Journal of Hydrology 553 (2017) 722–734

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Representative parameter estimation for hydrological models using a lexicographic calibration strategy

Marlene Gelleszun *, Phillip Kreye, Günter Meon

TU Braunschweig, Leichtweiß-Institut für Wasserbau, Abteilung Hydrologie, Wasserwirtschaft und Gewässerschutz, Beethovenstraße 51a, 38106 Braunschweig, Germany

ARTICLE INFO

Article history:
Received 9 December 2016
Received in revised form 7 July 2017
Accepted 10 August 2017
Available online 17 August 2017
This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Yu Zhang, Associate Editor

Keywords:
Hydrological modelling
Calibration
Lexicographic approach
Optimisation
Sensitivity analysis

ABSTRACT

We introduce the developed lexicographic calibration strategy to circumvent the imbalance between sophisticated hydrological models in combination with complex optimisation algorithms. The criteria for the evaluation of the approach were (i) robustness and transferability of the resulting parameters, (ii) goodness-of-fit criteria in calibration and validation and (iii) time-efficiency. An order of preference was determined prior to the calibration and the parameters were separated into groups for a stepwise calibration to reduce the search space. A comparison with the global optimisation method SCE-UA showed that only 6% of the calculation time was needed; the conditions total volume, seasonality and shape of the hydrograph were successfully achieved for the calibration and for the cross-validation periods. Furthermore, the parameter sets obtained by the lexicographic calibration strategy for different time periods were much more similar to each other than the parameters obtained by SCE-UA. Besides the similarities of the parameter sets, the goodness-of-fit criteria for the cross-validation were better for the lexicographic approach and the water balance components were also more similar. Thus, we concluded that the resulting parameters were more representative for the corresponding catchments and therefore more suitable for transferability. Time-efficient approximate methods were used to account for parameter uncertainty, confidence intervals and the stability of the solution in the optimum.

1. Introduction

Many hydrological modelling applications deal with long-term simulations up to 100 years. This includes climate impact research or extreme value statistics. In this context, reliable statements can only be obtained if the model and its calibration are representative for the whole time period. The transferability of calibration parameters to independent validation periods is considered to be an important issue in modern hydrologic modelling tasks, because it is going hand in hand with robust optimisation techniques. Common practice to calibrate a hydrological model is to estimate model parameters iteratively. Due to the increased performance of modern computers, semi- or fully automatic optimisation algorithms have been established for calibration, especially with regard to scientific questions (Efstratiadis and Koutsoyiannis, 2010). Simultaneously, enhanced physically based process descriptions as well as improved available input data were more and more integrated into the spatially highly resolved models. Consequently, these developments induce increased calculation times and effort in calibration, which can hardly be solved with conventional "trial and error" methods (Hogue et al., 2000). Alternatively, automated optimisations methods can be used, albeit the disadvantages of long computation times and insufficient user participation. For this reasons, optimisation methods are mostly used in combination with conceptual models, often on a daily time step or purely for research purposes (Zhang et al., 2009). Applications of both highly developed hydrological models and complex optimisation methods are challenging for operational hydrology due to the enormous computing time (Zhang et al., 2009; Vaze et al., 2011). To overcome this imbalance, Zhang et al. (2016) proposed the use of a parallel optimisation approach by using a high-performance computer (HPC). Since HPCs are often not available, we introduce a lexicographic calibration strategy in this study, whereby the objectives are based on an order of preferences. It delivered representative parameter sets under the constraint of limited calculation effort, while keeping objectivity in contrast to a manual calibration. The hydrological community agrees that expert knowledge should be included in the calibration process (see. e.g. Moussa and Chahinian, 2009). Primarily, expert knowledge is necessary for the following processes: parameterisation of the catchment's properties, the correct choice of the objective function(s), selection of an appropriate optimisation algorithm, plausibility check of the obtained parameter sets

* Corresponding author.
E-mail addresses: M.Gelleszun@tu-bs.de (M. Gelleszun), P.Kreye@tu-bs.de (P. Kreye), G.Meon@tu-bs.de (G. Meon).

http://dx.doi.org/10.1016/j.jhydrol.2017.08.015
0022-1694/© 2017 The Authors. Published by Elsevier B.V.

and selection of the best parameter set. Boyle et al. (2000) coupled expert knowledge with automated optimisation methods by separating the hydrograph into different cases. The criteria of differentiation were periods with and without precipitation. For these two cases different objective functions have been defined. As a result, pareto optimal parameter sets were obtained by using the time-consuming MOCOM-algorithm[1] by Yapo et al. (1998). Hogue et al. (2000) considered the expert knowledge by performing a stepwise calibration. An order of preferences of the objectives has not been set and a global optimisation method (SCE-UA[2]) was used. Similarly, Fenicia et al. (2007) introduced a method for stepwise calibration by using a global optimisation algorithm. Model parameters were linked to processes and objective functions were defined for every one of those processes. Cullmann et al. (2008) grouped observed hydrographs based on hydrological characteristics in several classes. Parameter sets were then detected for each class set, representing the dominant process. The aim was to improve the model results for the flood forecast. For the operational application, a simultaneous use of all sets was not feasible due to the time factor. Hence, all parameter sets were used to train a neural network. This black box model was then applied for the event-based flood forecast. Other hydrological applications with respect to optimisation techniques are stated in the field of flow forecast with artificial neural network techniques, see e.g. Wu et al. (2009), Wang et al. (2015), Taormina and Chau (2015) or Chen et al. (2015). The aim was to improve the neural network performance in the estimate of daily flows. For parameter estimations in highly resolved process based distributed models these methods are not feasible.

Based on all these different approaches, we identified the deficits of representative and efficient parameter calibration: Either the calibration is dominated by mathematical solutions, which may struggle in validation. Alternatively, the calibration is performed manually by an expert (trial and error), which is maybe more representative but often not reproducible (even by the same person) and inefficient. Thus, the first objective of this study was to find representative and robust parameter sets. We incorporated expert knowledge at the beginning of the calibration process by determining an appropriate order of hydrologic objectives. The developed lexicographic calibration strategy (LCS) can be considered as an approach, where the order of preference depends on the scientific framework and hydrological model. Gelleszun et al. (2015) introduced a lexicographic calibration strategy, which delivered a single representative and optimal parameter set by defining an order of preference of the objective functions. The method was validated by using synthetic hydrographs and a distributed hydrological model. To achieve the representativeness of the estimated parameters, the second objective of this study was, to achieve good performances in calibration and particularly in validation. We did not intend to find the mathematical global minimum of one objective or multi-objective function during the calibration period solely, but we expected to identify parameter sets, which were valid for the hydrological system and thus for the validation periods respectively.

In general, we distinguish between optimisation algorithms and calibration strategies. The optimisation itself is considered as a mathematically definite process with the aim to find parameters in order to minimize the objective function. These approaches can be divided into local and global methods. Global optimisation methods include evolution strategies or genetic algorithms. Widespread methods for multi-objective optimisation in applied hydrology are MOCOM (Multi-Objective Complex Evolution) by Yapo et al. (1998), MOSCEM (Multi-Objective Shuffled Complex Evolution Metropolis algorithm) by Vrugt et al. (2003) or SCE-UA (Shuffled Complex Evolution algorithm) by Duan et al. (1992). A broad description of this subject is given in Efstratiadis and Koutsoyiannis (2010). The challenge of linking complex optimisation algorithms with computationally intensive hydrological models was analysed by Zhang et al. (2009). The parameter estimation of a hydrological model was implemented with different global and evolution-based optimisation methods. None of the tested methods required less than 500 iterations. The methods for multi-objective optimisation are computationally intensive, as complex structures within the objective function lead to many local minima (Abbaspour, 2005). Hence, there is generally a conflict between high-resolution models in conjunction with complex optimisation algorithms (Zhang et al., 2009). This leads to the third objective of this study, namely to achieve practicable calculation effort by minimizing the optimisation runs since the calculation time of a physically based distributed hydrological model is often high.

Summarised, in the study at hand, the main focus lays on the reproducibility of the parameter estimations in order to obtain representative parameter sets for gauged catchments. For two areas, the observed runoff time series of ten years length (01.11.2001 to 31.10.2011) were divided into five separate series of two years each. We applied the lexicographic calibration strategy to obtain for each time series an individual parameter set. We expected the resulting parameter sets to be similar to each other, as the individual time series originate from the identic runoff regime. To compare the overall quality of each parameter set, we cross-validated each set for (i) every other two-year period and (ii) the overall ten-year period. In addition to goodness-of-fit-criteria, such as model efficiency (Nash and Sutcliffe, 1970), we analysed the similarities of the obtained parameter sets and the resulted influences on the simulated water balance components. Further, we compared the obtained results with the lexicographic calibration strategy with results received by applying the global multi-criteria optimisation method SCE-UA by Duan et al. (1992). We additionally showed that an uncertainty analysis of complex hydrological models can be performed by applying the approximate first-order second-moment (FOSM) method.

2. Materials and methods

2.1. The hydrological model system PANTA RHEI

PANTA RHEI is a deterministic, semi-distributed, physically based hydrological model for long term or single event simulations. It has been developed by the Department of Hydrology, Water Management and Water protection, Leichtweiss Institute for Hydraulic Engineering and Water Resources, University of Braunschweig in co-operation with the "Institut für Wassermanagement IfW GmbH", Braunschweig (LWI-HYWAG und IfW, 2012). It has been successfully employed for scientific questions (Hölscher et al., 2012) and in numerous national and international projects (Meon and Pätsch, 2014; Wurpts et al., 2014). Furthermore, PANTA RHEI is applied in the operational flood forecast of the federal state Lower Saxony, Germany (Meyer et al., 2012). The temporal discretisation is adaptive and in many applications an hourly time step is used. The spatial differentiation is divided into three levels: HRUs (hydrologic response units), subcatchments and gauged catchments.

A modified Penman-Monteith method (Penman, 1948; Monteith, 1965) is used to estimate the evapotranspiration. It is one of the most established physically based methods to calculate evapotranspiration (Sentelhas et al., 2010; Chen et al., 2005). Our modification includes the dynamic calculation of vegetational

[1] Multi-objective complex evolution global optimization method.
[2] Shuffled Complex Evolution – University of Arizona.

parameters by means of the growing season index (Förster et al., 2012). An alternative to Penman-Monteith is e.g. the equation of Priestley and Taylor (1972). Other methods are based on empirical approaches like Thornthwaite (1948) or Turc (1961).

Snow accumulation and melting (as snow-water-equivalent) is calculated by energy balance approaches (Förster et al., 2014).

Soil types were parameterised by texture information, which are derived from soil maps (Edt.: J. Boess et al., 2004; Kreye and Meon, 2016). PANTA RHEI provides a physically based soil model (Kreye et al., 2010, 2012; Kreye, 2015). The soil model DYVESOM (dynamic vegetation soil model) accounts for root growths and vegetational feedback on soil properties by using the growing season index (Jolly et al., 2005). This parameterisation was successfully validated by comparing its course of year with phenological observation data of the German Weather Service and satellite based measurements, in particular the normalised differenced vegetation index (NDVI), see Förster et al. (2012). Simulated variables of DYVESOM, like soil water content or groundwater recharge, were successfully validated by different products of satellite based soil water measurements and by groundwater level measurements. Details of the parameterisation of DYVESOM are given in Kreye and Meon (2016).

PANTA RHEI is a state-of-the-art hydrological model, which can be adopted over a large bandwidth of spatial and temporal scales (Kreye, 2015; Kreye and Meon, 2016). It provides an API[3] to apply individual optimisation methods. Alternative models are e.g. WaSim-ETH (Schulla, 2012) or the MHM model (Samaniego et al., 2010).

The calculation of runoff concentration is based on four linear storages with individual storage constants: surface runoff, fast and slow interflow and base flow.

All used calibration parameters of PANTA RHEI are listed and explained in Table 1. The calibration parameters are valid for the sub-catchments and affect the processes in the HRUs.

In the evaluation of the final calibration results several quality criteria were used. The selection is based on common criteria used in hydrological science (Hall, 2001). E_{rel}, *E*, R^2, *RMSE*, *PBIAS* and *IoAd* were calculated in addition to E_{log}, which is the model efficiency based on logarithmic time series. E_{log} was additionally used as an objective function (see Section 2.3.2 "Lexicographic calibration strategy"). These criteria were calculated for both model areas and all time periods. *E* stands for model efficiency and E_{rel} for model efficiency with relative errors. R^2 is the coefficient of determination and *IoAd* the index of agreement. These criteria are dimensionless and have their optimum at a value of one. *PBIAS* gives the percentage deviation in total volume. *RMSE* is the root mean square error. More information regarding these criteria can be found in Legates and McCabe (1999), Hall (2001), Krause et al. (2005), Moriasi et al. (2007) or Dawson et al. (2007).

2.2. Study areas

The lexicographic calibration strategy was applied and tested in two physiogeographical different model areas at the mesoscale, which are located in northern Germany (Hellwege and Reckershausen). Hellwege is characterised by agricultural and livestock economy, while Reckershausen has a large proportion of arable land. The catchment Hellwege in the northern Lüneburger Heath[4] has a large proportion of sandy soils. In contrast, the catchment Reckershausen is characterised by fine-grained soils. Spatial data are available in high resolutions for both catchments. Further, we have high experience with the hydrological response in these areas due to several research projects.

Table 1
Properties of the model parameters of PANTA RHEI, which are used for the lexicographic calibration strategy and for SCE-UA.

Parameter	Description
O_s	Offset of infiltration function. This parameter shifts the function of infiltration capacity and therefore influences the amount of water that infiltrates into the first soil storage. The higher the value, the higher is the infiltration rate
D_r	Draining rate of the canopy storage. The draining rate influences the interception loss and therefore the volume of runoff. The higher the value, the faster the canopy storage drains, leading to more runoff
F_v	Factor that accounts for the influence of vegetation development on the soil properties. The higher the value, the more smoothed is the influence of vegetation (depending on the growing season index)
F_p	Factor that influences percolation. The higher the value, the more water percolates into deeper soil storages, leading to more base flow and less interflow
K_i	Linear storage constant of fast interflow. The higher the value, the steeper are the hydrographs of fast interflow
P_f	Factor that influences preferential flow in the soil and therefore accounts for the allocation between fast and slow interflow. The higher the value, the more preferential flow occurs

The relevant regional properties are summarised in Table 2. The meteorological input data (precipitation, average, minimum and maximum temperature, relative humidity, global radiation and wind speed) were available on a daily basis. The average processing time of a six-year simulation run for the catchment Hellwege is 41 min and for Reckershausen eight minutes on an i7-4930 @ 6 × 3.4 GHz.

2.3. Strategies for calibration of rainfall-runoff models

2.3.1. Multi-objective optimisation methods

An optimisation task can be posed in different ways. Either one single objective function is implemented for one criterion (in hydrology usually the catchment discharge) or various objective functions are implemented with one or several criteria. Different objective functions can be aggregated over random weights into a single objective function. Alternatively, the euclidean norm of the objective functions can be minimised (see Efstratiadis and Koutsoyiannis, 2010). Both options of these simultaneously optimised objective functions yield many results of pareto-optimal parameter sets, which have to be evaluated in the post-processing (Madsen and Khu, 2006). This leads to different options to select the "best" suitable parameter set (see Confesor and Whittaker, 2007), for which it is necessary to determine an order of preferences after the calibration. Consequently, the chosen "best" parameter set is always a compromise solution. In this context, broad definitions and summaries are given e.g. in Madsen (2000) or Boyle et al. (2003).

There is a general conflict between high resolution models, which need much calculation time, and complex global optimisation algorithms, which need many iterations (see chapter 1 introduction or Zhang et al., 2009 respectively Vaze et al., 2011). For this reason, we implemented time efficient local optimisation methods. Furthermore, a stepwise approach leads to a downsizing of the parameter search space and therefore allows the application of local optimisation methods.

The lexicographic calibration strategy can be implemented with different optimisation algorithms. We tested different methods, mainly direct search methods, where no information about the gradient of the objective function is required (e.g. downhill simplex method by Nelder and Mead (1965) or the pattern search method

[3] API – application programming interface.
[4] Lüneburger Heide.

Table 2
Properties of the two study areas Reckershausen and Hellwege.

Property	Reckershausen	Hellwege
Gauss-Krueger-coordinates of the gauge [m]	3564820/5697260	3513861/5882633
Area [km^2]	321	907
Main land use	61% arable land, 29% forest, 10% other	37% arable land, 32% pasture, 21% forest, 10% other
Soil properties	55% clayey loam, 45% sandy loam	80% pure and loamy sand, 20% organic soil
Height a.s.L [m]	186–470	12–121
MQ (1970–2000) [$m^3 s^{-1}$]	2.59	9.13
P (1970–2000) [mm]	785	837
T (1970–2000) [°C]	8.1	8.8
Number of HRU	294	1778
Spatial resolution [km^2/HRU]	1.06	0.51

(Audet and Dennis, 2002), which can be modified to the Sequential Line Search method, see Kuzmin et al. (2008). Since these two methods, as well as the additionally tested SCE-UA, delivered the same parameter sets for 10 different test cases, we selected the downhill simplex method for this study due to performance reasons. This algorithm converges reliably, provided that the surface of the search space is continuous and has an explicit minimum (Nelder and Mead, 1965). These requirements were often not valid, when this algorithm was applied for minimising the multi-objective function by optimising all parameters simultaneously. A simplex is a geometric construct, which consists of a set of n + 1 points in an n-dimensional parameter space. The smaller the dimension of the parameter space, the more reliable is the convergence of the downhill simplex method. If only one parameter has to be optimised, we used Brent's method (2002). This method is reliable for determining the minimum of a function in one dimension. The derivative-free parabolic interpolation is combined with the golden section search (Brent, 2002). Unlike the widely used local Levenberg-Marquardt algorithm (as for example implemented in PEST[5]), a computation of derivatives of the model is not necessary for the used methods of this study.

2.3.2. Lexicographic calibration strategy

From the mathematical point of view, an optimisation task is not the challenge, if the objective function is well formed and the search space has a definite minimum. But from the hydrologic point of view, there are many challenges that need to be taken into account when optimising parameters within a hydrologic model: The data (runoff observations and input data, especially rainfall) are uncertain. According to this fact, hydrologists demand robust methods in terms of insensitivity of the model against measurement errors. Next, the calculation time of large-scale models can be costly due to the high resolution. This leads to optimisation approaches, which are fast and thus requires less iteration steps. The last requirement refers to the transferability of the resulting parameter sets. For these reasons the intention of this study was to define a strategy for calibration of hydrological models with the criteria (i) transferability of the parameter set (ii) goodness-of-fit in calibration and validation and (iii) time-efficiency.

The developed lexicographic strategy determines the preference order of the objectives prior to the calibration. This is in contrast to traditional multi-objective optimisation, in which the most appropriate parameter set has to be identified by the experts afterwards. The individual prioritisation can be compared to the specific determination of the weights to aggregate the multi-objective goals to a single objective function in the multi-objective optimisation (see Efstratiadis and Koutsoyiannis, 2010). Hydrological knowledge is needed for an appropriate division of the parameters into smaller sub-groups within the lexicographical calibration. Furthermore, it is important to identify suitable objective functions for all parameter sub-groups individually.

There are many possibilities for the quantification of cause-effect relationships of model parameters and their associated processes by means of one or more target functions. The assignment sequence from model parameters over hydrological processes to an objective function is often undetermined, since the model parameters are rarely associated with only one single process (Kreye, 2015; Gelleszun et al., 2015). Despite that, the effects of complex processes can hardly be evaluated within one single objective function. The presented lexicographical approach circumvents the assignment sequence by developing suitable objective functions for model parameters directly. Based on Reusser et al. (2009) and Reusser et al. (2011), Reusser and Zehe (2011) analysed the temporal course of parameter sensitivities and the accompanying dominant error types. They concluded that the effects of parameters are both time- and state-dependent, while still allowing an aggregation of the parameters by their actions. Since objective functions evaluate various aspects of the time series differently, the definition of direct assignments from model parameter to objective function is consequential. This approach needs particular adjustments in dependence of the used hydrological model. Once an assignment is specified, various questions can be processed by setting an appropriate order of preference.

A hydrologist who wants to adjust parameters in a hydrological model has knowledge about his model in the way, that he can assess which parameter has specific influence on the modelling result respectively on different objective functions. In general, expert knowledge (=defining a suitable order of preferences prior to the calibration) can be compared to the selection of the "best" parameter set of a pareto optimum after a global multi-objective calibration. A change of the preference order would lead to different parameter sets, but they would be still representative concerning the determined objective function(s). Our aim was to improve the a priori subjectivity with the best possible objectivity via a stepwise calibration process.

In the context of hydrological model calibration the term "step-by-step calibration" is not defined clearly. In some cases it is defined as manual stepwise calibration, other studies (e.g. Hay and Umemoto, 2007) describe their step-by-step approach as a stepwise circle (starting with step 1, then 2, 3, n and then step 1 again). Alternatively, Ning et al. (2015) described a stepwise calibration method, but did not change the objective function within the different time steps. In contrast to these definitions, we use different objective functions in every calibration step and these functions are associated with model parameters that have sensitive influence. Further, the preference order, the objective functions and the parameters all can vary due to the overall scope of the project or application and scientific framework.

For this study, we defined the order of preference as presented in the following enumeration. The overall objective was a reliable

[5] Parameter estimation, Doherty (1994).

reproduction of the observed discharge at the gauge with its following criteria:

1. Volume and runoff peaks
2. Seasonality and low flow values
3. Shape of the hydrograph

Prior to the calibration, we performed sensitivity analyses of all parameters in order to assign each parameter to an objective function. Therefore we performed Latin hypercube samplings for a small test catchment. Next to the influence of each parameter on the simulated runoff, we analysed the direct influence on several objective functions. Descriptions of the resulting objective functions for each step are given in Table 3.

We generally recommend focusing on the volume as a first step. In many hydrological models, more than one model parameter has an influence on the simulated volume. In PANTA RHEI, we identified two model parameters (D_r, O_s). However, these parameters additionally influence the simulated peaks. Therefore, we chose the combination of an absolute error between simulated and observed peaks and an RMSE between logarithmic observed and simulated time series. The peaks were determined by means of time series analysis (identification of local maxima) based on the observation. Two model parameters (F_p, F_v) were identified to have high impact on the simulation at low values. Hence, in the second step we developed an objective function accounting for values that are smaller than the 40% percentile of the (observed) duration curve of non-exceedance. The last step of the lexicographic calibration strategy concentrates on an overall fitting of the shape of the simulated hydrograph by using the model efficiency. Values of observed hydrographs in moderate climate zones often follow a logarithmic Gaussian distribution (Bowers et al., 2012). Furthermore, we avoid focussing on high flow values solely. Therefore, logarithmic time series of observation and simulation were used to calculate the model efficiency (see Krause et al., 2005). The parameter P_f and K_i were identified to inherit particular influence on this objective function. From these analyses, we also developed one multi-objective function, which was used in the optimisation with the global method SCE-UA (see Section 2.4 Calibration scheme).

2.4. Calibration scheme

Parts of the time series of observed runoff from 01.11.2001 to 31.10.2011 were used to calibrate the hydrological model PANTA RHEI for both catchments. All simulations were started four years before the actual calibration periods began in order to minimize effects of initial conditions within the hydrological model. The total time series (Pcontrol, see Table 4) was split into five sub-periods of two years each (P1-P5, see Table 4).

The developed lexicographic calibration strategy was used to calibrate all six periods independently. The global optimisation algorithm SCE-UA (Duan et al., 1992) served as a reference calibration. We selected the SCE-UA algorithm as reference since it is an established approach in hydrological research and it was often proofed that SCE-UA performs very well to identify the global minimum of a multi-criteria objective function (see e.g. Duan et al., 1994). Twelve calibrations were performed in total (6x LCS and 6x SCE).

2.5. Uncertainty analysis

Established methods for estimating model uncertainties in hydrological rainfall-runoff models are often based on stochastic approaches like Monte Carlo methods (Ajami et al., 2007). Probability densities of the parameters and/or probabilities of occurrence of model predictions are based on random experiments (see. Freer et al., 1996; Beven and Freer, 2001). For applications

Table 3
Order of preference with the corresponding calibration steps, associated parameters of PANTA RHEI and applied objective functions.

	Step	Parameter	Objective function, description
Volume	1	D_r, O_s	Absolute error between observed and simulated peaks multiplied with root mean square error (RMSE) between logarithmic observed and simulated time series
Seasonality	2	F_p, F_v	RMSE between low flow values of the observed and simulated time series. The low flow values are determined in dependence of the 40% percentil of the (observed) duration curve of non exceedance
Shape	3	P_f, K_i	Model efficiency E_{log} of the logarithmic observed and simulated time series

Table 4
Periods of simulations and calibrations.

Label	Start of simulation	Period of calibration	End of simulation
P1	01.11.1997	01.11.2001–31.10.2003	31.10.2003
P2	01.11.1999	01.11.2003–31.10.2005	31.10.2005
P3	01.11.2001	01.11.2005–31.10.2007	31.10.2007
P4	01.11.2003	01.11.2007–31.10.2009	31.10.2009
P5	01.11.2005	01.11.2009–31.10.2011	31.10.2011
Pcontrol	01.11.1997	01.11.2001–31.10.2011	31.10.2011

with high-resolution models, these methods are unsuitable, due to enormous computational effort. Based on this premise, we applied approximate methods for determining the uncertainty. Kunstmann et al. (2002) showed that the approximated uncertainty intervals for simulated groundwater levels with an enhanced FOSM method (first-order second-moment) achieved good matches with those obtained by the time-consuming Monte Carlo method.

Hydrological modelling entails various sources of uncertainty, in this study we focused on the following uncertainties:

- Parameter uncertainty in relation to the chosen optimisation algorithm (empirical variance estimator, see Eq. (3)).
- Confidence intervals of the model with respect to the uncertainty of the parameters (variance propagation, see Eqs. (1) and (2)).
- Empirical standard deviation between observed and simulated runoff.

In the final LCS optimisation step (see Table 3), we maximised the model efficiency of the logarithmised observed and simulated runoff time series. This determination was made to increase the weight of the mean flow in the parameter estimation. Accordingly, we quantified the uncertainties by using the logarithmic runoff data. As reverse operation, we exponentiated the calculated confidence and prediction intervals.

The FOSM-method is based on the variance-covariance propagation, given by Eq. (1) (see Witte and Schmidt, 2004, page 149):

$$C_{yy} = AC_{xx}A^T \tag{1}$$

The variances and co-variances of the parameters are given by the co-variance matrix C_{xx} ($n \times n$). The matrix A ($m \times n$) is denoted as Jacobian, sensitivity-, or functional-matrix. This matrix contains the partial derivations of the model with respect to its parameters. The co-variance matrix C_{yy} ($m \times m$) of the calculated random variable y (in this case the simulated runoff) gives the variances of y on the diagonal. These variances can be calculated directly by Eq. (2):

$$var(y) = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}a_{ik}c_{jk} \qquad \begin{matrix} a_{ij} : elements\ of\ A \\ c_{ij} : elements\ of\ C_{xx} \end{matrix} \tag{2}$$

Similar to Monte Carlo simulations, the variances of the parameters are needed a priori. In most cases, the estimation of the variances is challenging. For this reason, we calculated the empirical co-variance matrix by means of Eq. (3) (see Witte and Schmidt, 2004, page 154). The result is then used in Eq. (1). In general, Eq. (3) is valid for linear models and therefore provides approximated values for C_{xx}:

$$C_{xx} = s_e^2 (A^T A)^{-1} \tag{3}$$

The empirical standard deviation s_e^2 is a scalar value, which can be obtained from the entire simulation period.

Commonly, a hydrological model cannot be derived analytically by its parameters. Consequently, we calculated the Jacobian matrix A by numerical derivation in the optimum (see Maskey and Guinot, 2003). We used central differences as an approximation of the derivatives. Each parameter was changed by ±1% of the optimum parameter value. The obtained two time-series of simulated runoff were subtracted for every time step and divided by 2% of the corresponding parameter value. The sensitivity matrix was not only used in the above equations, but also to determine dimensional quantities. Therefore, we calculated the one-percent scaled sensitivity by multiplying the columns of the sensitivity matrix with the corresponding parameter values (Hill, 1998) and divided them by 100. The dimension describes the variation of the simulated result if the parameter is changed by one percent (Hill, 1998).

2.6. Reliable parameter estimation

Hydrologists interpret the terms robust, stable and representative in different ways. Here, we distinguished between robust parameter estimation, well-posed optimisation problems, including stability, and representative parameter sets.

The idea of robust parameter estimation is the insensitivity of the resulting parameter-vector to measurement errors (Klute et al., 1994; Bárdossy and Singh, 2008). Bárdossy and Singh (2008) showed that erroneous discharge data have a significant effect on the resulting parameter set, when using a global optimisation method with one objective function in combination with a conceptual model. They used the half space depth method to circumvent this problem. The method is promising, but due to the complexity of physically based models and its concomitant calculation time, it is not applicable to common practice. In Gelleszun et al. (2015), we showed that the lexicographic approach leads to similar parameter sets, when adding a synthetic noise on the observed discharge data. This led us to the conclusion that LCS induces robust parameter sets.

The three most important criteria to evaluate the quality of an inverse problem are identifiability, uniqueness and stability (Carrera and Neuman, 1986). A parameter set is identifiable if different parameter sets produce different results. In this context, "results" refer exclusively to the direct calculations within the model. In contrast, the uniqueness considers the inverse problem. An inverse problem is precisely unique if the considered objective function has a defined minimum. The last criterion is the stability of the optimal parameter set. It is an indication of the robustness of a solution in the optimum. Small differences in this stable solution do not lead to significant differences in the results (Carrera and Neuman, 1986). The inverse problem is well-posed if the obtained parameter set fulfils all three conditions. We performed a sensitivity analysis and a collinearity analysis as sufficient condition for the identifiability. Since uniqueness is dependent on the objective function(s), it is obvious, that a single aggregated objective function leads to many "good" results. We circumvented this issue by using multiple objective functions in a stepwise approach, which has a defined minimum each. The stability of the final parameter set was proved by calculating the condition numbers of each of the Hessian matrixes.

3. Results

3.1. Lexicographic calibration results

In this section, we concentrate on results of the (individual) calibrations for the five different two-year periods and the total period (2001–2011).

High goodness-of-fit criteria regarding the simulated discharges in comparison to the observations were achieved for the total period 2001–2011, as well as for the five periods of two years each in both model areas (Reckershausen, Hellwege). Established quality criteria, based on simulated and observed discharge time series, are given in Table 5. Following these quality criteria, the (individual) calibration results mostly reach top scores. In Reckershausen, the time period 2005–2007 shows weaknesses regarding high flow (moderate values *of E* and R^2). However, the volume (*RMSE*, *PBIAS*) and mean/low flow (E_{log}, E_{rel}) are simulated reasonable. A similar constellation occurs for the period 2009–2011 in the model area Hellwege.

The observed and LCS-simulated discharge time series, as well as the prediction intervals (see Eqs. (2) and (3)) for the model area Reckershausen, are shown in Fig. 1. At the top, the calibration of the total period is visualised and the picture in the middle shows the calibration of the two-year period from 2007 to 2009. Both for the calibration of the total and of the two-year period, the simulated discharges are highly correlated with the observations. The a priori defined criteria of the preference order (volume, seasonality and shape) were fulfilled successfully. The qualities of the simulations are particularly good for low and mean flow conditions. The calibration results for Reckershausen based on the SCE-UA algorithm are shown for the period 2007 to 2009 in the graphic at the bottom of Fig. 1. The simulated time series is very close to the observed time series and the quality criteria are only marginally higher as for the lexicographic calibration.

For the model area Hellwege, the volume was slightly overestimated for the lexicographical calibration of the total period, see Fig. 2 (top), while low, mean and high flow values are met with high performance. The calibration of the two-year period from 2007 to 2009 of Hellwege is visualised in Fig. 2 (middle). Both, the time series and cumulative time series of the simulation reach the observation with very high accuracy. The shown prediction intervals are calculated by means of Eqs. (2) and (3). The relatively narrow bandwidth of the 68.3%-prediction interval results from a small empirical standard deviation caused by the large amount of available time series data. Due to our prioritisation to assign more weight to the low and mean flow, the prediction interval is smaller in those phases. The calibration results for Hellwege based on the SCE-UA algorithm are shown for the period 2007 to 2009 in the graphic at the bottom of Fig. 2. The volume was not fitted perfectly by SCE-UA, because high flow events were a bit underestimated, but the shape of the simulated time series looks very reasonable in comparison to the observation.

3.2. Validation of the lexicographic calibration and comparison with the SCE-UA calibration

We investigated the validation performances of the model parameter sets, which resulted from the different calibrations (see Section 3.1 "Lexicographic calibration results"). In addition to the lexicographic calibrations, we elaborated calibrations by means of the SCE-UA algorithms. The results of these two approaches were compared.

Every two-year period was calibrated individually (see Table 4). Hence, five different sets of model parameters were determined for LCS and for SCE-UA. The general hydrological response of the catchments did not change relevantly over the total (ten-year) per-

Table 5
Quality parameters of calibration results obtained by LCS for the total period (2001–2011) and the five individual two-year periods of the study areas Reckershausen and Hellwege. E_{log} represents the model efficiency based on logarithmic time series, E_{rel} the relative and E the standard model efficiency. R^2 is the coefficient of determination, RMSE the root mean square error, PBIAS the percentage volume error and IoAd the index of agreement. All quality criteria without units have their optimum at a value of one, the other two (PBIAS, RMSE) have their optimum at a value of zero.

		E_{log} [−]	E_{rel} [−]	E [−]	R^2 [-]	RMSE[$m^3 s^{-1}$]	PBIAS [%]	IoAd [−]
Reckershausen	2001–2011	0.92	0.96	0.87	0.87	0.91	−1.69	0.96
	2001–2003	0.94	0.96	0.90	0.90	1.12	−1.74	0.97
	2003–2005	0.93	0.94	0.88	0.89	0.55	−2.99	0.97
	2005–2007	0.86	0.92	0.71	0.72	0.97	0.66	0.90
	2007–2009	0.96	0.97	0.90	0.90	0.66	1.77	0.97
	2009–2011	0.95	0.97	0.91	0.91	0.91	−1.16	0.98
Hellwege	2001–2011	0.88	0.89	0.85	0.88	3.34	−5.61	0.97
	2001–2003	0.91	0.90	0.90	0.91	3.72	−2.56	0.98
	2003–2005	0.82	0.82	0.78	0.86	2.26	−5.54	0.95
	2005–2007	0.84	0.86	0.81	0.85	3.08	−3.31	0.96
	2007–2009	0.95	0.96	0.94	0.95	2.54	−0.95	0.99
	2009–2011	0.82	0.81	0.71	0.76	3.91	−4.94	0.93

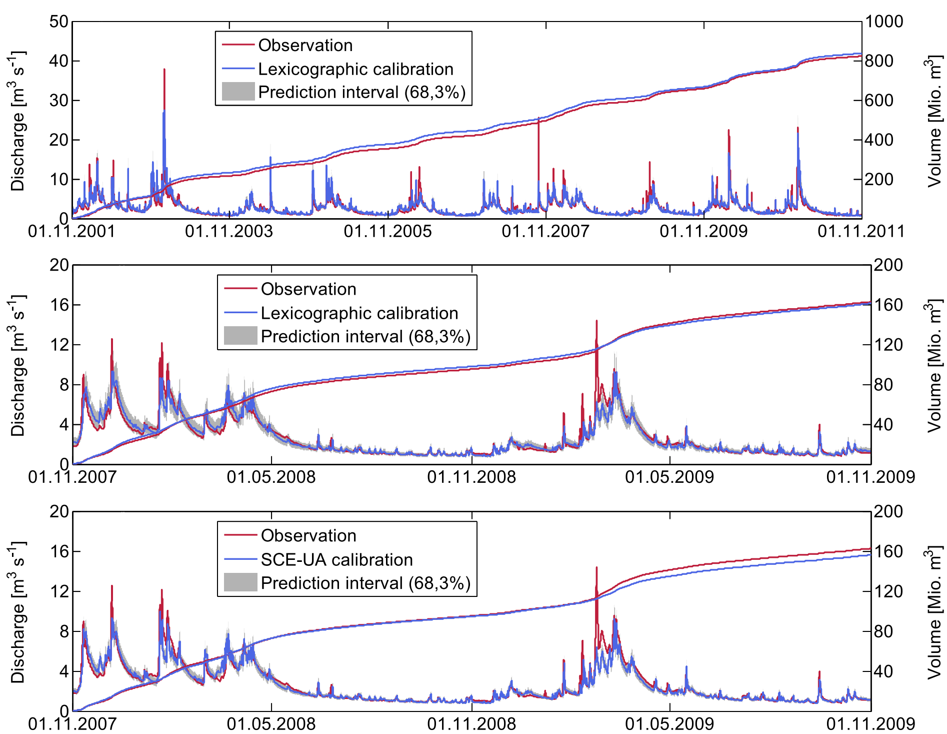

Fig. 1. Reckershausen: Discharge time series and cumulative time series of observation and lexicographic calibration, as well as the corresponding prediction interval for the total period (2001–2011, top graphic) and for the two-year period from 2007 to 2009 (middle graphic). The graphic below shows the result of the calibration with SCE-UA from 2007 to 2009.

iod (H÷lscher et al., 2012; Wurpts et al., 2014). Therefore, the model parameters, which resulted from the five different calibrations periods, are expected to be similar in their numeric values and in their effects on hydrological simulation results. Fig. 3 shows box-whisker plots based on the bandwidth of each parameter with different colours for the two calibration approaches (LCS, SCE-UA). In the study area Reckershausen (left hand side of Fig. 3), the ranges of the parameters O_s, F_p and P_f are significantly smaller for the lexicographic calibration strategy, compared to the calibration with SCE-UA. In the model area Hellwege (right hand side), the ranges of the parameters D_r and K_i are much smaller and the ranges of the parameters O_s and F_p are slightly smaller for the lexicographic calibration strategy. Out of these findings, the following statements can be pointed out:

- In general, the ranges of the parameters, based on the lexicographic calibration strategy, are smaller than the ranges based on calibration with SCE-UA.
- Not all parameters show smaller ranges, approx. 50% are smaller and 50% are more or less the same, when comparing the lexicographic calibration with SCE-UA.
- For different study areas, different parameters show smaller ranges. Due to different physiographical structure and hydrological behaviour of the catchment, the model parameters have various influences on simulation results. The higher the influence of the specific parameters, the smaller the (lexicographic) parameter ranges (see Chapter 3.3). For each calibration step, given by the order of preferences, the range of the dominant parameter (at this point) is very narrow. Hence, the

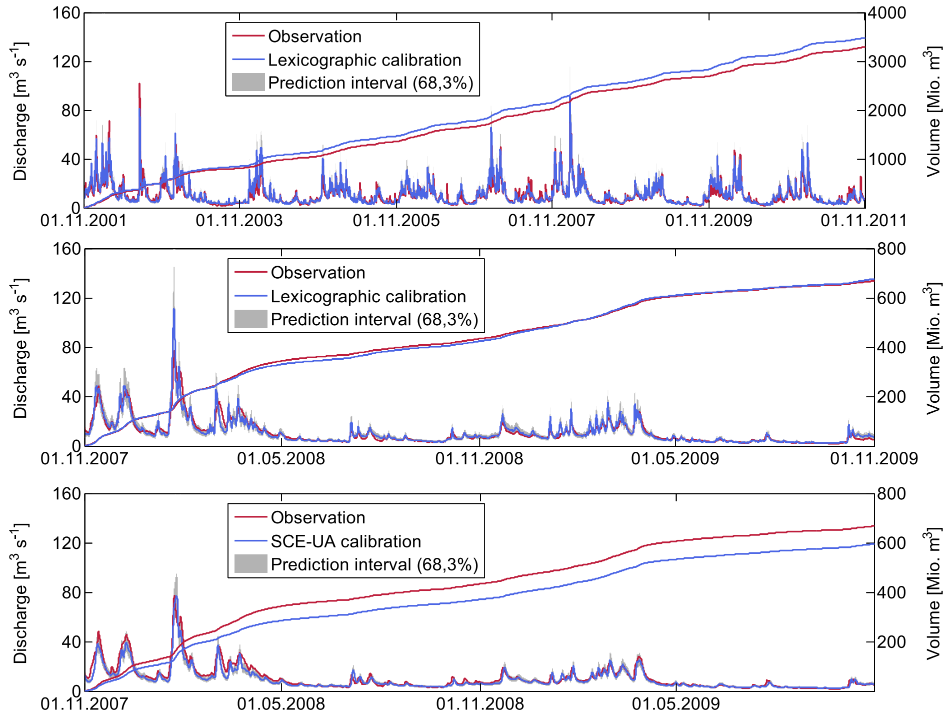

Fig. 2. Hellwege: Discharge time series and cumulative time series of observation and lexicographic calibration, as well as the corresponding prediction interval for the total period (2001–2011) and for the two-year period from 2007 to 2009. The graphic below shows the result of the calibration with SCE-UA from 2007 to 2009.

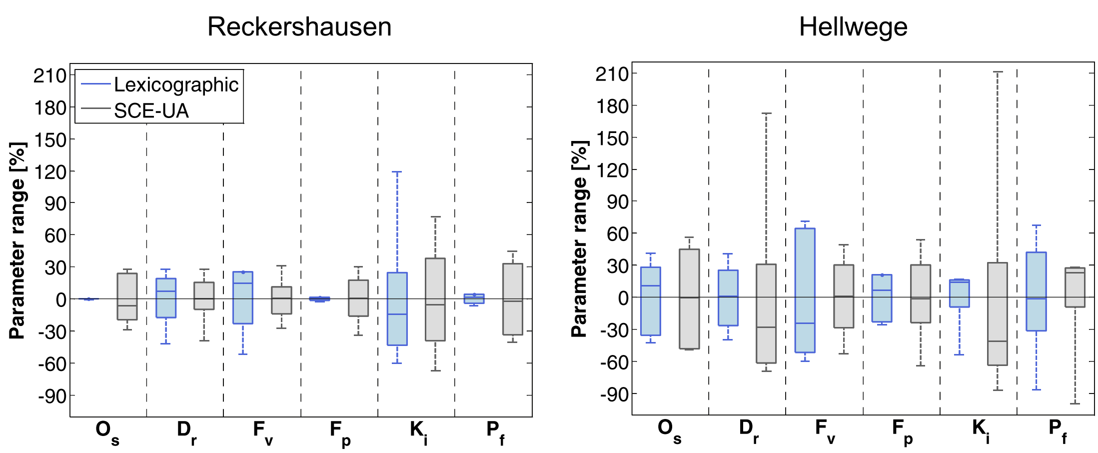

Fig. 3. Box-whisker plots for parameter ranges based on the five different two-year calibration periods. A parameter range of 0% means that all parameters have the same value. Results of the lexicographic calibration are shown in blue colour and results of calibration with SCE-UA are shown in grey colour. Left: Reckershausen, Right: Hellwege.

lexicographic calibration strategy increases the reliability of resulting parameters based on their specific influence on simulation results.

To further investigate the representativeness of the parameter sets, we performed a cross-validation. All five two-year parameter sets were applied to each of the other four time periods. In addition, the total time period of ten years was simulated with each of the five parameter sets. We performed this validation with the parameter sets obtained by the lexicographic calibration strategy, as well as with the parameter sets from the SCE-UA. In total, 60 (30 for each calibration approach) values of quality criteria (e.g. E_{log}) were calculated in this manner for every study area.

The resulting values of E_{log} are visualised in Fig. 4. Blue colour represents the lexicographic calibration strategy and grey colour the calibration with SCE-UA. Results of the catchment Reckershausen are shown on the left hand side and results of the catchment Hellwege on the right hand side of Fig. 4. The parameter sets, that show the valid calibration for the current time periods are each marked with a black cross. The rectangles without a black cross show the validation results. Hence, we obtained four validation-values of E_{log} for each time period and one value of E_{log}, which resulted from the calibration. In addition to these five values of E_{log} per time period, an average value was calculated arithmetically. The following statements were distinguished, based on the findings presented in Fig. 4:

- Both approaches (LCS, SCE-UA) achieved E_{log} values of high quality for calibration (rectangles marked with black crosses) and validation (rectangles without crosses).

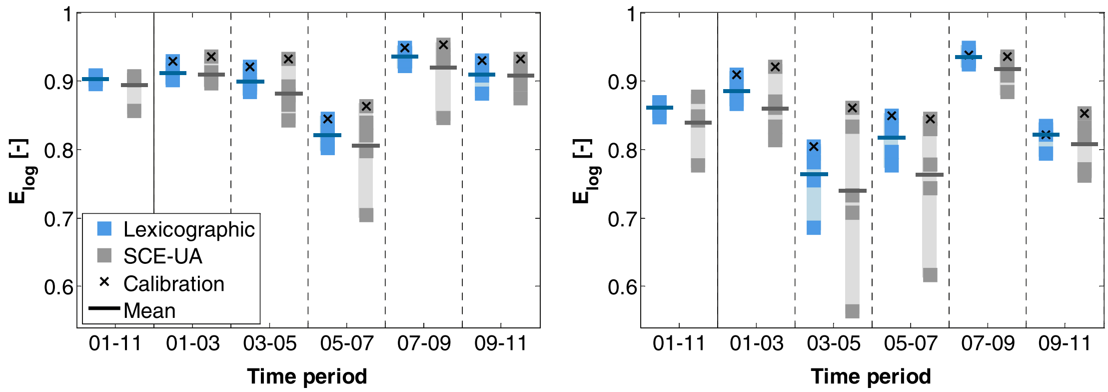

Fig. 4. Cross-validation of calibration results. Each parameter set was applied on each time period. Leading to overall five values of E_{log} per time period and per calibration strategy (blue colour: lexicographic, grey colour: SCE-UA). The values that are valid for the calibration of the current time periods are marked with a black cross, the other four values per time period show the validation results. The mean values are shown as horizontal lines. Left: Reckershausen, Right: Hellwege. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- Regarding calibration solely, SCE-UA achieved higher values of E_{log} than the lexicographic calibration. This behaviour was expected, since SCE-UA is a global optimisation algorithm and the lexicographic uses a stepwise local algorithm.
- The lexicographic calibration parameter sets achieved higher validation values for the two-year-validation periods and for the total period.
- The range of E_{log} values of the validations based on the lexicographic approach is much narrower than the range based on SCE-UA.

It can be summarised that the validation results, as well as the similarity of the parameter sets by the lexicographic calibration strategy, is higher than the results based on SCE-UA. These findings indicate the high representativeness of the parameter sets determined by the lexicographic calibration strategy.

The effects of the different parameterisations on components of the water balance are shown in Fig. 5. Each of the model parameter sets, which resulted from the five two-year calibrations, was used to simulate the total period of ten years. Hence, these two times five validation runs determine different water balances. We investigated mean yearly effects on the discharge, evapotranspiration, groundwater recharge and soil water content. As shown in Fig. 5, the variations in all of these components are much higher for the SCE-UA simulation runs compared to the variations of the runs based on the lexicographical calibration strategy. This indicates that the lexicographical calibration yielded in parameter sets of higher robustness. To explain this statement, we took a closer look to the groundwater recharge. Based on the SCE-UA validation runs, the yearly groundwater recharge varied between 78% to 120% for Reckershausen and between 88% to 135% for Hellwege. In other words: The highest value is more than 1.5 times the size of the smallest value (for the same area and time). The validation runs based on the lexicographical calibration produced a variation in groundwater recharge from 96% to 105% for Reckershausen and 95% to 107% for Hellwege. Here, the highest value is only increased by a factor of 1.09 or 1.15 of the smallest one. Similar constellations arised for the other components of the water balance (Fig. 5). Hence, the lexicographical calibration delivers parameter sets, which have similar effects on the hydrological system.

3.3. Sensitivity analysis

For all parameter sets (five LCS and five SCE-UA with six parameters each) we performed a sensitivity analysis for the overall ten-year validation period. For this, we used the one-percent scaled sensitivity, which has the advantage of the dimensions of the parameters being irrelevant. Fig. 6 shows the scaled sensitivities for the catchment Reckershausen in dependence of the time (shown here: 2007–2009). The six parameters are visualised by different colours and for each parameter five lines are drawn, which were derived from the five different calibration sets. Results based on the lexicographic calibration strategy are shown at the top of Fig. 6 and results based on SCE-UA are shown at the bottom. First to mention is the fact that the five different sets of the lexicographic calibration produced much more similar pictures of sensitivities for each of the parameters than the sets of the SCE-UA calibration: The coloured five lines (per parameter) of LCS have a very close range. This behaviour emphasises the statement that the lexicographic calibration strategy determines highly representative and robust parameters. The hydrological effects of these parameters are supposed to be similar, which lies in agreement to Fig. 5. The sensitivities in dependence of the five sets of the SCE-UA calibration on the other hand, have relatively high differences to each other. This leads to parameterisations of the hydrological system, which are potentially in disagreement to each other. Fig. 7 shows the scaled sensitivities of the catchment Hellwege with an equivalent structure to Fig. 6. For Hellwege, a similar constellation arose as for Reckershausen, but less distinctive.

Supplementary to these findings, it is noteworthy that the parameters with particularly high sensitivity (see Figs. 6 and 7) have very narrow ranges in their numeric values (see Fig. 3). For Reckershausen these parameters are O_s, F_p and P_f. In Hellwege we identified F_p and K_i. From a physical point of view this is entirely sensible: Those parameters that have a high influence on the result should have numerical values of the same magnitude. This again emphasises the reliability of the parameters determined with the lexicographic calibration strategy.

4. Discussion

We demonstrated that the lexicographic calibration strategy (LCS) delivered suitable parameter sets for the associated hydrologic system under the premises of robustness and minimal calculation effort. The stated constraint of limited calculation time originated from the complexity of the physically based hydrological model. In contrast to Zhang et al. (2016), where a HPC was used to circumvent this imbalance, we applied a lexicographic calibration strategy in combination with a local optimisation algorithm. The strength of LCS is an integration of expert knowledge from the very beginning of work (in the way of defining an appropriate order of preference) while keeping objectivity by using automatic

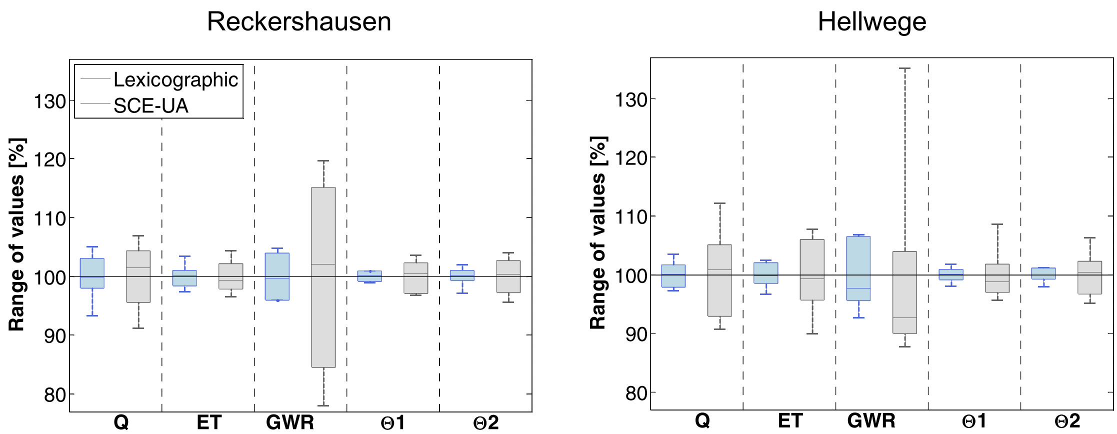

Fig. 5. Box-whisker plots of the effects on the main components of the water balance induced by the five different model parameter sets in the study areas Reckershausen (left) and Hellwege (right). The ten-year validation period was used as time frame of the simulations. Blue colour shows results of the lexicographic calibration strategy and grey colour the SCE-UA calibration. Q symbolizes discharge, ET evapotranspiration, GWR groundwater recharge, Θ_1 the upper and Θ_2 the deep soil water content. The ranges of the values are calculated as deviations of mean yearly values induced by the different model parameterisations.

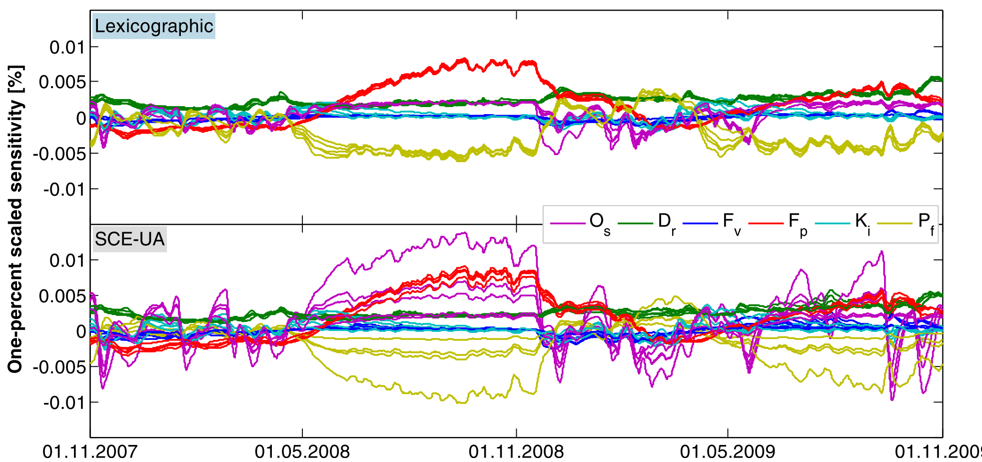

Fig. 6. One-percent scaled sensitivity of gauge Reckershausen for all six parameters for each of the five parameter sets (moving average over seven days). The one-percent scaled sensitivity is calculated by means of the sensitivity matrix with respect to the individual parameter values. If the value of a parameter is changed by one percent according to the individual optimum, the scaled sensitivity (each of the printed lines) visualizes the influence on the result (here simulated runoff).

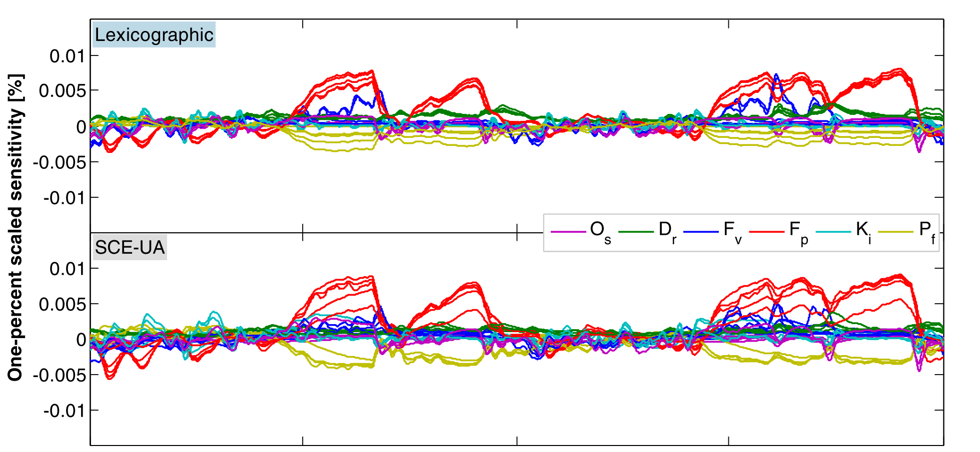

Fig. 7. One-percent scaled sensitivity of gauge Hellwege for all six parameters for each of the five parameter sets (moving average over seven days). The one-percent scaled sensitivity is calculated by means of the sensitivity matrix with respect to the individual parameter values. If the value of a parameter is changed by one percent according to the individual optimum, the scaled sensitivity (each of the printed lines) visualizes the influence on the result (here simulated runoff).

optimisation algorithms. We used five different time periods from one and the same gauged catchment(s). It was shown that the lexicographical calibration strategy lead to robust and representative parameter sets, when applied to the different validation periods. The results of LCS were additionally compared to the results obtained by the global optimisation method SCE-UA. We determined different criteria to evaluate the quality of the parameter estimation: These criteria took different aspects of the calibration process, the resulting parameters and the resulting simulations into account. Starting with goodness-of-fit criteria, as expected, the global optimisation algorithm SCE-UA achieved mathematical better results of the objective function than LCS during the calibration period. At the same time, the lexicographic calibration strategy averagely reached better scores for the four validation periods (see Fig. 4). Hence, the obtained lexicographic parameter sets were very robust when cross-validating with different time periods. We thus conclude that the transferability of the parameter sets for different purposes, such as planning tasks or operational flood forecast, is more suitable, and the parameters themselves are more reliable and representative. Furthermore, the five LCS parameter sets were comparable to each other, despite referring to different time periods. The general hydrological response for the catchment had not changed significantly over the considered ten years (Hölscher et al., 2012; Hölscher et al., 2014; Wurpts et al., 2014). Therefore it is plausible and remarkable, that the different LCS parameter sets were similar in their numeric values, their scores in the validation period and in their effect on hydrological simulation results such as discharge, evapotranspiration, groundwater recharge or soil moisture (see Fig. 5). It thus can be concluded, that the method is promising in regard to regionalisation approaches.

Measurements of water levels and discharges, as well as corresponding rating curves, cause uncertainties in observed data. This circumstance leads to the idea of robust parameter estimation. We distinguished between robust parameter estimation, a well-posed optimisation problem and representative resulting parameter vectors. All these aspects were considered in our calibration strategy. The stepwise approach allows minimizing the parameter search space connected with time-saving local optimisation methods instead of global optimisation methods. We considered various aspects of uncertainty, while focusing on uncertain model parameters due to the chosen optimisation algorithm. We applied the approximate first-order second-moment method to determine the confidence intervals of the estimated parameters, the corresponding confidence intervals of the simulated runoff, as well as the prediction intervals, which includes the standard deviation between observed and simulated runoff. The obtained parameter uncertainty could be transferred to the simulated runoff time series by using the variance–covariance propagation law. The calculated condition numbers of the Hessian were less than 100 for all LCS results. This demonstrates the stability of the lexicographic results. For the parameter sets obtained by SCE-UA, the value of the condition numbers were for two sets higher than 100 (120 and 117). These findings are emphasising the results of the one-percent scaled sensitivity (see Figs. 6 and 7), where we showed that the influences on the simulation results are highly depended on the based parameter sets. Regarding the effect on hydrological processes and therefore on water balance components, LCS delivered much more plausible and robust parameter sets than the tested global optimisation method.

By separating the parameters into groups and with a stepwise calibration, approximately 6% of the calculation time was needed, compared to the global optimisation method SCE-UA. One iteration run with our model (PANTA RHEI) and a standard desktop pc (i7 6×3.4 GHz) needs approx. 40 min for a medium size catchment like Hellwege. A global optimisation method needs approx. 1000 iterations, which results in more than 4 days of computing time for only one gauge. Usually many gauges have to be considered. In the Aller-Leine-Oker catchment, in which Reckershausen is located, more than 150 gauges (with different area sizes) are available.

The parameters of hydrological models are usually not quantifiable with field measurements or experiments. This challenge leads to fuzzy parameter definitions with the difficulty of clear determinations of the parameters with their blurry influence on model results. The clustering of parameters into independent groups with specific influence on the hydrograph is not straightforward feasible. This drawback limits the application of the presented LCS. Nevertheless, it is possible to group the parameters according to their main influence on the hydrograph (or other model result). In addition, we suggest already considering the calibration strategy, when developing the hydrological model, especially with respect to the calibration parameters.

The application of the lexicographic calibration strategy demands subjective adjustments, in the way, that the order of preferences needs to be determined. The order of preference has a deductive influence on the resulting parameter set, but usually this order is framed by the scientific question or planning task. The most important fact is that the results are reproducible, if the identic order is chosen. In general, the subjective definition of the "correct" order of preferences can be considered to be equal to finding the "best" parameter set of a pareto optimum after a global multi-objective calibration. The individual freedom within the LCS is considered as strength, because this "subjectivity" is taken into account from the beginning and not at the end. This reduces the search space from the beginning and focuses on the individual demands. In several projects, we had different demands on the model and thus on the parameters, e.g.: for the reproduction of flood statistics the most important criterion was the correct representation of high water events (Hölscher et al., 2012). The second preference was the long-time water balance and the last one referred to low water. Another project aimed in analysing the low flow events in the context of climate change (Hölscher et al., 2014). Obviously, the first preference was the correct representation of low flows, which we achieved by using the NMQ7D (minimum low flow of the moving average of 7 days) as criterion within the objective function. The second preference referred to the volumetric soil water content (fitting to e.g. satellite based soil moisture data) and the last criteria was a balanced long-term water budget. In a recent BMBF-project[6], we had to deliver resilient recommendations concerning the possible development of the water supply, which can be used to meet the future requirements of water resources management of Germany's coastal zones. Scenario-results from the project are directly used for decision-making by the local water suppliers. It is advantageous that the lexicographic calibration strategy results in only a single parameter set, as it is required. This is especially important, since parameters were used for simulations with numerous climate scenarios.

Acknowledgement

The authors would like to thank the editor and the anonymous reviewers for their valuable comments to improve the quality of the paper. The project NAWAK ("Development of sustainable adaptation strategies for the infrastructure of water resources under the conditions of climatic and demographic change", project duration 01.07.2013 to 31.01.2017) is funded by the German Federal Ministry for Education and Research (BMBF) in the framework of INIS (Funding code 033W007D).

[6] Project title: NAWAK ("Development of sustainable adaptation strategies for the infrastructure of water resources under the conditions of climatic and demographic change"), BMBF: German Federal Ministry for Education and Research.

References

Abbaspour, K., Calibration of hydrologic models: when is a model calibrated?, MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, 2005.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43. http://dx.doi.org/10.1029/2005WR004745.

Audet, C., Dennis, J.E., 2002. Analysis of generalized pattern searches. SIAM J. Optim. 13, 889–903. http://dx.doi.org/10.1137/S1052623400378742.

Bárdossy, A., Singh, S.K., 2008. Robust estimation of hydrological model parameters. Hydrol. Earth Syst. Sci. Discuss. 5, 1641–1675. http://dx.doi.org/10.5194/hessd-5-1641-2008.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249, 11–29.

Bowers, M.C., Tung, W.W., Gao, J.B., 2012. On the distributions of seasonal river flows: Lognormal or power law?, Water Resour. Res., 48, n/a-n/a, doi:10.1029/2011WR011308.

Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. Water Resour. Res. 36, 3663–3674. http://dx.doi.org/10.1029/2000WR900207.

Boyle, D.P., Gupta, H.V., Sorooshian, S., 2003. Multicriteria Calibration of Hydrologic Models 2003, 185–19, doi:10.1002/9781118665671.ch14.

Brent, R.P., Algorithms for minimization without derivatives, Dover Publications, Mineola, N.Y, 1 online resource (x, 195, 2002).

Carrera, J., Neuman, S.P., 1986. Estimation of Aquifer parameters under transient and steady state conditions: 2. uniqueness, stability, and solution algorithms. Water Resour. Res. 22, 211–227. http://dx.doi.org/10.1029/WR022i002p00211.

Chen, D., Gao, G., Xu, C.Y., Guo, J., Ren, G., 2005. Comparison of the Thornthwaite method and pan data with the standard Penman-Monteith estimates of reference evapotranspiration in China. Clim. Res. 28, 123–132. http://dx.doi.org/10.3354/cr028123.

Chen, X.Y., Chau, K.W., Busari, A.O., 2015. A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model. Eng. Appl. Artif. Intell. 46, 258–268. http://dx.doi.org/10.1016/j.engappai.2015.09.010.

Confesor, R.B., Whittaker, G.W., 2007. Automatic calibration of hydrologic models with multi-objective evolutionary algorithm and Pareto optimization1. JAWRA J. Am. Water Resour. Assoc. 43, 981–989. http://dx.doi.org/10.1111/j.1752-1688.2007.00080.x.

Cullmann, J., Krauße, T., Philipp, A., 2008. Enhancing flood forecasting with the help of processed based calibration. Phys. Chem. Earth, Parts A/B/C 33, 1111–1116. http://dx.doi.org/10.1016/j.pce.2008.03.001.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environ. Model. Softw. 22, 1034–1052. http://dx.doi.org/10.1016/j.envsoft.2006.06.008.

Doherty, J., PEST: A Unique Computer Program for Model-independent Parameter Optimisation, in: Water down under 94: Preprints of papers, Adelaide Convention Centre, Adelaide, South Australia, 21-25 November 1994, NCP, 94/10, 14-15, Institution of Engineers, Australia, Barton, A.C.T, 1994.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. http://dx.doi.org/10.1029/91WR02985.

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J. Hydrol. 158, 265–284. http://dx.doi.org/10.1016/0022-1694(94)90057-4.

Edt.: J. Boess, E. Gehrt, U. Müller, U. Ostmann, J. Sbresny, A. Steininger (Eds.): Erläuterungsheft zur digitalen nutzungsdifferenzierten Bodenkundlichen Übersichtskarte 1:50.000 (BÜK50n) von Niedersachsen, Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 2004.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrol. Sci. J. 55, 58–78. http://dx.doi.org/10.1080/02626660903526292.

Fenicia, F., Savenije, Hubert H.G., Matgen, P., Pfister, L., 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resour. Res., 43, n/a, doi:10.1029/2006WR005098.

Förster, K., Gelleszun, M., Meon, G., 2012. A weather dependent approach to estimate the annual course of vegetation parameters for water balance simulations on the meso- and macroscale. Adv. Geosci. 32, 15–21. http://dx.doi.org/10.5194/adgeo-32-15-2012.

Förster, K., Meon, G., Marke, T., Strasser, U., 2014. Effect of meteorological forcing and snow model complexity on hydrological simulations in the Sieber catchment (Harz Mountains, Germany). Hydrol. Earth Syst. Sci. 18, 4703–4720. http://dx.doi.org/10.5194/hess-18-4703-2014.

Freer, J., Beven, K., Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resour. Res. 32, 2161–2173. http://dx.doi.org/10.1029/95WR03723.

Gelleszun, M., Kreye, P., Meon, G., 2015. Lexikografische Kalibrierungsstrategie für eine effiziente Parameterschätzung in hochaufgelösten Niederschlag-Abfluss-Modellen. Hydrol. Wasserbewirtsch. 59, 84–95.

Hall, M.J., 2001. How well does your model fit the data? J. Hydroinform. 3, 49–55.

Hay, L.E., Umemoto, M., 2007. Multiple-objective stepwise calibration using luca: Open-file report 2006-1323, USGS.

Hill, M.C., 1998. Methods and guidelines for effective model calibration: With application to UCODE, a computer code for universal inverse modeling, and MODFLOWP, a computer code for inverse modeling with MODFLOW, Water-resources investigations report, 98-4005, U.S. Geological Survey; Branch of Information Services [distributor], Denver, Colo.

Hogue, T.S., Sorooshian, S., Gupta, H., Holz, A., Braatz, D., 2000. A multistep automatic calibration scheme for river forecasting models. J. Hydrometeor 1, 524–542. http://dx.doi.org/10.1175/1525-7541(2000) 001<0524:AMACSF>2.0.CO;2.

Hölscher, J., Petry, U., Anhalt, M., Meyer, S., Meon, G., Kreye, P., Haberlandt, U., Fangmann, A., Berndt, C., Wallner, M., Lange, A., Eggelsmann, F., 2014. Globaler Klimawandel: Wasserwirtschaftliche Folgenabschätzung für das Binnenland, Niedrigwasser, first ed., Oberirdische Gewässer, 36, NLWKN, Norden, 122.

Hölscher, J., Petry, U., Bertram, M., Anhalt, M., Schmidtke, S., Haberlandt, U., Müller, H., van der Heijden, S., Berndt, C., Verworn, A., Wallner, M., Belli, A., Dietrich, J., Meon, G., Förster, K., Gelleszun, M., Riedel, G., Lange, A., Eggelsmann, F., 2012. Globaler Klimawandel: Wasserwirtschaftliche Folgenabschätzung für das Binnenland, first ed., Oberirdische Gewässer, 33, NLWKN, Norden, 158.

Jolly, W.M., Nemani, R., Running, S.W., 2005. A generalized, bioclimatic index to predict foliar phenology in response to climate. Global Change Biol. 11, 619–632.

Klute, A., Weaver, R.W., Mickelson, S.H., Sparks, D.L., Bartels, J.M., Dane, J.H., Topp, G. C., Ulery, A.L., Drees, L.R., 1994. Methods of soil analysis, third ed., Soil Science Society of America book series, no. 5, Soil Science Society of America, Madison, Wis, volumes.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97. http://dx.doi.org/10.5194/adgeo-5-89-2005.

Kreye, P., 2015. Mesoskalige Bodenwasserhaushaltsmodellierung mit Nutzung von Grundwassermessungen und satellitenbasierten Bodenfeuchtedaten, Dissertation, Leichtweiß-Institut für Wasserbau, Technische Universität Braunschweig, Braunschweig, 284.

Kreye, P., Gelleszun, M., Meon, G., 2012. Ein landnutzungssensitives Bodenmodell für die meso- und makroskalige Wasserhaushaltsmodellierung In: Wasser ohne Grenzen, Weiler, M. (Ed.), 31.

Kreye, P., Gocht, M., Förster, K., 2010. Entwicklung von Prozessgleichungen der Infiltration und des oberflächennahen Abflusses für die Wasserhaushaltsmodellierung. Hydrologie und Wasserbewirtschaftung 54.

Kreye, P., Meon, G., 2016. Subgrid spatial variability of soil hydraulic functions for hydrological modelling. Hydrol. Earth Syst. Sci. 20, 2557–2571. http://dx.doi.org/10.5194/hess-20-2557-2016.

Kunstmann, H., Kinzelbach, W., Siegfried, T., 2002. Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling. Water Resourc. Res. 38.

Kuzmin, V., Seo, D.-J., Koren, V., 2008. Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search. J. Hydrol. 353, 109–128. http://dx.doi.org/10.1016/j.jhydrol.2008.02.001.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241. http://dx.doi.org/10.1029/1998WR900018.

LWI-HYWAG und IfW: Panta Rhei Benutzerhandbuch - Programmdokumentation zur hydrologischen Modellsoftware (unveröffentlicht), Abteilung Hydrologie, Wasserwirtschaft und Gewässerschutz am Leichtweiß-Institut für Wasserbau der TU Braunschweig in Kooperation mit dem Institut für Wassermanagement IfW GmbH, Braunschweig, 2012.

Madsen, H., 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. J. Hydrol. 235, 276–288. http://dx.doi.org/10.1016/S0022-1694(00)00279-1.

Madsen, H., Khu, S., 2006. On the use of Pareto optimization for multi-criteria calibration of hydrological models. IAHS Publ. 304, 93.

Maskey, S., Guinot, V., 2003. Improved first-order second moment method for uncertainty estimation in flood forecasting. Hydrol. Sci. J. 48, 183–196.

Meon, G., Pätsch, M., van Phuoc, N., 2014. EWATEC-COAST: Technologies for Environmental and Water Protection of Coastal Regions in Vietnam: Contributions to 4th International Conference for Environment and Natural Resources—ICENR 2014, 1., Aufl, HYWAG Schriftenreihe, 1, Cuvillier, E, Göttingen, Niedersachs, 232 S.

Meyer, S., Riedel, G., Lichtenberg, T., Meon, G., Lange, S., 2013. Operationelle Hochwasservorhersage in Niedersachsen mit dem Wasserhaushaltsmodell PANTA RHEI, in: Simulation hydrologischer Systeme – wie nah kommen wir der Realität?: Beiträge zum 3. Trierer Workshop zur Niederschlag-Abfluss-Modellierung am 17. und 18. September 2012. In: Trier, Casper, M., Gronz, O. (Eds.), Forum für Hydrologie und Wasserbewirtschaftung, 33, Dt. Vereinigung für Wasserwirtschaft, Abwasser und Abfall, Hennef.

Monteith, J.L., 1965. Evaporation and environment. In: Fogg, G. E. (Ed.), The state and movement of water in living organisms, Symposia of the Society for Experimental Biology, 19, Cambridge University Press, Cambridge, 205–234.

Moriasi, D.N., Arnold, J.G., 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. Trans. ASABE 50, 885–900.

Moussa, R., Chahinian, N., 2009. Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events. Hydrol. Earth Syst. Sci. 13, 519–535. http://dx.doi.org/10.5194/hess-13-519-2009.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: Part I: a discussion of principles.
Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.
Ning, S., Ishidaira, H., Wang, J., 2015. Calibrating a hydrologic model by step-wise method using grace tws and discharge data, J. Japan Soc. Civ. Eng., I_85-I_90.
Penman, H.L., 1948. Natural evaporation from open water bare soil and grass. Proc. Royal Soc. A: Mathem., Phys. Eng. Sci. 193, 120–145. http://dx.doi.org/10.1098/rspa.1948.0037.
Priestley, C., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters, Monthly Weather Review.
Reusser, D.E., Blume, T., Schaefli, B., Zehe, E., 2009. Analysing the temporal dynamics of model performance for hydrological models. Hydrol. Earth Syst. Sci. 13, 999–1018. http://dx.doi.org/10.5194/hess-13-999-2009.
Reusser, D.E., Buytaert, W., Zehe, E., 2011. Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test, Water Resour. Res., 47, n/a, doi:10.1029/2010WR009947.
Reusser, D.E., Zehe, E., 2011. Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, Water Resour. Res., 47, n/a, doi:10.1029/2010WR009946.
Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46, 1–25. http://dx.doi.org/10.1029/2008WR007327.
Schulla, J., 2012. Model Description WaSiM: (Water balance Simulation Model): completely revised version 2012. <http://www.wasim.ch/downloads/doku/wasim/wasim_2012_ed2_en.pdf> (last access: 24 June 2013).
Sentelhas, P.C., Gillespie, T.J., Santos, E.A., 2010. Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario Canada. Agric. Water Manage. 97, 635–644. http://dx.doi.org/10.1016/j.agwat.2009.12.001.
Taormina, R., Chau, K.-W., 2015. Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. J. Hydrol. 529, 1617–1632. http://dx.doi.org/10.1016/j.jhydrol.2015.08.022.
Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geogr. Rev. 38, 55. http://dx.doi.org/10.2307/210739.
Turc, L., 1961. Evaluation des besoins en eau d'irrigation, evaporation potentielle. Annales Agronomiques 12, 13–49.
Vaze, J., Jordan, P. W., Beecham, R., Frost, A.J., Summerell, G., 2011. Guidelines for Rainfall-runoff Modelling: towards Best-practice Model Application, eWater Cooperative Research Centre.
Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resour. Res. 39. http://dx.doi.org/10.1029/2002WR001746.
Wang, W.-C., Chau, K.-W., Xu, D.-M., Chen, X.-Y., 2015. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resour. Manage 29, 2655–2675. http://dx.doi.org/10.1007/s11269-015-0962-6.
Witte, B., Schmidt, H., 2004. Vermessungskunde und Grundlagen der Statistik für das Bauwesen, 5., völlig neu bearb. und erw. Aufl, Wichmann, Heidelberg, XIII, 714 S.
Wu, C.L., Chau, K.W., Li, Y.S., 2009. Methods to improve neural network performance in daily flows prediction. J. Hydrol. 372, 80–93. http://dx.doi.org/10.1016/j.jhydrol.2009.03.038.
Wurpts, A., Berkenbrink, C., Bruss, G., Dissayanake, P., Grabemann, I., Grashorn, S., Groll, N., Herold, M., Kaiser, R., Knaack, H., Kreye, P., Kuhn, K., Kuhn, M., Lettmann, K., Lüdecke, G., Mayerle, R., Meon, G., Miani, M., Ptak, T., Riedel, G., Ritzmann, A., Schlurmann, T., Schmidt, A., Striegnitz, M., von Storch, H., Weisse, R., Willert, M., Wolff, J.-O., Zorndt, A., Niemeyer, H., 2014. Forschungsthema 7: Küste und Küstenschutz – A-Küst. In: Beese, F., Aspelmeier, S. (Eds.), Abschlussbericht des Forschungsverbundes KLIFF – August 2014. Niedersächsisches Ministerium für Wissenschaft und Kultur, Göttingen.
Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. J. Hydrol. 204, 83–97. http://dx.doi.org/10.1016/S0022-1694(97)00107-8.
Zhang, A., Li, T., Si, Y., Liu, R., Shi, H., Li, X., Li, J., Wu, X., 2016. Double-layer parallelization for hydrological model calibration on HPC systems. J. Hydrol. 737–747.
Zhang, X., Srinivasan, R., Zhao, K., van Liew, M., 2009. Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. Hydrol. Process. 23, 430–441. http://dx.doi.org/10.1002/hyp.7152.