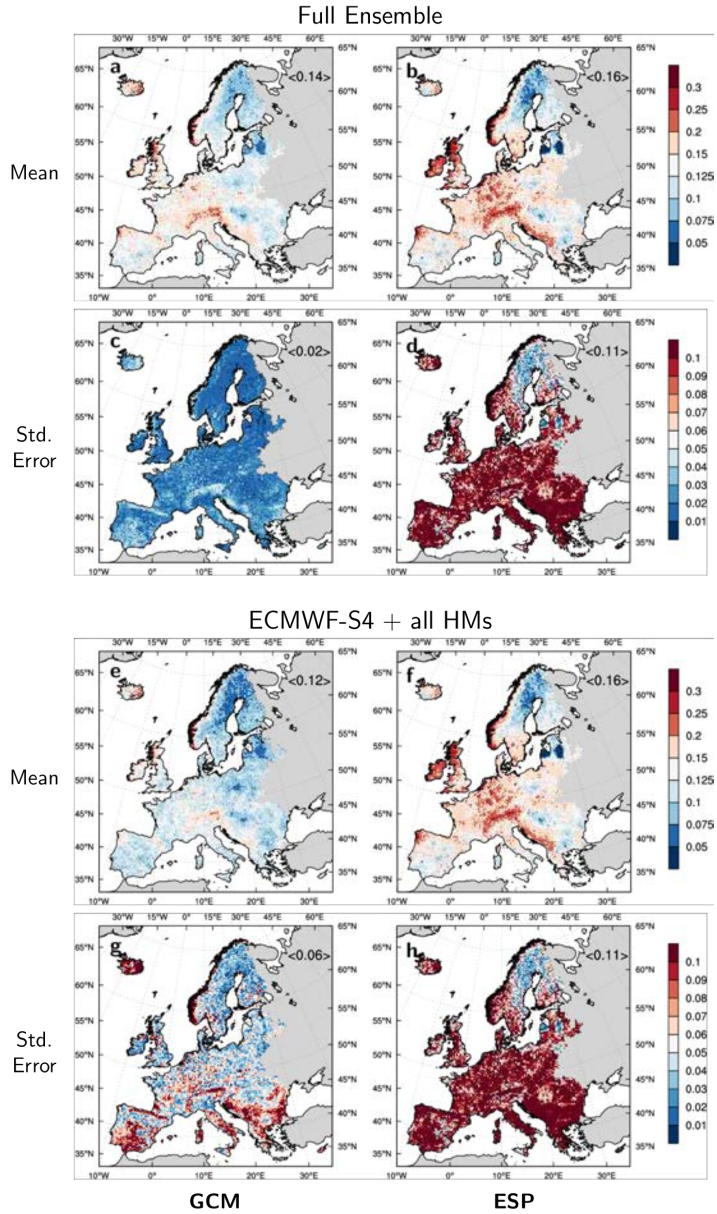

Figure 10.9 Effects of the sub-ensemble selection on the Brier Score (BS) skill in SF-mode for the 1st. soil moisture quantile indicator (drought events). Panels (a,c) depicts the ensemble mean and standard error of BS obtained with the full ensemble, which includes all SF-GCMs and all HMs/LSMs, respectively. Panel (e,g) shows the same statistics for the sub-ensemble of one SF-GCM (ECMWF-S4) with all HMs/LSMs. The right hand-side panels (b,d,f,h) show the corresponding statistics of BS obtained with ESP instead of the SF-GCM model combination. The number in the brackets of each panel denotes the mean value over space.

The seasonal forecast skill was calculated employing one of the most commonly used skill scores for seasonal forecasting in meteorology and hydrology: the Brier Score (BS, *Brier*, 1950). The BS uses categorical forecast thresholds to determine the quality of the forecast compared to a reference simulation. E-OBS-based simulations were used as reference. Quintile classes were defined for the qualification of the streamflow skill whereas a threshold soil moisture value was used to discriminate soil moisture drought events (a soil moisture value that is exceeded 80% of the time in a given calendar month and location). Forecasts that hit the reference class of the E-OBS reference simulation were considered as "skillful", whereas those that did not are denoted as "unskillful". The lower the BS is, the better the forecast. The spread of the BS values is estimated as a standard error: $\epsilon = t_\alpha(n-1)\frac{S}{\sqrt{n}}$. Here, n denotes the sample size (i.e., the ensemble size), S, the standard deviation of the ensemble statistic (BS), and t_α denotes the t-Student critical value for $1-\alpha$ confidence interval (95%) and $n-1$ degrees of freedom. Consequently, the uncertainty of the mean Brier score $\overline{BS}$ at a given location is $\overline{BS} \pm \epsilon$. A forecast with large standard error is deemed highly uncertain because the accuracy of the forecast is low.

The ESP approach (*Day*, 1985) is an often-used benchmark in seasonal hydrological forecasting (e.g., *Thober et al.*, 2015; *Wanders et al.*, 2019) and provides the forecast skill that can be obtained from the initial hydrological conditions. The method was implemented in EDgE using a hindcast starting at month m and year y, generated from 15 years randomly drawn from the E-OBS forcing data during the period 1993–2011, starting at the calendar month m. The sample size was selected to resemble the maximum number of ensemble members of the seasonal forecast models, so that the forecast quality cannot be influenced by different ensemble size. For shorter lead-times and in regions that have a long hydrological memory, the ESP can provide a highly skillful forecast because the impact of the initial hydrological conditions dominates the seasonal predictability (*Wanders et al.*, 2019). For longer lead-times, the ESP tends to become close to hydroclimatology, resulting in a decreasing forecast skill. Users can evaluate the added value of the dynamical (GCM-based) seasonal forecast by comparing it with ESP.

The full ensemble of GCMs and HMs/LSMs exhibits a BS of 0.14 over the entire domain for seasonal soil moisture droughts at one month lead time (Figure 10.9, panel a). The standard error of the BS values for the full ensemble is 0.02 on average. In comparison, the average BS values using ESP is slightly higher than that of the full ensemble (0.16), but has a standard error that is on average five times higher than that of the full ensemble. Notably, the spatial patterns in BS are comparable among the full ensemble and ESP. This indicates that forecasting skill does not only depend on the meteorological input, but also on other factors such as the persistence of initial hydrologic conditions (*Wanders et al.*, 2019). Notably, the skill of the full ensemble (SF-GCM + LSM/HM) is consistently higher than that of ESP at almost all locations in Europe (BS tends to have lower values). This is remarkable because the analysis is favorable to ESP for two reasons: 1) the SF-GCMs were not bias-corrected prior to being used as forcing data to the hydrological models, and 2) the reference values for the BS estimates are based on HMs/LSMs simulations using E-OBS forcing. As a consequence, the ESP forecasts and the reference values have the same climatology, which may not be the case for the SF-GCMs. The single best performing SF-GCM, ECWMF-S4, provides a minor improvement with respect to the full ensemble, but has a threefold standard error as the latter (Figure 10.9, panel e and g). Similarly, the skill of individual ECMWF-S4/HMs/LSMs combinations is slightly higher than that of the full ensemble and all of these outperform ESP (see Figure10-A.2 for individual combinations). However, the uncertainty for the individual model combinations cannot be estimated which lowers the credibility of their skill. Overall, any combination of SF-GCMs/HMs/LSMs provides a higher forecasting skill than ESP for soil moisture droughts at one month lead time. The full ensemble exhibits the highest forecasting skill with respect to both bias and uncertainty among all possible combinations. This may not be true everywhere, which is why users can choose the SF-GCMs/HMs/LSMs that provides the best ensemble forecast for a given location and time in the web interface of the demonstrator.

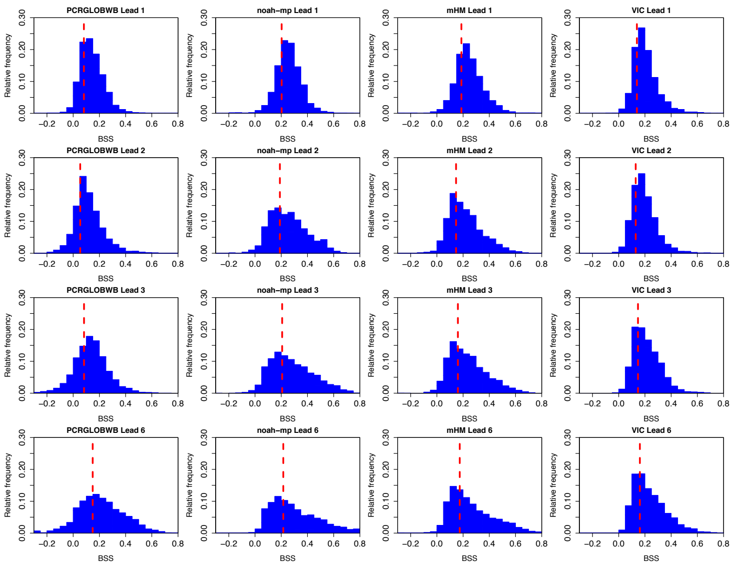

Figure 10.10 Brier skill score (BSS) for dynamical seasonal forecasts over Europe. BSS is calculated for each gauging station individually (shown in Figure 10-A.1), for each lead time and model separately, using ESP as reference (BS_{ESP}). No separation has been made with respect to the forecast season. The red dash line indicates the median of the distribution over 465 streamflow stations during the hindcast period. Positive BSS values indicate an improvement in the dynamical forecasts compared to the climatological forecast.

Improvement of skill using SF-GCM-driven forecasts instead of ESP can be measured using the Brier Skill Score (BSS), with a $BSS > 0$ showing an improvement. BSS was calculated at every location and for every SCII, and is available on the online demonstrator (*EDgE*, 2017). Figure 10.10 gives the histogram of the BSS for streamflow at 465 selected gauging locations, and Figure10-A.1 shows the LSM/HM model performance. Results show a strong relationship between the LSM/HM model performance and BSS, with high-performing hydrological models being associated with high median BSS values and a histogram skewed towards the right. For example, the median BSS are around 0.2 for mHM and Noah-MP but are slightly less than 0.2 for VIC and less than 0.1 for PCR-GLOBWB (except for 6-month lead time forecasts). This could be related to the initial model skill, but is also linked to the impact of the initial hydrological conditions. PCR-GLOBWB tends to show a long hydrological memory (*Wanders et al.*, 2019), which limits the impact of the dynamical forecast improvement. Hydrological models that respond rapidly to precipitation or temperature changes are more likely to benefit from accurate dynamical seasonal forecasts and thus show a strong improvement in the BSS. This suggests that future hydrological models with more accurate representations of the observed hydrology and applied in regions with shorter hydrological response times are likely to profit more from SF-GCMs.

10.8 Conclusions and Outlook

The EDgE project was one of the two proof-of-concepts (PoC) commissioned as precursors to the development of a fully operational system for the water sector in Europe. Three highlights of the EDgE approach are: 1) the unprecedented high-resolution and consistency of inputs of the multi-model hydrological simulations at time scales of seasonal forecasts and climate projections, 2) the systematic uncertainty estimation for 36 co-designed water impact indicators, and 3) the delivery of a high quality water information service tailored to the needs of end users. These characteristics of EDgE are preconditions for users to make informed decisions and therefore constitute the key for improved decision-making. The results shown here also highlight the relevance and value of having a multi-hydrological model ensemble capable of capturing the total uncertainty of the prediction chain.

Operationalization of the modeling chain would be straightforward because it was designed to be upgradable and scalable. Including new hydrological models or updated versions of Noah-MP, mHM, VIC, and PCR-GLOBWB would be possible due to the flexibility of the operational framework. Similarly, adding new climate models for the CP and the SF modeling chains would only be limited by the storage capacity and computational power available.

Skillful seasonal forecasts will depend on the quality of the initial conditions, the performance of the hydrological models against observations, the employed spatial resolution, and skill of the SF-GCM model. One limitation of this PoC is that no bias correction is applied to the SF-GCM data because of the short hindcast period of 19 years. A longer hindcast period should be used during an operationalization phase. As shown in this PoC, model cross-validation at gauged locations is an excellent diagnostic tool to assess model deficiencies (i.e., model parameterization and/or structure). Ideally, a diagnostic tool should be part of the operational system giving updated information on a regular basis, with updates on skill assessments regularly conveyed to end-users.

Next steps within the development of this water information system should focus on anthropogenic influences that alter the natural course and water balance of the hydrologic cycle in all hydrological models. These will greatly improve the quality of seasonal forecasts and the usability of the climate projections. For this reason, high resolution data of water bodies, dam systems and water distribution infrastructure in Europe should be further assembled. For climate projections, it is crucial to include dynamic land-cover/use models coupled with hydrological models. Further work should also provide indicators related to water quality and river temperature as needed by end users. All hydrological and land surface models in EDgE used the same underlying static data sets (e.g., soil and land cover). These were, however, processed differently for the individual models. Applying a seamless parameterization following *Samaniego et al.* (2010a, 2017), would help to further increase the consistency among hydrological model simulations. All these factors would improve the quality and realism of the impact indicators.

10.9 Acknowledgments

EDgE was performed under a contract (C3S 441 Lot1 NERC) for the Copernicus Climate Change Service (`edge.climate.copernicus.eu`). ECMWF implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. We acknowledge the E-OBS data set from the EU-FP6 project ENSEMBLES (`http://ensembles-eu.metoffice.com`) and the data providers in the ECA&D project (`http://www.ecad.eu`), the Global Runoff Data Centre (GRDC), CORINE data (Copernicus European Commission EEA), Global land cover data (ESA), Global 3D Soil Information System at 1 km resolution (ISRIC), Digital elevation model EU-DEM (EEA), Global 30 Arc-Second Elevation (USGS), Hydrogeological Map of Europe (BGR IHME). We also would like to thank the International Research Institute for Climate and Society (IRI; `http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/`) for making the NMME data set available. Last but not least, we would like to thank the many private and public stakeholders who participated in this study. Without them this project would not have been possible.

Appendix

(A) Bias correction and downscaling

The bias correction technique proposed by *Hempel et al.* (2013) corrects systematic deviations of simulations from historical observations, but preserves the absolute warming signal for temperature and the relative warming signal for precipitation. Daily variability around the monthly means has been adjusted by a quantile mapping assuming a normal and gamma distribution for temperature and precipitation residuals, respectively (*Hempel et al.*, 2013).

Coarse GCM daily values (CP- and SF-modes) of precipitation, daily mean temperature, daily maximum and minimum temperature were downscaled from their native resolution ($\ell_2 = 1°$, 0.75°, or 0.5°) to $\ell_1 = 5$ km using the external drift kriging (EDK) algorithm. EDK is an interpolation technique that provides the best linear unbiased estimation at unknown locations (*Kitanidis*, 1997). It also includes a drift governed by the terrain elevation. The advantage of this procedure over other estimation approaches is that it can account for the fine-scale orographic effects in interpolated precipitation and temperature fields. In this case, EDK can be considered as a simple but unbiased form of statistical disaggregation because it uses coarser predictors and terrain characteristics within a variance minimization scheme. The spatial weights determining the EDK interpolation depend on the spatial variability of the meteorological fields, which is quantified by a semi-variogram. The semi-variograms for the daily precipitation and temperature were derived from daily E-OBS station data (*Haylock et al.*, 2008) covering the entire domain following the procedure proposed by *Zink et al.* (2017). It should be noted that EDK does not modify the long-term trends, making this a suitable technique for climate change impact studies. After downscaling the precipitation and temperature fields, the Mountain Climate simulator tool (*Bohn et al.*, 2013) was used to generate the 3-hourly forcing data of air temperature, downward short-wave and long-wave radiation, specific humidity and surface pressure required for running the LSMs. Daily wind speed climatology was derived from the EFAS forcing data set.

B: Land Surface Data

A key feature of EDgE is the use of consistent land surface data at a high spatial resolution of 500 m. A summary of all open-source data used is listed in Table10-A.1. Terrain characteristics (e.g., elevation, slope, aspect, flow direction, and flow accumulation) were derived from the joined Europe-wide (EU) and Global (GOTOP30) Digital Elevation Model (DEM). The Global data set was used for delineating river basins at those locations that were not covered by the EU-DEM data set. All data sets were re-projected to the ETRS 1989 Lambert Azimuthal Equal Area Coordinate Reference System with a spatial resolution of 500 m for consistency. The spatial domain covers the entire drainage area of all rivers within the Pan-EU territory.

Table 10-A.1 Physiographic information used for the EDgE Project

Description	Data set name	Data Owner	Source
Elevation	EU-DEM	EEA	`http://www.eea.europa.eu/data-and-maps/data/eu-dem`
	GOTOPO30	USGS	`https://lta.cr.usgs.gov/GTOPO30`
Pan-European River and Catchment Database	CCM2 v2.1	EC -JRC	`http://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23`
Soils texture	SoilGrids1km	ISRIC	`https://www.isric.org/explore/soilgrids`
Land cover	GlobCOVER v2	ESA	`http://due.esrin.esa.int/page_globcover.php`
	CLC00, CLC06, CLC12, CLC90 v18.4	Copernicus, ESA	`http://land.copernicus.eu/pan-european/corine-land-cover`
Hydrogeology	IHME1500 v11/	BGR IHME	`http://www.bgr.bund.de/ihme1500`
Leaf Area Index	GIMMS	UMD	`https://iridl.ldeo.columbia.edu/SOURCES/.UMD/.GLCF/.GIMMS/.NDVIg/.global/.dataset_documentation.html`
World Register of Dams	WRD	CIGB-ICOLD	`https://www.icold-cigb.org/GB/world_register/world_register_of_dams.asp`

C: HMs/LSMs calibration and evaluation

Different parameter estimation strategies have been used for the individual HMs/LSMs, based on the expert knowledge of the different modeling teams involved, e.g., 1) estimation of global transfer function parameters; 2) manual tuning of selected (sensitive) model parameters, and 3) bilinear interpolation based on coarser resolution parameter sets. The first approach, which leads to a seamless parameterization of hydrological model parameters, as described in *Samaniego et al.* (2010a, 2017), was applied to mHM and PCR-GLOBWB. Given the resources and time constraints of the EDgE project, a manual calibration was applied for Noah-MP focusing on adjusting the surface evaporation resistance parameter, which was identified as highly sensitive by *Cuntz et al.* (2016). The parameters for the VIC model (*Liang et al.*, 1994) were mapped from a global simulation (*Sheffield and Wood*, 2007) so that they are consistent with the land surface parameters specified for the other models.

All models were driven with the downscaled 5×5 km^2 historical E-OBS data (*Haylock et al.*, 2008) and evaluated against observed GRDC streamflow data for 357 basins (`http://www.bafg.de/GRDC/`). It should be noted that the GRDC basins mainly consist of large rivers which are often heavily modified by anthropogenic activities and infrastructure (for example, large hydropower dams). As a result, it is difficult for models describing naturalized streamflows to simulate GRDC-data influenced by river regulations. Parameter estimation for each HMs/LSMs was conducted only on river basins without large hydro-infrastructure facilities because no dam management information was available at the time of the analysis.

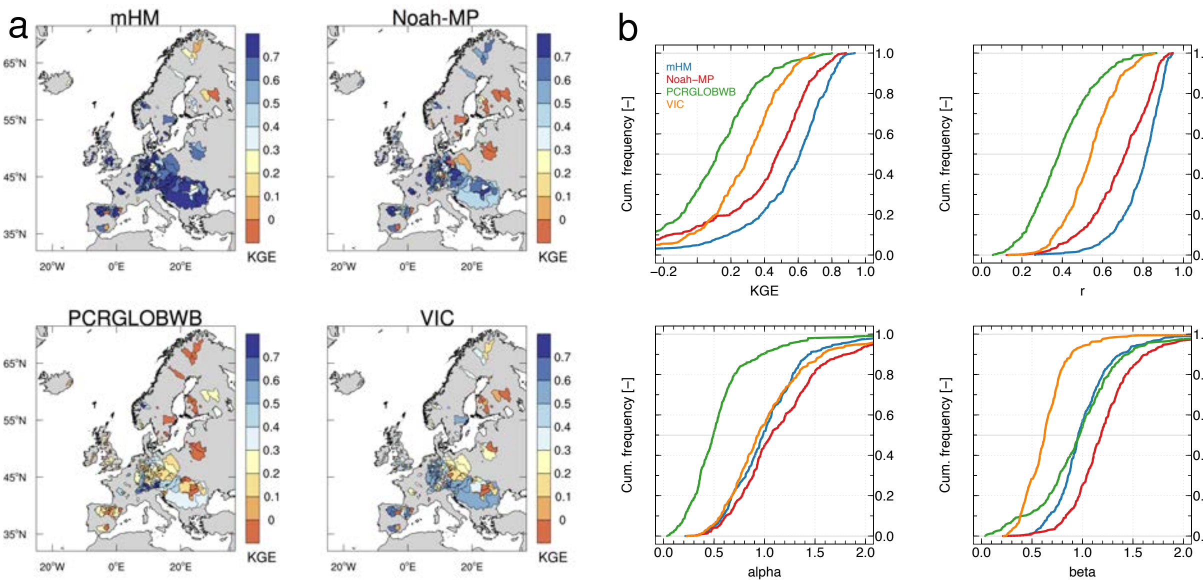

Figure 10-A.1 Evaluation of the hydrological models using the observed daily streamflow over 357 European basins forced with the E-OBS meteorological data: (a) spatial maps of the daily Kling-Gupta Efficiency (KGE) for HMs/LSMs, and (b) cumulative frequency of the daily KGE measure and decomposition into its three components (correlation - r, ratio of variability - $alpha$, ratio of bias - $beta$). Model statistics are based on the 30-year period (1966–1995).

Figure10-A.1 summarizes the model performance in terms of predicting daily streamflow data for all HMs/LSMs. In total, 357 diverse basins with a median area of around 1700 km^2, and a complete streamflow record for a 30-year period (1966–1995), are evaluated using the Kling-Gupta Efficiency (KGE, *Gupta et al.*, 2009). Figure10-A.1a presents the basin-wise spatial evaluation and reveals that the model performance based on historical forcing data strongly depends on model type and region, which highlights the added value of using multiple hydrological models. All models have some difficulties in capturing streamflow dynamics in the northeastern part of the domain, where snowmelt processes are dominant. Figure10-A.1b details quantitative estimates for KGE and its three components: correlation (r), ratio of variability (α), and ratio of mean (β). The median KGE varies between 0.1 and 0.6. The mHM and PCR-GLOBWB models provide unbiased streamflow estimates at the majority of the basins, while Noah-MP and VIC tend to overestimate and underestimate the mean flows, respectively. In the majority of the 357 basins, the variability of observed streamflow flow is well captured by all models except for PCR-GLOBWB. Overall, mHM exhibits the best model performance followed by Noah-MP, VIC and PCR-GLOBWB. Within the project, the stakeholders evaluated the model results in their target basins and found that the all models exhibited reasonable performance.

D: Performance of ECMWF-S4 driven forecasts

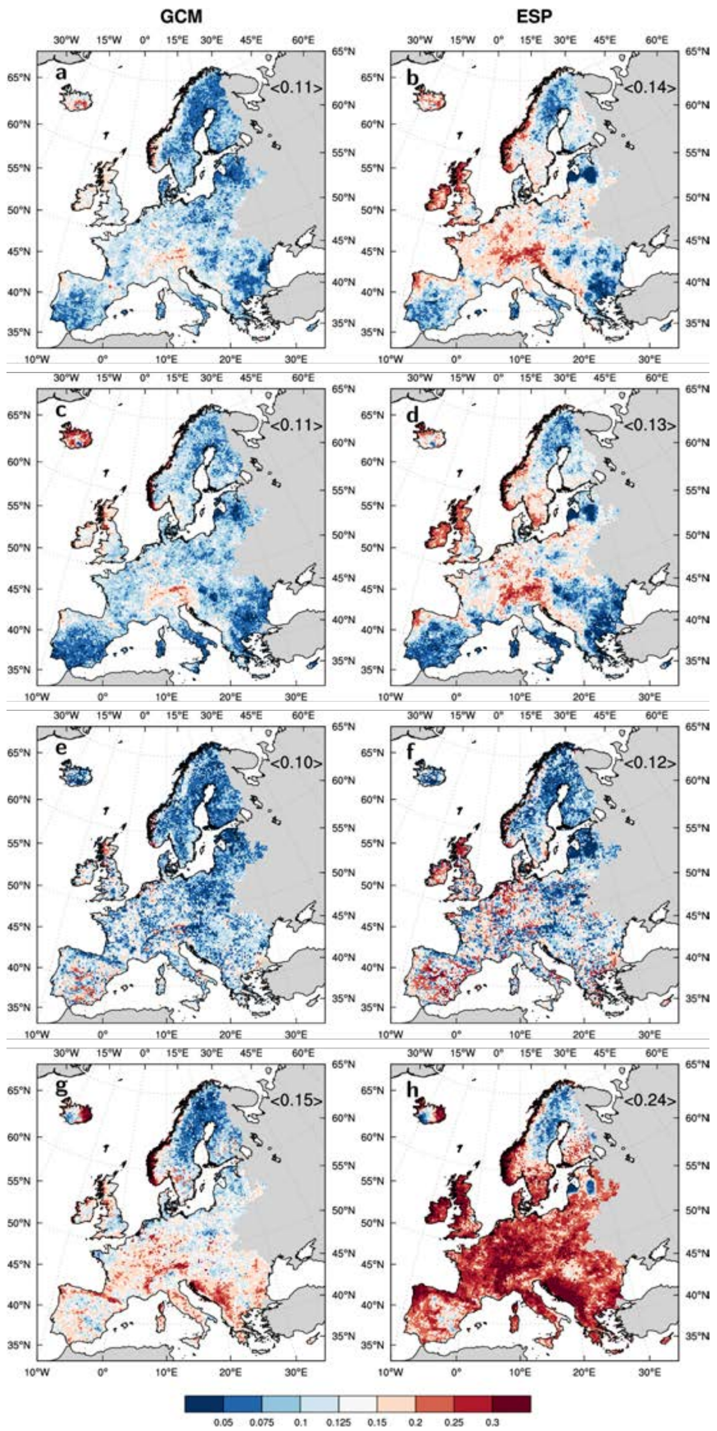

Figure 10-A.2 Effects of the sub-ensemble selection on the Brier Score (BS) skill in SF-mode for the 1st soil moisture quantile. Panel (a,c,e,g) depicts the BS obtained with ECMWF-S4 and the mHM, Noah-MP, PCR-GLOBWB, and VIC, respectively. The right hand-side panels (b,d,f,h, show the BS obtained with ESP and the corresponding HM/LSM. The number in the brackets of each panel denotes the mean value over space.

The individual HMs/LSMs (i.e., mHM, Noah-MP, PCR-GLOBWB, and VIC) driven by ECMWF-S4 show unique spatial patterns of the seasonal soil moisture drought forecasting skill at one month lead time (Figure10-A.2). This analysis provides insights on how the individual models contribute to the forecasting skill of the full ensemble.

The spatial distribution of BS values for the PCR-GLOBWB is very different from that of the other models (Figure10-A.2c). Notably, this model also has the highest forecasting skill both for ECMWF-S4 and ESP-based forecasts among all four models. It appears as a paradox that the model with the least ability to simulate observed streamflow receives the highest forecasting skill. PCR-GLOBWB shows a high underestimation of both the observed variability and correlation (α and r in Figure10-A.1). This indicates that this model has a high persistence, which implies that the ECWMF-S4 and the ESP-derived forecasts do not deviate too much from the E-OBS-based reference run. ECMWF-S4 based forecasts for mHM and Noah-MP outperform ESP and their forecasts are overall comparable (Figure10-A.2: panels a to d). Among all models, VIC shows the least skill for both ECMWF-S4 and ESP-based forecasts (Figure10-A.2: panels g and h). Notably, it also has the largest difference between these two forecasts. In other words, VIC shows the highest skill improvements of 37% for ECMWF-S4 forecasts with respect to ESP-based ones. It is unknown which characteristics of VIC are causing this behavior. Among the four models considered in the ECMWF-S4 ensemble (Figure 10.9: panels a and c), VIC contributes the most to the relatively higher BS values of this ensemble with respect to the ESP ensemble. Overall, there is no model that outperforms the other models at all locations, which highlights the necessity of a multi-model approach.

E: Stakeholder Focus Groups and Feedback

Stakeholder Focus Groups (FGs) in Norway, Spain and the UK were formed as part of the EDgE project, comprising representatives of national government agencies, regional and local government authorities, international water and hydropower companies, agricultural sector, river basin authorities, consultancies and academic sector.

To examine user needs in detail, each Focus Group considered a different aspect of water information and decision-making: water supply in the UK; catchment planning and agriculture in Spain; hydropower generation and local authority planning in Norway. The number of active Focus Group members totaled 29: 11 in the UK; 6 in Norway and 12 in Spain. Stakeholders in Germany were analyzed in a follow-up project (HOKLIM, `www.ufz.de/hoklim`).

Stakeholders welcomed the information provided by the EDgE demonstrator that combines seasonal forecasts and climate projections in a single platform, seeing it as a useful addition to the information they currently had access to. The international companies saw also a large potential in this system because it provides a consistent data set across Europe. In general, it was found that the seasonal forecasts need to have a better skill before they can be used operationally, although they would be used as additional information to climatology. However, the value of using multiple hydrological models to assess hydrological modeling uncertainty was generally appreciated by stakeholders. A European Sectoral Information System like EDgE was thought to be useful for countries that do not have national climate services providing seasonal forecasts and climate and hydrological projections but it should have higher spatial resolution to replace currently available national systems.

PART III

ESTIMATING DROUGHT IMPACTS

CHAPTER 11

THE GERMAN DROUGHT MONITOR

M. Zink[1], L. Samaniego[1] et al.

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Zink, M., L. Samaniego, R. Kumar, S. Thober, J. Mai, D. Schäfer, and A. Marx (2016), The German drought monitor, Environmental Research Letters, 11(7)

11.1 Abstract

The 2003 drought event in Europe had major implications on many societal sectors, including energy production, health, forestry and agriculture. The reduced availability of water accompanied by high temperatures led to substantial economic losses on the order of 1.5 Billion Euros, in agriculture alone. Furthermore, soil droughts have considerable impacts on ecosystems, forest fires and water management. Monitoring soil water availability in near real-time and at high-resolution, i.e., $4{\times}4$ km^2, enables water managers to mitigate the impact of these extreme events. The German Drought Monitor was established in 2014 as an online platform. It uses an operational modeling system that consists of four steps: (1) a daily update of observed meteorological data by the German Weather Service, with consistency checks and interpolation; (2) an estimation of current soil moisture using the mesoscale Hydrological Model (mHM); (3) calculation of a quantile-based Soil Moisture Index (SMI) based on a 60 year data record; and (4) classification of the SMI into five drought classes ranging from abnormally dry to exceptional drought. Finally, an easy to understand map is produced and published on a daily basis on `www.ufz.de/droughtmonitor`. Analysis of the ongoing 2015 drought event, which garnered broad media attention, shows that 75% of the German territory underwent drought conditions in July 2015. Regions such as Northern Bavaria and Eastern Saxony, however, have been particularly prone to drought conditions since autumn 2014. Comparisons with historical droughts show that the 2015 event is amongst the ten most severe drought events observed in Germany since 1954 in terms of its spatial extent, magnitude and duration.

11.2 Introduction

Drought is a natural phenomenon that results from deficiencies in precipitation compared to the expected or normal amount (*Wilhite*, 2005). It may translate to water scarcity, a discrepancy between the actual demand and the corresponding availability of water for environmental and societal needs. Compared to other natural disasters, droughts have the largest spatial extent and longest duration (*Sheffield and Wood*, 2011). These slowly developing events easily persist over several years and can reach national to continental spatial coverage (*Samaniego et al.*, 2013; *Sheffield and Wood*, 2011). According to the EM-DAT database (*Guha-Sapir et al.*, 2015), droughts affected 2.2 billion people worldwide between 1950 and 2014, thus making droughts the second most important natural disaster after floods (3.6 billion people affected). In Europe, for example, the costs per event during this period are estimated to be 621 Mio. EUR, the costliest amongst all natural disasters that occurred in this region (*Guha-Sapir et al.*, 2015). Droughts have impacts on many societal sectors, including forestry, water resources management, energy generation, and health. Their impacts can be divided into direct and indirect impacts (*Wilhite et al.*, 2007). Examples of direct impacts are reduced crop yield and forest productivity, increased forest fire hazard, reduced water levels, and increased mortality rates for livestock, wildlife and fish. They can usually be quantified, though the assessment of indirect impacts is often challenging. An example of indirect drought impact is variable food prices due to market effects in the agricultural sector. As a result, it is difficult to estimate the total costs and losses at the regional and national levels. Indirect losses of droughts often exceed those of the direct ones (*Wilhite et al.*, 2007).

According to the European Commission, the frequency of droughts has increased since 1980 and will, very likely, further increase (*EEA*, 2012). To date, 11% of the European population and 17% of the area of the EU have been affected by water scarcity (*European Commission*, 2007, 2010). For example, the 2003 drought event, which covered major parts of Europe, caused 7,000 fatalities in Germany alone (*European Commission*, 2012) and had an agro-economical impact of 1.5 billion EUR. On the European level, the death toll was estimated to exceed 70,000 (*Robine et al.*, 2008), and the agro-economical impact was estimated to be 15 billion EUR (*COPA-COGECA*, 2003). This severe drought impacted many components of societal life. It disrupted irrigation, inland navigation, and power plant cooling (*Fink et al.*, 2004; *Parry et al.*, 2007).

A precise and generally accepted definition of drought does not exist (*Wilhite*, 2005) because drought impacts are specific to the region of its occurrence and to the field of interest. According to the *WMO* (2006) and *Mishra and Singh* (2010), four different types of drought exist: meteorological, hydrological, agricultural, and socioeconomic droughts. Meteorological droughts relate to a deficiency of precipitation. Agricultural droughts arise as a consequence of this deficiency. They are characterized by low soil water availability for plants. Potential consequences of agricultural droughts are reduced biomass and yield or crop failure. Long-term soil water deficiencies diminish to surface and subsurface water availability, resulting in hydrological drought. It is denoted by reduced streamflow and low water levels of reservoirs and lakes. Hydrological droughts mainly affect water resources management, power plant cooling, irrigation and inland navigation. Groundwater droughts are a special case of hydrological droughts (*Kumar et al.*, 2016; *van Lanen and Peters*, 2000). They occur when water deficiencies reach deep subsurface storages resulting in exceptionally low groundwater levels, groundwater recharge and baseflow. They reduce the supply of fresh water, where groundwater is the major source for drinking water supply. Socio-economic drought can emerge from all of the aforementioned drought types. It is characterized by a shortfall of water supply (water scarcity) leading to monetary losses. In terms of duration, precipitation drought has the shortest occurrence, followed by agricultural drought and finally hydrological and groundwater droughts.

The German Drought Monitor (GDM) presented herein focuses on agricultural droughts, which are highly relevant for Germany because they may induce substantial agro-economic losses as shown by the 2003 drought event. Within this study we review existing drought monitoring systems and the advantages of the newly developed monitor for Germany. Furthermore, we present the technical implementation of the German Drought Monitor and an analysis of the drought event 2015 for which the GDM received broad attention from several media and the public. Finally, we provide an outlook on future developments of the German Drought Monitor.

11.3 Drought Monitoring

Drought monitoring and early warning systems are designed to identify water deficiencies in climatic or hydrologic variables. They aim to detect emergence, probability of occurrence and the potential severity of drought events (*WMO*, 2006). A drought monitoring system that delivers timely information about the onset, extent, and intensity of drought events could help to mitigate drought related impacts such as economic losses (Wilhite, 1993).

11.3.1 Existing Drought Monitoring Systems

Several drought monitors for large parts of the world are currently available to the public. On the continental scale, drought monitoring or forecasting systems exist for North America (*Lawrimore et al.*, 2002), Europe (*Horion et al.*, 2012), and Africa (*Sheffield et al.*, 2014). On a national scale, online platforms are available for India (*Shah and Mishra*, 2015), the Czech Republic (*Trnka et al.*, 2014), and the United States of America (*Luo and Wood*, 2007; *Svoboda et al.*, 2002; *Wood*, 2008). Efforts to monitor drought evolution on the global scale have been made by *Pozzi et al.* (2013) and *Hao et al.* (2014).

A variety of input data, spatial and temporal resolutions and estimated drought indices can be found among these monitoring systems. The longest established system is the US drought monitor launched in 1999. The weekly published map is a composite of different indices based on meteorological observations, i.e., standardized precipitation index, the Palmer drought severity index, soil moisture percentiles derived from hydrologic model simulations, and expert knowledge from more than 130 people (*Svoboda et al.*, 2002). Thus, local experts like agricultural and water resources managers can add information and help verify the drought map. The North American drought monitor was implemented in 2002 based on experience with the US drought monitor (*Lawrimore et al.*, 2002). It enlarges the investigated domain to Canada and Mexico and delivers monthly drought maps. The drought monitors of the University of Washington (*Wood*, 2008) and Princeton University (*Luo and Wood*, 2007) cover the continental United States, showing simulations and forecasts of soil moisture, snow and runoff at 1/8° spatial resolution derived using the Variable Infiltration Capacity (VIC) macroscale hydrologic model (*Liang et al.*, 1994).

Systems established for India (*Shah and Mishra*, 2015) and Africa (*Sheffield et al.*, 2014) are based on bias-corrected satellite precipitation with the latter including a seasonal forecasting capability. These systems are running on 1/4° resolution using the VIC model and provide drought indices based on precipitation, soil moisture, and streamflow. The Czech drought monitor (*Trnka et al.*, 2014) is based on modeled root zone soil moisture, which is derived from local meteorological observations. Maps are published on a weekly basis and have a spatial resolution of 500 m.

The European Drought Observatory (EDO) publishes the current drought status for Europe at a ten-day interval based on a combined drought indicator composed of the standardized precipitation index (SPI) as well as soil moisture and vegetation conditions (*Horion et al.*, 2012). The soil water and vegetation status are assessed by its anomalies. EDO uses local observations to derive the SPI and the hydrologic model LISFLOOD (*De Roo et al.*, 2000) to estimate soil moisture. The status of the vegetation is estimated based on the fraction of Absorbed Photosynthetically Active Radiation (fAPAR) retrieved from ENVISAT. The spatial resolutions of precipitation, soil moisture and fAPAR are 25 km, 5 km and 1 km, respectively, whereas their reference periods are 1981-2010, 1990-2010, and 1997-2010, respectively (*Horion et al.*, 2012).

11.3.2 Aims of the German Drought Monitor

The implementation of a national drought monitoring system goes beyond the capabilities of the existing systems. In our work with regional stakeholders from agriculture and forestry, the need for a high-resolution, near real-time, regional monitoring system was expressed. Therefore, the drought monitoring system presented herein is based on data provided by the German Meteorological Service (*Deutscher Wetterdienst (DWD)*, 2015), which are the most dense and reliable meteorological data available for this region. Furthermore, due to the long-term availability of these data, we are able to use a 60-year reference period for the estimation of drought indices for every grid cell and day of the year. This is substantially longer than those in other existing systems for this region.

The GDM addresses the need for daily up-to-date agricultural drought information. This broadens the decision base for local authorities complementing other available drought information based on e.g. precipitation or streamflow. Finally, the implementation of a national drought monitor encourages local experts, stakeholders and decision makers to take part in the future development. At the same time, it helps to validate the GDM.

11.4 Operational Drought Monitoring Framework

Ground-based monitoring of nation-wide soil moisture is to-date hardly possible. Hence, this study presents a drought identification and classification framework based on near real-time observed meteorological data and distributed hydrologic modeling. The German Drought Monitor estimates soil drought conditions on a high spatial resolution and allows for the evaluation of recent drought events with respect to historical events. A similar framework to that used in the GDM is applied to rank historic drought events in Germany (*Samaniego et al.*, 2013) and for seasonal drought predictions in Europe (*Thober et al.*, 2015).

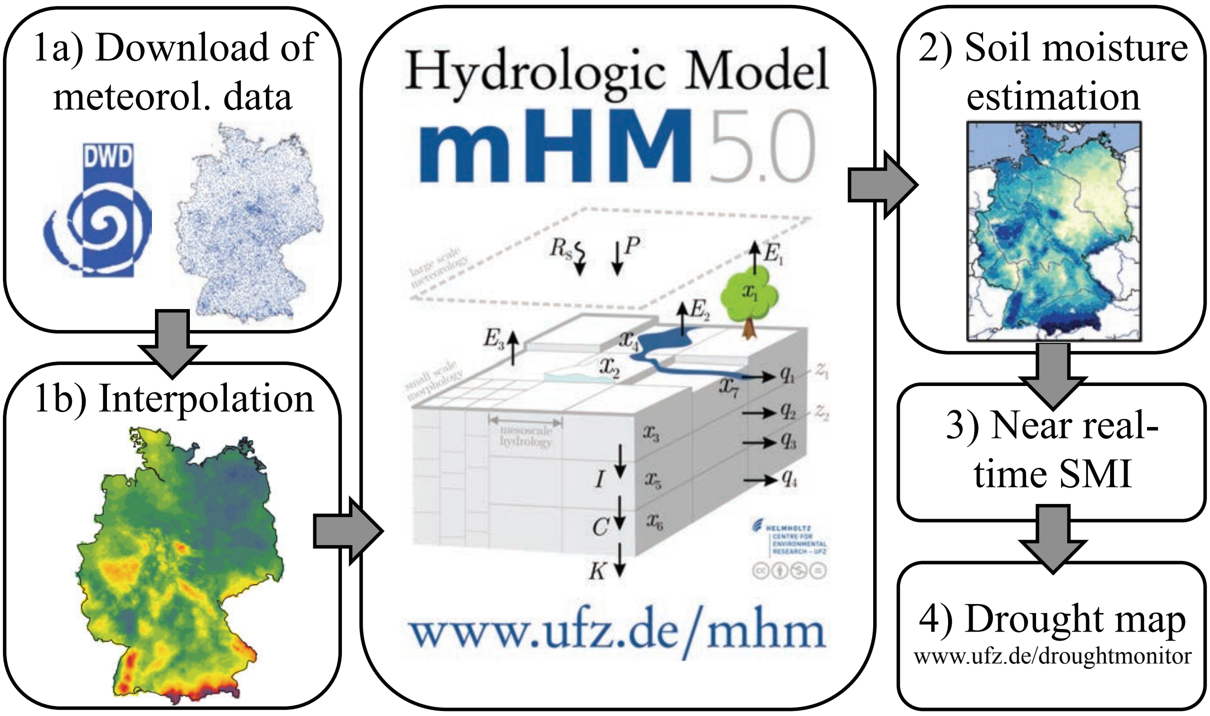

Figure 11.1 Framework of the German Drought Monitor. After 1) downloading and interpolating of the meteorological data from the National Weather Service (DWD) the data are fed to the hydrologic model mHM. 2) mHM estimates the soil moisture for the entire root zone on a daily basis which is used to 3) calculate the Soil Moisture Index (SMI). The SMI is 4) classified and visualized in a drought map published online.

The operational system consists of four processing steps (Figure 11.1). In the first step, local observations from the German Meteorological Service are retrieved every morning (*Deutscher Wetterdienst (DWD)*, 2015). These data are initially quality checked by the DWD. Nevertheless, the GDM checks the downloaded data for consistency and detects outliers as a supplementary quality control. Currently, approximately 1700 precipitation and 500 climate stations, which observe the minimum, maximum, and average daily temperatures, are used to derive daily fields of meteorological input data for the hydrologic model. The daily data are interpolated by external drift kriging using terrain elevation as external drift. The spatial resolution of the resulting meteorological fields is a compromise between the demands for highly resolved hydrological predictions, which are required by stakeholders and practitioners, and the lowest reasonable resolution supported by the input data. The average minimal distance between two neighboring precipitation stations is approximately 6 km in Germany. Thus, a target of 4×4 km^2 resolution was implemented, which would provide high-resolution information without facing the risk of over-interpreting of the meteorological observations. These data are available with a time lag of four days. Due to the high persistency of soil moisture, this near real-time estimation is considered sufficient for agricultural or water management purposes.

In the second step, these interpolated fields are used to force the hydrological model mHM, a process-based model that treats grid cells as unique hydrological units. It comprises hydrological processes such as interception, snow accumulation and melting, infiltration, soil water dynamics, groundwater recharge and storage. The generated discharge of a model cell consists of direct runoff, baseflow, slow and fast interflow, which, after aggregating its components, is routed through the model domain using the Muskingum-Cunge flood routing algorithm (*Chow et al.*, 1988; *Todini*, 2007). By using the Multiscale Parameter Regionalization (MPR) technique (*Kumar et al.*, 2013; *Samaniego et al.*, 2010a), mHM directly accounts for the sub-grid variability of physiographic characteristics. The model parameters are estimated in a preliminary step on the lowest possible input resolution of the physiographic variables, i.e., 100×100 m^2. In a second step, effective parameters at the hydrological modeling resolution of 4×4 km^2 are estimated by applying particular upscaling operators. This technique makes mHM scale- and location-independent because it connects effective parameters to physiographical inputs (*Kumar et al.*, 2013). In several studies, the model has shown to perform satisfactorily in a wide range of catchments with drainage areas ranging from 4 to 530,000 km^2 and with contrasting climatic regimes (Germany, India, USA, Europe; e.g., *Kumar et al.* (2013b); *Rakovec et al.* (2016c); *Samaniego et al.* (2011, 2013)).

A soil moisture field, updated daily, is estimated by running the model with an internal time step of one hour. The soil water availability is estimated in three different layers. The thicknesses of the upper two layers are 5 cm and 20 cm. A third layer is spatially variable in depth, depending on the soil horizon properties specified in the input data. This variable depth, is on average, 1.8 m in Germany. The estimated soil moisture of each single layer is used to estimate the total root zone soil moisture. The hydrological model stores specified state variables at the end of a model run. To calculate the soil moisture statistical reference, we performed a 60-year simulation from 1954 to 2013. Within the operational framework, we are currently performing hydrological simulations initialized with the model states of December 31, 2013. Thus computational time is minimized as the daily model simulation runs from January 1, 2014, onwards. An evaluation of the hydrologic model on the domain of Germany is provided by *Samaniego et al.* (2013).

The third step within the GDM is to transform the daily updated soil moisture into the Soil Moisture Index (SMI) by estimating the percentile of the updated soil moisture value with respect to its climatology. The daily updated soil moisture is estimated as the average of the soil conditions of the preceding 30 days. Therefore, it represents values that correspond to a time period of one month. The SMI is estimated using a non-parametric kernel-based cumulative distribution function based on a 60-year historic soil moisture reconstruction (1954-2013), as described by *Samaniego et al.* (2013). The SMI is bounded between 0-1 and can be easily transformed to the unbounded range of the standard normal distribution e.g. used for the Standardized Precipitation Index (SPI, *McKee et al.* (1993)). It is estimated on every grid cell and for the particular time of the year (i.e., the average of the 30 days preceding the estimation day). The one month running mean of soil moisture data for SMI derivation was chosen because it is well established in scientific literature (*Andreadis and Lettenmaier*, 2006; *Samaniego et al.*, 2013; *Sheffield and Wood*, 2007; *Vidal et al.*, 2010).

Table 11.1 The classification of droughts for the German Drought Monitor based on the Soil Moisture Index (SMI) (adapted from *Svoboda et al.* (2002)).

SMI class	Condition of the soil	Description of potential impacts
$0.2 < \text{SMI} \leq 0.3$	Abnormally Dry	conditions before or after a preceding drought
$0.1 < \text{SMI} \leq 0.2$	Moderate Drought	damages to crops and pastures possible
$0.05 < \text{SMI} \leq 0.1$	Severe Drought	losses in crops and pastures are likely
$0.02 < \text{SMI} \leq 0.05$	Extreme Drought	high probability of major losses in crops and pastures
$\text{SMI} \leq 0.02$	Exceptional Drought	high probability of exceptional losses in crops and pastures

Finally, the fourth step consists of categorizing the estimated SMI into several drought classes and visualizing the results. A main requirement for the appearance of the publicly available drought map is intelligibility. For the visualization of drought events, we adapted the appearance of the German Drought Monitor to that of the US Drought Monitor (*Svoboda et al.*, 2002), using five classes. Four classes define drought conditions, and the fifth class describes the pre-warning state of abnormally dry (Table 11.1). The four drought classes scale from moderate, (vegetation prone to water stress) to exceptional (high probability of losses of crops and increased forest fire risk).

The classes are derived using the thresholds of the Soil Moisture Index (Table 11.1). These thresholds reflect the occurrence of similar conditions in the past and thus indicate the potential impacts of these conditions. For example, the class of exceptional drought is defined by an upper threshold of 0.02. This implies that this soil moisture conditions were observed in less than 2% of the time within the 60-year reference period at this grid cell and time of the year. Thus, this drought condition only occurred in less than 1.2 cases over the last 60 years, which is equal to a return period of 50 a.

Because the SMI describes the status of the soil but not necessarily the impact on the vegetation, this classification scheme still requires further research. Crops cope with drought conditions better or worse at different stages of plant development and may not be influenced by heavy drought conditions. Revisiting this argument would mean that an effect of the Soil Moisture Index (SMI) on vegetation at different stages of plant development has to be investigated.

The resulting maps are visualized and published online in the GDM. Currently, an up-to-date drought map is published every morning at 3 am CET on `www.ufz.de/droughtmonitor`. This information is accompanied by historical, monthly drought maps starting in 2014. We provide detailed maps available since 1954, of particular regions as well as the underlying soil moisture and soil moisture index data on request.

11.5 Benchmark for the Recent 2015 Drought Event

Germany has experienced two drought events since the implementation of the GDM. The first took place in spring 2014, and the second occurred in summer/autumn 2015. These events are used to assess the performance of the GDM. The 2014 event (see Figure 11.2, upper row) had its largest spatial coverage in April 2014. In Germany, 70% of the area was under drought conditions (SMI$\leq$0.2), with 25% of the total area being under exceptional drought (SMI$\leq$0.02). The situation improved significantly in May 2014 due to above average rainfall, and the total drought area (moderate to exceptional drought) decreased to almost half of the area affected in April. Furthermore, the area under exceptional drought reduced to only 1%. As a consequence, the vegetation and, in particular, agricultural crops received sufficient amounts of water, especially during the crucial growing phase after seeding in April/May. In consequence, even the deterioration of drought conditions in June did not have a negative impact on yield in 2014. On the contrary, the Federal Ministry of Food and Agriculture (*BMEL*, 2014) reported that productivity of agriculture was superior to the preceding six years.

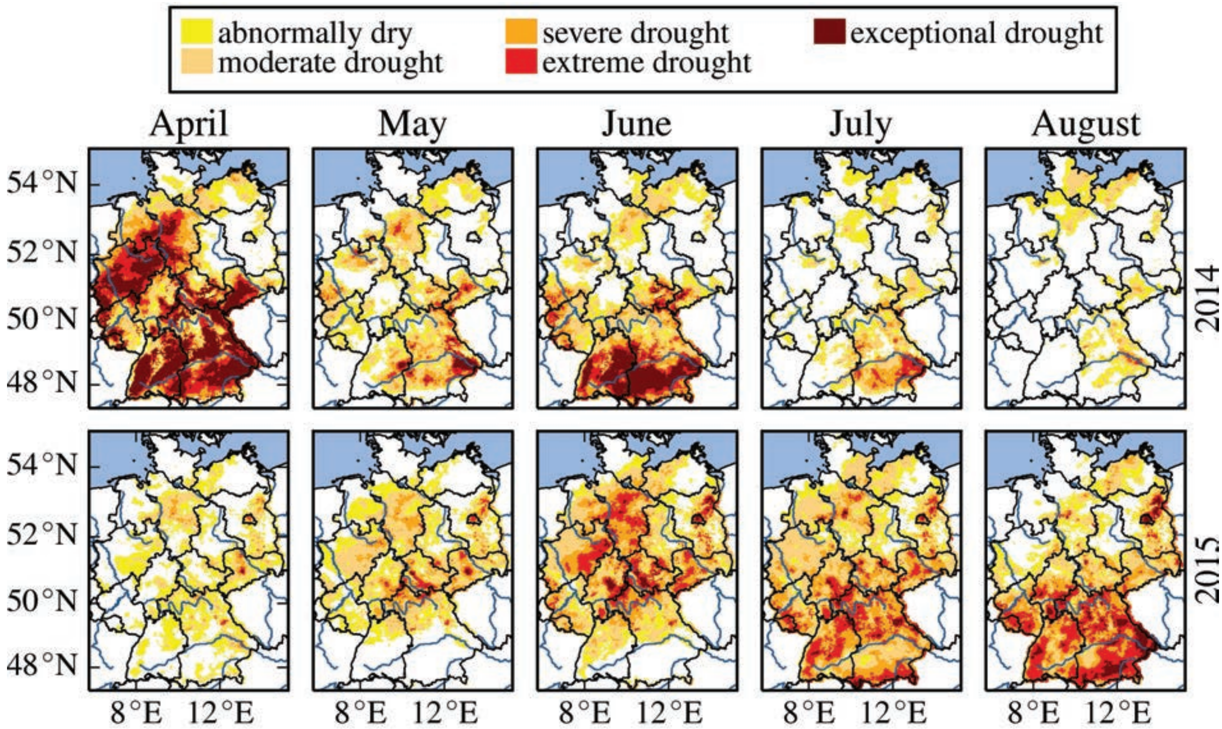

Figure 11.2 Soil water conditions from April to August in 2014 (upper row) and 2015 (lower row).

In 2015, the drought situation was different (Figure 11.2, lower row). In contrast to the situation in 2014, soils were not experiencing extreme to exceptional dry conditions in spring. The drought primarily evolved during spring and summer. Nevertheless, the growing phase of some crops was already delayed by water shortage in May (*BMEL*, 2015). In some regions of Northern Bavaria and Eastern Saxony, soils were under drought conditions since autumn 2014. These regions were especially prone to losses in crop yield and to increased forest-fire risk.

According to (*BMEL*, 2015), corn yield was 22% below the average yield between 2009 and 2014 in Germany. Additionally, some regions of Germany were prone to losses in animal food production, so they faced the decision of either buying additional food or reducing livestock (*BMEL*, 2015). Due to the low water levels, inland navigation was stopped on the Elbe River. A hotspot for very dry conditions was Berlin (Figure 11.2, lower row), where trees had already started shedding their leaves in the middle of August. Reports on economic consequences have not been published yet, but there were extensive fire watch activities due to very high forest fire risk and losses in crops such as corn, which led to increased expenses. Almost 75% of the area of Germany was under at least moderate drought in July 2015. During August 2015, the total area under drought decreased, but the areas of extreme and exceptional drought conditions increased to 22% and 5%, respectively.

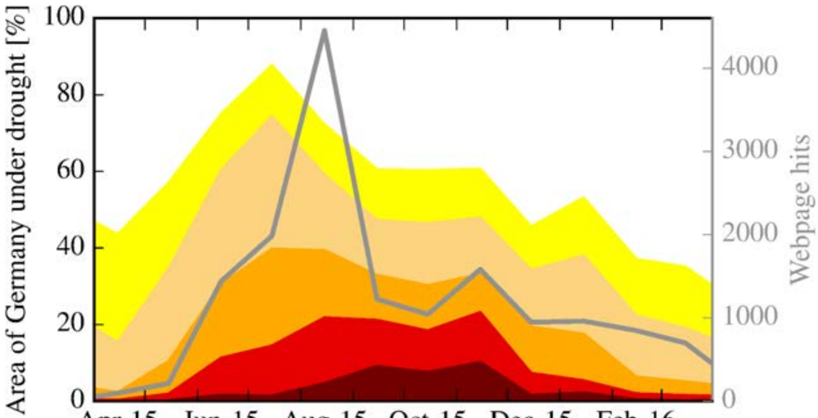

Figure 11.3 Percentage of area affected in Germany during the drought event in five drought classes (legend is show in Fig. 11.2) and total hits on our drought monitor webpage.

The recognition of the German Drought Monitor increased significantly during the 2105 event. The information provided by the GDM were used for public information and drought assessment in local authorities. We could identify users due to individual requests of several federal state agencies including the Saxon State Office for the Environment, Agriculture and Geology, the Thuringian State Office for Environment and Geology, the Bavarian State Office for Agriculture, and the North Rhine Westphalia Chamber of Agriculture. They used the GDM to inform agriculture and forest managers about the current soil moisture status. The drought monitor got attention in the public due to reports in several media ranging from regional to national newspapers as well as television broadcasters. The number of page views of the GDM website followed the estimated area under drought (Figure 11.3). This highlights that extreme events gather more public attention during periods when they do actually occur. Attention rose in April 2015 when newspapers in Saxony started reporting drought conditions due to negative impacts on forestry and agriculture, e.g., seeding of maize and dying tree seedlings. The highest interest was reached when several national media started republishing the maps of the GDM in August 2015. In the private sector, we got feedback from insurance and seed production companies.

The benchmark of the 2015 event with respect to historical drought events is shown in Figure 11.4. The left graph of this figure is created by applying the cluster identification algorithm proposed by *Samaniego et al.* (2013). This three-step algorithm uses the duration, spatial extent and drought intensity to calculate a dimensionless drought magnitude. The drought intensity is calculated as the negative deviation from the SMI value 0.2, whereas the magnitude is the integral of drought intensity over time and space. The results show that the ongoing 2014-2015 event ranks among the 10 largest events observed in Germany since 1954.

A more detailed insight can be obtained from the four panels on the right in Figure 11.4. In these graphs, drought events are evaluated for calendar months. The integral of drought intensity is based on monthly values. The probability is calculated from the empirical cumulative density function of the area under drought. The numbers next to the bubbles denote the respective year of the drought event. The drought conditions in June and July 2015, rank within the four largest events with respect to spatial extent. The magnitude is highly correlated to the area under drought; hence, between June and September, the 2015 event ranks among the 7 largest events for the respective months. The displayed 2003 event is well remembered in Germany due to its large socio-economic impacts. In 2003, the drought event evolved more slowly than the 2015 one did, but the former peaked in August, with a magnitude M=2067, which is greater than the maximum magnitude reached by the 2015 event in July (M=1770).

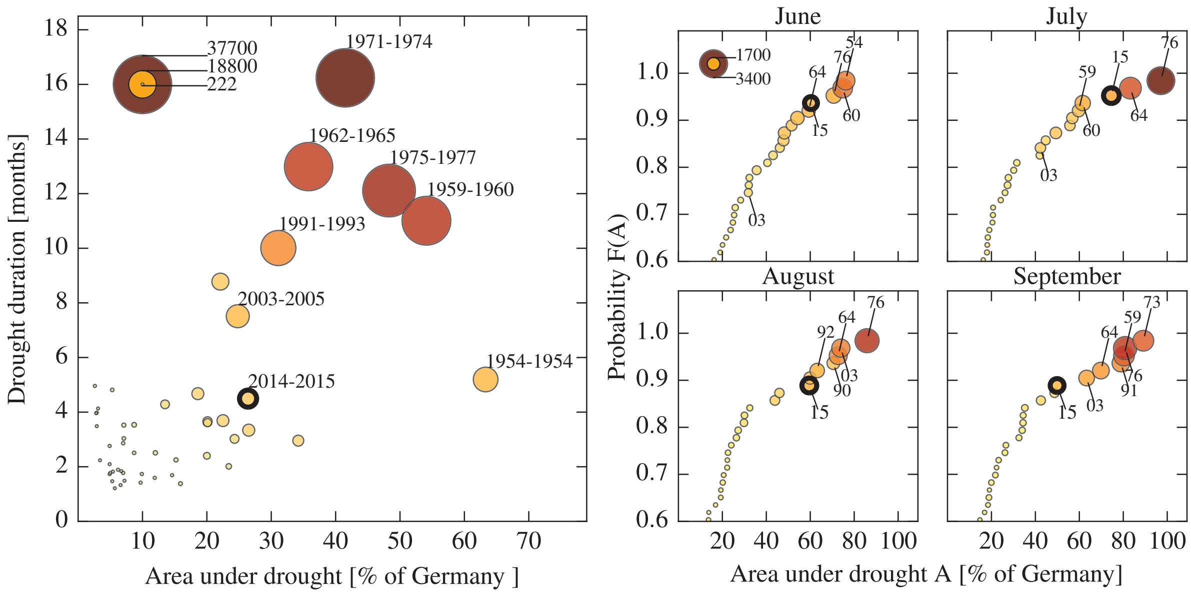

Figure 11.4 Ranking of the recently ongoing drought event in 2015. The panel on the left shows the relationship between the area, duration and magnitude of drought events since 1954. The 4 panels on the right show the ranking of drought area at specific months over the last 62 years. The magnitudes are represented by the size of the bubble and the color code. The reference period for this figure is 1954/01/01-2015/10/31.

11.6 Conclusion and Outlook

The German Drought Monitor (GDM) provides an easily accessible agricultural drought information system on both the regional and national level. It provides an added value through the daily, high-resolution availability of formerly unaccessible information. Stakeholder feedback indicates that the main user groups are from regional agencies and the agriculture and forestry sector. During the 2015 drought, the GDM was widely used by the media and stakeholders when drought consequences became visible (e.g., in tree leaf coloring in summer).

The GDM is driven by an observational dataset, which enables drought estimates on a higher spatial resolution (4×4 km^2) compared to other available products. A soil drought map for Germany is released to the public on a daily basis, with a latency of 4 days. This map is intended to be comprehensible and easy to access via a web browser. Additional information, e.g., the underlying SMI data, are available on request. The GDM information aims to support practitioners to optimize their actions.

A comparison of an ongoing event with historical drought events helps to understand their severity and to assess potential impacts. The sensitivity of plant growth to soil water availability depends of the timing within the year. This could be shown in the comparison of consequences of the drought situation in 2014 and 2015. Currently, the SMI data is used in our research to investigate the relationship between soil moisture and crop yield for different times of the year to gain more knowledge about the consequences of agricultural droughts.

Feedback from stakeholders has already been integrated in the GDM, e.g. in the publication of drought information for the uppermost soil layer with a depth of 25 cm. The future development of the German Drought Monitor will reflect the needs of stakeholders and decision makers. We use the Climate Office for Central Germany, a regional climate service center, to inform agencies, agricultural engineers, water resources managers, hydrologists and policy makers about the potential of the GDM. In this dialogue-based knowledge transfer, we identify 1) how to improve the visualization of drought information (e.g., readability and information content of the maps); 2) how to implement local expert knowledge into the daily published product, and 3) which additional information or combination of drought indices may be beneficial (e.g., Standardized Precipitation Index).

Currently, the drought maps are based on a 30 day soil moisture average, which is a well established procedure found in the literature. Shorter time aggregations may provide new information for particular crops. Thus, further research has to be attributed to determine the sensitivity of temporal aggregation on the soil moisture index and how this relates to agricultural crop development.

An additional field of work remains in handling predictive uncertainties. These uncertainties stem from the input data, the model structure and the model parameters (e.g., *Wagener et al.*, 2003). *Samaniego et al.* (2013) showed that parametric uncertainty alone can lead to significant classification errors in drought characteristics. A major challenge is to investigate how to communicate such uncertainties to the public and decision makers without counteracting the GDM's simplicity and intelligibility.

Providing forecasts may help to better mitigate drought consequences. Studies like *Thober et al.* (2015), however, showed that soil moisture drought forecasts underlie significant uncertainties at seasonal lead times. Nevertheless, we aspire to assess the potential of short and medium range forecasts.

The German Drought Monitor presented herein provides free, high-resolution, near real-time drought information for Germany and a contribution to mitigate negative effects of agricultural droughts.

Acknowledgements

We kindly acknowledge the German Meteorological Service (DWD), the Joint Research Center of the European Commission, the European Environmental Agency, the Federal Institute for Geosciences and Natural Resources (BGR), the Federal Agency for Cartography and Geodesy (BKG), the European Water Archive and the Global Runoff Data Centre at the German Federal Institute of Hydrology (BfG) for the provision of data. We especially thank Sven Petruschke and Ben Langenberg (UFZ) for their web and Linux cluster support. This work was supported by the "Helmholtz Alliance Remote Sensing and Earth System Dynamics" and is part of the Helmholtz Alliance Climate Initiative REKLIM. We would like to thank two anonymous reviewers for their valuable comments.

CHAPTER 12

THE EFFECT OF SOIL MOITURE ANOMALIES ON MAIZE YIELD IN GERMANY

M. PEICHL[1], ..., AND L. SAMANIEGO[1]

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Peichl, M., S. Thober, V. Meyer, and L. Samaniego (2018), The effect of soil moisture anomalies on maize yield in Germany, Natural Hazards and Earth System Science, 18(3), 889–906

12.1 Abstract

Crop models routinely use meteorological variations to estimate crop yield. Soil moisture, however, is the primary source of water for plant growth. The aim of this study is to investigate the intra-seasonal predictability of soil moisture to estimate silage maize yield in Germany. It is also evaluated how approaches considering soil moisture perform compared to those using only meteorological variables. Silage maize is one of the most widely cultivated crops in Germany because it is used as a main biomass supplier for energy production in the course of the German Energy Transition. Reduced form fixed effect panel models are employed to investigate the relationships in this study. These models are estimated for each month of the growing season to gain insights into the time varying effects of soil moisture and meteorological variables. Temperature, precipitation, and potential evapotranspiration are used as meteorological variables. Soil moisture is transformed into anomalies which provide a measure for the inter-annual variation within each month. The main result of this study is that soil moisture anomalies have predictive skills which vary in magnitude and direction depending on the month. For instance, dry soil moisture anomalies in August and September reduce silage maize yield more than 10 % other factors being equal. On the contrary, dry anomalies in May increase crop yield up to 7 % because absolute soil water content is higher in May compared to August due to its seasonality. With respect to the meteorological terms, models using both temperature and precipitation have higher predictabilities than models using only one meteorological variable. Also, models employing only temperature exhibit elevated effects.

12.2 Introduction

In the course of the German Energy Transition, the demand for biomass has increased considerably with silage maize being an important plant for high dry matter yields. The share of the total production in agriculture was 18 % in 2014 (*Die Landwirtschaft Band 1*, 2014), with an increasing share of agricultural area used for silage maize from 15.4 % in 2010 to 17.7 % in 2015 (*Statisitisches Bundesamt*, 2011, 2016). With that in mind, the observed susceptibility of silage maize towards extreme dry conditions during summer time supports the detection of relevant factors for yield variation (as for instance in 2015, *Becker et al.*, 2015; *BMEL*, 2015). Knowing the determinants of maize variation can help to mitigate welfare losses. For instance, detrimental effects of soil moisture shortage and abundance can be mitigated by the means of irrigation and drainage and thus are key for targeted and efficient development of adaptation measures (*Chmielewski*, 2011).

In general, two different kinds of modeling approaches are employed to assess the impact of weather or climate on the agricultural sector. These are structural (integrated assessment) models and reduced form models (*Auffhammer and Schlenker*, 2014). Whilst structural approaches specify the economic behavior based on theoretical models and assumptions and thus have "the ability to make predictions about counterfactual outcomes and welfare" (*Chetty*, 2009), the advantage of reduced form approaches is "transparent and credible identification" (*Chetty*, 2009) by exploiting the exogenous variation of key parameters (*Timmins and Schlenker*, 2009). Regression models are used to estimate the variation in the dependent variable within various sectors by the means of *damage* or *dose-response functions* (*Carleton and Hsiang*, 2016; *Hsiang*, 2016). In the agricultural sector, the major explanatory variables are temperature based (*Carleton and Hsiang*, 2016; *Lobell et al.*, 2008, 2011a; *Schlenker and Lobell*, 2010; *Schlenker et al.*, 2005). The use of temperature as the main explanatory variable is questioned in this study by using reduced form models to identify the impact of different determinants on crop yield.

In the agricultural context, most advances have been made regarding dose-response functions through the development of temperature estimates on high spatial and temporal resolutions (*Hsiang*, 2016). Based on this data, many studies employ a precise term which integrates cumulative exposure to specific temperature ranges over the growing period as major explanatory variable. Those are defined as growing degree days (*Deschenes and Greenstone*, 2007; *Schlenker et al.*, 2006) and accumulated measures of extreme heat above a certain threshold, as for instance extreme, heat, killing, or damage degree days (*Annan and Schlenker*, 2015; *Burke and Emerick*, 2016; *Butler and Huybers*, 2013, 2015; *Lobell et al.*, 2011b, 2013; *Ortiz-Bobea and Just*, 2013; *Roberts et al.*, 2013; *Schlenker and Roberts*, 2006, 2009; *Schlenker et al.*, 2013; *Teixeira et al.*, 2013; *Urban et al.*, 2012, 2015a). *Schlenker and Roberts* (2009) showed that the time in which a plant is exposed to a temperature above a threshold during each day of the growing season can explain almost half of its yield variations. For corn, this threshold is estimated to be 29 °Celsius. Thus, it is highly recommended to account for nonlinearity in temperature. This is particularly important in the context of climate change, as the likelihood of significant and non-marginal changes in relevant factors increases. Currently, non-linear measures with thresholds such as extreme degree days (EDD) are considered to be the best predictor of crop yield variation (*Auffhammer and Schlenker*, 2014; *Carleton and Hsiang*, 2016).

Recent research suggests, that the main reason of the importance of EDD is the high correlation with measures of cumulative evaporative demand (*Urban et al.*, 2015a), as for instance vapor pressure deficit (VPD, *Lobell et al.*, 2013; *Roberts et al.*, 2013). There is evidence, that the effect of EDD and measures for evapotranspirative demand is overstated when neglecting proper control for water supply (*Basso and Ritchie*, 2014; *Ortiz-Bobea*, 2013). For instance, soil moisture is considered a major limiting factor to maize growth (*Andresen et al.*, 2001). Extreme high temperature amplifies the impact of soil moisture deficit because of surface-atmosphere coupling (*Mueller and Seneviratne*, 2012), but the opposite is not necessarily the case as droughts occur independently of heat (*Basso and Ritchie*, 2014). *Urban et al.* (2015b) highlight the impact of interactive effects between VPD and water supply to further improve model predictability. In Germany, a recent statistical impact assessment of weather fluctuations affecting maize and winter wheat recognizes water shortage as major limiting factor (*Conradt et al.*, 2016; *Gornott and Wechsung*, 2015, 2016). These studies employ proxies to control for the primary source of water, such as precipitation and measures for evapotranspirative demand. The water holding capacity of the soil and the persistence of soil moisture is often not considered.

One basic assumption in EDD is that temperature effects are additive substitutable, which means that their impact is constant for all development stages of the plant. This assumption is rejected in both agronomic studies (*de Bruyn and de Jager*, 1978; *Sinclair and Seligman*, 1996; *Tubiello et al.*, 2007; *Wahid et al.*, 2007) and large-scale empirical analyses (*Berry et al.*, 2014; *Lobell et al.*, 2011b; *Ortiz-Bobea*, 2011; *Ortiz-Bobea and Just*, 2013). For example, the susceptibility to high temperatures is increased during flowering (i.e. tasseling, silkening, and pollination) and the reproductive period. Similar to heat measurements, the sensitivity to water stress is dependent on the development stage of the plant (*FAO Water*, 2016). For instance, it is shown for climate projections in India that a more uneven distribution of precipitation within a season overturns positive effects of an increase in total precipitation (*Fishman*, 2016). It is argued to control for intra-seasonal varying weather induced effects on crop yield variation. This issue is amplified for precipitation controls compared to temperature. The distribution of measures such as EDD partially overlaps with the sensitive phase of plant growth (see Figure A14 of *Schlenker and Roberts*, 2009), but precipitation, as control for water supply, is commonly aggregated for the entire growing season (*Annan and Schlenker*, 2015; *Burke and Emerick*, 2016; *Roberts et al.*, 2013; *Schlenker and Roberts*, 2006, 2009, among others). These studies do not explicitly account for seasonality of water supply related effects. Overall, controls for meteorological effects averaged over the entire season may bias the estimated dose-response function and diminish the predictive power of the models, because they do not account for the seasonal interaction between water supply and water demand (*Urban et al.*, 2015b).

Based on this analysis, it is the main aim of this study to investigate the intra-seasonal predictability of soil moisture to estimate silage maize yield in Germany. It is also evaluated how approaches considering soil moisture perform compared to those using meteorological variables. The examined hypothesis are, that a) models with soil moisture are better able to predict yield than meteorology-only approaches and that b) temporal patterns in the seasonal effects of the explanatory variables matter, i.e. there is no additive substitutability. In order to analyze these hypotheses, the intra-seasonal effects of soil moisture and meteorological variables for non-irrigated arable land in Germany are examined in this study. In detail, the following research questions are addressed: 1) Is there predictability of soil moisture additionally to meteorology? 2) If so, how does it compare to the one by meteorological determinants? 3) Is there temporal pattern in the seasonal effects of all explanatory variables (meteorology and soil moisture)? Along this analysis we also evaluate 4) how models based on different meteorological determinants perform compared to each other.

To answer this research questions, a reduced form panel approach is employed to examine the non-linear intra-seasonal partial effects of soil moisture anomalies and the meteorological variables temperature, potential evapotranspiration, and precipitation. For this purpose, we use a new data set which is additionally comprised of soil moisture anomaly data. The aim is to evaluate whether soil moisture anomalies have predictive skills and how the effects differ from those using only meteorological variables. Soil moisture and any derived index is highly autocorrelated in time and thus provide an integrated signal of the meteorological conditions in the preceding and subsequent months (e.g., *Orth and Seneviratne*, 2012; *Samaniego et al.*, 2013). This persistence does not allow for cumulative measures as those used for temperature, but it avoids the inflation of the error terms. Commonly, the predictive power of models only employing meteorological variables can be improved by accounting for the regional specific temporal distribution of the phenological stages (*Dixon et al.*, 1994). The integrated signal of the meteorological conditions provided by any measure derived from soil moisture, however, allows the employment of monthly averages to account for these intra-seasonal effects. In our study, it is implicitly controlled for the interaction of both variables controlling for water supply and water demand, because these show high correlation on a monthly basis. Different model configurations for each month of the growing season are compared by model selection criteria to allow conclusions about the effect of soil moisture anomalies on the explanatory power of the model and to test the assumption of additive substitutability. Further, the difference in explanatory power of models either using potential evapotranspiration or average temperature is evaluated. The partial effects of all covariates of the best model for each month are examined. For the purpose of a comprehensive examination, we also investigate the effects of wet anomalies.

12.3 Data

12.3.1 Yield Data

Annual yield data for silage maize are provided by the Federal Statistical Office of Germany for the administrative districts (rural districts, district-free towns, and urban districts) since the year 1999 (*Statistische Ämter des Bundes und der Länder*, 2017). The yield data are de-trended using linear regression for the period 1999 to 2015 to control for technical progress. A log transformation is applied to yield to better satisfy the normality assumption. This transformation also mitigates issues related to heteroscedasdicity and the estimates are less sensitive to outliers. All administrative districts with less than nine observations are removed from the analysis, because the influence of single observations points is too strong in these cases. The threshold nine has been chosen after exploring Cook's distance and evaluating the systematic omission of yield data by the administrative districts (*Cook*, 1977, 1979).

12.3.2 Soil Moisture Anomalies and Meteorology

The explanatory variables used in the study are the observed meteorological variables precipitation (P), average temperature (T), and potential evapotranspiration (E), as well as model-derived soil moisture. The mesoscale Hydrologic Model (mHM) has been used to estimate the soil moisture (*Kumar et al.*, 2013b; *Samaniego et al.*, 2010a). The model uses grid cells as primary unit and it accounts for various hydrological processes such as infiltration, percolation, evapotranspiration, snow accumulation, groundwater recharge and storage as well as fast and slow runoff. The parametrization process of the model is based on physical characteristic, as for instance soil texture. Three different forms of land cover are also integrated in the model, which are based on the CORINE Land Cover maps of 2006 (*European Environmental Agency*, 2009). However, no endogenous processes of land use management, as for instance drainage or irrigation, are considered within the model. The depth of the soil in each grid depends on the soil type used in mHM. Details can be found in *Zink et al.* (2017).

Soil moisture is further transformed into a soil moisture index (SMI), which is a non-parametric cumulative distribution function (cdf) derived from the absolute soil moisture estimated by mHM. A non-parametric kernel smoother algorithm has been used for the calculation of the cdf for each calendar month in accordance to the proposed method by *Samaniego et al.* (2013). It ranges from zero to one and represents an anomaly with respect to the monthly long term median in soil water (SMI $= 0.5$). Low values represent extreme dry soils and high values extreme wet ones. The SMI is calculated for entire Germany at a spatial resolution of 4 km. Monthly values of soil moisture are transformed to SMI for the period from 1951 to 2015. These values have also been used for drought reconstruction (*Samaniego et al.*, 2013). A similar procedure has been applied for the seasonal forecasts of agricultural droughts (*Thober et al.*, 2015).

The monthly SMI values are categorized into seven classes which follow the notion of the US drought monitor and the German Drought Monitor (*Zink et al.*, 2016). This stepwise approach allows to measure nonlinear effects of soil moisture. The dry categories SMI ≤ 0.1, $0.1 <$ SMI ≤ 0.2, and $0.2 <$ SMI ≤ 0.3 are denoted as severe drought, moderate drought and abnormally dry, respectively. The wet quantile intervals between $0.7 <$ SMI ≤ 0.8, $0.8 <$ SMI ≤ 0.9, and $0.9 <$ SMI are labeled as abnormally wet, abundantly wet and severely wet, respectively. The interval between $0.3 <$ SMI ≤ 0.7 serves as reference and characterizes normal situations. This classification uses location depend cdfs and thus allows comparison of classes across locations. In the rest of this, the terms soil moisture anomalies and soil moisture index (SMI) are used synonymously because of this categorization.

Daily data of precipitation and temperature are obtained from a station network operated by the German Weather Service (*Deutscher Wetterdienst*, 2017). Details on interpolation can be found in *Zink et al.* (2017). These daily values are also used to force mHM. For the analysis in this study, all daily values are aggregated to monthly ones conserving the mass and energy of the daily observations.

Further, we introduce Potential Evapotranspiration (E) as a measure for evaporative demand. E is calculated by the equation of *Hargreaves and Samani* (1985) based upon extraterrestrial radiation and temperature and is estimated in millimeter per day:

$$\mathrm{E} = \kappa \mathrm{R} \sqrt{\mathrm{T}_\delta}(\mathrm{T} + 17.8), \tag{12.1}$$

where κ is a free parameter ($^\circ\mathrm{C}^{-1.5}$) that compensates for advection of water vapor (mm d^{-1}), R is extraterrestrial radiation converted into equivalent water evaporation, and T_δ is the temperature difference between daily maximum and daily minimum temperature ($^\circ$C). The term $\mathrm{T} + 17.8$ is an approximation of saturated vapour pressure, whereas the term T_δ is an approximation of cloudiness. 17.8 is an empirical constant found by calibration.

More complex alternatives exist, as for instance the standard method of United Nations Food and Agriculture Organization after Penman and Monteith (*Monteith*, 1981). These data for example use net radiation that is more difficult to estimate at the national scale in comparison to temperature particularly due to the lack of consistent observations. Similar to Vapor Pressure Deficit, which has been introduced as effective crop yield predictor (*Lobell*, 2013; *Roberts et al.*, 2013), potential evapotranspiration has a more direct physical link to crop water requirements than temperature. One goal of this study is to evaluate whether potential evapotranspiration provides improved yield estimates in comparison to temperature.

All meteorological variables are standardized to ease the comparison among different months. After this transformation, the variables have a mean of zero and a standard deviation of one. The original mean and standard deviation of the meteorological variables are depicted in Table 12.1 for completeness.

Table 12.1 Mean and standard deviation of the meteorological variables, averaged over Germany. Data are obtained by the Germany Weather Service.

	May		June		July		August		September		October	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
P (monthly sum in mm)	75.74	39.84	69.71	33.15	89.48	39.72	84.04	43.68	63.88	32.62	57.72	27.28
T (monthly average in °C)	13.46	1.42	16.52	1.45	18.48	1.74	17.90	1.57	14.07	1.63	9.64	1.83
E (monthly average in mm)	115.23	12.15	133.42	12.21	139.10	16.52	115.24	13.55	70.33	8.73	36.82	4.69

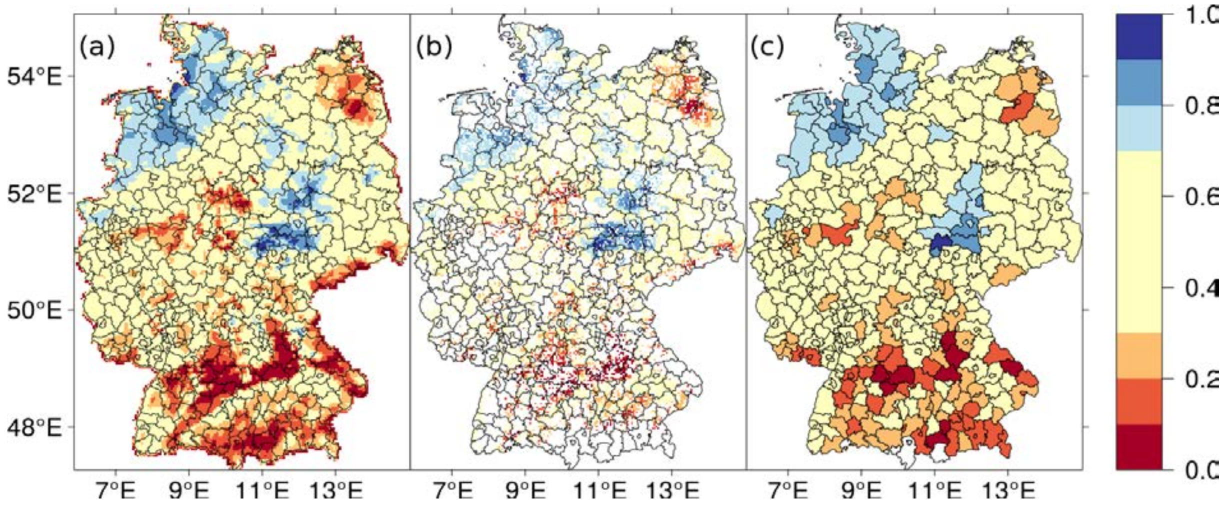

Figure 12.1 Illustration of the spatial processing of the SMI data of May 2003. On the left side, one can see the SMI with the $4 \times 4\ \mathrm{km}^2$ grids. In the middle, the data are masked with the $0.1 \times 0.1\ \mathrm{km}^2$ *non-irrigated arable land*-class of the CORINE Land Cover. Those data are than averaged over all the grid cells which are inside an administrative district. This is done for each district and the map on the right is derived. The processing steps shown in panel (a) and (b) are shown here exemplary for the soil moisture index for consistency, but these processing steps are applied to soil moisture fractions.

12.3.3 Spatial Processing

The explanatory variables (meteorology and soil moisture) are mapped onto the level of administrative districts to align with the spatial scale of the yield data. Maps st the different processing steps are shown in Fig. 12.1. Figure 12.1a depicts the $4 \times 4\ \mathrm{km}^2$ grid. These absolute soil moisture fractions are masked for *non-irrigated arable land*-class of the CORINE Land Cover (2006) at a $0.1 \times 0.1\ \mathrm{km}^2$ resolution to account for the variability due to heterogeneous land use within the administrative districts (Fig. 12.1b). The 0.1 km values are then averaged for each of the administrative district to obtain district level values (Fig. 12.1c). Blank administrative districts occur because of the absence of *non-irrigated arable land* grid cells. These processing steps are also applied to the meteorological variables (P, T, E). The soil moisture fractions of each administrative district is then transformed into a percentile index (SMI) using the kernel density estimator explained above (*Samaniego et al.*, 2013; *Thober et al.*, 2015; *Zink et al.*, 2016). An index reduces the probability of measurement errors and the estimated coefficients in the regression models are supposed to be less prone to attenuation bias (*Auffhammer and Schlenker*, 2014; *Fisher et al.*, 2012; *Hsiang*, 2016).

12.4 Regression Analysis

The main aim of this study is the identification of the monthly effects of soil moisture anomalies on crop yield. The model relates silage maize yield deviation (Y) to a stepwise function of soil moisture anomalies (SMI) and polynomials of the meteorological variables (P, T, E). Also, an error term is included which is composed of the fixed effects (c), a time-invariant categorical administrative district identifier, and the observation-specific zero-mean random-error term, which is allowed to vary over time (ϵ). Each monthly model can be written as:

$$
\begin{aligned}
Y_{ik} = & \sum_{n=1}^{6} \alpha_n \mathrm{I}(SMI_{ikm} \in \mathrm{C_n}) \\
& + \sum_{j=1}^{3} \beta_j (\mathrm{P}_{ikm})^j + \sum_{j=1}^{3} \gamma_j (\mathrm{T}_{ikm})^j + \sum_{j=1}^{3} \delta_j (\mathrm{E}_{ikm})^j \\
& + c_{im} + \epsilon_{ikm}.
\end{aligned} \tag{12.2}
$$

The index i represents the administrative districts, k the years, and m each month of the growing season, while the superscript j refers to degrees of the polynomials. $\mathrm{I}(\cdot)$ is the indicator function of the soil moisture categories $\mathrm{C_j}$, being 1 if the SMI belong to class n and 0 otherwise. The six classes are defined as severe drought (SMI ≤ 0.1), moderate drought ($0.1 <$ SMI ≤ 0.2), abnormally dry ($0.2 <$ SMI ≤ 0.3), abnormally wet ($0.7 <$ SMI ≤ 0.8), abundantly wet ($0.8 <$ SMI ≤ 0.9) and severely wet ($0.9 <$ SMI), respectively. The estimated coefficients of the model are α, β, γ, and δ and are constrained to be the same for all administrative districts. Time-invariant differences between administrative districts are taken into account by the fixed effects. These consist of the districts specific mean values of the individual variables on the right and left side of the equation.

Table 12.2 Comparison of Pearson Correlation Coefficients of the Exogenous Variables. Absolute values of the Pearson Correlation Coefficients are employed to calculated the averages presented in the last two columns.

	May	June	July	August	September	October	Average	Avg. June to Aug.
E / T	0.84	0.86	0.92	0.84	0.65	0.4	0.75	0.87
E / P	−0.38	−0.38	−0.52	−0.52	−0.56	−0.15	0.42	0.47
P / T	−0.31	−0.22	−0.54	−0.47	−0.47	−0.06	0.35	0.41
SMI / E	−0.27	−0.28	−0.44	−0.49	−0.46	−0.02	0.33	0.40
SMI / P	0.19	0.31	0.43	0.43	0.5	0.09	0.33	0.39
SMI / T	−0.04	−0.16	−0.35	−0.35	−0.27	0.13	0.22	0.29

The explanatory variables are correlated to each other (Table 12.2). Thus, higher non-orthogonal polynomials induce singularity in the moment matrix which cannot be inverted as required by the ordinary least-squares estimation of the coefficient. The polynomials are limited to degree three to avoid this and other detrimental consequences of multicollinearity such as the inflation of the standard errors. Additionally, E and T are treated as mutually exclusive because of the high correlation of E and T (Table 12.2). The coefficients γ or δ are set to 0, accordingly.

In addition to soil moisture, a meteorological and a fixed effect term is included. The fixed effects potentially reduce omitted variable bias, because they take into account the time-variant confounding factors specific to each spatial unit, such as average weather conditions and the water storage capacity of the respective soil. It is also assumed that farmers have optimized the entire production process at their location given their experience about that location. Soil and plant management, such as the choice of varieties, is adapted based on this long term experience. Therefore, the coefficients of the exogenous variables are determined on the basis of year-to-year variations. By restricting the coefficients to be same in all administrative districts, it is implicitly assumed that the response of plants to inter-annual stressors is the same across all locations. Differences in the sensitivity to exogenous weather and soil moisture fluctuations implied by the use of different silage maize varieties could thus be neglected by the model. If it is also assumed that these interannual fluctuations in weather and soil moisture are not fully taken into account by the farmer in the cultivation decisions, this corresponds to a randomised allocation of the farmer to a treatment group and can therefore be regarded as a natural experiment (*Auffhammer and Schlenker*, 2014; *Schlenker and Roberts*, 2009). The outlined interpretation of the coefficients is particularly suitable for SMI, because this index, which describes deviations from the median, is per definition an anomaly.

Endogenous variables are not included because these are considered as bad control in frameworks as those defined by *Angrist and Pischke* (2008). For instance, prices are affected by weather realizations and climate and are thus defined as endogenous (*Gornott and Wechsung*, 2015, 2016; *Hsiang*, 2016; *Hsiang et al.*, 2013). Other studies additionally use annual fixed effects and interaction terms of both time and entity specific fixed effects to control for time specific confounding factors (e.g., *Moore and Lobell*, 2014). These terms are not used in this study because annual variation should be explicitly accounted for by the weather variation of the exogenous variables. Annual fixed effects would diminish the entity specific inter-annual variation of the exogenous variables and thereby potentially amplify measurement errors (*Fisher et al.*, 2012).

Various estimation approaches are used to evaluate the quality of the models. Models can be distinguished by the explanatory variables they use and the degree of polynomials in the meteorological terms. The maximum number of parameters estimated in a model is 12. The Bayesian Information Criteria (BIC) is used for model selection in the next section. The BIC is composed of the maximum of the likelihood function for a particular set of variables as well as a penalty term (*Schwarz*, 1978). The latter adjusts the model selection criterion for the number of parameters to account for over-fitting. This allows to choose across models with different number of variables. The BIC criterion imposes a higher penalty on over-fitting compared to other model selection criteria based on maximum likelihood such as the Akaike Information Criterion (*Akaike*, 1973b). The penalty particularly affects the soil moisture anomaly term because it always adds six parameters. Overall, the model with the lowest BIC is preferred. To derive the BIC, a generalized linear model is fitted using the *glm* function (*R Core Team*, 2015).

Additionally, the models are evaluated according to their adjusted coefficient of determination (adj. R^2, Section 4.2). Ordinary least squares using the *lm* function (*R Core Team*, 2015) are employed with a dummy variable for each administrative districts to explicitly account for the fixed effects. As default, a demeaning framework (*Croissant and Millo*, 2008) has been applied to investigate the model performance in terms of R^2. The demeaning framework involves converting the data by subtracting the administrative district average from each variable. The estimated coefficients are the same for the least squares dummy variable regression, a demeaning framework, and maximum likelihood (BIC). This is in accordance to theory that normal distributed error terms estimators based on maximum likelihood and least squares are the same.

The standard errors of the coefficients are corrected for spatial autocorrelation. For this purpose, the Robust Covariance Matrix Estimator proposed by *Driscoll and Kraay* (1998) is employed to construct standard errors based on asymptotic formulas. No weights capturing decaying effects in space are used because the administrative districts have different areas and the spatial extent of SMI occurrences is heterogeneous. This can be regarded as comparable to block-bootstrapping on country-level, which has been used in many comparable studies relying on re-sampling methods (e.g. *Butler and Huybers*, 2015; *Moore and Lobell*, 2014, 2015; *Urban et al.*, 2015a,b). Further, serial correlation and heteroscedasdicity is also controlled for (*Arellano*, 1987; *White*, 1980). Overall, this approach is rather conservative but in alignment with the proposal of *Angrist and Pischke* (2008) to take the largest robust standard error as measure of precision.

12.5 Results and Discussion

12.5.1 Qualitative evaluation of different model configurations within the growing season

In this section, the Bayesian Information Criterion (BIC) is applied to evaluate the best combination with respect to soil moisture, meteorological variables, and the polynomial degrees of the latter. The BIC is calculated separately for each month to assess the intra-seasonal variability.

The distribution of the BIC for the various model configurations is presented in Fig. 12.2., which shows one panel for each month of the growing season. Within the panels, models with different variable combinations in the meteorological term are separated by vertical lines. A model configuration is defined by a set of meteorological variables, the polynomial degree of each variable, and the stepwise function of the soil moisture anomalies. The complexity of the configurations increase stepwise from the left to right within each panel. The model employing SMI as single explanatory variable is represented by a point on the left in each panel. The black markers indicate the models with soil moisture and gray markers without. The models 02 - 07 employ one meteorological variable each. These have three markers for the different degrees of the polynomials. The models 08 - 11 entail two meteorological variables and thus have nine markers.

The explanatory power is different across the months as indicated by the lowest marker within each panel. Overall, July has the highest explanatory power. Nonlinear meteorological terms improve the fit of the model on the data in all model configurations (not shown). The preferred polynomial in the meteorological term is of degree three. The only exception is June, where the best model employs a second degree polynomial for P. These observations are consistent with agronomic studies. Curvilinear relationships between maize yield and meteorological variables are already investigated in previous research. The rationale behind this is that optimal conditions exist for certain growth stages and deviations from them are detrimental. For example, *Thompson* (1969) found for corn in the U.S. Corn Belt that precipitation in July above and temperature in August below the monthly average is desirable. Nonlinear configurations have been neglected so far in comparable approaches employing constant elasticity models in Germany (*Conradt et al.*, 2016; *Gornott and Wechsung*, 2015, 2016).

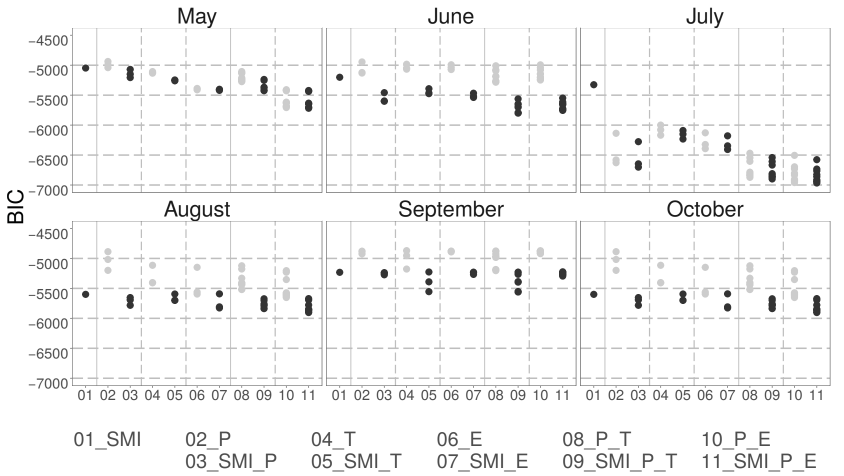

Figure 12.2 Each panel shows the BIC distribution of one month. Within the panels various models are compared, whilst the lowest marker is preferred. Each column represents a particular selection of variables. The markers represent different degrees of the polynomials in the meteorological term. The gray markers denote those models that neglect the SMI, whilst the black include it.

The composition of the meteorological term is evaluated by comparing the gray markers in Fig. 12.2. It is possible to asses the impact on the model fit of the single variables P, T, and E by the comparison of the configurations 02, 04, and 06, respectively. In May, most of yield variation is explained by E. In June and July, P contributes to model fit the most. In July, for instance, the explanatory power of a nonlinear P term is almost as good as the best combined configuration. September and October are determined by T. However, in most months, using more than one meteorological variable results in the highest explanatory power. The only exception is October, where model 05 (SMI & T) exhibits the lowest BIC.

The difference in BIC between configuration 08 (P & T) and 10 (P & E) is small from June to August. This result can be expected because T and E are highly correlated in our sample. The models with mixed meteorological terms in July and August slightly prefer E, while in June it is T. In the other months, the difference between T and E is comparatively larger. In May, E is preferred, and in September and October T is the better measure. Both measures, T and E, account for similar determinants of silage maize growth. The latter, however, is more complex because it contains information on sub-daily radiation additionally to daily temperature (*Hargreaves and Samani*, 1985). It can be assumed that this additional information are averaged out using monthly values and monthly temperature becomes a close estimate of monthly E. This is in alignment with results on different time resolutions, which indicate that measures of evapotranspirative demand are highly correlated with temperature extremes (*Lobell et al.*, 2013; *Roberts et al.*, 2013). Therefore, it is sufficient to account for temperature when simultaneously controlling for water supply (P, SMI) because it is easier to measure temperature data and there is a smaller chance of attenuation bias.

The extent of the model improvement by adding soil moisture anomalies varies across the months. This can be evaluated by comparing the gray and black markers in Fig. 12.2. Including soil moisture anomalies only improves model fit to a little extent in May and July. In all the other months, large improvement can be made when additionally controlling for soil moisture. In the second half of the season, i.e. August and September, the models using only SMI have a similar or even lower BIC compared to all meteorology-only models.

These results indicate that soil moisture builds memory over the season that adds relevant information, which are not integrated in the monthly meteorological variables. There are several reasons for this postulation remark. First, the seasonality of soil moisture must be considered. The fraction of the saturated soil changes over time and thus the base value for the index. For Germany, this seasonality is depicted in Fig. 4 in *Samaniego et al.* (2013). In March, soil water content is the highest while soils are usually driest in August and September. This also implies, that an agricultural drought has a lower absolute soil moisture in August and September compared to the preceding months. Second, the anomalies in the later months integrate information about the water balance in the preceding months because of the persistent character of soil moisture (evident from the autocorrelation of the soil moisture indexes). For instance, extreme dry conditions during flowering and grain filling are reflected in a dry soil moisture

Table 12.3 Comparison of the adjusted Coefficient of Determination R^2. The results from the demeaning framework serve as reference to the ones obtained by Least Square Dummy Variable Regression (LSDV). The latter explicitly accounts for the fixed effects. Additionally model configurations without either T, P, or SMI are shown.

	May	June	July	August	September	October	Average	June - August
(a) Adjusted R^2 demeaning	0.11	0.16	0.31	0.17	0.13	0.12	0.16	0.21
(b1) Adjusted R^2 LSDV	0.56	0.59	0.66	0.59	0.57	0.56	0.59	0.61
(b2) ((b1) − (a)) / (a) in %	409.1	268.8	112.9	247.1	338.5	366.7	290.5	209.6
(c1) Adjusted R^2 no T	0.07	0.13	0.28	0.16	0.08	0.08	0.13	0.19
(c2) ((c1) − (a)) / (a) in %	−36.4	−18.8	−9.7	−5.9	−38.5	−33.3	−23.7	−11.4
(d1) Adjusted R^2 no P	0.08	0.11	0.22	0.14	0.12	0.12	0.13	0.16
(d2) ((d1) − (a)) / (a) in %	−27.3	−31.3	−29.0	−17.6	−7.7	0.0	−18.8	−26.0
(e1) Adjusted R^2 no SMI	0.07	0.08	0.30	0.11	0.06	0.07	0.11	0.16
(e2) ((e1) − (a)) / (a) in %	−36.4	−50.0	−3.2	−35.3	−53.8	−41.7	−36.7	−29.5

anomaly in the second half of the agricultural season of silage maize. The observation, that the SMI represents additional information to the meteorology is also pronounced by the fact that the pairwise correlations including SMI are lower compared to any other combination of the exogenous variables (Table 12.2). Further, dry anomalies in the late part of the season may indicate a long lasting water shortage condition, as soil moisture drought lasts over several month or potentially even years (*Samaniego et al.*, 2013; *Sheffield and Wood*, 2011; *Zink et al.*, 2016).

Similar results may be achieved by cumulated measures of the meteorology or the climatic water balance. However, the comparison of soil moisture measurements and different cumulates of precipitation (one to six months) shows that it would be necessary to consider different precipitation accumulations for different sites in order to include the same information as for soil moisture (not shown). For example, Southern Germany exhibits higher water retaining capacities and also higher correlation with three month precipitation as compared to Eastern Germany. Further, a substantial share of the variability of soil moisture is not explained by precipitation (the mean coefficient of determination is at most 50 %). One advantage of using soil moisture in such a study is that the coefficients can be restricted to be the same at all locations, whilst assuming that the water retaining capacity is not the same everywhere.

In summary, soil moisture anomalies improve the model fit in all model configurations. This is the case even though soil moisture is strongly affected by the penalty for additional parameters within the BIC. Further, the evidence of nonlinear effects in the meteorological terms is confirmed. The results also indicate that there is substantial seasonal variability in the impact of exogenous variables. This is investigated further quantitatively in the next sections for the meteorological variables and soil moisture.

12.5.2 Quantitative Assessment: Coefficient of determination for models using different explanatory variables

In this and the next section (4.3), we present the quantitative results for the ”full” model with polynomials of degree three of the variables temperature (T) and precipitation (P) in the meteorological term and additionally the soil moisture anomalies (SMI). Using the same model configuration for each month allows the comparison of partial effects and ensures that the source of variation is the same within the meteorological term (*Auffhammer and Schlenker*, 2014). In this section, the coefficient of determination is employed to evaluate the share of the sample variation only explained by the exogenous variables. Additionally, it is used to assess the in-sample goodness of fit of the models 03 (SMI & P), 05 (SMI & T), 08 (P & T), and 09 (SMI & P & T), each using polynomials of degree three.

The coefficients of determination for two model settings are evaluated to show the ability of exogenous explanatory variables, e.g. the meteorological term and the soil moisture anomalies, to improve the in-sample goodness of fit of the full model: first, the model that only accounts for the variation in the exogenous explanatory variables, which is derived by the demeaning framework (row (a) in Table 12.3); second, the least squared dummy variable model that accounts for both the variation in the exogenous explanatory variables and the administrative district specific average yield (row (b1) in Table 12.3). The ratio of the coefficient of determination derived by these two model setups is investigated (row (b2) in Table 12.3) to quantify the share of variance explained only by the exogenous explanatory variables, e.g. the meteorological term and soil moisture anomalies. Expectedly, the exogenous variation in weather and soil moisture improves the model fit in all months, but the level of improvement varies. The month which gains the least in explanatory power when additionally accounting for the share of variation explained by the average crop yield of each administrative district is July (+ 112.9 %). This suggests that a large part of the yield variation is explained only by exogenous explanatory variables. The month with the greatest variation, which is explained only by the average yield of the districts, is May. During this month, 409.1 % of the explanatory power is added if the average yield of each county is explicitly taken into account in comparison to the models that only use soil moisture and weather variation as explanatory variables (line (b2) in table 3).

The adjusted R^2 presented in this study explicitly including fixed effects for each month of the period June (0.59), July (0.66), and August (0.59) is comparable to related approaches. *Urban et al.* (2015b), who employed a comparable period to estimate their results, reported R^2 of 0.65 and 0.67 for a model that successfully accounts for the interaction between heat and moisture for a 61 - 90 day period following sowing for Iowa, Illinois, and Indiana. Their study additionally employed time fixed effects which usually lead to higher R^2. The seminal approach employing extreme degree days (EDD, *Schlenker and Roberts*, 2009) reported R^2 between 0.77 and 0.78. In their sample, a comparatively large share of the variation was explained by the fixed effects and trend, which exhibited an R^2 of 0.66. A study using updated data of *Schlenker and Roberts* (2009) and controlling for evaporative demand in July and August achieved adjusted R^2 between 0.66 and 0.72 (*Roberts et al.*, 2013).

In the previous section, all the models have been evaluated with respect to the BIC criterion which penalizes over-fitting. The focus here is on reducing the sample bias of the model. The in-sample adjusted R^2 of the models is additionally compared when either one of the variables SMI, P, or T is not considered (rows (c1) - (e1) in Table 12.1). The according relative change in model fit when one variable is removed from the full model can be found in rows (c2) - (e2) of Table 12.3. In all months but May and July, the strongest loss in in-sample goodness of fit is seen for removing soil moisture (for instance - 50.0 % in June and - 35.3 % in August). In July, which is the month with the highest overall in-sample-goodness of fit, the largest effects is accounted for by precipitation (- 29.0 %). The average relative model loss is largest for soil moisture for the entire season (-36.7 %) as well as the period June to August (-29.5 %). As observed in the section before, the effect of each particular variable is dependent on the month. For instance, the largest relative loss in adjusted R^2 for SMI is estimated in June (- 50.0 %) and September (- 53.8 %). The largest effect of precipitation is observed in June (- 31.3 %) and July (- 29.0 %). Temperature is relevant the most in September (- 38.5 %) and May (- 36.4 %).

To summarize, the in-sample explanatory power of the full models are comparable to those reported in the previous literature. The largest average gain in goodness of fit is achieved by including SMI. In July, the month with the largest in-sample goodness of fit, most of the variation in yield is explained by precipitation. This section has only presented a quantitative analysis of the explanatory power in terms of adjusted R^2. A detailed assessment of the partial functional form of individual explanatory variables is presented in the next section to better understand their ceteris paribus impact on the crop yield.

12.5.3 Quantitative Assessment: Partial Effects of the Meteorological Variables

A better understanding of the relationship between individual explanatory variables allows to design effective adaptation measures. The partial functions of the meteorological covariates are presented in the next two sections and those of soil moisture in section 4.3.3. Those functional forms, which are significant at least in the first or second order, are presented for individual months in Fig. 12.3. The range of the meteorological variables is depicted from - 2 to + 2 standard deviations (SD). It can be assumed that larger deviations from the mean are related to higher uncertainties in the estimated crop yield. A table with the estimated coefficients and standard errors of all models can be found in Table 12.4.

Partial Effects of Precipitation The partial precipitation effects for the months May to August are shown in Panel a) of Fig. 12.3. Given constant soil moisture and temperature effects, negative precipitation anomalies are associated with reduced yield in these months. The largest effect is observed for June (- 5 % at - 1 SD) and July (- 6.5 % at - 1 SD). These are the overall most significant months, but with different patterns compared to the remaining two. In June and July, more than average precipitation is associated with comparatively higher yield (at 1 SD: + 2.2 % in June and + 2.1 % in July), whilst the opposite is the case for May and August.

Table 12.4 Results of Regression Models employing precipitation and temperature to account for meteorology (both with polynomials of degree 3, superscripts denote the degree of individual polynomials) and a stepwise function of SMI.

	Dependent Variable: log(Silage Maize)					
	Model of the month					
	May	June	July	August	September	October
Precipitation^1	0.004	0.036***	0.039***	−0.014	−0.011	−0.003
	(0.011)	(0.014)	(0.013)	(0.011)	(0.013)	(0.010)
Precipitation^2	−0.023*	−0.014*	−0.023***	−0.019***	−0.005	0.002
	(0.014)	(0.007)	(0.004)	(0.006)	(0.005)	(0.008)
Precipitation^3	0.004	0.001	0.005***	0.004***	0.002	−0.0001
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Temperature^1	0.024	−0.006	−0.036*	−0.003	0.038	−0.002
	(0.021)	(0.015)	(0.021)	(0.014)	(0.024)	(0.018)
Temperature^2	−0.005	−0.006	−0.007***	−0.008**	−0.009*	−0.016**
	(0.007)	(0.006)	(0.002)	(0.003)	(0.005)	(0.008)
Temperature^3	0.0004	−0.002	0.004*	−0.002	−0.013*	0.005
	(0.003)	(0.003)	(0.003)	(0.002)	(0.006)	(0.003)
SMI: severe drought	0.068***	0.024	−0.044**	−0.110***	−0.126***	−0.149***
	(0.012)	(0.020)	(0.019)	(0.035)	(0.028)	(0.037)
SMI: moderate drought	0.044***	0.016	−0.007	−0.055***	−0.041*	−0.024
	(0.011)	(0.017)	(0.011)	(0.017)	(0.023)	(0.030)
SMI: abnormal dry	0.011	0.023***	−0.005	−0.024**	−0.017	−0.005
	(0.011)	(0.007)	(0.007)	(0.011)	(0.015)	(0.017)
SMI: abnormal wet	−0.007	−0.034***	−0.011	0.026***	0.007	−0.006
	(0.014)	(0.011)	(0.007)	(0.008)	(0.011)	(0.019)
SMI: abundant wet	−0.014	−0.052**	−0.004	0.027***	0.012	−0.001
	(0.020)	(0.025)	(0.009)	(0.008)	(0.017)	(0.015)
SMI: severe wet	−0.009	−0.202***	−0.041***	0.037***	0.030	0.025
	(0.019)	(0.047)	(0.016)	(0.013)	(0.027)	(0.017)
Observations	5,376	5,376	5,376	5,376	5,376	5,376
R^2	0.113	0.173	0.326	0.179	0.136	0.129
Adjusted R^2	0.105	0.162	0.305	0.168	0.127	0.121
F Statistic	53.151***	87.531***	203.025***	91.409***	65.891***	62.296***

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The results indicate the importance of sufficient water supply provided to plants by precipitation, especially in June and July. In Germany, the begin of flowering is usually in July and extends into August (based on data provided by the German Weather Service - *Deutscher Wetterdienst*, 2017). Maize plants are susceptible to water stress during this growing phase (*Barnabás et al.*, 2008; *Bolaños and Edmeades*, 1996; *Fageria et al.*, 2006; *Grant et al.*, 1989). Despite the necessity to control for intra-seasonal variability of precipitation effects, explicitly controlling for this sensitive phase is not very common in recent reduced form studies (*Carleton and Hsiang*, 2016). Notable exceptions are *Lobell et al.* (2011b), who used precipitation centered around flowering (anthesis) in statistical models based on historical data of trials in Africa, and *Ortiz-Bobea and Just* (2013), who controlled for the vegetative, flowering, and grain-filling stages. Instead, many approaches employ total precipitation over the growing season (*Annan and Schlenker*, 2015; *Burke and Emerick*, 2016; *Roberts et al.*, 2013; *Schlenker and Roberts*, 2006, 2009), monthly mean growing season precipitation (*Urban et al.*, 2012) or the average of a subset of the season (*Urban et al.*, 2015a). Studies for Germany commonly separate the season into the periods May to July and August to October (*Conradt et al.*, 2016; *Gornott and Wechsung*, 2015, 2016), thus dividing exactly the time interval most susceptible to water stress and averaging over periods with diverse effects (e.g. May and June in Fig. 12.3a). This may hide water related effects. Other studies neglect precipitation entirely and only rely on temperature measures

(*Butler and Huybers*, 2013, 2015; *Schlenker et al.*, 2013). According to their results, the explanatory power is not improved when adding precipitation. This is contradictory to our observations that precipitation is particularly relevant (see also Section 12.5.1 & 12.5.2).

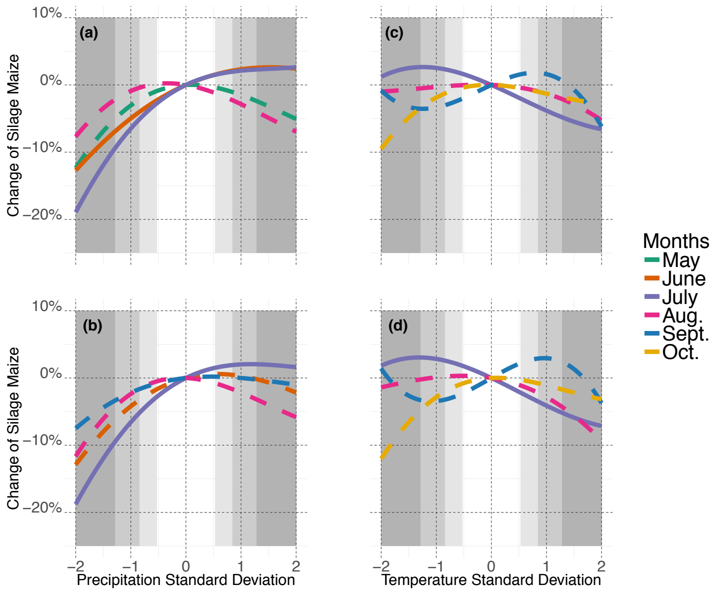

Figure 12.3 The partial dose-response functions of the meteorological variables are depicted for the range between - 2 and + 2 standard deviations (SD). The upper row represents those models considering SMI, whilst the lower row neglects SMI. A solid line is used for those variables which are significant in both the first and second degree polynomials. A dashed line is employed if only one of the first two polynomials is significant. The vertical axis represents the change in silage maize converted into % approximated by the formula $100(\exp(\sum_{j=1}^{3}\beta_j(\mathrm{x_{ikm}})^j)-1)$, where x_{ikm} is either precipitation or temperature. Under the assumption that the variables are normally distributed, the range depicted accounts for about 95 % of the observations. The dark gray areas denote the interval between the 0.023 % (- 2 SD) and the 10 % as well as the 90 % and 97.7 % (+ 2 SD) quantile. Similar, in medium gray the range between either the 10 % and the 20 % and the 80 % and 90 % quantiles is marked. The light gray quantifies the impact between the between either the 20 % and the 30 % and the 70 % and 80 % quantiles.

The models employed here do not explicitly account for interactions between the meteorological and the soil moisture terms. Nevertheless, soil moisture is a function of the meteorological variables and all effects are correlated to each other (see Table 12.2). The overall pattern in the effects of the meteorological variables only changes to a small extent when estimating the standard model configuration without the term for soil moisture anomalies (Fig. 12.3b). One of the most pronounced differences is that the positive effect of precipitation in June diminishes when not accounting for soil moisture. The coefficients in June are also less significant. The effects in September become significant in the second and third polynomial degree when not considering SMI (blue dashed line in Fig. 12.3b). On the contrary, May is less significant and thus not included in this panel. SMI improves the model fit but only slightly affects the functional form of precipitation, which highlights that soil moisture adds relevant but different information as those entailed in precipitation. The next section presents an analogue analysis for temperature.

Partial Effects of Temperature The significant partial temperature effects are depicted in Fig. 12.3c. A significant effect in all polynomials is only estimated for July, whilst in May and June, no significant coefficients can be found at all. In all months but September, higher than average temperatures are associated with reduced crop yield. The extent of the effects, however, varies over time. In July, less than average temperature is associated with above-normal crop yield. The estimated function peaks at - 1.24 SD, which is 16.18 °C (mean in July is 18.34 °C). Additional 2.66 % crop yield can be expected at this temperature, all other variables hold constant. In August, elevated temperatures are associated with negative effects. September exhibits a large but not significant linear effect, whilst the second and third polynomials are significant. Because maize is maturing during this time, higher temperatures up to a threshold are favorable as shown in Fig. 12.3c. Crop yield is reduced beyond this threshold, which might be related to heat waves. Cold temperatures have a negative effect in October, which is the strongest one observed. Harvesting commonly begins at the end of September within the period from 1999 to 2015 (*Deutscher Wetterdienst*, 2017). Thus, low temperatures may be related to early harvesting and result in lower yield.

When comparing the effects of precipitation and temperature in the months most relevant for meteorology, i.e. June and July, those of precipitation clearly outweigh temperature. The largest effects can be found for negative anomalies of precipitation in July (compare Fig. 12.3a and Fig. 12.3c). The limited effect of temperature is in alignment with agricultural literature, which states that maize is tolerant to heat as long as enough water is provided

(*FAO Water*, 2016). This is also the case in our study area given the fact that Germany lies in a rather temperate and marine climate zone. Additionally, sufficient provision of water is associated with prolonged grain filling and hence diminished heat sensitivity (*Butler and Huybers*, 2015). Recent literature often neglected precipitation and emphasized mostly extreme temperature instead (*Carleton and Hsiang*, 2016; *Lobell et al.*, 2008, 2011a; *Schlenker and Lobell*, 2010; *Schlenker et al.*, 2005), which may have lead to biased assessments.

The general functional form of temperature are hardly affected by neglecting SMI (Fig. 12.3d). For example, crop yield changes from one - 3.82 % with SMI to - 4.11 % without for one SD of elevated temperature in July. These effects are smaller than those seen for precipitation, which highlights again that soil moisture provides an information that is independent to the one provided by T.

As mentioned before, a substantial amount of studies employed temperature as the major explanatory variable neglecting knowledge about plant physiology and plant growth (*FAO Water*, 2016; *Wahid et al.*, 2007). The functional form of the partial temperature effects derived from different model configurations for July and August is presented in Fig. 12.4 to evaluate the magnitude of bias between the full model (presented in Fig. 12.3) and a temperature-only model.

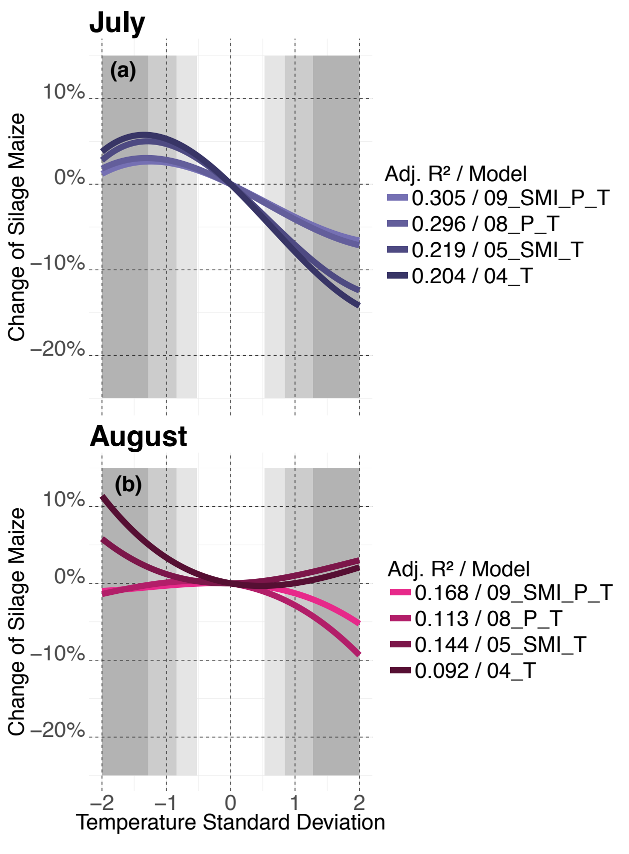

Figure 12.4 Sensitivity of the functional form of temperature partial effects for various controls for water supply.

In both months, the in-sample explanatory power is reduced compared to the full model when only using temperature as explanatory variables. In July, the model fit is - 34.2 % lower when employing the temperature only model compared to the full model, while it is - 45.9 % in August (Fig. 12.4). In July, the in-sample goodness of fit is affected stronger by removing precipitation (- 29.0 %) than by doing so for SMI (- 3.2 %), (Table 12.3). This is not surprising because the partial effect of precipitation in July is largest, whilst soil moisture anomalies only show negligible effect. On the contrary, considering SMI in August (- 35.3 %) exceeds the losses in Adjusted R^2 compared to a model without precipitation (- 17.6 %), (Table 12.3). In July, the functional form stays qualitatively the same across all model configurations (Fig. 12.4a). The magnitude of the effects is, however, larger when precipitation is not considered. In August, the temperature effect is elevated by not considering SMI. Taking out precipitation reverses the effects found for the full models. This observation clearly demonstrates that adequate control of water supply is necessary to derive non-biased estimates of partial temperature effects. These results also indicate that the biases seen for different model configuration depend on the month considered. Overall, a model using only temperature as explanatory variable has larger partial effects and potentially even different ones with regard to the direction compared to those of the full model. In the next section, the partial effects of the soil moisture index are investigated.

Partial Effects of the Soil Moisture Index (SMI) Similar to the meteorological terms, the susceptibility to SMI changes over the months (Fig. 12.5). In particular, a change in the general patterns can be observed. In May and June, dry conditions are associated with positive yield (up to + 7 % in May, and + 2.3 % in June), whilst wet conditions are harmful (up to - 18.3 % under severely wet conditions in June). In July, both extremes have negative impacts of around - 4 %. In all of the following months, dry conditions are associated with reduced crop yield (up to - 10.4 % in August, - 11.8 % in September, and - 13.8 % in October), whilst only extreme wet conditions in August are positive for annual silage maize yield (up to + 3.77 %). These deviations are as high as the ones observed for the meteorological variables (Fig. 12.3).

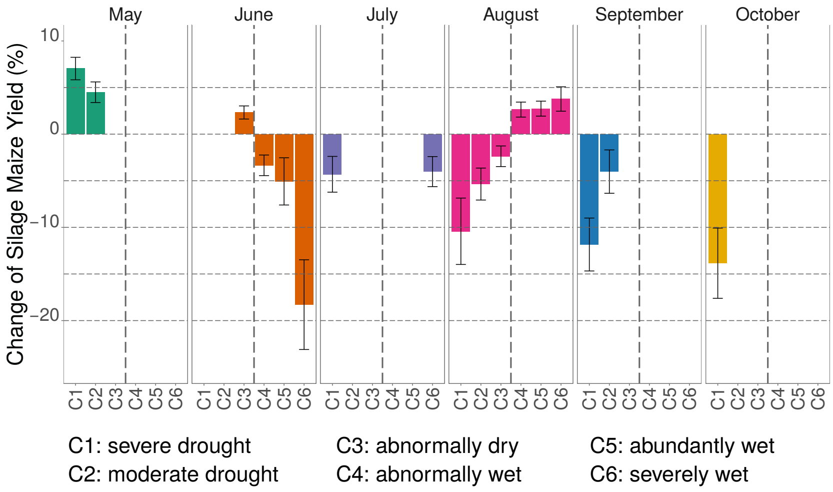

Figure 12.5 Percentage Change of silage maize yield caused by significant Soil Moisture Anomalies for each month. The vertical axis represents the change in silage maize converted into % approximated by the formula $100(\exp(\sum_{j=1}^{6} \alpha_j \mathrm{I}(\mathrm{SMI}_{ikm} \in \mathrm{C_j}) - 1)$, where C_j are the soil moisture classes. The standard errors are indicated by the black error bars.

For the interpretation of the results, the climatology of mean soil water content needs to be taken into account. The SMI of each month refers to different fractions of absolute water saturation in the soil. This seasonality is depicted in Fig. 4 in *Samaniego et al.* (2013) for different locations in Germany. In general, the optimal water content for plant development is defined by 60 % to 80 % of the available field capacity, whilst less than 40 % field capacity, as for instance in the year 2003, is associated with depression in crop yield (*Chmielewski*, 2011). In May and June, dry anomalies represent soil moisture fractions above critical water content because the soil has been replenished with water in preceding winter and spring. For silage maize, however, rather dry conditions are preferable during this time because high soil moisture saturation can induce luxury consumption and thus reduced root depths (*FAO Water*, 2016). This is particularly relevant for maize due to its capability to develop deep roots (*FAO Water*, 2016). This feature allows the plants to access deep soil water under dry conditions during the sensitive phase of flowering and grain filling. Empirical studies indicated that early wet conditions slow down the spreading of seeds and young plants can be damaged through indirect effects, such as fungus (*Urban et al.*, 2015a). A detailed analysis indicates that the large effect of severely wet conditions in June can be partly associated to the 2013 flood in Germany (not shown), which exhibited wet soils in large parts of the country. Starting in July, the level of soil water content decreases (see Fig. 4 in *Samaniego et al.*, 2013). As a consequence, dry anomalies represent damaging conditions because plant available soil water starts to be too low to provide enough water during the most susceptible phase. These effects are increasing over the subsequent months because of the seasonality, the particular growing stage, and the persistence of soil moisture. Lower levels in absolute soil water also explain why wet anomalies have a positive impact in August, but not in July. July exhibits the highest evapotranspiration among all months. This leads to a highly dynamic soil moisture in July which is characterized by a transition from a wet regime to a dry regime. Thus, small deviations from average soil moisture in this month have no significant effect on yield (Fig. 12.5). These are only observed for the very extreme conditions.

Additionally, the growing stage modifies the impact of soil moisture coefficients. In our sample, flowering commonly begins between mid- and end-July and milk ripening occurs in the second half of August (based on own calculation from data provided by *Deutscher Wetterdienst*, 2017). Plants exhibit an increased susceptibility to insufficient water supply during these development stages. As shown in section 4.3, July has the highest partial effect with respect to meteorological variables. In August, soil moisture anomalies show a significantly higher impact on annual silage maize yield than in July. Due its seasonality, absolute soil moisture values are in general lower in August than in July. Further, soil moisture in August integrates temperature and precipitation effects of the preceding months. Thus, dry soil moisture anomalies show harmful effects, while wet ones are beneficial. In September and October, soil moisture usually starts to refill (see Fig. 4 in *Samaniego et al.*, 2013). Maize is in the less susceptible phase to dryness of ripening in September and harvesting usually starts in the second half of this month (*Deutscher Wetterdienst*, 2017). This implies, that severe drought anomalies in September and October might be associated with extended periods of water stress over the sensitive growing stages in the months before.

In this section, it was shown that the seasonality of soil moisture underlying the soil moisture index needs to be considered to disentangled its temporal effects on silage maize yield. Thus, it is necessary to consider seasonality in soil moisture content and silage maize growth when assessing effects caused by soil moisture anomalies.

12.6 Conclusions

In this study, the intra-seasonal effects of soil moisture on silage maize yield in Germany are investigated. It is also evaluated how approaches considering soil moisture perform compared to meteorology-only ones. A demeaned reduced form panel approach is applied, which employs polynomials of degree three for variables of average temperature, potential evapotranspiration, precipitation, and a step wise function for soil moisture anomalies to capture nonlinearities. Potential evapotranspiration and average temperature are mutually exclusive. The model selection is based on the Bayesian Information Criterion (BIC) and the adjusted coefficient of determination (R^2).

This study provides a proof of concept, that a) soil moisture improves the capability of models to predict silage maize yield compared to meteorology-only approaches and that b) temporal patterns in the seasonal effects of the explanatory variables matter. It is shown that soil moisture anomalies improve the model fit in all model configurations according to both the BIC and R^2. SMI entails the highest explanatory power in all months but May (most explained by T) and July (most explained by P). This highlights that soil moisture adds different information than meteorological variables. All time invariant variables show seasonal patterns in accordance to each particular growing stage of silage maize. Furthermore, the dynamic patterns of the SMI effects originate from the seasonality in absolute soil moisture. Those results support the supposition that it is necessary to control for intra-seasonal variability in both the index for soil moisture and meteorology to derive valid impact assessments. Also, the comparison of various meteorological effects based on BIC showed that potential evapotranspiration adds no explanatory power compared to average temperature. Further, partial effects of precipitation outweigh those of temperature when controlling for intra-seasonal variability.

The temporal resolution for the meteorological and soil moisture data is months. This might be too low to accurately resolve the stage of plant growth. Future improvements will involve the use of daily data from high resolution remote sensing campaigns which would allow to determine growing seasons more accurately.

Our results have further implications for climate change impact assessment. First, it is shown that soil moisture can improve agricultural damage assessment and enrich the climate adaptation discourse in this realm, which is mostly based on temperature measures as major explanatory variable (*Carleton and Hsiang*, 2016). We recommend to control for at least those seasonal dependent pathways that affect plant growth presented in our study. Measures of soil moisture should be considered to derive evidence about climate impacts and adaptation possibilities. This particularly concerns climate econometrics, where frequently used reduced form approaches and dose-response functions should also control for soil moisture. For example, *Butler and Huybers* (2013) derived from a dose-response function only relying on temperature measures that the sensitivity to extreme degree days is lower in southern rather than northern U.S. counties. Based on these estimates they concluded that the south is better adapted to hot condition compared to the north. Transferring those adaptation potential to future impacts diminishes the estimated losses. However, various issues need to be considered when employing such an approach, such as the costs of adaptation and wrong institutional incentives (*Annan and Schlenker*, 2015; *Schlenker et al.*, 2013). Also, *Schlenker et al.* (2013) argued that higher average humidity levels in the south diminish the correlation between heat and measures based on evapotranspirative demand. Accordingly, it is recommended to directly control for evapotranspirative demand by vapour pressure deficit (VPD). As shown in section 4.1, no superior effect of potential evapotranspiration over temperature was found when controlling for either precipitation or both precipitation and SMI. Potential evapotranspiration and VPD both account for the water demand of the atmosphere. Instead, the results of this study show that controlling for water supply by measures of either soil moisture and precipitation avoids biased effects in a humid climate. This study further indicates, that it is necessary to account for the seasonal dynamics in both the meteorological and soil moisture effects that constitute the variation in crop yield to employ spatial adaptation as surrogate for future adaptation.

Second, the definition of an index as anomaly has general implications for climate econometrics. Such an index is less prone to systematic errors (Lobell2013, Gornott2015, Gornott2016), because any bias associated to the spatial processing and the meteorological or climatological modeling is minimized (*Auffhammer et al.*, 2013; *Conradt et al.*, 2016; *Lobell*, 2013). Also, the persistence in soil moisture and the resulting smoother distribution in comparison to the meteorological variables might deliver more reliable estimates than climate assessment based on meteorological variables because climate simulations only show robust trends at coarse temporal resolutions (*Gornott and Wechsung*, 2015). An index can also be interpreted as inter-annual variability beyond the demeaning framework. Any linear model employing a categorical variable for each spatial unit is equivalent to joint demeaning of both the dependent and the independent variables and thus the source of variation is the deviation from the mean. For instance, anomalies are used within the adaptation discourse to derive implications for short-term measures

(*Moore and Lobell*, 2014). Again, in such a setting soil moisture can serve as more comprehensive measure than the commonly used temperature.

Finally, this study has also several implications for the design of adaptation measures on weather realizations to reduce current welfare losses of climate events (*Kunreuther et al.*, 2009; *UNISDR*, 2015). First, indexes derived from soil moisture can be used in risk transfer mechanism. For instance, insurance schemes based on a particular weather indexes can be enhanced in both developed and developing countries (*Agriculture Risk Management Team*, 2011). Second, the detrimental effects of wet soil moisture anomalies might allow to extent the risk portfolio of multi-peril crop insurance and thus foster the advancement and implementation of those schemes in Germany (*Keller*, 2010). Third, the installation of agricultural infrastructure should be investigated because negative effects of soil moisture anomalies can be mitigated by irrigation and drainage. In 2010, only 2,34 % of the agricultural area used for silage maize is irrigated (own calculation from data provided by *Statisitisches Bundesamt* (2011)) and the latest numbers about drainage systems in Germany date back to 1993 (*ICID*, 2015).

Overall, an index of soil moisture considering intra-seasonal variability has relevant implications for current and future damage assessment and adaptation evaluation, which are supposed to gain importance in the course of climate change.

Acknowledgements

We kindly acknowledge the German Meteorological Service (DWD), the Joint Research Center of the European Commission, the European Environmental Agency, the Federal Institute for Geosciences and Natural Resources (BGR), the Federal Agency for Cartography and Geodesy (BKG), the European Water Archiv, the Global Runoff Data Centre at the German Federal Institute of Hydrology (BfG), and the Federal Statistical Office of Germany for the provision of data. We especially thank Matthias Zink (UFZ) for processing and providing the data. We also regard thanks to the authors of the R-packages used in this study (*Arnold* (2016); *Bivand and Lewin-Koh* (2017); *Bivand et al.* (2013, 2016); *Croissant and Millo* (2008); *Hijmans* (2016); *Hlavac* (2015); *Neuwirth* (2014); *Pierce* (2015); *Sarkar* (2008); *Sarkar and Andrews* (2016); *Wickham* (2007, 2011, 2016)). We express our thanks to Prof. Reimund Schwarze for his comments and the promotion of the project at the Helmholtz Alliance Climate Initiative REKLIM. This work is also part of the Integrated Project Water Scarcity at the UFZ - Helmholtz Centre for Environmental Research, Leipzig, Germany, which served as forum to present our work. Special thanks to Dr. Andreas Marx, head of the Climate Office for Central Germany, who supported us in the final steps of this study.

CHAPTER 13

ANTHROPOGENIC WARMING EXACERBATES EUROPEAN SOIL MOISTURE DROUGHTS

L. Samaniego, S. Thober, [1] et al.

[1] Helmholtz-Centre for Environmental Research-UFZ

Originally published at:
Samaniego, L., S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. Wood, and A. Marx (2018), Anthropogenic warming exacerbates European soil moisture droughts, Nat. Clim. Change, 4

13.1 Abstract

Anthropogenic warming is anticipated to increase soil moisture drought in the future. However, projections are accompanied by large uncertainty due to varying estimates of future warming. Here, using an ensemble of hydrological and land surface models, forced with bias-corrected downscaled GCM output, we estimate the impacts of 1–3 K global mean temperature increases on soil moisture droughts in Europe. Compared to the 1.5 K Paris target, an increase of 3 K – which represents current projected temperature change – is found to increase drought area by 40 % (± 24 %), affecting up to 42 % (± 22 %) more of the population. Furthermore, an event similar to the 2003 drought is shown to become twice as frequent; thus, due to their increased occurrence, events of this magnitude will no longer be classified as extreme. In the absence of effective mitigation, Europe will therefore face unprecedented increases in soil moisture drought, presenting new challenges for adaptation across the continent.

13.2 Introduction

Global warming is projected to increase evaporation and to reduce soil moisture where it is present, at several hotspot locations around the globe (*Dai et al.*, 2004; *Greve et al.*, 2017). Current research indicates that, although climate change may not create droughts, it may exacerbate them (*Berg et al.*, 2017; *Dai*, 2013; *Hirschi et al.*, 2010; *Huang et al.*, 2015; *Seneviratne et al.*, 2013; *Trenberth et al.*, 2014). Consequently, droughts may set in more quickly, be more intense and last longer (*Cook et al.*, 2015). The recent Paris climate change agreement focuses on holding the global temperature increase to well below 2 K or even 1.5 K above pre-industrial levels (*UNFCC*, 2015). It is worth noting that future global temperatures will likely exceed 2 K above pre-industrial levels by 2100 (*Raftery et al.*, 2017). Limiting global warming to these levels has unknown effects on the characteristics of soil moisture droughts (e.g., drought area and duration) because these characteristics have been quantified for different future periods using emission scenarios that cover a wide range of temperature projections (*Collins et al.*, 2012;

Cook et al., 2015; *James et al.*, 2017; *Prudhomme et al.*, 2014; *Wanders et al.*, 2015). Moreover, the definition of a drought under a non-stationary climate must be carefully chosen such that drought events represent dry anomalies with respect to reference conditions (*Wilhite*, 2000). The agricultural adaptation potential has been estimated for Europe, taking into account crop yield and profit per hectare (*Moore and Lobell*, 2014). Here, we quantify the extent and duration of future droughts and changes in aridity for different warming levels with and without adaptation (see Appendic Methods). We aim to provide information on the benefits of limiting global warming to 1.5 K relative to 3 K in terms of agricultural droughts, which have substantial impacts on vegetation stress, crop losses, the risk of forest fires, tourism (*van Lanen et al.*, 2016), ecosystem services and greenhouse gas emissions (*Ciais et al.*, 2005).

The uncertainty in climate projections and hydrological model parameterisations introduces considerable variability into the resulting projections of the characteristics of soil moisture drought (*Samaniego et al.*, 2013, 2016), thus highlighting the need for multi-model ensembles to enable comprehensive assessments of these events. However, studies of soil moisture droughts at continental and global scales are limited to a few ensemble members and/or employ a single hydrological model (*Lehner et al.*, 2017). Existing multi-model analyses of future droughts focus primarily on hydrological droughts (*Prudhomme et al.*, 2014; *Samaniego et al.*, 2016).

13.3 Methods

To address these shortcomings, we establish a modelling chain using multiple models to generate an unprecedentedly large (60-member) ensemble of high-resolution $5{\times}5\,\text{km}^2$ hydrological simulations that cover the European domain (see Methods). We use two hydrological models (HMs) and two land surface models (LSMs) that employ a consistent set of land-surface properties. The two hydrologic models use a temperature-based PET scheme, which has been criticised within the application of drought analysis using the Palmer Drought Severity Index (PDSI) (*Sheffield et al.*, 2013; *Trenberth et al.*, 2014). The soil moisture index (SMI) derived from these HMs, however, do not show the same deficiency as the PDSI because of methodological differences on how these indices are estimated (see Methods). All HMs/LSMs are driven by downscaled forcings obtained from five bias-corrected Coupled Model Intercomparison Project Phase 5 (CMIP5) projections (*Warszawski et al.*, 2014) that follow three representative concentration pathways (RCPs; RCP2.6, RCP6.0, and RCP8.5). To guarantee the comparability across the multi-model ensemble, all HMs and LSMs estimate soil moisture up to a depth of 2 m and the estimated soil moisture values are transformed into a monthly soil moisture index (SMI) (*Samaniego et al.*, 2013). These high resolution SMI fields are required to perform a spatio-temporal drought cluster analysis (*Samaniego et al.*, 2013) which enables to quantify the area-duration characteristics of every soil moisture drought event. Based on this cluster analysis, two key drought characteristics, the area under drought and the drought duration, are estimated for all drought events simulated by each general circulation model (GCM) and HM/LSM model combination (see Methods). These two characteristics are then analysed for the largest drought within each GCM-HM/LSM combination over specific 30-year periods that correspond to different warming levels under the three RCPs (*Samaniego*, 2017). A time sampling approach is used to extract future 30-year periods that correspond to global warming levels of 1.0, 1.5, 2.0, 2.5, and 3 K with respect to pre-industrial levels for each of the GCM/RCP projections (*James et al.*, 2017) (see Methods). The period from 1971 to 2000 is selected to represent present-day conditions.

13.4 Results

Based on our multi-model ensemble analysis, Figure 13.1a shows that the ensemble median of the largest drought areas increases from 18.7 % of the European territory under a warming of 1.5 K to 26.2 % under a warming of 3 K. The drought threshold from the reference period 1971-2000 is used to enable comparison with historic events; that is, adaptation to climate change is not considered. If adaptation is not considered, then only the top 9.9 % of simulated drought areas under a warming of 1.5 K exceed the ensemble median under a global warming of 3 K. Note that the percentage of ensemble members that exceed the median of the 3 K ensemble increases non-linearly with the degree of global warming. For example, this quantity increases by 13.3 % (2.5 % to 15.8 %) as the amount of global warming increases from 1 K to 2 K; however, it increases by 34.2 % as the amount of global warming increases from 2 K to 3 K.

Drought duration (Figure 13.1c) also exhibits substantial changes across the different warming levels. The median duration of exceptional drought events shows approximately a two- to three-fold increase between the

1.5 and 3 K warming levels (i.e., it increases from 20 months under a warming of 1.5 K to approximately 55 months under a warming of 3 K). Given these changes in the distributions of the areas and the durations of extreme drought events, these future events may no longer represent droughts, which are defined as deviations from normal conditions. This analysis indicates that, for amounts of global warming equal to or greater than 1.5 K, the normal conditions that are used to define typical drought characteristics must be reassessed.

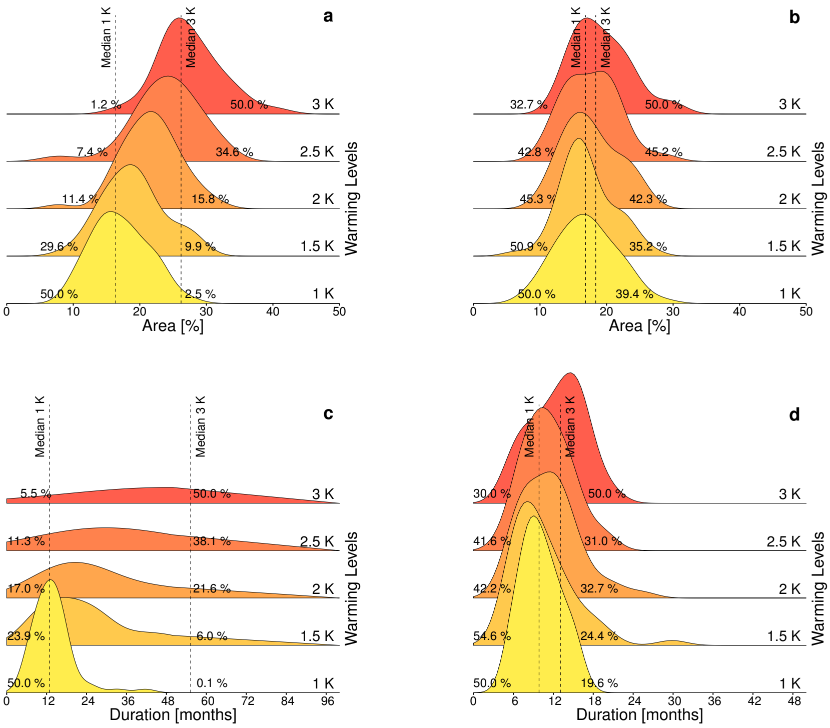

Figure 13.1 Distribution functions are displayed for both the drought areas (a,b) and durations (c,d) of the largest drought events over the 30-year periods corresponding to each global warming level. The results without adaptation are presented in the panels (a,c) and with adaptation in panels (b,d). The vertical dashed lines indicate the median values for global warming amounts of 1 K and 3 K. The fractions of ensemble members located towards the tails are also denoted as percentages. The x-axis limits are different for the duration with and without adaptation (c,d) for clarity.

The impact of climate change on drought characteristics is strongly diminished after adaptation (meaning that the drought threshold is re-calculated based on the projected soil moisture under different levels of global warming as indicated in the Methods section) to drought events is considered. Overall, the ensemble median drought area is estimated to be between 16 % and 18 % of the European territory, and the duration is approximately 9 to 12 months for all of the considered warming levels. A significant difference is only found between the warming levels of 3 K and at most 1.5 K (applying a Kolmogorov-Smirnov test with a significance level of 5 %, Figure 13.1d). Ideally, it is expected that drought area and duration remain unchanged if the soil moisture drought threshold is estimated for each warming level separately (representing adaptation to climate change). Small deviations may still occur because of the intrinsic uncertainty of the processes describing the soil moisture dynamics. It is worth noting that these increases are also obtained using other SMI drought thresholds (see Methods, compare Figure 13.1 and Figure 13-A.1.)

The substantial increases in drought area and duration without adaptation (Figure 13.1a,c) are not evenly distributed across the European domain. Figure 13.2 depicts strong spatial differences in the drought area and duration over six major environmental regions in Europe (i.e., the Alpine North, Atlantic, Boreal, Continental, Mediterranean, and Alpine South regions; see Figure 13.3a) (*Kovats et al.*, 2011; *Metzger et al.*, 2005). The exact values are provided in Table 13.1. The largest increases in the drought area and duration are projected to occur in the Mediterranean. Compared with the estimates for the historical period (1971–2000), the drought area will change from 28 % on average to 49 % under a warming of 3 K (Figure 13.2a,f). The increase in drought area is less than 10 % in the Atlantic, Continental, Alpine North and Alpine South regions. Increased precipitation will decrease the drought area in the Boreal region by about 3 % under a global warming of 3 K. Interestingly, the Alpine North region shows the highest percentage in drought area among all regions for the historic period 1971–2000 (Figure 13.2a), which highlights that droughts have a relatively higher spatial dependence in this region than in the other ones.

With the exception of the Alpine North and Boreal regions, the durations of the largest drought events are three to four times higher under a warming of 3 K compared to historical values (Table 13.1). The increases in drought duration are non-linearly related to climate change because they double (at most) under a global warming of 2 K. The longest droughts, which have durations exceeding 10 years (120 months), are projected to occur in the Mediterranean, Alpine South and Continental regions under a global warming of 3 K. Overall, our results show an alteration of the hydrologic regimes in the Mediterranean and Continental regions when a warming level of 3 K is approached.

Table 13.1 Multi-model ensemble median results for the area under drought ([% of total area]), drought duration [months], and months under drought conditions per year for different levels of global warming and stratified for the IPCC regions. The period of 1971–2000 is used as a reference.

Warming level	Atlantic	Continental	Boreal	Mediter.	Alpine North	Alpine South
			Drought area			
Reference	21.9	34.7	19.4	28.2	41.3	28.9
1.0 K	24.0	36.8	25.2	29.8	31.8	28.7
1.5 K	23.5	35.1	24.7	34.1	34.5	28.7
2.0 K	22.8	35.8	23.4	38.4	34.8	29.4
2.5 K	26.5	36.1	23.0	41.0	35.9	34.4
3.0 K	27.8	39.9	16.4	49.1	41.1	37.1
			Drought duration			
Reference	31.5	32.5	25.0	28.0	12.0	37.0
1.0 K	32.0	38.5	25.0	41.0	22.0	40.0
1.5 K	52.5	60.0	25.0	58.0	20.5	56.0
2.0 K	60.5	65.5	32.5	71.0	21.0	68.5
2.5 K	84.0	86.5	41.5	89.0	18.5	86.5
3.0 K	101.5	121.5	59.5	125.0	17.0	124.5
			Drought months per year			
Reference	2.0	2.0	1.9	2.1	1.9	2.0
1.0 K	2.0	2.1	2.0	2.6	1.7	1.9
1.5 K	2.7	2.6	2.4	3.2	1.9	2.3
2.0 K	3.0	2.8	2.5	3.7	2.0	2.7
2.5 K	3.3	3.1	2.7	4.5	2.2	3.2
3.0 K	3.8	3.9	2.9	5.6	2.4	3.9

The frequency of drought events (expressed in terms of the number of drought months occurring per year) also exhibits marked regional and sub-regional differences, due mainly to the influence of local physiographic and climatic characteristics (Figure 13.2 m-r). During the historical period, the mean drought frequency for all of the grid cells in all of the regions is approximately 2 months per year. This historically low value increases to an unprecedentedly high value under climate change, if no adaptation is considered. For example, the Mediterranean will experience a steady increase in this quantity as the warming level rises, reaching 5.6 months per year under 3 K. Note that some parts of the Iberian Peninsula are projected to experience more than seven drought months per year under the 3 K warming level (Figure 13.2r). These events may no longer be droughts, given that they occur half of the time. All HMs project increases in drought frequency in the Mediterranean, which is a result of the reduced precipitation in this region (see Figure 13-B.2 and 13-B.3). The Continental region shows a change from 1–2 months per year to 3–5 months. Most locations in the Alpine South region will experience a shift in drought frequency from 1–2 months under present-day conditions to 4 months per year under a warming of 3 K.

The previous two figures highlight the need for constant adaptation to the changing climate and indicate that historic drought thresholds may not apply in the future. Adaptation of society to the new normal is known to be associated with substantial costs (*Rötter et al.*, 2011). However, the crucial question for society as a whole and water planners in particular is what the new drought conditions that will occur under different warming levels imply for adaptation policies. To answer this fundamental question, the change in the drought threshold is estimated in a 2-m deep soil column in litres per square metre (i.e., in millimetres of soil water storage). This value is an indicator of the available soil water content under drought conditions and quantifies the change in aridity.

The resulting ensemble average change in the available soil water content is estimated over the six environmental regions for the different warming levels and seasons (i.e., winter, spring, summer, and autumn), including their variability and statistical significance. The magnitude of this change generally increases with increased global warming and is significant for changes larger than 3 % (Figure 13.3). Two major patterns are observed: 1) the Mediterranean and Atlantic regions experience decreases in soil water content in all seasons and under all warming levels; 2) the Alpine North, Alpine South, Boreal and Continental regions tend to become wetter in winter and spring and drier in summer and autumn.

The Mediterranean region is the most affected in all seasons (Figure 13.3e), with the largest increase in aridity appearing in the winter and spring under all warming levels. At the 3 K warming level, the available soil water decreases by 35 mm (±24 mm), which corresponds to a shortage of 35 000 $m^3 km^{-2}$. The Atlantic region exhibits

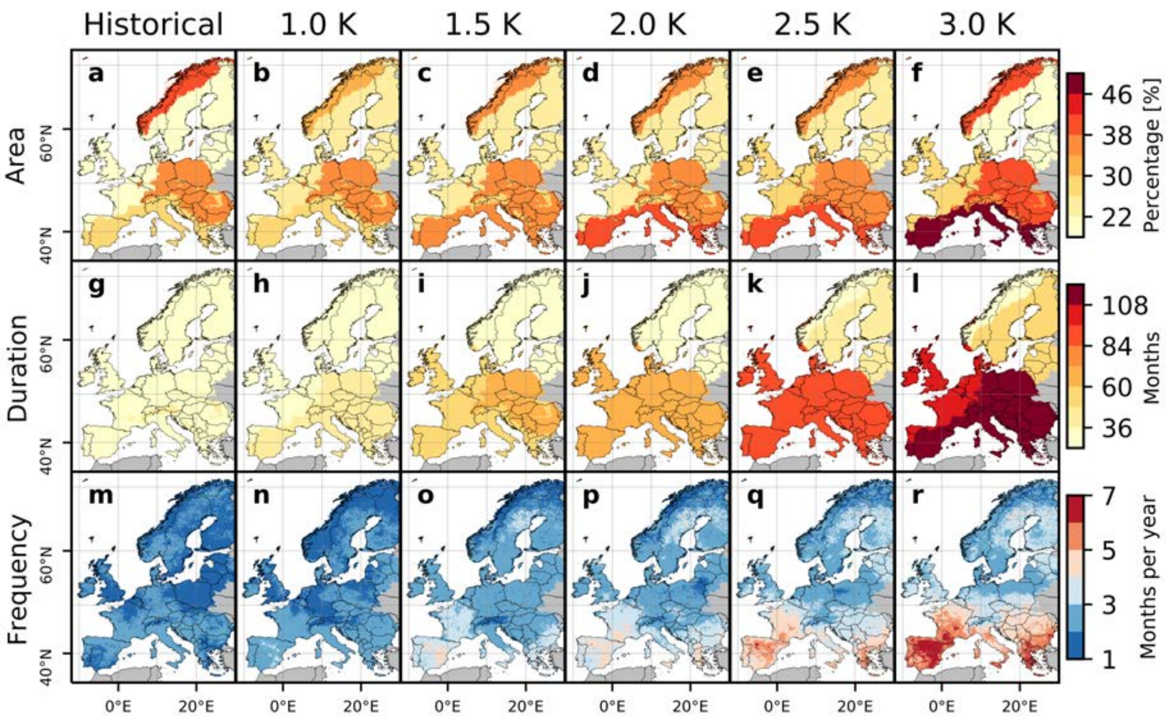

Figure 13.2 The area under drought is evaluated for the six IPCC AR5 regions (*Kovats et al.*, 2011) and quantified as a percentage of the total area of each region (a–f). The drought duration is shown for the same regions (g–l). The area under drought and the drought duration are calculated for the multi-model median of the largest drought events. The frequency of drought months is depicted at the individual grid cell level, which is calculated based on the multi-model median estimates (m–r). All of the results are calculated assuming no adaptation to climate change.

the smallest changes in the available soil water among all of the regions and for all of the warming levels (Figure 13.3b). The Continental region exhibits positive changes during the winter for warming amounts of up to 2 K (Figure 13.3g). In contrast, negative changes are observed for all of the warming levels above 1.5 K during the spring, summer and autumn. Earlier onsets of snowmelt cause increases in the available soil water in the winter and spring for all of the warming levels in the Alpine North and Boreal regions (Figure 13.3c and d). These earlier onsets also lead to increases in aridity in these regions of up to 20 mm in summer, when snowmelt is no longer a source of water.

13.5 Discussion and conclusions

Global warming leads to significant intensification of European droughts, which confirms previous work (*Trenberth et al.*, 2014). We show that climate change has diverse regional and seasonal impacts on soil water availability across Europe. An increase in surface water availability has been reported for different warming levels for the Alpine and Boreal regions (*Greve et al.*, 2017). However, this increase is unevenly distributed over the year. Moreover, soil water availability appears to decrease significantly throughout Europe during the growing season (i.e., summer and fall). Economic assessments of climate change adaptation for the agricultural sector are often based on temperature-related characteristic curves (*Moore and Lobell*, 2014). These analyses could benefit from incorporating soil moisture because it constitutes the primary source of water for plant growth.

The exacerbation of drought conditions in the Mediterranean under global warming of 1.5 K and 2 K will be unprecedented since the last millennium (*Lehner et al.*, 2017). If a global warming of 3 K is reached, southern Spain and probably Italy and Greece will turn “into a desert” (*Guiot and Cramer*, 2016). This unprecedented

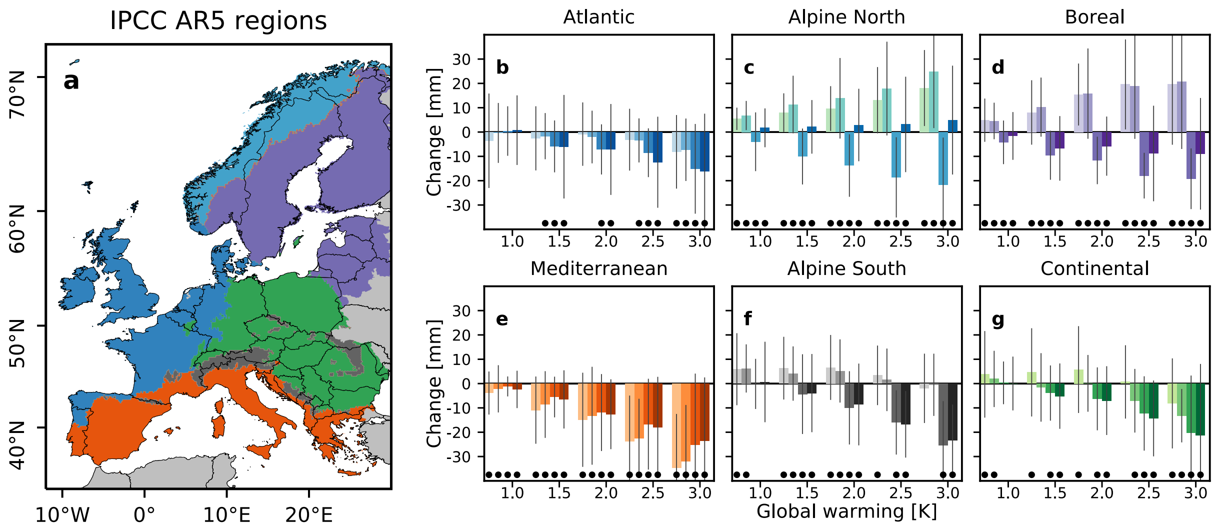

Figure 13.3 Changes in the soil water availability (increases in aridity) during drought events between a given warming level and the reference period, considering adaptation to climate change. The results are aggregated to the IPCC AR5 regions (*Kovats et al.*, 2011) for the different seasons (from left to right, DJF, MAM, JJA, and SON) and from for each warming level. The whiskers indicate the inter-quartile range of the multi-model ensemble results. The markers at the bottom of the plots indicate changes that differ significantly from zero, as determined using a Wilcoxon rank-sum test and a significance level of 5 %.

change will also have severe impacts on Mediterranean vegetation and biodiversity, and, thus on ecosystems and their services.

The strong reductions in soil water availability during dry periods are mostly related to decreases in precipitation and increases in evapotranspiration (*Greve et al.*, 2017) (see Figures 13-B.3 and 13-B.1). The relatively high decreases in soil water availability noted in this region are related to the relatively high increases in the maximum daytime temperatures compared to other regions (*Seneviratne et al.*, 2016). Whether economic adaptation assessments (*Moore and Lobell*, 2014) can properly assess such severe changes remains an open question. Note that, while we estimate soil moisture for a 2 m deep soil column, many plants, particularly crops, do not have roots that extend to that depth. Consequently, we likely underestimate the effects of soil moisture droughts in the top-soil layers because these layers tend to dry faster than the lower soil layers (*Berg et al.*, 2017).

We relate our results to the 2003 drought event (estimated based on historical observations, see Methods) to illustrate the severity of the projected changes. In water-limited regimes, agricultural droughts are intrinsically related to significant reductions of evapotranspiration and gross primary production (GPP), as well as the occurrence of heat waves. For example, Europe emitted an amount of carbon dioxide that corresponds to the amount that is normally sequestrated in four years during the 2003 drought event (*Ciais et al.*, 2005). In the future, drought events that are similar in magnitude and extent to that of 2003 will be twice as frequent. In detail, our results indicate that the increase in frequency, which is defined as the ratio of SMI under a warming of 3 K with respect to that of the reference period, is approximately 2.0 ($\pm$0.33). The estimated average soil water availability deficit during the 2003 drought event was 27.6 mm. The change in the drought threshold at a warming level of 3 K (Figure 13.3) is of the same order of magnitude as the average deficit during the 2003 event in most of the regions. This result implies that much of this event will not be classified as a drought in the future, and the projected droughts will be associated with substantially less available soil water than the 2003 event.

We estimate that 42 % ($\pm$ 22 %) more people will be located within areas enduring extreme droughts under a warming level of 3 K compared to a warming level of 1.5 K (170 million people vs. 120 million people, respectively; Figure 13.4). In contrast, 15 % of the population (83 million people) was located under drought affected areas during the 2003 event. At the peaks of the largest droughts, the population located within areas under drought increases from 336 to 400 million people (Figure 13.4), and these numbers correspond to 61 % and 73 % of the European population, respectively. The increases in population within drought prone areas mostly occur in the Atlantic, Continental and Mediterranean region, because drought area is increasing the most in these regions (Table 13.1). Global warming may constitute a new human health threat (*Robine et al.*, 2008) and extreme droughts, under particular situations, may trigger migration (*Wilbanks et al.*, 2007). For these reasons, further studies should be conducted to investigate the potential effects of future extreme droughts on the European society and potential mitigation strategies aiming at reducing their negative effects.

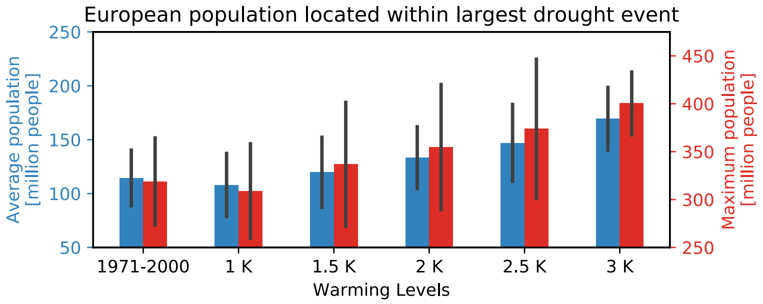

Figure 13.4 Average and maximum European population who are located within the area enduring the largest drought at a given warming level (i.e., experience an alteration in standard living conditions during an event). Population data for 2005 are used for reference, and these data were obtained from the SEDAC data set (http://sedac.ciesin.columbia.edu). Based on this data set, the population of the study area is estimated to be approximately 550 million people. Error bars represent the ensemble standard deviation.

Overall, Europe will face unprecedented increases in the area affected by the largest soil moisture drought and the duration of such droughts if no adaptation is implemented during the coming decades (with respect to the historical period). The magnitudes of these increases depend strongly on the level of global warming. If future global temperatures will exceed 2 K above preindustrial levels (*Raftery et al.*, 2017), our results show that drought areas will be up to 40 % larger under a warming level of 3 K compared to a warming level of 1.5 K. Similarly, the drought duration will increase by three times between these two warming levels. Decreases in aridity are found only in the Alpine and Boreal regions during the winter and spring. Even if adaptation measures are successfully implemented, aridity will increase throughout the continent during the summer from less than 10 mm at a global warming of 1.5 K to approximately 20–35 mm at a global warming of 3 K. Such an increase in aridity is comparable to the deficit during the 2003 drought event. Our study therefore highlights the need to adapt to new normal conditions to minimise the impact of extreme drought events. The European agricultural sector must adapt to summers with reduced soil water, and the risk of land degradation and desertification in sensitive environments exists. Further research is urgently needed to assess the degree of impact of future extreme drought events on the European society as a whole, if increased aridity threatens minimum living conditions (*Wilbanks et al.*, 2007).

Acknowledgements

This study was partially performed under a contract for the Copernicus Climate Change Service (`edge.climate.copernicus.eu`). ECMWF implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. This study has been mainly funded within the scope of the HOKLIM project (www.ufz.de/hoklim) by the German Ministry for Education and Research (grant number 01LS1611A). We would like to thank the editor Graham Simpkins, one anonymous reviewer and Rene Orth for their constructive comments that helped to improve the manuscript. We would also like to thank Peter Greve for providing data included in Figure 13-B.1. We acknowledge the funding from NWO Rubicon 825.15.003. We acknowledge the E-OBS dataset from the EU FP6 project ENSEMBLES (`http://ensembles-eu.metoffice.com`) and the data providers in the ECA&D project (`http://www.ecad.eu`). We would like to thank people from various organisations and projects for kindly providing us the data which were used in this study, which includes ISI-MIP, JRC, NASA, GRDC, BGR & ISRIC. This study was carried out within the Helmholtz-Association climate initiative REKLIM (`www.reklim.de`).

Appendix: (A) Supplementary figures

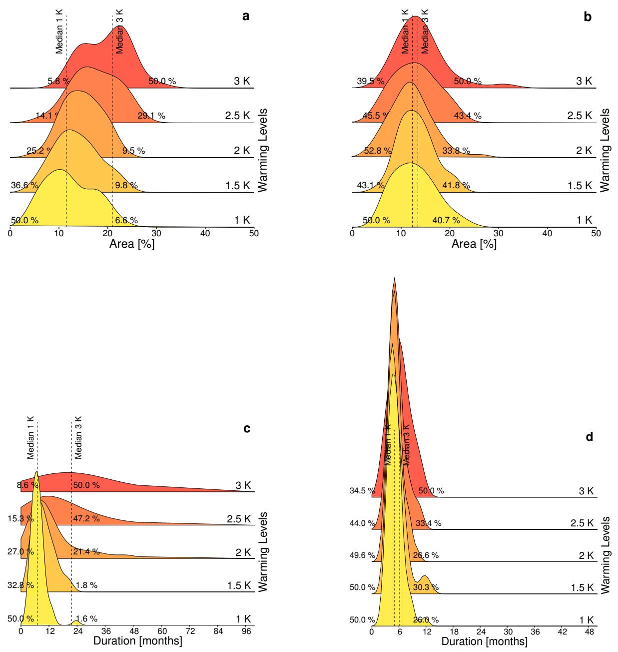

Figure 13-A.1 Same as Figure 1 in the main text, but using a drought threshold τ of 0.1 for the spatio-temporal clustering algorithm.

Appendix: (B) Methods

Modelling chain

Daily temperature and precipitation values for the period 1950 to 2099 obtained from five Coupled Model Intercomparison Project v5 (CMIP5) Global Climate Models (GCMs) (HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, GFDL-ESM2M and NorESM1-M) forced by three RCPs (RCP2.6, RCP6.0, and RCP8.5) are used as input to four hydrologic models (HMs). These GCM data were made available by the ISI-MIP project (*Warszawski et al.*, 2014) and are downscaled to a global resolution 0.5° and bias-corrected using a trend-preserving approach (*Hempel et al.*, 2013). These models cover a range of 0.55 of the uncertainty of the entire CMIP5 ensemble for precipitation and 0.75 for temperature (*McSweeney and Jones*, 2016). The uncertainty range of this 5-member ensemble is comparable to that of a larger CMIP5 model ensemble (Figure 13-B.1) (*Greve et al.*, 2017). The 0.5° data are further disaggregated within the EDgE project (edge.climate.copernicus.eu) to a 5-km grid over Europe using the external drift kriging (EDK) approach. EDK constitutes the best linear unbiased estimator of the selected meteorological variable. This key characteristic of EDK constraints the mean of the interpolated (downscaled) values to not differ from the expectation of the meteorological variable at this location. Thus, EDK does not introduce artefacts (e.g., trends) into the original forcing. Another advantage of this approach is that it introduces orographic effects of precipitation and temperature that are not present in GCMs at the coarse resolution, while maintaining the trend of the original data. The disadvantage of EDK is that it does not guarantee a conservation of mass and energy everywhere. Within the present study, however, the differences between original and downscaled values are in general less than 1 % (at most 5 %) for precipitation and 0.1 K (at most 0.23 K) for temperature. These differences are smaller than the differences between the individual GCMs and the changes induced by climate change.

Two hydrological models (HMs: mHM, PCR-GLOBWB) and two land surface models (LSMs: Noah-MP, VIC) are used to simulate soil moisture up to a depth of 2 m. The same morphologic, land cover, and soil data are used to setup these models; thus, the differences among the model simulations are due solely to differences in the representations of different processes used in the models. The mesoscale hydrological model (mHM; www.ufz.de/mhm) is a process-based hydrologic model that was developed for use at scales ranging from 1 km to 50 km (*Kumar et al.*,

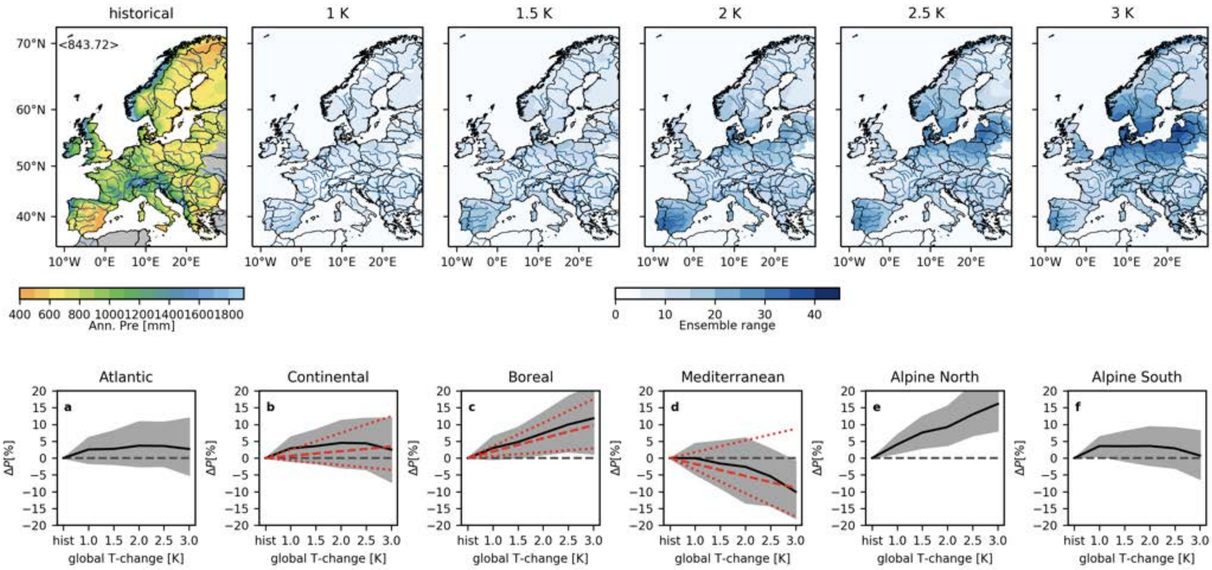

Figure 13-B.1 Uncertainty in precipitation projections. In the top row, the map for the historical period shows the long-term annual precipitation of the 5 GCMs used in this study. For the different warming levels, the map show the ensemble range between the percentage change occurring with a probability of 90% and 10% according to the five GCMs used under three RCPs. These values are then averaged for the different European regions and depicted in the lower row. For the Continental, Boreal, and Mediterranean region, the red dashed line denotes the median change reported by Greve et. al. (2017) and the dotted lines the 10% and 90% percentiles. These values are added for comparison although a different method (i.e., pattern-scaling) has been used to derive them.

2013; *Samaniego et al.*, 2010a). PCR-GLOBWB was developed to represent the terrestrial water cycle, including artificial water management, at global and continental scales, and it places special emphasis on the groundwater component (*Sutanudjaja et al.*, 2018). Noah-MP is the land-surface component of the Weather Research and Forecast model, and it represents both the terrestrial water and energy cycles (*Niu et al.*, 2011). VIC was developed to provide a simplified representation of land-surface hydrological processes that would be suitable for implementation in a GCM (*Liang et al.*, 1994). The model parameters are calibrated using the E-OBS meteorological data (*Haylock et al.*, 2008) for nine distinct catchments located in Spain, the United Kingdom, and Norway. An automatic calibration scheme is employed for mHM and PCR-GLOBWB (*Rakovec et al.*, 2016c). Noah-MP is calibrated manually by adjusting the parameter describing surface evaporation resistance based on previous analyses (*Cuntz et al.*, 2016). The VIC parameters are taken from global simulation runs and are not calibrated using the E-OBS or observed river discharge datasets over the EU domain.

Drought frequencies related to changes of meteorological forcings

Figure 13-B.2 provides a comparison of the number of drought months for the individual hydrologic models, considering no adaptation to climate change for various levels of global warming. All hydrologic models show a similar increase in drought frequency in the Mediterranean region in southern Europe. This may be related to the relatively large decrease in annual precipitation of up to 25 % at a warming level of 3 K (Figure 13-B.3). In central Europe, all models exhibit a smaller increase of drought frequencies in comparison to those in the Mediterranean, which can be expected given the relatively smaller changes in projected precipitation (Figure 13-B.3). Projected temperature is increasing similarly in central Europe and the Mediterranean region, which highlights that the simulated evapotranspiration in this model ensemble is limited by water availability rather than by energy in this region. In contrast, precipitation is projected to increase in the Scandinavian region in northern Europe up to 20 %. In this region, the hydrologic models differ in their projections of drought frequencies. For example, VIC and mHM show increases in this region, PCR-GLOBWB shows a mixed pattern, and drought frequencies simulated by Noah-MP remain unchanged by global warming. Because all models are forced with the same meteorological data, the pa-

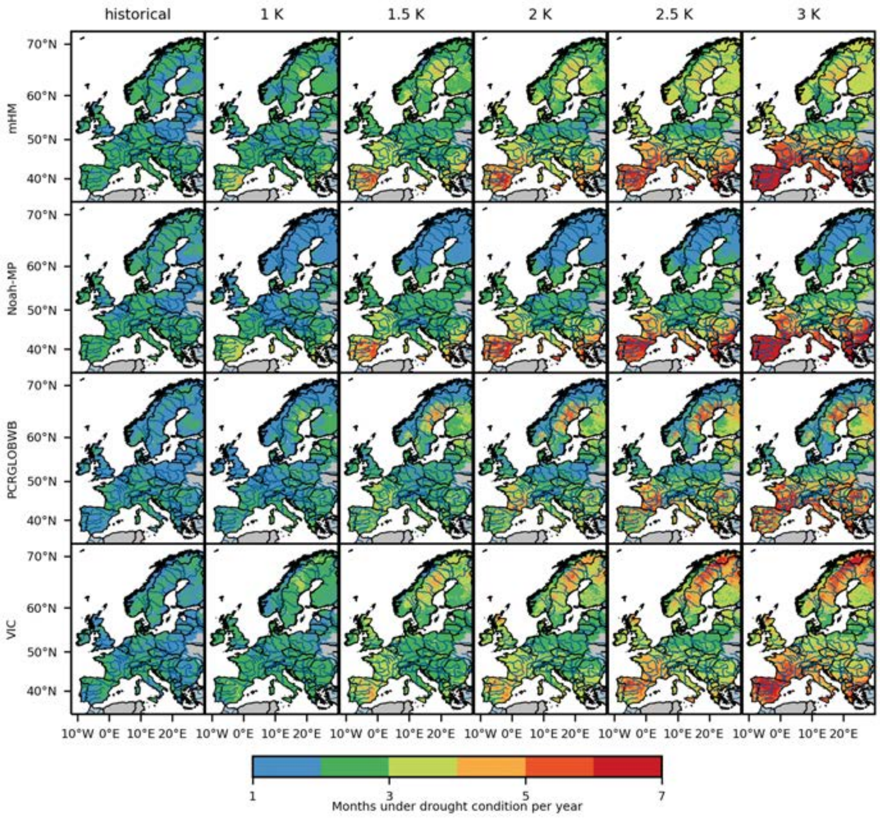

Figure 13-B.2 Drought frequency for every hydrological model for various global warming levels. The number of drought months per year are calculated using the distribution functions of the SMI of the reference period, thus assuming no adaptation to climate change.

rameterization of snow processes in this cold region and the parameterization of ET have a strong impact on soil drought characteristics. For example, mHM allows ET when the surface is covered with snow, which is based on the model assumption that snow cover has a large subgrid variability. On the contrary, Noah-MP explicitly considers snow cover fractions within the calculation of evaporation. These results show that the hydrological models have relative larger differences over various regions. For this reason, we consider it fundamental to use a multi-model ensemble for climate change drought analysis.

Model verification

Streamflow simulations from the four hydrologic models, driven by five GCMs, were compared against observations during the historical 30-year period (1966–1995). Here, we analyse the model skill for reproducing the median daily flows (p50) over 357 gauging stations located across the EU domain (Figure 13-B.4). The gauges have been selected from the Global Runoff Data Centre database. All gauges have complete 30-year period (1966–1995) of daily observations across the modelling domain, which allows for a robust statistical analysis. Additionally, these basins have an error of less than 10 % in the basin delineation and the median basin area is 1680 km^2. Overall, the ensemble model simulations show reasonably high skill in capturing the observed variability of p50, with a correlation coefficient value of 0.92 (Figure 13-B.4e) and the mean relative bias is 35 %. In general, the model combinations (GCM/HM) appear to slightly overestimate the observed p50 values, with mHM being closest to observations compared to the Noah-MP, PCR-GLOBWB and VIC model simulations. The basins in the central EU region and in the Iberian peninsula generally exhibit a positive bias (Figure 13-B.4f). We note that these verifications are quite rigorous as the hydrologic models are forced with GCM simulated datasets, rather than observed

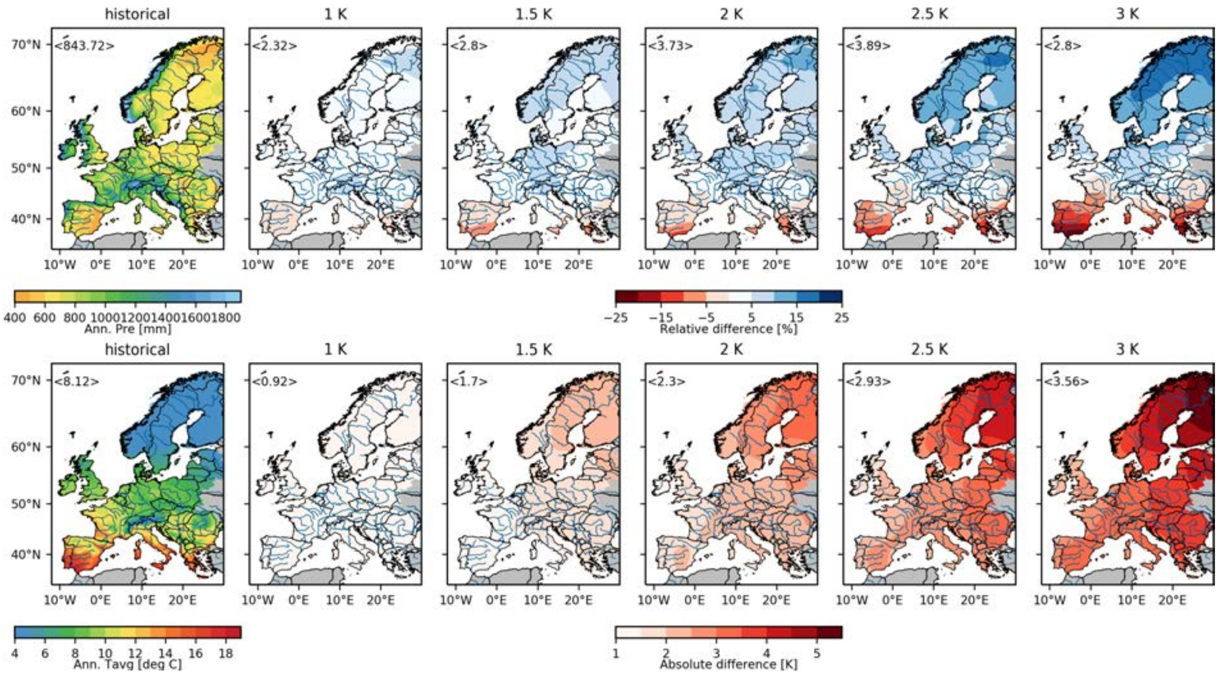

Figure 13-B.3 Projected changes in precipitation and temperatures for different levels of global warming. Upper panels show the historical annual average precipitation for the reference period and the corresponding anomalies for various global warming levels. Lower panels show the same but for annual average temperature.

meteorological datasets. This implies that a comparison of simulated and observed streamflow for specific time points is not feasible because GCM-based simulations do not reproduce observed weather and thus events.

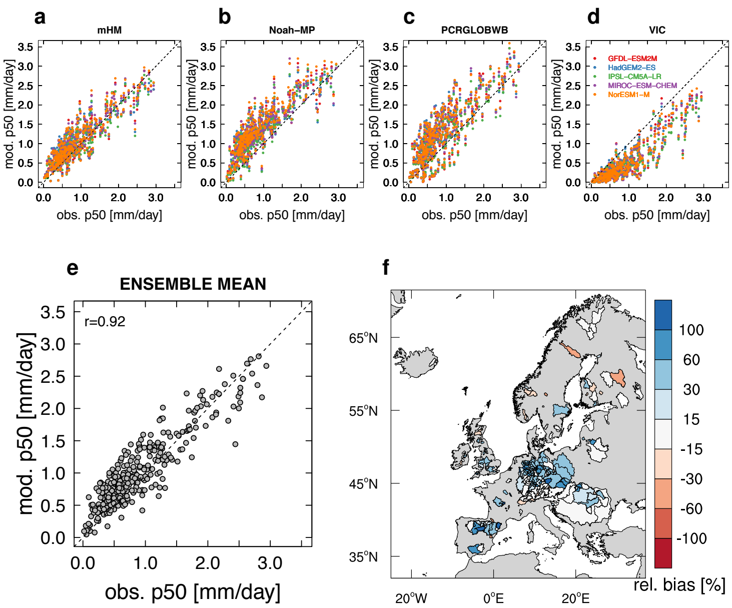

Figure 13-B.4 Verification of simulated streamflow. Scatter plots of the median daily streamflow (p50) between observations and simulations for individual GCM/HM combinations (panels a to d) and the multi-model mean (panel e). Hydrologic model simulations are obtained using the forcings based on five GCMs during the period (1966–1995) over 357 EU river basins. Also shown are geographical location of the river basins with colours indicating relative bias between simulations and observations (panel f).

Estimation of warming levels

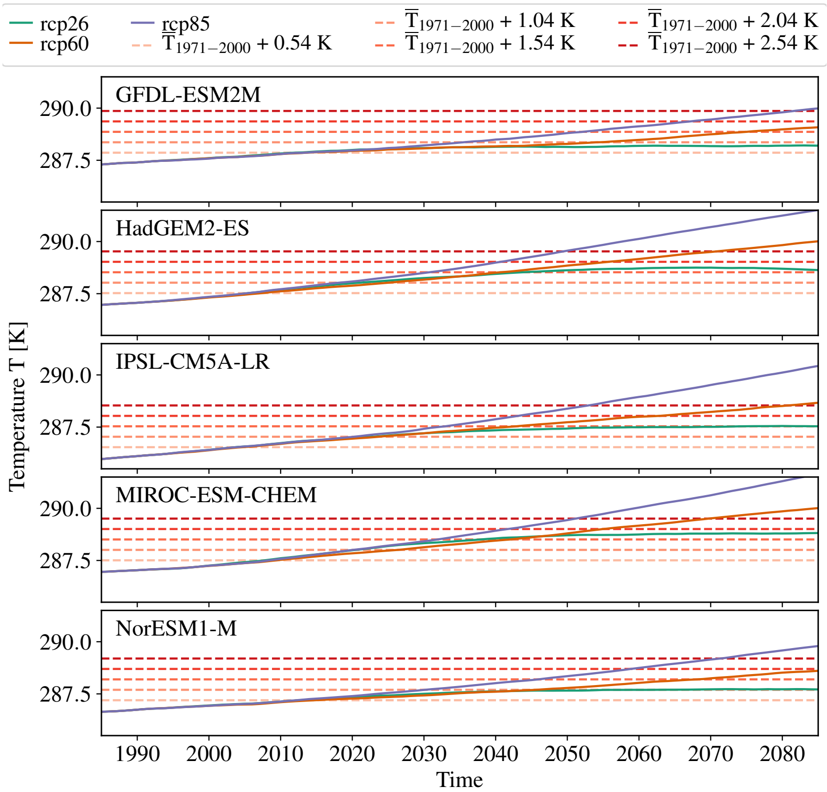

Figure 13-B.5 Projected global mean temperatures. Development of centered 30-year global average temperatures for all five General Circulation Models (GCMs) included in this study. The horizontal lines mark when the global warming of 1, 1.5, 2, 2.5, and 3 K are reached. The coloured lines indicate the different Representative Concentration Pathways (RCPs): RCP2.6, RCP6.0, and RCP8.5.

Within this study, the global warming levels for 1, 1.5, 2, 2.5, and 3 K are identified employing a time sampling approach (*James et al.*, 2017). The 30-year average temperature of 1971–2000 is used as a reference. The pre-industrial warming between the periods 1881–1910 and 1971–2000 is assumed to be 0.46 K (*Vautard et al.*, 2014). This offset is subtracted from the warming levels for determining the 30-year periods for the specific global warming. These periods are identified as follows. For each Global Climate Model (GCM) and representative concentration pathway (RCP), calculate the 30-year global average temperature for all 30-year periods between 1960 and 2099 (prepending the historical data to each RCP). Note down the period when a 30-year global average temperature first reaches or exceeds a given global warming (1, 1.5, 2, 2.5, and 3 K minus 0.46 K offset) (*James et al.*, 2017). The procedure is illustrated in Figure 13-B.5 for all GCMs and RCPs. It is worth mentioning that other periods than 1881–1910 have been suggested to represent pre-industrial conditions, which might lead to offsets that are 0.11 K higher than the one used in this study (*Hawkins et al.*, 2017). We recalculated the periods based on this adjusted threshold and found shifts of 2 to 6 years (not shown). Given the fact that our analysis is using simulated soil moisture of 30 year periods, we expect little influence of the adjusted offset on our results.

In total, 15 GCM realisations reach 1 K, 14 reach 1.5 K, 13 reach 2 K, and 8 reach 2.5 K and 3 K global warming. As four HMs are used in this study, the obtained sample sizes are sufficiently large to quantify extreme soil moisture droughts for each level of global warming.

Soil moisture index and drought characteristics

The soil moisture index (SMI) for a given cell and month is estimated as

$$\mathrm{SMI}_t = \hat{F}_T(x_t), \tag{B.1}$$

and it represents the quantile at the soil moisture fraction value x (normalised against the saturated soil water content). x_t denotes the simulated monthly soil moisture fraction at a time t and $\hat{F}_T$ is the empirical distribution function estimated using the kernel density estimator $\hat{f}_T(x)$ of the corresponding calendar month at time t. $\hat{f}_T(x)$

is estimated as

$$\hat{f}_T(x) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{x - x_k}{h}\right). \tag{B.2}$$

Here, $x_1, \ldots, x_n$ represents the simulated soil moisture fraction of a given calendar month during the reference period T; n denotes the number of calendar months within a given period (i.e., 30 for a 30-year period); and K represents a Gaussian kernel function with a bandwidth h. The bandwidth is estimated by minimising a cross-validation error estimate (*Samaniego et al.*, 2013) for the reference period separately for each calendar month, grid cell, LSM/HM and GCM combination to ensure comparability across time, space and model combinations. A cell at time t is under drought when $\mathrm{SMI}(t) < \tau$. Here, τ denotes that the soil water content in this cell is less than the values occurring $\tau \times 100$ % of the time. In this study, τ is set to 0.2. All drought events are identified using a multi-temporal clustering algorithm (*Samaniego et al.*, 2013). This algorithm first masks all cells at each time step that fulfil SMI $\leq \tau$ and consolidates adjacent cells to a drought event. Second, drought events at consecutive time steps that share a minimum overlapping area are consolidated into a single event. Third, drought statistics (e.g., areal extent, duration) are estimated for all identified drought events. The *mean duration* (D) of a drought event is then defined as the mean of the drought duration estimated over every cell affected by a drought event. This statistic is given in months. The *mean areal extent* (A) is defined as the average of the region under drought from the onset until the end of the drought event, which is then expressed as a percentage of the total surface area of the region. It should be noted that the value of the threshold τ certainly determines A and D. Sensitivity analysis, however, shows that the rate of increase of these characteristics between two warming levels is invariant of the value of τ (compare Figure 1 and Figure S1). The reference period T within the estimation of the SMI $\hat{F}_T$ is chosen in two ways to quantify the effect of adaptation to climate change: 1) T is chosen as 1971–2000 to calculate the drought area and duration for all warming levels, which represents no adaptation to climate change, 2) T is identical to the period when a global warming level has been reached, which represents adaptation to climate change (*Wanders et al.*, 2015). In the latter case, it depends on the amount of global warming, the GCM and the RCP considered.

Estimation of available soil water (aridity)

The changes in the water soil storage (aridity) that occur at the different warming levels is estimated by varying the reference period from T_0 to T_Δ, where T_0 denotes the historical reference period (1971–2000), and T_Δ denotes the period until a particular value of ΔK is reached in a given RCP and GCM combination. Based on these two periods, the change in aridity within a region (as represented by the average over all of the cells within the region) for a given RCP-GCM-HM combination is estimated as

$$\delta x_\Delta = \langle \overline{\hat{F}^{-1}_{T_\Delta}(\tau)} \rangle - \langle \overline{\hat{F}^{-1}_{T_0}(\tau)} \rangle. \tag{B.3}$$

The operator $\langle \cdot \rangle$ denotes the ensemble mean, and the overline indicates the spatial average. Finally, the seasonal averages are estimated from the values obtained for each month. This index is depicted in Figure 3. Note that the threshold τ is kept constant (e.g., 0.2) for T_0 and T_Δ. The absolute soil moisture thresholds (e.g, $\hat{F}^{-1}_{T_\Delta}(\tau)$), on the other hand, depend on the period.

Estimation of soil water deficit for the 2003 event

For a given drought event occurring in a period T, the soil water deficit in a given grid cell is estimated by

$$d_i^T(t) = \left[\hat{F}^{-1}_{T,i}(\tau) - x_i(t)\right]_+. \tag{B.4}$$

The average deficit estimated over the lifespan of a drought event occurring in a period T is given as

$$d^{\,T} = \overline{\frac{1}{n_T} \sum_{t \in T} d_i^T(t)}. \tag{B.5}$$

Here, n_T denotes the number of months under drought in the period T and the overline indicates the spatial average. The operator $\left[\,\cdot\,\right]_+$ denotes the positive part function. The soil water deficit for the 2003 event is estimated as indicated above with every hydrological model forced with the E-OBS (*Haylock et al.*, 2008) meteorological data (1950–2015). The period T corresponds to 1960–2002. The ensemble average is afterwards estimated to be 27.6 mm.

Comparison of SMI and PDSI

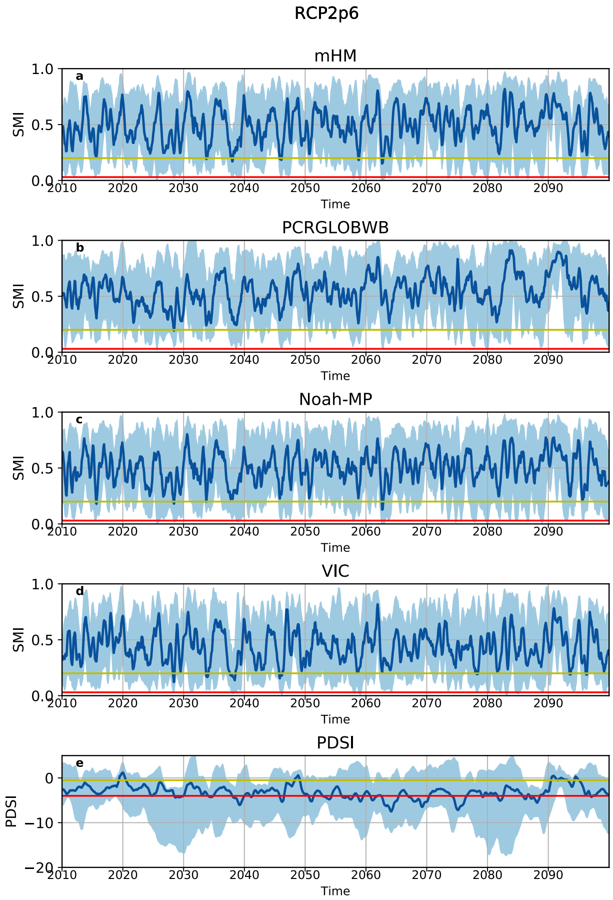

Figure 13-B.6 Comparison between SMI and PDSI. Panels a) to d): Soil moisture index (SMI) of the four hydrologic models used in this study for one location in Eastern Germany (Saxony). Each panel contains five realizations under RCP 2.6 (one for each considered General Circulation Model, GCM). For clarity, only the median (solid blue line) and the range from minimum to maximum are shown. Panel e): Same as a) to d), but for PDSI instead of a hydrologic model. For both indices SMI and PDSI, red and yellow lines depict thresholds for drought events having an exceedance probability of 95% and 80%, respectively.

Numerous studies on drought research used the Palmer Drought Severity Index (PDSI) (*Dai et al.*, 2004; *Palmer*, 1965; *Sheffield et al.*, 2013; *Trenberth et al.*, 2014). The PDSI is a water budget accounting index that cumulates soil moisture anomalies derived from monthly precipitation and temperature. Here, we use the self-calibrating version of PDSI (*Wells et al.*, 2004) at the monthly timescale. PDSI requires two input parameters for every grid cell. These are the latitude of the considered location and the available water holding capacity (AWC). The latter is derived using the same soil dataset used for the hydrologic models and the Multiscale Parameter Regionalization (MPR) method used in the mesoscale Hydrologic Model (mHM) (*Samaniego et al.*, 2010a). The calibration period for the PDSI is set to 1971 to 2000, which is consistent with the period for the estimation of the kernel density function of the soil moisture index (SMI). Subsequently, both indices (SMI and PDSI) are evaluated during the period 2010 to 2099. We present results for one location in Eastern Germany (lat: 51.09 °N, lon: 12.89 °E) to discuss the differences between the PDSI and the SMI. However, the same features discussed below were also observed at locations in Southern France, Spain, and England.

The RCP 2.6 scenario results in stationary SMI and PDSI data without any significant trend (Figure 13-B.6). This can be expected because the RCP 2.6 scenario leads to a projected increase in global mean temperature of 0.3–1.7 K until the end of the 21st century. All indices detect relatively more droughts under RCP 6.0 (Figure 13-B.7) and RCP 8.5 (Figure 13-B.8) as compared to RCP 2.6. However, there are substantial differences between the PDSI and SMI. Most importantly, the median PDSI is indicating extreme drought conditions for the last third of the 21st century for both RCP 6.0 and RCP 8.5.

In the latter case, the median PDSI shows a strong negative trend. For the same period, the median SMI is indicating non-drought conditions for the majority of time points. This indicates that the PDSI is extremely sensitive to the projected climate change in this region. It is worth noting that climate change in this region is mostly increasing temperature, whereas annual precipitation is increased by less than 10 % (Figure 13-B.3). It is known that the PDSI method using the temperature-based Thornthwaite potential evapotranspiration scheme is oversensitive to changes in temperature and that the Penman-Monteith method provides a less biased estimate (*Sheffield et al.*, 2013). The hydrologic models mHM and PCR-GLOBWB also use a temperature-based PET formulation (i.e., the Hargreaves-Samani equation (*Hargreaves and Samani*, 1985)), but show a similar behaviour as Noah-MP and VIC (Figures 13-B.6–13-B.8), which do not use a PET approach and calculate the full energy balance at the land surface.

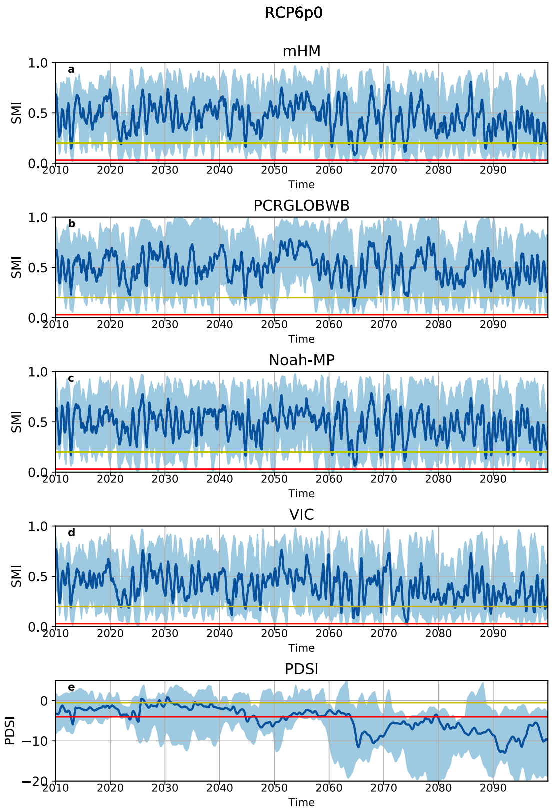

Figure 13-B.7 Comparison between SMI and PDSI. Same as Figure 13-B.6, but for RCP 6.0.

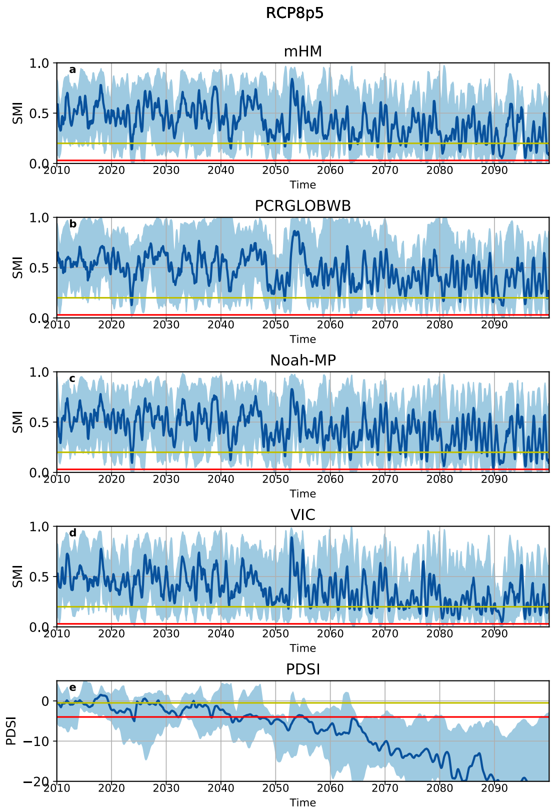

Figure 13-B.8 Comparison between SMI and PDSI. Same as Figure 13-B.6, but for RCP 8.5.

These results highlight that the combination of a temperature-based PET approach with the conceptualisation of the PDSI leads to an overestimation of drought conditions. On the contrary, a drought index derived from hydrologic models (i.e., mHM and PCR-GLOBWB) that use a temperature-based PET scheme, do not exhibit such behaviour. The reason for this difference stems from the way these indices are estimated. PDSI is an autoregressive model of the type

$$X_t = pX_{t-1} + qZ_t \tag{B.6}$$

that estimates the current PDSI value (X_t) based on the previous value of the index and the current soil moisture anomaly Z_t (*Wells et al.*, 2004). Here p and q are the so-called Palmer "duration" factors to be determined empirically for every location. Z_t is determined with a two layer water balance model and several empirically parameters that "allow for accurate comparisons of PDSI values over time and space" (*Wells et al.*, 2004). The autoregressive conceptualisation of PDSI under a non-stationary climate (i.e., increasing temperature, PET, and soil moisture anomalies under RCP6.0, RCP8.5) induces a negative drift from the long-term mean. On the contrast, SMI is by definition bounded between zero and one because it corresponds to the respective quantiles of the simulated soil moisture (see Section above).

Population in drought areas

For each member of the multi-model ensemble, the spatio-temporal evolution of the largest drought event is identified during the reference period T_0 and all of the 30-year periods representing different levels of global warming T_Δ. This information is then overlaid with the population density to estimate the population located in the area under drought at a given point in time. Based on these results, we estimate the average and maximum populations affected over the lifespan of the drought. To identify the effect of future droughts, we use the distribution of the population of Europe in 2005. The UN-adjusted Gridded Population of the World, data set, version 4, was obtained from SEDAC (http://sedac.ciesin.columbia.edu). The year 2005 is selected because it best represents the population distribution during the 2003 event, which is used in this study as a reference. According to this data set, the population of the entire domain is approximately 550 million people. This analysis does not account for demographic changes.

Appendix: (C) Data availability

All information used in this study has been obtained from the following open sources: 1) Terrain elevation EU-DEM and GOTOPO30 from https://lta.cr.usgs.gov/GTOPO30 and http://www.eea.europa.eu/data-and-maps/data/eu-dem; the river database CCM2 v2.1 from http://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23; Soils texture maps SoilGrids1km from http://www.isric.org/content/input-data-soilgrids; the land cover product GlobCOVER v2 from http://due.esrin.esa.int/page_globcover.php; the land cover products CLC00, CLC06, CLC12, CLC90 v18.4 from http://land.copernicus.eu/pan-european/corine-land-cover; the hydrogeology map IHME1500 v11 from http://www.bgr.bund.de/ihme1500; the climate projections CMIP5 from https://www.isimip.org/outputdata/isimip-data-on-the-esgf-server/; the historical forcings E-OBS v12 from www.ecad.eu/E-OBS/; the GRDC streamflow data from http://www.bafg.de/GRDC. Finaly, the EDgE simulations http://edge.climate.copernicus.eu, and the data that support the findings of this study are available from the corresponding author upon request.

PART IV

LESSONS AND OUTLOOK

CHAPTER 14

LESSONS AND OUTLOOK

"Intelligence is the ability to adapt to change."

—Stephen Hawking

14.1 Subject and aims of the thesis

This thesis summarizes a decade of research carried out at the Helmholtz Centre for Environmental Research on the subject of drought monitoring, modeling, and forecasting, from local to continental scales. This topic constitutes one of the grand challenges of contemporary hydro-meteorology due to its complexity and socio-economic impacts.

The overarching objectives of this study, systematically addressed in the twelve previous chapters, are:

Objective 1 Create the capability to seamless monitor and predict water fluxes at various spatial resolutions and temporal scales varying from days to centuries.

Objective 2 Develop and test a modeling chain for monitoring, forecasting and predicting drought events and related characteristics at national and continental scales.

Objective 3 Develop drought indices and impact indicators that are useful for end-users.

14.2 General challenges and scientific relevance

The subject covered in this thesis was scrutinized with a holistic approach aimed at investigating the optimal design of the various components of the modeling chain depicted in Fig. 1.6, to understand how to improve its predictability of the overall chain, to understand how uncertainties affect final results, and to provide guidelines on how to develop operational tools for monitoring and forecasting that are useful for decision making in the water sector.

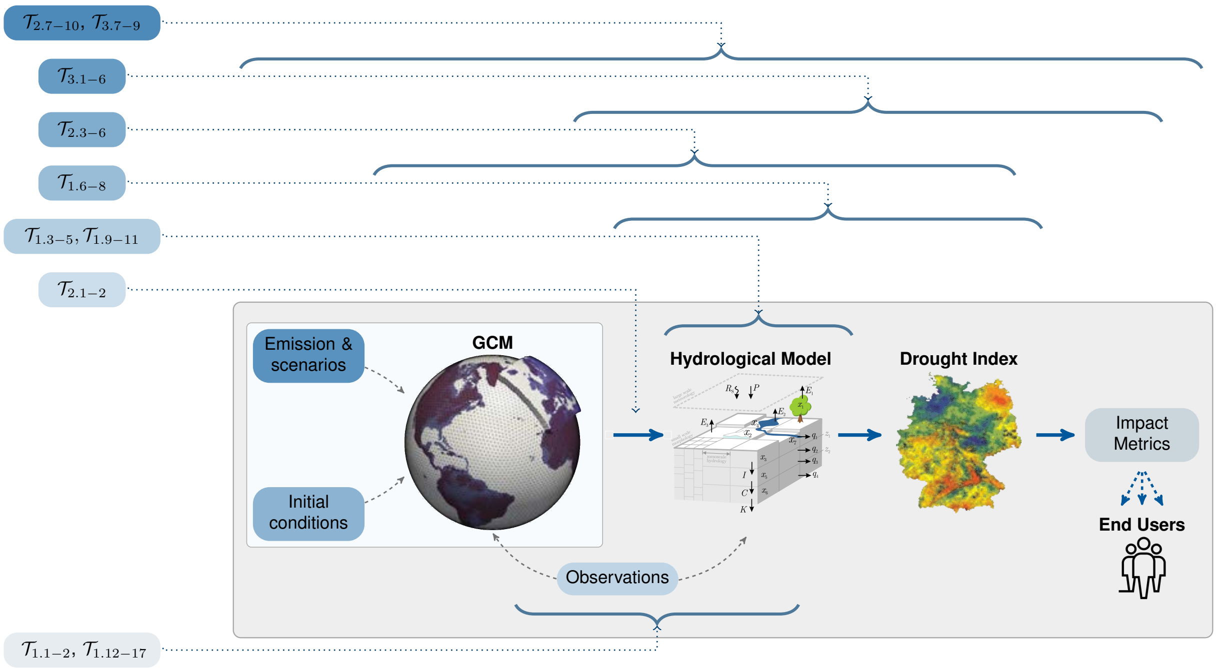

Figure 14.1 Schema depicting the relationship of the theses with the elements of the modeling chain.

The Figure 14.1 depicts the theses stated in Section 1.9.1 and their relation to a specific element of the modeling chain or the chain as a whole. As can be seen in this figure, most components have been investigated thoroughly and in relation with other ones. It should be noted that there is additional research carried out on this subject, but due to space limitations, have not been included in this thesis. Interested readers should refer to the references of the papers included in this thesis.

This study covered several major issues of contemporary hydro-meteorology as listed below:

- The detection of changes and causal relationships for drought characteristics at the mesoscale. These issues were mainly dealt with in *Samaniego and Bárdossy* (2007), and are related to theses $\mathcal{T}_{1.1-2}$.

- The development of a novel parameterization technique (MPR) and the corresponding model interphases leading to satisfactory cross-validation results across locations and scales. These developments provided significant insight on how to improve predictions in ungauged locations. The key for the development of MPR was the reduction of overparameterization in HMs/LSMs and the embracing of the scaling problem. These issues were mainly dealt with in *Samaniego et al.* (2010a, 2017) and are related to theses $\mathcal{T}_{1.3-5}$, $\mathcal{T}_{1.9-11}$.

- The development of novel bias-insensitive pattern matching objective functions to allow the conditioning of HM/LSM with remotely sensed data. The assimilation of RS data is expected to improve the parameterization of a HM/LSM. These issues were mainly dealt with in *Rakovec et al.* (2016c); *Zink et al.* (2018) and are related to theses $\mathcal{T}_{1.12-17}$.

- The development of robust drought indicators aimed at describing the evolution of hydrological and agricultural droughts. These issues were covered in *Samaniego et al.* (2013, 2016) and are related to theses $\mathcal{T}_{1.6-8}$ and $\mathcal{T}_{2.4-6}$.

- The qualification of the uncertainty related with drought characteristics and indices and their propagation along a modeling chain. These issues were mainly dealt with in *Samaniego et al.* (2013, 2016) and are related to theses $\mathcal{T}_{1.6-8}$ and $\mathcal{T}_{2.4-6}$.

- The development and investigation of preprocessing and sub-ensemble selection techniques to improve the skill of dynamic forecasts compared with climatological based-forecasts (ESP) used as reference. These issues were mainly covered in *Thober et al.* (2015) and related to theses $\mathcal{T}_{2.1-2}$.

- The evaluation of the skill and efficiency of modeling chains aimed at monitoring, forecasting, and projecting drought events across spatial and temporal scales. These issues were the subject of *Samaniego et al.* (2018); *Thober et al.* (2015); *Zink et al.* (2016) and *Samaniego et al.* (2019a), which are related to theses $\mathcal{T}_{2.3}$, $\mathcal{T}_{2.7-10}$, $\mathcal{T}_{3.1-3}$, $\mathcal{T}_{3.7-9}$.

- The design and operationalization of drought (and flood) monitoring and forecasting modeling chains. These issues were the subject of *Zink et al.* (2016) and *Samaniego et al.* (2019a), which are related to theses $\mathcal{T}_{2.7-10}$ and $\mathcal{T}_{3.1-3}$.

- The design of robust statistical impact models and metrics that are suitable for end-users. These topics were the subject of *Peichl et al.* (2018) and are related to theses $\mathcal{T}_{3.4-6}$.

14.3 Conclusions

This systematic work presented here has allowed us to derive several significant conclusions that are relevant for the hydrological science community, water resources planners and practitioners. Below a synthesis of the main lessons learnt in this series of papers.

- Machine learning and statistical techniques can help us to identify causal relationships and stochastic dependency among a chain of variables. Based on observed data alone, it is possible to find significant evidence that anthropogenic land cover and climatic changes induce impacts on low-flow streamflow time series, especially in summer. This implies, that all variables related with the hydrological cycle are modified in one way or another. Findings derived from these studies were fundamental to develop and parameterize the mHM model.

- Recognizing the role of subgrid variability of model parameters and their relationship with geo-physiographic characteristics was a fundamental insight to developing the multiscale parameter regionalization (MPR) scheme. We have demonstrated that any model (e.g., mHM, VIC, PCR-GLOBWB) that is parameterized with this technique outperforms, ceteris paribus, the same model but parameterized with standard techniques such as hydrological response units (HRUs), standard regionalization techniques, or brute-force calibration. MPR can be interpreted as a sophisticated statistical regularization technique that leads to a parsimonious model parameterization with excellent transferability performance across scales and locations. MPR is, in conclusion, an excellent alternative to estimating parameters at the REA-scale of any land surface or hydrological model (LSM/HM). It also has the advantage of generating flux-matching simulations across spatial scales and therefore of making a model quasi-scale invariant.

- By using the MPR technique in mHM, it was possible to estimate water fluxes seamlessly across spatial scales. This technique allowed us to run this model at the native scale of remotely sensed (RS) observations such as GRACE TWS, MODIS ET, H-SAF LST. Using mHM and MPR, it was then possible to demonstrate that streamflow assimilation (via inverse modeling) was a necessary but not sufficient condition to improve the skill of the model to reproduce model states such as soil moisture, land surface temperature, terrestrial water storage anomalies, and water fluxes such as evapotranspiration. Consequently, the assimilation of additional information is of great importance to improving the model's performance and transferability. Assimilating RS-products is, on the contrary, quite cumbersome due to their intrinsic bias and uncertainties. The development of a novel bias-insensitive pattern matching technique was a fundamental step to easing the use of these products into LSMs/HMs.

- There are countless drought indices and combinations of them. Most of these indices are based on precipitation data only (e.g., SPI) or precipitation and potential evaporation based on temperature observations (e.g., SPEI) or empirical and over-parameterized indices such as the Palmer Drought Index. In this work, it was proven that SPI and as an extension SPEI are not suitable for describing hydrological, agricultural, and groundwater droughts. It was also shown that the self-calibrating Palmer Drought Index is not suitable for climate projection studies.

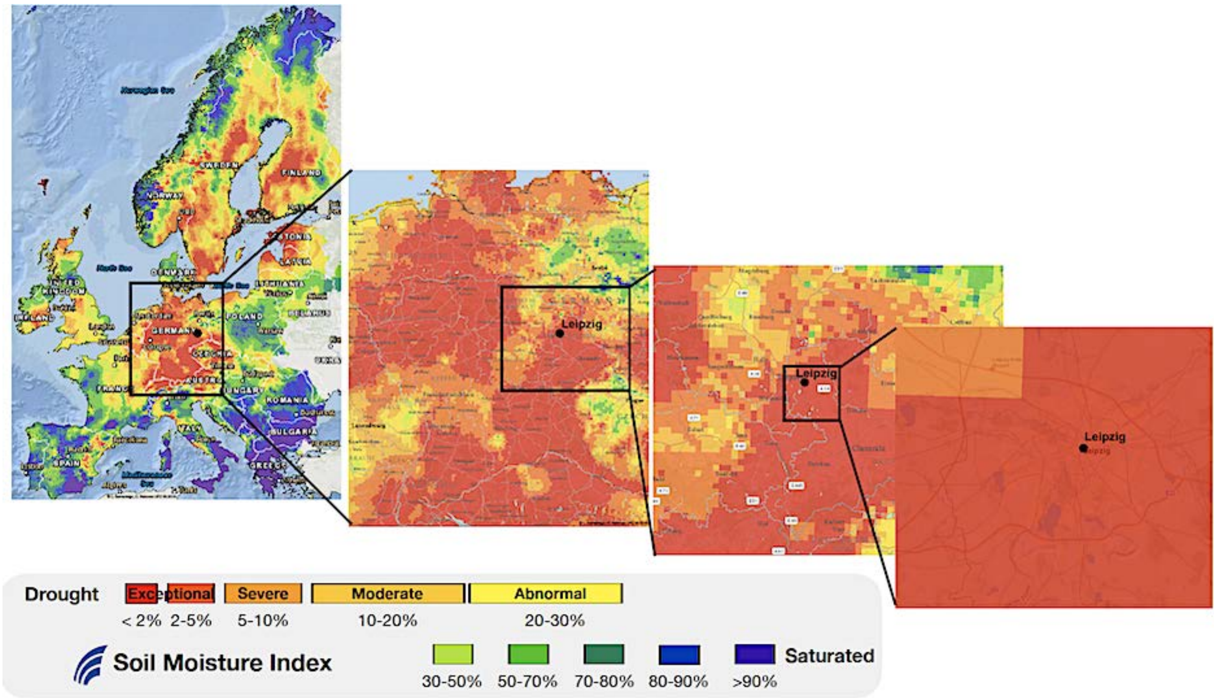

Figure 14.2 Seamless soil moisture index (SMI) at multiple spatial resolutions and scales based on mHM simulations using E-OBS v18 forcing data for August 2018. The SMI drought classes indicate exceedance probabilities of the fraction of soil moisture w.r.t. saturation, for a given cell and calendar month.

Instead, we demonstrated that a LSM or a HM (e.g., mHM) can be used to estimate the past, current and future soil moisture states at any spatial resolution (see Figure 14.2). We also showed that for comparison across location and time, a percentile-based index is appropriate for defining drought characteristics such as duration, area, magnitude, intensity, and severity. We also showed that standard approaches to fit a theoretical distribution function of soil moisture (or runoff) at a given location and calendar month (or day) is not appropriate. Instead a non-parametric kernel-based approach was proposed.

- It has been argued in the literature that the epistemic uncertainty (input, parametric, structural) in hydrological models (or in LSMs) is significant. Little, however, has been done to estimate the implications of the parametric uncertainty into drought identification and estimation. In this respect, we demonstrated that parametric uncertainty plays a fundamental role in drought modeling and developed a non-parametric algorithm to decompose the uncertainty originating from GCMs and HMs in seasonal forecasts and climate projections.
- Data availability plays a critical role for the development of drought modeling chains. We demonstrated that the existing open-source data sets can be used to develop both drought monitoring systems, and seasonal forecasting or climate projection systems. An example of the former type is the German Drought Monitor, and for the latter the EDgE project portal.
- Current seasonal forecast products have still coarser temporal and spatial resolutions. GCM model outputs are also biased. Consequently, we needed to develop tools to bias-correct, disaggregate and select ensemble members to maximise the model-chain skill scores. In this respect, we proposed the external drift Kriging technique for spatial disaggregation and developed a stochastic temporal disaggregation approach based on a multiplicative cascade for seamless temporal disaggregation of GCM forcings. Considering that not all GCMs have the same skill in a given region, we proposed a forward/backward elimination ensemble search algorithm to estimate the optimal ensemble size for a given region or time.
- We also showed that the ensemble prediction is in all cases and locations a better predictor than the single best HM, and that optionally selected subensembles only show performance losses less than 1% on average in

comparison to the full ensemble but with a saving of 60% in computational demand. It was also demonstrated that GCM seasonal forecasts have better skills than climatological forecast used as benchmarks (e.g., ESP). These findings are of great significance for the development of an effective operational seasonal forecast service at global scale, e.g., in the project ULYSSES commissioned by the Copernicus Climate Change Service - ECMWF to the UFZ, under my coordination.

- We showed that the forecasting skill is mainly determined by initial hydrologic conditions, and hence it is of crucial importance to know how a given model processes the existing geo-physiographical datasets to derive model parameters at a given target resolution.
  In the EDgE project, for example, it was possible to demonstrate that the mHM model, the only model that at that moment was fully parameterized with the MPR technique, has the best performance over the 300+ gauging stations across Europe. Similar experiences have been obtained in the ISI-MIP2a project, over the MOPEX-CONUS dataset, and currently in a Canadian inter-comparison project (*Mai et al.*, 2019).

- We demonstrated that by using existing data sets and GCM outputs, it is possible to predict soil moisture droughts over Europe, with a lead time of up to two months.

- The high-resolution multi-model simulations carried out over Europe allowed us, for the first time, to estimate the sensitivities of anthropogenic global warming on aridity, drought characteristics such as extent, duration, and frequency, changes in low- and high-flow percentiles of streamflow as well as potential impacted population by the near- future, middle and end of the century. The sensitivities of the drought characteristics were estimated with a time sampling approach because every GCM under an emission scenario reaches a fixed temperature increase with respect to the global average at different points in time. The large ensemble size allowed us to also estimate the uncertainty of each drought characteristic. In this study, we also showed the necessity of adapting the reference period used to define the empirical distribution functions of the target variables (e.g., soil moisture), which are used to estimate the percentile-based drought indices (e.g., SMI) as a direct consequence of the non-stationarity of the meteorological forcings. As a result, we showed that events similar to the 2003 drought in Europe will become twice as frequent be the end of this century.

- The econometric climate impact yield models found in the literature mainly use climatic variables such as precipitation and temperature, among other location specific variables, as predictors of yield change in the future. In this work, we demonstrated that this is not enough and proposed to include the soil moisture index as a better indicator of plant stress induced by agricultural droughts. As a result, the efficiency of the standard models were significantly improved over Germany.

- Finally, we concluded that the success of any water-oriented decision-support system used as monitoring, early-warning, seasonal forecasting or prediction platform, should be co-designed with key stakeholders to gain their acceptance and insights on the key variables and indicators that are relevant for their decision making activities. The success of the German drought monitor, measured by the continuous increase of web visitors, media references, as well as the positive evaluation of the focus-groups on the functionalities of the EDgE online platform corroborate this assertion.

14.4 Outlook

14.4.1 Towards robust high-resolution GHMs

The recent IPCC SR-15 report states that anthropogenic activities have already caused a 1.0 °C increase in global temperatures above pre-industrial levels. Hydrological impact studies for Europe listed in this IPCC report (among them e.g., *Marx et al.*, 2018; *Samaniego et al.*, 2018; *Thober et al.*, 2018) showed that the potential impact of these changes will become substantial for ecosystems and society during the coming decades. These facts reinforce the urgent need for supporting the United Nations' Sustainable Development Goals towards developing tools and impact models that help decision makers to implement ambitious adaptation and mitigation measures against global changes, from local to continental scales. The quintessence of such tools are the global hydrological models (GHM).

Three decades ago, *Eagleson* (1986) foresaw "the emergence of global-scale hydrology" as a fundamental need to provide answers to emergent environmental changes and the increasing demand for long range hydrologic forecasting at global rather than catchment scale. Eagleson also envisioned the need to develop integrated models accounting for physical and bio-chemical process related with the water, carbon, and nutrient cycles on Earth. Recently, *Bierkens* (2015) presented an appraisal of the state, trends, and directions in global hydrology, and one of the striking conclusions of his commentary is that the "runoff processes in GHMs are still represented rudimentarily". One year later, this assertion was confirmed by *Beck et al.* (2016), who evaluated available GHMs against observed streamflow at 1113 river basins across the globe and reported that, on average, these models have a Nash-Sutcliffe efficiency (NSE) of -0.09 if compared against monthly streamflow records! We have come a long way since the first blueprints for primitive land surface and hydrological models (*Freeze and Harlan*, 1969; *Manabe*, 1969) but the state-of-the-art of the majority of the GHMs, as demonstrated by *Beck et al.* (2016), is still not satisfactory. In fact, the capability of these models to close the water balance over a basin is practically null. Moreover, their spatial resolution is still too coarse (0.5°) for practical applications.

The crucial question is why these models performed so poorly? The hydrological community has struggled with these facts and issues over the last four decades (*Bierkens*, 2015; *Blöschl et al.*, 2013; *Dooge*, 1982; *Hrachowitz et al.*, 2013; *Li et al.*, 2012b; *Sivapalan et al.*, 2003). Potential answers to explain the poor performance are: 1) the forcing data does not represent the meteorological spatio-temporal conditions of the region; 2) the model has poor parameterizations and parameter fields that do not account for the subgrid variability of the physiographic variables; 3) the observed streamflow reveal large observational errors; and/or 4) the model misses key hydrological processes.

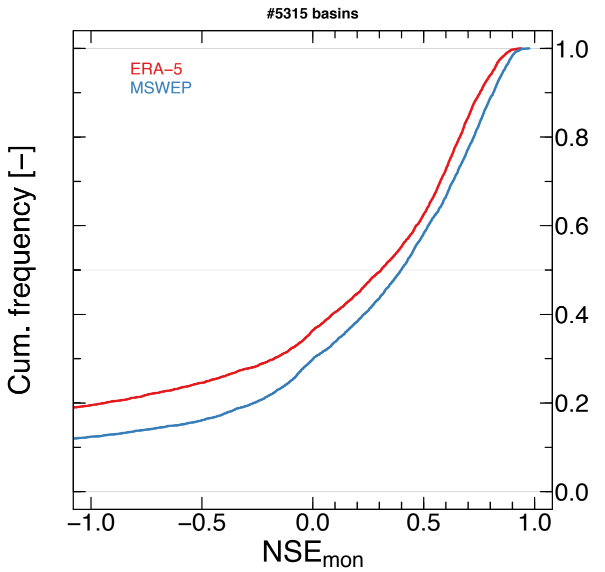

Figure 14.3 Monthly NSE for the uncalibrated mHM model over 5500 GRDC stations (*Samaniego et al.*, 2019b). ERA-5 forcing was obtained from the CDS at ECMWF. MSWEP forcing was kindly provided by E. F. Wood and H. Beck.

Considering that GHMs are basically using the same water balance equations, then the root of problem is very likely the parameterization of such models as pointed out by (*Dooge*, 1982) (see Section 1.6) and (*Bierkens*, 2015), who also remarked that a "sophisticated regionalization scheme, such as the multiscale approach by *Samaniego et al.* (2010a) could be a way forward" in this respect. Of course, other issues afecting model efficiency should be also taken seriously if we aim to improve the performance of the GHMs in the future.

To demonstrate the scalability and transferability of the mHM model and its parameterization technique (MPR), we setup this model on selected GRDC stations located in all continents and covering major hydro-climatic regimes. The model was set up at two spatial resolutions: 0.25° and 1.0° to match the spatial resolution of the ERA-5 forcings (ECMWF) and the GRACE (NASA) terrestrial water storage anomaly. The mHM model was not calibrated at these stations but used the default transfer-function parameter set. The results of such experiments is shown in Figure 14.3. The median of the monthly NSE for the uncalibrated mHM model over 5500 GRDC stations reaches a value of 0.40 for monthly streamflow (simulations from 1950-2016). In other words, an improvement of 544% with respect to the median value reported in *Beck et al.* (2016). The model performance of mHM obtained with MSWEP is slightly better than that obtained with ERA-5.

The comparison of terrestrial water storage (TWS) anomaly of GRACE against the corresponding anomalies derived from mHM simulations during the period from 2004 to 2016 at the resolution of 1.0° reveal an overall good agreement. In this case too, mHM is uncalibrated and uses the same parameter set as in the example above. The Pearson correlation between these two variables is depicted in Figure 14.4. This figure shows that there are hotspots of weaker model performance, and mHM, in general, performing better in humid regions than in semi-arid regions. Poor performance corresponds to regions were the water balance closure error is the largest. These errors could also be associated with poor forcing data sets in those regions.

In summary, these two examples show that the proposed MPR approach has the great potential to estimate parameters for high resolution mHM across the globe. Currently, an extraordinary large parameter estimation ex-

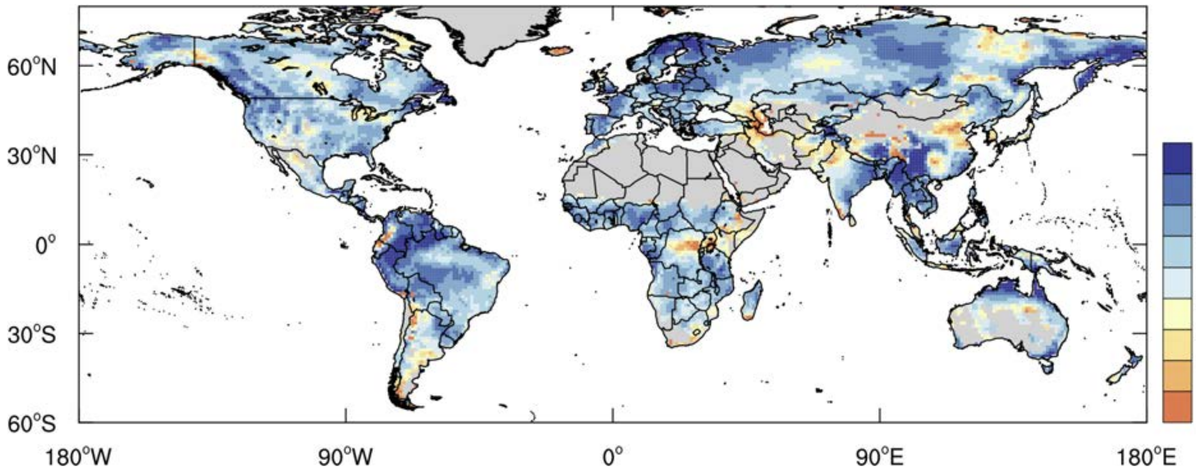

Figure 14.4 Pearson correlation of monthly TWS anomalies of GRACE and mHM from 2004-2016 (*Samaniego et al.*, 2019b).

periment to improve the performance of the global mHM beyond the current state-of-the-art is ongoing (*Samaniego et al.*, 2019b).

14.4.2 Towards a global drought monitoring and forecasting system

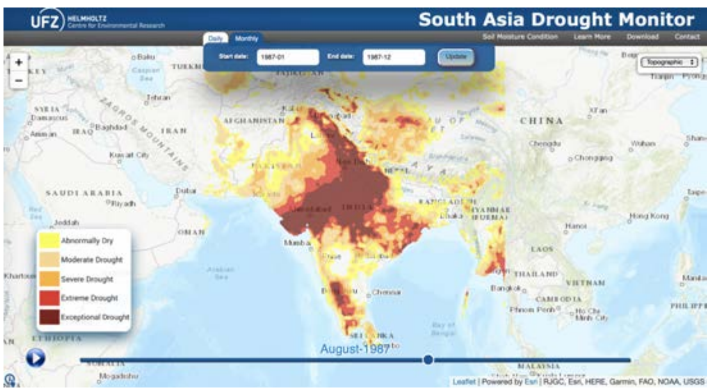

Figure 14.5 Prototype of the South-Asian drought monitor. The figure depicts the 1987 drought event, one of the largest since 1950. The SMI is simulated with mHM using ERA-5 forcings (*Saha et al.*, 2020). Source: `http://southasiadroughtmonitor.pythonanywhere.com`

The studies reported in this Thesis have been carried out mainly in Europe, where the historical meteorological records and geo-physiographical data is among the best in the world. Consequently, the remaining question to answer is, how well the proposed drought modelling chain performs in a different hydro-climatological region where the data situation is not as favorable as in Europe? To answer this question, the South-Asia subcontinent, covering an area of about 5.1 millon km^2, was selected for two reasons: first, 24% of the world's population live here and second, it is known that droughts have caused enormous impacts on regional agriculture, food storage, and livelihood in the past decades. Recent studies, for example, showed that the droughts of 1987 and 2002-2003 affected at least 50% of the crop area in India (*Wassmann et al.*, 2009) and similarly, the drought events from 1979 and 1982 incurred a loss of about 2 million and 53 000 tons of rice in Bangladesh, respectively (*Rakib et al.*, 2015).

The main meteorological drivers leading to extended droughts in this region are related to the increase in seasonality of precipitation and extreme heat. These, in turn, disrupt the vegetation growing season in Pakistan, India, and Bangladesh (*Vinke et al.*, 2017). Consequently, it is expected that by 2100, 50% of the wheat production in the Indo-Gangetic Plains will be reduced due to this natural hazard (*Vinke et al.*, 2017).

Currently, the only readily available drought-related information for the subcontinent are the SPI and SPEI indices provided by the Indian Meteorological service (`www.imdpune.gov.in`) and the International Water Management Institute `http://dms.iwmi.org`. Pakistan Meteorological Department (`http://www.pmd.gov.pk/`), only provides reports on droughts based on monthly rainfall data and satellite-based vegetation index. It is also known that SPI and SPEI are empirical indices that do not adequately represent extreme water stress conditions on vegetation (*Keyantash and Dracup*, 2002; *Mishra and Singh*, 2010).

Consequently, the application of a model chain consisting of a quasi-real time reanalysis product, the mHM model and the SMI index, could offer a simple but effective solution comparable to that of the GDM (*Zink et al.*, 2016). This chain also constitutes an effective proof-of-concept of the transferability of the proposed model chain to other locations. As forcings, the ERA-5 was selected due to its high resolution and latency. The hydrological model mHM was setup at the 0.25° resolution based on the global open-data sources described in *Samaniego et al.* (2019b). With this information, the daily soil moisture for the region was reconstructed from 1950 to 2019. Subsequently, the SMI, as described in *Samaniego et al.* (2013), was estimated. For the operationalization of the system, *Saha et al.* (2020) used the same scripts used by *Zink et al.* (2016) for the GDM. In addition to that, an interactive web-portal was designed by Toma Saha, as shown in Fig.14.5. These results are very encouraging considering that the model has not been calibrated, yet it is able to reproduce well the extension of the 1987 drought in the subcontinent, as simulated by the VIC model (*Mishra et al.*, 2014), although both reconstructions used different reanalysis datasets. Consequently, it can be concluded that the proposed system is ready to be extended to other regions of the world suffering from this natural hazard. The operationalization of the South-Asia and global drought monitoring systems at high resolution are ongoing projects.

14.4.3 The new generation of "smart" hydrological models

The scientific hydrologic community have pointed out that one of the unsolved grand challenges in hydrological sciences is to be able to monitor and simulate water fluxes at the land surface at higher resolutions so that they become available "everywhere and [are] locally relevant" (*Bierkens et al.*, 2014; *Wood et al.*, 2011). Equally important will be also to inform stakeholders of the related uncertainties of these hydrological predictions.

The papers presented in this thesis have contributed to this overarching goal by providing tools (e.g., MPR, mHM, mRM), a prototype of a model chain design and a protocol for parameter estimation and verification at multiple scales. The proposed model chain is aimed at generating key terrestrial Essential Climate Variables (tECVs) such as streamflow and soil moisture. These variables, in turn, are the basis for estimating user-specific sectorial climate impact indicators such as soil moisture (or runoff) index, drought extension and severity, changes in peak flows, among others.

The proposed model chain (Figure 1.6), however, should be seen as a proof-of-concept because of the following reasons. First, it is the initial attempt to model a chain of effects leading to the evolution of a drought event with its respective uncertainty. Second, the proposed model chain is scale-independent but does not account for feedback from the land surface processes or human activities to the GCM processes. Ergo, it constitutes a first-order simplification of the whole system. Future developments of this impact modeling chain should include all kinds of feedback to account for secondary effects that were neglected in this first attempt.

Future hydrological models should also be improved with respect to efficiency and skill, and become locally relevant but globally applicable. LSM/HMs should have an adequate complexity to be computational tractable and fed by existing datasets; and should be transferable and robust but not over-parameterized. Moreover, next generation LSM/HMs should be fully integrated into a new kind of "Earth System Model", which should account for all relevant physical, and geo-chemical processes related to water and matter cycles as well as accounting for man-made interactions taken place at local, regional, and global scales.

The overall goal of future modeling efforts should be to keep the balance between process complexity, data availability, and predictive uncertainty. This new kind of optimal-complexity land surface/hydrological models, hereafter denoted as "smart" models, should help decision makers to develop adaptation and mitigation measures at multiple scales and be globally aplicable. Smart models should become operational decision support tools.

New data-science methods and machine learning (ML) algorithms will play a key role in the future of hydrological sciences and for developing the new generation of "smart" global hydrologic models at high-resolution. ML-algorithms will be used to extend existing hydrological models by linking them to robust impact models targeting specific sectors such agro-production, energy generation, water resources operations, hazard relief and early-warning systems. ML-based modules will also need to be locally relevant but globally applicable.

The development of new "smart" impact models will largely depend on the available skills and the capability to harness the recently available large data sets (ranging from tera- to petabytes). Data-Science methods are fundamental to distilling stochastic and causal dependency relationships as well as heuristic rules that will be the kernel of such models. Suitable multi-variate inference methods applied to large data-cubes will be key to estimating the predictive uncertainty of these new ML-based impact models. For example, the new generation of global hydrological models should include a more realistic integration of man-made driven activities affecting the water cycle such as large-scale dams, diversion, groundwater pumping, and irrigation infrastructure.

Future generation of LSMs/HMs require large-scale modules for sediment transport, water quality, water temperature, groundwater, and cold processes (e.g., permafrost and glacier evolution). This new generation of models should routinely use multi-scale datasets to improve their parameterization. Among potential data sets are the global networks of eddy covariance stations (e.g., ICOS or FLUXNET), cosmic ray neutron sensors, gravimeters, as well as the new generation of remotely sensed products such as GRACE-FO, Sentinel 1-6, Copernicus LSTM, CHIME and SWOT. In this regard, the MPR technique should be further developed to become model-agnostic (on going development at `https://git.ufz.de/chs/MPR`) to support the parameterization of new processes and the integration with LSM/HM of varying complexity.

After more than a decade of work on this subject, based on the results obtained in my research and that of many other colleagues, and given the promising technological developments in the near future, I envision a situation, not far in the future, in which humans, empowered by fully operational "smart" models for everywhere, will able to "defeat" the "Bull of Heaven" and minimize its socio-economic and ecological impacts that have affected mankind since its origins.

Bibliography

Aarts, E., and J. Korst (1990), *Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing*, Wiley, Chichester.

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell, and J. Rasmussen (1986), An Introduction to the European Hydrological System - Systeme Hydrologique Europeen, She .1. History and Philosophy of a Physically-Based, Distributed Modeling System, *J. Hydrol.*, *87*(1-2), 45–59.

Abdulla, F., and D. Lettenmaier (1997), Development of regional parameter estimation equations for a macroscale hydrologic model, *J. Hydrol.*, *197*, 230–257.

Addor, N., O. Rössler, N. Köplin, M. Huss, R. Weingartner, and J. Seibert (2014), Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments, *Water Resources Research*, *50*(10), 7541–7562.

Agriculture Risk Management Team (2011), Weather Index Insurance for Agriculture: Guidance for Development Practitioners, *Tech. Rep. November*, The World Bank, Washington.

Ahmed, S., and G. De Marsily (1987), Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity, *Water Resources Research*, *23*(9), 1717–1737.

Aich, V., S. Liersch, T. Vetter, S. Huang, J. Tecklenburg, P. Hoffmann, H. Koch, S. Fournet, V. Krysanova, E. N. Müller, and F. F. Hattermann (2014), Comparing impacts of climate change on streamflow in four large african river basins, *Hydrology and Earth System Sciences*, *18*(4), 1305–1321.

Akaike, H. (1973a), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, *19*, 716–723.

Akaike, H. (1973b), Information theory and an extension of the maximum likelihood principle, in *International Symposium on Information Theory*, pp. 267–281, Springer New York.

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger (2013), GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth. Syst. Sc.*, *17*(3), 1161–1175.

Alfieri, L., F. Pappenberger, F. Wetterhall, T. Haiden, D. Richardson, and P. Salamon (2014), Evaluation of ensemble streamflow predictions in Europe, *J. Hydrol.*, *517*, 913–922.

Allen, R. G. R., L. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, *Tech. Rep. 56*, FAO - Food and Agriculture Organization of the United Nations, Rome.

Andersen, O. B., S. I. Seneviratne, J. Hinderer, and P. Viterbo (2005), GRACE derived terrestrial water storage depletion associated with the 2003 European heat wave, *Geophys. Res. Lett.*, *32*(18).

Andreadis, K. M., and D. P. Lettenmaier (2006), Trends in 20th century drought over the continental United States, *Geophys. Res. Lett.*, *33*(10), L10,403.

Andreadis, K. M., E. A. Clark, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier (2005), Twentieth-Century Drought in the Conterminous United States, *Journal of Hydrometeorology*, *6*(6), 985–1001.

Andréassian, V., F. Bourgin, L. Oudin, T. Mathevet, C. Perrin, J. Lerat, L. Coron, and L. Berthet (2014), Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency, *Water Resources Research*, *50*(10), 8356–8366.

Andresen, J. A., G. Alagarswamy, C. A. Rotz, J. T. Ritchie, and A. W. LeBaron (2001), Weather impacts on maize, soybean, and alfalfa production in the Great Lakes region, 1895-1996, *Agronomy Journal*, *93*(5), 1059–1070.

Angrist, J. D., and J.-S. Pischke (2008), *Mostly harmless econometrics : an empiricist's companion*, March, Princeton Univers. Press.

Annan, F., and W. Schlenker (2015), Federal Crop Insurance and the Disincentive to Adapt to Extreme Heat, *American Economic Review: Papers and Proceedings*, *105*(5), 262–266.

Arellano, M. (1987), Computing Robust Standard Errors for Within Group Estimators.

Arnal, L., H. L. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski, and F. Pappenberger (2018), Skilful seasonal forecasts of streamflow over Europe?, *Hydrology and Earth System Sciences*, *22*(4), 2057–2072.

Arnell, N. W. (2011), Uncertainty in the relationship between climate forcing and hydrological response in UK catchments, *Hydrology and Earth System Sciences*, *15*(3), 897–912.

Arnold, J. B. (2016), ggthemes: Extra Themes, Scales and Geoms for 'ggplot2' [R package ggthemes version 3.3.0].

Auffhammer, M., and W. Schlenker (2014), Empirical studies on agricultural impacts and adaptation, *Energy Economics*, *46*, 555–561.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013), Using Weather Data and Climate Model Output in Economic Analyses of Climate Change, *Review of Environmental Economics and Policy*, *7*(2), 181–198.

Ball, J. T., I. E. Woodrow, and J. A. Berry (1987), A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions, in *Progress in Photosynthesis Research*, pp. 221–224, Springer Netherlands, Dordrecht.

Bárdossy, A. (1993), *Stochastische Modelle zur Beschreibung der raum-zeitlichen Variabilität des Niederschlages*, vol. 44, Institut für Hydrologie und Wasserwirtschaft der Universität Karlsruhe, Karlsruhe.

Bárdossy, A., and F. Filiz (2005), Identification of flood producing atmospheric circulation patterns, *Journal of Hydrology*, *313*, 48–57.

Bárdossy, A., and E. Plate (1992), Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resources Research*, *28*, 1247–1259.

Bárdossy, A., J. Stehlik, and H.-J. Caspary (2002), Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimised fuzzy rules, *Climate Research*, *23*, 11–22.

Barnabás, B., K. Jäger, and A. Fehér (2008), The effect of drought and heat stress on reproductive processes in cereals, *Plant, Cell and Environment*, *31*(1), 11–38.

Barrios, M., and F. Francés (2011), Spatial scale effect on the upper soil effective parameters of a distributed hydrological model, *Hydrological Processes*, *26*(7), 1022–1033.

Basso, B., and J. Ritchie (2014), Temperature and drought effects on maize yield, *Nature Climate Change*, *4*(April), 233.

Bastiaanssen, W., M. Menenti, R. Feddes, and A. Holtslag (1998), A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation, *Journal of Hydrology*, *212-213*(1-4), 198–212.

Batalla, R. J., C. M. Gómez, and G. Kondolf (2004), Reservoir-induced hydrological changes in the Ebro River basin (NE Spain), *J. Hydrol.*, *290*(1-–2), 117–136.

Batjes, N. H. (1996), Development of a world data set of soil water retention properties using pedotransfer rules, *Geoderma*, *71*(1-2), 31–52.

Bauer, P., A. Thorpe, and G. Brunet (2015), The quiet revolution of numerical weather prediction, *Nature*, *525*(7567), 47–55.

Beck, H. E., A. I. J. M. van Dijk, A. de Roo, D. G. Miralles, T. R. McVicar, J. Schellekens, and L. A. Bruijnzeel (2016), Global-scale regionalization of hydrologic model parameters, *Water Resources Research*, *52*(5), 3599–3622.

Becker, A., and J. J. McDonnell (1998), Topographical and ecological controls of runoff generation and lateral flows in mountain catchments, *IAHS Publication*, *248*, 199–206.

Becker, P., F. Imbery, K. Friedrich, M. Rauthe, A. Matzarakis, A. Grätz, and W. Janssen (2015), Klimatologische Einschätzung des Sommer 2015, *Tech. rep.*, Deutscher Wetter Dienst.

Beldring, S., K. Engeland, L. A. Roald, N. R. Saelthun, and A. Vokso (2003), Estimation of parameters in a distributed precipitation-runoff model for Norway, *Hydrol. Earth Syst. Sci.*, *7*(3), 304–316.

Benito, G., R. Brázdil, J. Herget, and M. J. Machado (2015), Quantitative historical hydrology in Europe, *Hydrology and Earth System Sciences*, *19*(8), 3517–3539.

Berg, A., J. Sheffield, and P. C. D. Milly (2017), Divergent surface and total soil moisture projections under global warming, *Geophysical Research Letters*, *44*(1), 236–244, 2016GL071921.

Berger, K. P., and D. Entekhabi (2001), Basin hydrologic response relations to distributed physiographic descriptors and climate, *J. Hydrol.*, *247*(3–4), 169–182.

Bergström, S. (1995), The HBV Model, in *Computer Models of Watershed Hydrology*, edited by V. Singh, pp. 443–476, Water Resources Publications, Colorado, USA.

Berry, S. T., M. J. Roberts, and W. Schlenker (2014), Corn Production Shocks in 2012 and Beyond: Implications for Harvest Volatility, in *The Economics of Food Price Volatility*, edited by J.-P. Chavas, D. Hummels, and B. D. Wright, pp. 59–81, University of Chicago Press.

Best, M. J., M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding (2011), The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, *Geoscientific Model Development*, *4*(3), 677–699.

Beven, K. (1993), Prophesy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, *16*, 41–51.

Beven, K. (1995), Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models, *Hydrological Processes*, *9*(5-6), 507–525.

Beven, K. (2001), How far can we go in distributed hydrological modelling?, *Hydrol. Earth Sys. Sci.*, *5*(1), 1–12.

Beven, K. (2002), Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, *Hydrological Processes*, *16*(2), 189–206.

Beven, K., P. J. Smith, and A. Wood (2011), On the colour and spin of epistemic error (and what we might do about it), *Hydrology and Earth System Sciences*, *15*(10), 3123–3133.

Beven, K. J., and H. L. Cloke (2011), Defining grand challenges in hydrology: A comment on Wood et al. (2011) Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*

Bierkens, M. F. P. (2015), Global hydrology 2015: State, trends, and directions, *Water Resources Research*, *51*(7), 4923–4947.

Bierkens, M. F. P. (2018), Calibrating Hydrological Models by Satellite., *Earth & Space Science News Eos*, https://eos.org/editor-highlights/calibrating-hydrological-models-by-satellite.

Bierkens, M. F. P., V. A. Bell, P. Burek, N. Chaney, L. E. Condon, C. H. David, A. de Roo, P. Döll, N. Drost, J. S. Famiglietti, M. Flörke, D. J. Gochis, P. Houser, R. Hut, J. Keune, S. Kollet, R. M. Maxwell, J. T. Reager, L. Samaniego, E. Sudicky, E. H. Sutanudjaja, N. van de Giesen, H. Winsemius, and E. F. Wood (2014), Hyper-resolution global hydrological modelling: what is next? "Everywhere and locally relevant", *Hydrological Processes*, pp. n/a–n/a.

Binley, A. M., K. J. Beven, and J. Elgy (1989), A Physically-Based Model of Heterogeneous Hillslopes II. Effective Hydraulic Conductivities, *Water Resour. Res.*, *25*(6), 1227–1233.

Bivand, R., and N. Lewin-Koh (2017), maptools: Tools for Reading and Handling Spatial Objects [R package version 0.8-41].

Bivand, R., E. Pebesma, and V. Gómez-Rubio (2013), *Applied spatial data analysis with R*, 405 pp., Springer.

Bivand, R., T. Keitt, B. Rowlingson, E. Pebesma, M. Sumner, R. Hijmans, and E. Rouault (2016), Bindings for the Geospatial Data Abstraction Library: Package 'rgdal'.

Bjerknes, V. (1904), Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik (The problem of weather prediction, considered from the viewpoints of mechanics and physics), *Meteorologische Zeitschrift*, *21*, 1–7, (translated and edited by Volken E. and S. Brönnimann (2009), Meteorol. Z., 18, 663–667).

Blöschl, G. (1999), Scaling issues in snow hydrology, *Hydrol. Processes*, *13*(14–15), 2149–2175.

Blöschl, G. (2001), Scaling in hydrology, *Hydrol. Process.*, *15*(4), 709–711.

Blöschl, G., and A. Montanari (2010), Climate change impacts—throwing the dice?, *Hydrol Process*, *24*, 374–381.

Blöschl, G., R. B. Grayson, and M. Sivapalan (1995), On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, *Hydrol. Processes*, *9*(3–4), 313–330.

Blöschl, G., C. Reszler, and J. Komma (2008), A spatially distributed flash flood forecasting model, *Environ. Model. Softw.*, *23*(4), 464–478.

Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione, and S. H (Eds.) (2013), *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, Cambridge University Press, ISBN: 978-1107028180.

Blöschl, G., M. F. Bierkens, ..., E. F. Wood, R. Woods, Z. Xu, K. K. Yilmaz, and Y. Zhang (2019), Twenty-three unsolved problems in hydrology (uph) – a community perspective, *Hydrological Sciences Journal*, *64*(10), 1141–1158.

BMEL (2014), Erntebericht 2014: Mengen und Preise, *Tech. Rep. August*, Federal Ministry of Food and Agriculture, Bonn.

BMEL (2015), Erntebericht 2015 : Mengen und Preise, *Tech. Rep. August*, Federal Ministry of Food and Agriculture, Bonn/Berlin.

Bock, A. R., L. E. Hay, G. J. McCabe, S. L. Markstrom, and R. D. Atkinson (2016), Parameter regionalization of a monthly water balance model for the conterminous United States, *Hydrology and Earth System Sciences*, *20*(7), 2861–2876.

Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier (2013), Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models, *Agr. Forest. Meteorol.*, *176*, 38–49.

Bolaños, J., and G. O. Edmeades (1996), The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize, *Field Crops Research*, *48*(1), 65–80.

Bonan, G. B. (1995), Land Atmosphere Interactions for Climate System Models - Coupling Biophysical, Biogeochemical, and Ecosystem Dynamical Processes, *Remote Sens. Environ.*, *51*(1), 57–73.

Bonan, G. B., P. J. Lawrence, K. W. Oleson, S. Levis, M. Jung, M. Reichstein, D. M. Lawrence, and S. C. Swenson (2011), Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, *Journal of Geophysical Research-Atmospheres*, *116*(G2), GB1008.

Boni, G., D. Entekhabi, and F. Castelli (2001), Land data assimilation with satellite measurements for the estimation of surface energy balance components and surface control on evaporation, *Water Resources Research*, *37*(6), 1713.

Booij, M. J. (2005), Impact of climate change on river flooding assessed with different spatial model resolutions, *J. Hydrol.*, *303*, 176–198.

Boone, A., F. Habets, J. Noilhan, D. Clark, P. Dirmeyer, S. Fox, Y. Gusev, I. Haddeland, R. Koster, D. Lohmann, et al. (2004), The Rhone-Aggregation land surface scheme intercomparison project: An overview, *J. Climate*, *17*(1), 187–208.

Bosshard, T., M. Carambia, K. Goergen, S. Kotlarski, P. Krahe, M. Zappa, and C. Schar (2013), Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, *Water Resources Research*, *49*(3), 1523–1536.

Boughton, W., and F. Chiew (2007), Estimating runoff in ungauged catchments from rainfall, pet and the awbm model, *Environ. Model. Soft.*, *22*, 476–487.

Brázdil, R., A. Kiss, J. Luterbacher, D. J. Nash, and L. Řezníčková (2018), Documentary data and the study of past droughts: a global state of the art, *Climate of the Past*, *14*(12), 1915–1960.

Brier, G. (1950), Verification of forecasts expressed in terms of probability, *Mon. Wea. Rev.*, *78*(1), 1–3.

Briffa, K. R., G. Van Der Schrier, and P. D. Jones (2009), Wet and dry summers in Europe since 1750: evidence of increasing drought, *International Journal of Climatology*, *29*(13), 1894–1905.

Bronstert, A., B. Glüsing, and E. Plate (1998), Physically-based hydrological modelling on the hillslope and micro-catchment scale: Examples of capabilities and limitations, *IAHS-AISH Publication*, *248*, 207–215.

Brooks, R. H., and A. T. Corey (1964), Hydraulic properties of porous media, *Tech. Rep. 3*, Colorado State University, Fort Collins.

Brunet, G., S. Jones, and B. Mills (2015), Introduction, in *Seamless prediction of the Earth system: from minutes to months*, edited by M. Bëland and A. Thorpe, 1156, chap. 1, p. 471, World Meteorological Organization, Geneva.

Brynjarsdottir, J., and A. O'Hagan (2014), Learning about physical parameters: the importance of model discrepancy, *Inverse Problems*, *30*(11).

Bryson, R. A., and T. J. Murray (1977), *Climates of hunger : mankind and the world's changing weather*, Australian National University Press, Canberra, ACT, Australia.

Budyko, M. (1974), *Climate and Life*, International geophysics series, Academic Press.

Buizza, R. (2002), *Chaos and weather prediction*, Meteorological Training Course Lecture Series, ECMWF.

Büntgen, U., V. Trouet, D. Frank, H. H. Leuschner, D. Friedrichs, J. Luterbacher, and J. Esper (2010), Tree-ring indicators of German summer drought over the last millennium, *Quaternary Science Reviews*, *29*(7-8), 1005–1016.

Burke, M., and K. Emerick (2016), Adaptation to Climate Change: Evidence from US Agriculture, *American Economic Journal: Economic Policy*, *8*(3), 106–140.

Burnash, R., R. Ferral, and R. McGuire (1973a), A Generalized Streamflow Simulation System Conceptual Modeling for Digital Computers, *Report*, Joint Fed. State River Forecast Center, U.S. Natl. Weather Serv. and Calif. Dep. of Water Resour., Sacramento, CA.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973b), *A generalized streamflow simulation system: Conceptual modeling for digital computers*, U.S. Dept. of Commerce, National Weather Service.

Butler, E. E., and P. Huybers (2013), Adaptation of US maize to temperature variations, *Nature Climate Change*, *3*(1), 68–72.

Butler, E. E., and P. Huybers (2015), Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase, *Environmental Research Letters*, *10*(034009), 8.

Cai, X., Z.-L. Yang, C. David, G.-Y. Niu, and M. Rodell (2014), Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin, *J. Geophys. Res. Atmos.*, *119*(1), 23–38.

Campbell, G., and S. Shiozawa (1994), Prediction of hydraulic properties of soils using particle-size distribution and bulk density data, in *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, University of California*, edited by M. van Genuchten et al., pp. 317–328, Riverside, CA.

Cannon, A., and P. Whitfield (2002), Downscaling recent streamflow conditions in british columbia, canada using ensemble neural network models., *J. Hydrol.*, *259*, 136–151.

Cardano, G. (1993), *Ars Magna or the rules of algebra*, NY: Dover Publications, New York, 291 p., ISBN: 0–486-67811-3.

Carleton, T., and S. Hsiang (2016), Social and Economic Impacts of Climate Change, *Science*.

Chaney, N. W., P. Metcalfe, and E. F. Wood (2016a), HydroBlocks: a field-scale resolving land surface model for application over continental extents, *Hydrological Processes*, *30*(20), 3543–3559.

Chaney, N. W., E. F. Wood, A. B. McBratney, J. W. Hempel, T. W. Nauman, C. W. Brungard, and N. P. Odgers (2016b), POLARIS: A 30-meter probabilistic soil series map of the contiguous United States, *Geoderma*, *274*, 54–67.

Chauhan, N. S., S. Miller, and P. Ardanuy (2003), Spaceborne soil moisture estimation at high resolution: a microwave-optical/ir synergistic approach, *Int. J. Remote Sens.*, *24*(22), 4599–4622.

Chen, T. H., and Coauthors (1997), Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes, *J. Climate*, *10*, 1194–1215.

Cherkauer, K. A., L. C. Bowling, and D. P. Lettenmaier (2003), Variable infiltration capacity cold land process model updates, *Global and Planetary Change*, *38*(1-2), 151–159.

Chetty, R. (2009), Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods, *Annual Review of Economics*, *1*(1), 451–488.

Chmielewski, F. M. (2011), Wasserbedarf in der Landwirtschaft, in *WARNSIGNAL KLIMA: Genug Wasser für alle?*, 3 ed., pp. 149–156, Universität Hamburg, Institut f. Hydrobiologie.

Chow, V. (Ed.) (1964), *Handbook of Applied Hydrology: a compendium of water resources technology*, McGraw-Hill, New York.

Chow, V. T., D. R. Maidment, and L. W. Mays (1988), *Applied hydrology*, 540 pp., McGraw-Hill, New York.

Christensen, N. S., and D. P. Lettenmaier (2007), A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, *Hydrology and Earth System Sciences*, *11*(4), 1417–1434.

Ciais, P., M. Reichstein, N. Viovy, A. Granier, J. Ogee, V. Allard, M. Aubinet, N. Buchmann, C. Bernhofer, A. Carrara, F. Chevallier, N. De Noblet, A. D. Friend, P. Friedlingstein, T. Grunwald, B. Heinesch, P. Keronen, A. Knohl, G. Krinner, D. Loustau, G. Manca, G. Matteucci, F. Miglietta, J. M. Ourcival, D. Papale, K. Pilegaard, S. Rambal, G. Seufert, J. F. Soussana, M. J. Sanz, E. D. Schulze, T. Vesala, and R. Valentini (2005), Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, *437*(7058), 529–533.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resources Research*, *14*(4), 601–604.

Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, *47*(9), 1–16.

Clark, M. P., B. Nijssen, J. D. Lundquist, D. Kavetski, D. E. Rupp, R. A. Woods, J. E. Freer, E. D. Gutmann, A. W. Wood, L. D. Brekke, J. R. Arnold, D. J. Gochis, and R. M. Rasmussen (2015), A unified approach for process-based hydrologic modeling: 1. Modeling concept, *Water Resources Research*, pp. n/a–n/a.

Clark, M. P., B. Schaefli, S. J. Schymanski, L. Samaniego, C. H. Luce, B. M. Jackson, J. E. Freer, J. R. Arnold, R. D. Moore, E. Istanbulluoglu, and S. Ceola (2016), Improving the theoretical underpinnings of process-based hydrologic models, *Water Resources Research*, *52*(3), 2350–2365.

Clark, M. P., M. F. P. Bierkens, L. Samaniego, R. A. Woods, R. Uijlenhoet, K. E. Bennett, V. R. N. Pauwels, X. Cai, A. W. Wood, and C. D. Peters-Lidard (2017), The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, *Hydrology and Earth System Sciences*, *21*(7), 3427–3440.

Clarke, R. (1994), *Statistical modelling in hydrology*, Wiley, Chichester.

Cloke, H. L., and F. Pappenberger (2008), Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures, *Meteorological Applications*, *15*(1), 181–197.

Collins, M., R. E. Chandler, P. M. Cox, J. M. Huthnance, J. Rougier, and D. B. Stephenson (2012), Quantifying future climate change, *Nature Climate Change*, *2*(6), 403–409.

Conradt, T., C. Gornott, and F. Wechsung (2016), Extending and improving regionalized winter wheat and silage maize yield regression models for Germany: Enhancing the predictive skill by panel definition through cluster analysis, *Agricultural and Forest Meteorology*.

Cook, B. I., T. R. Ault, and J. E. Smerdon (2015), Unprecedented 21st century drought risk in the American Southwest and Central Plains, *Science Advances*, *1*(1), e1400,082–e1400,082.

Cook, R. D. (1977), Detection of Influential Observation in Linear Regression, *Technometrics*, *19*(1), 15–18.

Cook, R. D. (1979), Influential Observations in Linear Regression, *Journal of the American Statistical Association*, *74*(365), 169–174.

COPA-COGECA (2003), Assessment of the impact of the heat wave and drought of the summer 2003 on agriculture and forestry, *Report*, Committee of Professional Agricultural Organisations in the European Union. General Confederation of Agricultural Co-operatives in the European Union (COPA-COGECA), Brussels.

Corbari, C., and M. Mancini (2014), Calibration and Validation of a Distributed Energy-Water Balance Model Using Satellite Data of Land Surface Temperature and Ground Discharge Measurements, *Journal of Hydrometeorology*, *15*(1), 376–392.

Corbari, C., J. a. Sobrino, M. Mancini, and V. Hidalgo (2010), Land surface temperature representativeness in a heterogeneous area through a distributed energy-water balance model and remote sensing data, *Hydrology and Earth System Sciences*, *14*(10), 2141–2151.

Corbari, C., M. Mancini, J. Li, and Z. Su (2015), Can satellite land surface temperature data be used similarly to river discharge measurements for distributed hydrological model calibration?, *Hydrological Sciences Journal*, *60*(2), 202–217.

Cosby, B. J., G. M. Hornberger, R. B. Clapp, and T. R. Ginn (1984), A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, *Water Resources Research*, *20*(6), 682–690.

Crawford, N. H., and R. K. Linsley (1966), Digital simulation in hydrology: Stanford watershed model iv, *Tech. Rep. 39*, Stanford Univ. Dept. of Civil Engineering.

Cretat, J., and B. Pohl (2012), How Physical Parameterizations Can Modulate Internal Variability in a Regional Climate Model, *Journal of Atmospheric Sciences*, *69*(2), 714–724.

Cristea, N. C., S. K. Kampf, and S. J. Burges (2012), Revised coefficients for Priestley-Taylor and Makkink-Hansen equations for estimating daily reference evapotranspiration, *J. Hydraul. Eng. ASCE*, *18*(10), 1289–1300.

Croissant, Y., and G. Millo (2008), Panel data econometrics in R: The plm package, *Journal of Statistical Software*, *27*(2).

Crow, W. T., E. Wood, and M. Pan (2003), Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals, *Journal of Geophysical Research*, *108*(D23), 4725.

Cuntz, M., J. Mai, M. Zink, S. Thober, R. Kumar, D. Schäfer, M. Schrön, J. Craven, O. Rakovec, D. Spieler, V. Prykhodko, G. Dalmasso, J. Musuuza, B. Langenberg, S. Attinger, and L. Samaniego (2015), Computationally inexpensive identification of noninformative model parameters by sequential screening, *Water Resources Research*, *51*(8), 6417–6441.

Cuntz, M., J. Mai, L. Samaniego, M. Clark, V. Wulfmeyer, O. Branch, S. Attinger, and S. Thober (2016), The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, *Journal of Geophysical Research-Atmospheres*, *121*(18), 10,676–10,700.

Dagan, G. (1989), *Flow and transport in porous media*, Springer Verlag, New York.

Dai, A. (2011), Drought under global warming: a review, *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), 45–65.

Dai, A. (2013), Increasing drought under global warming in observations and models, *Nature Climate Change*, *3*(1), 52–58.

Dai, A., K. Trenberth, and T. Qian (2004), A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming, *J. Hydrometeor.*, *5*(6), 1117–1130.

Daniell, J., F. Wenzel, A. McLennan, K. Daniell, T. Kunz-Plapp, B. Khazai, A. Schaefer, M. Kunz, and T. Girard (2016), The global role of natural disaster fatalities in decision-making: statistics, trends and analysis from 116 years of disaster data compared to fatality rates from other causes, in *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, pp. EPSC2016–2021.

Davison, A. C., and D. V. Hinkley (1997), *Bootstrap Methods and Their Applications*, Cambridge University Press, Cambridge, iSBN 0-521-57391-2.

Dawdy, D. R., and R. W. Lichty (1968), Methodology of hydrologic mode building - digital computers in hydrology, *Int. Ass. Sci. Hydrol. Publ.*, *81*, 123–137.

Day, G. (1985), Extended Streamflow Forecasting Using NWSRFS, *Journal of Water Resources Planning and Management*, *111*(2), 157–170.

de Bruyn, L. P., and J. M. de Jager (1978), A meteorological approach to the identification of drought sensitive periods in field crops, *Agricultural Meteorology*, *19*(1), 35–40.

De Roo, A., and C. G. Wesseling (2000), Physically based river basin modelling within a GIS: the LISFLOOD model, *Hydrological Processes*, *14*(1112), 1981–1992.

De Roo, A. P. J., C. G. Wesseling, and W. P. A. Van Deursen (2000), Physically based river basin modelling within a GIS: the LISFLOOD model, *Hydrological Processes*, *14*(11-12), 1981–1992.

Demargne, J., L. Wu, S. K. Regonda, J. D. Brown, H. Lee, M. He, D.-J. Seo, R. Hartman, H. D. Herr, M. Fresch, J. Schaake, and Y. Zhu (2014), The Science of NOAA's Operational Hydrologic Ensemble Forecast Service, *Bulletin of the American Meteorological Society*, *95*(1), 79–98.

Dembélé, M., M. Hrachowitz, H. H. G. Savenije, G. Mariéthoz, and B. Schaefli (2020), Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, *Water Resources Research*, *56*(1), 490–26.

Demuth, S., and B. Heinrich (1997), Temporal and spatial behaviour of drought in south germany, *Friend97 Regional Hydrology Concepts and Models for Sustainable Water Resource Management*, *246*, 151–157.

Deschenes, O., and M. Greenstone (2007), The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather, *The American Economic Review*, *97*(1), 354–385.

Deutscher Wetterdienst (2017), Climate Data Center.

Deutscher Wetterdienst (DWD) (2011), Global irradiance data.

Deutscher Wetterdienst (DWD) (2015), Climate station data.

Diamond, J. M. (2011), *Collapse: How Societies Choose to Fail or Survive*, London: Penguin.

Dickinson, R. (1984), Modelling evapotranspiration for three-dimensional global climate models, in *Climate Processes and Climate Sensitivity*, *Geophysical Monograph Series*, vol. 29, edited by J. E. Hansen and T. Takahashi, pp. 58–72, AGU.

Die Landwirtschaft Band 1 (2014), *Landwirtschaftlicher Pflanzenbau*, 1040 pp., BLV/LVH.

Dingman, S. (2004), *Physical Hydrology*, Prentice-Hall Inc.

Dirmeyer, P., P.-L. C., and G. Balsamo (2015), Land-atmosphere interactions and the water cycle, in *Seamless prediction of the Earth system: from minutes to months*, edited by M. Bëland and A. Thorpe, 1156, chap. 8, p. 471, World Meteorological Organization, Geneva.

Dirmeyer, P. A., Z. Guo, and X. Gao (2004), Comparison, validation, and transferability of eight multiyear global soil wetness products, *J. Hydrometeor*, *5*(6), 1011–1033.

Dixon, B. L., S. E. Hollinger, P. Garcia, and V. Tirupattur (1994), Estimating Corn Yield Response Models to Predict Impacts of Climate Change, *Journal of Agricultural and Resource Economics*, *19*(1), 58–68.

Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, *270*(1–2), 105 – 134.

Donnelly, C., J. C. M. Andersson, and B. Arheimer (2015), Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, *Hydrological Sciences Journal*, *61*(2), 255–273.

Dooge, J. C. (1959), Quantitative Hydrology in the 17th century, *La Houille Blanche*, *nn*(6), 799–807.

Dooge, J. C. (1982), Parameterization of hydrologic processes, in *Land surface processes in atmospheric general circulation models. Proceedings of the World Climate Research Programme study conference held in Greenbelt, Maryland.*, edited by P. Eagleson, pp. 243–288, Cambridge University Press, new York, N.Y.

Dooge, J. C. (1986), Looking for hydrologic laws, *Water Resour. Res.*, *22*.

Dooge, J. C. (2001), Concepts of the hydrological cycle. Ancient and Modern, in *International Symposium OH Origins and History of Hydrology*, pp. 9–10, Dijon.

Dorigo, W., A. Gruber, R. D. Jeu, W. Wagner, T. Stacke, A. Loew, C. Albergel, L. Brocca, D. Chung, R. Parinussa, and R. Kidd (2014), Evaluation of the ESA CCI soil moisture product using ground-based observations, *Remote Sens. Environ.*

Döring, S., J. Döring, H. Borg, and F. Böttcher (2011), Vergleich von trockenheitsindizes zur nutzung in der landwirtschaft unter den klimatischen bedingungen mitteldeutschlands, *Hercynia N. F.*, *44*, 145–168.

Dracup, J., K. Lee, and E. Paulson (1980), On the definition of droughts, *Water Resour. Res.*, *16*(2), 297–302.

Dracup, J. A. (1991), Drought monitoring, *Stochastic Hydrology and Hydraulics*, *10*(3), 111–266.

Dracup, J. A., and E. Kahya (1994), The relationships between U.S. streamflow and La Niña events, *Water Resources Research*, *30*(7), 2133–2141.

Draper, N. R., and H. Smith (1981), *Applied regression analysis*, 2nd ed., Wiley, New York.

Driscoll, J. C., and A. C. Kraay (1998), Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data, *Economic Systems Research*, *10*(4), 307–324.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.

Duan, Q., V. Gupta, and S. Sorooshian (1993), Shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*(3), 501–521.

Duckstein, L. (1984), Multiobjective Optimization in Structural Design: The Model Choice Problem, in *New directions in optimum structural design*, edited by E. Atrek, R. H. Gallagher, K. M. Ragsdell, and O. C. Zienkiewicz, pp. 459–481, John Wiley, New York.

Duckstein, L., and S. Opricovic (1980), Multiobjective optimization in river basin development, *Water Resources Research*, *16*(1), 14–20.

Dutra, E., L. Magnusson, F. Wetterhall, H. L. Cloke, G. Balsamo, S. Boussetta, and F. Pappenberger (2013), The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products, *International Journal of Climatology*, *33*(7), 1720–1729.

Dyck, S., and G. Peschke (1995), *Grundlagen der Hydrologie*, 3 ed., 536 pp., Verlag für Bauwesen, Berlin.

Eagleson, P. S. (1986), The emergence of global-scale hydrology, *Water Resources Research*, *22*(9 S), 6S–14S.

ECMWF (2016), Ifs documentation – cy41r2 operational implementation 8 march 2016, *Tech. rep.*, European Centre for Medium-Range Weather Forecasts, Access date: 2017/02/02.

EDgE (2017), End-to-end Demonstrator for improved decision making in the water sector in Europe.

Edijatno, N. de Oliveira Nascimento, X. Yang, Z. Makhlouf, and C. Michel (1999), GR3J: a daily watershed model with three free parameters, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *44*(2), 263–277.

Edwards, P. N. (2010), *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, The MIP Press.

EEA (2012), Climate change, impacts and vulnerability in Europe 2012: an indicator-based report., *Tech. Rep. 12*, European Environment Agency, Copenhagen, Denmark.

Efron, B. (1982), The Jackknife, the Bootstrap and Other Resampling Plans, *Society for Industrial and Applied Mathematics*, *VII*, regional conference series in applied mathematics; 38.

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert (2012), HESS Opinions "Should we apply bias correction to global and regional climate model data?", *Hydrol Earth Syst Sci*, *16*, 3391–3404.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *108*(D22).

Emerton, R. E., E. M. Stephens, F. Pappenberger, T. C. Pagano, A. H. Weerts, A. W. Wood, P. Salamon, J. D. Brown, N. Hjerdt, C. Donnelly, C. A. Baugh, and H. L. Cloke (2016), Continental and global scale flood forecasting systems, *Wiley Interdisciplinary Reviews: Water*, *3*(3), 391–418.

Entin, J., A. Robock, K. Vinnikov, S. Hollinger, S. Liu, and A. Namkhai (2000), Temporal and spatial scales of observed soil moisture variations in the extratropics, *J. Geophys. Res.-Atmos.*, *105*, 11,865–11,877.

ESA (2015), Soil Moisture CCI, Accessed 1 July 2015. [Available online at http://www.esa-soilmoisture-cci.org/node/136.].

EUMETSAT (2016), Land Surface Analysis Satellite Applications Facility (LSA SAF).

European Commission (2007), Water Scarcity and Droughts - In-Depth Assessment, *Tech. rep.*, European Commission.

European Commission (2010), Water Scarcity and Drought in the European Union.

European Commission (2012), Final Report Gap Analysis of the Water Scarcity and Droughts Policy in the EU European Commission, *Tech. rep.*, European Commission.

European Environmental Agency (2009), CORINE Land Cover 1990, 2000 and 2006.

European Environmental Agency (EEA) (2009), CORINE Land Cover 1990, 2000 and 2006.

European Water Archive (EWA) (2011), Runoff data.

Euser, T., H. C. Winsemius, M. Hrachowitz, F. Fenicia, S. Uhlenbrook, and H. H. G. Savenije (2013), A framework to assess the realism of model structures using hydrological signatures, *Hydrol. Earth Syst. Sci.*, *17*(5), 1893–1912.

Evans, N. P., T. K. Bauska, F. Gázquez-Sánchez, M. Brenner, J. H. Curtis, and D. A. Hodell (2018), Quantification of drought during the collapse of the classic Maya civilization, *Science*, *361*(6401), 498–501.

Fageria, N. K., V. C. Baligar, and R. B. Clark (2006), *Physiology of crop production*, CRC Press.

Famiglietti, J., and E. Wood (1994), Multiscale modeling of spatially variable water and energy balance processes, *Water Resources Research*, *30*(11), 3061–3078.

Famiglietti, J. S., and E. F. Wood (1995), Effects of Spatial Variability and Scale on Areally Averaged Evapotranspiration, *Water Resources Research*, *31*(3), 699–712.

FAO Water (2016), Crop Water Information: Maize.

FAO/IIASA/ISRIC/ISSCAS/JRC (2012), Harmonized World Soil Database (version 1.2), FAO, Rome, Italy and IIASA, Laxenburg, Austria, *Tech. rep.*, FAO/IIASA/ISRIC/ISSCAS/JRC.

Favre, A., S. E. Adlouni, L. Perreault, N. Thiémonge, and B. Bobée (2004), Multivariate hydrological frequency analysis using copulas, *Water Resources Research*, *40*, W01101.

Feddes, R. A., P. Kowalik, K. Kolinska-Malinka, and H. Zaradny (1976), Simulation of field water uptake by plants using a soil water dependent root extraction function, *Journal of Hydrology*, *31*(1-2), 13–26.

Federal Agency for Cartography and Geodesy (BKG) (2010), Digital Elevation Model (DEM).

Federal Institute for Geosciences and Natural Resources (BGR) (1998), Digital soil map of Germany 1:1,000,000 (BUEK 1000).

Federal Institute for Geosciences and Natural Resources (BGR) (2009), Hydrogeological map of Germany 1:200,000 (HUEK 200).

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2008), Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, *44*, W01402.

Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011), Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, *Water Resources Research*, *47*(11), n/a–n/a.

Fernandez, W., R. Vogel, and A. Sankarasubramanian (2000), Regional calibration of a watershed model, *Hydrolog. Sci. J.*, *45*(5), 689–707.

Field, J. O. (2000), *Drought: A Global Assessment*, chap. Drought, the Famine Process, and the phasing of interventions, Routledge, ed. Wilhite, D. A.

Fink, A. H., T. Br ucher, A. Kr uger, G. C. Leckebusch, J. Pinto, and U. Ulbrich (2004), The 2003 european summer heatwaves and drought - synoptic diagnosis and impacts, *Weather*, *59*, 209–216.

Fisher, A. C., M. W. Hanemann, M. J. Roberts, and W. Schlenker (2012), The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment, *The American Economic Review*, *102*(7), 3749–3760.

Fisher, J. B., D. N. Huntzinger, C. R. Schwalm, and S. Sitch (2014), Modeling the Terrestrial Biosphere, *Annual Review of Environment and Resources*, *39*(1), 91–123.

Fishman, R. (2016), More uneven distributions overturn benefits of higher precipitation for crop yields, *Environmental Research Letters*, *11*(2), 024,004.

Flügel, W.-A. (1995), Delineating Hydrological Respons Units by Geographical Information System Analysis for Regional Hydrological Modelling using PRMS/MMS in the Drainage Basin of the River Bröl, Germany, *Hydrological Processes*, *9*(January 1994), 423–436.

Forman, B. A., R. H. Reichle, and M. Rodell (2012), Assimilation of terrestrial water storage from GRACE in a snow-dominated basin, *Water Resour. Res.*, *48*(1), 1–14.

Franke, J., V. Goldberg, U. Eichelmann, E. Freydank, and C. Bernhofer (2004), Statistical analysis of regional climate trends in saxony, germany, *Climate Res.*, *27*(1), 145–150.

Freeze, R., and R. Harlan (1969), Blueprint for a physically-based, digitally-simulated hydrologic response model, *Journal of Hydrology*, *9*(3), 237–258.

Freeze, R. A. (1974), Streamflow generation, *Reviews of Geophysics*, *12*(4), 627–647.

Gelhar, L. W. (1993), *Stochastic Subsurface Hydrology*, Prentice Hall.

Gentine, P., T. J. Troy, B. R. Lintner, and K. L. Findell (2012), Scaling in surface hydrology: progress and challenges, *Journal of Contemporary Water research & education*, *147*(1), 28–40.

Gentleman, W. M. (1974), Basic procedures for large, sparse or weighted linear least squares problems, *Applied Statistics*, *23*, 448–454.

Giuntoli, I., B. Renard, J.-P. Vidal, and A. Bard (2013), Low flows in France and their relationship to large-scale climate indices, *J. Hydrol.*, *482*, 105–118.

Giuntoli, I., J. P. Vidal, C. Prudhomme, and D. M. Hannah (2015), Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models, *Earth System Dynamics*, *6*(1), 267–285.

Global Runoff Data Centre (2017), Global Runoff Database.

Göckede, M., T. Foken, M. Aubinet, M. Aurela, J. Banza, C. Bernhofer, J. M. Bonnefond, Y. Brunet, A. Carrara, R. Clement, E. Dellwik, J. Elbers, W. Eugster, J. Fuhrer, A. Granier, T. Grünwald, B. Heinesch, I. A. Janssens, A. Knohl, R. Koeble, T. Laurila, B. Longdoz, G. Manca, M. Marek, T. Markkanen, J. Mateus, G. Matteucci, M. Mauder, M. Migliavacca, S. Minerbi, J. Moncrieff, L. Montagnani, E. Moors, J.-M. Ourcival, D. Papale, J. Pereira, K. Pilegaard, G. Pita, S. Rambal, C. Rebmann, A. Rodrigues, E. Rotenberg, M. J. Sanz, P. Sedlak, G. Seufert, L. Siebicke, J. F. Soussana, R. Valentini, T. Vesala, H. Verbeeck, and D. Yakir (2008), Quality control of CarboEurope flux data – Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems, *Biogeosciences*, *5*(2), 433–450.

Godfree, R. C., N. Knerr, D. Godfree, J. Busby, B. Robertson, and F. Encinas-Viso (2019), Historical reconstruction unveils the risk of mass mortality and ecosystem collapse during pancontinental megadrought, *Proceedings of the National Academy of Sciences of the United States of America*, *116*(31), 15,580–15,589.

Goehler, M., J. Mai, and M. Cuntz (2013), Use of eigendecomposition in a parameter sensitivity analysis of the Community Land Model, *Journal of Geophysical Research-Biogeosciences*, *118*(2), 904–921.

Gornott, C., and F. Wechsung (2015), Niveauneutrale Modellierung der Ertragsvolatilität von Winterweizen und Silomais auf mehreren räumlichen Ebenen in Deutschland, *Journal für Kulturpflanzen*, *65*(June), 248–254.

Gornott, C., and F. Wechsung (2016), Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany, *Agricultural and Forest Meteorology*, *217*, 89–100.

Gotzinger, J., and A. Bárdossy (2007), Comparison of four regionalisation methods for a distributed hydrological model, *J. Hydrol.*, *333*(2-4), 374–384.

Grant, R. F., B. S. Jackson, J. R. Kiniry, and G. F. Arkin (1989), Water Deficit Timing Effects on Yield Components in Maize, *Agronomy Journal*, *81*(1), 61–65.

Grayson, R., and G. Blöschl (Eds.) (2000), *Spatial patterns in catchment hydrology—observations and modelling*, Cambridge University Press, 423 pp, ISBN 0-521-63316-8.

Greve, P., L. Gudmundsson, and S. I. Seneviratne (2017), Regional scaling of annual mean precipitation and water availability with global temperature change, *Earth System Dynamics Discussions*, pp. 1–24.

Gudmundsson, L., L. M. Tallaksen, K. Stahl, D. B. Clark, E. Dumont, S. Hagemann, N. Bertrand, D. Gerten, J. Heinke, N. Hanasaki, F. Voss, and S. Koirala (2012), Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, *Journal of Hydrometeorology*, *13*(2), 604–620.

Guha-Sapir, D., R. Below, and P. Hoyois (2015), EM-DAT: The OFDA/CRED International Disaster Database.

Guiot, J., and W. Cramer (2016), Climate change: The 2015 Paris Agreement thresholds and Mediterranean basin ecosystems, *Science*, *354*(6311), 465–468.

Gupta, H. V., S. Sorooshian, T. S. Hogue, and D. P. Boyle (2002), Advances in automatic calibration of watershed models, in *Calibration of Watershed Models*, *Water Science and Application*, vol. 6, edited by Q. Duan, H. Gupta, S. Sorooshian, A. Rousseau, and R. Turcotte, pp. 9–28, AGU.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1–2), 80–91.

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, *48*(8), 1–16.

Gupta, H. V., C. Perrin, G. Blöschl, A. Montanari, R. Kumar, M. Clark, and V. Andréassian (2014), Large-sample hydrology: a need to balance depth with breadth, *Hydrology and Earth System Sciences*, *18*(2), 463–477.

Gutmann, E. D., and E. E. Small (2010), A method for the determination of the hydraulic properties of soil from MODIS surface temperature for use in land-surface models, *Water Resources Research*, *46*(6), W06,520.

Haddeland, I., B. V. Matheussen, and D. P. Lettenmaier (2002), Influence of spatial resolution on simulated streamflow in a macroscale hydrologic model, *Water Resour. Res.*, *38*(7).

Haddeland, I., D. B. Clark, W. Franssen, F. Ludwig, F. Voss, N. W. Arnell, N. Bertrand, M. Best, S. Folwell, D. Gerten, S. Gomes, S. N. Gosling, S. Hagemann, N. Hanasaki, R. Harding, J. Heinke, P. Kabat, S. Koirala, T. Oki, J. Polcher, T. Stacke, P. Viterbo, G. P. Weedon, and P. Yeh (2011), Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, *Journal of Hydrometeorology*, *12*(5), 869–884.

Hao, Z., A. AghaKouchak, N. Nakhjiri, and A. Farahmand (2014), Global integrated drought monitoring and prediction system, *Scientific Data*, *1*, 1–10.

Hargreaves, G., and Z. Samani (1982), Estimating potential evapotranspiration, *Journal of the Irrigation & Drainage Division - ASCE*, *108*(IR3), 225–230.

Hargreaves, G. H., and Z. A. Samani (1985), Reference crop evapotranspiration from temperature, *Applied Engineering in Agriculture*, *1*(2), 96–99.

Haug, G. H., D. Günther, L. C. Peterson, D. M. Sigman, K. A. Hughen, and B. Aeschlimann (2003), Climate and the Collapse of Maya Civilization, *Science*, *299*(5613), 1731–1735.

Haughton, N., G. Abramowitz, A. J. Pitman, D. Or, M. J. Best, H. R. Johnson, G. Balsamo, A. Boone, M. Cuntz, B. Decharme, P. A. Dirmeyer, J. Dong, M. Ek, Z. Guo, V. Haverd, B. J. J. van den Hurk, G. S. Nearing, B. Pak, J. A. Santanello Jr., L. E. Stevens, and N. Vuichard (2016), The Plumbing of Land Surface Models: Is Poor Performance a Result of Methodology or Data Quality?, *Journal of Hydrometeorology*, *17*(6), 1705–1723.

Haverd, V., M. Cuntz, R. Leuning, and H. Keith (2007), Air and biomass heat storage fluxes in a forest canopy: Calculation within a soil vegetation atmosphere transfer model, *Agricultural and Forest Meteorology*, *147*(3-4), 125–139.

Hawkins, E., P. Ortega, E. Suckling, A. Schurer, G. Hegerl, P. Jones, M. Joshi, T. J. Osborn, V. Masson-Delmotte, J. Mignot, P. Thorne, and G. J. van Oldenborgh (2017), Estimating Changes in Global Temperature since the Preindustrial Period, *Bulletin of the American Meteorological Society*, *98*(9), 1841–1856.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New (2008), A european daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006, *Journal of Geophysical Research: Atmospheres*, *113*(D20).

Heathcote, R. L. (2016), *Drought and the Human Story: Braving the Bull of Heaven*, 1 edition ed., Routledge, London.

Held, I. M., and B. J. Soden (2006), Robust responses of the hydrological cycle to global warming, *Journal of Climate*, *19*(21), 5686–5699.

Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek (2013), A trend-preserving bias correction - the ISI-MIP approach, *Earth System Dynamics*, *4*(2), 219–236.

Hengl, T., J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen (2017), Soilgrids250m: Global gridded soil information based on machine learning, *PLOS ONE*, *12*(2), 1–40.

Hess, P., and H. Brezowsky (1969), Katalog der Großwetterlagen Europas. Berichte des Deutschen Wetterdienst, *Tech. Rep. 113(15)*, Selbstverlag des Deutschen Wetterdienst, Offenbach a. Main, 2 neu bearbeitete und ergänzte Aufl.

Heuvelmans, G., B. Muys, and J. Feyen (2006), Regionalization of the parameters of a hydrologic model: comparison of linear regression models with artificial neural nets, *J. Hydrol.*, *319*, 245–265.

Hijmans, R. J. (2016), Geographic Data Analysis and Modeling [R package raster version 2.5-8].

Hirschi, M., S. I. Seneviratne, V. Alexandrov, F. Boberg, C. Boroneant, O. B. Christensen, H. Formayer, B. Orlowsky, and P. Stepanek (2010), Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, *Nature Geoscience*, *4*(1), 17–21.

Hlavac, M. (2015), stargazer: Well-Formatted Regression and Summary Statistics Tables [R package version 5.2].

Hodell, D. A., J. Curtis, and M. Brenner (1995), Possible role of climate in the collapse of Classic Maya, *Nature*, *375*(6530), 391–394.

Hoeg-Guldberg, O., J. D, M. Taylor, B. M, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K. L. Ebi, F. Engelbrecht, J. Guiot, Y. Hijioka, S. Mehrotra, A. Payne, S. I. Seneviratne, A. Thomas, R. Warren, and G. Zhou (2019), Impacts of 1.5°C Global Warming on Natural and Human Systems , in *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, edited by V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Peean, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield, pp. 1–138, World Meteorological Organization.

Hofstra, N., M. Haylock, M. New, and P. D. Jones (2009), Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res. Atmos.*, *114*(D21).

Horion, S., A. Singleton, P. Barbosa, and J. Vogt (2012), JRC experience on the development of drought information systems, *Tech. rep.*, Joint Research Centre, technical Report.

Horton, R. (1935), *Surface Runoff Phenomena. Part 1. Analysis of the Hydrograph.*, Horton Hydrologic Laboratory Publication 101. Edward Bros: Ann Arbor, MI.

Houghton, J. T., Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. Johnson (Eds.) (2001), *Climate Change 2001: The Scientific Basis*, Cambridge University Press.

Hrachowitz, M., H. Savenije, G. Blöschl, J. McDonnell, M. Sivapalan, J. Pomeroy, B. Arheimer, T. Blume, M. Clark, U. Ehret, F. Fenicia, J. Freer, A. Gelfan, H. Gupta, D. Hughes, R. Hut, A. Montanari, S. Pande, D. Tetzlaff, P. Troch, S. Uhlenbrook, T. Wagener, H. Winsemius, R. Woods, E. Zehe, and C. Cudennec (2013), A decade of Predictions in Ungauged Basins (PUB)-A Review, *Hydrological Sciences Journal*, *58*(6), 1198–1255.

Hsiang, S. M. (2016), Climate Econometrics, *Annual Review of Resource Economics*.

Hsiang, S. M., M. Burke, and E. Miguel (2013), Quantifying the influence of climate on human conflict, *Science (New York, N.Y.)*, *341*(6151), 1235,367.

Huang, J., H. Yu, X. Guan, G. Wang, and R. Guo (2015), Accelerated dryland expansion under climate change, *Nature Climate Change*, *316*(2), 847–171.

Huang, S., R. Kumar, M. Flörke, T. Yang, Y. Hundecha, P. Kraft, C. Gao, A. Gelfan, S. Liersch, A. Lobanova, M. Strauch, F. Ogtrop, J. Reinhardt, U. Haberlandt, and V. Krysanova (2016), Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide, *Climatic Change*, pp. 1–17.

Hundecha, Y., and A. Bárdossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *Journal of Hydrology*, *292*(1-4), 281–295.

Hundecha, Y., B. Arheimer, C. Donnelly, and I. Pechlivanidis (2016), A regional parameter estimation scheme for a pan-European multi-basin model, *Journal of Hydrology: Regional Studies*, *6*, 90–111.

Hurst, H. E. (1951), Long-term storage of reservoirs: an experimental study, *Transactions of the American Society of Civil Engineers*, *116*, 770–799.

ICID (2015), Agricultural Water Management for Sustainable Rural Development: Annual Report, *Tech. rep.*, INTERNATIONAL COMMISSION ON IRRIGATION AND DRAINAGE, New Delhi, India.

Intsiful, J., and H. Kunstmann (2008), Upscaling of Land-Surface Parameters Through Inverse Stochastic SVAT-Modelling, *Boundary-Layer Meteorology*, *129*(1), 137–158.

IPCC (2007), Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change, (eds Solomon, S. et al.). Cambridge Univ. Press.

Jacob, T., J. Wahr, W. T. Pfeffer, and S. Swenson (2012), Recent contributions of glaciers and ice caps to sea level rise, *Nature*, *482*(7386), 514–518.

James, R., R. Washington, C.-F. Schleussner, J. Rogelj, and D. Conway (2017), Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets, *Wiley Interdisciplinary Reviews-Climate Change*, *8*(2), 1–23.

Jarvis, A., H. Reuter, A. Nelson, and E. Guevara (2008), Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database, *Tech. rep.*, CGIAR-CSI.

Jung, M., M. Reichstein, H. A. Margolis, A. Cescatti, A. D. Richardson, M. A. Arain, A. Arneth, C. Bernhofer, D. Bonal, J. Chen, D. Gianelle, N. Gobron, G. Kiely, W. Kutsch, G. Lasslop, B. E. Law, A. Lindroth, L. Merbold, L. Montagnani, E. J. Moors, D. Papale, M. Sottocornola, F. Vaccari, and C. Williams (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.: Biogeosci.*, *116*(G3), g00J07.

Kauffeldt, A., F. Wetterhall, F. Pappenberger, P. Salamon, and J. Thielen (2016), Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level, *Environ. Modell. Softw.*, *75*(c), 68–76.

Kavetski, D., G. Kuczera, and S. W. Franks (2003), Semidistributed hydrological modeling: A “saturation path” perspective on TOPMODEL and VIC, *Water Resources Research*, *39*(9), n/a–n/a.

Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*, W03407.

Keller, J. (2010), Ernteversicherungen als Risikomanagementinstrument - Eine Analyse von Versicherungstypen und Tarifierungsmodellen -, Ph.D. thesis, Justus-LIebig-Universität Giessen.

Kessomkiat, W., H.-J. H. Franssen, A. Graf, and H. Vereecken (2013), Estimating random errors of eddy covariance data: An extended two-tower approach, *Agricult. Forest Meterol.*, *171*, 203–219.

Keyantash, J., and J. A. Dracup (2002), The quantification of drought: An evaluation of drought indices, *Bull. Am. Meteorol. Soc.*, *83*(8), 1167–1180.

Kim, U., and J. Kaluarachchi (2008), Application of parameter estimation and regionalization methodologies to ungauged basins of the upper blue nile river basin, ethiopia, *J. Hydrol.*, *362*, 39–56.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, *42*(3), 1–16.

Kirkby, M. (1988), Hillslope runoff processes and models, *Journal of Hydrology*, *100*(1-3), 315–339.

Kirtman, B. P., D. Min, J. M. Infanti, J. L. Kinter, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, M. Rienecker, M. Suarez, Z. E. Li, J. Marshak, Y.-K. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis, and E. F. Wood (2014), The North American Multi-Model Ensemble (NMME): Phase-1 Seasonal to Interannual Prediction, Phase-2 Toward Developing Intra-Seasonal Prediction, *Bulletin of the American Meteorological Society*.

Kitanidis, P. K. (1997), *Introduction to Geostatistics*, Applications to Hydrogeology, Cambridge University Press.

Kitanidis, P. K., and E. G. Vomvoris (2010), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resources Research*, *19*(3), 677–690.

Klemeš, V. (1986), Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, *31*(1), 13–24.

Koch, J., K. H. Jensen, and S. Stisen (2015), Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study, *Water Resources Research*, *51*(2), 1225–1246.

Köhli, M., M. Schrön, M. Zreda, U. Schmidt, P. Dietrich, and S. Zacharias (2015), Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons, *Water Resour. Res.. Accepted Author Manuscript*.

Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker (1999), A parameterization of snowpack and frozen ground intended for ncep weather and climate models, *J. Geophys. Res.*, *104*(D16), 19,569–19,585.

Koren, V., S. Reed, M. Smith, Z. Zhang, and D. J. Seo (2004), Hydrology Laboratory Research Modeling System (HL-RMS) of the U.S. National Weather Service, *J. Hydrol.*, *291*(3-4), 297–318.

Koren, V., M. Smith, and Q. Duan (2013), *Use of a Priori Parameter Estimates in the Derivation of Spatially Consistent Parameter Sets of Rainfall-Runoff Models*, pp. 239–254, American Geophysical Union.

Koster, R. D., M. J. Suarez, P. Liu, U. Jambor, A. Berg, M. Kistler, R. Reichle, M. Rodell, and J. Famiglietti (2004), Realistic initialization of land surface states: Impacts on subseasonal forecast skill, *J. Hydrometeor*, *5*(6), 1049–1063.

Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma (2009), On the Nature of Soil Moisture in Land Surface Models, *Journal of Climate*, *22*(16), 4322–4335.

Kovats, R., R. Valentini, L. Bouwer, E. Georgopoulou, D. Jacob, E. Martin, M. Rounsevell, and J.-F. Soussana (2011), Europe, in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by V. Barros, C. Field, D. Dokken, M. Mastrandrea, K. Mach, T. Bilir, M. Chatterjee, K. Ebi, Y. Estrada, R. Genova, B. Girma, E. Kissel, A. Levy, S. MacCracken, P. Mastrandrea, and L. White, pp. 1267–1326, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kowalczyk, E. A., Y. P. Wang, and R. M. Law (2006), The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, *CSIRO. Marine and Atmospheric Research*, *13*.

Krysanova, V., and F. Hattermann (2016), Overview of applied models and summary of results on intercomparison of climate impact assessment, *Climatic Change*, *x*(x), x, submitted to this Special Issue.

Kuichling, E. (1889), The relation between the rainfall and the discharge of sewers in populous districts, *Transactions, American Society of Civil Engineers*, *20*, 1–56.

Kumar, R., L. Samaniego, and S. Attinger (2010), The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, *Journal of Hydrology*, *392*(1-2), 54–69.

Kumar, R., L. Samaniego, and S. Attinger (2013), Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, *Water Resources Research*, *49*(1), 360–379.

Kumar, R., B. Livneh, and L. Samaniego (2013b), Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, *Water Resources Research*, *49*(9), 5700–5714.

Kumar, R., J. Mai, O. Rakovec, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schaefer, M. Schrön, and L. E. Samaniego (2015), Regionalized Hydrologic Parameters Estimates for a Seamless Prediction of Continental scale Water Fluxes and States, *AGU Fall Meeting Abstracts*.

Kumar, R., J. L. Musuuza, A. F. Van Loon, A. J. Teuling, R. Barthel, J. Ten Broek, J. Mai, L. Samaniego, and S. Attinger (2016), Multiscale evaluation of the standardized precipitation index as a groundwater drought indicator, *Hydrol. Earth Syst. Sci.*, *20*(3), 1117–1131.

Kunreuther, H. C., E. O. Michel-Kerjan, N. A. Doherty, M. F. Grace, R. W. Klein, and M. V. Pauly (2009), *At War With the Weather: Managing Large-Scale Risks in a New Era of Catastrophes*, 464 pp., The MIT Press, Cambridge, MA.

Kutilek, M., and D. Nielsen (1994), *Soil hydrology*, Catena Verlag, Cremlingen-Destedt.

Kutsch, W. L., O. Kolle, C. Rebmann, A. Knohl, W. Ziegler, and E.-D. Schulze (2008), Advection and resulting CO2 exchange uncertainty in a tall forest in central Germany, *Ecological Applications*, *18*(6), 1391–1405.

Laio, F., A. Porporato, L. Ridolfi, and I. Rodriguez-Iturbe (2002), On the seasonal dynamics of mean soil moisture, *J. Geophys. Res.*, *107*(D15), 8–1, 8–9.

Lakshmi, V. (2000), A simple surface temperature assimilation scheme for use in land surface models, *Water Resources Research*, *36*(12), 3687.

Landerer, F. W., and S. C. Swenson (2012), Accuracy of scaled grace terrestrial water storage estimates, *Water Resour. Res.*, *48*(4).

Lawrimore, J., R. R. Heim, M. Svoboda, V. Swail, and P. J. Englehart (2002), BEGINNING A NEW ERA OF DROUGHT MONITORING ACROSS NORTH AMERICA, *Bull. Amer. Meteor. Soc.*, *83*(8), 1191–1192.

Le Moine, N., V. Andréassian, C. Perrin, and M. C. (2007), How can rainfall-runoff models handle intercatchment groundwater flows? theoretical study based on 1040 french catchments, *Water Resour. Res.*, *43*, W06428.

Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, and M. Prather (2007), Historical overview of climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and M. H.L., chap. 1, pp. 1–36, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Leavesley, G. H., R. W. Lichty, B. M. Troutman, and L. G. Saindon (1983), *Precipitation-Runoff Modeling System: User's Manual*, U.S. Geological Survey Water-Resources Investigations, Denver, Colorado, 83-4238 ed.

Lehner, F., S. Coats, T. F. Stocker, A. G. Pendergrass, B. M. Sanderson, C. C. Raible, and J. E. Smerdon (2017), Projected drought risk in 1.5°C and 2°C warmer climates, *Geophysical Research Letters*, *44*(14), 7419–7428.

Lerat, J., V. Andréassian, C. Perrin, J. Vaze, J.-M. Perraud, P. Ribstein, and C. Loumagne (2012), Do internal flow measurements improve the calibration of rainfall-runoff models?, *Water Resources Research*, *48*(2).

Li, B., M. Rodell, B. F. Zaitchik, R. H. Reichle, R. D. Koster, and T. M. van Dam (2012a), Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe, *J. Hydrol.*, *446–447*, 103–115.

Li, D., E. Bou-Zeid, M. Barlage, F. Chen, and J. A. Smith (2013a), Development and evaluation of a mosaic approach in the WRF-Noah framework, *Journal of Geophysical Research-Atmospheres*, *118*(21), 11,918–11,935.

Li, H., M. Sivapalan, and F. Tian (2012b), Comparative diagnostic analysis of runoff generation processes in Oklahoma DMIP2 basins: The Blue River and the Illinois River, *Journal of Hydrology*, *418-419*(C), 90–109.

Li, X., and D. Sailor (2000), Application of tree-structured regression for regional precipitation prediction using general circulation model output, *Climate Research*, *16*(1), 17–30.

Li, Z.-L., B.-H. Tang, H. Wu, H. Ren, G. Yan, Z. Wan, I. F. Trigo, and J. Sobrino (2013b), Satellite-derived land surface temperature: Current status and perspectives, *Remote Sensing of Environment*, *131*, 14–37.

Liang, X., D. Lettenmaier, E. Wood, and S. Burges (1994), A Simple Hydrologically Based Model of Land-Surface Water and Energy Fluxes for General-Circulation Models, *Journal of Geophysical Research-Atmospheres*, *99*, 14,415–14,428.

Liang, X., E. F. Wood, and D. P. Lettenmaier (1996a), Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, *Global and Planetary Change*, *13*(1–4), 195 – 206, soil Moisture Simulation.

Liang, X., D. P. Lettenmaier, and E. F. Wood (1996b), One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model, *Journal of Geophysical Research-Atmospheres*, *101*(D16), 21,403–21,422.

Liang, X., J. Guo, and L. R. Leung (2004), Assessment of the effects of spatial resolutions on daily water flux simulations, *J. Hydrol.*, *298*, 287–310.

Lievens, H., A. A. Bitar, N. E. C. Verhoest, F. Cabot, G. J. M. D. Lannoy, M. Drusch, G. Dumedah, H.-J. H. Franssen, Y. Kerr, S. K. Tomer, B. Martens, O. Merlin, M. Pan, M. J. van den Berg, H. Vereecken, J. P. Walker, E. F. Wood, and V. R. N. Pauwels (2015), Optimization of a radiative transfer forward operator for simulating smos brightness temperatures over the upper mississippi basin, *Journal of Hydrometeorology*, *16*(3), 1109–1134.

Lindsey, J. K. (1999), *Applying generalized linear models*, Springer, New York.

Lindstrom, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *Journal of Hydrology*, *201*, 272–288.

Lindström, G., C. Pers, J. Rosberg, J. Strömqvist, and B. Arheimer (2010), Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, *Hydrology research*, *41*(3-4), 295–26.

Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. De Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans (2011), Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, *Hydrol. Earth Syst. Sci.*, *15*(2), 425–436.

Livneh, B., and D. Lettenmaier (2012), Multi-criteria parameter estimation for the unified land model, *Hydrol. Earth Syst. Sci.*, *16*(8), 3029–3048.

Livneh, B., and D. P. Lettenmaier (2013), Regional parameter estimation for the unified land model, *Water Resources Research*, *49*(1), 100–114.

Livneh, B., R. Kumar, and L. Samaniego (2015), Influence of soil textural properties on hydrologic fluxes in the Mississippi river basin, *Hydrological Processes*, *29*(21), 4638–4655.

Lobell, D. B. (2013), Errors in climate datasets and their effects on statistical crop models, *Agricultural and Forest Meteorology*, *170*, 58–66.

Lobell, D. B., M. B. Burke, C. Tebaldi, M. D. Mastrandrea, and W. P. Falcon (2008), Needs for Food Security in 2030 Region, *Science*, *319*(February).

Lobell, D. B., W. Schlenker, and J. Costa-Roberts (2011a), Climate Trends and Global Crop Production Since 1980, *Science*, *333*(6042), 616–620.

Lobell, D. B., M. Bänziger, C. Magorokosho, and B. Vivek (2011b), Nonlinear heat effects on African maize as evidenced by historical yield trials, *Nature Climate Change*, *1*(1), 42–45.

Lobell, D. B., G. L. Hammer, G. McLean, C. Messina, M. J. Roberts, and W. Schlenker (2013), The critical role of extreme heat for maize production in the United States, *Nature Climate Change*, *3*(March), 497–501.

Lorenzo-Lacruz, J., S. Vicente-Serrano, J. López-Moreno, E. Morán-Tejeda, and J. Zabalza (2012), Recent trends in Iberian streamflows (1945–2005), *J. Hydrol.*, *414–415*, 463–475.

Lourenço, T. C., R. Swart, H. Goosen, and R. Street (2015), The rise of demand-driven climate services, *Nature*, *6*(1), 13–14.

Luo, L., and E. F. Wood (2007), Monitoring and predicting the 2007 U.S. drought, *Geophysical Research Letters*, *34*(22).

Luo, L., E. F. Wood, and M. Pan (2007), Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions, *Journal of Geophysical Research: Atmospheres*, *112*(D10), n/a–n/a.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner (2004), European seasonal and annual temperature variability, trends, and extremes since 1500, *Science*, *303*(5663), 1499–1503.

Lynch, P. (2008), The origins of computer weather prediction and climate modeling, *Journal of Computational Physics*, *227*(7), 3431–3444.

Mai, J., B. Tolson, H. Shen, and et al. (2019), The Runoff Model-Intercomparison Project Over Lake Erie and the Great Lakes, *AGU Fall Meeting Abstract N. H32E-03*.

Manabe, S. (1969), Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface, *Mon. Wea. Rev.*, *97*(11), 739–774.

Martina, M. L. V., E. Todini, and Z. Liu (2011), Preserving the dominant physical processes in a lumped hydrological model, *Journal of Hydrology*, *399*(1-2), 121–131.

Marx, A., R. Kumar, S. Thober, O. Rakovec, N. Wanders, M. Zink, E. F. Wood, M. Pan, J. Sheffield, and L. Samaniego (2018), Climate change alters low flows in Europe under global warming of 1.5, 2, and 3°C, *Hydrol. Earth. Syst. Sc.*, *22*(2), 1017–1032.

Mauder, M., T. Foken, R. Clement, J. A. Elbers, W. Eugster, T. Grünwald, B. Heusinkveld, and O. Kolle (2008), Quality control of CarboEurope flux data – Part 2: Inter-comparison of eddy-covariance software, *Biogeosciences*, *5*(2), 451–462.

McCabe, G. J., and M. A. Palecki (2006), Multidecadal climate variability of global lands and oceans, *Int. J. Climatol.*, *26*(7), 849–865.

McCabe, M. F., J. D. Kalma, and S. W. Franks (2005), Spatial and temporal patterns of land surface fluxes from remotely sensed surface temperatures within an uncertainty modelling framework, *Hydrology and Earth System Sciences*, *9*(5), 467–480.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, *21*(2), 239–245.

McKee, T., N. J. Doesken, and J. Kleist (1993), The relationship of drought frequency and duration to time scales, in *in: Proceedings of the 8th Conference of Applied Climatology*, pp. 179–184, American Meterological Society, Anaheim, CA.

McSweeney, C. F., and R. G. Jones (2016), How representative is the spread of climate projections from the 5 CMIP5 GCMs used in ISI-MIP?, *Climate Services*, *1*(C), 24–29.

Mendoza, P. A., M. P. Clark, M. Barlage, B. Rajagopalan, L. Samaniego, G. Abramowitz, and H. Gupta (2015), Are we unnecessarily constraining the agility of complex process-based models?, *Water Resources Research*, *51*(1), 716–728.

Meng, L., and S. M. Quiring (2008), A comparison of soil moisture models using soil climate analysis network observations, *J. Hydrometeor.*, *9*(4), 641–659.

Merz, R., and G. Blöschl (2004), Regionalisation of catchment model parameters, *Journal of Hydrology*, *287*(1-4), 95–123.

Merz, R., J. Parajka, and G. Blöschl (2009), Scale effects in conceptual hydrological modeling, *Water Resour. Res.*, *45*(9).

Metzger, M. J., R. G. H. Bunce, R. H. G. Jongman, C. A. Mücher, and J. W. Watkins (2005), A climatic stratification of the environment of Europe, *Global Ecology and Biogeography*, *14*(6), 549–563.

Miller, E. E., and R. D. Miller (1956), Physical Theory for Capillary Flow Phenomena, *Journal of Applied Physics*, *27*(4), 324–332.

Mishra, A. K., and V. P. Singh (2010), A review of drought concepts, *J. Hydrol.*, *391*(1-2), 202–216.

Mishra, V., K. A. Cherkauer, and S. Shukla (2010), Assessment of Drought due to Historic Climate Variability and Projected Future Climate Change in the Midwestern United States, *J. Hydrometeor.*, *11*(1), 46–68.

Mishra, V., R. Shah, and B. Thrasher (2014), Soil Moisture Droughts under the Retrospective and Projected Climate in India*, *Journal of Hydrometeorology*, *15*(6), 2267–2292.

Mishra, V., H. L. Shah, R. Kumar, L. Samaniego, S. Eisner, and T. Yang (2016), Multimodel Assessment of Sensitivity and Uncertainty of Water Availability under Climate Change, *Climatic Change*, submitted to this Special Issue.

Mizukami, N., M. Clark, A. Newman, A. Wood, E. Gutmann, B. Nijssen, L. Samaniego, and O. Rakovec (2017), Towards seamless large domain parameter estimation for hydrologic models, *Water Resources Research*, submitted.

Mo, K. C. (2011), Drought onset and recovery over the United States, *Journal of Geophysical Research: Atmospheres*, *116*(D20).

Mo, K. C., and D. P. Lettenmaier (2014), Hydrologic prediction over Conterminous U.S. using the National Multi Model ensemble, *Journal of Hydrometeorology*.

Mo, K. C., and B. Lyon (2015), Global meteorological drought prediction using the north american multi-model ensemble, *J. Hydrometeor*.

Mo, K. C., L.-C. Chen, S. Shukla, T. J. Bohn, and D. P. Lettenmaier (2012a), Uncertainties in North American Land Data Assimilation Systems over the Contiguous United States, *J. Hydrometeor.*, *13*(3), 996–1009.

Mo, K. C., S. Shukla, D. P. Lettenmaier, and L.-C. Chen (2012b), Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction?, *Geophysical Research Letters*, *39*(23).

Monteith, J. L. (1981), Evaporation and surface temperature, *Quarterly Journal of the Royal Meteorological Society*, *107*(451), 1–27.

Montgomery, D. C., and E. A. Peck (1982), *Introduction to linear regression analysis*, Wiley, New York.

Moore, F. C., and D. B. Lobell (2014), Adaptation potential of European agriculture in response to climate change, *Nature Climate Change*, *4*(7), 610–614.

Moore, F. C., and D. B. Lobell (2015), The fingerprint of climate trends on European crop yields, *Proceedings of the National Academy of Sciences*, *112*(9), 2670–2675.

Morris, M. D. (1991), Factorial Sampling Plans for Preliminary Computational Experiments, *Technometrics*, *33*(2), 161–174.

Mosley, M. (1981), Delimitation of New Zealand hydrologic regions, *J. Hydrol.*, *49*, 173–192.

Mu, Q., F. A. Heinsch, M. Zhao, and S. W. Running (2007), Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, *Remote Sensing of Environment*, *111*(4), 519–536.

Mueller, B., and S. I. Seneviratne (2012), Hot days induced by precipitation deficits at the global scale, *Proc Natl Acad Sci U S A*, *109*(31), 12,398–12,403.

Mueller, B., and X. Zhang (2015), Causes of drying trends in northern hemispheric land areas in reconstructed soil moisture data, *Climatic Change*, *134*(1-2), 255–267.

Mueller, B., M. Hirschi, C. Jimenez, P. Ciais, P. A. Dirmeyer, A. J. Dolman, J. B. Fisher, M. Jung, F. Ludwig, F. Maignan, D. G. Miralles, M. F. McCabe, M. Reichstein, J. Sheffield, K. Wang, E. F. Wood, Y. Zhang, and S. I. Seneviratne (2013), Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, *Hydrology and Earth System Sciences*, *17*(10), 3707–3720.

Müller Schmied, H., S. Eisner, D. Franz, M. Wattenbach, F. T. Portmann, M. Flörke, and P. Doll (2014), Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, *Hydrology and Earth System Sciences*, *18*(9), 3511–3538.

NASA (2015), GRACE monthly mass grids – land. Accessed 1 July 2015. [Available online at http://grace.jpl.nasa.gov/data/gracemonthlymassgridsland.].

Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models part I–A discussion of principles, *Journal of hydrology*, *10*(3), 282–290.

Nash, J. E. (1957), The Form of the Instantaneous Unit Hydrograph, *International Association of Scientific Hydrology Publication*, *45*(3), 114–121.

National Research Council (2001), *A Climate Services Vision: First Steps Toward the Future*, The National Academies Press, Washington, DC.

Nearing, G. S., Y. Tian, H. V. Gupta, M. P. Clark, K. W. Harrison, and S. V. Weijs (2016), A philosophical basis for hydrological uncertainty, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *61*(9), 1666–1678.

Nelsen, R. B. (2006), *An introduction to copulas*, Springer-Verlag New York.

Neuman, S. P. (2010), Universal scaling of hydraulic conductivities and dispersivities in geologic media, *Water Resources Research*, *26*(8), 1749–1758.

Neuwirth, E. (2014), RColorBrewer: ColorBrewer Palettes [R package version 1.1-2].

Nicholson, S. (2000), Land surface processes and Sahel climate, *Rev. Geophys.*, *38*(1), 117–139.

Niclòs, R., J. M. Galve, J. a. Valiente, M. J. Estrela, and C. Coll (2011), Accuracy assessment of land surface temperature retrievals from MSG2-SEVIRI data, *Remote Sensing of Environment*, *115*(8), 2126–2140.

Nijzink, R. C., L. Samaniego, J. Mai, R. Kumar, S. Thober, M. Zink, D. Schäfer, H. H. G. Savenije, and M. Hrachowitz (2016), The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, *Hydrology and Earth System Sciences*, *20*(3), 1151–1176.

Niu, G.-Y. (2011), THE COMMUNITY NOAH LAND-SURFACE MODEL (LSM) WITH MULTI-PHYSICS OPTIONS, *Tech. rep.*, National Centers for Environmental Prediction (NCEP), Oregon State University, Air Force, and Hydrology Lab - NWS, Access date: 2017/02/02.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *J. Geophys. Res.*, *116*(D12109).

NMME (2014), nmme phase-2 data plan, *Tech. rep.*, NOAA, uRL: http://www.cpc.ncep.noaa.gov/products/ctb/nmme/NMME-PhaseII-DataPlan-27May.pdf, last access: 25th June 2015.

Nykanen, D. K., and E. Foufoula-Georgiou (2001), Soil moisture variability and scale-dependency of nonlinear parameterizations in coupled land-atmosphere models, *Adv. Water Resour.*, *24*(9-10), 1143–1157.

Nykanen, D. K., E. Foufoula-Georgiou, and W. M. Lapenta (2001), Impact of small-scale rainfall variability on larger-scale spatial organization of land-atmosphere fluxes, *J. Hydrometeorol.*, *2*, 105–120.

Oki, T., and S. Kanae (2006), Global hydrological cycles and world water resources, *Science*, *313*, 1068–1072.

Ol'dekop, E. M. (1911), *On evaporation from the surface of river basins In Transactions on meteorological observations*, vol. 4, 200 pp., University of Tartu., Tartu, Estonia.

Oleson, K., D. Lawrence, G. Bonan, B. Drewniak, M. Huang, C. Koven, S. Levis, F. Li, W. Riley, Z. Subin, S. Swenson, P. Thornton, A. Bozbiyik, R. Fisher, E. Kluzek, J.-F. Lamarque, P. Lawrence, L. Leung, W. Lipscomb, S. Muszala, D. Ricciuto, W. Sacks, Y. Sun, J. Tang, and Z.-L. Yang (2013), Technical Description of version 4.5 of the Community Land Model (CLM), *Tech. rep.*, Ncar Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO, Access date: 2017/02/02.

Orth, R., and S. I. Seneviratne (2012), Analysis of soil moisture memory from observations in Europe, *Journal of Geophysical Research Atmospheres*, *117*(15), 1–19.

Orth, R., and S. I. Seneviratne (2015), Introduction of a simple-model-based land surface dataset for Europe, *Environ. Res. Lett.*, *10*(4), 044012.

Ortiz-Bobea, A. (2011), Improving Agronomic Structure in Econometric Models of Climate Change, in *Agricultural and Applied Economics Association's 2011 AAEA and NAREA Joint Annual Meeting, unpublished*.

Ortiz-Bobea, A. (2013), Is Weather Really Additive in Agricultural Production?, *Working Paper*, *1*(December).

Ortiz-Bobea, A., and R. E. Just (2013), Modeling the structure of adaptation in climate change impact assessment, *American Journal of Agricultural Economics*, *95*(2), 244–251.

Palmer, W. C. (1965), Meteorological drought, *Tech. rep.*, Office of Climatology Research Paper 45, Weather Bureau, Washington, D.C., 58 pp.

Papale, D., M. Reichstein, M. Aubinet, E. Canfora, C. Bernhofer, W. Kutsch, B. Longdoz, S. Rambal, R. Valentini, T. Vesala, and D. Yakir (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*(4), 571–583.

Pappenberger, F., and K. Beven (2006), Ignorance is bliss: Or seven reasons not to use uncertainty analysis, *Water Resources Research*, *42*(5), n/a–n/a.

Parajka, J., R. Merz, and G. Blöschl (2005), A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Sys. Sci.*, *9*(3), 157–171.

Parry, M., O. Canziani, J. Palutikof, P. van der Linden, C. Hanson, and Eds. (2007), The impact of the European 2003 heatwave, *Tech. rep.*, Intergovernmental Panel on Climate Change (IPCC), Cambridge, UK.

Patterson, D. E., and M. W. Smith (1981), The measurement of unfrozen water content by time domain reflectometry: results from laboratory tests., *Can. Geotech. J.*, *18*, 131–144.

Pechlivanidis, I. G., B. Arheimer, C. Donnelly, Y. Hundecha, S. Huang, V. Aich, L. Samaniego, S. Eisner, and P. Shi (2016), Analysis of hydrological extremes at different hydro-climatic regimes under present and future conditions, *Climatic Change*.

Peichl, M., S. Thober, V. Meyer, and L. Samaniego (2018), The effect of soil moisture anomalies on maize yield in Germany, *Natural Hazards and Earth System Science*, *18*(3), 889–906.

Peters-Lidard, C. D., M. Clark, L. Samaniego, N. E. C. Verhoest, T. van Emmerik, R. Uijlenhoet, K. Achieng, T. E. Franz, and R. Woods (2017), Scaling, similarity, and the fourth paradigm for hydrology, *Hydrology and Earth System Sciences*, *21*(7), 3701–3713.

Pfister, L., H. Savenije, and F. Fenicia (2009), *Leonardo Da Vinci's Water Theory: On the Origin and Fate of Water*, IAHS special publication, International Association of Hydrological Sciences.

Piechota, T., and J. Dracup (1996), Drought and Regional Hydrologic Variations in the United States: Associations with the El Niño-Southern Oscillation, *Water Resources Research*, *32*(5), 1359–1373.

Pielke Sr, R. (2013), *Mesoscale meteorological modeling*, Academic Press, Elsevier, International Geophysics, 3 Rev ed.

Pierce, D. (2015), ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data. [R package version 1.15].

Pokhrel, P., and H. V. Gupta (2010), On the use of spatial regularization strategies to improve calibration of distributed watershed models, *Water Resources Research*, *46*(1).

Pokhrel, P., H. V. Gupta, and T. Wagener (2008), A spatial regularization approach to parameter estimation for a distributed watershed model, *Water Resources Research*, *44*(12), W12,419.

Pokhrel, P., K. K. Yilmaz, and H. V. Gupta (2012), Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, *418–419*(0), 49–60.

Popper, K. (1935), *The Logic of Scientific Discovery*, Routledge Classics, Taylor & Francis.

Pozzi, W., J. Sheffield, R. Stefanski, D. Cripe, R. Pulwarty, J. V. Vogt, R. R. Heim Jr., M. J. Brewer, M. Svoboda, R. Westerhoff, A. I. J. M. van Dijk, B. Lloyd-Hughes, F. Pappenberger, M. Werner, E. Dutra, F. Wetterhall, W. Wagner, S. Schubert, K. Mo, M. Nicholson, L. Bettio, L. Nunez, R. van Beek, M. Bierkens, L. G. G. de Gonçalves, J. G. Z. de Mattos, and R. Lawford (2013), Toward Global Drought Early Warning Capability: Expanding International Cooperation for the Development of a Framework for Monitoring and Forecasting, *Bull. Amer. Meteor. Soc.*, *94*(6), 776–785.

Prudhomme, C., I. Giuntoli, E. L. Robinson, D. B. Clark, N. W. Arnell, R. Dankers, B. M. Fekete, W. Franssen, D. Gerten, S. N. Gosling, S. Hagemann, D. M. Hannah, H. Kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada, and D. Wisser (2014), Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, *Proceedings of the National Academy of Sciences*, *111*(9), 3262–3267.

Quenouille, M. (1949), Approximate tests of correlation in time series, *Journal of the Royal Statistical Society*, *11B*.

R Core Team (2015), R: A Language and Environment for Statistical Computing.

Raftery, A. E., A. Zimmer, D. M. W. Frierson, R. Startz, and P. Liu (2017), Less than 2C warming by 2100 unlikely, *Nature Climate Change*, *109*, 13,915–7.

Rajaram, H., J. M. Bahr, G. Blöschl, X. Cai, D. Scott Mackay, A. M. Michalak, A. Montanari, X. Sanchez-Villa, and G. Sander (2015), A reflection on the first 50 years of Water Resources Research, *Water Resources Research*, *51*(10), 7829–7837.

Rakib, M., M. S. Akter, M. Elahi, M. Ali, and M. B. Hossain (2015), An overview of drought hazards and prospective mitigation approach in bangladesh, *Advances in Research*, *5*(6), 1–16.

Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, and R. Uijlenhoet (2014), Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, *Water Resour. Res.*, *50*, 1–18.

Rakovec, O., R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego (2016a), Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, *Journal of Hydrometeorology*, *17*(1), 287–307.

Rakovec, O., R. Kumar, S. Attinger, and L. Samaniego (2016b), Improving the realism of hydrologic model functioning through multivariate parameter estimation, *Water Resources Research*, *52*(10), 7779–7792.

Rakovec, O., R. Kumar, J. Mai, M. Cuntz, S. Thober, M. Zink, S. Attinger, D. Schäfer, M. Schrön, and L. Samaniego (2016c), Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, *J. Hydrometeorol.*, *17*(1), 287–307.

Rakovec, O., N. Mizukami, R. Kumar, A. J. Newman, S. Thober, A. W. Wood, M. P. Clark, and L. Samaniego (2019), Diagnostic Evaluation of Large-Domain Hydrologic Models Calibrated Across the Contiguous United States, *Journal of Geophysical Research-Atmospheres*, *124*(24), 13,991–14,007.

Rawls, W. J. (1983), Estimating soil bulk densiry from particle size analysis and organic matter content, *Soil Sci.*, *135*, 123–125.

Rebmann, C., M. Zeri, G. Lasslop, M. Mund, O. Kolle, E.-D. Schulze, and C. Feigenwinter (2010), Treatment and assessment of the CO2-exchange at a complex forest site in Thuringia, Germany, *Agricultural and Forest Meteorology*, *150*(5), 684–691.

Reddmont, T., and R. Koch (1991), Surface climate and stream flow variability in the Western United States and their relationship to large-scale circulation indices, *Water Resources Research*, *17*, 2381–2399.

Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, , and D. Participants (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, *298*(1–4), 27–60.

Refsgaard, J., and B. Storm (1995), MIKE SHE, in *Computer Models of Watershed Hydrology*, edited by V. Singh, pp. 809–846, Water Resources Publications, Colorado, USA.

Reggiani, P., M. Sivapalan, and S. Majid Hassanizadeh (1998), A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, *Advances in Water Resources*, *22*(4), 367–398.

Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, *31*(19).

Reichle, R. H., S. V. Kumar, S. P. P. Mahanama, R. D. Koster, and Q. Liu (2010), Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models, *Journal of Hydrometeorology*, *11*(5), 1103–1122.

Reichstein, M., E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havránková, H. Ilvesniemi, D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, J.-M. Ourcival, J. Pumpanen, S. Rambal, E. Rotenberg, M. Sanz, J. Tenhunen, G. Seufert, F. Vaccari, T. Vesala, D. Yakir, and R. Valentini (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, *Glob. Change Biol.*, *11*(9), 1424–1439.

Relief (2011), Horn of africa crisis: 2011-2012, last access: March 26th, 2015.

Richardson, L. F. (1922), *Weather Prediction by Numerical Process*, Cambridge University Press.

Ritchie, H., and M. Roser (2019), Global deads from natural disasters (1900-2016), https://ourworldindata.org/uploads/2019/11/Annual-deaths-by-natural-disaster.png.

Rivera Villarreyes, C. A., G. Baroni, and S. E. Oswald (2011), Integral quantification of seasonal soil moisture changes in farmland by cosmic-ray neutrons, *Hydrol. Earth Syst. Sc.*, *15*(12), 3843–3859.

Roberts, M. J., W. Schlenker, and J. Eyer (2013), Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change, *American Journal of Agricultural Economics*, *95*(2), 236–243.

Robine, J., S. L. K. Cheung, S. Le Roy, H. Van Oyen, C. Griffiths, J.-P. Michel, and F. R. Herrmann (2008), Death toll exceeded 70,000 in Europe during the summer of 2003, *Comptes Rendus Biologies*, *331*(2), 171–178.

Rodell, M., P. R. Houser, and A. A. Berg (2005), Evaluation of 10 methods for initializing a land surface model, *J. Hydrometeor.*, *6*, 146–155.

Rodriguez-Iturbe, I., and J. B. Valdes (1979), The geomorphologic structure of hydrologic response, *Water Resour. Res.*, *15*(6), 1409–1420.

Rodriguez-Iturbe, I., D. Entekhabi, and R. Bras (1991), Nonlinear dynamics of soil moisture at climate scales: 1. stochastic analysis, *Water Resour. Res.*, *27*(8), 1899–1906.

Rosero, E., L. E. Gulden, Z.-L. Yang, L. De Goncalves, G.-Y. Niu, and Y. H. Kaheil (2011), Ensemble evaluation of hydrologically enhanced noah-lsm: Partitioning of the water balance in high-resolution simulations over the little washita river experimental watershed, *J. Hydrometeor.*, *12*(1), 45–64.

Rötter, R. P., T. R. Carter, J. E. Olesen, and J. R. Porter (2011), Crop–climate models need an overhaul, *Nature Climate Change*, *1*(4), 175–177.

Rubel, F., and M. Kottek (2010), Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification, *Meteorol. Z.*, *19*, 135–141.

Saha, T. R., R. Rakovec, P. Shrestha, S. Thober, and L. Samaniego (2020), Monitoring and Assessing Soil Moisture Drought by Using Hydro-Meteorological Model in South Asia, *Environmental Research Letters*, in preparation.

Sakumura, C., S. Bettadpur, and S. Bruinsma (2014), Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models, *Geophys. Res. Lett.*, *41*, 1389–1397.

Samaniego, L. (2003), *Hydrological Consequences of Land Use/ Land Cover Change in Mesoscale Catchments*, Institute of Hydraulic Engineering, University of Stuttgart, Faculty of Civil Engineering, Stuttgart, Ph.D. dissertation No. 118, ISBN 3-9337 61-21-2.

Samaniego, L. (2017), EDgE model chain and development of Sectoral Climate Impact Indicators, Online http://edge.climate.copernicus.eu, last visited, 11.09.2017.

Samaniego, L., and A. Bárdossy (2005), Robust parametric models of runoff characteristics at the mesoscale, *Journal of Hydrology*, *303*(1-4), 136–151.

Samaniego, L., and A. Bárdossy (2006), Simulation of the impacts of land use/cover and climatic changes on the runoff characteristics at the mesoscale, *Ecological modelling*, *196*(1-2), 45–61.

Samaniego, L., and A. Bárdossy (2007), Relating macroclimatic circulation patterns with characteristics of floods and droughts at the mesoscale, *J. Hydrol.*, *335*, 109–123.

Samaniego, L., A. Bárdossy, and K. Schulz (2008), Supervised classification of remotely sensed imagery using a modified k-nn technique, *IEEE T. Geosci. Remote*, *46*(7), 2112–2125.

Samaniego, L., R. Kumar, and S. Attinger (2010a), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, *46*(5), w05523.

Samaniego, L., A. Bárdossy, and R. Kumar (2010b), Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, *Water Resources Research*, *46*(2), W02,506.

Samaniego, L., R. Kumar, and C. Jackisch (2011), Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data, *Hydrology research*, *42*(5), 338–355.

Samaniego, L., R. Kumar, and M. Zink (2013), Implications of Parameter Uncertainty on Soil Moisture Drought Analysis in Germany, *Journal of Hydrometeorology*, *14*(1), 47–68.

Samaniego, L., O. Rakovec, R. Kumar, D. Schaefer, M. Cuntz, J. Mai, S. Thober, and S. Attinger (2014), Multiscale prediction and verification of water fluxes and states over large river basins, in *Scientific Program of the AGU General Assembly 2014*.

Samaniego, L., R. Kumar, L. Breuer, A. Chamorro, M. Flörke, I. G. Pechlivanidis, D. Schäfer, H. Shah, T. Vetter, M. Wortmann, and X. Zeng (2016), Propagation of forcing and model uncertainties on to hydrological drought characteristics in a multi-model century-long experiment in large river basins, *Climatic Change*, *141*(3), 435–449.

Samaniego, L., R. Kumar, S. Thober, O. Rakovec, M. Zink, N. Wanders, S. Eisner, H. Müller Schmied, E. H. Sutanudjaja, K. Warrach-Sagi, and S. Attinger (2017), Toward seamless hydrologic predictions across spatial scales, *Hydrology and Earth System Sciences*, *21*(9), 4323–4346.

Samaniego, L., S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. Wood, and A. Marx (2018), Anthropogenic warming exacerbates European soil moisture droughts, *Nat. Clim. Change*, *4*.

Samaniego, L., S. Thober, N. Wanders, M. Pan, O. Rakovec, J. Sheffield, E. F. Wood, C. Prudhomme, G. Rees, H. Houghton-Carr, M. Fry, K. Smith, G. Watts, H. Histal, T. Estrella, C. Buontempo, A. Marx, and R. Kumar (2019a), Hydrological forecasts and projections for improved decision-making in the water sector in Europe, *Bull. Am. Meteorol. Soc.*, *100*, 2451–2472.

Samaniego, L., M. Kaluza, S. Thober, and O. Rakovec (2019b), Seamless reconstruction of global hydrological fluxes and states at high resolution, *AGU Fall Meeting Abstract N. H43M-2223*.

Samaniego, L. E., K. Warrach-Sagi, M. Zink, and V. Wulfmeyer (2012), Verification of High Resolution Soil Moisture and Latent Heat in Germany, *AGU Fall Meeting Abstracts*, provided by the SAO/NASA Astrophysics Data System.

Sankarasubramanian, A., and R. M. Vogel (2002), Comment on the paper: Basin hydrologic response relations to distributed physiographic descriptors and climate by Karen Plaut Berger, Dara Entekhabi, 2001. Journal of Hydrology 247, 169–182, *J. Hydrol.*, *263*(1), 257–261.

Sarkar, D. (2008), *Lattice: Multivariate Data Visualization with R*, Springer, New York.

Sarkar, D., and F. Andrews (2016), latticeExtra: Extra Graphical Utilities Based on Lattice [R package version 0.6-28].

Savenije, H. H. G., and M. Hrachowitz (2017), HESS Opinions ”Catchments as meta-organisms — a new blueprint for hydrological modelling”, *Hydrology and Earth System Sciences*, *21*(2), 1107–1116.

Schaake, J., Q. Duan, V. Koren, K. Mitchell, P. Houser, E. Wood, A. Robock, D. P. Lettenmaier, D. Lohmann, B. Cosgrove, and et al. (2004), An intercomparison of soil moisture fields in the north american land data assimilation system (nldas), *J. Geophys. Res.*, *109*(D1), 1–16.

Schewe, J., J. Heinke, D. Gerten, I. Haddeland, N. W. Arnell, D. B. Clark, R. Dankers, S. Eisner, B. M. Fekete, F. J. Colón-González, S. N. Gosling, H. Kim, X. Liu, Y. Masaki, F. T. Portmann, Y. Satoh, T. Stacke, Q. Tang, Y. Wada, D. Wisser, T. Albrecht, K. Frieler, F. Piontek, L. Warszawski, and P. Kabat (2014), Multimodel assessment of water scarcity under climate change, *Proceedings of the National Academy of Sciences*, *111*(9), 3245–3250.

Schindler, D., and H. Mayer (2007), Forest meteorological investigation of the drought 2003 in the southwest of germany, *Allg. Forst Jagdztg.*, *178*(2-3), 21–37.

Schindler, U., J. Steidl, L. Müller, F. Eulenstein, and J. Thiere (2007), Drought risk to agricultural land in northeast and central germany, *J. Plant Nutr. Soil Sc.*, *170*(3), 357–362.

Schlenker, W., and D. B. Lobell (2010), Robust negative impacts of climate change on African agriculture, *Environmental Research Letters*, *5*(1), 14,010.

Schlenker, W., and M. J. Roberts (2006), Nonlinear Effects of Weather on Corn Yields, *Review of Agricultural Economics*, *28*(3), 391–398.

Schlenker, W., and M. J. Roberts (2009), Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change, *Proceedings of the National Academy of Sciences*, *106*(37), 15,594–15,598.

Schlenker, W., W. M. Hanemann, and A. C. Fisher (2005), Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach, *American Economic Review*, *95*(1), 395–406.

Schlenker, W., W. M. Hanemann, and A. C. Fisher (2006), The impact of global warming on US agriculture: an econometric analysis of optimal growing conditions, *Review of Economics and Statistics*, *88*(February), 113–125.

Schlenker, W., M. J. Roberts, and D. B. Lobell (2013), US maize adaptability, *Nature Climate Change*, *3*(8), 690–691.

Scholz, C. A., T. C. Johnson, A. S. Cohen, J. W. King, J. A. Peck, J. T. Overpeck, M. R. Talbot, E. T. Brown, L. Kalindekafe, P. Y. O. Amoako, R. P. Lyons, T. M. Shanahan, I. S. Castaneda, C. W. Heil, S. L. Forman, L. R. McHargue, K. R. Beuning, J. Gomez, and J. Pierson (2007), East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins, *Proceedings of the National Academy of Sciences*, *104*(42), 16,416–16,421.

Schreiber, P. (1904), Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa, *Zeitschrift Für Meteorologie*, *21*, 441–452.

Schulla, J., and K. Jasper (2007), *Model Description WaSiM-ETH (Water balance Simulation Model ETH)*, ETH, Zürich.

Schultz, G. (1988), Remote-Sensing in Hydrology, *Journal of Hydrology*, *100*(1-3), 239–265.

Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, *6*(2), 461–464.

Schweizer, B., and E. F. Wolff (1981), On nonparametric measures of dependence for random variables, *Ann. Stat.*, *9*, 879–885.

Scott, D., and S. Sain (2005), Multidimensional density estimation, *Handbook of Statistics*, *24*, 229–261.

SCS (1973), *A Method for Estimating Volume and Rate of Runoff in Small Watersheds*, U.S. Department of Agriculture. Soil Conservation Service.

Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agric. Fores. Meteorol.*, *98-99*, 279 – 293.

Seibert, J. (2000), Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, *4* (2), 215–224.

Sellers, P. J., R. E. Dickinson, D. A. Randall, A. K. Betts, F. G. Hall, J. A. Berry, G. J. Collatz, A. S. Denning, H. A. Mooney, C. A. Nobre, N. Sato, C. B. Field, and A. Henderson-Sellers (1997), Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, *Science*, *275*(5299), 502–509.

Seneviratne, S., and R. Stöckli (2008), The role of land-atmosphere interactions for climate variability in europe, in *Climate Variability and Extremes during the Past 100 years*, *Advances in Global Change Research*, vol. 33, edited by S. Brönnimann, J. Luterbacher, T. Ewen, H. Diaz, R. Stolarski, and U. Neu, Springer Verlag, xVI.

Seneviratne, S. I., N. Nicholls, D. Easterling, C. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera, and X. Zhang (2012), *Changes in climate extremes and their impacts on the natural physical environment*, chap. Changes in climate extremes and their impacts on the natural physical environment, IPCC, in: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 109-230.

Seneviratne, S. I., M. Wilhelm, T. Stanelle, B. van den Hurk, S. Hagemann, A. Berg, F. Cheruy, M. E. Higgins, A. Meier, V. Brovkin, M. Claussen, A. Ducharne, J.-L. Dufresne, K. L. Findell, J. Ghattas, D. M. Lawrence, S. Malyshev, M. Rummukainen, and B. Smith (2013), Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment, *Geophysical Research Letters*, *40*(19), 5212–5217.

Seneviratne, S. I., M. G. Donat, A. J. Pitman, R. Knutti, and R. L. Wilby (2016), Allowable CO2 emissions based on regional and impact-related climate targets, *Nature*, *529*(7587), 477–483.

Sessa, R., and H. Dolman (Eds.) (2008), *Terrestrial Essential Climate Variables for climate change assessment, mitigation and adaptation*, vol. Global terrestrial Observing System, GTOS 52, Food and Agriculture Organization of the United Nations, Rome, http://www.fao.org/gtos/doc/pub52.pdf visited 2018-03-21.

Shah, R. D., and V. Mishra (2015), Development of an Experimental Near-Real-Time Drought Monitor for India, *Journal of Hydrometeorology*, *16*(1), 327–345.

Shahrbanou Madadgar, and Hamid Moradkhani (2013), A Bayesian Framework for Probabilistic Seasonal Drought Forecasting, *Journal of Hydrometeorology*, *14*(6), 1685–1705.

Sheffield, J., and E. F. Wood (2007), Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle, *Journal of Geophysical Research*, *112*(D17), 7449–21.

Sheffield, J., and E. F. Wood (2008a), Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, *Climate Dynamics*, *31*(1), 79–105.

Sheffield, J., and E. F. Wood (2008b), Global Trends and Variability in Soil Moisture and Drought Characteristics, 1950–2000, from Observation-Driven Simulations of the Terrestrial Hydrologic Cycle, *Journal of Climate*, *21*(3), 432–458.

Sheffield, J., and E. F. Wood (2011), *Drought: Past Problems and Future Scenarios*, 192 pp., Earthscan.

Sheffield, J., G. Goteti, F. Wen, and W. E. F. (2004), A simulated soil moisture based drought analysis for the united states, *J. Geophys. Res.*, *109*, D24108.

Sheffield, J., K. M. Andreadis, E. F. Wood, and D. P. Lettenmaier (2009), Global and Continental Drought in the Second Half of the Twentieth Century: Severity-Area-Duration Analysis and Temporal Variability of Large-Scale Events, *J. Climate*, *22*(8), 1962.

Sheffield, J., B. Livneh, and E. F. Wood (2012), Representation of terrestrial hydrology and large-scale drought of the continental united states from the north american regional reanalysis, *J. Hydrometeor*, *13*(3), 856–876.

Sheffield, J., E. F. Wood, and M. L. Roderick (2013), Little change in global drought over the past 60 years, *Nature*, *491*(7424), 435–438.

Sheffield, J., E. F. Wood, N. Chaney, K. Guan, S. Sadri, X. Yuan, L. Olang, A. Amani, A. Ali, S. Demuth, and L. Ogallo (2014), A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security, *Bull. Amer. Meteor. Soc.*, *95*(6), 861–882.

Sherman, L. K. (1932), Streamflow from Rainfall by the Unit Graph Method, *Eng. News Rec.*, *108*, 501–505.

Shevenell, L. (1999), Regional potential evapotranspiration in arid climates based on temperature, topography and calculated solar radiation, *Hydrol. Process.*, *13*(13), 577–596.

Shorthouse, C., and N. Arnell (1997), Spatial and temporal variability in European river flows and the North Atlantic Oscillation, *FRIEND'97: International Association of Hydrological Science Publications*, *246*, 77–85.

Shukla, S., and D. P. Lettenmaier (2011), Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill, *Hydrology and Earth System Sciences*, *15*(11), 3529–3538.

Shukla, S., and A. W. Wood (2008), Use of a standardized runoff index for characterizing hydrologic drought, *Geophysical Research Letters*, *35*(2).

Shukla, S., A. C. Steinemann, and D. P. Lettenmaier (2011), Drought monitoring for washington state: Indicators and applications, *J. Hydrometeor.*, *12*(1), 66–83.

Shukla, S., J. Sheffield, E. F. Wood, and D. P. Lettenmaier (2013), On the sources of global land surface hydrologic predictability, *Hydrology and Earth System Sciences*, *17*(7), 2781–2796.

Shukla, S., A. McNally, G. Husak, and C. Funk (2014), A seasonal agricultural drought forecast system for food-insecure regions of East Africa, *Hydrology and Earth System Sciences*, *18*(10), 3907–3921.

Shuttleworth, W. J. (2012), *Terrestrial Hydrometeorology*, Wiley.

Silvestro, F., S. Gabellani, F. Delogu, R. Rudari, and G. Boni (2013), Exploiting remote sensing land surface temperature in distributed hydrological modelling: the example of the Continuum model, *Hydrology and Earth System Sciences*, *17*(1), 39–62.

Silvestro, F., S. Gabellani, R. Rudari, F. Delogu, P. Laiolo, and G. Boni (2015), Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data, *Hydrology and Earth System Sciences*, *19*(4), 1727–1751.

Sinclair, T. R., and N. G. Seligman (1996), Crop modeling: From infancy to maturity, *Agronomy Journal*, *88*(5), 698–704.

Singh, R., S. A. Archfield, and T. Wagener (2014), Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments – A comparative hydrology approach, *Journal of Hydrology*, *517*, 985–996.

Singh, S. K., A. Bárdossy, J. Götzinger, and K. P. Sudheer (2012), Effect of spatial resolution on regionalization of hydrological model parameters, *Hydrological Processes*, *26*(23), 3499–3509.

Sinha, A., L. Stott, M. Berkelhammer, H. Cheng, R. L. Edwards, B. Buckley, M. Aldenderfer, and M. Mudelsee (2011), A global context for megadroughts in monsoon Asia during the past millennium, *Quaternary Science Reviews*, *30*(1-2), 47–62.

Sivapalan, M., K. Takeuchi, S. W. Franks, V. K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. J. McDonnell, E. M. Mendiondo, P. E. O'Connell, T. Oki, J. W. Pomeroy, D. Schertzer, S. Brook, and E. Zehe (2003), IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrological Sciences Journal*, *48*(6), 857–880.

Sivapalan, M., R. Grayson, and R. Woods (2004), Scale and scaling in hydrology, *Hydrological Processes*, *18*(8), 1369–1371.

Smith, A., and R. Katz (2013), US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases, *Natural Hazards*, *67*(2), 387–410.

Smith, A., and J. Matthews (2015), Quantifying uncertainty and variable sensitivity within the us billion-dollar weather and climate disaster cost estimates, *Natural Hazards*, pp. 1–23.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulation*, *55*(1–3), 271–280.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, *16*(2), 430–442.

Soubeyroux, J., E. Martin, L. Franchisteguy, and F. Habets (2008), Safran-Isba-Modcou (SIM) : Un outil pour le suivi hydrométéorologique opérationnel et les études, *La Météorologie*, *63*, 40–45.

Stahl, K., and S. Demuth (1999), Linking streamflow drought to the occurrence of atmospheric circulation patterns, *Hydrol. Sci. J.*, *44*(3), 467–482.

Statisitisches Bundesamt (2011), Land- und Forstwirtschaft, Fischerei - Bodenbearbeitung, Bewässerung, Landschaftselemente - Erhebung über landwirtschafliche Produktionsmethoden (ELPM) -, *Tech. rep.*, Statistisches, Wiesbaden.

Statisitisches Bundesamt (2016), Weizen und Silomais dominieren mit 45 % den Anbau auf dem Ackerland Statistisches Bundesamt.

Statistische Ämter des Bundes und der Länder (2017), The Regional Database Germany (”Regionaldatenbank Deutschland”).

Steffen, W., J. Rockström, K. Richardson, T. M. Lenton, C. Folke, D. Liverman, C. P. Summerhayes, A. D. Barnosky, S. E. Cornell, M. Crucifix, J. F. Donges, I. Fetzer, S. J. Lade, M. Scheffer, R. Winkelmann, and H. J. Schellnhuber (2018), Trajectories of the Earth System in the Anthropocene., *Proceedings of the National Academy of Sciences of the United States of America*, *2*(33), 201810,141–8.

Stensrud, D. J. (2007), *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*, Cambridge University Press.

Stisen, S., M. F. McCabe, J. C. Refsgaard, S. Lerer, and M. B. Butts (2011), Model parameter analysis using remotely sensed pattern information in a multi-constraint framework, *Journal of Hydrology*, *409*(1-2), 337–349.

Stöckli, R., P. L. Vidale, A. Boone, and C. Schär (2007), Impact of scale and aggregation on the terrestrial water exchange: Integrating land surface models and rhone catchment observations, *J. Hydrometeorol.*, *8*(5), 1002–1015.

Stone, R. (2009), ECOLOGY: Tree Rings Tell of Angkor's Dying Days, *Science*, *323*(5917), 999b–999b.

Su, H., Z.-L. Yang, R. E. Dickinson, C. R. Wilson, and G.-Y. Niu (2010), Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information, *J. Geophys. Res.*, *115*(D10), 1–14.

Sutanudjaja, E., L. Van Beek, S. De Jong, F. Van Geer, and M. Bierkens (2014), Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, *Water Resour. Res.*, *50*(1), 687–705.

Sutanudjaja, E., J. Bosmans, N. Chaney, M. P. Clark, L. E. Condon, C. H. David, A. P. J. De Roo, P. M. Doll, N. Drost, S. Eisner, J. S. Famiglietti, M. Floerke, J. M. Gilbert, D. J. Gochis, R. Hut, J. Keune, S. J. Kollet, R. M. Maxwell, M. Pan, O. Rakovec, J. T. Reager, II, L. E. Samaniego, H. Mueller Schmied, T. Trautmann, L. P. van Beek, N. van de Giesen, E. F. Wood, M. F. Bierkens, and R. Kumar (2015), The HyperHydro (Hˆ2) experiment for comparing different large-scale models at various resolutions, *AGU Fall Meeting Abstracts*.

Sutanudjaja, E., R. van Beek, Y. Wada, J. Bosmans, N. Drost, I. de Graaf, K. de Jong, P. Lopez Lopez, S. Pessenteiner, O. Schmitz, M. Straatsma, N. Wanders, D. Wisser, and M. Bierkens (2016), PCR-GLOBWB_model: PCR-GLOBWB version v2.1.0_alpha, Note that this is still a 'pre-release' version. Please test it and, if you find any, please report any bugs and/or issues.

Sutanudjaja, E. H., R. van Beek, N. Wanders, Y. Wada, J. H. C. Bosmans, N. Drost, R. J. van der Ent, I. E. M. de Graaf, J. M. Hoch, K. de Jong, D. Karssenberg, P. López López, S. Peßenteiner, O. Schmitz, M. W. Straatsma, E. Vannametee, D. Wisser, and M. F. P. Bierkens (2018), PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, *Geoscientific Model Development*, *11*(6), 2429–2453.

Svoboda, M., D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, D. Miskus, and S. Stephens (2002), The drought monitor, *Bulletin of the American Meteorological Society*, *83*(8), 1181–1190.

Swain, D. L., M. Tsiang, M. Haugen, D. Singh, A. Charland, B. Rajaratnam, and N. S. Diffenbaugh (2014), The Extraordinary California Drought of 2013/2014: Character, Context, and the Role of Climate Change, *Bulletin of the American Meteorological Society*, *95*(9), S3–S7.

Swenson, S. C., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, *33*, L08402.

Tan, L. (2015), A Chinese cave links climate change, social impacts, and human adaptation over the last 500 years, *Nature Publishing Group*, pp. 1–10.

Taylor, A. L., S. Dessai, and W. B. de Bruin (2015), Communicating uncertainty in seasonal and interannual climate forecasts in europe, *Philos. Trans. A Math. Phys. Eng. Sci.*, *373*.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res. Atmos.*, *106*(D7), 7183–7192.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An Overview of CMIP5 and the Experiment Design, *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Teixeira, E. I., G. Fischer, H. van Velthuizen, C. Walter, and F. Ewert (2013), Global hot-spots of heat stress on agricultural crops due to climate change, *Agricultural and Forest Meteorology*, *170*, 206–215.

Teng, J., J. Vaze, F. H. S. Chiew, B. Wang, and J. Perraud (2012), Estimating the Relative Uncertainties Sourced from GCMs and Hydrological Models in Modeling Climate Change Impact on Runoff, *Journal of Hydrometeorology*, *13*(1), 122–139.

Tetzlaff, D., S. K. Carey, H. Laudon, and K. McGuire (2010), Catchment processes and heterogeneity at multiple scales–benchmarking observations, conceptualization and prediction, *Hydrol. Processes*, *24*(16), 2203–2208.

Tewolde, M. H., and J. C. Smithers (2006), Flood routing in ungauged catchments using muskingum methods, *Journal of Water South Africa*, *32*(3), 379–388.

Therrell, M. D., D. W. Stahle, and R. A. Soto (2004), Aztec drought and the ”curse of one rabbit”, *Bulletin of the American Meteorological Society*, *85*(9), 1263–+.

Thielen, J., J. Bartholmes, M. H. Ramos, and A. de Roo (2009), The European Flood Alert System - Part 1: Concept and development, *Hydrol. Earth. Syst. Sc.*, *13*(2), 125–140.

Thober, S., and L. Samaniego (2014), Robust ensemble selection by multivariate evaluation of extreme precipitation and temperature characteristics, *Journal of Geophysical Research: Atmospheres*, *119*(2), 594–613.

Thober, S., J. Mai, M. Zink, and L. Samaniego (2014), Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets, *Water Resources Research*, *50*(11), 8714–8735.

Thober, S., R. Kumar, J. Sheffield, J. Mai, D. Schäfer, and L. Samaniego (2015), Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME), *Journal of Hydrometeorology*, *16*(6), 2329–2344.

Thober, S., R. Kumar, N. Wanders, A. Marx, M. Pan, O. Rakovec, L. Samaniego, J. Sheffield, E. F. Wood, and M. Zink (2018), Multi-model ensemble projections of european river floods and high flows at 1.5, 2, and 3 degree global warming, *Environ. Res. Lett.*, pp. 1–22.

Thober, S., G. Schweppe, S. Attinger, and L. Samaniego (2019a), Application of the Multiscale Parameter Regionalization (MPR) to the land-surface model HTESSEL, *Geophysical Research Abstracts, EGU General Assembly Vol. 21, EGU2019-10122* .

Thober, S., M. Cuntz, M. Kelbling, R. Kumar, J. Mai, and L. Samaniego (2019b), The multiscale routing model mRM v1.0: simple river routing at resolutions from 1 to 50 km, *Geoscientific Model Development*, *12*(6), 2501–2521.

Thober, S., M. Kelbling, F. Pappenberger, C. Prudhomme, G. Balsamo, G. Schweppe, S. Attinger, and L. Samaniego (2020), Improvement of the simulation of the water and energy cycle using Multiscale Parameter Regionalization (MPR) , *Geophysical Research Abstracts, EGU General Assembly, EGU2020-x*.

Thompson, L. M. (1969), Weather and Technology in the Production of Corn in the U. S. Corn Belt, *Agronomy Journal*, *61*(5).

Timmins, C., and W. Schlenker (2009), Reduced-Form Versus Structural Modeling in Environmental and Resource Economics, *Annual Review of Resource Economics*, *1*(1), 351–380.

Todini, E. (2007), A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach, *Hydrology and Earth System Sciences Discussions*, *4*(3), 1549–1592.

Tolson, B. a., and C. a. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resources Research*, *43*(1), 1–16.

Tolson, B. A., and C. A. Shoemaker (2008), Efficient prediction uncertainty approximation in the calibration of environmental simulation models, *Water Resour. Res.*, *44*, WR005,869.

Trenberth, K. E., A. Dai, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield (2014), Global warming and changes in drought, *Nature Climate Change*, *4*(1), 17–22.

Trigo, I. F., I. T. Monteiro, F. Olesen, and E. Kabsch (2008), An assessment of remotely sensed land surface temperature, *Journal of Geophysical Research*, *113*(D17), 1–12.

Trnka, M., P. Hlavinka, D. Semerádová, J. Balek, M. Možný, P. Štěpánek, P. Zahradníček, M. Hayes, J. Eitzinger, and Z. Žalud (2014), Drought monitor for the Czech Republic - www.intersucho.cz, in *Mendel and Bioclimatology*, pp. 630–638.

Troy, T. J., E. F. Wood, and J. Sheffield (2008), An efficient calibration method for continental-scale land surface modeling, *Water Resources Research*, *44*(9).

Tubiello, F. N., J.-F. Soussana, and S. M. Howden (2007), Crop and pasture response to climate change., *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19,686–90.

Twedt, T., J. J. Schaake, and E. Peck (1977), National weather service extended streamflow prediction, in *45th Annual Western Snow Conference*, Western Snow Conference, Albuquerque, New Mexico.

UNFCC (2015), Adoption of the Paris agreement, Proposal by the President, *Tech. rep.*, United Nations, Geneva, Switzerland.

UNISDR (2015), *Global Assessment Report on Disaster Risk Reduction - Making Development Sustainable: The Future of Disaster Risk Management* -, 311 pp., United Nations Office for Disaster Risk Reduction (UNISDR), Geneva, Switzerland.

Urban, D., M. J. Roberts, W. Schlenker, and D. B. Lobell (2012), Projected temperature changes indicate significant increase in interannual variability of U.S. maize yields: A Letter, *Climatic Change*, *112*(2), 525–533.

Urban, D. W., M. J. Roberts, W. Schlenker, and D. B. Lobell (2015a), The effects of extremely wet planting conditions on maize and soybean yields, *Climatic Change*, *130*, 247–260.

Urban, D. W., J. Sheffield, and D. B. Lobell (2015b), The impacts of future climate and carbon dioxide changes on the average and variability of US maize yields under two emission scenarios, *Environmental Research Letters*, *10*(4), 045,003.

van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens (2011), Global monthly water stress: 1. Water balance and water availability, *Water Resour. Res.*, *47*(7), 1–25.

van Dijk, A. I. J. M., H. E. Beck, R. S. Crosbie, R. A. M. de Jeu, Y. Y. Liu, G. M. Podger, B. Timbal, and N. R. Viney (2013), The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resources Research*, *49*(2), 1040–1057.

van Lanen, H. A., G. Laaha, D. G. Kingston, T. Gauster, M. Ionita, J.-P. Vidal, R. Vlnas, L. M. Tallaksen, K. Stahl, J. Hannaford, C. Delus, M. Fendekova, L. Mediero, C. Prudhomme, E. Rets, R. J. Romanowicz, S. Gailliez, W. K. Wong, M.-J. Adler, V. Blauhut, L. Caillouet, S. Chelcea, N. Frolova, L. Gudmundsson, M. Hanel, K. Haslinger, M. Kireeva, M. Osuch, E. Sauquet, J. H. Stagge, and A. F. van Loon (2016), Hydrology needed to manage droughts: the 2015 European case, *Hydrological Processes*, *30*(17), 3097–3104.

van Lanen, H. A. J., and E. Peters (2000), Definition, effects and assessment of groundwater droughts, in *Drought and Drought Mitigation in Europe*, edited by J. V. Vogt and F. Somma, pp. 49–61, Kluwer Academic Publishers, Dordrecht, Netherlands.

van Loon, A. F., E. Tijdeman, N. Wanders, H. A. van Lanen, A. J. Teuling, and R. Uijlenhoet (2014), How climate seasonality modifies drought duration and deficit, *Journal of Geophysical Research: Atmospheres*, *119*(8), 4640–4656.

Vautard, R., A. Gobiet, S. Sobolowski, E. Kjellstrom, A. Stegehuis, P. Watkiss, T. Mendlik, O. Landgren, G. Nikulin, C. Teichmann, and D. Jacob (2014), The European climate under a 2 degrees C global warming, *Environmental Research Letters*, *9*(3).

Velpuri, N., G. Senay, R. Singh, S. Bohms, and J. Verdin (2013), A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET, *Remote Sens. Environ.*, *139*, 35–49.

Vereecken, H., J. A. Huisman, H. Bogena, J. Vanderborght, J. A. Vrugt, and J. W. Hopmans (2008), On the value of soil moisture measurements in vadose zone hydrology: A review, *Water Resources Research*, *44*(4), n/a–n/a.

Vetter, T., S. Huang, V. Aich, T. Yang, X. Wang, V. Krysanova, and F. Hattermann (2015), Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, *Earth System Dynamics*, *6*(1), 17–43.

Vidal, J.-P., E. Martin, L. Franchistéguy, F. Habets, J.-M. Soubeyroux, M. Blanchard, and M. Baillon (2010), Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, *Hydrology and Earth System Sciences*, *14*(3), 459–478.

Vinke, K., M. A. Martin, S. Adams, F. Baarsch, A. Bondeau, D. Coumou, R. V. Donner, A. Menon, M. Perrette, K. Rehfeld, A. Robinson, M. Rocha, M. Schaeffer, S. Schwan, O. Serdeczny, and A. Svirejeva-Hopkins (2017), Climatic risks and impacts in south asia: extremes of water scarcity and excess, *Reg Environ Change*, *17*, 1569–1583.

Viterbo, P., and C. M. Beljaars (1995), An improved land surface parameterization scheme in the ECMWF model and its validation, *Journal of Climate*, *8*, 2716–2748.

Viviroli, D., M. Zappa, J. Gurtz, and R. Weingartner (2009), An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, *Environ. Model. Softw.*, *24*, 1209–1222.

von Bertalanffy, L. (1968), *General System Theory: Foundations, Development, Applications*, George Braziller, New York.

Vrugt, J., C. Diks, H. Gupta, W. Bouten, and J. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, *41*, W01017.

Wada, Y., and M. F. P. Bierkens (2014), Sustainability of global water use: past reconstruction and future projections, *Environmental Research Letters*, *9*(10), 104,003–15.

Wada, Y., L. P. H. van Beek, C. M. van Kempen, J. W. T. M. Reckman, S. Vasak, and M. F. P. Bierkens (2010), Global depletion of groundwater resources, *Geophys. Res. Lett.*, *37*(20).

Wagener, T., and J. Kollat (2007), Numerical and visual evaluation of hydrological and environmental models using the monte carlo analysis toolbox, *Environ. Model. Soft.*, *22*, 1021–1033.

Wagener, T., and H. S. Wheater (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *Journal of Hydrology*, *320*(1-2), 132–154.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis, *Hydrological Processes*, *17*(2), 455–476.

Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment Classification and Hydrologic Similarity, *Geography Compass*, *1*(4), 901–931.

Wagner, W., G. Blöschl, P. Pampaloni, J.-C. Calvet, B. Bizzarri, J.-P. Wigneron, and Y. Kerr (2007), Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, *Nordic Hydrology*, *38*(1), 1–20.

Wahid, A., S. Gelani, M. Ashraf, and M. R. Foolad (2007), Heat tolerance in plants: An overview, *Environmental and Experimental Botany*, *61*(3), 199–223.

Wanders, N., and Y. Wada (2015), Human and climate impacts on the 21st century hydrological drought, *Journal of Hydrology*, *526*, 208–220.

Wanders, N., D. Karssenberg, A. de Roo, S. M. de Jong, and M. F. P. Bierkens (2014), The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrol. Earth Syst. Sci.*, *18*(6), 2343–2357.

Wanders, N., Y. Wada, and H. A. J. van Lanen (2015), Global hydrological droughts in the 21st century under a changing hydrological regime, *Earth System Dynamics*, *6*(1), 1–15.

Wanders, N., S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego, and E. F. Wood (2019), Development and evaluation of a pan-european multi-model seasonal hydrological forecasting system, *Journal of Hydrometeorology*, *20*, 99–115.

Wang, A., T. J. Bohn, S. P. Mahanama, R. D. Koster, and D. P. Lettenmaier (2009), Multimodel Ensemble Reconstruction of Drought over the Continental United States, *Journal of Climate*, *22*(10), 2694–2712.

Wang, A., D. P. Lettenmaier, and J. Sheffield (2011), Soil moisture drought in china, 1950-2006, *J. Climate*, *24*(13), 3257–3271.

Wang, L., J. J. Qu, S. Zhang, X. Hao, and S. Dasgupta (2007), Soil moisture estimation using MODIS and ground measurements in eastern China, *Int. J. Remote Sens.*, *28*(6), 1413–1418.

Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe (2014), The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework, *P. Natl. Acad. Sci. USA*, *111*(9), 3228–3232.

Wassmann, R., S. Jagadish, K. Sumfleth, H. Pathak, G. Howell, A. Ismail, R. Serraj, E. Redona, R. Singh, and S. Heuer (2009), *Advances in Agronomy*, vol. 102, chap. Chapter 3 Regional Vulnerability of Climate Change Impacts on Asian Rice Production and Scope for Adaptation, pp. 91–133, Academic Press.

Weedon, G. P., S. Gomes, P. Viterbo, W. J. Shuttleworth, E. Blyth, H. Österle, J. C. Adam, N. Bellouin, O. Boucher, and M. Best (2011), Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, *Journal of Hydrometeorology*, *12*(5), 823–848.

Wells, N., S. Goddard, and M. J. Hayes (2004), A self-calibrating Palmer Drought Severity Index, *Journal of Climate*, *17*(12), 2335–2351.

Werth, S., and A. Güntner (2010), Calibration analysis for water storage variability of the global hydrological model wghm, *Hydrol. Earth Syst. Sci.*, *14*(1), 59–78.

White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity, *Econometrica1*, *48*(4), 817–838.

Wickham, H. (2007), Reshaping Data with the reshape Package, *Journal of Statistical Software*, *21*(12), 1–20.

Wickham, H. (2011), The split-apply-combine strategey for data analysis, *Journal of Statistical Software*, *40*(1), 1–29.

Wickham, H. (2016), *ggplot2 : elegrant graphics for data analysis*, XVI, 260 pp., Springer, New York.

Wilbanks, T., P. Romero Lankao, M. Bao, F. Berkhout, S. Cairncross, J.-P. Ceron, M. Kapshe, R. Muir-Wood, and R. Zapata-Marti (2007), Industry, settlement and society, in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by M. Parry, O. Canziani, J. Palutikof, P. van der Linden, and C. Hanson, pp. 357–390, Cambridge University Press, Cambridge, UK.

Wilhite, D. (2000), Drought as a natural hazard: concepts and definitions, in *Drought: A Global Assessment, Vol. I*, edited by D. Wilhite, chap. 1, pp. 3–18, Routledge, London.

Wilhite, D., and M. Glantz (1985), Understanding the drought phenomenon: The role of definitions, *Water Int.*, *10*, 111–120.

Wilhite, D. A. (Ed.) (2005), *Drought and Water Crisis : Science, Technology, and Management Issues*, 432 pp., Taylor & Francis, Boca Raton.

Wilhite, D. a., M. D. Svoboda, and M. J. Hayes (2007), Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness, *Water Resources Management*, *21*(5), 763–774.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, 3rd ed., Academic Press, Amsterdam.

WMO (2006), Drought monitoring and early warning: concepts, progress and future challenges, *Tech. Rep. 1006*, World Meteorological Organisation.

Wöhling, T., L. Samaniego, and R. Kumar (2013), Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment, *Environmental Earth Sciences*, *69*(2), 453–468.

Wood, A., and N. Mizukami (2014), Project Summary Report: CMIP5 1/8 Degree Daily Weather and VIC Hydrology Datasets for CONUS, *Tech. rep.*, B. o. R. U.S. Department of the Interior, Technical Services Center, Denver, Colorado, access date: 2017/01/24.

Wood, A. W. (2008), The University of Washington Surface Water Monitor: An experimental platform for national hydrologic assessment and prediction, in *22nd conference on hydrology*, p. 13, New Orleans.

Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophysical Research Letters*, *35*(14), n/a–n/a.

Wood, E. F. (Ed.) (1990), *Land Surface – Atmosphere Interactions for Climate Modeling: observations. models, and analysis*, vol. 12, Kluwer Academic Publishers, Dordrecht, Reprinted from Surveys in Geophysics, Nos. 1-3.

Wood, E. F. (1995), Scaling behaviour of hydrological fluxes and variables: empirical studies using a hydrological model and remote sensing data, *Hydrol. Processes*, *9*(3–4), 331–346.

Wood, E. F. (1997), Effects of soil moisture aggregation on surface evaporative fluxes, in *Journal of Hydrology*, pp. 397–412.

Wood, E. F., M. Sivapalan, K. Beven, and L. Band (1988), Effects of Spatial Variability and Scale with Implications to Hydrologic Modeling, *Journal of Hydrology*, *102*, 29–47.

Wood, E. F., M. Sivapalan, and K. Beven (1990), Similarity and scale in catchment storm response, *Reviews of Geophysics*, *28*(1), 1–18.

Wood, E. F., D. Lettenmaier, X. Liang, B. Nijssen, and S. W. Wetzel (1997), Hydrological modeling of continental-scale basins, *Annu. Rev. Earth Pl. Sc.*, *25*, 279–300.

Wood, E. F., D. P. Lettenmaier, X. Liang, D. Lohmann, A. Boone, S. Chang, F. Chen, Y. Dai, R. E. Dickinson, Q. Duan, M. Ek, Y. M. Gusev, F. Habets, P. Irannejad, R. Koster, K. E. Mitchel, O. N. Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. B. Shmakin, D. Verseghy, K. Warrach, P. Wetzel, Y. Xue, Z.-L. Yang, and Q. C. Zeng (1998), The project for intercomparison of land-surface parameterization schemes (PILPS) phase

2(c) Red-Arkansas River basin experiment: 1. Experiment description and summary intercomparisons, *Global and Planetary Change*, *19*(1-4), 115–135.

Wood, E. F., J. K. Roundy, T. J. Troy, L. P. H. van Beek, M. F. P. Bierkens, E. Blyth, A. de Roo, P. Doell, M. Ek, J. Famiglietti, D. Gochis, N. van de Giesen, P. Houser, P. R. Jaffé, S. Kollet, B. Lehner, D. P. Lettenmaier, C. Peters-Lidard, M. Sivapalan, J. Sheffield, A. Wade, and P. Whitehead (2011), Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resources Research*, *47*.

Woods, R., M. Sivapalan, and M. Duncan (1995), Investigating the representative elementary area concept: An approach based on field data, *Hydrological Processes*, *9*(3-4), 291–312.

Wösten, J. H. M., Y. A. Pachepsky, and W. J. Rawls (2001), Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, *Journal of Hydrology*, *251*(3-4), 123–150.

Xia, Y., J. Sheffield, M. B. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, and E. F. Wood (2014), Evaluation of multi-model simulated soil moisture in NLDAS-2, *J. Hydrol.*, *512*, 107–125.

Xia, Y., M. T. Hobbins, Q. Mu, and M. B. Ek (2015), Evaluation of NLDAS-2 evapotranspiration against tower flux site observations, *Hydrol. Processes*, *29*(7), 1757–1771.

Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Advances in Water Resources*, *30*(8), 1756–1774.

Yarnal, B. (1993), *Synoptic climatology in environmental analysis*, Studies in Climatology Series, Belhaven Press, London.

Yevjevich, V. (1968), Misconceptions in hydrology and their consequences, *Water Resources Research*, *4*(2), 225–232.

Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resour. Res.*, *44*(9), 1–18.

Yuan, X., and E. F. Wood (2012a), On the clustering of climate models in ensemble seasonal forecasting, *Geophysical Research Letters*, *39*(18), n/a–n/a.

Yuan, X., and E. F. Wood (2012b), Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast, *Water Resources Research*, *48*(12).

Yuan, X., and E. F. Wood (2013), Multimodel seasonal forecasting of global drought onset, *Geophysical Research Letters*, *40*(18), 4900–4905.

Yuan, X., E. F. Wood, L. Luo, and M. Pan (2011), A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction, *Geophysical Research Letters*, *38*(13).

Yuan, X., E. F. Wood, J. K. Roundy, and M. Pan (2013a), CFSv2-Based Seasonal Hydroclimatic Forecasts over the Conterminous United States, *Journal of Climate*, *26*(13), 4828–4847.

Yuan, X., E. F. Wood, N. W. Chaney, J. Sheffield, J. Kam, M. Liang, and K. Guan (2013b), Probabilistic Seasonal Forecasting of African Drought by Dynamical Models, *Journal of Hydrometeorology*, *14*(6), 1706–1720.

Yuan, X., J. K. Roundy, E. F. Wood, and J. Sheffield (2015), Seasonal Forecasting of Global Hydrologic Extremes: System Development and Evaluation over GEWEX Basins, *Bulletin of the American Meteorological Society*, *96*(11), 1895–1912.

Zacharias, S., and G. Wessolek (2007), Excluding Organic Matter Content from Pedotransfer Predictors of Soil Water Retention, *Soil Sci. Soc. Am. J.*, *71*(1), 43–50.

Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008), Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin, *J. Hydrometeorol.*, *9*(3), 535–548.

Zehe, E., H. Lee, and M. Sivapalan (2006), Dynamical process upscaling for deriving catchment scale state variables and constitutive relations for meso-scale process models, *Hydrol. Earth Sys. Sci.*, *10*, 981–996.

Zehe, E., U. Ehret, L. Pfister, T. Blume, B. Schroeder, M. Westhoff, C. Jackisch, S. J. Schymanski, M. Weiler, K. Schulz, N. Allroggen, J. Tronicke, L. van Schaik, P. Dietrich, U. Scherer, J. Eccard, V. Wulfmeyer, and A. Kleidon (2014), HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, *Hydrology and Earth System Sciences*, *18*(11), 4635–4655.

Zhu, J., and B. P. Mohanty (2002), Spatial Averaging of van Genuchten Hydraulic Parameters for Steady-State Flow in Heterogeneous Soils: A Numerical Study, *Vadose Zone J.*, *1*(2), 261–272.

Zink, M., L. Samaniego, R. Kumar, S. Thober, J. Mai, D. Schäfer, and A. Marx (2016), The German drought monitor, *Environmental Research Letters*, *11*(7).

Zink, M., R. Kumar, M. Cuntz, and L. Samaniego (2017), A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, *Hydrology and Earth System Sciences*, *21*(3), 1769–1790.

Zink, M., J. Mai, M. Cuntz, and L. Samaniego (2018), Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature, *Water Resources Research*, *54*, 2976–2998.

Zreda, M., W. Shuttleworth, X. Zeng, C. Zweck, D. Desilets, T. Franz, and R. Rosolem (2012), COSMOS: the cosmic-ray soil moisture observing system, *Hydrol. Earth Syst. Sci.*, *16*(11), 4079–4099.

Index

Abbott, M.B., 8
Abdulla, F., 34
Aerodynamic resistance, 121
Alfieri, L., 166
Andreadis, K.M., 58, 59, 63, 142
Aristotle, 3
Arnal, L., 166
Arnell, N.W., 154
Attinger, S., 33

Bárdossy, A., 19, 20
Bauer, P.A., 9
Beck, H., 233
Berg, A., 212
Bergström, S., 98
Beven, K., 8, 34, 39, 84, 116
Bierkens, M., 78, 233, 236
Binley, A.M., 39
Bjerknes, V., 7
Blöschl, G., 9, 10, 34
Blueprint, of a hydrologic model, 1st, 8
Bonan, G.B., 8, 78
Boni, G., 116
Bosshard, T., 154
Brazdil, R., 2
Bronstert, A., 8
Brynjarsdottir, J., 78
Bryson, R.A., 2
Budyko plot, 127
Buizza, R., 154
Bull of Heaven, mythical beast, 237

Chen, T.H., 58
Ciais, Ph., 212
Clark, M., 6, 84
Climate projection
    Workflow, 169
Climate projections
    Uncertainty, 212
Climatological indices, 20
Crawford, N.H., 9, 98
Crop yield modelling
    Demeaning framework, 201
    Endogenous variables, 200
    Partial effects analysis, 204
    SMI, predictor, 203
    Spatial autocorrelation, 201
Crow, W.T., 116
Cuntz, M., 9, 78

da Vinci, L., 3
Dagan, G., 78
Dai, A., 4, 154, 212
Day, G., 140
Demargne, J., 166
Dembélé, M., 10
Demuth, S., 156
Dooge, J.C., 3, 6, 9, 34, 78
Dracup, J.A, 58
Drough monitoring
    EDO, 166
    German drought monitoring, 166
    USA, 166
Drought
    2003 European event, 58
    Affected population, 217
    Area under drought forecast, 149
    Areal extent, 63
    Aridity change, 215
    Aridity estimation, 223
    Available soil water, 223
    Benchmark events, Germany, 190
    Bull of Heaven, mythical beast, 2
    CDF drought magnitude, 160
    Characteristics, 63, 156, 222
    Characteristics, sensitivity, 212

Drought Modeling and Forecasting, First edition.

COPACOGECA, loseses, 140
Curse, 2
Death toll, 4
Definition, 1
Development and recovery, forecast, 150
Dryness index, 21, 32
Duration, 63, 212
Duration, Europe, 213
Economic losses, 4
EDgE project, 218
Estimating Impacts, 15
Event identification, 69
False positives, identification, 71
Forecast, 140
Forecasting and predicting, 14
Forecasting skill, 148
Frequency, 219
Frequency, Europe, 186
Gilgamesh, Epic, 2
Global drought monitoring, 235
Hunger stones, 3
Hydrological, 154, 212
Identification, 63
Indices, 62
Indices, comparison, 67
Intensity, 63
Living conditions, 217
Low flow duration, 30
Magnitude, 63
Main modeling elements, 6
Mega-droughts, 2
Modeling, 6
Modeling across scales, 11
Modeling chain, 6, 218
Multi-model ensemble statistics, 214
Number of ISI listed articles, 5
PDF adaptation, 212
PDF drought areas, 212
PDSI, 212
Population in drought areas, 225
Prayers, 2
Projections, 212
Projections confidence, global, 154
Projections, uncertainty, 174
RI uncertainty, 158
Runoff characteristics, 22
Runoff index, 156
SAD curves, 64
SAD-curve uncertainty, 73
Severity, 64
SMI, 62, 189, 222
SMI vs. PDSI, 224
Socioeconomic losses, 186
Soil water decrease, 215
Soil water deficit, 223
Songs, 2
Spatio-temporal clustering, 212
SPI, 189
Trends runoff index, 157
Types, 1, 58
Warming sensitivity, 214
Drought monitor
GDM's aims, 187
Germany, 186
Operational chain, 187
USA, 187
Druyan. A., 1
Duckstein, L., 44, 122
Dutra, E., 140

Eagleson, P.S., 6, 233
Earth System Model, 236
EDgE, project, 87, 218
Edwards, P.N., 78
Effective parameter, 78
Efron, B., 25
Emerton, R.E., 166
Error measure, LST, 122
ESA-CCI, 98

Feynman, R., 77
Field, J.O., 58
Floods, 22
Forcing data
E-OBS, 167
Forecast
CFS v2, 140
Drought skill, 148
ECMWF, System-3, -4, 140
Ensemble average, 146
Equitable Threat Score, SMI, 146
ESP, 140
Initial conditions, hydrologic, 140
mHM, 141
NMME ensemble, 141
NMME subemsemble, 147
Skill, precipitation & temperature, 145
Skill, SMI, 144
Skill, spatio-temporal distribution, 148
SMI, 142
Subensemble selection, 143
Verification, SMI, 144
Freeze, R.A., 8

GCM, 78
Bias correction, 154, 156, 167
CMIP5 forcing data, 155, 212
Downscaling, 21
Downscaling methods, 20
ENIAC, 8
Evolution of the spatial resolution, 8
Forcing data, 155
Forescast, 140
Initial conditions, uncertainty, 154
Land-atmosphere interactions, 10
NMME ensemble, 140
Parameterization, uncertainty, 154
RCP uncertainty, 155
Scenarios, 20
The parameterization problem, 9
The weather factory, 7
Uncertainty, 156
Gelhar, L.W., 78
Generalized model, 24
Giuntoli, I., 154
Global warming
Levels 1.0 K,...,3.0 K, 222
Projected changes, 219
Gotzinger, J., 34
GRACE, 98
Correlation with mHM, global, 234
TWS anomaly, 100
Grayson, R., 9, 10
Greve, P., 212

Gudmundsson, L., 154
Gupta, H., 34

Habilitation
    Challenges, 229
    Conclusions, 231
    Objectives, 229
    Research Statements, 11
    Scientific relevance, 229
Haddeland, I., 78, 154
Harlan, R., 8
Haylock, M.R., 141
Heathcote, R.L., 2
Hempel, S., 156
Hirschi, M., 212
Horton, R., 98
Huang, J., 212
Hundecha, Y., 34
Hurst, H.E, 98
Hydrological cycle
    Mundus Subterraneus, 3
    Representation, 3
Hydrological extremes
    Uncertainty, 154
Hydrological models
    Intercomparison Project, 47
    Blueprint, 98
    CABLE, 81
    CLM, 81
    Conceptual, 34
    GCM performance, 233
    HBV, 34, 81, 155
    HBV-IWS, 35
    High resolution GCMs, 233
    HTESSEL, 81
    HYPE, 155
    JULES, 81
    LISFLOOD, 81
    Machine learning, role, 236
    mHM, 10, 36, 81, 91, 155, 168, 188
    MIKE-SHE, 34
    New generation, 236
    Noah-MP, 8, 78, 81, 168
    Parametric uncertainty, 154
    PCR-GLOBWB, 81, 91, 168
    Poor performance, reasons, 234
    Routing, 168
    SAC, 34
    Stanford Watershed Model, 8
    Statistical, 21
    Streamflow calibration, 98
    SWIM, 155
    Uncertainty, 156
    Verification, 220
    VIC, 35, 155, 168
    VIC-3L, 34
    WASIM-ETH, 34
    WaterGAP, 81, 155
    WaterGAP mHM, 91
Hydrology
    Emergence global hydrology, 233
    Grand challenge, 5
    Mesoscale basins, 34
    Misconceptions, 6
    The parameterization problem, 9
    The scale problem, 8
    The unifying theory, 10

Inverse modeling, 116
IPCC, 5, 10
    AR5, 154
    Reports, 8
ISI-MIP, 155

James, R., 212

Kircher, A., 3
Kirchner, J., 34
Kirkby, M., 8
Kirtman, B.P., 140
Kitanidis, P.K., 78
Kling–Gupta efficiency, 86
Koch, J., 116
Koster, R.D., 62
Kovats, R.S., 214
Kumar, R., 10, 33, 57, 77, 153, 156, 189

Land surface temperature, 116
Lawrimore, J., 166
Lettenmaier, D.P., 34, 59, 140
Li, Z.-L., 116
Liang, X., 78
Linsley, R.K., 9
Livneh, B., 86
LSA-SAF, 117
LSM, 8
    PLUMBER project, 81
    Porosity fields, 81

McCabe, M.F., 116
Mendoza, P., 8, 78
Merz, R., 116
Meteosat Second Generation, 117
mHM
    Basic equations, 36
    Budyko plot, German basins, 127
    Conceptualization, 36
    Crosvalidation, streamflow and LST, 126
    Efficiency over Europe, 105
    Ensemble SM, Germany, 60
    Evaluation at eddy flux stations, 132
    Evaluation Germany, 64
    Evaluation, ESA-CCI, 106
    Evaluation, FLUXNET, 106
    Evaluation, GRACE, 106
    Germany, 60
    Hierarchy of spatial scales, 37
    Land surface temperature, 119
    Local scale evaluation, 111
    Parameter sensitivity, 128
    Parametrization, 37
    Simplified energy balance, 120
    Simulated ET, 130
    Simulated vs. observed LST, 130
    SMI, 141
    Taylor diagrams, 109
    Water balance, 109
Mishra, V., 59
Mizukami, N., 83
Mo, K.C., 140
Morris, M.D., 85
MPR, 79, 234
    Transferability, ungauged basins, 53
    Across scales, 86
    Effect of the sub-grid variability, 45

Flux-matching test, 40
Flux-matching, MPR vs. SR, 49
Global parameters, 45
Land surface temperature, 121
Limitations, 87
Mass Balance, 49
Parameter Identification, 44
Parameter Regionalization, 38
PCR-GLOBWB, 91
PCR-GLOBWB, implementation, 93
Pedo-transfer functions, 35
Performance, scales, 47
Predictors, 40
Regionalization functions, 41
Scale experiments, 87
Schema, 35
Seamless parameters, 86
Spatial patterns, 46, 89
Sub-grid effects, 49
Transferability, scales, 45
Upscaling operators, 39
mRM, 168
Mueller, B., 154

Nash, J.E., 98
Natural disasters database, 4
Nearing, G., 84
Net radiation, 120
Neuman, S.P., 78
Nijzink, R., 85
Niu, G.-Y., 78
Non-parametric kernel density estimation, 142
NWP, 78

Operational forcasting
Africa, 166
EDgE PoC Copernicus C3S, 166
EFAS, 166
GLOFAS, 166
Optimization
Compromise objective functions, 122
Compromise-metric, 86
Dynamically dimensioned search, 85
Hydrologic signatures, 85
Pattern similarity, 123
Shuffled complex evolution, 85
Simulated Annealing, 85
Simulated Anneling, 44
Stratified sampling, 124

Pappenberger, F., 154
Parajka, J., 34
Parameterization
Definition, 8
Effective parameter, 34, 78
Effects of calibration, 88
Effects on model dynamics, 90
Estimation, 102
Experimental design, 102
Flux-matching, 83, 91
Flux-matching, ET flux, 91
Global parameter, 79
HRU, 9, 79
MPR, 10, 35, 83
MPR protocol, 85
Non-seamless, 81, 87
Overparameterization, 34
Post regionalization, 34
REA concept, 10
REA, role, 83
Reasons, poor performance, 78
Regionalization, 34
Scale invariant, 85
Scaling problem, 10
Seamless prediction framework, 83
seamless verification, 102
Simultaneous regionalization, 34
Spatial patterns, 89
Subgrid, 78
Techniques, 79
Transfer-function, 34
Pechlivanidis, I.G., 154
Peichl, M., 7, 195
Peters-Lidard, C.D., 6
Pielke, R., 78
PILPS, 58
Pokhrel, P., 34
Popper, K., 3, 6
Pozzi, W., 187
Precipitaion change, Europe, 220
Precipitation change, Europe, 219
Projections
Modeling chain, 167
Uncertainty, 154, 172
Uncertainty GCM/HM ratio, 158
Prudhomme, C., 154, 212

Raftery, A.E., 217
Rajaram, H., 6
Rakovec, O., 10, 83, 85, 97, 116
RCP, 154
Reed, S., 47
Regionalization
MPR, 34
MPR vs. the standard method, 40
Reichle, R., 116
Richardson, L.F., 7, 8
Robine, J.S., 4
Robine, J.-M., 186

Samaniego, L., 1, 5, 7, 10, 19, 33, 57, 77, 97, 115, 139, 142, 153, 156, 165, 185, 189, 195, 208, 211, 234
Schewe, J., 155
Schlenker, W., 196
Scholz, C.A., 2
Schultz, G., 6
Seasonal forecast
Brier Score, 175
ESP, 172
Modeling chain, 167
Skill, 172, 175
Workflow, 169
Sectoral Climate Impact Indicators, 169
Seibert, J., 34
Sellers, P.J., 78
Seneviratne, S.I, 212
Sensible heat, 120
Sensitivity analysis, 85
Sheffield, J., 59, 62, 140, 154, 166
Sherman, L.K., 98
Shukla, S., 140
Silvestro, F., 116
Sivapalan, M., 9
SMI, 59, 198

Estimation, 62, 222
Persistence, 74
Seamless across spatial scales, 232
Trends, 74
Uncertainty, 68
Soil moisture
Reconstruction, Germany, 65, 189
Seasonality, 66
Soil moisture drought index, 59
Stahl, K., 20, 63
Statistical modeling
Crop yield, 196
Crop yield predictors, 198
Explanatory variables, 25
Heteroscedastic errors, 31
Hypothesis, 26
Non-parametric test, 25, 157
Parameter estimation, 25
Variable combinations, 26
Stensrud, D.J., 8
Stisen, S., 116
Stone, R., 2
Svoboda, M., 187

Taylor, A.L., 172
Temperature change, Europe, 219
Therrell, M.D., 2
Thober, S., 10, 139, 165, 198, 211
Todini, E., 189
Trenberth, K.E., 154, 212
Trigo, I.F., 116
Trnka, M., 187
Troy, T.J., 35

Uncertainty
Drought duration, 159
Drought magnitude, 159

van Lanen, H.A.J., 212
Vereecken, H., 116
Vetter, T., 154
Vidal, J.-P., 58, 156

Wöhling, T., 10
Wagener, T., 34
Wanders, N., 78, 172
Wang, A., 59, 140, 159
WATCH dataset, 154
Weather
Circulation patterns, 21
ENSO, 140
Prediction, 7
Wilbanks, T.J., 5, 217
Wilhite, D.A., 58
Wilks, D., 62
WMO, 1
Rain gauge density, 106
Wood, A.W., 187
Wood, E.F., 5, 10, 59, 78, 87, 140, 142, 236
Woods, R., 10

Yield data, Germany, 197
Yuan, X., 140

Zink, M., 4, 10, 57, 115, 166, 185, 198