Spatial validation of a semi-distributed hydrological nutrient transport model

Salman Ghaffar[1,*], Seifeddine Jomaa[1], Günter Meon[2] and Michael Rode[1,3]

[1]Department of Aquatic Ecosystem Analysis and Management, Helmholtz Centre for Environmental Research – UFZ, Magdeburg, Germany

[2]Department of Hydrology, Water Management and Water Protection, Leichtweiss Institute for Hydraulics and Water Resources, University of Braunschweig, Braunschweig, Germany

[3]Institute of Environmental Science and Geography, University of Potsdam, Potsdam-Golm, Germany

[*]Corresponding author: Salman Ghaffar (salman.ghaffar@ufz.de)

Abstract

Semi-distributed hydrological and water quality models are increasingly used as innovative and scientific-based management tools. However, their application is usually restricted to the gauging stations where they are originally calibrated, limiting their spatial capability. In this study, the semi-distributed hydrological water quality model HYPE (HYdrological Predictions for the Environment) was tested spatially to represent nitrate-N (NO_3^--N) and total phosphorus (TP) concentrations and loads of the nested and heterogeneous Selke catchment (463 km^2) in central Germany. First, an automatic calibration procedure and uncertainty analysis were conducted using the DiffeRential Evolution Adaptive Metropolis (DREAM) tool to simulate discharge, NO_3^--N and TP concentrations. A multi-site and multi-objective calibration approach was applied using three main gauging stations, covering the most important hydro-meteorological and physiographical characteristics of the whole catchment. Second, the model's capability was tested to represent further internal stations, which were not initially considered for calibration. Results showed that discharge was well represented by the model at all three main stations during both calibration (1994-1998) and validation (1999-2014) periods with lowest Nash-Sutcliffe Efficiency (NSE) of 0.71 and maximum Percentage BIAS (PBIAS) of 18.0%. The model was able to reproduce the seasonal dynamics of NO_3^--N and TP concentrations with low predictive uncertainty at the three main stations, reflected by PBIAS values in the ranges from -16.1% to 6.4% and from -20.0% to 11.5% for NO_3^--N and TP load simulations, respectively. At internal stations, the model could represent reasonably well the seasonal variation of nutrient concentrations with PBIAS values in the ranges from -9.0% to 14.2% for NO_3^--N and from -25.3% to 34.3% for TP concentration simulations. Overall, results suggested that the spatial validation of a nutrient transport model can be better ensured when a multi-site and multi-objective calibration approach using archetypical gauging stations is implemented. Further, results revealed that the delineation of sub-catchments should put more focus on hydro-meteorological conditions than on land-use features.

Keywords: HYPE model, Nitrate-N, Phosphorus, Internal validation, Uncertainty analysis, Archetypical gauging station.

Highlights

- The HYPE model reproduces well the spatiotemporal variability of NO_3^--N and TP
- Multi-site calibration increases the spatial capability of nutrient catchment model
- Hydro-meteorological sub-catchment delineation is important for nutrient prediction

1. Introduction

High exports of nitrogen (N) and phosphorous (P) from agriculture are continuously threatening the aquatic ecosystem in surface waters and coastal areas throughout the world (Reusch et al., 2018). River nutrient loads are highly impacted by agricultural practices and land-use characteristics (Rode et al., 2009). The main contribution of N loads in Europe stems from agriculture, while the dominant sources of P are wastewater and dwellings (Delgado and Scalenghe, 2008; Whiters et al. 2014). The N to P ratio in the freshwater ecosystem is also much influenced by human activities on a global scale (Beusen et al., 2016). Hydrological transport has a strong impact on N exports, which are mostly regulated by subsurface flow (Lam et al., 2012). The high input of N (through mineral fertilizer and manure application) stimulates the rate of N processing in both terrestrial and aquatic ecosystems (Hall et al., 2009). Coastal algal blooms are also induced due to excess N inputs (Le et al., 2019) and P contributions. Recent studies have suggested that the export of P from terrestrial to stream systems is limited by the occurrence of storm events that exacerbate soil erosion (Lee et al., 2013). Previous studies revealed a higher inter-field variation of P influenced by soil characteristics and soil moisture conditions rather than by effects from the application of crop-rotation fertilizers (Kistner et al., 2013; Haygarth et al., 2014).

Catchment modeling has been widely used for hydrology sciences and environment-related research studies under different objectives. Distributed and semi-distributed process-based catchment models offer an opportunity to improve the physical understanding of processes and to formalize the knowledge of catchment systems thereby gained, and thus can be used as complex catchment management tools

(Jackson-Blake et al., 2016). Such models can be used to identify and address data gaps, to help in the design of monitoring strategies (McIntyre and Wheater, 2004; Jackson-Blake and Starrfelt, 2015) and testing and evaluation of environmental management strategies (Hashemi et al., 2016). Numerous catchment water quality models such as SWAT (Soil and Water Assessment Tool), INCA (INtegrated Catchment), HSPF (Hydrological Simulation Program Fortran) and HBV (Hydrologiska Byråns Vattenbalansavdelning) have been used in recent decades depending on researchers' specific objectives (Wellen et al., 2015). During the development and testing of the HYPE (HYdrological Predictions for the Environment) model Lindström et al. (2010) made a comparison with the most commonly used hydrological and water quality models (SWAT, INCA, MIKE BASIN (an integrated hydrological modeling system developed by the company DHI Water and Environment), and MONERIS (modeling Nutrient Emissions in RIver Systems) based on model complexity, input data constraints, ease of application, time effectiveness and performance of the model. The SWAT model needs an enormous amount of input data and neglects the entrance of groundwater into aquifers during hydrology-related simulations (Chahinian et al., 2011; Glavan et al., 2011), limiting its application without further adjustment. Wade et al. (2002) have reported that the INCA model is only focused on the river part of the aquatic system. Although some improvements have been incorporated in the latest version of the INCA model regarding nutrient exports from terrestrial parts (Jackson-Blake et al., 2016), this model is still overly complex for catchment studies (Jackson-Blake et al., 2017). The MIKE BASIN model is a water resource management tool featuring advances in user interface linked to ArcView GIS but is limited regarding nutrient transformation descriptions (Kumar et al., 2018).

MONERIS describes nutrient pathways and is based on a regression investigation without a description of detailed processes and water balance (Behrendt et al., 2002). With regard to water balance, water compartments, soil nutrient balance and dynamics, the HYPE model justifies all these characteristics as a process-based hydrological and water quality model (Lindström et al., 2010) that achieves a balance between complexity and process representation. Refsgaard (2001) stated that for operational use, conceptual models have major potential when compared to physically distributed models, though the latter show potential for research purposes.

Previous studies argued for the importance of gathering informative data from more than one location (e.g., catchment outlet) regarding spatial reference (Wellen et al., 2015; Moussa et al., 2007). Relying on observed data from a single location leads to the possibility of poorly simulated fluxes, especially in heterogeneous catchments (Beven, 2006). Calibration only at the outlet resulted in an over-optimistic evaluation of the model's capability to generate the dynamics at internal stations, which includes the source provenance and land-use consequences (Wellen et al., 2015). Jiang et al. (2014) suggested calibrating the process-based model at more than one station within the catchment using a multi-site calibration approach to minimize uncertainties of predictions of water quality. Studies related to HBV models concluded with the same suggestions (Pettersson et al., 2001). Besides, numerous studies have reported that a multi-objective calibration approach improves the optimization of model parameters, refines the internal processes and reduces the uncertainty in hydrological water quality modeling (Gupta et al., 1999; Lu et al., 2014; Van Griensven et al., 2006), compared to the traditionally used stepwise method.

Generally speaking, hydrological and water quality models are kept restricted to the gauging stations where they are calibrated (mostly at the outlet), and these are likely to fail to represent internal stations. There are limited studies that have validated the model performance internally within sub-catchments (e.g., Dunn et al., 2013). Wellen et al., (2015) also found in their study evaluating catchment nutrient water quality models that only 19% of studies from their database conducted calibration at more than one station; they therefore concluded that using data from more than one gauge station can lead to considerably improved identification of spatially distributed model parameters. They also suggested practicing the use of data from multi-sites for parameterization, calibration and validation to overcome overconfident assessment of the models. This additional internal validation of semi-distributed models can also increase confidence in applying these models for management purposes.

As mentioned above, the HYPE model was developed based on achieving an effective balance between data requirements and reasonable process representation. The HYPE model parameters are based on physiographical characteristics (such as land use and soil type) of the catchment rather than sub-catchment divisions. This enhances the transferral of model parameters to non-gauged catchments, which means that HYPE is not overly dependent on resolution or on the scale of the model (Lindström et al., 2010) when compared to other distributed physically-based models that are sensitive to multi-scale problems (Refsgaard, 1997). In terms of water quality and hydrology, the HYPE model has been shown to represent the measured discharge and nutrient concentrations reliably in distinct catchments that are subject to different climatic and anthropogenic conditions (Strömqvist et al., 2012; Jiang et al., 2014; Pechlivanidis and Arheimer, 2015;

Hundecha et al., 2016; Jomaa et al., 2016; Veinbergs et al., 2017). However, the model has not been tested widely at internal stations, which were not used for model calibration. Thus, the objective of this study was (i) to set up the HYPE model for NO_3^--N and TP concentration calculation in the heterogeneous Selke catchment using multi-site and multi-objective calibration, (ii) to test the capability of the model to represent the measured NO_3^--N and TP concentrations at eight internal gauging stations that were not considered for calibration, and (iii) to analyse the predictive uncertainty of the model for NO_3^--N and TP concentrations. To this end, the HYPE model was set up for the Selke catchment, which was delineated according to its internal stations at the outlet of sub-catchments. The DiffeRential Evolution Adaptive Metropolis (DREAM) tool (Vrugt et al., 2009) was used to calibrate the model and analyse the predictive uncertainty at three main gauging stations (Silberhuette, Meisdorf, and Hausneindorf) for discharge, and for NO_3^--N and TP concentrations. After this, the set of parameters obtained from the calibration process was further tested at eight internal gauge stations.

2. Methodology

2.1. Study area

The Selke catchment (463 km^2) is located in the lower range of the Harz Mountains in central Germany. The Selke discharges at its station at Hausneindorf into the River Bode, which continues into the River Elbe until it reaches the North Sea. Monitoring data have been available since 1993 at three gauging stations at the main stem of the Selke stream (Silberhuette, Meisdorf, and Hausneindorf). These three stations were used for the calibration of the HYPE model for discharge, NO_3^--N and TP concentrations. The elevation of the Selke catchment varies from 53 to 605 metres (Figure 1). Land use of the upper part is mainly dominated by three types of forest (broad-leaved, coniferous, and mixed) and in the lower parts most of the area is dominated by arable land (Figure 1), resulting in a 52% area of the catchment being covered by arable land and 35% by forest. A decrease in forest share from upstream towards downstream can be observed with an increase in arable land use. The mountain area is covered by cambisols (brown soils), whereas the lowland areas are dominated by chernozems (black soils). The annual mean precipitation in the mountain part is 792 mm y^{-1}. It then decreases to 450 mm y^{-1} towards downstream in lowland areas, resulting in an average of 660 mm y^{-1} precipitation for the whole Selke catchment (Haberlandt and Ebner, 2008). Compared to winter, in summer there is more precipitation, with a ratio of 1.35 between both periods. 9°C of mean temperature is recorded, with an average monthly high of 15.5°C in July and -1.8°C in January. From the mountains upstream towards the downstream area of the Selke catchment, the temperature increases due to lower elevation. Prevailing crops are winter wheat, triticale, winter barley, rye, corn, and rape. In the fertile lowland area, additionally, sugar beets are planted. Application of fertilizer in the Selke catchment range from 130-190 kgN ha^{-1} y^{-1} to 20-30 kgP ha^{-1} y^{-1}, according to a survey of farmers.

The long-term average discharge of the fourth-order Selke stream is 1.54 m^3 s^{-1} (1994-2014 is considered). Mean NO_3^--N and TP concentrations recorded at Hausneindorf are 3.15 mgN l^{-1} and 0.190 mgP l^{-1}, respectively. There is temporal variation in streamflow caused by high flows during winter periods (rainfall with additional snowmelt) and low flows with infrequent high flows characterized by extreme rainfall events in summer. Within areas close to the Selke catchment there were 16 precipitations and two climate stations. The density of precipitation stations was higher in mountain areas relative to lowland areas. The source and resolution of spatial and temporal data used in the model are presented in Table 2. There is a variation from weeks to a month in sampling frequency. For discharge, NO_3^--N and TP concentration, the time series data from 1994 to 1998 were used for calibration and from 1999 to 2014 for validation at three main stations. In addition, measured discharge data from 2001 to 2008 at sub-catchment 2 (Schäfertal, Figure 1) were used for spatial validation of discharge at this station. A summary of Selke catchment characteristics is given in Table 1.

Figure 1 is near here

Table 1 is near here

Eight internal stations represented as outlets of sub-catchments were used for the spatial validation of the HYPE model. These internal stations are monitored by State Agency for Flood Protection and Water Management of Saxony-Anhalt (LHW) at biweekly to monthly time steps. These were selected based on their locations and data availabilities regarding NO_3^--N and TP concentration observations. Stations 1, 2, 3, 4, and 5 are located in the forest-dominant part and Stations 6, 7, and 8 represent the downstream arable-land part of the Selke catchment (Figure 1). Station 2 (Schäfertal) represents an agriculture-dominated headwater sub-catchment. Station 6 is located at the outlet of a mixed agriculture- and urban-dominated sub-catchment in the lowlands. These internal stations (Table S1, Supplementary material) represent exports from all different land-use characteristics for the whole Selke catchment. For the internal stations, the duration of observed time series data varied between 1994 and 2014, depending on the station.

Table 2 is near here

2.2. HYPE model approach

The HYPE model is a process-based and semi-distributed model that simulates discharge, nutrient transport and transformation. Description of the model and governing equations are available in detail elsewhere (e.g., Lindström et al., 2010) and here only a summary of the model is given. For application of the HYPE model, the whole catchment was delineated into sub-catchments based on the Digital Elevation Model (DEM). Each sub-catchment was further divided into different combinations of land use and soil type units, jointly called soil land-use classes (SLCs), and also commonly known as hydrological response units (HRUs). Water flow and concentration of nutrients for each sub-catchment are accumulated as the area-weighted sum of respective values from all SLCs. Simulation of variables from every sub-catchment is routed between sub-catchments and then finally to the outlet of the catchment through flow connections (Lindström et al., 2010).

2.3. Setup of model and calibration

HYPE was set up for a period of 21 years (1994-2014) for the simulation of discharge and NO_3^--N and TP concentrations. According to the split-sample approach, the model was calibrated in the period 1994-1998 and validated in the period 1999-2014. Simulation from 1993 was excluded from the model evaluation because that year was used as a warming-up period for the model. For the study, the catchment was divided into 19 soil types and ten land-use classes. In total, the Selke catchment was categorized into 117 soil land-use classes (SLCs) and divided into 11 sub-catchments (three main Stations for model calibration and eight internal Stations for model evaluation). All mean daily discharge data were calculated from 15 minute high frequency measurements. During the calibration period (1994-1998) we used biweekly observed N and P concentration data. Daily data of precipitation and mean temperature for discharge simulation were taken from the nearby monitoring stations for the relevant sub-catchments. Data relating to different agricultural practices, main crops and sowing/harvesting time were taken from previous data published for the Selke catchment (Kistner et al., 2013) and kept constant for the whole simulation period from 1993-2015. Fertilizer application rates did not show significant changes since 1993 in the study area (Häußermann et al. 2019). Residue amounts from plants and animals and their dates of the application were defined on the basis of livestock types and previous model applications in the Selke catchment (Jiang et al., 2015). Three-point source input data sets were used from six sewage treatment plants in the Selke catchment from 1994 to 2014, depending on their availability.

For calibration of the model, a multi-site and multi-objective method was implemented using the DREAM tool for parameter optimization of both discharge and water quality parameters. The DREAM tool was coupled with HYPE using MATLAB scripts. Discharge and water quality parameters were calibrated simultaneously (multi-objective) at Silberhuette, Meisdorf, and Hausneindorf at the same time (multi-site) using 10,000 iterations. The DREAM tool is based on the Markov chain Monte Carlo (MCMC) approach, developed by Vrugt et al. (2009). It runs multiple trajectories in parallel to explore targeted posterior distribution. This works on the principle of self-adaptive random sampling. It is a globally used tool for the research and optimization for Bayesian inference of the posterior probability density function of model parameters (Schoups and

Vrugt, 2010). It has been successfully applied in various model calibration studies (Jiang et al., 2015; Decker et al. 2012). The multi-site approach was applied to account for the impact on hydrological and N processes by spatial inconsistencies in climate patterns, land use, topography and soil type.

HYPE model parameters are divided into three categoriesas a) general parameters, b) land use dependent parameters and c) soil dependent parameters. (Jiang et al., 2014). Discharge and NO_3^--N-related sensitive parameters were identified in Jiang et al. (2014). TP-related sensitive parameters were taken from Namugize et al. (2017) and identified through manual calibration. Optimization of the sensitive parameters was done by using DREAM. In this step, the initial values of the parameters and the range of calibrated parameters were based on values taken from previous application of HYPE (Jiang et al., 2014; Namugize et al., 2017). Details of the parameters are given in Table 3. Two calibration schemes were used for this study: Scheme 1: calibration only at Hausneindorf, and Scheme 2: calibration at Hausneindorf, Meisdorf, and Silberhuette. Improved model performance was obtained by Scheme 2 due to the application of a multi-site and multi-objective approach. The detailed model performance results from Scheme 2 are discussed in section 4. Results from Scheme 1 are given in Section S2 (Supplementary material) for gauge stations Hausneindorf, Meisdorf, and Silberhuette as well as for internal stations.

2.4. Model performance criteria

The capability of the model to predict discharge, NO_3^--N and TP concentrations was evaluated. Statistical methods and graphical observations were used for the assessment of the model's performance. For the evaluation of hydrological modeling, the Nash-Sutcliffe Efficiency (NSE) measure is widely used. NSE is dependent on factors like size of the samples, magnitudinal bias, outliers, time-offset of hydrograph models and intervals between hydrological sampling data (Jain and Sudheer, 2008; McCuen et al., 2006), resulting in an incomplete evaluation of model performance. Thus, two additional statistical criteria were considered in the analysis, percentage bias (PBIAS) and mean values (mean observed vs mean simulated). A more detailed description of these criteria has been extensively covered in many previous studies (e.g., Gupta et al., 1999; McCuen et al., 2006; Moriasi et al., 2007; Nash and Sutcliffe, 1970; Ullrich and Volk, 2010). The above-mentioned criteria were evaluated by the following formulas:

$$NSE=1-\frac{\sum_{i=1}^{n}\left(Y_i^{sim}-Y_i^{obs}\right)}{\sum_{i=1}^{n}\left(Y_i^{obs}-\overline{Y}^{obs}\right)} \quad (1)$$

$$PBIAS=\frac{\sum_{i=1}^{n}\left(Y_i^{sim}-Y_i^{obs}\right)}{\sum_{i=1}^{n} Y_i^{obs}} \quad (2)$$

where Y_i^{sim} and Y_i^{obs} are the i^{th} simulated and observed values for the criteria being evaluated, respectively, $\overline{Y}^{obs}$ and $\overline{Y}^{sim}$ are the mean values of observed and simulated results for the whole duration, respectively, and *n* is the total number of observations.

2.5. Uncertainty approach

DREAM was used for the uncertainty analysis of the HYPE model for discharge, NO_3^--N and TP concentration simulations. It has been used successfully in other studies for the uncertainty analysis of hydrological and water quality models (Jiang et al., 2015).

Estimation and assessment, both of parameters and total uncertainty, were conducted by DREAM. Parameter uncertainty is related to the prediction uncertainty resulting from the interaction of parameters and the complexity of the model. The total uncertainty is related to the prediction uncertainty resulting from the non-unique parameters' behavior and structure of the model. In this study 95% confidence interval band of parameter and total uncertainty was obtained from 10,000 MCMC estimations while the total uncertainty range was generated with random errors (normal distribution). Three different criteria were used for quantification of prediction uncertainty of the model for discharge, NO_3^--N and TP concentration simulations. Assessment of 95% confidence interval sharpness was done by Average Relative Interval Length (Jin et al., 2010). The percentage of observations embodied by the 95% predicted confidence intervals (PCI) was used for reliability assessment. Percentage of observed concentrations connected by Unit Confidence Interval (PUCI) was used to assess the credibility of 95% confidence intervals and was calculated on the basis of the average relative interval length (ARIL) and PCI (Lu et al., 2011). ARIL and PUCI were calculated according to equations (3) and (4).

$$ARIL = \frac{1}{n}\sum\frac{(Limit_{Upper,t}-Limit_{Lower,t})}{Q_{obs,t}} \quad (3)$$

$$PUCI = \frac{(1.0-Abs(PCI-0.95))}{ARIL} \quad (4),$$

Where $Limit_{Upper,\ t}$ and $Limit_{Lower,\ t}$ are upper and lower boundaries, respectively, n is the time interval and $Q_{obs,\ t}$ is the measured observation at t_{th} time.

Table 3 is near here

3 Results

3.1 Calibration schemes

Validation PBIAS values at three main stations and at internal stations for these two schemes are shown in Table 4. Calibration using Scheme 2 (calibration using the three main stations) shows better results at main stations as well as at internal stations than calibration using Scheme 1 (where only observation from the outlet of the Selke was considered). Respective means of PBIAS values for NO_3^--N and TP concentrations were 20.2% and -3.0% (Table 4) at main stations and 5.1% and 20.9% (Table 4) at internal stations using calibration Scheme 1. However, all PBIAS values that were obtained in calibration Scheme 2 showed better results than calibrated Scheme 1, except for the PBIAS value of TP concentration. The latter was lower at three main stations with a PBIAS value of -7.6% because TP concentration was mainly controlled by point source data, which were not consistent for the whole validation period at all point source data locations. Due to better comparative results obtained by calibration Scheme 2 at the main three stations as well as internal stations, results from calibration Scheme 2 were used for further discussion and conclusion.

Table 4 is near here

3.2 Calibrated parameters

The calibrated model parameters of discharge, NO_3^--N and TP concentration are listed in Table 3 with their physical interpretation, initial values and range as well as their optimized values. The most sensitive parameters (Jiang et al., 2014) for discharge are *wcep* (for brown soil which is the dominant soil type in mountain areas), *rivvel* (as a general parameter) and *cevp* (for the most dominant land use: arable land and forest). Velocity of flow (*rivvel*) in the river is responsible for the presentation of the hydrograph. *Epotdist* and *cevp* are important as these control evapotranspiration. From NO_3^--N-related processes, the *uptsoil* parameter (for arable-dominant land) was more sensitive than other parameters and represents the share of uptake by plants from the first layer of soil. Parameter *denitr* is important and sensitive as it controls the denitrification rate of NO_3^--N. Among TP-related parameters, *sedexp* is the most sensitive, and is responsible for the sedimentation factor. The second most sensitive parameter for TP was *pnratio*, which is responsible for the relationship between N and P for plant uptake. It was observed that for discharge, NO_3^--N and TP, most of the calibrated parameters have optimized values close to the initial given values and not close to minimum and maximum limits. All optimized values of calibrated parameters are given in Table 3.

3.3 Discharge simulation

Discharge simulation results criteria (Table 5) showed that the model performance for both calibration (1994-1998) and validation (1999-2014) functioned reasonably well. The HYPE model was able to capture the seasonal behavior for all three main gauging stations (Silberhuette, Meisdorf, and Hausneindorf) for both calibration and validation periods and during low- and high-flow conditions (Figure S1, Supplementary material). During the calibration period, the highest NSE value (0.87) was recorded in the uppermost station Silberhuette and the lowest NSE value (0.84) at the outlet station Hausneindorf. The PBIAS of water balance for the calibrated period at all three main stations was between -4.8% and 2.1%, which showed the best representation of discharge by the model. During the validation period (1999-2014), the highest and lowest NSE values of 0.76 (at Silberhuette) and 0.71 (at Hausneindorf) were observed. The water balance was better represented by the model during the calibration period compared to the validation period. During the validation period, the water balance PBIAS values at Silberhuette, Meisdorf, and Hausneindorf were 11.9%, 3.0%, and 18.0%, respectively. Overall discharge performance during both calibration and validation was very good and good, respectively, according to the evaluation criteria of Moriasi et al. (2007).

Table 5 is near here

3.4 Nitrogen simulation

The model performance results of NO_3^--N load simulations are presented in Table 5. Load simulations were well represented by the model during both calibration and validation periods. The highest and lowest NSE values of NO_3^--N loads were 0.93 and 0.70 during both calibration and validation periods. During the calibration, NO_3^--N load simulations were well represented by the model with the highest performance at Silberhuette (NSE = 0.93, PBIAS = -2.1%) and the lowest performance at Hausneindorf (NSE = 0.74, PBIAS = -5.7%). For the simulation of NO_3^--N, the model covered both higher and lower values of observed NO_3^--N concentrations, represented in Figure 2. The model performance at Silberhuette was better when compared to Meisdorf due to the clear pattern of seasonal behavior. The highest values of NSE = 0.64 and 0.49 for NO_3^--N concentrations were found at Silberhuette during calibration and validation periods, respectively (Table S3.1).

Figure 2 is near here

NSE values of 0.58 and 0.41 for NO_3^--N concentrations (Table S3.1) were attained at the Meisdorf station during the calibration and validation period, respectively. PBIAS values for NO_3^--N concentrations were between 2.8% and 11.0% during both calibration and validation periods, respectively (Table S3.1), which shows the satisfactory performance of the model at two upper stations. The model performance was not as good at the outlet (Hausneindorf), compared to the upper part of the Selke catchment (PBIAS = -21.0% for NO_3^--N concentration). Mean simulated NO_3^--N concentration of 3.10 mgN l^{-1} was detected against the mean observed 3.18 mgN l^{-1} at Hausneindorf station. Mean observed and mean simulated NO_3^--N concentrations at three main stations are given in the supplementary material (Table S3.2).

3.5 Phosphorous simulations

Model performance for TP loads is given in Table 5, and TP simulated concentrations are presented in Figure 3. For TP loads, the best model performances were obtained at the upper two stations (Silberhuette and Meisdorf). During the calibration period, NSE values of 0.48 and 0.53 were achieved at Silberhuette and Meisdorf, respectively. However, during the validation period, Silberhuette had a higher NSE (0.52) compared to Meisdorf (0.46). PBIAS values of TP loads for two upper stations were observed within a satisfactory range (-20.0 to 11.5%) for both calibration and validation periods (Table 5).

Figure 3 is near here

TP measured concentration values for both upper gauging stations were predicted well for both the calibration and validation periods. PBIAS values for TP concentration were 0.9% and -11.0% for Silberhuette and 2.6% and -25.0% for Meisdorf during calibration and validation, respectively. Mean simulated and mean observed TP concentration values testify to the satisfactory performance of the model (Table S3.2). At Hausneindorf, the performance was affected by the combination of unknown point sources and farming activities in the agricultural downstream area of the Selke catchment. NSE values for TP loads were 0.13 for calibration and 0.20 for validation. The performance of the model was disturbed by the high observed concentration during low flow, which was underestimated by the model (Figure 3). The PBIAS values for TP concentration during calibration and validation at Hausneindorf were -6.6% and -13.0%, respectively. A mean simulated TP concentration of 0.190 mgP l^{-1} was recorded against 0.210 mgP l^{-1} mean observed concentration at Hausneindorf station during the calibration period (Table S3.2, Supplementary material).

3.6 Spatial validation of NO_3^--N and TP concentration at internal stations

Calibrated model parameters were further tested at internal stations that were not included in the calibration mode. The calibrated discharge parameter set was used at the Schäfertal headwater (Station 2, Figure 1) for discharge simulation. Model validation resulted in NSE = 0.25 and PBIAS = -20.0% for the period from 2001 to 2008. The mean simulated discharge was 7.5 l s^{-1} and the mean observed discharged was 10.0 l s^{-1} for the whole period.

Observed and simulated NO_3^--N concentrations at five internal stations are shown in Figure 4, and performance criteria are given in Table 6. The temporal frequency of data was different for internal stations. These nutrient concentrations were simulated by using the same single set of parameters that were obtained after the multi-site calibration process at three main stations (Silberhuette, Meisdorf, and Hausneindorf).

Table 6 is near here

Figure 4 presents NO_3^--N concentration simulations of five sub-catchments (Stations 1, 2, 4, 8, and 10). For most of the stations, a good agreement between simulated and observed NO_3^--N concentrations was found (Table 5), reflecting the capability of HYPE to represent NO_3^--N concentration at different land-use dominated sub-catchments. The difference between mean simulated (9.47 mgN l^{-1}) and mean observed concentration (8.46 mgN l^{-1}) values was greater at Station 8 (Figure 4d) due to agriculture-dominant land use in that particular sub-catchment.

The model was able to represent well the NO_3^--N concentration of the agricultural headwater catchment 2 (Station 2, Figure 4e), with simulated and observed mean NO_3-N concentration of 4.35 mgN l^{-1} and 4.52 mgN l^{-1}, respectively. The model was able to accurately capture the NO_3-N concentration dynamics of forest-dominated sub-catchments (Stations 1, 4, and 5), simulating a mean concentration of 1.67 mgN l^{-1} and 2.00 mgN l^{-1} for Stations 1 and 4 (Figures 4a-b), respectively. PBIAS (%) values for these five sub-catchments are given in Table 5, with the highest performance at sub-catchment 1 (PBIAS = 2.5%) and the lowest at sub-catchment 8 (PBIAS = 14.2%).

Figure 4 is near here

The model performance criteria of TP concentration are given in Table 5 (Stations 3, 4, 5, 6, and 7) and the simulated time series of five sub-catchments are shown in Figure 5. A good agreement of observed and simulated mean TP concentrations was obtained.

Sub-catchments of Stations 3, 4, and 5 (Stations at the main stem) are forest-dominated with some share of arable land, and the measured TP concentrations were between 0.031 mgP l^{-1} and 0.051 mgP l^{-1} (Table 6).

The largest difference between mean measured (0.051 mgP l^{-1}) and simulated (0.036 mgP l^{-1}) TP concentrations was observed at sub-catchment 4, which is located downstream of a point source impacted reach. The point source impacted sub-catchment 6 (headwater catchment) showed a significant difference between its observed (0.200 mgP l^{-1}) and simulated (0.300 mgP l^{-1}) TP concentrations. Sub-catchment 7 is located at the main stem in the agriculture part of the Selke catchment and represents two different periods of TP concentration. Differences between mean observed and simulated TP concentrations were primarily caused by simulation errors in the first period of higher point source pollution. The PBIAS values for all sub-catchments range from -25.3% to 34.3%. Overall, the model performance at the stations situated at the main stream is better than that located in the tributaries (Figure 1 and Table 5).

Figure 5 is near here

3.7 Uncertainty Analysis

Predicted uncertainty and total uncertainty (95%) ranges of daily discharge, NO_3^--N concentration and TP concentration are shown for the outlet of the Selke catchment (Hausneindorf) in Figure 7 and the corresponding ARIL, PCI and PUCI of 95% predicted confidence intervals are listed in Table 7. The band for parameter uncertainty (black shaded area) of discharge simulation is narrow, indicating low uncertainty related to parameter optimization and showing a very similar variation to the observed values represented by red dots. This is confirmed by the small value of ARIL (0.093) characterizing the narrow range (black range) of 95% confidence intervals. The total uncertainty of discharge prediction was much higher than the parameter uncertainty, and this is reflected by the high value of ARIL (4.139) and a large range of 95% confidence intervals (grey band in Figure 7).

Table 7 is near here

Parameter and total uncertainty ranges for NO_3^--N concentration simulations are much wider, which indicates higher uncertainty when compared to discharge simulations. Parameter-related uncertainty ranges are much narrower than the total uncertainty, indicating that the predicted uncertainty is mainly caused by the uncertainty of model structure error and measurement error. This is confirmed by ARIL values of total uncertainty (1.242) and parameter uncertainty (0.150). Many of the observed concentration values are contained in the total prediction confidence interval (92%). A lower PUCI value of total uncertainty (0.783) was evidence of higher uncertainty in the predictions of NO_3^--N simulations compared to discharge simulations.

Figure 6 is near here

In comparison to uncertainty for NO_3^--N concentration simulation, higher uncertainty was revealed for the simulation of TP concentrations with wider ranges of 95% parameter and the total uncertainty confidence intervals. The PCI value for total uncertainty of TP concentration simulations (96.1%) was higher than the PCI (92.0%) of NO_3^--N concentration simulations. A lower value of PUCI for TP concentration simulations showed that the main drivers of uncertainty stem from model structure and measurement errors, a very similar result to the one obtained from discharge and NO_3^--N concentration simulations.

4 Discussion

Our findings indicate a strong impact of the selected calibration scheme on NO_3^--N and TP concentration evaluation at the catchment scale. When three gauging stations that accurately reflected the catchment heterogeneity from upstream to downstream in the calibration mode were considered, the HYPE model was highly effective at representing the measured discharge and nutrient concentrations at internal stations. In other words, good agreement between the measured and simulated NO_3^--N concentration was achieved only when three main stations for model calibration were considered, instead of using only the data from the catchment outlet at gauging station Hausneindorf. Interestingly, this improvement was very similar for the main stations and those additional internal stations that often provide for more uniform land use and which hence showed much larger variation in NO_3^--N concentration when compared to the main stations. This improvement can be explained by the significant hydrogeographical differences between the three sub-catchments represented by the three main gauge stations. The uppermost sub-catchment represented by the gauge station Silberhuette is the wettest one with an average long-term discharge of 414 mm and mixed agricultural and forest land use. The two downstream gauge stations represent much drier conditions with only 72 mm as an average long-term discharge in the intermediate sub-catchment between gauge stations Silberhuette and Meisdorf and 36 mm in the lowest sub-catchment representing the area between gauge stations Meisdorf and Hausneindorf. Moreover, these two lower sub-catchments differ markedly in geology and land use. Calibrating the model only at the catchment outlet does not allow for indicating a representative NO_3^--N parameterization for these heterogeneous catchment conditions. It is important to note that internal catchments representing single land-use types like forest (sub-catchment 1) or arable land (sub-catchments 2 and 8) showed acceptable or good model validation results, even though model calibration was carried out at gauge stations with mixed land-use patterns. This indicates that the consideration of meteorological and hydrogeological area properties seems to be of significantly greater importance for NO_3^--N model calibration than consideration of land use. Urban areas did not contribute significantly to NO_3^--N load and point source inputs were controlled by sewage systems, as previous studies like those of Jiang et al. (2015) and Rode et al. (2016) have also found.

After increasing the calibration gauging stations for TP simulation from one (calibration Scheme 1) to three (calibration Scheme 2), the validation results improved considerably only for internal stations. Looking at the three main stations, the PBIAS of the validation results even became slightly worse. These small changes in model performance when shifting from calibration Scheme 1 to calibration Scheme 2 are likely caused by the variability of TP values at the three stations. These were highest at the most downstream gauging station Hausneindorf. Adding the other two upstream gauge stations to the calibration procedure will not drastically change the parameters because of the lower weight of these stations on the optimization process. Nevertheless, including the two upstream stations allowed the model to also better consider very low TP concentrations in those sub-catchments not impacted by point source inputs from sewage systems. This is reflected in the markedly improved mean PBIAS of the internal stations. The results show that, in contrast to NO_3^--N concentrations, TP concentrations are much more strongly affected by point sources. Even sub-catchments with a share of nearly 92% of agricultural land use (sub-catchment 2) did not reveal considerably different concentration ranges than the mostly forested (sub-catchment 3) areas, at least during low-flow conditions. This was well captured by the model.

The model performance for TP concentration simulations at internal stations was lower than for NO_3^--N concentration simulations. A rise in TP concentration from upstream to downstream can be explained by an increased share of urban and arable land use. The HYPE model showed an underestimation of TP concentration in the forest-dominated sub-catchments 3 and 4 (Figures 5a and 5b) because of an underestimation of some high-flow events since the main export of TP occurs as a result of overland flow (Jiang and Rode, 2012). Higher TP concentrations were found at sub-catchments 6 and 7 (Figures 5d and 5e) as these stations provide coverage over a higher share of urban and arable land. TP concentrations in these sub-catchments (6 and 7) were 2.5 times higher than in sub-catchments 3 and 4, which may be explained by agricultural sediment and urban soluble P inputs (Lee et al., 2013). The HYPE model overestimates TP concentration in sub-catchments 6 and 7, which was possibly caused by uncertainties in point source data. Simulated high-flow TP concentrations were only sporadically captured because of mostly low-frequency sampling, which reveals considerable uncertainties in assessing the model’s performance (Yin et al., 2016). Our findings suggest that only rough estimates for internal stations can be achieved when the calibration of the model is conducted exclusively at the catchment outlet. Model performance of internal catchment stations could be considerably improved if additional stations (Silberhuette and Meisdorf), representing the upper and middle forest parts of the Selke catchment, were included in the calibration of discharge, NO_3^--N and TP parameters. These findings are in line with results from the meta-analysis of Wellen et al. (2015). They argued that calibrating a model only at the outlet may lead to an over-optimistic model evaluation for capturing internal catchment process dynamics.

4.1 Importance of input data

The availability of accurate input data has an impact from the very beginning of a study, and this can help to improve a model's performance. In the validation period, the number of precipitation stations decreased by almost half. This was the main reason for the lower performance of the model at three main stations for the validation period. Precipitation and discharge data were established from 2001 to 2008 for Station 2, which is a 1.45 km^2 agriculture catchment. In the beginning, interpolated precipitation data were used for this station, and the mean annual precipitation used for this station was 605 mm, which led to a large underestimation of discharge, with a PBIAS value of -45.0%. Later, station-specific precipitation data from here were used from 2001 to 2008 which leads to mean annual precipitation of 677 mm. By using accurate precipitation data for this station, the HYPE model showed much better performance for discharge (without calibration) at this station with a PBIAS value of -20.0% and NSE of 0.25. This improved discharge results in much more accurate NO_3^--N concentration simulations at this sub-catchment.

4.2 Uncertainty analysis

Regarding uncertainties surrounding our study, more than 98.3% of discharge, 92.0% of NO_3^--N and 96.1% of TP concentration observations were included in the range of 95% predicted confidence intervals (PCI). The total prediction uncertainty was less influenced by parameter uncertainty with a high value of PUCI than from other uncertainty sources like structural and measurement uncertainties. This reflects the sensible parameterization of discharge and water quality parameters. These results are in line with former studies that also used Bayesian uncertainty analysis (e.g., Yang et al., 2007; Jiang et al., 2019). Total uncertainty increased from discharge to NO_3^--N and TP concentrations. Prediction uncertainty relating to water quality simulations can be improved by using high-frequency observation data during calibration (Jackson-Blake and Starrfelt, 2015; Jiang et al., 2019). The higher total uncertainty of TP simulations can be explained by the daily simulation time step (resulting in higher short-term variability of TP concentrations), the lack of available event data and uncertainties in point source TP concentration data (Dean et al., 2008). Uncertainty analysis of water quantity as well as for water quality simulation showed the acceptable and justifiable performance of model assessment.

5 Conclusions

Our spatial validation of the HYPE model suggests that consistent spatially-distributed results can be achieved only when enough observations from representative sub-catchments of the hydrological characteristics of the whole catchment are considered in the calibration mode. This is true for NO_3^--N simulations and is even more important for TP concentrations. For the latter, uncertainties of simulated concentrations are higher than for NO_3^--N and for discharge. Our findings suggest that hydro-meteorological catchment characteristics are more crucial than the land-use patterns consideration for delineating uniform sub-catchments that allow for a reasonable NO_3^--N simulation within these subunits. This has key implications for the choice of calibration gauge stations within a given larger catchment. We assume that in mesoscale catchments, it will mostly not be sufficient to choose only one station at the outlet of the catchment to calibrate the model if reasonable spatial simulations are also needed within the whole catchment. Due to low agricultural TP losses during high-flow events, catchment point sources clearly dominated TP loss in our study, giving us reasonable proof of the model with sparse but long-term data. If agricultural TP losses are higher, acceptable distributed-model testing is only possible if more high-frequency data are available to capture TP losses during high-flow events in more detail. Likely, these findings are also valid for other distributed nutrient transport models as well as the HYPE model selected for this study. Bearing these requirements in mind, it is possible to support the development and evaluation of nutrient management and mitigation strategies using semi-distributed hydrological nutrient models, which are also valid for smaller subunits within a given catchment.

Acknowledgement

Salman Ghaffar is funded by the Higher Education Commission of Pakistan (HEC). This work was also partly supported by the BONUS MIRACLE project (Call 2014-73) under the joint Baltic Sea Research and Development Programme (Art 185), co-funded jointly by the European Union.

References

Behrendt, H., Kornmilch, M., Opitz, D., Schmoll, O., Scholz, G., 2002. Estimation of the nutrient inputs into river systems – experiences from German rivers. Regional Environmental Change, https://doi.org/10.1007/s10113-002-0042-3.

Beusen, A.H.W., Bouwman, A.F., Van Beek, L.P.H., Mogollón, J.M., Middelburg, J.J., 2016. Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. Biogeosciences, https://doi.org/10.5194/bg-13-2441-2016.

Beven, K., 2006. A Manifesto for the Equifinality Thesis. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2005.07.007.

Bouwman, A.F., Beusen, A.H.W., Billen, G., 2009. Human alteration of the global nitrogen and phosphorus soil balances for the period 1970–2050. Global Biogeochemical Cycles, https://doi.org/10.1029/2009GB003576.

Chahinian, N., Tournoud, M.-G., Perrin, J.-L., Picot, B., 2011. Flow and nutrient transport in intermittent rivers: A modeling case-study on the Vène River using SWAT 2005. Hydrological Sciences Journal, https://doi.org/10.1080/02626667.2011.

Dean, S., Freer, J., Beven, K., Wade, A.J., Butterfield, D., 2008. Uncertainty assessment of a process-based integrated catchment model of phosphorus. Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-008-0273-z.

Dekker SC, Vrugt JA, Elkington RJ. 2012. Significant variation in vegetation characteristics and dynamics from ecohydrological optimality of net carbon profit. Ecohydrology, https://doi.org/10.1002/eco.177.

Delgado, A., and R. Scalenghe (2008), Aspects of phosphorus transfer from soils in Europe, Journal of Plant Nutrition and Soil Science, https://doi.org/10.1002/jpln.200625052.

Dunn, S.M., Johnston, L., Taylor, C., Watson, H., Cook, Y., Langan, S.J., 2013. Capability and limitations of a simple grid-based model for simulating land use influences on stream nitrate concentrations. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2013.10.016.

Glavan, M., White, S., Holman, I., 2011. Evaluation of river water quality simulations at a daily time step – experience with SWAT in the Axe catchment, UK. CLEAN Soil Air Water, https://doi.org/10.1002/clen.200900298.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. Journal of Hydrologic Engineering, https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135).

Haberlandt, U., Ebner, A.-D., 2008. A space-time hybrid hourly rainfall model for derived flood frequency analysis. Hydrology and Earth System Sciences, https://doi.org/10.5194/hess-12-1353-2008.

Hall, R.O., Tank, J.L., Sobota, D.J., 2009. Nitrate removal in stream ecosystems measured by 15N addition experiments: Total uptake. American Society of Limnology and Oceanography, https://doi.org/10.4319/lo.2009.54.3.0653.

Hashemi, F., Olesen, J. E., Dalgaard, T., Børgesen, C. D., 2016. Review of scenario analyses to reduce agricultural nitrogen and phosphorus loading to the aquatic environment. Science of the Total Environment, https://doi.org/10.1016/j.scitotenv.2016.08.141.

Häußermann, U., Bach, M., Klement, L., Breuer, L., 2019. Nitrogen areal balance for Germany with regional structuring in states and counties – 1995-2017; UBA Texte 131/2019; Umweltbundesamt: Dessau-Roßlau, Germany.

Haygarth, P.M., Jarvie, H. P., Powers, S. M., Sharpley, A. N., 2014. Sustainable phosphorus management and the need for a long-term perspective: The Legacy

Hypothesis. Environmental Science & Technology, https://doi.org/10.1021/es502852s.

Hundecha, Y., Arheimer, B., Donnelly, C., Pechlivanidis, I., 2016. A regional parameter estimation scheme for a pan-European multi-basin model. Journal of Hydrology, https://doi.org/10.1016/j.ejrh.2016.04.002.

Jackson-Blake, L.A., Sample, J.E., Wade, A.J., Helliwell, R.C., Skeffington, R.A., 2017. Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, SimplyP, and INCA-P. Water Resources Research, https://doi.org/10.1002/2016WR020132.

Jackson-Blake, L.A., Starrfelt, J., 2015. Do higher data frequency and Bayesian auto-calibration lead to better model calibration? Insights from an application of INCA-P, a process-based river phosphorus model. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2015.05.001.

Jackson-Blake, L.A. Wade, A.J., Futter, M.N., Butterfield, D., Couture, R.M., Cox, B.A., Crossman, J., Ekholm, P., Halliday, S.J., Jin, L., Lawrence, D.S L., Lepistö, A., Lin, Y., Rankinen, K., Whitehead, P.G., 2016. The INtegrated CAtchment model of phosphorus dynamics (INCA-P): Description and demonstration of new model structure and equations. Environmental modeling and Software, https://doi.org/10.1016/j.envsoft.2016.05.022 .

Jain S.K., Sudheer, K.P., 2008. Fitting of hydrologic models: a close look at the nash–sutcliffe index. Journal of Hydrologic Engineering, https://doi.org/10.1061/(ASCE)1084-0699(2008)13:10(981).

Jiang, S., Jomaa, S., Büttner, O., Meon, G., Rode, M., 2015. Multi-site identification of a distributed hydrological nitrogen model using Bayesian uncertainty analysis. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2015.09.009.

Jiang, S., Jomaa, S., Rode, M., 2014. modeling inorganic nitrogen leaching in nested mesoscale catchments in central Germany. Ecohydrology, https://doi.org/10.1002/eco.1462 .

Jiang, S., Rode, M., 2012. Modeling water flow and nutrient losses (nitrogen, phosphorus) at a nested meso scale catchment, Germany.

Jiang, S.Y., Zhang, Q., Werner, A.D., Wellen, C., Jomaa, S., Zhu, Q.D., Büttner, O., Meon, G., Rode, M., 2019. Effects of stream nitrate data frequency on watershed model performance and prediction uncertainty. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2018.11.049.

Jin, X., Xu, C.-Y., Zhang, Q., Singh, V.P., 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2009.12.028.

Jomaa, S., Jiang, S., Thraen, D., Rode, M., 2016. Modeling the effect of different agricultural practices on stream nitrogen load in central Germany. Energy, Sustainability and Society, https://doi.org/10.1186/s13705-016-0077-9.

Kistner, I., Ollesch, G., Meissner, R., Rode, M., 2013. Spatial-temporal dynamics of water soluble phosphorus in the topsoil of a low mountain range catchment. Agriculture, Ecosystems and Environment, https://doi.org/10.1016/j.agee.2013.05.016.

Kumar, K.S., Galkate, R.V., Tiwari, H.L., 2018. River basin modeling for Shipra River using MIKE BASIN. Journal of Hydraulic Engineering, https://doi.org/10.1080/09715010.2018.1534219 .

Lam, Q.D., Schmalz, B., Fohrer, N., 2012. Assessing the spatial and temporal variations of water quality in lowland areas, Northern Germany. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2012.03.011.

Lee, T.Y., Huang, J.C., Kao, S.J., Tung, C.P., 2013. Temporal variation of nitrate and phosphate transport in headwater catchments: the hydrological controls and land use alteration. Biogeosciences, https://doi.org/10.5194/bg-10-2617-2013.

Le Moal, M., Gascuel-Odoux, C., Manesguen, A., Souchon, Y., 2019. Eutrophication: A new wine in an old bottle? Science of the total environment, https://doi.org/10.1016/j.scitotenv.2018.09.139.

Lu, L., Xu, C.Y., Xia, J., Engeland, K., Reggiani, P., 2011. Uncertainty estimates by Bayesian method with likelihood of AR (1) plus Normalmodel and AR (1) plus Multi-Normal model in different time-scaleshydrological models, Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2011.05.052.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology Research, https://doi.org/10.2166/nh.2010.007 .

Lu, S., Kayastha, N., Thodsen, H., van Griensven, A., Andersen, H.E., 2014. Multiobjective calibration for comparing channel sediment routing models in the

Soil and Water Assessment Tool. Journal of Environmental Quality, https://doi.org/10.2134/jeq2011.0364.

Malagó, A., Bouraoui, F., Grizzetti, B., De Roo, A., 2019. modeling nutrient fluxes into the Mediterranean Sea. Journal of Hydrology: Regional Studies, https://doi.org/10.1016/j.ejrh.2019.01.004.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe Efficiency Index. Journal of Hydrologic Engineering, http://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597).

McIntyre, N.R., Wheater, H.S., 2004. Calibration of an in-river phosphorus model: prior evaluation of data needs and model uncertainty. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2003.12.003.

Moussa, R., Chahinian, N., Bocquillon, C., 2007. Distributed hydrological modeling of a Mediterranean mountainous catchment – Model construction and multi-site validation. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2007.01.028.

Moriasi, D.N., Arnold, J.G., Van Liew. M.W., Bingner, R.L., Harmel, R.D., Veith. T.L., 2007. Mode evaluation guidlines for systematic quantification in watershed simulations. Transactions of the ASABE, https://doi.org/10.13031/2013.23153.

Namugize, J.N., Jewitt, G.P.W., Clark, D., Strömqvist, J., 2017. Assessment of the Hype Model for Simulation of Water and Nutrients in the Upper uMngeni River Catchment in South Africa. Hydrology and Earth System Sciences Discussions, https://doi.org/10.5194/hess-2017-365.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. Journal of Hydrology, https://doi.org/10.1016/0022-1694(70)90255-6.

Pechlivanidis, I., Arheimer, B., 2015. Large-scale hydrological modeling by using modified PUB recommendations: the India-HYPE case. Hydrology and Earth System Sciences, https://doi.org/10.5194/hess-19-4559-2015.

Pettersson, A., Arheimer, B., Johansson, B., 2001. Nitrogen concentrations simulated with HBV-N: new response function and calibration strategy: paper presented at the Nordic Hydrological Conference (Uppsala, Sweden – June, 2000). Hydrology Research, https://doi.org/10.2166/nh.2001.0014.

Reusch, T.B.H. , Dierking, J., Andersson, H.C., Bonsdorff, E., Carstensen, J., Casini, M., Czajkowski, M., Hasler, B., Hinsby, K., Hyytiäinen, K., Johannesson, K., Jomaa, S., Jormalainen, V., Kuosa, H., Kurland, S., Laikre, L., MacKenzie, B.R., Margonski, P., Melzner, F., Oesterwind, D., Ojaveer, H., Refsgaard, J.C., Sandström, A., Schwarz, G., Tonderski, K., Winder, M., Zandersen, M., 2018. The Baltic Sea as a time machine for the future coastal ocean. Science Advances, https://doi.org/10.1126/sciadv.aar8195.

Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. Journal of Hydrology, https://doi.org/10.1016/S0022-1694(96)03329-X.

Refsgaard, J.C., 2001. Discussion of model validation in relation to the regional and global scale. Model Validation: Perspectives in Hydrological Science.

Rode, M., Thiel, E., Franko, U., Wenk, G., Hesser, F., 2009. Impact of selected agricultural management options on the reduction of nitrogen loads in three representative meso scale catchments in Central Germany. The Science of the Total Environment, https://doi.org/10.1016/j.scitotenv.2009.01.053.

Rode, M., Halbedel Nee Angelstein, S., Anis, M.R., Borchardt, D., Weitere, M., 2016. Continuous in-stream assimilatory nitrate uptake from high-frequency sensor measurements. Environmental Science and Technology, https://doi.org/10.1021/acs.est.6b00943.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, https://doi.org/10.1029/2009WR008933.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. Hydrological Sciences Journal, https://doi.org/10.1080/02626667.2011.637497.

Ullrich, A., Volk, M., 2010. Influence of different nitrate-N monitoring strategies on load estimation as a base for model calibration and evaluation. Environmental Monitoring and Assessment, https://doi.org/10.1007/s10661-009-1296-8.

Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2005.09.008.

Veinbergs, A., Lagzdins, A., Jansons, V., Abramenko, K., Sudars, R., 2017. Discharge and nitrogen transfer modeling in the Berze River: A HYPE setup and calibration. Environmental and Climate Technologies, https://doi.org/10.1515/rtuect-2017-0005.

Vrugt, J.A., Robinson, B.A., Hyman, J.M., 2009. Self-adaptive multimethod search for global optimization in real-parameter spaces. IEEE Transactions on Evolutionary Computation, https://doi.org/10.1109/TEVC.2008.924428.

Wade, A.J., Whitehead, P.G., Butterfield, D., 2002. The integrated catchments model of phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river systems: model Structure and equations, https://hal.archives-ouvertes.fr/hal-00304711.

Wellen, C., Kamran-Disfani, A.R., Arhonditsis, G.B., 2015. Evaluation of the current state of distributed watershed nutrient water quality modeling. Environmental Science and Technology, https://doi.org/10.1021/es5049557.

Withers, P.J.A., Neal, C., Jarvie, H. P., Doody D. G. 2014. Agriculture and Eutrophication: Where Do We Go from Here?, Sustainability, 6(9), 5853-5875. https://doi.org/10.3390/su6095853.

Yang, J., Reichert, P., Abbaspour, K.C., 2007. Bayesian uncertainty analysis in distributed hydrologic modeling: A case study in the Thur River basin (Switzerland). Water Resources Research, https://doi.org/10.1029/2006WR005497.

Yin, Y., Jiang, S., Pers, C., Yang, X., Liu, Q., Yuan, J., Yao, M., He, Y., Luo, X., Zheng, Z., 2016. Assessment of the Spatial and Temporal Variations of Water Quality for Agricultural Lands with Crop Rotation in China by Using a HYPE Model.

International Journal of Environmental Research and Public Health, https://doi.org/10.3390/ijerph13030336.

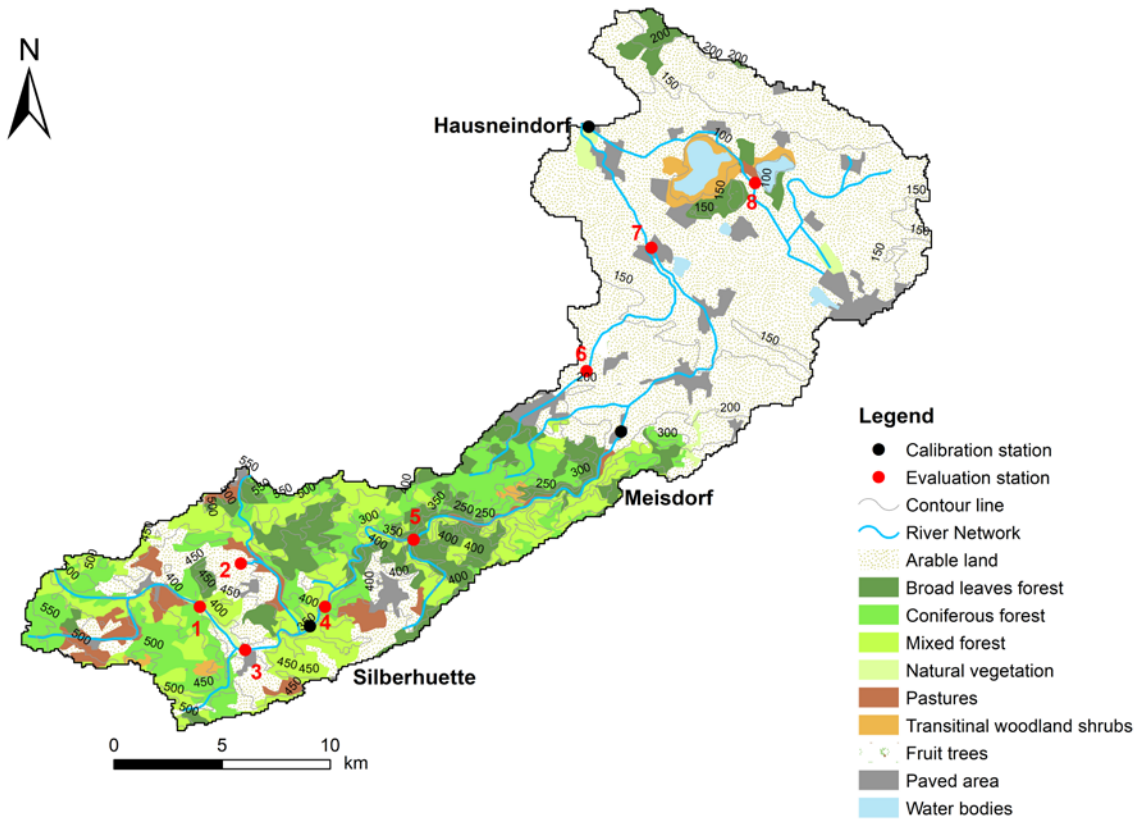

Figure 7. The Selke catchment and its dominant land-use classes. The black dots indicate the three main gauging stations (Silberhuette, Meisdorf and Hausneindorf) used for the model calibration. The red dots correspond to the location of the eight internal stations used for the spatial validation of the model. The grey lines shows the contour elevation.

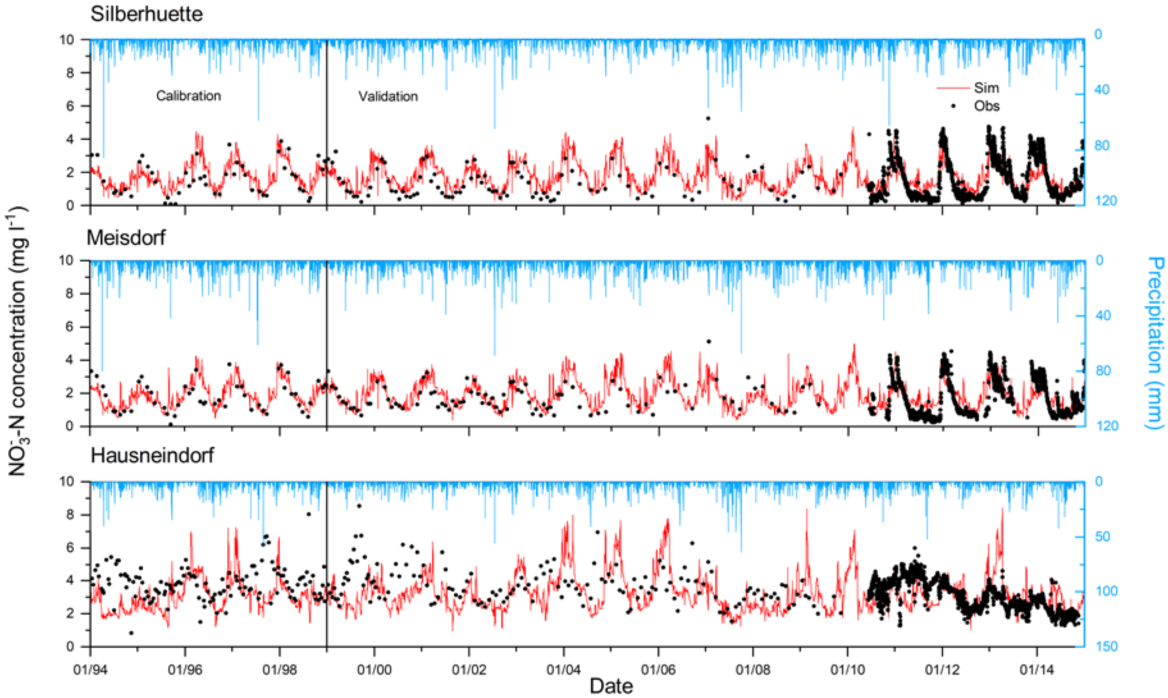

Figure 8. Simulated and observed nitrate-N (NO_3^--N) concentrations at three main stations (Silberhuette, Meisdorf and Hausneindorf) during calibration (1994-1998) and validation (1999-2014). The black dots and red lines represent the observed and simulated NO_3^--N concentrations, respectively.

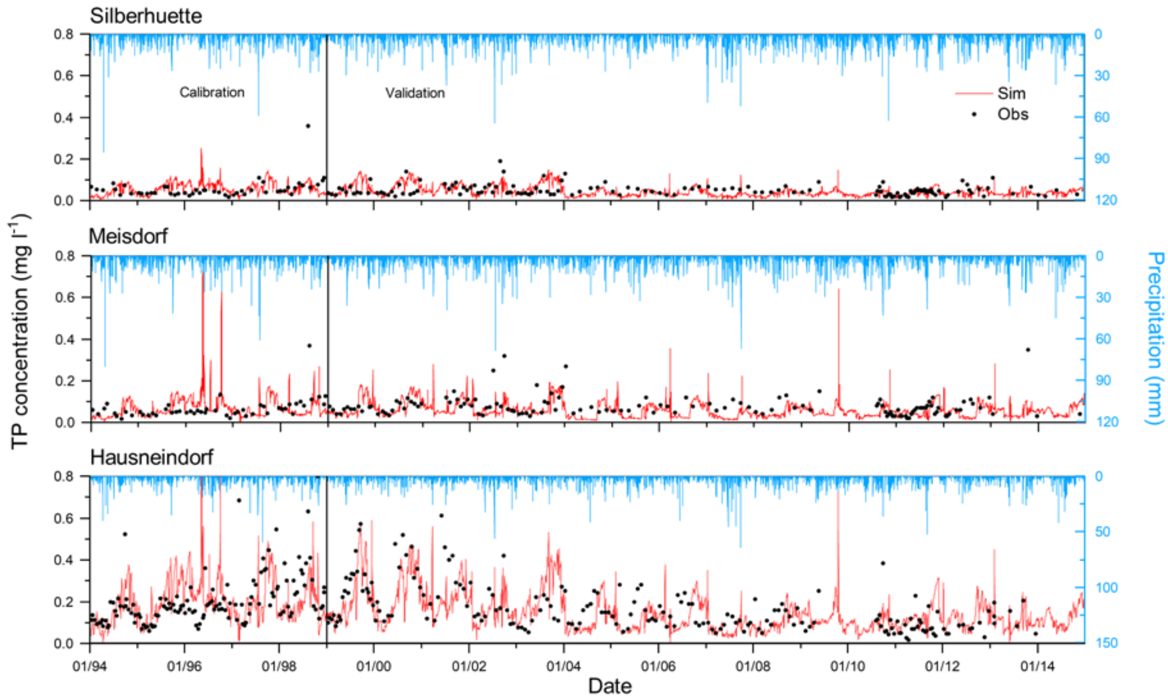

Figure 3. Simulated and observed total phosphorus (TP) concentrations at three main stations (Silberhuette, Meisdorf and Hausneindorf) during the calibration (1994-1998) and validation (1999-2014) periods. The black dots and red lines represent the observed and simulated TP concentrations, respectively.

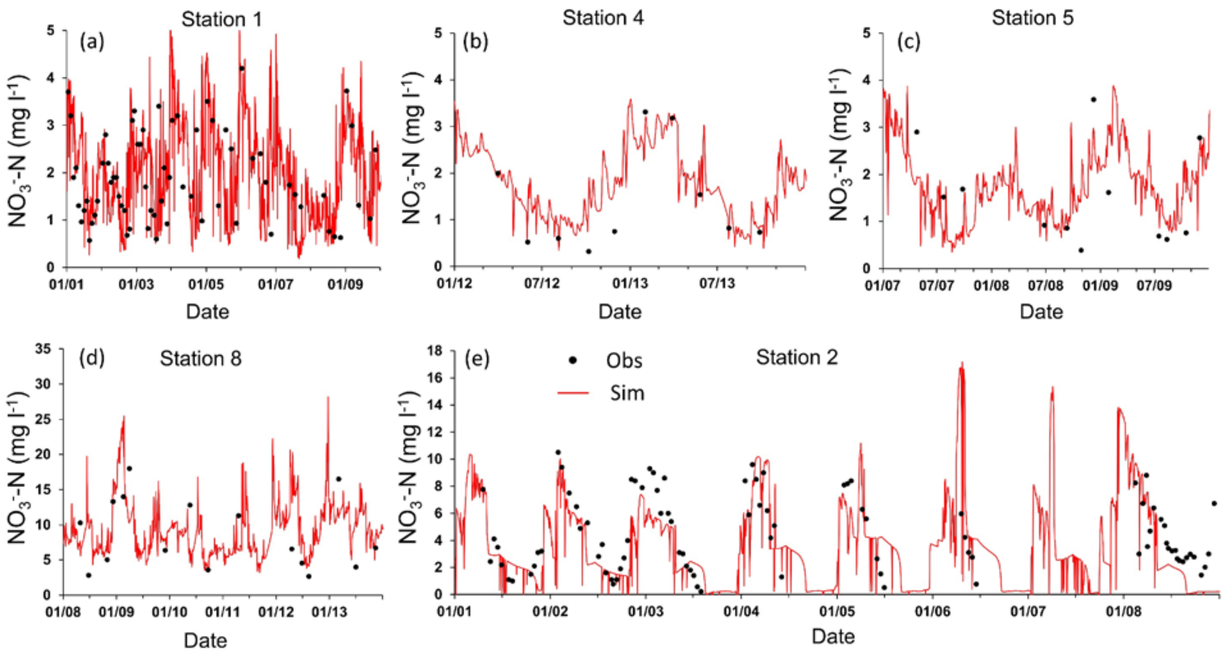

Figure 9. Simulated and observed nitrate-N (NO_3^--N) concentrations at internal stations. The black dots and red lines represent the observed and simulated NO_3^--N concentrations, respectively.

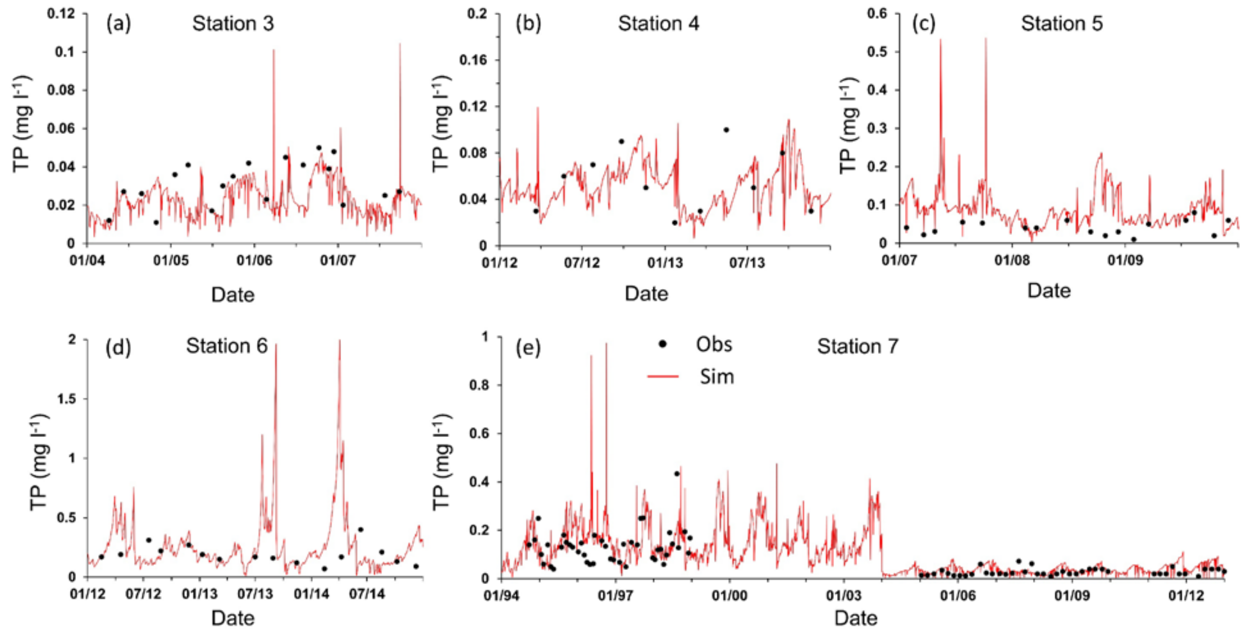

Figure 10. Simulated and observed total phosphorus (TP) concentrations at internal stations. The black dots and red lines represent the observed and simulated TP concentrations, respectively.

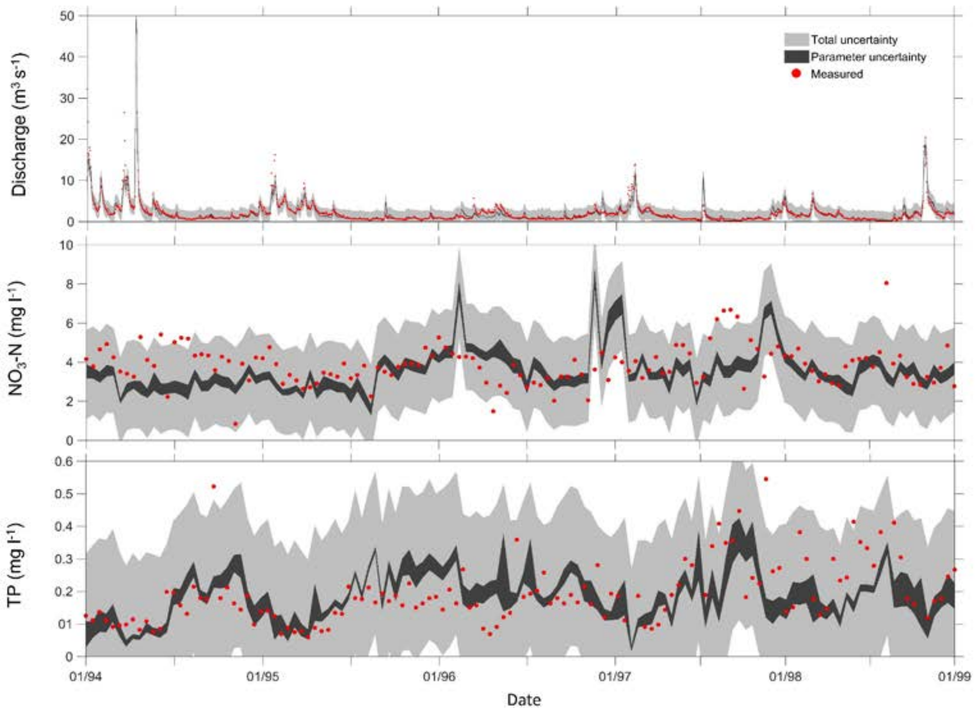

Figure 11. Estimated 95% prediction confidence intervals of daily mean discharge, nitrate-N (NO_3^--N) and total phosphorus (TP) at catchment outlet (gauging station Hausneindorf) during calibration period (1994–1998).

Table 3. Characteristics of three main catchments: 1) Hausneindorf, 2) Meisdorf and 3) Silberhuette with mean specific discharge and mean NO_3^--N and TP concentrations.

Data type	Hausneindorf	Meisdorf	Silberhuette
Mean elevation (m)	104-469	212-469	409-469
Area (km^2)	463	176.3	98.7
Soil type	Cambisols in the upper area and Chernozems in the lowland area	Cambisols	Cambisols
Forest Share (%)	35.4	71.9	60.4
Arable land share (%)	52.3	16.9	25.3
Mean annual precipitation (mm y^{-1})	Mountain areas: 625 Lowland areas: 450	640	653
Mean discharge (l s^{-1} km^{-2})	3.99	8.40	13.15
Mean NO_3^--N concentration (mg l^{-1})	3.91	1.75	1.44
Mean TP concentration (mg l^{-1})	0.18	0.07	0.05

Table 4. Description of spatial and time series input data for the HYPE model setup in the Selke catchment.

Data type	Data description/properties	Resolution	Source
Geographical data	Elevation	90 m	State Survey Office
	Stream network	-	State Survey Office
	Soil type	50 m	State Survey Office
	Land use	25 m	Corrine Land Cover 2006
Meteorological data	Daily precipitation and mean air temperature	16 rainfall and 2 climate stations	German Weather Service-DWD
Agricultural practices	Manure and inorganic fertiliser application, timing and amount for fertilisation, sowing and harvesting	-	Field survey and literature
Soil nitrogen content	Initial nitrogen storage	-	Literature review
Sewage treatment plants	Water flow and NO_3^--N concentration TP concentration	Constant daily loadings from 6 sewage treatment plants	Operating reports of sewage treatment plants

Table 3. Physical meanings, initial values and ranges and optimized values of discharge (Q), nitrate-N (NO_3^--N) and total phosphorus (TP) parameters.

	Physical meaning	Initial value	Initial range	Optimized value
	Discharge parameters (Q)			
cevp	Potential evapotranspiration rate (mm d^{-1} °C^{-1})			
Agriculture land		0.234	0.01-1	0.308
Coniferous forest		0.170	0.01-1	0.172
Mixed forest		0.116	0.01-1	0.200
rrcs1 brown soil	Soil runoff coefficient for the uppermost soil layer (d^{-1})	0.104	0.001-1	0.350
rivvel	Maximum velocity in the stream channel (m s^{-1})	0.202	0.001-1	0.264
rcgrw	Runoff coefficient for regional groundwater flow (d^{-1})	0.004	0.0001-0.1	0.005
epotdist	Decrease of evapotranspiration with soil depth (m^{-1})	6.574	1-10	9.530
wcfc Brown soil	Fraction of soil layer where water is available for evapotranspiration but not for runoff (-)	0.700	0.01-0.7	0.690
wcep Brown soil	Fraction of soil layer where water is available for runoff (-)	0.120	0.001-0.2	0.191
pcadd	Correction parameter for precipitation (-)	0.094	0.001-1	0.468
	Nitrate-N parameters (NO_3^--N)			
denitr	Denitrification rate in soil (d^{-1})	0.022	0.001-0.1	0.050
uptsoil1				
Agriculture land		1.000	0.001-1.0	0.990
Coniferous forest	Fraction of nutrient uptake in the uppermost soil	1.000	0.001-1.0	0.587
Mixed forest	layer (-)	0.940	0.001-1.0	0.002
fertdays	Number of days that fertilizer applications occur counting from application day 1 and forward using the same amount every day	60	10-150	62
denitw	Parameter for the denitrification in water (kg m^2 d^{-1})	2×10^{-6}	1×10^{-6}-0.1	7×10^{-6}
wprod	Production/decay of N in water (kg m^{-3} d^{-1})	0.055	0.0001-0.1	0.001
rivvel2	Parameter for calculating the velocity of water in the stream channel	1.000	0.001-1.0	0.521
	Total Phosphorous parameters (TP)			
Sedexp	Parameter for sedimentation	2.441	0.1-10	3.672
Pnratio				
Agriculture land		0.112	0.001-1	0.240
Coniferous forest	N and P relationship for nutrient uptake	0.181	0.001-1	0.792
Mixed forest		0.134	0.001-1	0.223
freund3	Desorption speed (l d^{-1})			
Brown soil		0.366	0.001-1	0.852
Sandy soil		0.252	0.001-1	0.521
Black soil		0.141	0.001-1	0.561
Hphalf				
Agriculture land		0.125	0.001-1	0.334
Coniferous forest	Halving depth for humus P pool (m)	0.313	0.001-1	0.652
Mixed forest		0.051	0.001-1	0.401

Table 4. Model validation results using mean of PBIAS (%) values and PBIAS ranges of NO_3^--N and TP concentration of calibration Schemes: (1) Calibration at catchment outlet (Hausneindorf) and (2) Calibration at three main gauge stations (Silberhuette, Meisdorf and Hausneindorf).

Calibration Scheme	Average PBIAS (%) at three main stations (PBIAS range)		Average PBIAS (%) at internal stations (PBIAS range)	
	NO_3^--N	TP	NO_3^--N	TP
Calibration at Hausneindorf	20.2 (12.0 to 33.6)	-3.0 (-19.5 to 6.7)	5.1 (-34.0 to 16.8)	20.9 (-38.3 to 81.8)
Calibration at three stations	5.1 (-2.0 to 11.0)	-7.6 (-25.0 to -11.0)	1.7 (-9.0 to 14.2)	4.0 (-25.3 to 34.3)

Table 5. Model evaluation of discharge (Q), nitrate-N (NO_3^--N) and total phosphorous (TP) simulations at the stations Silberhuette, Meisdorf and Hausneindorf for calibration and validation period using calibration Scheme 2.

Variable	Station	Calibration (1994-1998)		Validation (1999-2014)	
		NSE	PBIAS (%)	NSE	PBIAS (%)
Q	Silberhuette	0.87	-4.8	0.76	11.9
	Meisdorf	0.85	0.5	0.73	3.0
	Hausneindorf	0.84	2.1	0.71	18.0
NO_3^--N Load	Silberhuette	0.93	-2.1	0.72	2.4
	Meisdorf	0.90	6.4	0.77	-16.1
	Hausneindorf	0.74	-5.7	0.70	-2.5
TP Load	Silberhuette	0.48	-20.0	0.52	-10.0
	Meisdorf	0.53	11.5	0.46	-20.0
	Hausneindorf	0.13	-19.1	0.20	6.5

Table 6. Model evaluation of nitrate-N (NO_3^--N) and total phosphorous (TP) concentration simulations at internal stations. Sim and Obs are refering to Simulated and Observed concentrations.

	NO_3^--N (mgN l^{-1})			TP (mgP l^{-1})		
Station	PBIAS (%)	Mean (Sim)	Mean (Obs)	PBIAS (%)	Mean (Sim)	Mean (Obs)
1	2.5	2.00	1.90	-	-	-
2	-9.0	4.35	4.52	-	-	-
3	-	-	-	-25.3	0.023	0.031
4	3.9	1.67	1.60	-22.1	0.036	0.051
5	-3.0	1.90	1.92	20.1	0.049	0.042
6	-	-	-	34.3	0.300	0.200
7	-	-	-	13.2	0.092	0.081
8	14.2	9.47	8.46	-	-	-

Table 7. ARIL, PCI and PUCI for discharge, nitrate-N (NO_3^--N) and total phosphorous (TP) of 95% prediction confidence interval for calibration period at Hausneindorf.

Variable	Criterion	Parameter Uncertainty	Total Uncertainty
Discharge	ARIL	0.093	4.139
	PCI	0.136	0.983
	PUCI	2.003	0.234
NO_3^--N Concentrations	ARIL	0.150	1.242
	PCI	0.270	0.920
	PUCI	2.120	0.783
TP Concentrations	ARIL	0.310	3.112
	PCI	0.330	0.961
	PUCI	1.228	0.317