Accepted Manuscript

Combined uncertainty of hydrological model complexity and satellite-based forcing data evaluated in two data-scarce semi-arid catchments in Ethiopia

Malte Knoche, Christian Fischer, Eric Pohl, Peter Krause, Ralf Merz

Please cite this article as: Knoche, M., Fischer, C., Pohl, E., Krause, P., Merz, R., Combined uncertainty of hydrological model complexity and satellite-based forcing data evaluated in two data-scarce semi-arid catchments in Ethiopia, *Journal of Hydrology* (2014), doi: http://dx.doi.org/10.1016/j.jhydrol.2014.10.003
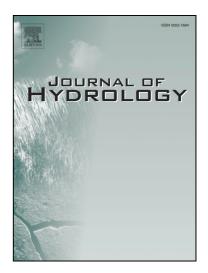
Title:

Combined uncertainty of hydrological model complexity and satellite-based forcing data evaluated in two data-scarce semi-arid catchments in Ethiopia

Authors:

Malte Knoche[1], Christian Fischer[2], Eric Pohl[3], Peter Krause[4] and Ralf Merz[1]

[1] Department Catchment Hydrology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser-Str. 4, D-06120 Halle, Germany

[2] Department of Geoinformatics, Hydrology and Modelling, Friedrich Schiller University of Jena, Löbdergraben 32, D-07743 Jena, Germany

[3] Remote Sensing Group, Geology Institute, Bergakademie Freiberg University of Mining and Technology, Bernhard-v-Cotta Str. 2, D-09599 Freiberg, Germany

[4] Thüringen State Institute for Environment and Geology, Göschwitzer Str. 41, D-07745 Jena, Germany

[1] Department Catchment Hydrology, Helmholtz Centre for Environmental Research – UFZ, Theodor-Lieser-Str. 4, D-06120 Halle, Germany

Corresponding author: M. Knoche (Helmholtz UFZ) malte.knoche@ufz.de;

phone: +49 345 558 5490

Co-author email addresses: christian.fischer.2@uni-jena.de; eric.pohl@geo.tu-freiberg.de; peter.krause@tlug.thueringen.de; ralf.merz@ufz.de

Abstract.

In water resources modeling, meteorological data scarcity can be compensated by various global data sets, but those data sets can differ tremendously. In the literature, hydrological models of differing complexity are proposed for estimating the water resources of semi-arid catchments, and also to evaluate rainfall data sets. The goal of this paper is to provide a joint analysis of modeling uncertainty due to different input data and increasing model complexity. Impacts of mutually concealed uncertainties on model performance and model outputs are exemplified in two data sparse semi-arid catchments in Ethiopia. We applied a semi-distributed and a fully distributed hydrological model, having different levels of complexity. Three different satellite-based rainfall data sets and two temperature products were used as model inputs. The semi-distributed model demonstrated good validation performances, while the fully distributed model was more sensitive to data uncertainties. The application of TRMM version 6 completely failed and the high-resolution CMORPH precipitation estimate outperformed TRMM version 7. In contrast, the use of high-resolution temperature data did not improve the model results. The large differences in remotely sensed input data were buffered inside the hydrological models. Therefore, data set evaluations regarding only the simulated hydrographs were less meaningful. In contrast, the investigation of parameter evolution and distributed outputs' variability appeared to be a valuable tool to uncover the interdependencies of data and model uncertainties. We suggest this procedure to be applied by default in water resources estimations that are affected by data scarcity, but especially when data sets are evaluated using hydrological models. Our case study demonstrates that estimations of groundwater recharge and actual evapotranspiration vary largely, depending on the applied data sets and models. The joint analysis reveals large interdependencies between data and model evaluations. It shows that traditional studies focusing only on one part of uncertainty, either the input uncertainty or the uncertainty arising from the choice of model

structure are limited in their explanatory power of the modeling performance, particularly in poorly gauged regions.

Keywords:

hydrological modeling; data scarcity; uncertainty; remote sensing precipitation; model complexity; East Africa

1. Introduction

East Africa is often threatened by severe water scarcity. Millions rely on unsecure availability of fresh-water. This problem is further enhanced over the last decade by an eminent economic growth [Shiklomanov, 2000; AfDB et al., 2013]. The resulting water demand and the ongoing growth of population ask for reliable estimations of water resources. However, the hydro-meteorological infrastructure is old and sparsely distributed across the country with its complex topography and subsurface. Hydrological models are a common tool for water resources management in data sparse regions such as East Africa. Modelers are challenged by the following questions: (i) which climate data should be used to drive the model, (ii) what model should be used and (iii) how to verify that data and model are sufficient to be used for water resources management. The first two questions have individually been discussed in numerous publications, illustrating a high complexity of the depicted problem:

i) Which climate data should be used?

When ground-based rainfall data are missing or rare, satellite based rainfall estimates (e.g. TRMM, CMORPH) are frequently applied to drive hydrological models [e.g. Velpuri et al., 2012]. However, those data sets are affected by high uncertainties resulting from physical sensor limitations, limited space-time coverage and spatial resolution [Prigent, 2010; Yong et al., 2014]. The known uncertainties raise the question of the extent of their impact on water resource estimations. Therefore, data sets are sometimes indirectly tested with hydrological models by comparing the model performances [e.g. Xue et al., 2013]. This partly leads to

critical judgments on the applicability of satellite based precipitation estimates [e.g. Meng et al., 2014].

There are numerous studies in the literature analyzing the impacts of different remotely sensed rainfall products in hydrological modeling, including some studies focusing on poorly gauged African river basins. E.g. Khan et al. [2011] reported underestimations of simulated runoff using TRMM3B42 V6 when modeling the Nzoia Basin, Lake Victoria, Kenya. Gebregiorgis et al. [2012] compared three satellite rainfall estimates by modeling the Mississippi River Basin. They found a high correlation (0.85) between runoff error and the hit bias of the precipitation data and furthermore a correlation of 0.75 between missed precipitation and soil moisture error. Bitew and Gebremichael [2011] discovered CMORPH and the TRMM real-time product perform better than post-processed TRMM6 and infrared-based rainfall estimates in an Ethiopian highland's catchment. Milzow et al. [2011] used the SWAT model to find out major differences between TRMM, FEWS (Famine early warning system) and ERA-Interim rainfall data in the Okavango basin. Stisen and Sandholt [2010] compared different rainfall data sets inside the MIKE SHE model in the Senegal River Basin. Their study reveals that satellite data used in pre-calibrated models (forced by rain gauges) leads to weak performances. In contrast, models calibrated and forced by satellite inputs have Nash-Sutcliffe efficiencies (NSE) between 0.63 and 0.87. Therefore, the judgment of input data sets which is based solely on the assessment of simulated runoff after calibration might be insufficient because other water balance components might be affected negatively. However, water balance components other than runoff remain mostly disregarded when data sets are tested directly in hydrological models. An exception is represented by Stisen et al. [2008] who perform a comparison of infrared-based rainfall estimates with rain gauge data within the MIKE-SHE model. They discovered that actual evapotranspiration is highly sensitive to different rainfall inputs in the Senegal Basin.

Temperature estimates are also required to drive hydrological models. The use of remote sensing products and reanalysis data are an appealing approach in data scarce regions, but large uncertainties may accompany these data sets. For example, Becker et al. [2010] applied ECMWF temperatures with a spatial resolution of 0.75° (ca. 6800 km² per pixel) in modeling the East African Rift System's hydrology. In contrast, Deus et al. [2013] applied MODIS land surface temperatures (LST) in 5.6 km² resolution in the data-scarce Lake Manyara Basin, Rift Valley, Tanzania. The first application cannot cover the spatial heterogeneity of air temperature (AT), especially over complex terrain, while the latter application raises the question whether LST is an adequate replacement for AT in water resource estimations. The impact of largely differing spatial resolutions and temperature magnitudes remains unclear. Alternatively, air temperatures can be modeled from LST, but with a significantly higher effort [e.g. Cristóbal et al., 2008] since the derivation of AT from LST "is far from straight forward" as Vancutsem et al. [2010] figured out in East Africa.

In contrast to precipitation data, evaluations of temperature data sets by using hydrological models are unusual. E.g. Wang et al. [2009] and Wang et al. [2011] evaluated MODIS LST and GLDAS AT by using the WEB-DHM distributed model in a semi-arid basin. The influence of other uncertainties (model, precipitation) has not been discussed.

ii) What model should be used?

In addition to the need for meteorological data, finding an adequate model structure for data-sparse semi-arid catchments reveals the agony of choice. Subsurface complexity and consequently the model complexity correspond with scarcity of soil and catchment information. There are diverging approaches to set up models for such an environment. Collick et al. [2009] developed a simple semi-distributed model for semi-arid regions under monsoonal influence (Ethiopian highlands). Güntner and Bronstert [2004] highlight the importance of taking into account the re-infiltration processes in large scale modeling of semi-arid catchments (Brazilian Caatinga). They suggest the application of fully distributed models

to account for lateral redistribution processes. Love et al. [2011] confirm these findings and show their semi-distributed model is not able to reproduce hydrology of more ephemeral and drier catchments. They could not extend a Nash-Sutcliffe efficiency of 0.3 when applying the HBVx model in 15 semi-arid catchments in Zimbabwe. In contrast, Reed et al. [2004] found distributed models not to perform better than lumped models in the data-rich central USA. However, detailed studies on what is an appropriate model structure and model complexity for data-sparse semi-arid regions are rare and focus on hydrograph simulation only [e.g. Jothityangkoon et al., 2001; Ghavidelfar et al., 2011]. Model choice is often affected by other factors. For example, Van Griensven et al. [2012] carried out an extensive review of 20 applications of the Soil and Water Assessment Tool (SWAT) in East Africa and concluded that the main reason for the application of SWAT by numerous scientists was much more due to its handy applicability inside GIS software than its physical representation of the depicted catchments.

iii) How to verify data and model, both?

Although the limitations of hydrological models in data-sparse semi-arid catchments are clearly mentioned by many authors [e.g. Love et al., 2011], strict calibration and validation procedures [e.g. Klemes, 1986] are often missing or weakly documented in model applications in East African catchments [Van Griensven et al., 2012].

To our knowledge, what is missing in the literature, are analyses of the combined effect of input data and model structure on the performance and outputs of hydrological models in data sparse semi-arid catchments. Some studies focus exclusively on data set evaluations, others on model evaluation. However, the compensation of data set uncertainty by hydrological models and the concealment of model uncertainty by data uncertainties remain underexposed in the meantime. The objective of this paper is to fill this gap and to combine data set evaluation and model evaluation. Uncertainties resulting from data of land cover or subsurface properties are ignored in this study (see Branger et al., 2013).

On the example of two data-sparse catchments in Ethiopia, we specifically address two scientific questions: A) What is the impact of different input data-model-combinations on the efficiency of reproducing runoff? B) What is the impact of different input data-model-combinations on the evolution of model parameters and the spatial patterns of groundwater recharge and evapotranspiration?

The paper is organized as follows. First, we present an overview on the catchment characteristics and different remote sensing rainfall and temperature products used to drive the hydrological models in this study. Following is a section on model structures, parameter derivation and calibration approaches used in this study. We then present an evaluation of how different input data-model type combination can reproduce observed runoff. Finally, we analyze the evolution of parameters and differences in the model outputs, such as groundwater recharge. The paper concludes with remarks on hydrological modeling in data sparse regions.

[-- Figure 1 --]

2. Study area

The study was carried out in the upper Awash River Basin and the Kessem River Basin, located in the central Main Ethiopian Rift, between 7.75° N and 9.75° N, and 37.75° E and 40.75° E (Figure 1). Here, the Great African Rift Valley intersects the Ethiopian highlands. The Kessem River passes an altitude drop from more than 2000 m on a flow length of 130 km (Figure 1). The upper Awash River passes an altitude drop of ca. 1000 m on 100 km flow length and therefore is moderately steep in comparison to the Kessem River Basin. The catchments have sizes of 4531 km² (upper Awash) and 2950 km² (Kessem) and their cross profiles and river profiles are shown in Figure 1.

The climate is mostly influenced by the movement of the intertropical convergence zone (ITCZ). During its movement northwards in March/April and its retreat southwards, ITCZ

creates two rainy seasons, a shorter one around March, and a longer one between June and September, which partly fall into one longer rain-season [Schultz, 2005; Seleshi and Zanke 2004]. The time between October and March is a dry season. Semi-arid to arid conditions prevail in the Rift Valley. In contrast, the highlands partly receive more than 1600 mm of rainfall in ca. six months per year [Dinku et al., 2008].

The Kessem River has recently been dammed (2011) upstream of Melka Werer town for the production of hydropower and irrigation of sugar cane plantations. In the present study we study the catchment in its undammed status (until 2010).

The dominating vegetation in the study area is part of the Seasonal Tropics ecozone. At high altitudes this ecozone forms individual humid alpine microclimates. The Tropical Arid Lands ecozone predominates at low levels in the Rift. Primary vegetation underlies a strong anthropogenic impact. Vertisols are the predominant soil type of the upper Awash Basin, covering more than 80 percent of the catchment. In contrast, approx. 80 percent of the steep Kessem basin is covered by Leptosols, which are thin-layered, immature soils [FAO et al., 2009].

The escarpments consist of mainly basaltic and rhyolithic fractured aquifers, where the rifting process formed large tilted block formations. Their flat plateaus are slightly tilted in opposite direction compared to the altitude drop [Kurz et al., 2007; Kebede et al., 2007] and partly force surface water to run parallel to the rift. In the Rift Valley the geological setting is very heterogeneous. Ignimbrites on top of alluvial and lacustrine sediments form the deepest aquifers. Shallow aquifers lie mostly in quaternary basaltic rocks [Kebede et al. 2007; Bretzler et al. 2011]. The structure geology is characterized by numerous rift-parallel as well as transversal faults and fissures [Kurz et al., 2007; Kebede et al., 2007].

[-- Figure 2 --]

3. Forcing data

The study area is characterized by data scarcity concerning meteorological time series. The data of three meteorological stations can be accessed from the Global Historical Climatology Network (GHCN). More ground-based data might be available on request [ethiomet.gov.et; see Dinku et al., 2008], but we were not successful in receiving those data. However, we focused on testing open access meteorological data because these data are frequently used in many studies.

We evaluated three remote sensing-based rainfall products in two hydrological models: TRMM3B42-daily version 6, TRMM3B42-daily version 7 and CMORPH in the 8 km resolution (hereinafter called: TRMM6, TRMM7, and CMORPH; tab. 1). We have focused on microwave-based estimates because those have been reported to give more realistic results than infrared-based rainfall estimates [Bitew and Gebremichael, 2011].

The TRMM6 data set, provided by NASA and JAXA since 1998, is a composition of precipitation-related passive microwave sensor data and geosynchronous earth orbit infrared sensor data, calibrated on a global scale using rain gauge stations [Huffman et al. 2007]. For TRMM7 the data set calibration has been improved by using additional satellite retrievals. The NOAA Climate Prediction Center Morphing Technique (CMORPH) data set is also produced based on infrared and passive microwave data and has a spatial resolution of 8km at the equator [Joyce et al. 2004]. We accumulated hourly rainfall to daily rainfall. CMORPH data is available since December 2002. Romilly et al. [2011] observed a generally good performance of the 0.25° CMORPH product in Ethiopia, with an overall underestimation of 11%.

In addition, we evaluated two different temperature products: MOD11C1 Version 4.1 land surface temperature [Wan 1999, 2008] and the air temperature from the Global Land Data Assimilation System (GLDAS) NOAH land surface model [Rodell et al. 2004]. Both data sets are available from 24th February 2000 onwards (tab. 1).

MOD11C1 is a remote sensing product of the Moderate-resolution Imaging Spectroradiometer (MODIS) instrument mounted on board the Terra satellite and has a resolution of 0.05° (see Wan [1999, 2008] for a detailed description). We applied version 4.1 which has been documented to be better in arid and semi-arid regions than the latest version 5 [Hulley and Hook 2009]. The day-LST is used as daily maximum temperature and the night-LST as daily minimum temperature, respectively. Daily mean temperature was calculated as average of both extremes. The data set has detailed quality assurance flags due to different sources of distortions (mainly cloud contaminations). We used the "Mandatory QA flag" to reject all pixels without good quality. Resulting gaps in the time series were closed by applying a linear time series interpolation. In general, LST must be seen as a canopy temperature in vegetated areas and a soil or rock surface temperature in unvegetated areas [Wan, 2008].

The GLDAS NOAH data set is provided in a spatial resolution of 0.25 degrees in 3-hourly steps. Daily minimum, maximum and mean temperatures were calculated from the 3-hourly data. Both temperature data sets are hereinafter designated by the short-terms "MODIS" and "GLDAS".

Specific humidity and wind speed were also obtained from the GLDAS Noah Land Surface Model L4. Relative humidity was calculated from specific humidity according to equations, provided by the LP DAAC (Land Processes Distributed Active Archive Center) on request (appendix equations A1 – A4), using further information available in the GLDAS NOAH data set. The daily sunshine duration was derived from the ERA-INTERIM reanalysis data set [Dee et al., 2011].

Table 1: Meteorological forcing data sets evaluated in this study

variable	data set	resolution		availability	Citation	access
		space	time			
precipitation	TRMM-3B42 V6	0.25°	3 hours	1-1998	Huffman et al. [2007]	http://mirador.gsfc.nasa.gov
precipitation	TRMM-3B42 V7	0.25°	3 hours	1-1998	Huffman et al. [2007]	http://mirador.gsfc.nasa.gov
precipitation	CMORPH	0.07277°	1 hour	12-2002	Joyce et al. [2004]	Robert Joyce
land surface temperature	MOD11C1	0.05°	day, night	24-02-2000	Wan [1999, 2008]	http://reverb.echo.nasa.gov/
air temperature	GLDAS NOAH L4	0.25°	3 hours	24-02-2000	Rodell et al. [2004]	http://mirador.gsfc.nasa.gov
relative humidity, wind speed	GLDAS NOAH L4	0.25°	3 hours	24-02-2000	Rodell et al. [2004]	http://mirador.gsfc.nasa.gov
sunshine duration	ERA-INTERIM	0.75°	1 day	1-1979	Dee et al. [2011]	http://data-portal.ecmwf.int

4. Model structures and set-up

We set up two different conceptual hydrological models to study the effect of model complexity on performance, parameter evolution, and distributed outputs. The two models (J2000, J2000g) differ in the conceptualization of hydrological processes, distribution concepts and linkage of their process components, as illustrated in Figure 3.

Both models represent the catchment by distributed spatial model units. We applied gridded model units with a cell-size of 0.02° by 0.02° (~2200m) using Nearest Neighborhood Resampling of the raster scenes described in section 5. Actual evapotranspiration (AET) is calculated by reducing potential ET [Allen et al. 1998] depending on the soil moisture condition and a calibration parameter.

J2000 is a spatially fully distributed conceptual hydrological model [Krause, 2002] using a routing topology to distribute lateral and surface water budgets between spatial model units along the topographical gradient. Its process-based soil water balance module functions as the central "regulation and distribution system" [Krause et al., 2006], and mutually interacts with nearly all other J2000 process modules (Figure 3). Spatial model units contain two soil storages: The mid-size pore storage (MPS) represents the effective field capacity water budget which is reduced by the AET only. The large pore storage (LPS) cannot hold water against

gravity and therefore, "is considered as the source of all subsurface flow processes in the J2000 model" [Krause et al., 2006]. Infiltration water is distributed to the MPS and the LPS based on a distribution coefficient until these storages are filled or the maximal infiltration rate is reached. Infiltration excess water is stored as depression storage (DPS) at the surface. When DPS is exceeded, surface runoff is generated and routed to the adjacent downslope spatial unit. The LPS outflow is distributed into lateral runoff and percolation depending on the slope and a calibration parameter. There are two groundwater storages for each spatial model unit, one having a quick hydrological reaction and one having longer residence times. The percolation water is distributed between the two groundwater storages depending on a calibration parameter and the slope [Krause et al., 2006]. The J2000 runoff concentration and flood routing is calculated for the spatial model units and a network of river reaches. Lateral flows calculated for each grid cell are passed to downslope grid cells until a river reach is connected, where the lateral runoff is transmitted to the streamflow budget. "Flood routing in the river network is calculated by a simplified kinematic wave approach, using Manning's formula to calculate flow velocity" [Krause et al., 2006].

[-- Figure 3--]

The J2000g model is a simplified version of the J2000 model with a significantly reduced number of calibration parameters [Krause and Hanisch, 2009]. It is a semi-distributed conceptual hydrological model and does not account for lateral flow processes, nor has a streamflow routing (Figure 3). Precipitation is transferred to a simple soil water storage (SWS) device, without simulation of interception and infiltration processes. The storage capacity of each model unit results from the field capacity of the dominant soil type of that unit. The soil water budget can only be depleted by evapotranspiration, which is calculated similar to the J2000 model. Excess water is partitioned into surface runoff and percolation based on the slope, the underlying hydrogeological unit, and a calibration parameter.

Percolation is partitioned between two groundwater storages by a distribution parameter. The spatial modeling units of J2000g are not connected with each other. Therefore, the total streamflow at the outlet of a catchment results from the summation of the direct runoff and the baseflow from each modeling unit [Krause and Hanisch, 2009]. A description of all calibration parameters and its mathematical implementation is given in table A1 (J2000) and table A2 (J2000g) in the appendix.

5. Parameter data

The models require several data sets about the physical characteristics of the watershed. For the extraction of elevation, aspect, slope, flow directions, river reaches and the topological routing the 30 by 30 m digital elevation model ASTER GDEM was used. To extract this information the MATLAB-based TopoToolbox [Schwanghart and Kuhn, 2010] was modified. Leaf area index and albedo were obtained from the MODIS land product subsets platform [ORNL DAAC, 2012]. The stomata resistances were taken out of several literature resources [Heilmeier et al., 2007; Lambin, 1999; Yucel, 2006] and the plant parameter database "PlaPaDa" [Breuer et al., 2003]. Based on this information, the annual biome specific variation was adapted according to Dorman and Sellers [1989]. Individual land cover classes were derived from the MODIS MCD12Q1 version 5 IGBP global vegetation classification scheme of the year 2005 [LP DAAC, 2012].

Soil physical properties to be implemented into J2000 are the field capacity, minimal and maximal hydraulic conductivity and air capacity. All of these properties are provided by the Harmonized World Soil Database [FAO et al., 2009], except of the air capacity, which had to be estimated out of the pore volume. The simpler model J2000g only requires the field capacity.

In the Hydrogeological Map of Ethiopia [Tesfaye, 1988] aquifers are divided into sedimentary rocks, metamorphic rocks, and volcanic rocks and sediments, classified into three levels of

productivity. This rough information was complemented by more recent and more detailed publications [Ayenew et al., 2008a; Ayenew et al., 2008b; Bretzler et al., 2011; Demlie et al., 2008; Kebede et al. 2007]. Storage volume and a storage coefficient were estimated for a fast and a slow groundwater layer in the J2000 model. For the J2000g model the same data source was used to state a maximal percolation rate for each aquifer type. Due to the lack of more precise aquifer information, we focused on a qualitative parameterization, estimating the differences between different aquifer types, more than their absolute properties. Recession coefficients were automatically estimated during the calibration.

6. Calibration and validation

Observed discharge from the hydrometric stations (Figure 1) Melka Kunture (Awash River) and Awara Melka (Kessem River), which was provided by the Ethiopian Ministry of Water Resources, is used to calibrate the models. We generated rating curves by applying non-linear regressions between cross section area (A [m²]) and water level (WL [m]), and between flow velocity (V [m/s]) and water level. These regressions were used to calculate continuous daily observed discharges using the following equation:

$$Q = A(WL) * V(WL) \left[\frac{m^3}{s}\right]$$ Eq. (1)

Such rating curves can be subject of large uncertainties [Shao et al., 2014]. Therefore, the correlations of the regressions used for the rating curves are shown in table 2.

Table 2: Correlations of the regressions used for the rating curves

regression between		R^2	
X	Y	Awash (n=64)	Kessem (n=115)
water level [m]	cross section area [m²]	0.94	0.79
water level [m]	velocity [m/s]	0.63	0.76
measured flow [m³/s]	flow based on rating curve [m³/s]	0.99	0.85

The correlation between measured flows and those calculated based on a rating curve show a very good correlation for the Awash River gauge and a moderate correlation for the Kessem River gauge (tab. 2). The rating curves of both gauges rely on numerous single measurements (Awash: 64; Kessem: 115), which is why we consider the uncertainties to be acceptable. The time periods of calibration and validation are shown in table 3. The start date is defined by the start date of the MODIS and GLDAS temperatures we have used. For automatic calibration we used the Shuffled Complex Evolution (SCE) optimization algorithm [Duan et al., 1993] with a maximum number of iterations of 50,000, a worst acceptable improvement of 0.01 in the last 10 iterations and a minimal geometric population range of $1*10^{-5}$. The Nash-Sutcliffe efficiency (hereinafter called NSE) was chosen as efficiency criterion to be maximized during the optimization.

The calibration parameters of both models are succinctly described in tables A1 and A2. All eight parameters of the J2000g model have been optimized during calibration. For the J2000 model, we selected 20 out of 22 parameters for auto-calibration. The two initial groundwater storage parameters were treated as fixed parameters, to ensure that all data-model combinations start with the same initialization.

Both models forced by TRMM6, TRMM7 and GLDAS AT have been cross-validated in the time-periods T_1 and T_2 (tab. 3). The calibration of the models using CMORPH data had to start at 1st December 2002 (tab. 3). Therefore cross-validations for the model setups with CMORPH were not possible due to insufficient temporal coverage. In a second step, all data-model combinations, including MODIS LST and CMORPH, were calibrated over the complete available runoff time series, hereinafter called "T_{all} calibration" (tab. 3).

Table 3: Calibration and validation time-frames

	T_1	gap	T_2	T_{all}
Awash	24th Feb 2000 – 30th Nov 2002	No	1st Dec 2002 – 16th Oct 2004	$T_1 + T_2$
Kessem	24th Feb 2000 – 19th Feb 2002	137 days	7th Jul 2002 - 1st May 2004	$T_1 + T_2$
CMORPH$_{Awash}$	No	No	1st Dec 2002 – 16th Oct 2004	T_2 only
CMORPH$_{Kessem}$	No	No	1st Dec 2002 – 1st May 2004	T_2 only

In addition to NSE, the volume error (VE), the modified Kling-Gupta Efficiency (KGE'), and the average water balance components were depicted for evaluating the model performance. The KGE' criterion allows decomposing model efficiency into the correlation coefficient r, the bias ratio β and the variability ratio γ, according to Kling et al. [2012]. To evaluate the plausibility of the groundwater recharge simulations our results were compared to values based on the Chloride Mass Balance Method (CMBM) [Demlie et al. 2007]. Plausibility of AET simulations was evaluated by using the satellite-based MOD16A2 data set [ORNL DAAC, 2012; Mu et al., 2007; 2011], which provides estimations of actual evapotranspiration in a very high resolution.

7. Results and Discussion

7.1 Comparison of temperature data sets

On average, the MODIS LST is 4.4 °C (Awash) and 5.5°C (Kessem) higher than GLDAS AT (Figure 4). The long-term average of daily maximum LST is 8°C higher than the GLDAS daily maximum temperatures in both basins. Daily minimum temperatures of MODIS are on average 1.4 (Awash) and 2.3 °C (Kessem) higher than the GLDAS daily minimum temperatures. The daily and seasonal variance of MODIS LST is significantly higher than the one of GLDAS AT. For the temperature minima the GLDAS data set shows higher seasonal variances (Figure 4). In a summary, the temperature data sets differ in daily mean, maximum and minimum temperatures and the spatial patterns of long-term averages. Both data sets are

inconsistent and the concise spatial discretization of the MODIS LST data (Figure 2, bottom row) is contradicted by its absolute over-estimation in comparison to re-analyses air temperatures.

[-- Figure 4 --]

7.2 Comparison of rainfall data sets

By comparing space and time distributions of three rainfall data sets we could identify four types of differences that may impact on hydrological modeling: a) Different spatial distributions of long-term mean values (Figure 2), where CMORPH reveals a more concise spatial gradient than both TRMM products. b) Differences in long-term annual rainfall volumes (Figure 5): TRMM7 provides ca. 1.5-fold precipitation amounts than TRMM6 and CMORPH. c) TRMM6 shows rainfall events were no runoff is observed (Figures 6, 7, 8). d) Different temporal trends (Figure 5): TRMM6 shows a strongly decreasing temporal trend in precipitation, while TRMM7 does not and CMORPH outlines a slightly decreasing temporal trend. Furthermore, maxima and minima of annual rainfall amounts occur in different years, depending on the data set which is used. Therefore, the modeler's choice of a specific rainfall data set might highly affect the resulting water resource estimations. It is surprising that large differences occur, although the three data sets consist of more or less the same data sources [Joyce et al. 2004; Huffman et al. 2007].

[-- Figure 5 --]

7.3 Cross-validation

A judgment of what is an adequate rainfall data set is difficult, when ground-based rainfall measurements are missing or rare. The comparison of different data sets (Figures 2, 5) raises the question, how biased data sets can (easily) be identified. Here, we carry out bias identification by cross-validation of the hydrological models forced by different data sets (tab.

4). The CMORPH-driven simulations are not included in this cross-validation because their time series are too short (tab. 3).

[-- Figure 6 --]

<u>Awash Basin:</u> When applying TRMM6, performances decline to partly negative efficiencies during validation. Using TRMM7, the good performances during calibration can be reproduced during validation when applying the semi-distributed J2000g model. This is not true when applying the J2000 model with TRMM7, but validation efficiencies of 0.61 and 0.38 are still acceptable in comparison to the TRMM6-based model validation, which completely fails (tab. 4). Volume errors are ranging from -3.2% to 19.4% in the calibration period and from 26% up to 106% in the validation period. Only the application of TRMM7 inside the J2000g model has an increase of volume errors of less than 10% from calibration to validation (tab. 4).

The KGE' performances during validation of TRMM6-based models have poor bias ratios, while correlation coefficients and variability ratios show rather good performances in the Awash basin. Therefore, the negative temporal trend of TRMM6 can be identified as bias. Visual interpretation of the J2000g cross-validations (Figure 6) allows for decomposition of error sources. Applying TRMM6, runoff is over-estimated in the validation time-frame T1, and under-estimated in the validation time-frame T2 in the Awash basin. Here, the negative temporal trend of TRMM6 (Figure 5) causes model validation to fail. In contrast, the TRMM7-based simulated hydrographs of the calibration and validation are much closer to each other in the Awash basin.

Table 4: Cross-validation of two models in two basins forced by TRMM6 and TRMM7 using split runoff time series

Awash Basin		calibration							validation						
model	Precipitation	time	NSE	VE [%]	KGE'	r	β	γ	time	NSE	VE [%]	KGE'	r	β	γ
J2000	TRMM V6	t1	0.81	-12.1	0.84	0.91	0.88	0.98	t2	0.21	90.7	0.26	0.63	0.38	1.16
J2000	TRMM V6	t2	0.73	-3.2	0.83	0.86	0.97	0.91	t1	0.13	-61.6	0.06	0.89	1.91	0.78
J2000g	TRMM V6	t1	0.82	4.5	0.87	0.91	1.04	0.92	t2	0.36	106.3	0.39	0.70	0.48	1.10
J2000g	TRMM V6	t2	0.72	10.4	0.78	0.86	1.10	0.87	t1	-0.21	-52.1	-0.10	0.84	2.06	0.75
J2000	TRMM V7	t1	0.77	16.4	0.81	0.91	1.16	0.99	t2	0.61	76.4	0.62	0.82	0.67	1.03
J2000	TRMM V7	t2	0.74	6.4	0.87	0.88	1.06	1.01	t1	0.38	-33.2	0.22	0.89	1.76	0.92
J2000g	TRMM V7	t1	0.79	16.3	0.81	0.91	1.16	0.96	t2	0.79	26.1	0.75	0.92	0.77	0.95
J2000g	TRMM V7	t2	0.78	-19.4	0.77	0.92	0.81	0.90	t1	0.76	-22.85	0.68	0.90	1.26	0.85
Kessem Basin		calibration							validation						
model	Precipitation	time	NSE	VE [%]	KGE'	r	β	γ	time	NSE	VE [%]	KGE'	r	β	γ
J2000	TRMM V6	t1	0.52	-3.5	0.62	0.72	0.96	0.74	t2	0.25	74.0	0.40	0.55	0.61	1.08
J2000	TRMM V6	t2	0.63	-11.1	0.71	0.80	0.89	0.82	t1	0.1	-39.2	-0.01	0.51	1.74	0.51
J2000g	TRMM V6	t1	0.53	-30	0.59	0.75	0.70	1.11	t2	0.01	109.3	-0.89	0.44	0.21	2.62
J2000g	TRMM V6	t2	0.62	-8.4	0.76	0.79	0.92	0.92	t1	0	-78.9	-0.28	0.55	2.09	0.50
J2000	TRMM V7	t1	0.64	-5.6	0.74	0.80	0.94	0.84	t2	0.14	56.6	0.36	0.44	0.75	0.81
J2000	TRMM V7	t2	0.58	-20.6	0.67	0.77	0.79	0.90	t1	-0.1	-24.8	0.11	0.44	1.57	0.60
J2000g	TRMM V7	t1	0.56	-30	0.61	0.75	0.70	1.07	t2	-0.03	69.7	-0.09	0.40	0.19	1.41
J2000g	TRMM V7	t2	0.53	-22.9	0.65	0.74	0.77	0.96	t1	-0.14	-80.8	0.03	0.46	1.70	0.59

[20]

Kessem Basin: The performances range from 0.52 to 0.64 NSE during calibration and -0.14 to 0.25 NSE during validation and TRMM6 and TRMM7 perform similarly (tab. 4). None of the data-model combinations can be validated successfully. During calibration volume errors range from -3.5 to -30 %, and are all negative. The volume errors during validation are high, in the range of -24.8 to 109.3 %. Simulated temporal dynamics in the validation period are having weak correlation coefficients of 0.4 to 0.55 (tab. 4). Bias ratio and variability ratio also show values being far away from unity, making a further decomposition of error sources difficult.

Visual interpretation (Figure 6) shows a major timing error between observed and simulated runoff in the Kessem basin in 2003, which leads to an unreliable calibration in the time-frame T2. The resulting unrealistic recession curves are also visible during validation in T1, highlighted by the horizontal arrows (Figure 6). Furthermore, TRMM6 causes runoff peaks where no runoff events where observed (diagonal arrows) in both basins.

Summarizing the cross-validations, just one out of four data-model combinations achieves good results in only one of two basins. This is the application of TRMM7 inside the semi-distributed J2000g model in the Awash catchment (tab. 4). TRMM6 is clearly identified as biased by a negative temporal trend. Furthermore, TRMM6 delivers rainfall where no runoff occurs and vice versa. The weak performance of TRMM6-based models illustrates one reason of failings in previous studies, e.g. Khan et al. [2011], Milzow et al. [2011], and Samaniego et al. [2011]. Our results confirm the findings of Bitew and Gebremichael [2011], who figured out that the TRMM near-real-time product performs better than post-processed TRMM6 product in an Ethiopian highland's catchment. Some major post-processing problems of TRMM6 have been solved in TRMM7. Nevertheless, seasonal volume errors resulting from TRMM7 are still present (Figures 6, 7, 8), but the stable temporal trend is well reflected by the hydrographs' trend (Figures 5, 6).

[21]

The mismatch between simulated and observed hydrograph in 2003 in the Kessem basin (Figures 6, 8) causes all models to fail during validation. In the Awash Basin as well, the rainy season of 2003 shows the largest mismatch between observation and simulations (Figures 6, 7). This might be caused by ongoing light rain events which might not have been detected by the TRMM satellite sensors, as documented by Huffman et al. [2007] for the subtropical highs. Habib et al. [2012] also reported "significant missed-rain volumes" for CMORPH.

We conclude that split-sample cross-validations can be a strong and simple tool to identify biased data sets when evaluating them inside hydrological models. A less complex model seems to be better suited for this purpose, especially when runoff time-series are short. A single rainy season of mismatched precipitation lets model validation completely fail. This should be taken into regard when models are tested for their capability of modeling (semi-) arid catchments [e.g. Love et al. 2011; Samaniego et al. 2011], which are commonly data-scarce.

7.4 Water balance simulation

To study the effect of both, model complexity and data sets on the simulation of water balance components, we calibrated the models over the complete available time-frame of runoff data. During the "full-time" calibration (tab. 3), all available data sets are applied in all possible combinations: CMORPH, TRMM6 and TRMM7 for precipitation and GLDAS and MODIS for temperature, resulting in 12 different data-model combinations (tab. 5).

[22]

Table 5: Calibration results of two models in two basins forced by three precipitation and two temperature data sets using the complete runoff time series. Abbreviations: RD – direct runoff; RG – subsurface runoff; SWS – soil water storage; GWS – groundwater storage; % – percent of annual precipitation

Awash Basin

model	precipitation	temperature	time	NSE	VE [%]	KGE'	r	β	γ	AET [%]	RD [%]	RG [%]	GWR [%]	SWS [mm]	GWS [mm]
J2000	CMORPH	GLDAS	t2*	0.75	-4.9	0.83	0.87	0.91	0.95	75	20	3	11	32	131
J2000	CMORPH	MODIS	t2*	0.82	-1.8	0.90	0.91	0.97	1.03	69	23	1	0	76	135
J2000g	CMORPH	GLDAS	t2*	0.79	-0.2	0.87	0.89	1.00	0.94	72	0	25	26	51	12
J2000g	CMORPH	MODIS	t2*	0.79	0.4	0.88	0.89	1.01	0.94	74	0	25	26	20	11
J2000	TRMM V6	GLDAS	t_all	0.57	-5.7	0.77	0.78	0.95	1.05	75	15	3	19	209	156
J2000	TRMM V6	MODIS	t_all	0.65	2.5	0.82	0.82	1.02	0.98	78	8	12	45	18	177
J2000g	TRMM V6	GLDAS	t_all	0.65	-0.2	0.76	0.81	1.00	0.86	81	0	19	20	7	12
J2000g	TRMM V6	MODIS	t_all	0.67	8.1	0.71	0.82	1.07	0.79	79	0	20	22	12	24
J2000	TRMM V7	GLDAS	t_all	0.77	5.2	0.88	0.89	1.05	0.98	75	14	2	10	380	128
J2000	TRMM V7	MODIS	t_all	0.76	4.6	0.87	0.88	1.04	0.98	77	13	2	9	282	122
J2000g	TRMM V7	GLDAS	t_all	0.79	-1.7	0.88	0.89	0.99	0.95	85	0	15	15	9	12
J2000g	TRMM V7	MODIS	t_all	0.81	-3.1	0.89	0.90	0.97	0.97	85	0	15	15	8	11

Kessem Basin

model	precipitation	temperature	time	NSE	VE [%]	KGE'	r	β	γ	AET [%]	RD [%]	RG [%]	GWR [%]	SWS [mm]	GWS [mm]
J2000	CMORPH	GLDAS	t2*	0.68	-3.2	0.70	0.84	0.93	0.76	79	18	12	0	27	118
J2000	CMORPH	MODIS	t2*	0.64	-13.2	0.65	0.83	0.70	0.94	81	18	1	0	4	153
J2000g	CMORPH	GLDAS	t2*	0.71	-8.4	0.74	0.85	0.81	1.08	46	1	26	28	144	11
J2000g	CMORPH	MODIS	t2*	0.71	-8.4	0.74	0.85	0.81	1.08	62	0	26	28	55	11
J2000	TRMM V6	GLDAS	t_all	0.45	-17.7	0.61	0.68	0.84	0.84	77	7	13	10	86	199
J2000	TRMM V6	MODIS	t_all	0.41	-15.7	0.58	0.64	0.86	0.83	75	15	6	8	158	189
J2000g	TRMM V6	GLDAS	t_all	0.45	-11.4	0.59	0.67	0.90	0.78	74	2	19	20	112	19
J2000g	TRMM V6	MODIS	t_all	0.35	-31.4	0.49	0.60	0.72	0.84	83	4	13	14	9	12
J2000	TRMM V7	GLDAS	t_all	0.5	-11.8	0.60	0.70	0.89	0.75	73	1	19	28	227	238
J2000	TRMM V7	MODIS	t_all	0.44	-35.5	0.54	0.68	0.68	0.95	80	7	8	25	191	223
J2000g	TRMM V7	GLDAS	t_all	0.4	9.1	0.46	0.63	1.08	0.61	67	3	22	23	263	52
J2000g	TRMM V7	MODIS	t_all	0.39	-11.9	0.56	0.63	0.89	0.78	79	0	19	20	42	14

[23]

Awash Basin: The application of TRMM7 and CMORPH performs significantly better than the application of TRMM6 (Figure 7, tab. 5). The simple J2000g model has higher NSE, although the differences in NSE are smaller than 15%. Total volume errors are small for all data sets and range from -5.7% to 8.1 %. KGE' is worst for TRMM6, while CMORPH- and TRMM7-based models perform similarly well. In all data-model combinations, bias ratio and variability ratio are close to unity, while the temporal dynamics (measured by r) are the main limiter of KGE'.

It is apparent that the J2000g model does not simulate any surface runoff, accordingly, groundwater recharge makes up between 15 % and 25 % of total precipitation. In opposite, in the J2000 model GWR largely differs between different data-model combinations, but is generally smaller (with one exception applying TRMM6 and MODIS). Applying both TRMM data sets, J2000 simulates subsurface storages (SWS and GWS) one order of magnitude higher than the J2000g model. Using CMORPH, this ratio is true for the groundwater storage (GWS) only (tab. 5).

Visual interpretation of the hydrographs shows partly better matching of smaller runoff events by the J2000 model (Figure 7). The semi-distributed model J2000g noticeably over-estimates falling limbs and also simulates un-observed runoff peaks during relatively small rain events in the early smaller rain seasons, when the soils have their highest infiltration capacity (Figure 7). The fully distributed J2000 model does not model un-observed runoff peaks (Figure 7) and generally reproduces the falling limbs better, particularly for smaller events.

In summary, the semi-distributed model simulates large infiltration rates with short residence times, while the fully distributed model simulates only little infiltration with long residence times. Although invisible in the performance metrics, the fully distributed model, which accounts for lateral re-distribution, captures baseflow recession and small runoff events better than the semi-distributed model. This is in line with Güntner and Bronstert [2004], who

[24]

mentioned the importance of re-infiltration and lateral re-distribution in model approaches for large semi-arid catchments.

[-- Figure 7 --]

Kessem Basin: TRMM6- and TRMM7-based model performances do not differ significantly, but do not exceed NSE of 0.5 (tab. 5, Figure 8). CMORPH-based model performances are much better, having NSE from 0.64 to 0.71, but in a shorter time-frame. Eleven of 12 simulations have negative volume errors, up to -35.5%. The KGE' bias ratio and variability ratio are less good in comparison to the Awash basin, but timing errors are the main limiter of model performance, too. Although the Kessem catchment has a steep gradient, the semi-distributed model simulates direct runoff close to zero. This is only partly true for the TRMM-based J2000 model which simulates subsurface runoff to be larger as surface runoff in seven out of eight cases (tab. 5). Comparing the data-model combinations, subsurface storage volumes are differing in absolute amounts and in the partitioning between SWS and GWS. All large runoff events (> 250 m³/s) are underestimated by all data-model combinations, while the falling limbs of the rainy seasons in 2000 and 2001 are over-estimated. TRMM7 used in the J2000g model leads to an over-estimation of baseflow, while this effect is not visible in the hydrographs of the fully distributed J2000 model.

[-- Figure 8 --]

In summary, simulated hydrographs are similar, while fractions of water balance components differ tremendously. Generally, the models and data applied to the Kessem basin outline only small consistency and indicate larger equifinality problems. A timing error present in only one rainy season (2003) makes further model and data evaluation impossible and might be the reason for unreliable groundwater recharge estimations.

[25]

7.5 Distributed outputs

Hydrological models in semi-arid regions are often used to estimate spatial pattern of water budget components such as groundwater recharge (GWR) or actual evapotranspiration (AET) to better manage available water resources. To fully evaluate the effects of different input data-model combinations, we also compare the spatial patterns of AET and GWR. AET and GWR simulations of all data-model combinations from the "full-time" calibration were applied over the period 2003 – 2010 and accumulated to long-term annual averages. In this period all data sets are available. TRMM6 is excluded from this procedure, because it failed during cross-validation (tab. 4).

[-- Figure 9 --]

The distributed outputs of AET are almost identical when comparing the two different models, J2000g and J2000 (Figure 9). The J2000 model simulated only slightly higher AET in the Kessem Basin than the J2000g model. Major differences occur when using different precipitation data sets. For the Awash Basin TRMM7-based AET is 600 – 700 mm higher than CMORPH-based AET values. For the Kessem Basin the differences still exceed 500 mm. In contrast, there are only small differences between AET simulations based on MODIS LST and those which are based on GLDAS AT (Figure 9). We make the following conclusions: a) Uncertainties in precipitation data sets can be identified as the dominating source of uncertainty in actual ET simulations, while model complexity and temperature data have a minor impact. b) TRMM7 appears to over-estimate precipitation amounts, because actual ET is over-estimated in comparison to satellite-based estimates (Figure 15) and partly baseflow is over-estimated, too (Figure 8). c) CMORPH-based AET simulations show much better agreement with satellite-based AET data (Figure 15).

[-- Figure 10 --]

[26]

The distributed outputs of groundwater recharge explicitly differ depending on the applied model. GWR estimations of the semi-distributed J2000g model show a coherent picture between the different applied data set combinations (Figure 10, left column). Differences are visible but small. Groundwater recharge varies between values exceeding 350 mm in the upper highlands and almost no recharge in the lowest lowlands. These results are in line with values from the literature [Demlie et al. 2007; Ayenew et al. 2008b].

In contrast, the results of the fully distributed J2000 model are much more sensitive to the forcing data sets (Figure 10, right column). Estimations based on TRMM7 have realistic annual recharge rates [Demlie et al. 2007; Ayenew et al. 2008b]. CMORPH-based models significantly under-estimate recharge, especially in the Kessem Basin and along with MODIS LST (Figure 15). Estimations based on MODIS LST show lower annual recharge than those based on GLDAS AT.

Based on this comparison, we can determine main drivers for uncertainty in groundwater recharge simulations: a) Uncertainty towards precipitation and temperature data is highly dependent on the model complexity. b) Different rainfall amounts of differing data sets remain invisible in the simulated hydrographs (Figures 7, 8), but have a large impact on GWR estimations in the fully distributed model. Therefore, GWR estimation is sensitive towards both: Input data and model complexity. c) When applying the simple model, CMORPH appears to be a promising data source for rugged terrain catchments: Simulations of GWR and AET are realistic (Figure 15) and its high spatial resolution outperforms TRMM7 which creates foot-prints in the steep catchment. d) In contrast, GWR simulations by the complex model were unrealistically low when CMORPH was applied (Figure 10): Here, the short calibration time-frame (1.5 years; tab. 3) seems to be insufficient for optimizing 20 parameters.

Concerning temperature data, we conclude: a) Applying MODIS LST might result in under-estimations of groundwater recharge. b) The higher spatial resolution of MODIS LST does

[27]

not improve distributed outputs of GWR. Therefore, the straight-forward application of LST

in hydrological models, as e.g. done by Deus et al. [2013] appears to be limited. We suggest

the application of LST only for very specific needs, or when air temperatures can be modeled

based on land-surface temperatures [Cristóbal et al., 2008; Vancutsem et al. 2010].

7.6 Parameter evolution

The comparison of the different input data sets (Figures 2, 4, 5) and the hydrographs resulting

from the data set-model combinations (Figures 7, 8) show that the models compensate the

data sets' inconsistencies to a large extend. The lively scientific discussion on parameter

equifinality (see e.g. Beven and Binley [1992]) in the last decades demonstrates that the

concept of an optimum parameter set must be superseded by the idea that many possible

parameter sets may provide acceptable simulations. To evaluate the effect of equifinality in

the different data set-model combinations we analyze the evolution of the parameters during

automatic calibration. In Figure 11 the parameter ranges for each data set-model combination

for the best optimization runs is shown. The best optimization runs are selected to have an

efficiency higher than maximum NSE minus 0.05 for the full-time calibration period (tab. 5,

Figure 11).

[-- Figure 11 --]

[-- Figure 12 --]

Additionally, three calibration parameters (ETred, LatVertDist and maxPerc) are selected to

show the parameter evolution during the optimization procedure (Figure 12; see tables A1 and

A2 for parameter descriptions.) Those parameters are exemplary selected because they control

the calculation of AET and GWR, and have similar functions in both models. In principle,

model parameters as representatives of catchment functioning should converge to a clear

[28]

optimum during auto-calibration, independent from input data and models. However, the

evolution of parameters shows that parameter sets are input dependent:

Different climate data sets used in the same model for the same catchment cause different

parameter sets. In fact, the input dependency of model calibration does also affect those

parameters responsible for subsurface processes, like soil water distribution or recession

(Figure 11; Figure 12). Parameters show a wide scattering in the range of the best NSE

(Figure 11). Some data-model combinations show a weak convergence of parameters to a

global maximum (Figure 12). Hence, hundreds of diverging parameter combinations lead to

similar model efficiencies. Choosing "the best" parameter set for modeling e.g. groundwater

recharge contains the risk to choose an outlier. This fact is model-specific as well as data set-

specific. Parameter convergence is more distinct when using the semi-distributed model

J2000g (Figure 12).

7.7 Distributed outputs' variability

As can be seen in the parameter evolution plots (Figure 11; Figure 12) during auto-calibration

many differing parameter combinations are found to perform the highest NSE. While the

resulting simulated hydrographs can be seen as similarly "efficient" or plausible, it is

important to evaluate what is the effect on simulated model outputs of groundwater recharge.

Therefore, we randomly select 100 parameter sets out of the best model evaluations in the

range of maximum NSE minus 0.05 to maximum NSE. The ranges are shown by the red lines

in Figure 12. These 100 randomly selected parameter sets are used to model 100 distributed

outputs of groundwater recharge. We subsequently calculate the mean (Figure 13) and the

standard deviation (Figure 14) of those GWR estimations.

The results show that the simple J2000g model applied in the upper Awash Basin performs

only small differences in simulated GWR resulting from the alternating "best" parameter sets.

This is true for all data set combinations. Using the J2000g model in the steep-gradient

[29]

Kessem Basin, the mean GWR also differs not much from the best parameter set (Figure 10), but the standard deviation of simulated GWR reaches partly more than 40 mm per year. Here smallest standard deviation occurs when applying TRMM7 rainfall data in combination with MODIS LST in the semi-distributed model.

[-- Figure 13 --]

[-- Figure 14 --]

When the fully distributed J2000 model is applied, differences between the simulated GWR of 100 similar efficient "best" parameter sets are larger. The standard deviation of GWR partly exceeds 150 mm per year, in some data-combinations (Figure 14). Especially the CMORPH-based simulations of GWR show major differences between the "best" parameter set and 100 similar ones. Based on this comparison, groundwater recharge simulations are uncertain towards parameter equifinality, in the complex model much more than in the simple model. Therefore, groundwater recharge simulations that are calibrated by runoff data only, have to be treated with care. E.g. hydro-chemical studies can be used to constrain the wide spectrum of recharge estimations [Demlie et al. 2007; Rödiger et al. 2014].

[-- Figure 15 --]

8. Conclusions

The combined analyses of modelling uncertainty due to different rainfall and temperature input data and increasing model complexity shows that the main challenge in modeling the water resources of data-scarce basins is still the reliability of the precipitation data. The analysis in two semi-arid Ethiopian catchments shows that:

A) The TRMM6 data set is identified to be biased in our study area by a negative temporal trend and it displays precipitation events that most probably did not occur. Model failure

[30]

reported by several studies using TRMM6 in data-sparse catchments appears to be attributed to the TRMM6 imperfections [e.g. Khan et al. 2011; Samaniego et al. 2011].

B) The latest TRMM7 satellite product seems to be not ready for use in the study area, to answer the question asked by Xue et al. [2013]. Total precipitation amounts are over-estimated by TRMM7 and it incorporates seasonal volume errors. Nevertheless, TRMM7 shows major improvements towards TRMM6, especially the long-term trend is realistic.

C) The CMORPH high resolution data set (8km) seems to be a promising data set. It captures topographic precipitation pattern better than TRMM does. Simulations of actual evapotranspiration and groundwater recharge showed realistic values, when using the semi-distributed model. Under-estimations of groundwater recharge when using the fully distributed model might be caused by unreliable calibration due to runoff time-series being too short.

D) The use of high resolution LST does not improve model efficiency and the better spatial discretization of LST is contradicted by large over-estimations in comparison to reanalysis air temperatures. Latter are sufficient for hydrological modeling on the catchment scale, despite their coarser spatial resolution.

Beside the high dependency of modeling results on input data, the analysis clearly shows that model complexity increases the sensitivity towards uncertain climate data. Advantages of the fully distributed model become visible in the simulation of falling limbs, baseflow recession and runoff after intermediate rain events, but hardly in efficiency metrics. Estimations of direct runoff are more realistic, too (tab. 5). On the other hand, the fully distributed model calibrated over just a 1.5 years (CMORPH) shows unrealistic groundwater recharge simulations. Further, cross-validation of the complex model using split runoff time-series (of ca. 2 years) is less successful. This shows that results can be unreliable when a complex model is calibrated over just one or two rainy seasons (e.g. Jung et al. [2012]). Therefore, we

[31]

suggest adapting model complexity to the availability of runoff data, which is a common

shortcoming in (semi-) arid catchments.

The analysis shows that studies focusing only on one part of uncertainty, either the input

uncertainty or the uncertainty arising from the choice of model structure are limited in their

explanatory power of the modeling performance, particular in poorly gauged regions.

Contradicting results in the literature which type of model to be used (e.g. Collick et al.

[2009]; Love et al. [2011]) may arise from the fact, that in these studies input data of different

quality is used which highly affects modeling performance. Regarding only one input data set

in our study, one would prefer the more complex model as hydrographs of single events are

better simulated by the fully distributed model (Figure 7). However, regarding the large

differences in the input data, the joint analysis of input and modeling uncertainty clearly

reveals that the performance stems from the fact that more complex models can better

compensate for input data errors. The simulated distributed outputs of GWR and AET

(Figures 9, 10, 13, 14) shows a large variability for the fully distributed model using different

input data sets, while the variation in outputs are much smaller using the simpler model. The

input dependence of the models is certainly only visible by analyzing multiple data sets. By

looking at only one data set, one would see only different outputs, but it would be difficult to

judge, which one is more realistic. The analysis of different data sets shows that output pattern

are to a large extend artefacts of the model ability to compensate input uncertainty. On the

other hand, analyzing the plausibility of different remote sensing data sets by hydrological

modeling highly benefits from using different model types. The level of model compensation

of input data errors is only visible when using more than one model type.

In summary, the joint analysis of model and data uncertainty shows a high interdependence of

input data and model structure on modeling performance. Cross-validation, analysis of

[32]

parameter evolution and investigations of distributed outputs (of GWR, AET) can help to easily uncover concealed data-model uncertainties. We suggest this procedure to apply by default in water resource estimations that are affected by data scarcity, but especially when data sets are evaluated using hydrological models.

A next step towards a more reliable choice of input data and model structure is to use independent data sets on water balance components, to constrain model parameters. E.g. actual evapotranspiration can be estimated by satellite data independent from hydrological models (MOD16 [Mu et al., 2007; 2011], SEBAL [Bastiaansen et al., 1998a,b]). Those estimates of single water budget components can be used for post-calibration evaluation or multi-response calibration [Rientjes et al., 2013]. Other techniques, such as isotope and hydro-chemical surveys, can be used for multi-response calibration of e.g. groundwater recharge or baseflow. Such approaches, e.g. presented by Rödiger et al. [2014], are laborious but highly advisable.

[33]

References

AfDB, OECD, UNDP and the Economic Commission of Africa (2013), African Economy Outlook. [online] Available from: http://www.africaneconomicoutlook.org/en

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, FAO, Rome.

Ayenew, T., S. Kebede, and T. Alemyahu (2008a), Environmental isotopes and hydrochemical study applied to surface water and groundwater interaction in the Awash River basin, Hydrol. Process., 22(10), 1548–1563, doi:10.1002/hyp.6716.

Ayenew, T., M. Demlie, and S. Wohnlich (2008b), Hydrogeological framework and occurrence of groundwater in the Ethiopian aquifers, J. African Earth Sci., 52(3), 97–113, doi:10.1016/j.jafrearsci.2008.06.006.

Bastiaanssen, W. G. M., M. Menenti, R. a. Feddes, and A. a. M. Holtslag (1998a), A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation, *J. Hydrol.*, *212-213*, 198–212, doi:10.1016/S0022-1694(98)00253-4.

Bastiaanssen, W. G. M., H. Pelgrum, J. Wang, Y. Ma, J. F. Moreno, G. J. Roerink, and T. van der Wal (1998b), A remote sensing surface energy balance algorithm for land (SEBAL) 2. Validation, *J. Hydrol.*, *212-213*, 213–229, doi:10.1016/S0022-1694(98)00254-6.

Becker, M., W. LLovel, A. Cazenave, A. Güntner, and J.-F. Crétaux (2010), Recent hydrological behavior of the East African great lakes region inferred from GRACE, satellite altimetry and rainfall observations, Comptes Rendus Geosci., 342(3), 223–233, doi:10.1016/j.crte.2009.12.010.

[34]

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6(3), 279–298, doi:10.1002/hyp.3360060305.

Bissinger, V. and O. Kolditz (2008). Helmholtz Interdisciplinary Graduate School for Environmental Research (HIGRADE). GAIA 1(2008):71-73.

Branger, F., S. Kermadi, C. Jacqueminet, K. Michel, M. Labbas, P. Krause, S. Kralisch, and I. Braud (2013), Assessment of the influence of land use data on the water balance components of a peri-urban catchment using a distributed modelling approach, *J. Hydrol.*, *505*, 312–325, doi:10.1016/j.jhydrol.2013.09.055.

Bretzler, A., K. Osenbrück, R. Gloaguen, J. S. Ruprecht, S. Kebede, and S. Stadler (2011), Groundwater origin and flow dynamics in active rift systems – A multi-isotope approach in the Main Ethiopian Rift, J. Hydrol., 402(3-4), 274–289, doi:10.1016/j.jhydrol.2011.03.022.

Breuer, L., K. Eckhardt, and H.-G. Frede (2003), Plant parameter values for models in temperate climates, Ecol. Modell., 169(2-3), 237–293, doi:10.1016/S0304-3800(03)00274-6.

Bitew, M. M., and M. Gebremichael (2011), Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model, Water Resour. Res., 47(6), W06526, doi:10.1029/2010WR009917.

Collick, A. S., Z. M. Easton, T. Ashagrie, B. Biruk, S. Tilahun, E. Adgo, S. B. Awulachew, G. Zeleke, and T. S. Steenhuis (2009), A simple semi-distributed water balance model for the Ethiopian highlands, Hydrol. Process., 23(7517), n/a–n/a, doi:10.1002/hyp.7517.

Cristóbal, J., M. Ninyerola, and X. Pons (2008), Modeling air temperature through a combination of remote sensing and GIS data, J. Geophys. Res., 113(D13), D13106, doi:10.1029/2007JD009318.

[35]

Dee, D. P. et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828.

Demlie, M., S. Wohnlich, B. Gizaw, and W. Stichler (2007), Groundwater recharge in the Akaki catchment, central Ethiopia: evidence from environmental isotopes (δ18O, δ2H and3H) and chloride mass balance, Hydrol. Process., 21(6), 807–818, doi:10.1002/hyp.6273.

Demlie, M., S. Wohnlich, and T. Ayenew (2008), Major ion hydrochemistry and environmental isotope signatures as a tool in assessing groundwater occurrence and its dynamics in a fractured volcanic aquifer system located within a heavily urbanized catchment, central Ethiopia, J. Hydrol., 353(1-2), 175–188, doi:10.1016/j.jhydrol.2008.02.009.

Deus, D., R. Gloaguen, and P. Krause (2013), Water Balance Modeling in a Semi-Arid Environment with Limited in situ Data Using Remote Sensing in Lake Manyara, East African Rift, Tanzania, Remote Sens., 5(4), 1651–1680, doi:10.3390/rs5041651.

Dinku, T., S. Chidzambwa, P. Ceccato, S. J. Connor, and C. F. Ropelewski (2008), Validation of high-resolution satellite rainfall products over complex terrain, Int. J. Remote Sens., 29(14), 4097–4110, doi:10.1080/01431160701772526.

Dorman, J. L., and P. J. Sellers (1989), A Global Climatology of Albedo, Roughness Length and Stomatal Resistance for Atmospheric General Circulation Models as Represented by the Simple Biosphere Model, J. Appl. Meteorol., 28(9), 833–855.

Duan Q. Y., V.K. Gupta, and S. Sorooshian (1993), Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, J. Optim. Theory Appl., 76(3), 501–521.

Gebregiorgis, A. S., Y. Tian, C. D. Peters-Lidard, and F. Hossain (2012), Tracing hydrologic model simulation error as a function of satellite rainfall estimation bias components and land

[36]

use and land cover conditions, Water Resour. Res., 48(11), n/a–n/a, doi:10.1029/2011WR011643.

Ghavidelfar, S., S. R. Alvankar, and A. Razmkhah (2011), Comparison of the Lumped and Quasi-distributed Clark Runoff Models in Simulating Flood Hydrographs on a Semi-arid Watershed, Water Resour. Manag., 25(6), 1775–1790, doi:10.1007/s11269-011-9774-5.

Güntner, A., and A. Bronstert (2004), Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas, J. Hydrol., 297(1-4), 136–161, doi:10.1016/j.jhydrol.2004.04.008.

Habib, E., A. T. Haile, Y. Tian, and R. J. Joyce (2012), Evaluation of the High-Resolution CMORPH Satellite Rainfall Product Using Dense Rain Gauge Observations and Radar-Based Estimates, J. Hydrometeorol., 13(6), 1784–1798, doi:10.1175/JHM-D-12-017.1.

Heilmeier, H., E.-D. Schulze, J. Fan, and W. Hartung (2007), General relations of stomatal responses to xylem sap abscisic acid under stress in the rooting zone – A global perspective, Flora - Morphol. Distrib. Funct. Ecol. Plants, 202(8), 624–636, doi:10.1016/j.flora.2007.06.002.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker (2007), The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, J. Hydrometeorol., 8(1), 38–55, doi:10.1175/JHM560.1.

Hulley, G. C., and S. J. Hook (2009), Intercomparison of versions 4, 4.1 and 5 of the MODIS Land Surface Temperature and Emissivity products and validation with laboratory measurements of sand samples from the Namib desert, Namibia, Remote Sens. Environ., 113(6), 1313–1318, doi:10.1016/j.rse.2009.02.018.

[37]

Jothityangkoon, C., M. Sivapalan, and D. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, J. Hydrol., 254(1-4), 174–198, doi:10.1016/S0022-1694(01)00496-6.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, P. Xie, and J. E. T. Al (2004), CMORPH : A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, J. Hydrometeorol., 5, 487–503, doi:http://dx.doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Jung, G., S. Wagner, and H. Kunstmann (2012), Joint climate–hydrology modeling: an impact study for the data-sparse environment of the Volta Basin in West Africa, Hydrol. Res., 43(3), 231, doi:10.2166/nh.2012.044.

Kebede, S., Y. Travi, A. Asrat, T. Alemayehu, T. Ayenew, and Z. Tessema (2007), Groundwater origin and flow along selected transects in Ethiopian rift volcanic aquifers, Hydrogeol. J., 16(1), 55–73, doi:10.1007/s10040-007-0210-0.

Khan, S. I., P. Adhikari, Y. Hong, H. Vergara, R. F Adler, F. Policelli, D. Irwin, T. Korme, and L. Okello (2011), Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data, Hydrol. Earth Syst. Sci., 15(1), 107–117, doi:10.5194/hess-15-107-2011.

Klemes, V. (1986), Operational testing of hydrological simulation models, Hydrol. Sci. J., 31(1), 13–24, doi:10.1080/02626668609491024

Kling, H., M. Fuchs, and M. Paulin (2012), Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *J. Hydrol.*, *424-425*, 264–277, doi:10.1016/j.jhydrol.2012.01.011.

[38]

Krause, P. (2002), Quantifying the impact of land use changes on the water balance of large catchments using the J2000 model, Phys. Chem. Earth, Parts A/B/C, 27(9-10), 663–673, doi:10.1016/S1474-7065(02)00051-7.

Krause, P., F. Bäse, U. Bende-Michl, M. Fink, W. Flügel, and B. Pfennig (2006), Multiscale investigations in a mesoscale catchment – hydrological modelling in the Gera catchment, Adv. Geosci., 9(53), 53–61, doi:10.5194/adgeo-9-53-2006.

Krause, P., and S. Hanisch (2009), Simulation and analysis of the impact of projected climate change on the spatially distributed waterbalance in Thuringia, Germany, Adv. Geosci., 21, 33–48, doi:10.5194/adgeo-21-33-2009.

Kurz, T., R. Gloaguen, C. Ebinger, M. Casey, and B. Abebe (2007), Deformation distribution and type in the Main Ethiopian Rift (MER): A remote sensing study, J. African Earth Sci., 48(2-3), 100–114, doi:10.1016/j.jafrearsci.2006.10.008.

Lambin, E. F. (1999), Monitoring forest degradation in tropical regions by remote sensing: some methodological issues, Glob. Ecol. Biogeogr., 8(3-4), 191–198, doi:10.1046/j.1365-2699.1999.00123.x.

Love, D., S. Uhlenbrook, and P. van der Zaag (2011), Regionalising a meso-catchment scale conceptual model for river basin management in the semi-arid environment, Phys. Chem. Earth, Parts A/B/C, 36(14-15), 747–760, doi:10.1016/j.pce.2011.07.005.

LP DAAC (2012), MODIS MCD12Q1 Global Land Cover Type, NASA Land Processes Distributed Active Archive Center, USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. [online] Available from: http://webmap.ornl.gov/wcsdown/dataset.jsp?ds_id=10004

[39]

Meng, J., L. Li, Z. Hao, J. Wang, and Q. Shao (2014), Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River, *J. Hydrol.*, *509*, 320–332, doi:10.1016/j.jhydrol.2013.11.049.

Milzow, C., P. E. Krogh, and P. Bauer-Gottwein (2011), Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, Hydrol. Earth Syst. Sci., 15(6), 1729–1743, doi:10.5194/hess-15-1729-2011.

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 Sens. Environ., 111(4), 519–536, doi:10.1016/j.rse.2007.04.015.

Mu, Q., M. Zhao, and S. W. Running (2011), Improvements to a MODIS global terrestrial evapotranspiration algorithm, *Remote Sens. Environ.*, *115*(8), 1781–1800, doi:10.1016/j.rse.2011.02.019.

FAO, IIASA, ISRIC, ISSCAS and JRC (2009), Harmonized World Soil Database (version 1). FAO, Rome, Italy and IIASA, Laxenburg, Austria. [online] Available from: http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML

Nepal, S., P. Krause, W. Flügel, M. Fink, and C. Fischer (2013), Understanding the hydrological system dynamics of a glaciated alpine catchment in the Himalayan region using the J2000 hydrological model, Hydrol. Process., n/a–n/a, doi:10.1002/hyp.9627.

ORNL DAAC (2012), MODIS subsetted land products, Collection 5, Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), Oak Ridge, Tennessee, U.S.A. [online] Available from: http://daac.ornl.gov/MODIS/modis.html

[40]

Prigent, C. (2010), Precipitation retrieval from space: An overview, Comptes Rendus Geosci., 342(4-5), 380–389, doi:10.1016/j.crte.2010.01.004.

Rientjes, T. H. M., L. P. Muthuwatta, M. G. Bos, M. J. Booij, and H. a. Bhatti (2013), Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration, *J. Hydrol.*, *505*, 276–290, doi:10.1016/j.jhydrol.2013.10.006.

Rödiger, T., S. Geyer, U. Mallast, R. Merz, P. Krause, C. Fischer, and C. Siebert (2014), Multi-response calibration of a conceptual hydrological model in the semiarid catchment of Wadi al Arab, Jordan, *J. Hydrol.*, *509*, 193–206, doi:10.1016/j.jhydrol.2013.11.026.

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, Hydrol. Res., 42(5), 338, doi:10.2166/nh.2011.156.

Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and DMIP Participants (2004), Overall distributed model intercomparison project results, J. Hydrol., 298(1-4), 27–60, doi:10.1016/j.jhydrol.2004.03.031.

Rodell, M. et al. (2004), The Global Land Data Assimilation System, Bull. Am. Meteorol. Soc., 85(3), 381–394, doi:10.1175/BAMS-85-3-381.

Romilly, T. G., and M. Gebremichael (2011), Evaluation of satellite rainfall estimates over Ethiopian river basins, *Hydrol. Earth Syst. Sci.*, *15*(5), 1505–1514, doi:10.5194/hess-15-1505-2011.

Schultz, J. (2005), The Ecozones of the World: The Ecological Divisions of the Geosphere, 2nd edition, Springer, Berlin, Heidelberg, Germany

[41]

Schwanghart, W., and N. J. Kuhn (2010), TopoToolbox: A set of Matlab functions for topographic analysis, Environ. Model. Softw., 25(6), 770–781, doi:10.1016/j.envsoft.2009.12.002.

Seleshi, Y., and U. Zanke (2004), Recent changes in rainfall and rainy days in Ethiopia, Int. J. Climatol., 24(8), 973–983, doi:10.1002/joc.1052.

Shao, Q., J. Lerat, G. Podger, and D. Dutta (2014), Uncertainty estimation with bias-correction for flow series based on rating curve, *J. Hydrol.*, *510*, 137–152, doi:10.1016/j.jhydrol.2013.12.025.

Shiklomanov, I. a. (2000), Appraisal and Assessment of World Water Resources, Water Int., 25(1), 11–32, doi:10.1080/02508060008686794.

Stisen, S., K. H. Jensen, I. Sandholt, and D. I. F. Grimes (2008), A remote sensing driven distributed hydrological model of the Senegal River basin, J. Hydrol., 354(1-4), 131–148, doi:10.1016/j.jhydrol.2008.03.006.

Stisen, S., and I. Sandholt (2010), Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling, Hydrol. Process., 24(7), 879–891, doi:10.1002/hyp.7529.

Tesfaye, C. (1988): Hydrogeological Map of Ethiopia, Ethiopian Institute of Geological Surveys, Addis Ababa, Ethiopia

Vancutsem, C., P. Ceccato, T. Dinku, and S. J. Connor (2010), Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa, Remote Sens. Environ., 114(2), 449–465, doi:10.1016/j.rse.2009.10.002.

[42]

Van Griensven, A., P. Ndomba, S. Yalew, and F. Kilonzo (2012), Critical review of SWAT applications in the upper Nile basin countries, Hydrol. Earth Syst. Sci., 16(9), 3371–3381, doi:10.5194/hess-16-3371-2012.

Velpuri, N. M., G. B. Senay, and K. O. Asante (2012), A multi-source satellite data approach for modelling Lake Turkana water level: calibration and validation using satellite altimetry data, Hydrol. Earth Syst. Sci., 16(1), 1–18, doi:10.5194/hess-16-1-2012.

Wan, Z. (1999). MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD), Version 3.3, Institute for Computational Earth System Science, University of California, Santa Barbara, California, USA

Wan, Z. (2008), New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products, Remote Sens. Environ., 112(1), 59–74, doi:10.1016/j.rse.2006.06.026.

Wang, L., T. Koike, K. Yang, and P. J.-F. Yeh (2009), Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin, J. Hydrol., 377(1-2), 21–34, doi:10.1016/j.jhydrol.2009.08.005.

Wang, F., L. Wang, T. Koike, H. Zhou, K. Yang, A. Wang, and W. Li (2011), Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model, J. Geophys. Res., 116(D21), D21108

Yong, B., B. Chen, J. J. Gourley, L. Ren, Y. Hong, X. Chen, W. Wang, S. Chen, and L. Gong (2014), Intercomparison of the Version-6 and Version-7 TMPA precipitation products over high and low latitudes basins with independent gauge networks: Is the newer version better in both real-time and post-real-time analysis for water resources and hydrologic extremes?, *J. Hydrol.*, *508*, 77–87, doi:10.1016/j.jhydrol.2013.10.050.

[43]

Yucel, I. (2006), Effects of Implementing MODIS Land Cover and Albedo in MM5 at Two Contrasting U.S. Regions, J. Hydrometeorol., 7(5), 1043–1060, doi:10.1175/JHM536.1.

Xue, X., Y. Hong, A. S. Limaye, J. J. Gourley, G. J. Huffman, S. I. Khan, C. Dorji, and S. Chen (2013), Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?, *J. Hydrol.*, *499*, 91–99, doi:10.1016/j.jhydrol.2013.06.042

[44]

Appendix

Table A1: Parameters of the J2000 model and their mathematical implementation

parameter	function source: Krause [2002]; Krause et al. [2006]; Nepal et al. [2012]	equation
ACA	initial adaptation factor for large pore storage capacity (maxLPS)	$maxLPS = ACA * AC$ [L] $AC = air\ capacity$
FCA	adaptation factor for mid-size pore storage capacity (maxMPS)	$maxMPS = FCA * FC$ [L] $FC = effective\ field\ capacity$
flowRouteTA	routing coefficient for MANNING-STRICKLER equation, representing the runtime of a hydrograph after a rain event.	$Rk = \dfrac{v}{fl} * flowRouteTA * 3600$ [–] v = velocity [m/s] fl = flow path length [m] $Q = q * e^{-\frac{1}{Rk}}\ [\dfrac{m^3}{s}]$ q = Q$_{reach}$
ETred	polynomial coefficient for PET reduction to achieve AET in relation to actual mid-size pore storage (actMPS)	$AET = \left(10^{-10*(1-\Theta MPS)^{ETRed}}\right) * PET$ ΘMPS = rel. saturation of MPS
soilMaxDPS	depression storage capacity	threshold [mm]
soilMaxInfWinter	maximal infiltration rate per day in winter	threshold [mm/d]
soilMaxInfSummer	maximal infiltration rate per day in summer	threshold [mm/d]
soilImpGT80	relative infiltration capability of impervious areas of >80 % sealing (urban)	[-]
soilImpLT80	relative infiltration capability of impervious areas of <80 % sealing	[-]
LatVertDist	coefficient for the distribution of large pore storage outflow to interflow and percolation	$perc = LPSout * (1 - \tan(slope) * LatVertDist)$

[45]

		$inter = LPSout * (\tan(slope) * LatVertDist)$ $perc = percolation$ [L] $inter = interflow$ [L]
soilOutLPS	outflow coefficient for the large pore storage (LPS)	$LPSout = soilsat^{soilOutLPS} * LPSact$ $LPSout = outflow\ from\ LPS$ [L] $soilsat = \dfrac{MPSact + LPSact}{MPSmax + LPSmax}[-]$ $MPSact, LPSact = actual\ volume\ of\ MPS, LPS$ [L] $MPSmax, LPSmax = maximal\ volume\ of\ MPS, LPS$ [L]
maxPerc	maximal percolation rate per day to groundwater (mm/d)	Threshold
soilConcRD1	retention coefficient for direct runoff (overland flow)	$RO_{t+1} = \dfrac{1}{soilConcRD1} * RO_t$ $RO = surface\ runoff$ [L] $t = time\ step$
soilConcRD2	retention coefficient for interflow	$RI_{t+1} = \dfrac{1}{soilConcRD2} * RI_t$ $RI = Interflow\ 1$ [L] $t = time\ step$
soilDistMPSLPS	coefficient for the distribution of infiltration water to large pore storage (LPS) and mid-size pore storage (MPS)	$MPSin = Inf * \left(1 - e^{\frac{-soilDistMPSLPS}{\Theta MPS}}\right)$ $MPSin = inflow\ into\ MPS$ [L] $\Theta MPS - rel.saturation\ of\ MPS$ $[-]$ $Inf = infiltration\ water$ [L]

[46]

soilDiffMPSLPS	coefficient for the amount of diffusion from large pore storage (LPS) and mid-size pore storage (MPS)	$LPS2MPS = LPSact * \left(1 - e^{\frac{-soilDiffMPSLPS}{\Theta MPS}}\right)$ $LPS2MPS = LPS\ to\ MPS\ diffusion$ [L] $LPSact = actual\ LP\ storage$ [L] $\Theta MPS - rel.saturation\ of\ MPS$ [-]
gwRG1RG2dist	coefficient for the distribution of the percolation water to groundwater storage (GWS) 1 and 2	$inGWS1 = perc * \tan(slope) * gwRG1RG2dist$ $inGWS2 = perc * (1 - \tan(slope)) * gwRG1RG2dist$ perc = percolation [L] $inGWS1 = inflow\ into\ upper\ groundwater\ storage$ $inGWS2 = inflow\ into\ lower\ groundwater\ storage$
gwRG1Fact	recession coefficient for groundwater cascade no.1	$outGWS1 = \dfrac{1}{gwRG1Fact * recGWS1} * actGWS1$ $outGWS1 = outflow\ from\ GWS1$ [L] $recGWS1 = recession\ coefficient\ of\ geological\ units$ $actGWS1 = actual\ volume\ of\ GWS1$
gwRG2Fact	recession coefficient for groundwater cascade no.2	$outGWS2 = \dfrac{1}{gwRG2Fact * recGWS2} * actGWS2$
gwCapRise	factor for the capillary rise of groundwater to mid-size pore storage (MPS)	$Rise = \Delta MPS * \left(1 - e^{-\frac{gwCapRise}{\Theta MPS}}\right)$ [mm] $Rise = capillary\ rise\ from\ groundwater\ to\ MPS$ $\Delta MPS = MPSmax - MPSact$ $MPSact = actual\ MPS$ $MPSmax = maximal\ MPS$ $\Theta MPS = relative\ saturation\ of\ MPS$

[47]

Table A2: Parameters of the J2000g model and their mathematical implementation

parameter	function (source: Krause and Hanisch [2009])	equation
FCA	adaptation factor for maximum soil water storage (maxSWS)	$maxSWS = FCA * FC$ [mm] $FC = effective\ field\ capacity$
ETred	coefficient for PET reduction to achieve AET in relation to actual soil water storage (actSWS)	$AET = \left(\dfrac{actMPS}{ETRed * maxSWS}\right) * PET$ [mm]
PETmult	PET adaption factor	$PET = PETmult * PET_0$ $PET_0 = PET\ after\ Allen\ et\ al.\,(1998)$
LatVertDist	coefficient for the lateral/vertical distribution of excess water from soil water storage (SWS) to surface runoff and percolation	$RO = \tan(slope) * LatVertDist * EW$ $perc = (1 - \tan(slope)) * LatVertDist * EW$ $perc \leq maxPerc * maxPercRate$ maxPercRate = maximal percolation rate from geological units $RO = surface\ runoff$ [mm] $EW = excess\ water\ from\ SWS$ [mm] $perc = percolation$ [mm]
maxPerc	adaptation of maximal percolation rate	see LatVertDist
gwAlpha	distribution coefficient for the two groundwater components	$GWS1inflow = gwAlpha * perc$[mm] $GWS2inflow = (1 - gwAlpha) * perc$[mm]
gwK1	recession coefficient for groundwater cascade no.1	$BQ1 = \dfrac{1}{gwK1} * actGWS1$ $BQ1 = baseflow1$ $actGWS1 =$ $actual\ storage\ of\ groundwater\ tank\ 1$
gwK2	recession coefficient for groundwater cascade no.2	analog to gwK2

[48]

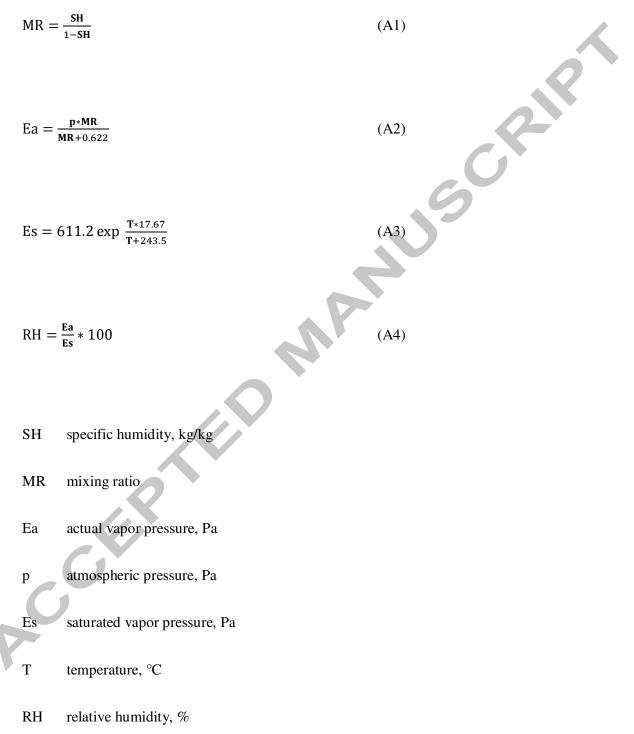Eq. A1 – A4: Calculation of relative humidity from specific humidity (source: LP DAAC, on request)

$$MR = \frac{SH}{1-SH} \qquad\qquad (A1)$$

$$Ea = \frac{p*MR}{MR+0.622} \qquad\qquad (A2)$$

$$Es = 611.2 \exp \frac{T*17.67}{T+243.5} \qquad\qquad (A3)$$

$$RH = \frac{Ea}{Es} * 100 \qquad\qquad (A4)$$

SH   specific humidity, kg/kg

MR   mixing ratio

Ea   actual vapor pressure, Pa

p   atmospheric pressure, Pa

Es   saturated vapor pressure, Pa

T   temperature, ℃

RH   relative humidity, %

[49]

Figure 1. Study area in the central Main Ethiopian Rift. Red: The upper Awash Catchment; Black: The Kessem Catchment

Figure 2. Forcing data: A) Annual average precipitation 2003 – 2010; B) Daily average temperature 2001 – 2010

Figure 3. Graphical representation of the model structures. Left: J2000 model, adapted after Nepal et al. [2012]: Right: J2000g model. Abbreviations: DPS – depression storage; MPS – mid-size pore storage; LPS – large pore storage; GWS – groundwater storage; SWS – soil water storage.

Figure 4. Spatial averages of daily temperature of two different data sets

Figure 5. Annual precipitation rates of three different data sets: Boxplots showing spatial variance

Figure 6. Uncertainty decomposition by applying a split-sample cross-validation to the J2000g model, forced by TRMM6 and TRMM7. (Discharge representation has been truncated at 250 m³/s.) Vertical arrows: Major difference between calibration and validation, caused by trend bias in the TRMM6 rainfall data. Horizontal arrows: Major timing error between observed and simulated hydrograph. Diagonal arrows: Major rain events in TRMM6 data set, where no runoff was observed.

Figure 7. Awash Basin: Calibration of the J2000g and the J2000 model being forced by three different precipitation data sets, using the complete runoff time series. Due to considerations of simplicity, only simulated hydrographs driven by GLDAS temperature data are shown.

Figure 8. Kessem Basin: Calibration of the J2000g and the J2000 model being forced by three different precipitation data sets, using the complete runoff time series. Due to considerations of simplicity, only simulated hydrographs driven by GLDAS temperature data are shown.

[50]

Figure 9. Annual actual evapotranspiration, modeled with two different models, forced by four different data set combinations. Left catchment: Awash; Right catchment: Kessem

Figure 10. Annual groundwater recharge, modeled with two different models, forced by four different data set combinations. Left catchment: Awash; Right catchment: Kessem

Figure 11. Parameter range of all optimization runs in the range of the highest 0.05 Nash-Sutcliffe efficiencies. Note, that highest NSE is different for each data-model-combination. Parameters have been normalized to the range of 0 – 1. TRMM stands for TRMM version 7 (TRMM version 6 failed during cross-validation).

Figure 12. Evolution of selected parameters during optimization, sorted by ascending Nash-Sutcliffe efficiencies (red line: Best 0.05 NSE; oblique number shows the total number of model evaluations, until the abort criterion was reached; horizontal number shows number of model evaluations where best NSE start). Upper left: Fully distributed J2000 model applied to the Awash basin. Lower left: J2000 model applied to the Kessem basin. Upper right: Semi-distributed J2000g model applied to the Awash Basin. Lower right: J2000g model applied to the Kessem basin.

Figure 13. Modeled groundwater recharge: Mean of 100 randomly selected best optimization runs in the range of the highest 0.05 NSE
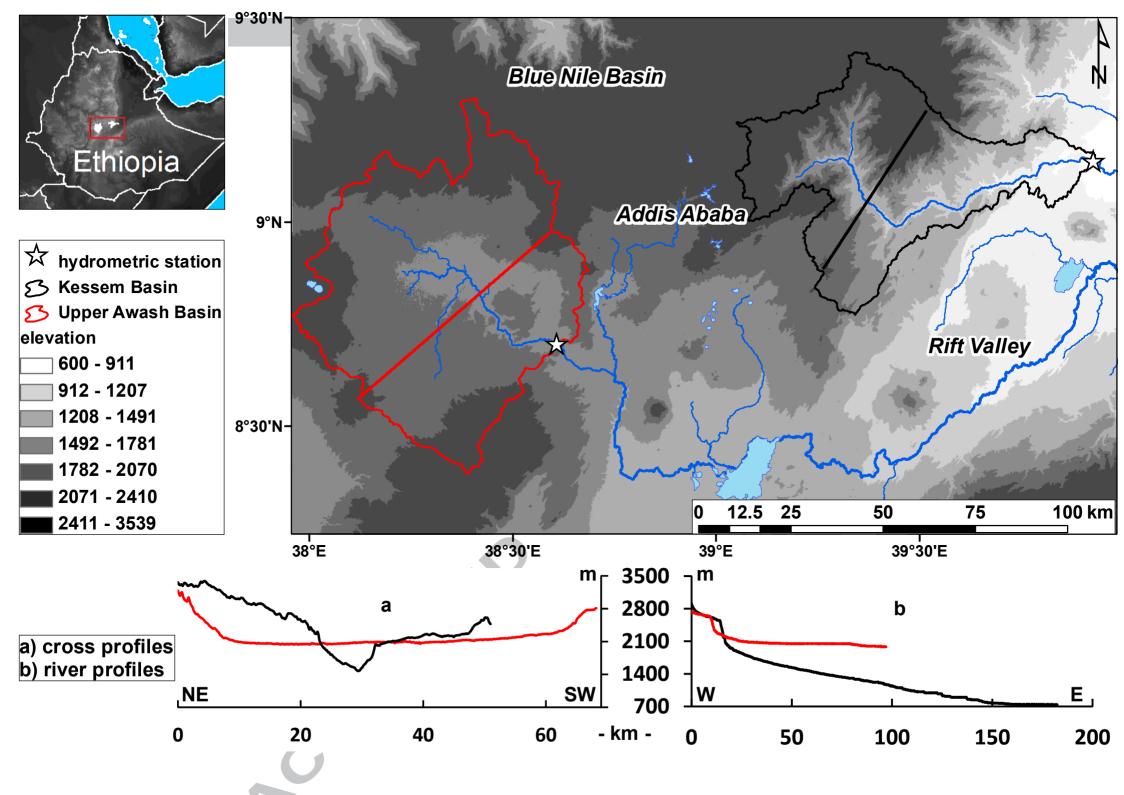
Figure 14. Modeled groundwater recharge: Standard deviation of 100 randomly selected best optimization runs in the range of the highest 0.05 NSE
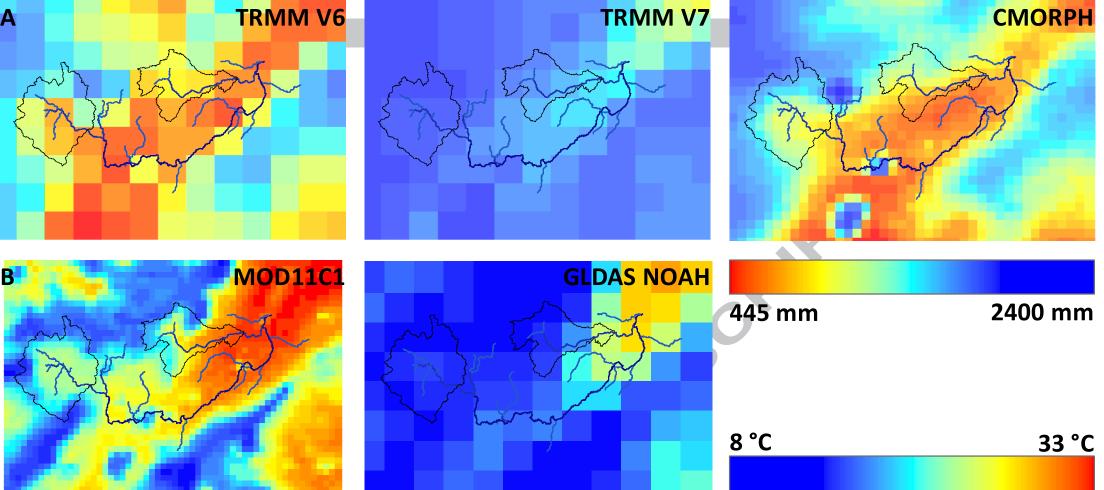
Figure 15. Comparison of modeled GWR and AET with alternative data sources. Boxplots are showing spatial variance. Demlie et al. [2008] applied the Chloride Mass Balance Method (CMBM) at
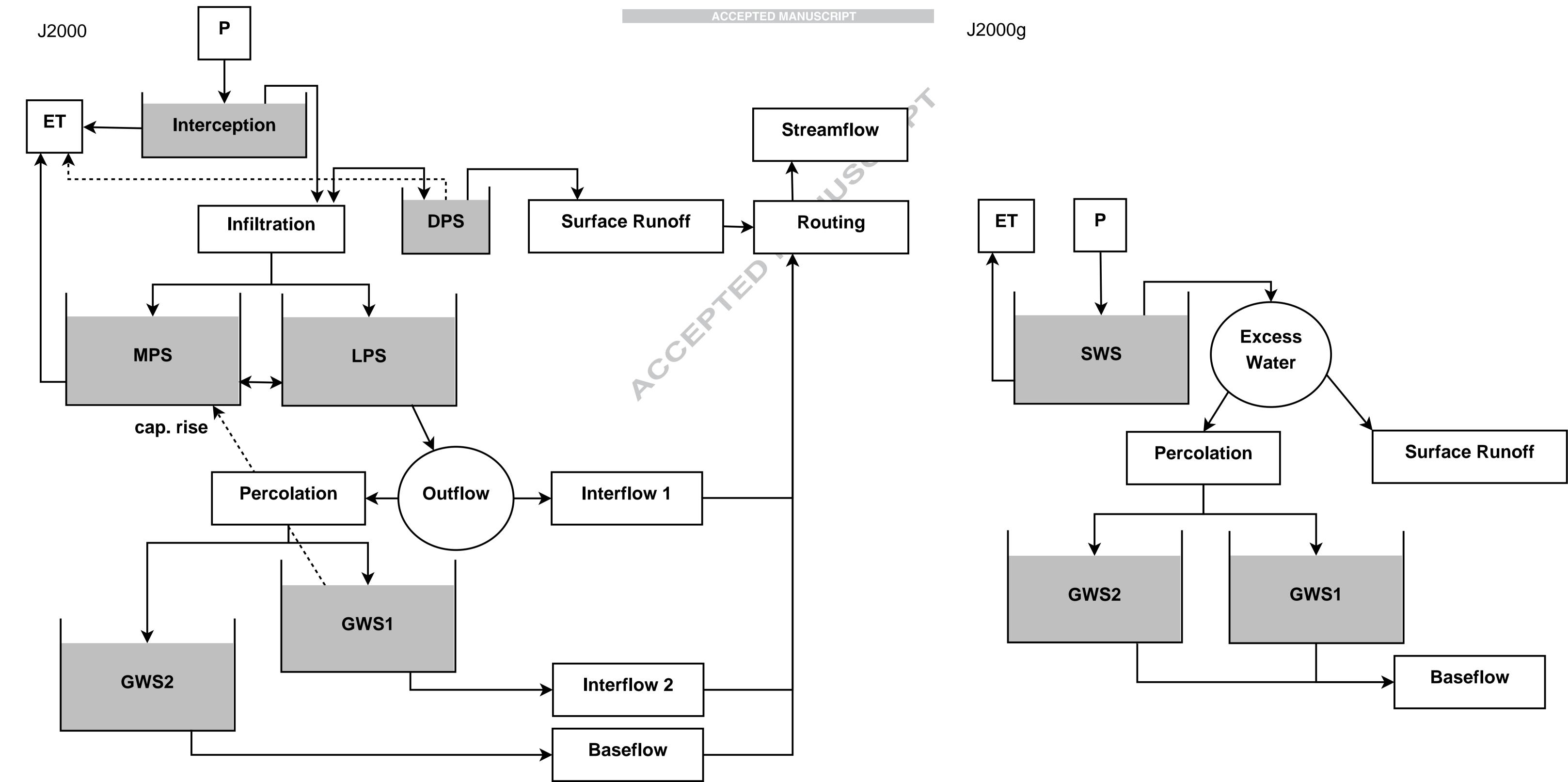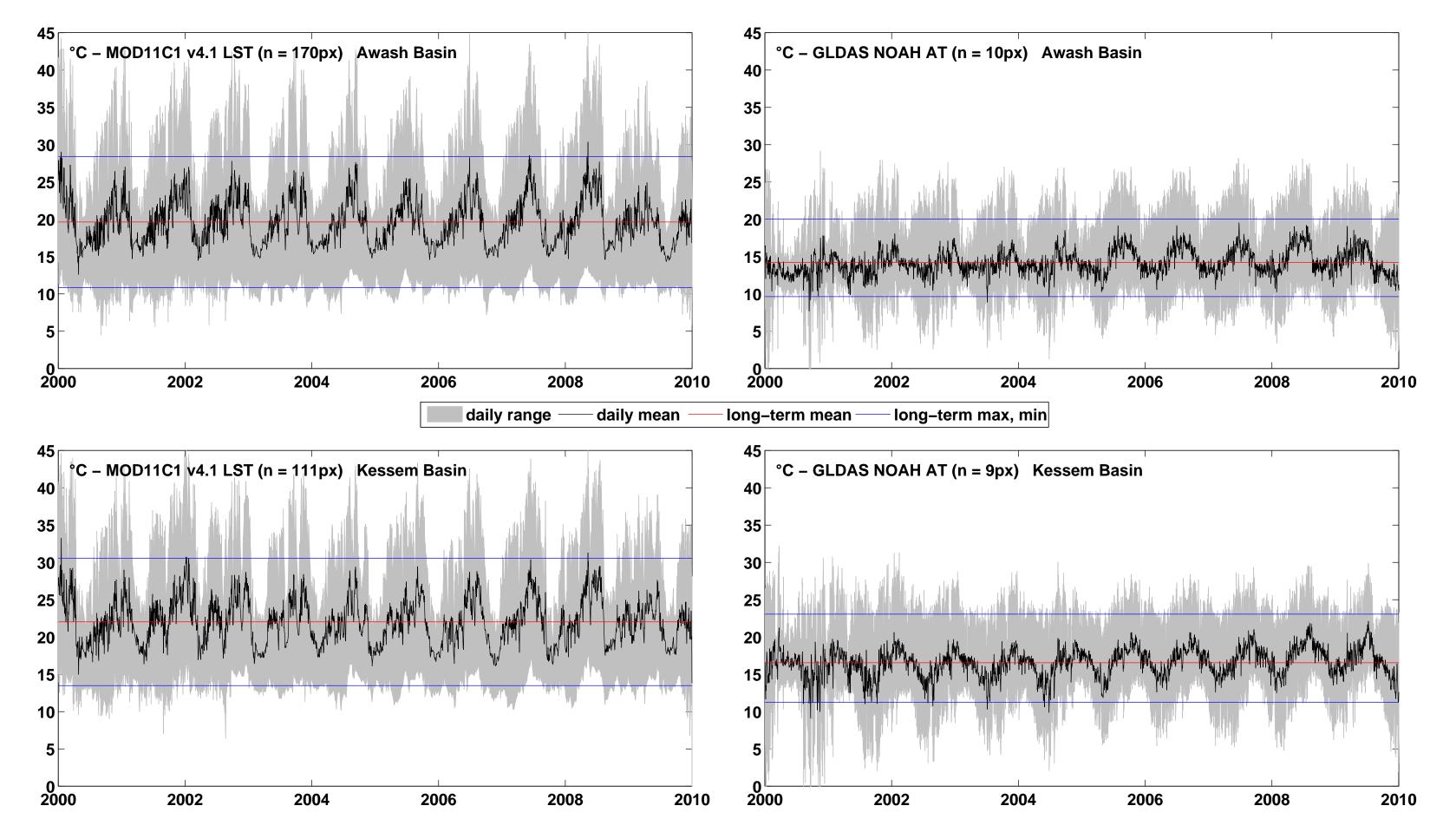
[51]

7 locations in the adjacent Akaki catchment between 2300 and 2605 m a.s.l. MOD16A2 is a satellite-

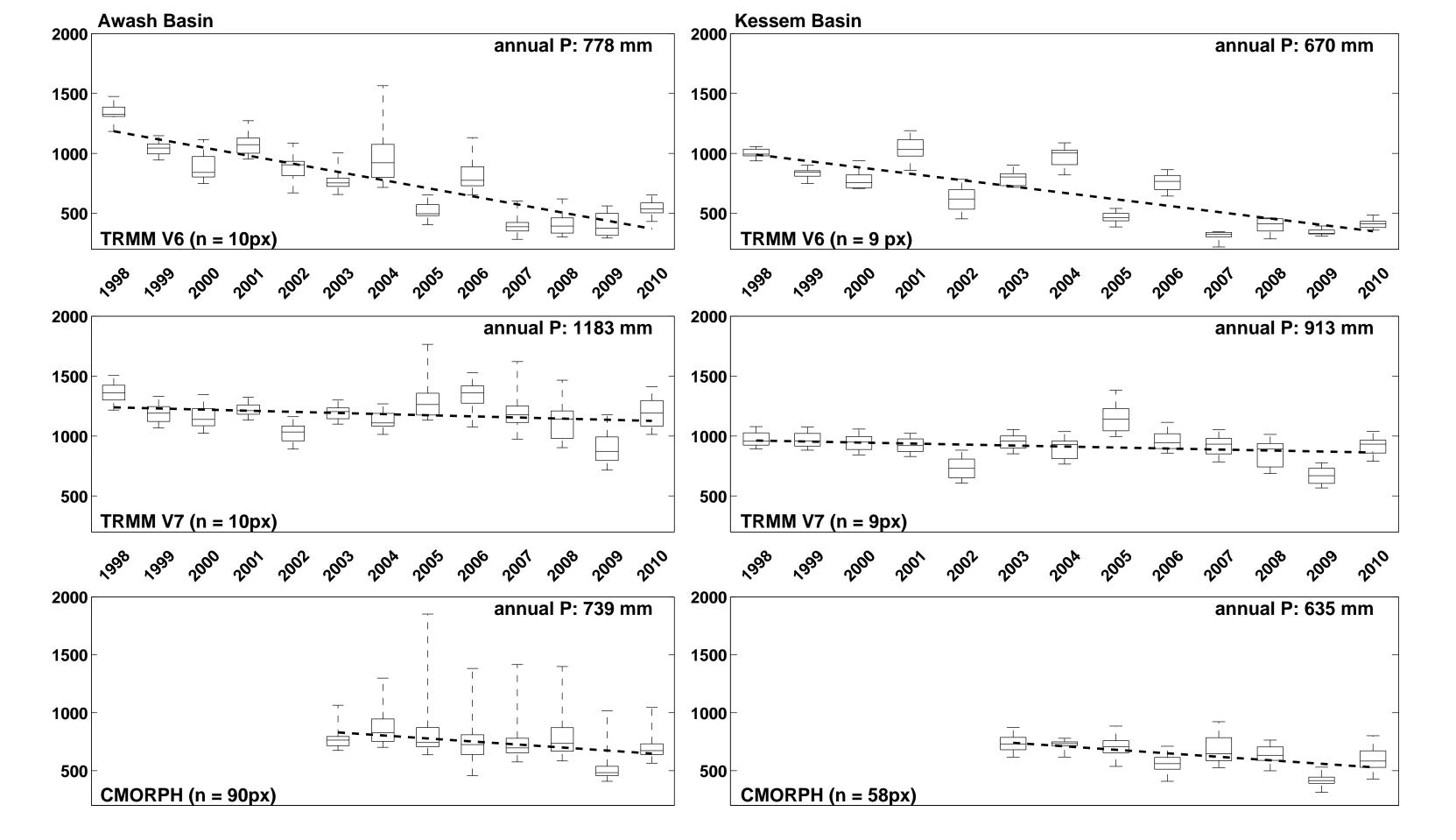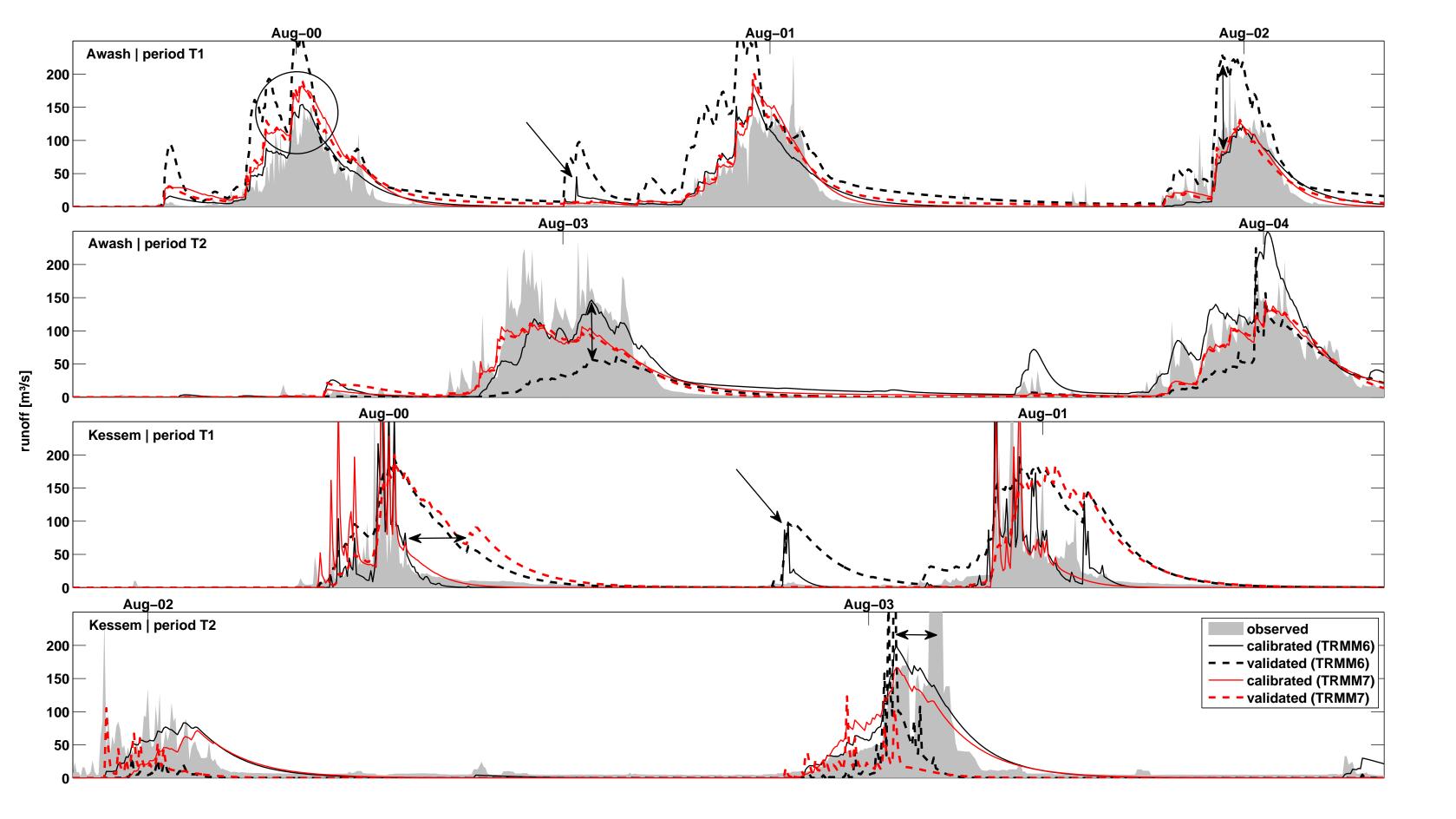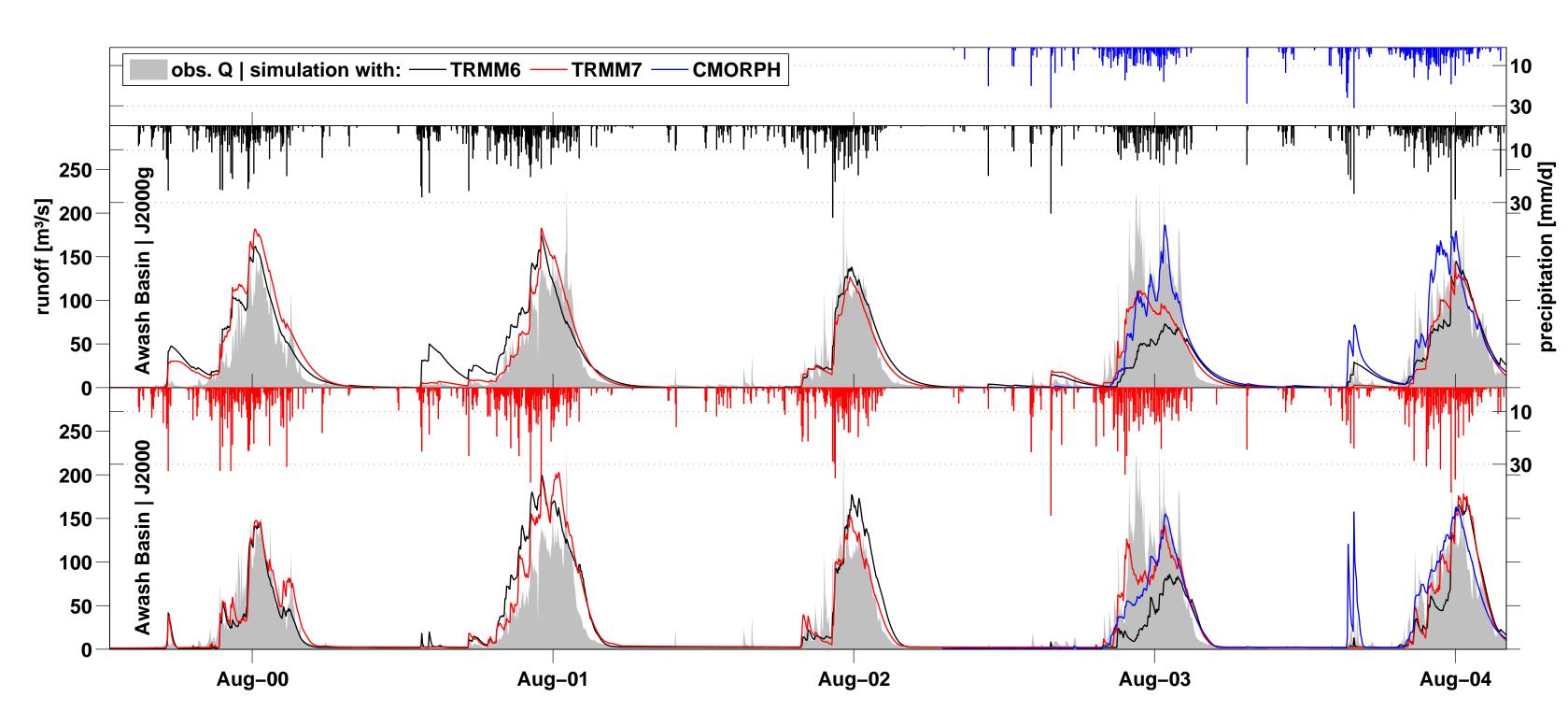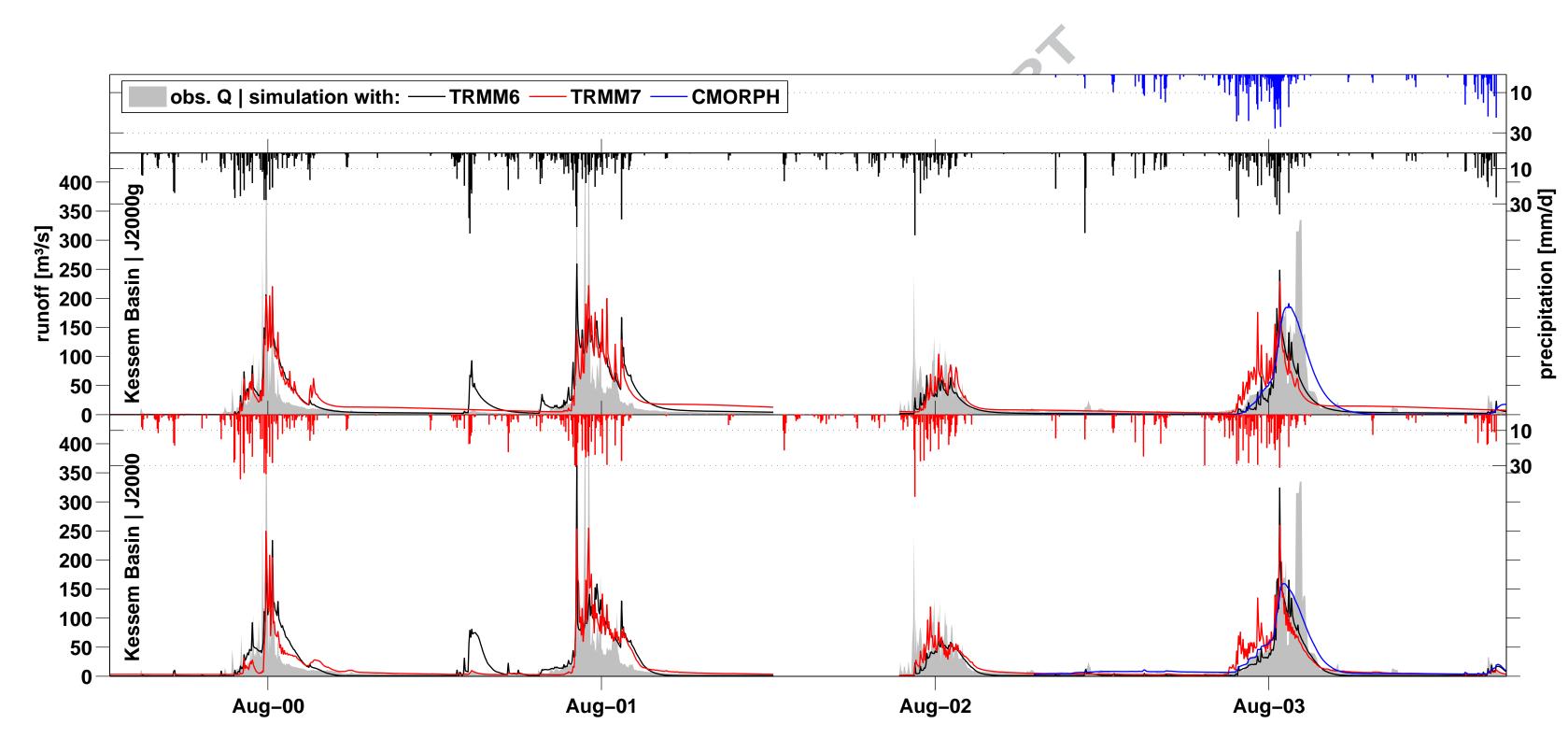based global ET product (1km², 8-day space-time resolution).

9°30'N

9°N

8°30'N

Blue Nile Basin

Addis Ababa

Rift Valley

Ethiopia

☆ hydrometric station

⬡ Kessem Basin

⬡ Upper Awash Basin

elevation

	600 - 911
	912 - 1207
	1208 - 1491
	1492 - 1781
	1782 - 2070
	2071 - 2410
	2411 - 3539

N

0 12.5 25 50 75 100 km

38°E   38°30'E   39°E   39°30'E

a

a) cross profiles
b) river profiles

NE   SW

0   20   40   60   - km -

m
3500
2800
2100
1400
700

b

W   E

0   50   100   150   200

Figure: runoff [m³/s] time series for Awash (period T1, period T2) and Kessem (period T1, period T2) catchments, comparing observed, calibrated (TRMM6), validated (TRMM6), calibrated (TRMM7), and validated (TRMM7) discharge.

Legend:
- observed
- calibrated (TRMM6)
- validated (TRMM6)
- calibrated (TRMM7)
- validated (TRMM7)

J2000g

J2000

1384

1000

700

500

211 mm actual ET/year

TRMM-V7
MOD11C1 LST

CMORPH
MOD11C1 LST

TRMM-V7
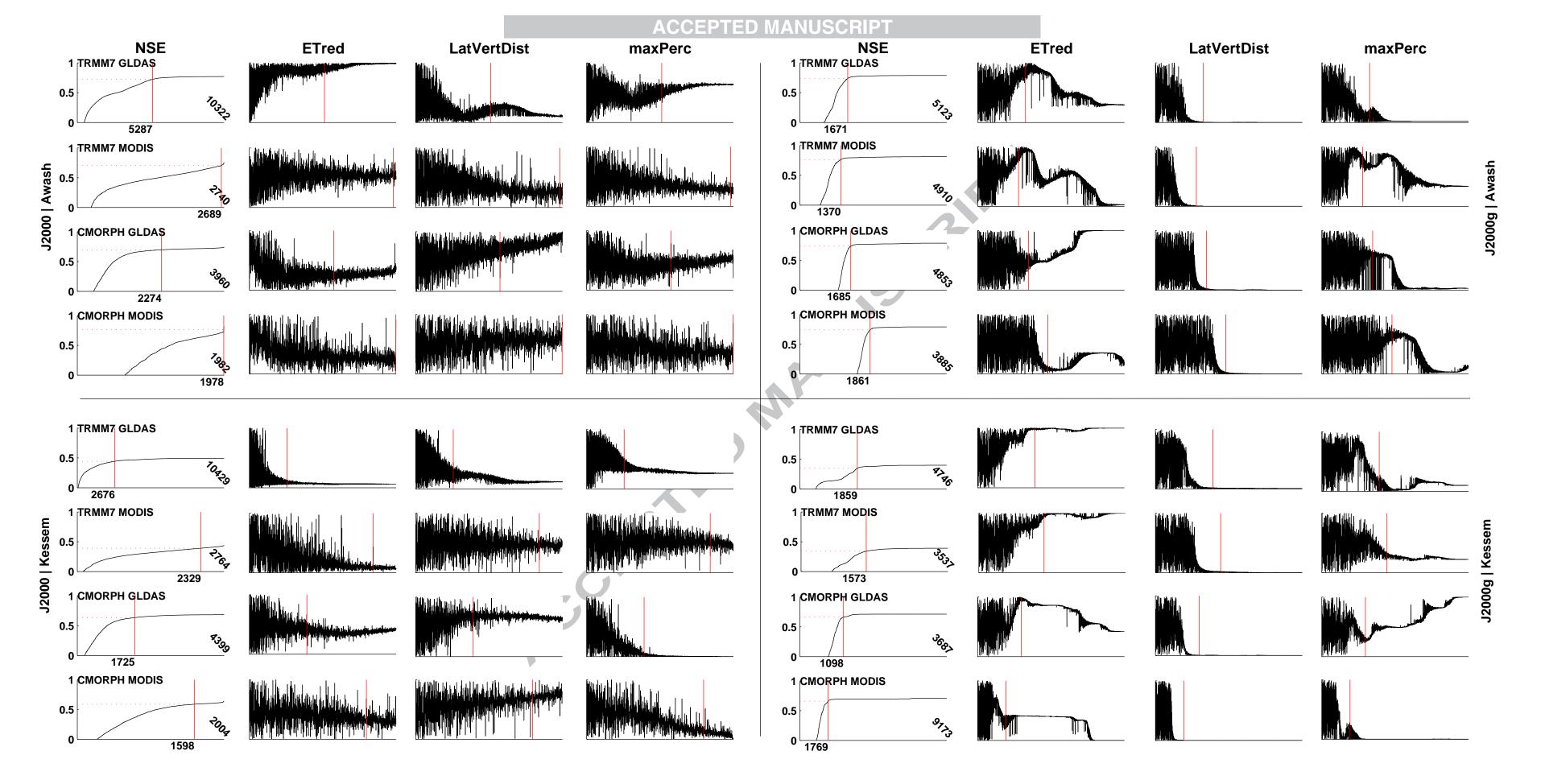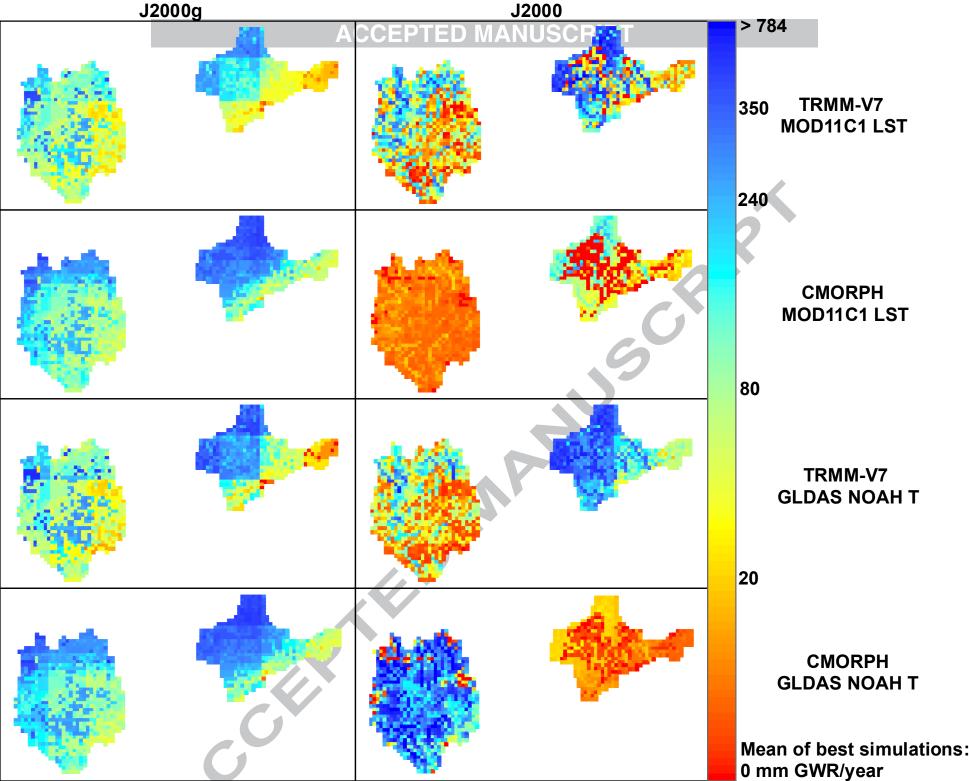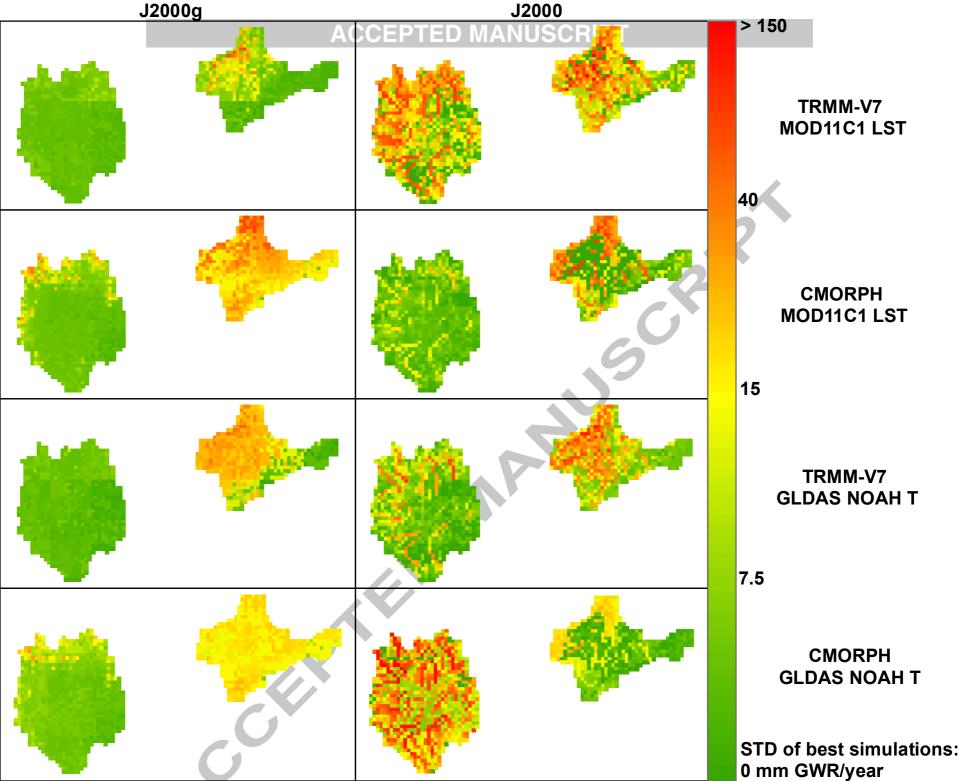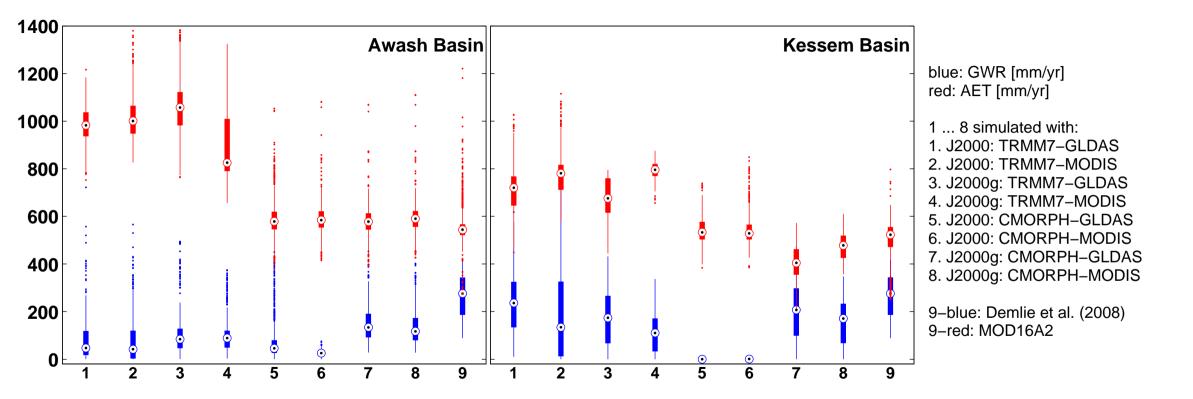GLDAS NOAH T

CMORPH
GLDAS NOAH T

J2000g: 1 NSE | 2 ETred | 3 gwAlpha | 4 PETmult | 5 FCA | 6 maxPerc | 7 LatVertDist | 8 gwK1 | 9 gwK2

J2000: 1 NSE | 2 ACA | 3 FCA | 4 soilDiffMPSLPS | 5 soilDistMPSLPS | 6 soilOutLPS | 7 LatVertDist | 8 soilMaxDPS | 9 soilConcRD1 | 10 soilConcRD2 | 11 gwRG1Fact | 12 gwRG2Fact | 13 gwRG1RG2Dist | 14 gwCapRise | 15 maxPerc | 16 ETred
17 flowRouteTA | 18 soilMaxInfWinter | 19 soilMaxInfSummer | 20 soilImpGT80 | 21 soilImpLT80

J2000g	J2000	

TRMM-V7
MOD11C1 LST

CMORPH
MOD11C1 LST

TRMM-V7
GLDAS NOAH T

CMORPH
GLDAS NOAH T

> 784

350

240

80

20

Mean of best simulations:
0 mm GWR/year

blue: GWR [mm/yr]
red: AET [mm/yr]

1 ... 8 simulated with:
1. J2000: TRMM7−GLDAS
2. J2000: TRMM7−MODIS
3. J2000g: TRMM7−GLDAS
4. J2000g: TRMM7−MODIS
5. J2000: CMORPH−GLDAS
6. J2000: CMORPH−MODIS
7. J2000g: CMORPH−GLDAS
8. J2000g: CMORPH−MODIS

9−blue: Demlie et al. (2008)
9−red: MOD16A2

[52]

Key Points:

data set uncertainties and model uncertainties are mutually concealed

data set evaluation and hydrological model evaluation should be applied together

hydrological models buffer data set differences to a large extend

simulated hydrographs are meaningless for data set evaluations

parameter evolution and distributed outputs uncover data and model uncertainties