

Water Resources Research

RESEARCH ARTICLE

10.1029/2019WR025605

Key Points:

- A novel sensitivity-weighted performance criterion to improve parameter identifiability was developed
- The information gain as obtained from temporally resolved parameter sensitivity analysis increased
- Identifiability was improved for parameters whose corresponding processes are dominant for a short period or with seasonal variations

Correspondence to:

B. Guse,
bguse@hydrology.uni-kiel.de

Citation:

Guse, B., Pfannerstill, M., Fohrer, N., & Gupta, H. (2020). Improving information extraction from simulated discharge using sensitivity-weighted performance criteria. *Water Resources Research*, 56, e2019WR025605. <https://doi.org/10.1029/2019WR025605>

Received 25 JUN 2019

Accepted 9 JUL 2020

Accepted article online 13 JUL 2020

Improving Information Extraction From Simulated Discharge Using Sensitivity-Weighted Performance Criteria

B. Guse^{1,2} , M. Pfannerstill¹ , N. Fohrer¹ , and H. Gupta³ 

¹Department of Hydrology and Water Resources Management, Institute of Natural Resource Conservation, Christian-Albrechts-University of Kiel, Kiel, Germany, ²Section Hydrology, GFZ German Research Centre for Geosciences, Potsdam, Germany, ³Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA

Abstract Due to seasonal or interannual variability, the relevance of hydrological processes and of the associated model parameters can vary significantly throughout the simulation period. To achieve accurately identified model parameters, temporal variations in parameter dominance should be taken into account. This is not achieved if performance criteria are applied to the entire model output time series. Even when using complementary performance criteria, it is often only possible to identify some of the model parameters precisely. We present an innovative approach to improve parameter identifiability that exploits the information available regarding temporal variations in parameter dominance. Using daily parameter sensitivity time series, we construct a set of sensitivity-weighted performance criteria, one for each parameter, whereby periods of higher dominance of a model parameter and its corresponding process are assigned higher weights in the calculation of the associated performance criterion. These criteria are used to impose constraints on parameter values. We demonstrate this approach by constraining 12 model parameters for three catchments and examine ensemble hydrological simulations generated using these constrained parameter sets. The sensitivity-weighted approach improves in particular the identifiability for parameters whose corresponding processes are dominant only for short periods of time or have strong seasonal patterns. This results overall in slight improvement of model performance for a set of 10 contrasting performance criteria. We conclude that the sensitivity-weighted approach improves the extraction of hydrologically relevant information from data, thereby resulting in improved parameter identifiability and better representation of model parameters.

1. Introduction

The relative dominance of various hydrological processes varies in space between different catchments and in time between different seasons and phases of the hydrograph. To represent the hydrological system accurately, such spatiotemporal process variability should be taken into account when developing and using hydrological models. While spatial variability is sometimes assessed when using models in contrasting catchments, the temporal variability in dominance of hydrological processes is often not explicitly considered.

Hydrological models typically contain several model parameters that control the representation of hydrological processes and their temporal dynamics. To obtain reliable hydrological predictions for any given study catchment, appropriate values for these parameters must be specified so that the model is well adapted to both spatial and temporal process variability.

Precise estimation of model parameters is achieved if behavioral model runs are obtained using narrow uncertainty ranges for their estimated values. In general, “high parameter identifiability” can be defined as the ability to obtain parameter estimates with low uncertainty (within a relative narrow range of suitable values), while “low parameter identifiability” is associated with high parameter uncertainty (Abebe et al., 2010; Kelleher et al., 2013; Shin et al., 2015; Wagener et al., 2003).

However, it is also necessary for these narrow parameter uncertainty ranges to correspond to realistic (i.e., behavioral) representation of the spatiotemporal dynamics of the corresponding process. Accordingly,

a relevant model parameter might incorrectly be deemed nonidentifiable if the nature of its corresponding process is not adequately considered in the design of the parameter identification method. Consequently, inappropriate settings for the values of these model parameters may result in the corresponding processes not being properly represented and with the model response in these phases being incorrectly controlled by a different process (Fenicia et al., 2007, 2008; Gharari et al., 2013; Wagener et al., 2003).

To guide the identification of parameter values, performance criteria are typically used that compare simulated and observed discharge time series. In general, each such criterion tends to emphasize performance with respect to specific parts of the hydrograph; for example, the Nash-Sutcliffe Efficiency (NSE, Nash & Sutcliffe, 1970) tends to emphasize the fitting of larger flow values. Accordingly, the selection of any performance criterion implicitly leads to a focus on one particular aspect of the hydrograph (Krause et al., 2005; Reusser et al., 2009).

To address this issue, it has been recommended that a set of contrasting performance criteria be used so as to accurately represent all phases of the hydrograph via a multiobjective procedure (Gupta et al., 1998). For example to obtain balanced performance on both high and low flows simultaneously in the same model run, a multiobjective optimization that seeks to minimize both the NSE and its logarithmic version (NSElog) can be conducted. More granularity can be achieved with regard to the hydrograph by further distinguishing between the rising and falling limbs (Boyle et al., 2000). In addition, to optimize the soil water or evapotranspiration parameters, performance criteria that assess water balance (e.g., the bias component of KGE [KGE_beta] or Percent Bias [PBIAS]) are often used (Guse et al., 2017; van Werkhoven et al., 2008).

It can be challenging to simultaneously achieve both a high degree of parameter identifiability and high model performance (Wagener et al., 2001), and improvements in parameter identification therefore depend on improving the nature of the set of criteria used for model performance evaluation. When attempting to do so, it can be challenging to achieve an appropriate overall representation of all of the phases of model response; for example, when examining hydrographs it can be difficult to achieve good performance with regard to both high and low flows simultaneously in the same model run (Pechlivanidis et al., 2012; Pfannerstill et al., 2014b).

One way to achieve a balanced representation of various hydrological processes, and to consequently identify their parameter values precisely, is to select each performance criterion to be individually strongly related to a specific process (a so called “signature-based” performance criterion). This results in a signature-based multicriteria method and has been proposed as a way to achieve better diagnostic assessment of the process representations in a model (Gupta et al., 2008). Examples of this approach include the use of separate performance criteria to assess performance with respect to the different segments of the flow duration curve (FDC) (Euser et al., 2013; Pfannerstill et al., 2014b; Shafii & Tolson, 2015; Yilmaz et al., 2008). However, it may not be readily apparent how to construct appropriate signature-based measures for all of the hydrological processes of interest. Given that no “standard” set of signatures has been established, there is an ongoing debate on what would constitute a suitable set of such signatures for any given application.

It is common for model performance criteria to be computed as integrated values over the entire discharge time series. Given a set of performance criteria, model evaluation can be seen as a backward approach in which changes in performance criteria are related back to changes in the model parameters to guide a search for their optimal values. In doing so, unless somehow reweighted (i.e., by modifying the design of the specific performance criterion), each time step is considered to have equal importance when computing the value of the model performance criterion. One simple way to achieve an appropriate reweighting is to adjust the mathematical form of the performance criterion so that different discharge magnitudes are emphasized differently (Madsen et al., 2002; van Werkhoven et al., 2008; Sorooshian et al., 1983).

A problem with computing a single performance criteria over the entire simulation time period is that reliable parameter identification is often only achieved for the subset of parameters that are associated with the processes that are dominant during that period. In other words, the focus is on the processes and parameters that have the largest time-period-average impact on the simulated discharge for the prevailing catchments (Wagener et al., 2003). Even when using signature-type measures (Gupta et al., 2008), all of the relevant model parameters may not be adequately well identified (Guse et al., 2017).

A major reason for this limitation is that conventional approaches neglect the fact that the relevance of hydrological processes (and their associated parameters) changes dynamically over time, often following

seasonal patterns (such as for evapotranspiration or snow processes) or due to their dependence on event occurrence (such as for surface runoff) (Boyle et al., 2000; Tang et al., 2007; Wagener et al., 2003). Accordingly, model parameters tend to be better identifiable during periods of associated high process dominance (Wagener et al., 2003). The parameter dominance patterns will also vary between different catchments (Cibin et al., 2010; Guse, Pfannerstill, Strauch, et al., 2016).

When constant weighting of time steps is used in computing the values of the performance criteria, the available information regarding temporal variations in the dominance of processes and their parameters is not explicitly considered. Consequently, the importance of certain parameters can be obscured due to the overriding impact of a subset of “major” parameters that tend to play a more dominant role throughout the year (Herman et al., 2013; Wagener et al., 2003). The result can be “apparent poor identifiability” associated with parameters whose impacts are relevant during only a short time period or specific phase of the year (Guse, Pfannerstill, Gafurov, et al., 2016), even if they are highly dominant during those short time periods. Thus, low parameter identifiability as assessed by a certain performance criterion does not necessarily mean that a certain model parameter is irrelevant. A perception of low parameter identifiability can also result from the use of an inappropriate performance criterion computed over an inadequate time period (Wagener et al., 2002, 2003). An approach that accounts for the temporally varying nature of process dominance will lead to more consistency in the identification of parameter values than a model optimization without considering the temporal variability in parameter dominance (Clark et al., 2008; Fenicia et al., 2008; Yilmaz et al., 2008).

To resolve this issue, a possibility (considered herein) is to design the parameter identification process so that it accounts for the fact that process and parameter dominance varies temporally. For example, one way would be to partition the entire time series into subperiods, by subdivision into different clusters (de Vos et al., 2010), into surface runoff and baseflow segments (Zhang et al., 2011), or based on dominant processes (He et al., 2015). However, these approaches require a discrete (noncontinuous) partitioning of the time series.

In regard to continuous variability in the time series, Wagener et al. (2003) have clearly shown, using dynamic identifiability analysis (DYNIA), that model parameters exhibit overlapping phases of higher identifiability due to temporal variation in the dominant hydrological processes. As pointed out by Abebe et al. (2010), information gain can be increased by focusing on the part of the data that is most relevant to the question at hand, concluding that model parameters are better identified when focusing on their most relevant parts of the entire time series.

A number of ways to detect temporal changes in parameter relevance have been proposed, including temporally resolved parameter sensitivity analysis (Gupta & Razavi, 2018; Herman et al., 2013; Massmann et al., 2014; Reusser et al., 2011). In particular, the temporal dynamics of parameter sensitivity approach (TEDPAS) can be used to efficiently estimate the partial sensitivities for each parameter for each time step (Guse et al., 2014; Reusser et al., 2011).

The aim of this study is to show that this information of daily parameter sensitivities can be incorporated as weights into the design and computation of performance criteria, thereby reflecting the temporal variations in parameter dominance in the parameter identification process. Therefore, a process-based parameter identification methodology is developed and tested that makes use of information regarding temporal variations in parameter sensitivity. This accounts for the fact that the importance of some parameters may last for only short periods of time or may vary seasonally. By doing so, we seek to achieve a higher degree of identifiability for a larger set of model parameters, because when more model parameters are precisely identified, we can expect the representation of the hydrological system to be better. We then use the constrained parameter ranges thus obtained to evaluate the performance of a corresponding ensemble of model simulations, using a set of 10 performance criteria. Thus, we test whether the resulting overall model performance associated with these selected parameter ranges is improved.

2. Material and Methods

2.1. Hydrological Model and Set-Up in the Three Catchments

In this study, we used a set of three catchments (Table 1). For a more detailed description of the catchments and their dominant hydrological processes, please see Guse et al. (2019).

Table 1
Characteristics of the Three Catchments

	Treene	Saale	Kinzig
Region	Lowland	Midrange mountains	Mid-range mountains
Catchment size (km ²)	481	1,019	925
Hydrological station	Treia	Blankenstein	Hanau
Dominant processes	Groundwater and tile flow	Surface and subsurface processes	Surface and subsurface processes
Minimum elevation (m asl)	2	415	98
Maximum elevation (m asl)	80	856	628
Mean annual precipitation (mm/a)	995	922	899

Note. The model set-up is described for Treene and Saale by Guse, Pfannerstill, Strauch, et al. (2016) and for Kinzig by Kakouei et al. (2018).

The semidistributed eco-hydrological Soil and Water Assessment Tool (SWAT) model (Arnold et al., 1998) was used to generate spatially explicit hydrological simulations for each subbasin. The model simulates all of the major hydrological processes and generates estimates of several different runoff components (surface runoff, subsurface flow, tile flow, and groundwater flow). We used model version SWAT3s that includes two active aquifers (Pfannerstill et al., 2014a).

For our analysis, we focused on the 12 model parameters listed in Table 2 and used the same parameter ranges for all three catchments. Some of these model parameters can be characterized in terms of typical periods of parameter dominance in the case that the associated process is of high relevance in the catchment. Based on former sensitivity studies with the SWAT model (Guse et al., 2019; Guse, Pfannerstill, Gafurov, et al., 2016; Guse, Pfannerstill, Strauch, et al., 2016), specific periods of parameter dominance can be derived regarding specific phases of the year and of the hydrograph as shown in Table 2. Snow parameters are dominant in winter up to the end of snow melting period. Dominant phases of evapotranspiration parameters are related to warmer seasons and resaturation phases. Fast occurring processes are characterized by high fluctuations in their relevance, and the associated parameters are dominant for a short-period after strong, intensive precipitation events. In contrast, groundwater parameters are dominant for a longer period due to its longer retention time in dry phases.

Table 2
Description of the SWAT Model Parameters

Parameter abbreviation	Description	Range type	Lower range	Upper range	Dominant period	Catchment
SFTMP	Snow fall temperature	r	-2.5	2.5	Cold season	K, S, T
SMTMP	Snow melt temperature	r	-2.5	2.5	Cold season	S, K, T
SNOCOVMX	Minimum snow cover for 100% coverage	r	1	50	Cold season	S, T
CN2	Curve number	a	-15	5	High precipitation	S, K, T
SURLAG	Surface runoff lag time	r	0.8	4	High precipitation	S, K
LATTIME	Lateral flow lag time	r	0.2	8	No preference	K, S
GDRAIN	Tile flow lag time	m	0.5	1.5	No preference	T
SOL_AWC	Available water capacity of a soil layer	a	-0.02	0.1	No preference	S, T, K
ESCO	Soil evaporation compensation factor	r	0.2	1	Warm and dry	K, T, S
GW_DELAYfsh	Delay time to flow into fast aquifer	r	1	50	Low-medium flow	T, K, S
RCHRGssh	Fraction coefficient between fast and slow aquifers	r	0.2	0.8	Low-medium flow	T, K, S
ALPHA_BFssh	Baseflow alpha factor for the timing of slow aquifer	r	0.001	0.2	Dry	T, K, S

Note. Their lower and upper ranges are added as absolute range (r), additive (a), or multiplicative (m) values. Dominant periods of the model parameters are presented regarding phases of the year and of the hydrograph based on former sensitivity studies (Guse et al., 2019; Guse, Pfannerstill, Gafurov, et al., 2016; Guse, Pfannerstill, Strauch, et al., 2016). The parameter ranges are taken from Guse et al. (2019). The catchments (T = Treene, S = Saale, and K = Kinzig) are ordered by decreasing sensitivity. No preference means that no period of clear parameter dominance can be derived. More details of the model parameters can be found in the theoretical documentation of the SWAT model (Neitsch et al., 2011).

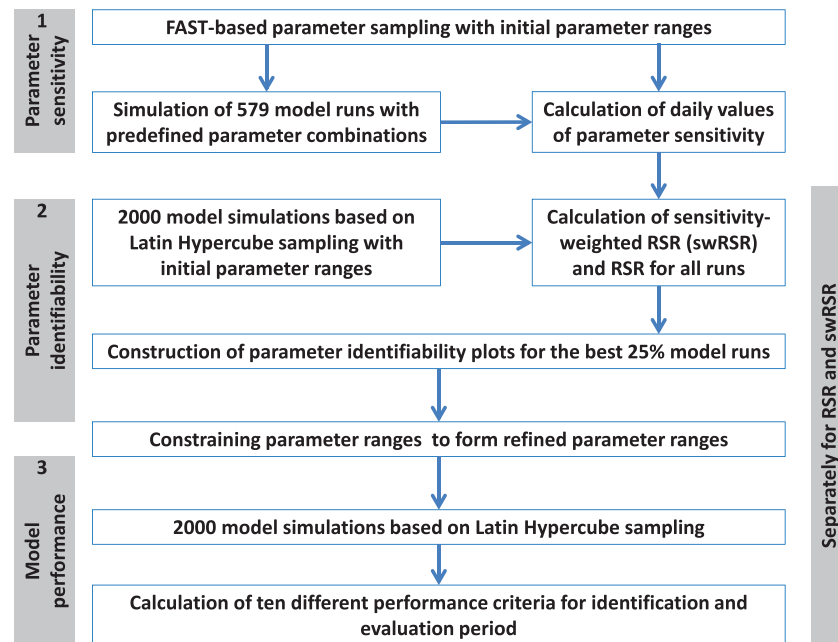


Figure 1. Framework of methodological approach (shown for RSR and swRSR).

Simulations were run for the period 1997–2010, with the first 3 years used for warm-up, 6 years (2000–2005) for parameter identification, and 5 years (2006–2010) for model performance evaluation.

2.2. Methodological Design

An overview of our three-stage methodological approach is presented in Figure 1 and described in detail in the following subsections. We propose to extract information regarding temporal variations in the dominance of parameters provided by the TEDPAS approach and to incorporate this information into the design of the performance criteria to be used for parameter identification. In the first stage, we generate 579 model runs for a set of parameter values selected across the initial ranges of 12 model parameters and use a temporally resolved parameter sensitivity analysis to calculate daily sensitivity time series for each of the 12 model parameters. Next, we generated 2,000 parameter sets across these same parameter ranges via Latin Hypercube sampling (LHS) (Pfannerstill et al., 2014b; Soetaert & Petzoldt, 2010). The resulting 2,000 simulated discharge time series were used to calculate both a classical performance criterion and a set of sensitivity-weighted versions of it (that use the daily parameter sensitivity values as weights) for each of the model runs (parameter sets) mentioned above. These results were then used to construct two sets of parameter identifiability plots for the “best” model runs, one set for each of the two approaches. This step imposes constraints on the parameters and results in different parameter uncertainty ranges. In stage three, we use these two sets of parameter ranges to generate two different sets of 2,000 model runs (parameter sets) again using LHS. The resulting simulated discharge time series were evaluated using a set of 10 performance criteria to compare the model performance between the two approaches.

2.3. Temporal Dynamics of Parameter Sensitivity

To conduct the temporally resolved parameter sensitivity analysis, we used the TEDPAS approach as presented in Reusser et al. (2011), to efficiently obtain daily time series of sensitivity estimates for each model parameter. From this, the dominant model parameters on each day were identified. This type of sensitivity analysis, with a focus on identifying dominant model parameters, is categorized as factor prioritization by Saltelli et al. (2006). For this, a global sensitivity analysis was used to explore the entire parameter space using the Fourier Amplitude Sensitivity Test (FAST; Cukier et al., 1973) implemented in R (Reusser, 2013). During FAST parameter sampling, all model parameters are changed simultaneously. Importantly, each parameter is changed within a different frequency. In FAST, to calculate the parameter sensitivities, the time series of modeled discharge are transformed into a Fourier series. The FAST coefficient is then used to estimate the partial variance of a given parameter. Thus as the measure of sensitivity, we selected the first-order partial variance, which is the ratio between the first-order partial variance and the total variance (Guse et al., 2014;

Reusser et al., 2011). The partial variance is computed for each day to provide daily sensitivities for each model parameter. In the FAST approach, the number of required model runs depends on the number of parameters; for 12 model parameters, 579 model simulations are required for the estimation of the parameter sensitivity and to guarantee independence of the Fourier frequencies (Reusser et al., 2011).

Note that TEDPAS uses the modeled discharge directly as the target variable for assessing parameter sensitivity analysis, rather than a performance criteria (Gupta & Razavi, 2018; Herman et al., 2013; van Werkhoven et al., 2008). Using TEDPAS, typical temporal patterns of parameter sensitivities can be identified. These parameter dominance patterns can be explained in relation to seasonal variations in the target variable of the sensitivity analysis (Guse, Pfannerstill, Gafurov, et al., 2016; Guse, Pfannerstill, Strauch, et al., 2016).

To investigate the variability in parameter sensitivity between study catchments, the cumulative distribution functions of the daily parameter sensitivity values were computed for each parameter and compared among the catchments.

2.4. Parameter Sampling

Next, the 12 parameters were randomly varied across their parameter ranges (same as used in the sensitivity analysis) to generate 2,000 parameter sets via Latin Hypercube sampling (LHS). Based on this, 2,000 model runs were carried out in the three catchments with the same parameter sets.

In this regard, one can generally distinguish between probabilistic approaches (in which the use of prior and posterior parameter distributions is appropriate and relevant) and behavioral approaches (as conceptually introduced by Hornberger & Spear, 1981; of which the GLUE methodology of Beven & Binley, 1992, can sometimes be an example). To be clear, the approach discussed in our manuscript is a behavioral approach.

2.5. Identifiability Metrics

These model runs were then used to constrain the parameter ranges. As a benchmark against which to compare our sensitivity-weighted approach, we also used two performance criteria with equal weighting in time.

For use in constraining the parameter ranges, two versions of the RSR metric were used (see Moriasi et al., 2007). RSR is the ratio of the root mean square error (RMSE) to the standard deviation of the observed discharge time series. In the first version, RSR is equally weighted (Equation 1) in each time step:

$$RSR = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{o,i} - Q_{s,i})^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)^2}} \quad (1)$$

Hereby, Q_o is the observed and Q_s the simulated discharge as well \bar{Q}_o the mean of observed discharge. N is the length of time series (number of days).

The second version is a time-weighted performance criterion, using the daily parameter sensitivity values to calculate a weighting factor w . This results in separate swRSR metrics for each model parameter (Equation 2). Hereby, w varies between each day and between each parameter.

$$swRSR = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N w * (Q_{o,i} - Q_{s,i})^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N w * (Q_{o,i} - \bar{Q}_o)^2}} \quad (2)$$

To calculate w , the sensitivities values for a given parameters are weighted along the entire time series (Equation 3) and multiplied by N . Thus, the sum of all weights is identical with the length of the time series. Consequently, if all of the weights are set to one, swRSR becomes equal to RSR.

$$w_i = \frac{S_i}{\sum S_i} * N \quad (3)$$

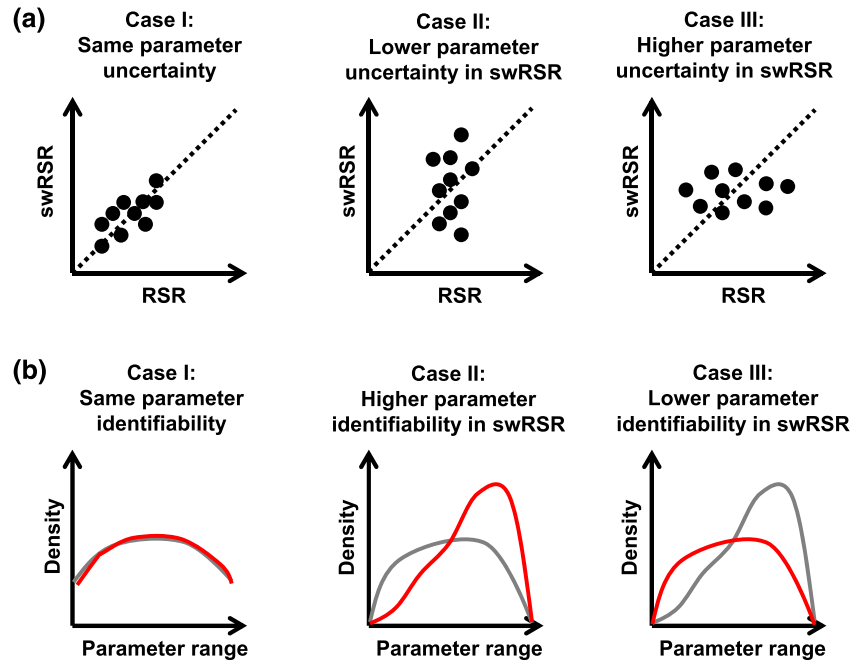


Figure 2. Three cases illustrating interpretation of (a) the correlation patterns between the values of RSR and swRSR and (b) the parameter identifiability plots for RSR (gray) and swRSR (red).

Similarly, this approach was repeated for a second performance criterion which does not use a squaring of the residuals. For this, we used the mean absolute error (MAE, Equation 4) and its sensitivity-weighted version (Equation 5).

$$MAE = \frac{\sum_{i=1}^N [Q_{o,i} - Q_{s,i}]}{N} \quad (4)$$

$$swMAE = \frac{\sum_{i=1}^N w * [Q_{o,i} - Q_{s,i}]}{N} \quad (5)$$

On the one side, as it is shown in Equations 2 and 5, the model error becomes high if both the weights, that is, the parameter sensitivity of that day and the deviation between observed and simulated discharge, are high. This would mean that the error is high in times of high parameter sensitivity. As a consequence, model runs with a coincidence of high parameter sensitivity and deviation between observed and simulated discharge are lower ranked, and it is less probable that they belong to the best model runs. On the other side, model runs with high sensitivity in times of good model performance and low sensitivity on days with poor model performance are better ranked.

For each of the 2,000 model runs, both RSR and swRSR as well as MAE and swMAE were computed separately for each model parameter for an identification period from 2000 to 2005.

These values of RSR and swRSR were first analyzed using correlation plots whereby similarities and differences between them were assessed to determine whether or not the information provided by swRSR is already captured by the classical RSR metric. Better performance with regard to swRSR compared to RSR for a parameter indicates that ignoring temporal weighting results in poorer identifiability.

Hereby, we distinguish three general cases as demonstrated in Figure 2a. In Case I, RSR and swRSR are similar in values and in variability. Since swRSR does not perform better than RSR for a given parameter, considering temporal variations in parameter dominance in the calculation of a performance criteria does not improve the results. In Case II, for a given value of RSR we see a relatively high degree of variability in

Table 3
Overview of Selected Performance Criteria and Their Focus on the Hydrograph

	Performance criteria	Part of the hydrograph	Optimum
NSE	Nash-Sutcliffe Efficiency	High flows	1
KGE	Kling-Gupta Efficiency	Entire hydrograph	1
KGE_alpha	KGE variability component	Variability	1
KGE_beta	KGE bias component	Water balance	1
KGE_r	KGE correlation component	Timing	1
FDC_very_high	RSR for FDC segment (0–5% exceedance)	Very high flows	0
FDC_high	RSR for FDC segment (5–20% exceedance)	High flows	0
FDC_mid	RSR for FDC segment (20–70% exceedance)	Middle flows	0
FDC_low	RSR for FDC segment (70–95% exceedance)	Low flows	0
FDC_very_low	RSR for FDC segment (95–100% exceedance)	Very low flows	0

swRSR (different parameter values across the range result in different values of the criterion). This means that the use of the latter as the performance criterion can help to reduce the perception of equifinality associated with. This indicates the benefit of using a sensitivity-weighted identifiability metric. Case III shows the opposite case, wherein the variability in RSR is higher, and therefore, parameter uncertainty and equifinality increases by using a sensitivity-weighted approach.

2.6. Parameter Identifiability Plots

Using RSR and swRSR metric values obtained for the 2,000 model runs, parameter identifiability plots were constructed for each of the approaches, using the Kernel density estimation routine in R. These plots provide information regarding the parts of the initial parameter range that are most strongly associated with “good” model simulations as assessed by a particular performance criterion. When constructed for the entire set of model simulations, this distribution is uniform but changes in shape when selecting the “best” model runs; a small range of parameter values with a strong peak indicates high parameter identifiability (Case II/III in Figure 2b), while a flat or unsystematic density curve with little peakedness indicates low parameter identifiability (Case I in Figure 2b) (Wagener et al., 2001, 2003). Here we used the best 500 (25%) of the model runs to construct the constrained parameter ranges. At first, the procedure is applied to RSR. Then, swRSR is calculated separately for each parameter. Hereby, each parameter gets its own best 500 model runs. This results in 13 selections of the best run.

Figure 2b (Case II) shows higher parameter identifiability in the sensitivity-weighted approach using swRSR, while in Case III, the classical approach of using RSR leads to higher parameter identifiability.

2.7. Parameter Constraints

The parameter identifiability plots were used to infer constrained parameter ranges for each of the model parameters (for each catchment). For each of the three catchments, we constrained the ranges of each parameter according to the density function in the parameter identifiability plot. At first, we detected the maximum (peak) density for a parameter. Then, the parameter ranges were constrained to parameter values that are greater or equal to the half of the maximum density. In other words, parameter areas below the half of the maximum density were excluded. Thus, the steeper the density curve, the smaller is the parameter range and the higher is the parameter identifiability. Note that only data from the identification period was used for this step. See Guse et al. (2020) for further details.

In the first approach, the ranges of all parameters were selected based on RSR. In the second approach, the ranges of each parameter were identified based on the individual sensitivity-weighted metric (swRSR). Thus, each of the two approaches (RSR and swRSR) resulted in two different parameter ranges.

2.8. Assessing Model Performance

Next, we evaluated and compared the resulting model performance thus obtained. These constrained parameter ranges were used for a second round of model runs. Again, we selected 2,000 parameter sets using

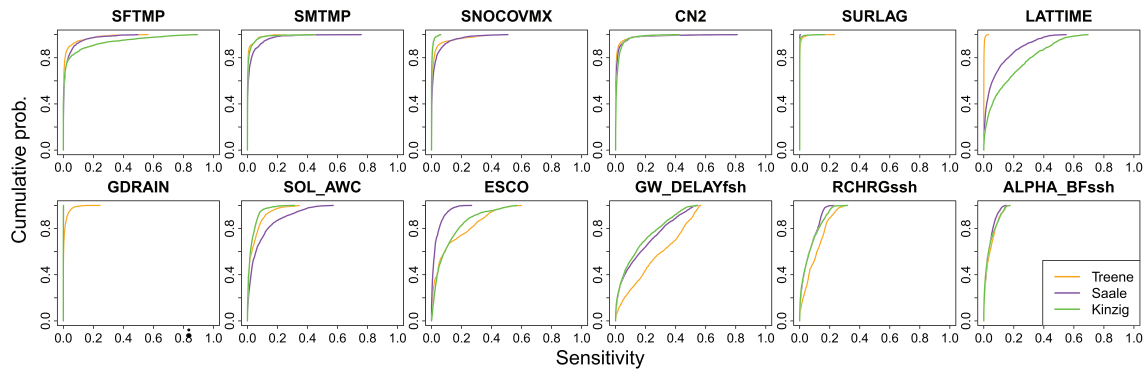


Figure 3. Cumulative distribution of daily values of sensitivity for the calibration period for the three catchments.

Latin Hypercube sampling, but this time with separate parameter ranges for RSR and swRSR. The distributions of model performance associated with these 2,000 model simulations were then assessed for both the identification period (2000–2005) and the independent evaluation period (2006–2010).

Model performance was assessed using a set of 10 complementary performance criteria as used in Guse et al. (2017) (see details in Table 3). Five of these were statistical model performance criteria, namely, the Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) and the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009), to evaluate the overall behavior. Moreover, the three components of the KGE are used to analyze the model performance regarding variability (KGE_alpha), bias (KGE_beta), and correlation (KGE_r). Using them, the dynamics of the hydrograph are adequately considered. The other five performance criteria are signature-based measures based on the flow duration curve to capture model errors in relation to discharge magnitudes (Euser et al., 2013; Shafii & Tolson, 2015; Yilmaz et al., 2008; Zhang et al., 2014). RSR is separately computed for five segments of the flow duration curve (FDC): FDC_very_high (0–5% exceedance), FDC_high (5–20%), FDC_mid (20–70%), FDC_low (70–95%), and FDC_very_low (95–100%) (Haas et al., 2016; Pfannerstill et al., 2014b; Yilmaz et al., 2008).

3. Results

3.1. Parameter Sensitivity

Figure 3 presents cumulative distribution functions of the daily parameter sensitivities computed for each of the three catchments using TEDPAS for the identification period (2000–2005) (see major results of this study in Guse et al., 2020). Two groups of parameters can be distinguished. The first group includes the snow and surface runoff parameters that are only sensitive for a short period of the year (for the majority of the time, their sensitivities are close to zero, but for a few days a high sensitivity is detected).

The second group includes lateral and groundwater flow and soil parameters, which are characterized by sensitivities over longer time periods, but with maximum values that are lower than for the parameters in the first group. Partly, specific parameter patterns were detected such as low sensitivity of lateral flow lag time (LATTIME) in the Treene and evaporation (ESCO) in the Saale.

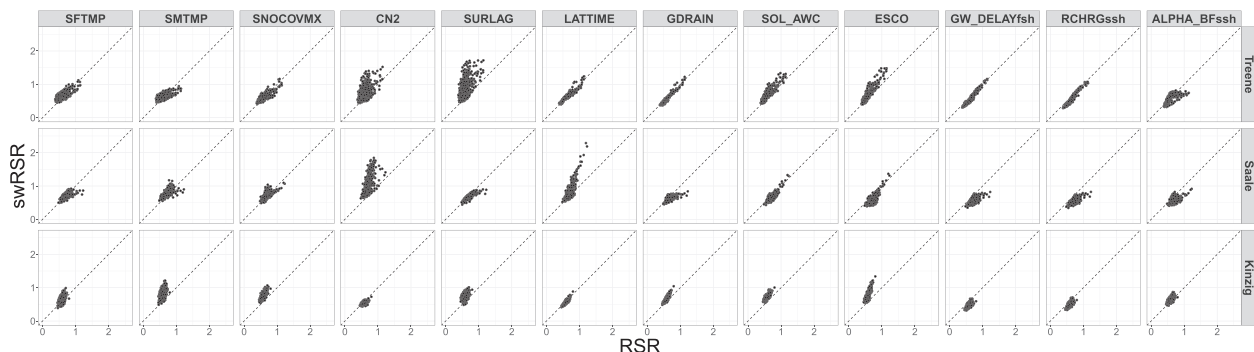


Figure 4. Comparison of RSR with RSR weighted with the sensitivity time series.

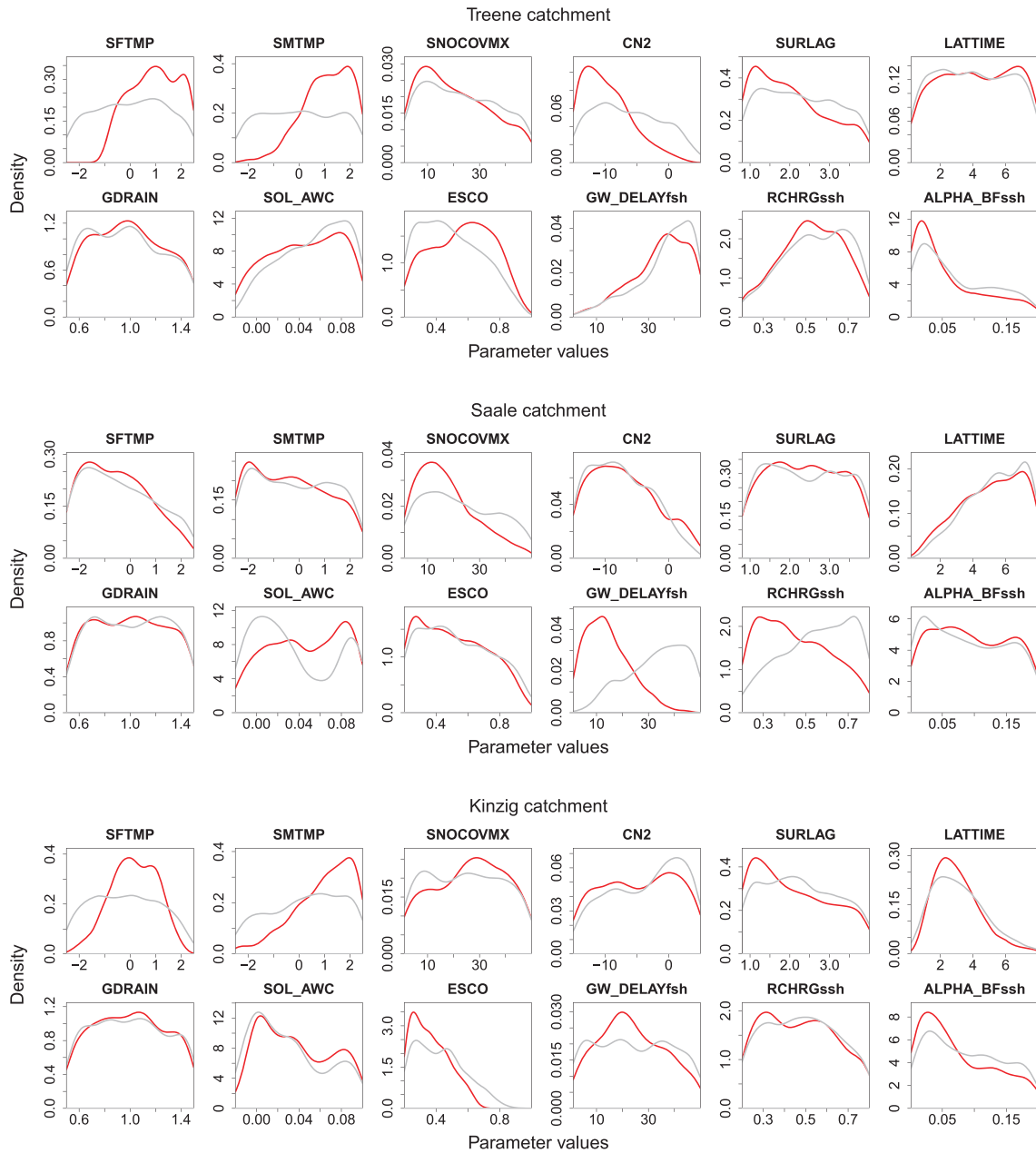


Figure 5. Parameter identifiability plots for all three catchments. Comparison of RSR (in gray) with RSR weighted with the sensitivity time series (in red).

3.2. Parameter Identifiability

3.2.1. Variability in Identifiability Metrics

The variability in the values of RSR and swRSR within the 2,000 model runs is shown in Figure 4. In the Treene catchment, tight grouping of the results as diagonal lines going from bottom left toward the top right is seen for groundwater parameters GW_DELAYfsh and RCHRGssh. This indicates that the impact of the parameters associated to the most dominant processes on model performance is already well captured using RSR as the performance metric (see case I in Figure 2a). In contrast, larger variability is seen for the surface runoff parameters, with the variability in swRSR being larger than RSR for parameters CN2 (SCS, 1972) and SURLAG (see case II in Figure 2a). High variability in a performance criterion facilitates the decision regarding suitable model runs. Since the use of sensitivity-weighted RSR results in the high flow phases being more highly weighted for the surface runoff parameters, this indicates that RSR does not provide adequate information about the effects of those parameters on the flow peaks. Moreover, since an RSR value

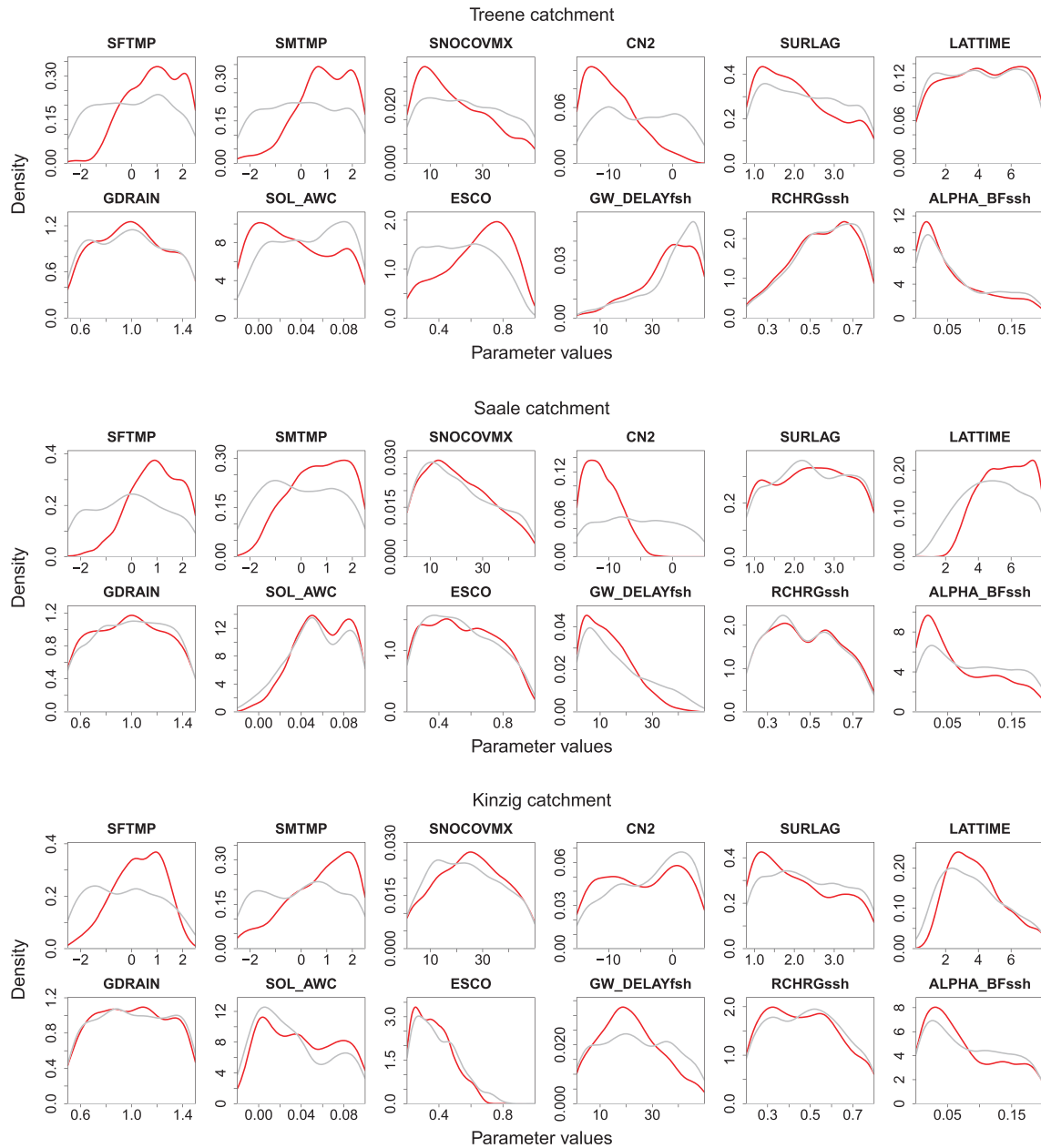


Figure 6. Parameter identifiability plots for all three catchments. Comparison of MAE (in gray) with MAE weighted with the sensitivity time series (in red).

of about 0.5 results in distinctly different swRSR values for parameters CN2 and SURLAG, we can infer that it is necessary to identify these parameters separately that are associated to a process of moderate relevance (using the weighted approach that targets each parameter individually).

Similarly, in the Saale catchment, the relative variability associated with swRSR (compared to RSR) is larger for parameters CN2 and LATTIME but small for the snow and groundwater parameters, indicating that sensitivity-based weighting leads to a improved model performance of the high flows, whereas performance with regard to the low flows is already good using RSR. In the Kinzig catchment, the relative variability associated with swRSR (compared to RSR) is larger for the snow and evaporation parameters. The variability in swRSR is lower for CN2 but higher for SURLAG.

3.2.2. Parameter Identifiability Plots

Figure 5 shows the 12 parameter identifiability plots obtained for each of the three catchments for the identification period for RSR and swRSR. In the Treene catchment, the use of swRSR results in improved

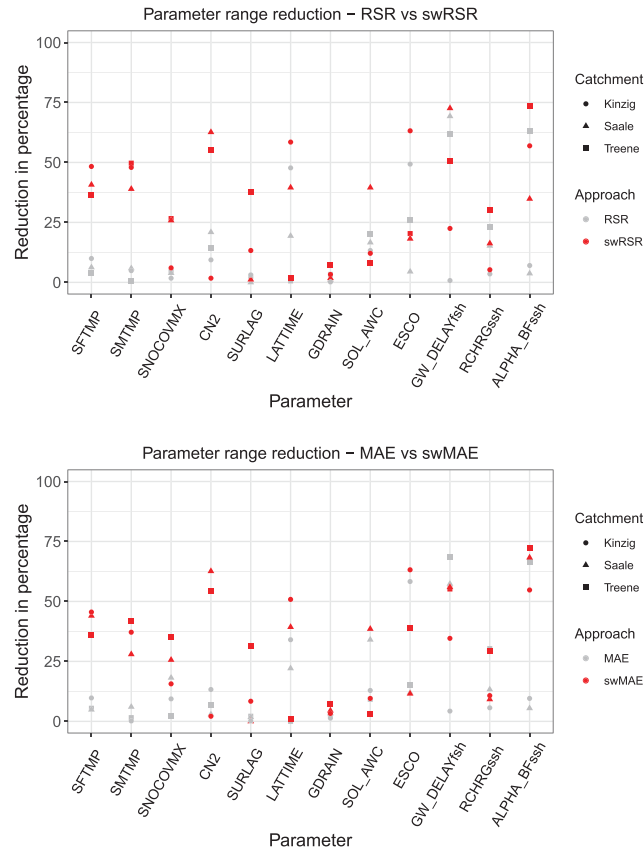


Figure 7. Range reduction for the twelve model parameters compared above between RSR and swRSR (weighted with the sensitivity time series) and below between MAE and swMAE in the three catchments.

parameter identifiability for snow (SFTMP, SMTMP, and SNOCOVMX) and surface runoff parameters (CN2 and SURLAG), and the peak of the parameter identifiability plot for evaporation regulation (ESCO) and for a groundwater parameter (RCHRGssh) moving to different optimum values compared to the classical approach. The sensitivity-weighted parameter identification approach leads to less precise parameter identification only for two parameters—SOL_AWC and GW_DELAYfsh—in which case the classical approach resulted in better parameter identifiability. In the Saale catchment, the use of swRSR results in improved parameter identifiability for parameters associated with snow (SNOCOVMX) and groundwater (GW_DELAYfsh). The peaks are different between RSR and swRSR for two groundwater parameters (GW_DELAYfsh and RCHRGssh). In the Kinzig catchment, the use of swRSR results in improved parameter identifiability for nine of the model parameters, with the strongest improvement being for the snow (SFTMP and SMTMP) and evaporation (ESCO) parameters.

Parameter identifiability plots for MAE are shown in Figure 6. The results are similar for Treene and Kinzig with slight differences. For both catchments, the most appropriate parameter values as derived from the parameter identifiability plots are similar. In contrast, the parameter identifiability is evaluated partly differently between MAE and swMAE compared to RSR and swRSR for the Saale catchment. Precise identifiability is obtained using swMAE for the snow parameters SFTMP and SMTMP as well as the surface runoff parameter CN2. Partly, also the location of the peak of the parameter identifiability plot varies between swMAE and swRSR as it could be expected in a parameter optimization with different performance criteria. However, the overall pattern that the parameter identifiability plots are more precise in the sensitivity-weighted approach is supported both by swRSR and swMAE.

For all of the catchments, the parameter ranges are more strongly constrained using the sensitivity-weighted approach. The largest range reduction was obtained for the three snow parameters (SFTMP, SMTMP, and SNOCOVMX) and the two surface runoff parameters (CN2 and SURLAG). For the snowfall (SFTMP) and

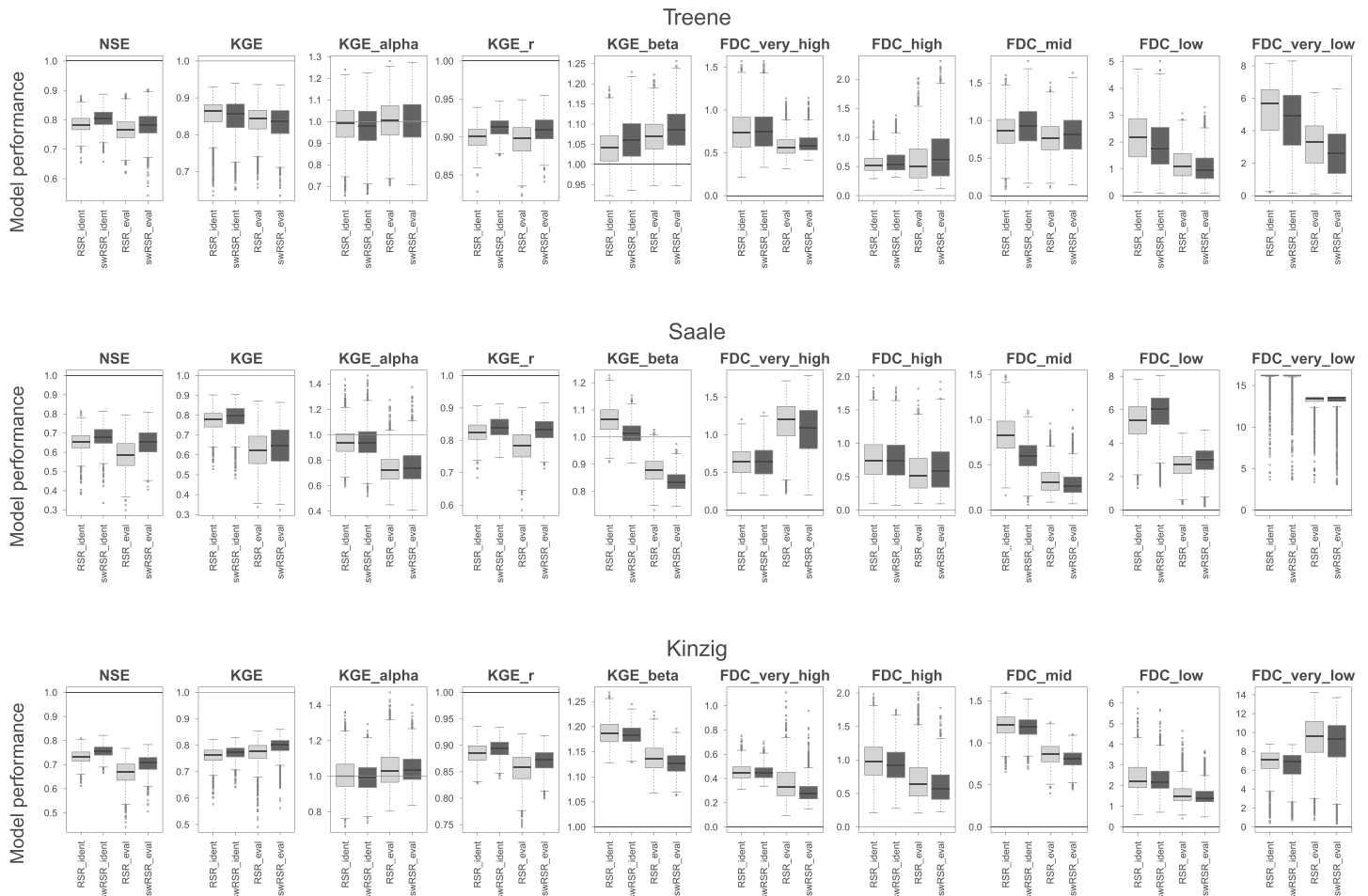


Figure 8. Comparison of model performance for 10 contrasting performance criteria between model simulations with constrained parameter ranges based on RSR and swRSR (weighted with the sensitivity time series). The abbreviations ident and eval stands for identification and evaluation period.

snow melt (SMTMP) temperature parameters, the range reduction achieved using swRSR and swMAE is distinctly higher for all three catchments.

Figure 7 shows the percentage range reduction (reduced uncertainty) achieved for each of the 12 parameters achieved using either RSR or swRSR and MAE or swMAE, respectively. For all of the catchments, the parameter ranges are more strongly constrained using the sensitivity-weighted approach. The largest range reduction was obtained for the three snow parameters (SFTMP, SMTMP, and SNOCOVX) and the two surface runoff parameters (CN2 and SURLAG). For the snowfall (SFTMP) and snow melt (SMTMP) temperature parameters, the range reduction achieved using RSR_w is distinctly higher for all three catchments (see strong separation of the red and grey markers in Figure 5).

3.3. Model Performance

The distributions of model performance, assessed using 10 different criteria, obtained for the final set of 2,000 model simulations representing parameter sets sampled uniformly from the final parameter space obtained using either the classical (RSR) or the sensitivity-weighted (swRSR) approaches is shown in Figure 8 and for MAE and swMAE in Figure 9. When using the sensitivity-weighted approach to constrain the parameter ranges, model performance was improved for the Treene catchment only for four of the performance criteria while being lower for five of them both for swRSR and swMAE. Model performance in the Saale and Kinzig catchments tends to be superior or is on the same level in the majority of the cases (on both identification and evaluation periods). It appears that use of the sensitivity-weighted approach results in improved constraints mainly with regard to the low and very low flow segments of the FDC as well as for NSE and KGE_r. In contrast, it is lower for KGE_{alpha} and RSR for the high flow segment of the FDC. A special pattern was

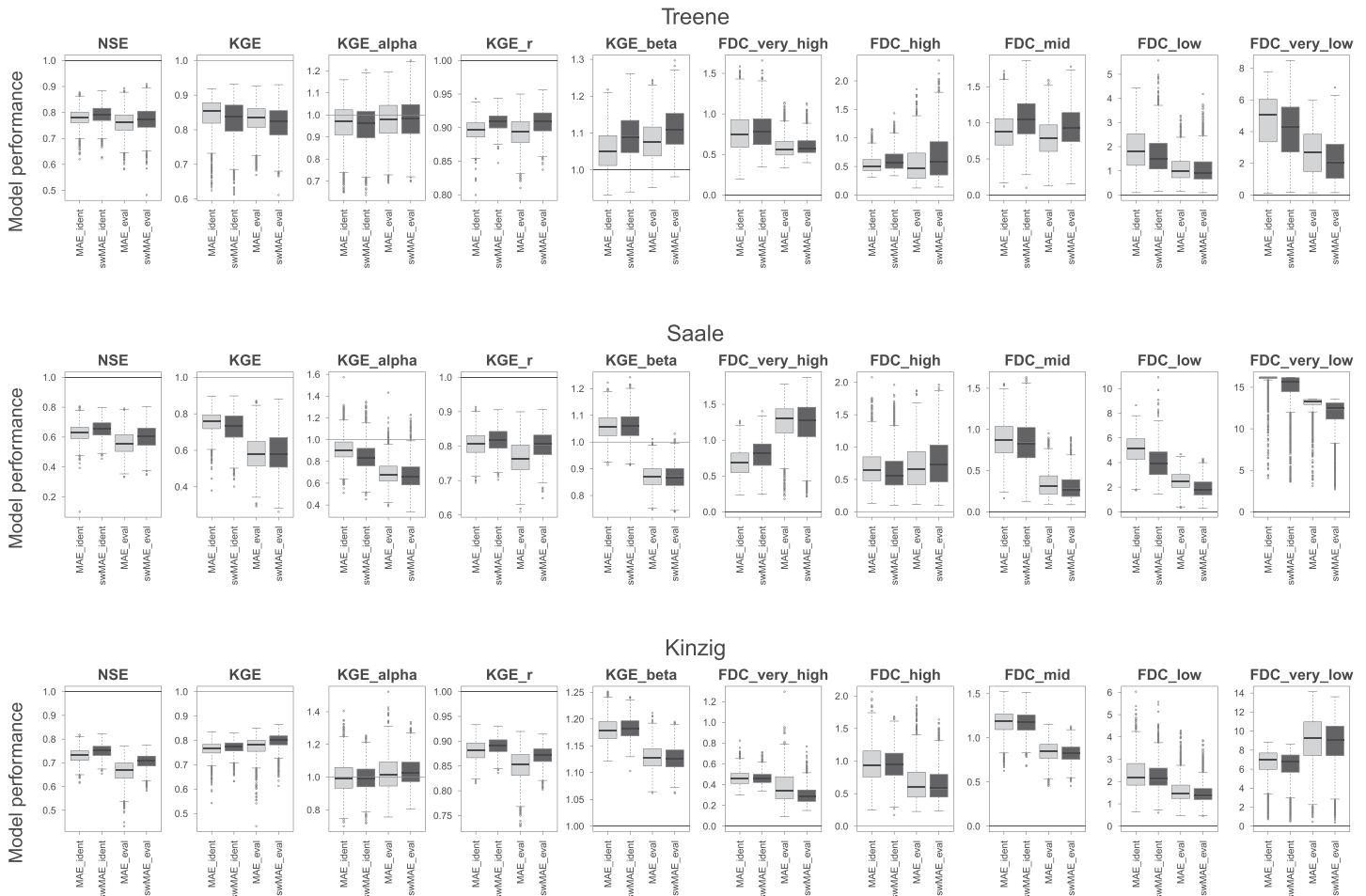


Figure 9. Comparison of model performance for 10 contrasting performance criteria between model simulations with constrained parameter ranges based on MAE and swMAE (weighted with the sensitivity time series). The abbreviations ident and eval stands for identification and evaluation period.

observed for the KGE_beta statistic that evaluates the goodness of the simulated water balance. For the Saale and Kinzig catchments, the identification period is wetter than the evaluation period, and higher KGE_beta values are obtained using both RSR and swRSR as well as MAE and swMAE. We do not see major differences in using RSR and swRSR or MAE and swMAE, respectively.

4. Discussion

4.1. Parameter Identifiability

Regarding parameter identifiability, our main point was to show which parameters are better constrained in their values by a sensitivity-weighted metric. Overall, our results indicate that for the three catchments tested, the sensitivity-weighted approach results in better parameter identifiability for more of the model parameters than is achieved using the nonweighted approach. This suggests that the parameter constraints as identified in the sensitivity-weighted approach improves the parameter representation.

For model parameters that tend to be continuously relevant through the year, and therefore exert dominant overall influences on the modeled discharge, the sensitivity-weighted approach does not offer much benefit over the nonweighted approach. This manifests as similar parameter identifiability plots for groundwater and evapotranspiration for the Treene lowland catchment, where the hydrological behavior is strongly controlled by those parameters throughout the year.

In contrast, snow parameters have a strong seasonal dynamic (since their relevance is focused on the snow period), and surface runoff parameters that are activated only during periods of high precipitation intensity are better identified using the sensitivity-weighted approach. In particular, the sensitivity time series

associated with the surface runoff parameters fluctuate highly from low sensitivity for the majority of the time and high sensitivity during very short time periods. For parameters such as these, the use of a separate performance criterion is required to properly extract information regarding their optimal values from the data. Typical performance criteria that do not account for the temporal variations in parameter dominance are inadequate in the presence of such strong temporal sensitivity dynamics. The hydrological responses of the Kinzig and Saale catchments are controlled by a mixture of processes that act over both the long and short terms. The important issue is that moderately important parameters with temporally varying dominance are typically not identifiable using metric approaches that employ fixed temporal weighting of the entire time series.

The issue of parameter correlation has low impact on our approach since we calculated separate sensitivity-weighted performance criteria for each parameter to constrain its ranges. Thus, even in the case of higher correlation among the parameters, the major outcomes are not affected.

The improvement in parameter identifiability using the sensitivity-weighted approach is achieved for two different performance criteria. The results using MAE instead of RSR does not alter the main message of the study. There are different ways to select the sampling algorithm, the identifiability method, and the metric, and which one to use is a matter of subjective choice, based on which ones better meet the needs of the user.

In principle, any approach to generating sensitivity-time-weights and not only TEDPAS can be applied to construct sensitivity-weighted performance criteria. Given the generic nature of the methodology, in future work, it could be examined how the selection of a sensitivity analysis method impact the results.

4.2. Model Performance

The use of appropriate temporal weighting results in a set of performance criteria that improves parameter identifiability without having to design specific signature-type metrics for each parameter (or process). This is addressed in our approach by introducing a sensitivity-weighted performance criterion.

This results in a multicriteria replacement to the single criterion optimization problem. In our manuscript we have presented a sensitivity-weighted metric. We think that a comparison to a single metric optimization (in our case RSR and MAE) is the most apparent way of analyzing the benefit of the sensitivity-weighted metric.

The typical limitation of multiobjective studies, where poor informativeness, parameter interdependence, and objective function interdependence can jointly conspire to complicate identifiability (Gupta et al., 1998), is not necessarily a major problem in our case. By constructing a different objective function for each parameter (based on time series of specific sensitivities to that parameter and that parameter alone), we are able to assess each parameter separately. This is in contrast to other multiobjective studies in which each criterion is applied to the entire set of parameters and thus contradicting and compensating results are to be expected.

We seriously considered a comparison to a signature-based approach but see some problems. Signatures are intended to achieve a stronger relationship to processes and to better represent processes in models (Gupta et al., 2008; Yilmaz et al., 2008). Each signature is related to a certain process or to a part of the hydrograph. Thus, it is always required to select a complete set of signatures to fully capture the hydrological behavior in models. The specification of performance criteria for each process and the associated model parameters is a great challenge, in particular since the adequate set of performance criteria can change among different models and requires thus a model-specific search for appropriate performance criteria. There is not yet consensus or agreement on a suitable set of signatures for model optimization of hydrological models. Therefore, any decision regarding which signature to select will unavoidably be subjective and results in a focus on specific processes (or parts of hydrograph) while neglecting others.

We used a set of 10 informative statistical and signature-based criteria to evaluate the model performance obtained for an ensemble of 2,000 model simulations generated using the constrained parameter ranges obtained by the classical and sensitivity-weighted approaches. The use of these criteria helps to ensure that all phases of the hydrograph are specifically considered (Pfanterstill et al., 2014b; Yilmaz et al., 2008). The information extraction gains are very clearly apparent in the parameter identifiability plots. Improved precision in the identification of a larger number of model parameters (here, the snow and surface runoff parameters) and the resulting parameter constraints tend to result in an overall improvement in model performance in two of the three catchments. This finding is consistent with other studies that have

demonstrated the benefits of constraining parameters to achieve more behavioral model simulations (Gharari et al., 2014; Hogue et al., 2006; Wagener, 2007).

So to reiterate, we present our approach for improving parameter identifiability as being applicable prior to the implementation of a model calibration procedure and as a method suited to the selection of appropriate parameter ranges for model calibration. The consequence of a reduced parameter range (and the consequent overall parameter space) allows for more efficient parameter sampling and search. Hereby, it was not the intention of our manuscript to present a method for identifying the best model run(s) such as is the general goal of model optimization.

5. Conclusion

In this study, a sensitivity-weighted performance criterion was developed and overall parameter identifiability is improved as a result of incorporating sensitivity time series information as weights into the calculation of the performance criterion. The main outcome from our study is that the use of a sensitivity-weighted metric improves the identifiability of parameters that are seasonally dominant (e.g., snow parameters) or that are only dominant during a short period of time (e.g., parameters determining surface runoff). Thus, improvement in parameter identifiability is most notable for catchments where the most relevant hydrological processes tend to have strong seasonal variations or only active for relatively short periods of time. Our approach enables the identification of parameters that are of major dominance and also parameters that may only be of moderate relevance most of the time but are of high relevance for short periods of time. It leads to an improvement of parameter identifiability and in the majority of the cases also of model performance as assessed in using 10 contrasting performance criteria. Increased parameter identifiability leads to more precision (lower uncertainty ranges) in the estimated values of the model parameters. In conclusion, the sensitivity-weighted approach helps to maximize the information extracted from model simulations, leading to improved parameter identifiability by avoiding the need to design targeted signature-type criteria for every process or parameter.

Acknowledgments

We thank Schleswig-Holstein Agency for Coastal Defence, National Park and Marine Conservation of Schleswig-Holstein (LKN-SH), the State Institute for Environment and Geology of Thuringia (TLUG) and Hessian Agency for Nature Conservation, Environment and Geology (HLNUG) for providing discharge data. We thank our cooperating scientists Jens Kiesel (IGB Berlin) for the SWAT model set-up for the Kinzig catchment (GLANCE project [Global change effects in river ecosystems; 01LN1320A] supported by the German Federal Ministry of Education and Research [BMBF]) as well as Martin Volk and Michael Strauch (UFZ Leipzig) for their support in the set-up of the SWAT model for the Saale catchment (Helmholtz Programme Terrestrial Environmental Research). The first author (B. G.) thanks the DFG for financial support (project GU 1466/1-1 Hydrological consistency in modeling). The last author (H. G.) acknowledges partial support by the Australian Centre of Excellence for Climate System Science (CE110001028). We would like to thank the community of the open source software R, which was used for these analyses. We thank the Associated Editor Juliane Mai and four anonymous reviewers for their feedback to our article that have raised fruitful discussions and improved the article substantially.

Data Availability Statement

The major results will be provided via GFZ Data Services (see the preliminary link: <https://doi.org/10.5880/GFZ.4.4.2019.004>).

References

- Abebe, N. A., Ogden, F. L., & Pradhan, N. R. (2010). Sensitivity and uncertainty analysis of the conceptual HBV rainfall-runoff model: Implications for parameter estimation. *Journal of Hydrology*, *389*, 301–310.
- Arnold, J. G., Srinivasan, R., Mutiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: Model development. *Journal of the American Water Resources Association*, *34*, 73–89.
- Beven, K. J., & Binley, A. M. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*, 279–298.
- Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, *36*(12), 3663–3674.
- Cibin, R., Sudheer, K. P., & Chaubey, I. (2010). Sensitivity and identifiability of stream flow generation parameters of the SWAT model. *Hydrological Processes*, *24*, 1133–1148.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008). Framework for understanding structural errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*, W00B02. <https://doi.org/10.1029/2007WR006735>
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients I. Theory. *The Journal of Chemical Physics*, *59*(8), 3873–3878.
- de Vos, N. J., Rientjes, T. H., & Gupta, H. V. (2010). Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering. *Hydrological Processes*, *24*, 2840–2850.
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, *17*, 1893–1912.
- Fenicia, F., McDonnell, J. J., & Savenije, H. H. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, *44*, W06419. <https://doi.org/10.1029/2007WR006386>
- Fenicia, F., Savenije, H. H., Matgen, P., & Pfister, L. (2007). A comparison of alternative multiobjective calibration strategies for hydrological modeling. *Water Resources Research*, *43*, W03434. <https://doi.org/10.1029/2006WR005098>
- Fenicia, F., Savenije, H. H., & Winsemius, H. C. (2008). Moving from model calibration towards process understanding. *Physics and Chemistry of the Earth*, *33*, 1057–1060.
- Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. (2013). An approach to identify time consistent model parameters: Sub-period calibration. *Hydrology and Earth System Sciences*, *17*, 149–161.
- Gharari, S., Shafiei, M., Hrachowitz, M., Kumar, R., Fenicia, F., Gupta, H. V., & Savenije, H. H. (2014). A constraint-based search algorithm for parameter identification of environmental models. *Hydrology and Earth System Sciences*, *18*, 4861–4870.

- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*, 80–91.
- Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for Dynamical Earth System Models. *Water Resources Research*, *54*(11), 8692–8717. <https://doi.org/10.1029/2018WR022668>
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, *34*(4), 751–763.
- Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*, 3802–3813.
- Guse, B., Kiesel, J., Pfannerstill, M., & Fohrer, N. (2020). Assessing parameter identifiability for multiple performance criteria to constrain model parameters. *Hydrological Sciences Journal*, *65*, 1158–1172.
- Guse, B., Pfannerstill, M., Gafurov, A., Fohrer, N., & Gupta, H. V. (2016). Demasking the integrated information of discharge: Advancing sensitivity analyses to consider different hydrological components and their rates of change. *Water Resources Research*, *52*, 8724–8743. <https://doi.org/10.1002/2016WR018894>
- Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., & Fohrer, N. (2017). Identifying the connective strength between model parameters and performance criteria. *Hydrology and Earth System Sciences*, *21*, 5663–5679. <https://doi.org/10.5194/hess-21-5663-2017>
- Guse, B., Pfannerstill, M., Kiesel, J., Strauch, M., Volk, M., & Fohrer, N. (2019). Analysing spatio-temporal process and parameter dynamics in models to characterise contrasting catchments. *Journal of Hydrology*, *507*, 863–874. <https://doi.org/10.1016/j.jhydrol.2018.12.050>
- Guse, B., Pfannerstill, M., Kiesel, J., Strauch, M., Volk, M., Gupta, H., & Fohrer, N. (2020). Simulated sensitivity time series and model performance in three german catchments. *GFZ Data Services*. https://doi.org/10.5880/GFZ.4.4.2019_004
- Guse, B., Pfannerstill, M., Strauch, M., Reusser, D. E., Volk, M., Gupta, H. V., & Fohrer, N. (2016). On characterizing the temporal dominance patterns of model parameters and processes. *Hydrological Processes*, *30*, 2255–2270. <https://doi.org/10.1002/hyp.10764>
- Guse, B., Reusser, D. E., & Fohrer, N. (2014). How to improve the representation of hydrological processes in SWAT for a lowland catchment—Temporal analysis of parameter sensitivity and model performance. *Hydrological Processes*, *28*, 2651–2670. <https://doi.org/10.1002/hyp.977>
- Haas, M. B., Guse, B., Pfannerstill, M., & Fohrer, N. (2016). A joined multi-metric calibration of river discharge and nitrate loads with different performance measures. *Journal of Hydrology*, *536*, 534–545. <https://doi.org/10.1016/j.jhydrol.2016.03.001>
- He, Z. H., Tian, F. Q., Gupta, H. V., Hu, H. C., & Hu, H. P. (2015). Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning. *Hydrology and Earth System Sciences*, *19*, 1807–1826.
- Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). From maps to movies: High resolution time-varying sensitivity analysis for spatially distributed watershed models. *Hydrology and Earth System Sciences*, *17*, 5109–5125.
- Herman, J. D., Reed, P. M., & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, *49*, 1400–1414. <https://doi.org/10.1002/wrcr.20124>
- Hogue, T. S., Gupta, H., & Sorooshian, S. (2006). A user-friendly approach to parameter estimation in hydrologic models. *Journal of Hydrology*, *320*, 202–217.
- Hornberger, G., & Spear, R. (1981). Approach to the preliminary analysis of environmental systems. *Journal of Environmental Economics and Management*, *12*, 7–18.
- Kakouei, K., Kiesel, J., Domisch, S., Irving, K. S., Jaehnis, S. C., & Kail, J. (2018). Projected effects of climate-change-induced flow alterations on stream macroinvertebrate abundances. *Ecology and Evolution*, *8*(6), 3393–3409. <https://doi.org/10.1002/ece3.3907>
- Kelleher, C., Wagener, T., McGlynn, B., Ward, A. S., Gooseff, M. N., & Payn, R. A. (2013). Identifiability of transient storage model parameters along a mountain stream. *Water Resources Research*, *49*, 5290–5306. <https://doi.org/10.1002/wrcr.20413>
- Krause, P., Boyle, D. P., & Baese, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97.
- Madsen, H., Wilson, G., & Ammentorp, H. C. (2002). Comparison of different automated strategies for calibration of rainfall-runoff models. *Journal of Hydrology*, *261*, 48–59.
- Massmann, C., Wagener, T., & Holzmann, H. (2014). A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales. *Environmental Modelling & Software*, *51*, 190–194. <https://doi.org/10.1016/j.envsoft.2013.09.033>
- Moriasi, D. N., Arnold, J. R., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*, 885–900.
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models: Part I—A discussion of principles. *Journal of Hydrology*, *10*, 282–290.
- Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). Soil and water assessment tool—Theoretical documentation version 2009. *Texas Water Resources Institute Technical Report*, *406*.
- Pechlivanidis, I. G., Jackson, B. M., McMillan, H. K., & Gupta, H. V. (2012). Using an information entropy-based metric as a diagnostic of flow duration to drive model parameter identification. *Global NEST*, *14*(3), 325–334.
- Pfannerstill, M., Guse, B., & Fohrer, N. (2014a). A multi-storage groundwater concept for the SWAT model to emphasize nonlinear groundwater dynamics in lowland catchments. *Hydrological Processes*, *28*, 5599–5612.
- Pfannerstill, M., Guse, B., & Fohrer, N. (2014b). Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *Journal of Hydrology*, *510*, 447–458.
- Reusser, D. E. (2013). FAST: Implementation of the Fourier amplitude sensitivity test (FAST). R package version 0.63. <http://CRAN.R-project.org/package=fast>
- Reusser, D. E., Blume, T., Schaeffli, B., & Zehe, E. (2009). Analysing the temporal dynamics of model performance for hydrological models. *Hydrology and Earth System Sciences*, *13*, 999–1018.
- Reusser, D. E., Buytaert, W., & Zehe, E. (2011). Temporal dynamics of model parameter sensitivity for computationally expensive models with FAST (Fourier Amplitude Sensitivity Test). *Water Resources Research*, *47*, W07551. <https://doi.org/10.1029/2010WR009947>
- Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*, *91*, 1109–1125.
- SCS (1972). Chapter 4 Hydrology. *National engineering handbook*. Soil Conservation Service. Washington DC: United States Department for Agriculture.
- Shafii, M., & Tolson, B. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, *51*, 3796–3814. <https://doi.org/10.1002/2014WR016520>
- Shin, M.-J., Guillaume, J. H., Croke, B. F., & Jakeman, A. J. (2015). A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. *Journal of Hydrology*, *520*, 1–16.

- Soetaert, K., & Petzoldt, T. (2010). Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *Journal of Statistical Software*, *33*(3), 1–28.
- Sorooshian, S., Gupta, V. K., & Fulton, J. L. (1983). Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models—Influence of calibration data variability and length on model credibility. *Water Resources Research*, *19*(1), 251–259.
- Tang, Y., Reed, P., van Werkhoven, K., & Wagener, T. (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis. *Water Resources Research*, *43*, W06415. <https://doi.org/10.1029/2006WR005813>
- van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, *44*, W01429. <https://doi.org/10.1029/2007WR006271>
- Wagener, T. (2007). Can we model the hydrological impacts of environmental change? *Hydrological Processes*, *21*, 3233–3236.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, *5*(1), 13–26.
- Wagener, T., Camacho, L. A., & Wheater, H. S. (2002). Dynamic identifiability analysis of the transient storage model for solute transport in rivers. *Journal of Hydroinformatics*, *4*, 199–211.
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, *17*, 455–476.
- Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*, W09417. <https://doi.org/10.1029/2007WR006716>
- Zhang, H., Huang, G. H., Wang, D., & Zhang, X. (2011). Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. *Advances in Water Resources*, *34*, 1292–1303.
- Zhang, Y., Vaze, J., Chiew, F. H., Teng, J., & Li, M. (2014). Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall runoff modelling? *Journal of Hydrology*, *517*, 936–948.