Parameter calibration and uncertainty estimation of a simple rainfall-runoff model in two case studies

X. Zhang, G. Hörmann, N. Fohrer and J. Gao

ABSTRACT

The simple rainfall-runoff conceptual KIDS (Kielstau Discharge Simulation) model using PCRaster is applied to simulate continuously daily discharge of the Kielstau and XitaoXi basins. This work focuses on parameter calibration procedure and, in particular, assessment of model prediction uncertainty. We employ a simplistic analysis routine SUFI-2, coupled with the implementation of a Monte Carlo based sampling strategy for the joint investigation of parameter calibration and uncertainty estimation. The scatter plots of model performance and parameter exhibit high equifinality of parameter sets in fitting observations, while their histogram distribution patterns imply that most parameters can be well defined. This study investigates parameter sensitivities and finds interesting local results: soil and groundwater parameters are more sensitive in Kielstau models than in XitaoXi models, and only the soil parameters 'S_{fk}' and 'K_c' are found strongly correlated. Finally, the uncertainty bounds are always thin and the global shape of the hydrograph is well approximated for both basins. As the validated uncertainty bounds also represent the desired coverage (P factor $>50\%$) of the observations, and the calculated R factor values are in the targeted range (R factor < 1), it demonstrates the efficiency and suitability of this revised SUFI method for the two case studies.

Key words | hydrologic modeling, KIDS model, Kielstau in Germany, parameter uncertainty estimation, random sampling, XitaoXi in China

X. Zhang (corresponding author)
G. Hörmann
N. Fohrer
Ecology Centre,
Department of Hydrology and Water Resources Management,
Institute for the Conservation of Natural Resources,
Christian-Albrechts-University of Kiel,
Olshausenstr. 40,
D-24098 Kiel,
Germany
E-mail: *xzhang@ecology.uni-kiel.de*

J. Gao
Nanjing Institute of Geography & Limnology,
CAS,
Nanjing 210008,
China

INTRODUCTION

In the last few decades, there has been an increasing use of dynamic simulation in hydrological models. In order to improve the accuracy of model simulations, many calibration methodologies have been developed intending to locate the values of unknown parameters and to verify the usefulness and power of models. The laborious nature of conventional manual calibration in the 'trial and error' adjustment style has motivated the development of automatic calibration techniques, including gradient-based methods like the Gauss-Levenberg-Marquardt method (Doherty & Johnston 2003), population-evolution-based algorithms like the shuffled complex evolution method (Duan *et al*. 1992), and regionalization or spatial generalization (Lamb & Kay 2004). The main difficulties preventing the determination of the best parameter set by any automatic calibration schemes are the presence of non-uniqueness in parameter optimization, nonlinear parameter interaction and the complex shape of the response surface defined by the objective function (Feyen *et al*. 2007). Moreover, since process-based hydrological models consist, at least partially, of an empirical combination of mathematical relationships describing some observable features of idealized hydrological processes (Kuczera & Parent 1998), parameter estimates are subject to uncertainty, which leads to uncertainty in model predictions. To overcome these problems, many studies have shifted the research emphasis on identifying the model prediction uncertainty, instead of searching for one absolute global optimum. Parameter calibration and prediction uncertainty of a model are then intimately related (Blasone *et al*. 2008; Laloy *et al*. 2010).

doi: 10.2166/hydro.2012.084

A large number of methods have been published recently to derive feasible calibration techniques and uncertainty analysis in hydrologic modeling, involving the optimization algorithms such as the shuffled complex evolution algorithm (Duan *et al*. 1992; Vrugt *et al*. 2003b; Feyen *et al*. 2007), simulated annealing (Sumner *et al*. 1997) and genetic algorithms (Wang 1997; Cheng *et al*. 2006). With the aim of simulating various important characteristics of the observed data, automatic routines using multiple criteria or objectives have been introduced in solving the calibration problem (Yapo *et al*. 1998; Boyle *et al*. 2000; Cheng *et al*. 2002; Madsen 2003; Vrugt *et al*. 2003a; Schuol & Abbaspour 2006; Li *et al*. 2010). Other approaches focusing on assessing global uncertainty in rainfall-runoff modeling include the (pseudo-) Bayesian methods (Freer *et al*. 1996; Thiemann *et al*. 2001), such as the generalized likelihood uncertainty (GLUE) framework (Beven & Binley 1992; Aronica *et al*. 2002; Uhlenbrook & Sieber 2005; Blasone *et al*. 2008), and the meta-Gaussian approach (Krzysztofowicz & Kelly 2000; Montanari & Brath 2004). These methods contributed to the development of more complex or sophisticated uncertainty analysis tools, but some weaknesses may be inevitable at the same time, such as the fact that global algorithms are mostly complex and computationally expensive (Li *et al*. 2010), and other approaches such as GLUE also requires a large sample of model runs and adequate reliable data describing watershed characteristics (Choi & Beven 2007).

In practice, however, the complex approaches do not always provide more accurate results relative to simpler and low-dimensional estimation problems, depending on the watershed scale, the number of parameters to calibrate, or the quantity and quality of calibration data. A simple approach may be adequate and efficient in the cases of ungauged basins, or lumped models with less parameters.

Abbaspour *et al*. (2004) proposed a simple inverse modeling routine SUFI-2 (Sequential Uncertainty Fitting, version 2) for an uncertainty estimation of the SWAT (Soil and Water Assessment Tool) model (Arnold *et al*. 1998). Another application in Schuol & Abbaspour (2006) showed that SUFI-2 is an efficient parameter optimization-uncertainty analysis procedure for a multi-site large-scale water quantity investigation. The sequential fitting scheme in SUFI-2 helps to maintain an appropriate sampling density in the Monte Carlo based sampling method. The easy-to-setup approach is attractive for low-dimensional uncertainty estimation problems, such as modeling with restricted data availability, a relatively simple model structure, or less parameters. The calibration procedure we have applied in this paper belongs to this group, but with adjustments adapted to the PCRaster modeling environment (Van Deursen 1995; Wesseling *et al*. 1996). The current study has two objectives. First, we test the SUFI-2 method with a simple rainfall-runoff conceptual model, KIDS (Kielstau Discharge Simulation) (Hörmann *et al*. 2007; Zhang *et al*. 2007). The KIDS model is raster based using a dynamic modeling language PCRaster, which has more flexibility in input data requirements and less complexity in model structure than the SWAT model applied in Abbaspour *et al*. (2004). Second, we assess the parameter uncertainty for the KIDS model and quantify its effects on model simulations for daily discharge of two river basins. One is a small lowland Kielstau catchment in Germany, and the other is a mesoscale mountainous XitaoXi basin in China. We hypothesize that the approach in this two-site comparison would do better to examine the influence pattern of each parameter on model performance, and yield better results in uncertainty assessment than one could get from extrapolating parameters from a single site. To evaluate our hypothesis, we selected six main parameters for each basin, compared their distribution patterns and correlations, and applied the adapted SUFI-2 methodology to map and quantify simulation uncertainty in the modeling process onto the parameter space.

SITES DESCRIPTION AND DATA PROCESSING

The study sites include two basins with remarkable differences in hydrologic features – Kielstau in Germany and XitaoXi in China.

Kielstau is a lowland watershed in Northern Germany, with a drainage area of 51.5 km^2 (see Figure 1(a)). As the development of the landscape was mainly influenced by the Saale and the Weichselian ice ages (Eggemann *et al*. 2001), the whole catchment is rather flat, with elevation ranging from sea level to 50 m. Land use is dominated by agricultural (55.8%) and grassland (26.1%). Mean annual precipitation is circa 800 mm, and evaporation approximately 400 mm (Schmidtke 1995). The main soil texture is

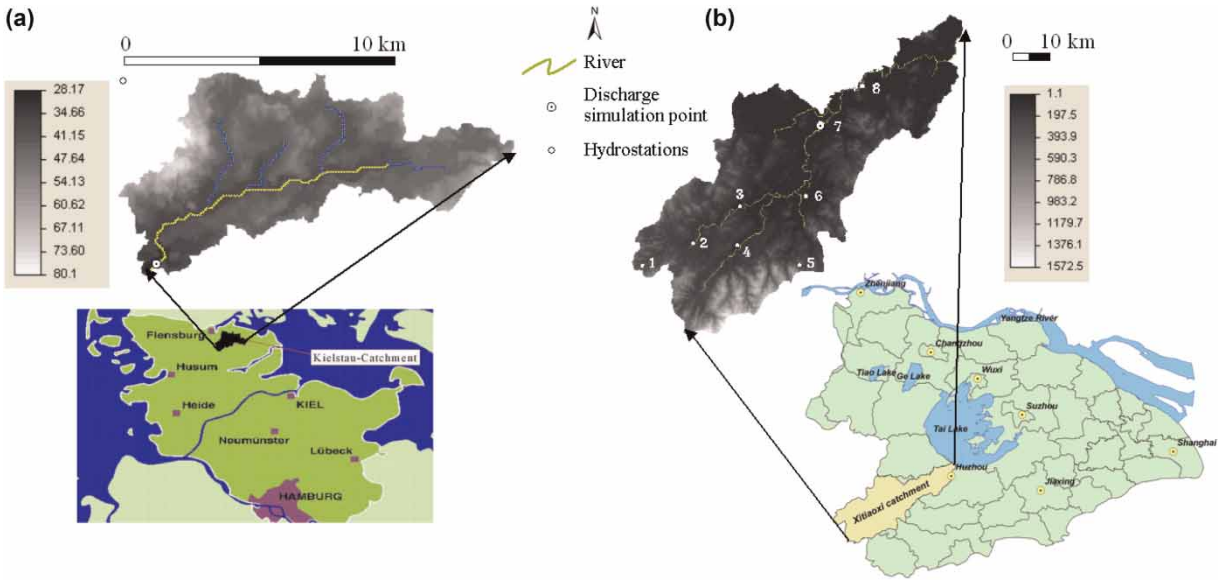

Figure 1 | Location of (a) the Kielstau catchment and (b) the XitaoXi catchment, with gray-scale overlay of the topography (discharge simulation point: Soltfelt-Kielstau, Hengtangcu-XitaoXi; weather station in Kielstau – Flensburg, hydrostations in XitaoXi – 1: Tianjintang, 2: Hanggai, 3: Fushishuiku, 4: Laoshikan, 5: Yinkeng, 6: Dipu, 7: Hengtangcun, 8: Fanjiacun).

sandy loam and the dominating soil types are Gleysol and Luvisol (Sponagel 2005). A large fraction of wetland area and the near-surface groundwater level are observed in this region.

The second study watershed, XitaoXi, is a 2,271 km^2 sized mountainous basin (Figure 1(b)), which is located in the semitropical monsoon zone in Southern China. It is a sub-basin of the Taihu Lake. In the XitaoXi region, 63.4% of land use is forest and grass, 20% paddy rice land (Wan *et al.* 2007). Average precipitation in the watershed is 1,466 mm annually, with 75% of rain falling between April and October. Average evaporation from water surface ranges from 800 to 900 mm annually. The dominant soil types are red soil and rocky soil. Since these soils tend to have limited water storage capacity, most portions of the river discharge are probably from the saturation excess surface runoff (Xu *et al.* 2007).

The KIDS model in PCRaster requires climatic and topographic data as basic data input for all model runs. The DEMs are derived originally from the topographic maps provided by local authorities (Landesvermessungsamt Kiel and the Administrative Bureau of TaiHu Basin) and gridded with an resolution of 50 m for the Kielstau basin and 200 m for the XitaoXi basin according to the input dap scale. Although the drainages in Kielstau area are relatively indistinct due to the flatness of the basin surface, the KIDS model still allows for efficient runoff routing based on the flow accumulation calculated from DEM.

The climatic data for Kielstau catchment are taken from the data set of Flenburg station, the official weather station nearest Kielstau basin. The LANU (Landesamt für Natur und Umwelt) provided river discharge values from 1983 to 1999 (at the official Soltfeld gauge station). Daily precipitation data in the XitaoXi basin are available from eight stations within the watershed area, while evaporation data are only recorded at two stations – Fushishuiku and Hengtangcun. The discharge data set from the selected Hengtangcun gauge station is available from 1978 to 1987. The reason to select Hengtangcun instead of the river outlet as discharge simulation point is that Hengtangcun is the nearest station to river outlet and are not greatly influenced by the backflow from Taihu Lake. Only the contributing area is considered in the model calculation. Other data including land-use and soil maps are converted from its originally coarser data resolution to a raster map with the same cell length of DEM.

The Kielstau and XitaoXi watersheds have quite different hydrologic regimes (Zhang *et al.* 2009), thereby

providing diverse data sets to test the adapted SUFI-2 method used in this study. For example, the average daily runoff of the XitaoXi river ($35.09\,m^3\,s^{-1}$) is much higher than that of the Kielstau stream ($0.45\,m^3\,s^{-1}$). The difference is significant as well when considering the different discharge area: the runoff rate per unit area is $23.02\,l\,s^{-1}\,km^{-2}$ for XitaoXi, and $8.82\,l\,s^{-1}\,km^{-2}$ for Kielstau. The calibration data considered in this paper are the daily records of river discharge ($m^3\,s^{-1}$) at the selected gauge station. The current work also serves as one part of a Sino German integrated geohydrological study of these two basins. All analysis steps will be made in parallel for the Kielstau and XitaoXi catchments.

KIDS HYDROLOGIC MODEL

The KIDS model (Kielstau Discharge Simulation model; Hörmann *et al.* 2007; Zhang *et al.* 2007, 2011) is a simple rainfall-runoff model developed for practical purposes to facilitate water resource management in the field of Kielstau. It is basically driven by elevation map and meteorological input data, and then simulates river discharge in given river basins as a dynamic function of spatial information using PCRaster modeling language. Additional inputs like soils and land cover can be integrated as extended submodels to the basic model structure. All derived models form the KIDS model ensembles. Figure 2 gives an overview of the basic model structure with solid lines and added submodels in dashed lines. The model is spatially distributed and space was discretized in 50×50 m grid size for Kielstau and 200×200 m for XitaoXi.

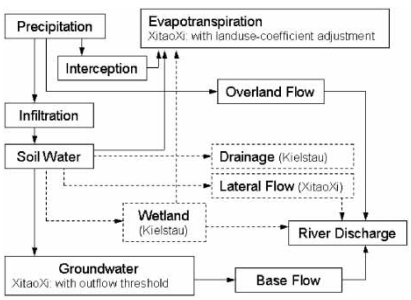

Figure 2 | Illustration of the KIDS model structure.

Runoff is calculated on each grid cell based on the water balance equation (Equation (1)), taking into account interception, precipitation, evapotranspiration, and the flows to other compartments. We use 'mm' as unit of measure for all the water amount expressions included in equations of this paper, and calculate with a daily time step.

$$S_t = S_{t-1} + P_t - \mathrm{ET}_t - I_t - \mathrm{Qo}_t - \mathrm{Sp}_t \quad (1)$$

where S is the soil water content, t is the modeling time step (daily), P is precipitation and ET is evapotranspiration, I is interception, Qo is surface runoff (overland flow), Sp is percolation or seepage.

Precipitation and potential evapotranspiration are input forcing data. Interception is calculated with the parameter 'Im' (Equation (2)), such that

$$I_t = \min(P_t, \mathrm{I}m) \quad (2)$$

where Im is the maximum interception amount of vegetation cover.

Surface runoff 'Qo' (Equation (3)) describes soil percolation and storage calculation on the basis of derived soil parameters like field capacity and infiltration rate of soil water deficit,

$$\mathrm{Qo}_t = \max\{[P_t - I_t - K_c(S_{fk} - S_t)], 0\} \quad (3)$$

where Qo represents the surface runoff, K_c is the infiltration parameter, and S_{fk} is the wetness at field capacity.

The whole river basin is assumed to have one soil type with unified water storage capacity in the basic model structure. Sub-surface flow is modeled as 1 D bucket flow with a lateral flow rate parameter 'K_s', as shown in Equation (4). The value of parameter K_s is set equal to zero in the basic model for both basins, so

$$\mathrm{Qs}_t = K_s S_t \quad (4)$$

where Qs is subsurface flow, and K_s is lateral flow rate.

The groundwater layer is represented as linear storage and its discharge is set with a groundwater outflow rate 'K_g'. Combining Equations (5), (6) and (7) yields the

daily groundwater dynamics,

$$G_t = G_{t-1} + \mathrm{Ig}_t - \mathrm{Qg}_t \quad (5)$$

$$\mathrm{Ig}_t = \rho S_t \quad (6)$$

$$\mathrm{Qg}_t = K_g G_t \quad (7)$$

where G is groundwater storage, Ig the inflow to groundwater aquifer, Qg the groundwater discharge to runoff, ρ the water seepage rate from soil to groundwater, K_g the groundwater outflow rate.

The flow path is then derived from topography through a flow accumulation grid calculated in PCRaster. The routing of the runoff is modeled with the fully dynamic kinematic wave function (Chow *et al.* 1988).

Considering the greatly differing hydrology of the two basins, the appropriate model structure for each basin is selected from the KIDS model ensembles (Zhang *et al.* 2008, 2011): Model 'DW' for Kielstau – basic KIDS model with drainage (Equation (8)) and integration of wetland (Equation (9)); Model 'LTG' for XitaoXi – basic KIDS model with landuse-coefficient adjusted ET distribution, additional subsurface flow with $K_s > 0$ and groundwater outflow threshold (Equation (10)).

$$D_t = K_d * S_t + L_d \quad (8)$$

where D is the drained water volume, K_d is drainage factor, S is available soil water storage, L_d is the lateral inflow volume (lateral seepage from irrigation canals and drainage channels),

$$W_t = W_{t-1} + \mathrm{Iw}_t - \mathrm{Ew}_t - \mathrm{Qw}_t \quad (9)$$

where W is wetland water storage; Iw the incoming water volume influenced by precipitation, interception and soil moisture; Ew the water loss from wetland, mainly evapotranspiration; Qw the wetland water seepage contributing to runoff, and

$$\mathrm{Qg}_t = \mathrm{Min}(K_g * G_t, \mathrm{Gm}) \quad (10)$$

where Gm is the maximum daily groundwater outflow or groundwater outflow threshold.

Owing to the general nature and flexible structure of the KIDS model, its application to any study area requires that certain parameters be identified for the particular basin. In the current model version, six main parameters need to be determined by calibration using daily discharge observations. Table 1 lists an overview of the calibration parameters with their upper and lower value ranges.

CALIBRATION SCHEME

SUFI-2 revised with random sampling strategy

The SUFI-2 scheme (Abbaspour *et al.* 2004) is designed to perform multi-site, semi-automated global search procedure with the SWAT hydrologic models (Arnold *et al.* 1998). Schuol & Abbaspour (2006) provided another application example using SUFI-2 for large-scale water quantity investigations. It combines parameter calibration and uncertainty prediction. The procedure is simple and the uncertainty is worked out by calculating the likelihood of the parameter set using the simulated and observed discharge. The choice of the likelihood function relies on the user-defined hypothesis. However, the Nash-Sutcliffe (NS) efficiency (Nash & Sutcliffe 1970) is assumed as the informal likelihood measure, which is widely used as a performance measure in hydrological modeling (Beven *et al.* 2007).

Table 1 | Description of parameters included in the KIDS hydrological model calibration procedure, with their upper and lower bounds

Parameter (unit)	Range	Definition
Im [mm]	0–10	Maximum water amount intercepted by vegetation cover
S_{fk} [mm]	1–800	Soil water storage capacity
K_c [–]	0.01–0.6	Soil infiltration parameter
ρ [–]	0.01–0.5	Water seepage rate from soil to groundwater
K_g [–]	0.001–0.1	Groundwater discharge rate to the river baseflow
K_d [–] in Kielstau	0–0.01	Water drainage rate from available soil water storage
K_s [–] in XitaoXi	0.001–0.05	Soil water percolation rate to river discharge

Each set of parameters is assigned a likelihood value (the NS index) to quantify how well that particular parameter combination simulates the system. Higher NS values typically indicate better correspondence between the model prediction and observations. A preliminary sensitivity test of six parameters is conducted to decide the initial sampling space (as shown in Table 1) by sequentially varying one parameter while keeping all other constant. In order to adapt the sampling scheme in SUFI-2 to PCRaster environmental modeling language, Schmitz *et al.* (2009) developed a specialized client-server software toolbox for the KIDS model. It has extendable interfaces and a Python binding allowing the user to add calibration algorithms and define specific objective functions. This software framework is suited for the PCRaster environment to assign the KIDS model files, and it is applied to the parameter sampling in this study. It is a Monte Carlo based method as it evaluates the objective function at randomly spaced points in the defined parameter space. This random sampling scheme is easy to install and requires two assumptions when used in practical applications: low-parameterized models and the uniform prior distributions for the tested parameters. Our conjecture is that the random sampling scheme adopted here can provide a sufficiently large sample of solutions for the simple KIDS models with relative low-dimensional parameter estimation problems.

Analysis procedure

The process of parameter calibration and uncertainty analysis of the SUFI-2 algorithm is depicted graphically in Figure 3. Starting from models with the initially large parameter sampling range, each iteration run generates 2,000 model simulations ($s = 2{,}000$). With the purpose of identifying a group of behavioral parameter sets within the resulting possible model parameter combinations from the first iteration, we derived a subjective method to differentiate between behavioral and non-behavioral simulations. The term 'behavioral' is used here to characterize those parameter sets that are judged to be 'acceptable' on the basis of available data and knowledge (Tang *et al.* 2007). As each parameter set is assigned with a NS value, we sorted the sample population in order of decreasing NS value, and chose half the results ($s/2 = 1{,}000$) with higher NS

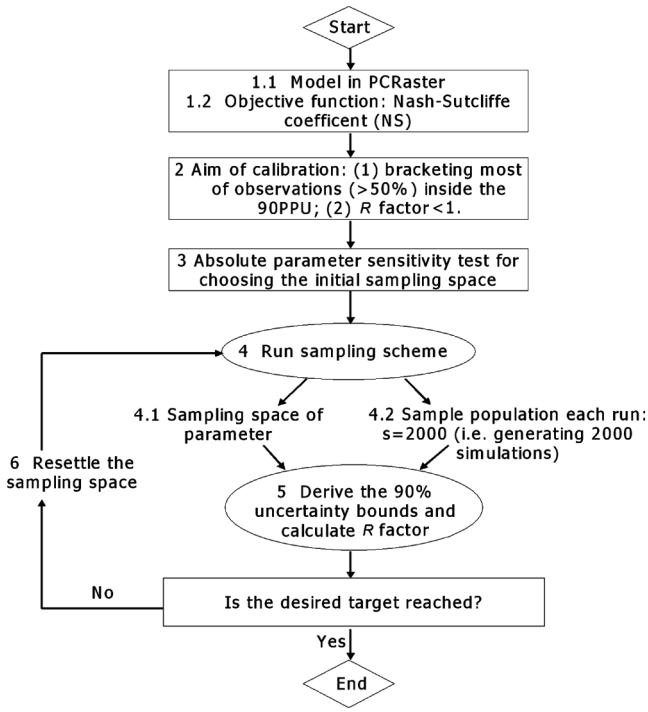

Figure 3 | Flowchart of the calibration strategy.

values as behavioral parameter combinations. From the chosen 1,000 simulations result, the parameter distribution pattern, parameter correlation and identifiability can be inferred. In SUFI-2, parameter uncertainty is depicted as uniform distributions. We construct the probability distributions for each parameter by dividing its sampling range into ten equivalents. When a further iteration is required, we narrow the parameter sampling range by neglecting those parameter value ranges that has low probability (it is defined here <5%). The resettled parameter value range is used as the new sampling range in the next SUFI-2 iteration.

Two measures are defined in SUFI-2 to quantify the model uncertainty and to decide the calibration target: *P* factor and *R* factor (Equations (11) and (12)). The percentage of measured data bracketed by the 90% simulation uncertainty bound is referred as the *P* factor. From the simulation results of each sampling run, the 90% uncertainty bound can be derived by calculating the 0.05 and 0.95 quantiles of the simulated discharge values. We defined the associated parametric uncertainty *R* factor as the ratio of the average distance of the 90% uncertainty intervals and

the standard deviation of the measured data. The ideal situation is to have an R factor value close to zero, while at the same time to cover all the observation data within the 90% prediction uncertainty bounds (P factor = 100%).

$$P\text{ factor} = \frac{m}{n} \times 100\% \tag{11}$$

where m is the number of model time steps counted when the observed river discharge value is within the modeled simulation uncertainty bounds, n is the time steps of the selected flow period, and

$$R\text{ factor} = \frac{\sum_{i=1}^{n}\left[(S_{95}-S_5)_1+(S_{95}-S_5)_2+\ldots+(S_{95}-S_5)_n\right]/n}{stdev\{Q1,Q2,\ldots,Qn\}} \tag{12}$$

where S_{95} and S_5 denote the 95 and 5% percentiles for each simulated variables, n is the time steps of the selected flow period, and *stdev* is standard deviation of the observed flows within the selected period.

The values of the P and R factors reflect the uncertainty about the model parameters after taking into account the discharge observations. These two measures also specify the stopping rules of SUFI iterations. When the calculated R factor value reaches a small ratio – usually less than 1, and most of the observations (>50%) can be bracketed inside the uncertainty bound, we defined it as a state of being sufficiently calibrated. The sampling run will be iterated several times by resettling, usually narrowing the parameter space though the posterior parameter distribution, until the calibration target is achieved.

RESULTS AND DISCUSSION

The SUFI-2 scheme is implemented for the Kielstau and XitaoXi basin. Starting from an initially large parameter sampling space, the calibration procedure was iterated three times until the desired target was reached. This section presents analyses for the last iteration that reaches the desired calibration targets.

Parameter uncertainty and correlation

The result of the last iteration is plotted in Figure 4 (Figure 4 (a) for Kielstau and Figure 4(b) for XitaoXi) to assess the parameter behavior on model performance in Nash-Sutcliffe values. In general, the scatter plots of model performance versus parameter values exhibit a high degree of equifinality of parameter sets in fitting the observations. This indicates that a wide range of parameter values can be included in behavioral parameter sets to make the model's performance similar or close to optimal. This equifinality problem has been universally found and accepted as a working paradigm in hydrological models (Zak & Beven 1999; Beven 2007; Li *et al*. 2009).

For Kielstau, the parameter, ρ (water seepage rate from soil to groundwater), is the most sensitive to model performance, with a peaked band of dots ranging 0.05–0.1. This means that better model performances occurred for parameter sets having ρ values between 0.05 and 0.1. For XitaoXi, a tendency of achieving better results is observed when the parameter K_c and ρ have small values. The maximum efficiency occurred in Kielstau model when the parameter S_{fk} has higher value (around 450 mm), while in XitaoXi model it was with a lower value of S_{fk} near 225 mm. An inspection of other plots in Figure 4 reveals low sensitivity of those parameters, where smooth response surfaces are clearly displayed. The scatter plots describe adequately the high equifinality of acceptable parameter sets, and imply a large uncertainty from parameter estimations. The high uncertainty depicted in the figures is mainly due to the iteration, which leads to a narrower sampling space taken from a wide interval of possible parameter values. Further explorations are thus needed by examining parameter distributions and calculating associated uncertainty bounds.

For a closer look at the parameter distribution pattern, we construct the posterior parameter distribution from the result of the last iteration by setting a subjective cutoff threshold (50% of all simulations) for behavior parameter sets. This posterior distribution can be used to directly resettle, usually narrow, the parameter sampling space when a further iteration is needed. Figure 5 presents the probability distributions for the calibration parameters constructed using the top-ranked (NS-index) 1,000 samples. For a good

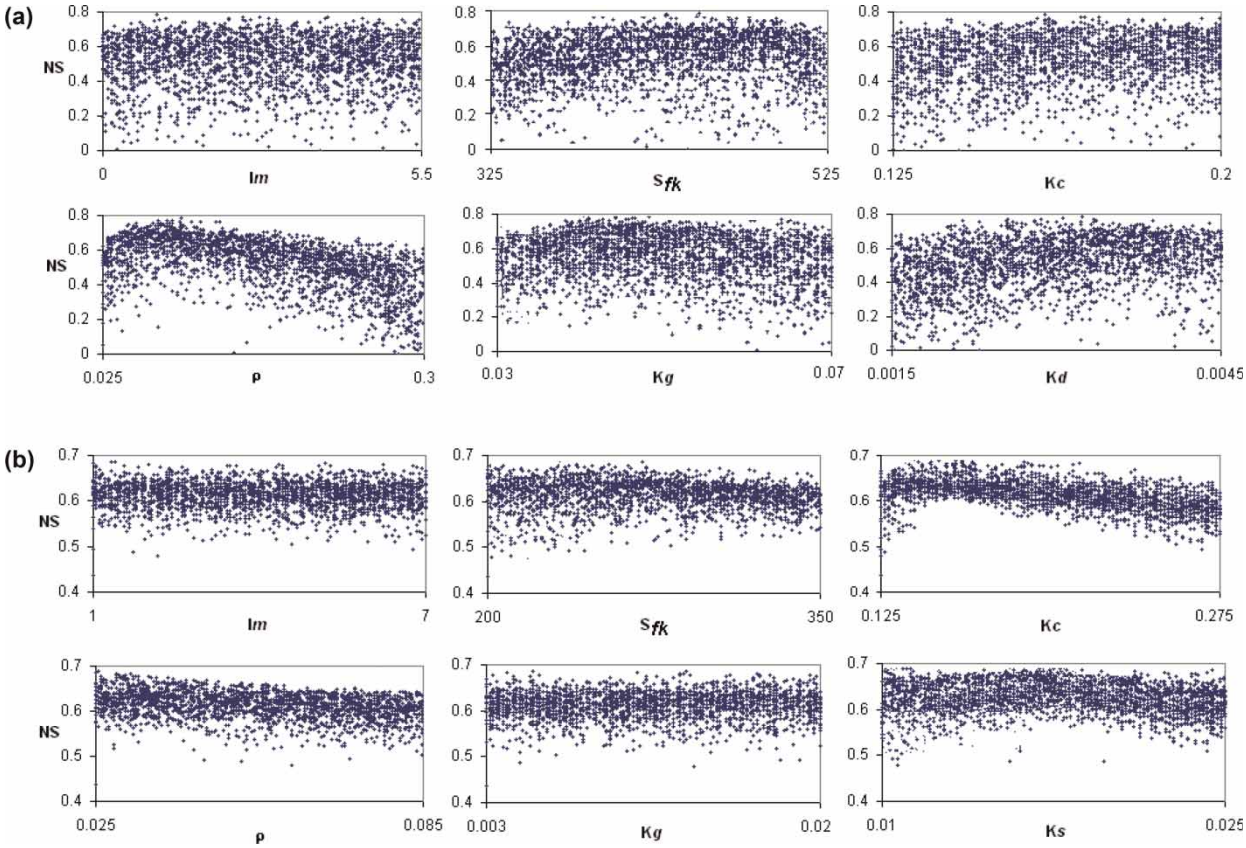

Figure 4 | Scatter plots of model performance in Nash-Sutcliffe index within the parameter space of the last sampling iteration for (a) the Kielstau basin and (b) the XitaoXi basin.

comparison of the two study basins, results for the same parameter are plotted in one diagram, where the x-axis depicts its original bound range stated in Table 1. The upper and lower boundary values of each parameter in its distribution histogram are showed along the x-axis, where numbers with brackets are values for the XitaoXi basin. The response surface of each parameter can help to identify its optimum value range.

The approximately flat response surface for the interception parameter Im corresponds to a uniform distribution for both basins. It indicates a nearly negligible influence contribution of this part in water processes to the overall model performance. The distributions of soil field capacity parameter S_{fk} and hydraulic infiltration parameter K_{c} show a log-normal shape. Given that the peak of their posterior distribution around the most likely value is sharp or nearly sharp, these parameters are well defined for the study area. The most likely value of the soil water storage capacity parameter S_{fk} is well defined but its optimal value range differs in the two basins. Most acceptable simulations are achieved when the S_{fk} in XitaoXi has lower values from 230 to 275, while in Kielstau it is in the higher values ranging from 405 to 485. Owing to the very different soil types in two basins, it reflects the relatively high water storage capacity of the loamy soils or wetland soils in the lowland area of Kielstau, and meanwhile limited soil water capacity of the typical red soil and rocky soil in the mountainous region of XitaoXi. Distributions of another soil parameter K_{c} still have a large probability at their peaks indicating more uncertainty on their most likely value for both basins. The density of the soil groundwater flux parameter ρ approximates a normal distribution centered around the sharp peak, despite the fact that its left lower boundary is truncated by the limits of the x-axis. The higher ρ optimal values in Kielstau reveal a more interactive relationship between soil and groundwater layers. Moreover, the most

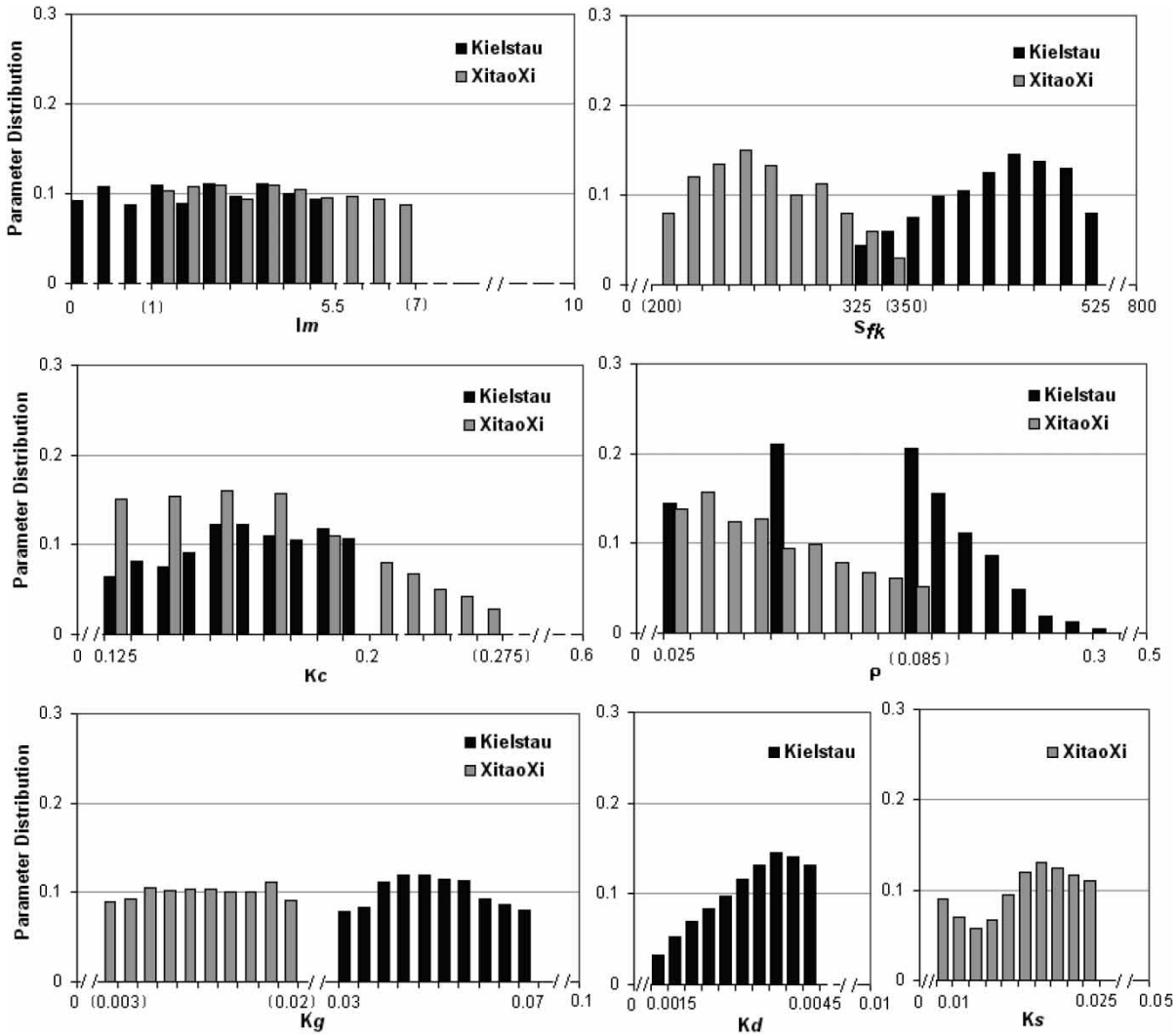

Figure 5 | Histograms of parameter distribution in the comparison of the two study basins. Parameter units are listed in Table 1.

likely value of groundwater recharge parameter K_g is not easily identifiable in XitaoXi, but in Kielstau it shows a moderate normal distribution. This indicates that groundwater in Kielstau plays a more important role in river runoff producing processes. Furthermore, the added submodel parameters describing drainage rate K_d in Kielstau and lateral flow rate K_s in XitaoXi have high possibilities in small values comparing their original bound ranges. The optimal value of K_d is well defined, but the distribution of K_s is slightly bimodal. One must also note that for all parameters, the intervals between the lower and upper bounds are not identical for different study basins, in order to have a good comparison on the same range of the x-axis.

As an illustration, Figure 6 presents correlation plots in two dimensions of parameter space obtained from samples with good performance (10% of the maximum NS) in the two study areas. It indicates that moderate to strong linear correlations exist between some parameters. For example, the plots of soil parameters $\{S_{fk}, K_c\}$ show a clear negative correlation in both cases. It is especially closely related in the XitaoXi catchment. The two parameters affect the fast response of the model either through overland flow or subsurface runoff. We also note in panels $\{S_{fk}, K_d\}$ for Kielstau and $\{S_{fk}, K_s\}$ for XitaoXi, these parameters are slightly negative correlated, while in $\{K_c, K_d\}$ for Kielstau and $\{K_c, K_s\}$ for XitaoXi, they are slightly positive correlated. It can be

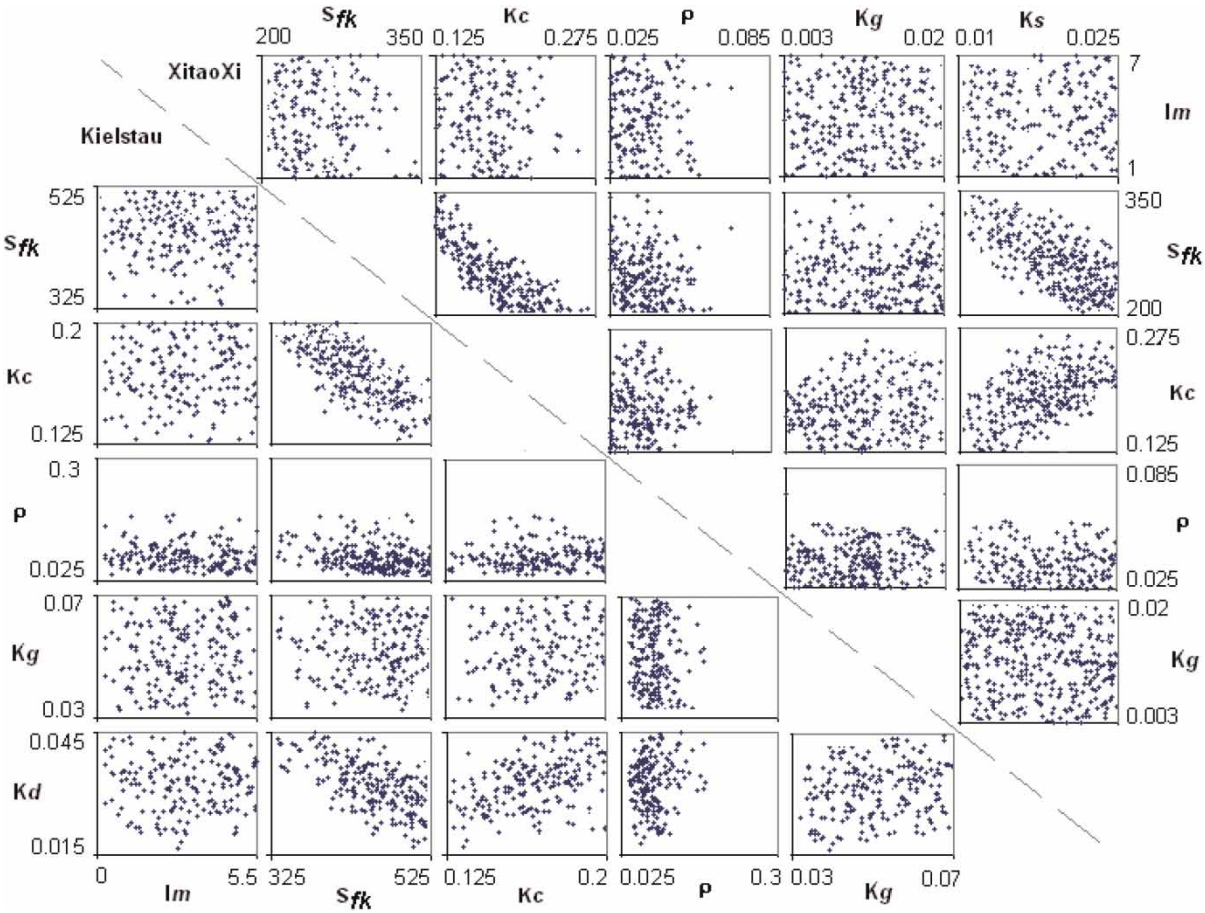

Figure 6 | Correlation plots of parameters in two study areas (lower left: Kielstau, upper right: XitaoXi).

directly explained by their effect on model response. Indeed, an increase of S_{fk} will have the same effect on runoff production as a decrease of K_c, or K_d for Kielstau (K_s for XitaoXi). The plots for parameter ρ in both basins are restrained within the half range of lower values, suggesting a high probability to have optimal values near lower boundary. Other plots concerning parameters Im and K_g show very low correlations, which is consistent with their large uncertainty demonstrated in Figures 4 and 5.

The SUFI-2 uncertainty bounds in calibration and validation results

Accurate probabilistic forecasting requires that the uncertainty bounds are statistically meaningful and exhibit the appropriate coverage. Instead of focusing on the goodness-of-fit of the median output estimate, we examine the properties of each iteration result by checking if the two measures of P factor and R factor could reach the desired targets. When the calculated results from the total generated simulations of the last SUFI-2 iteration meet our calibration targets, the same run was conducted for validation. The calibration and validation results for both study basins are derived here in the sense that they represent the desired range of simulation uncertainty instead of one 'optimum' parameter set. The 90% uncertainty bound describes the parameter uncertainty resulting from the non-uniqueness of effective model parameters. As we can see from Figures 7 and 8, hydrograph simulations obtained from the 90% uncertainty bounds can reproduce the observed discharges with reasonable accuracy for both catchments. The calculated values of P factor and R factor are also shown in the figures. This exhibits the appropriate coverage, i.e. the simulations bracket the observations most of the time

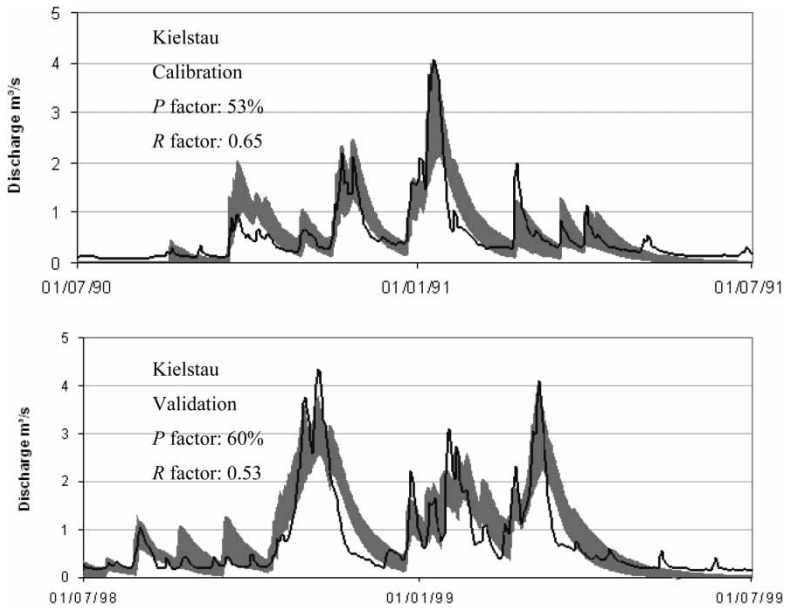

Figure 7 | Calibration and validation results for Kielstau basins showing the 90% simulation uncertainty intervals (gray shadow) along with the measured discharge (black line).

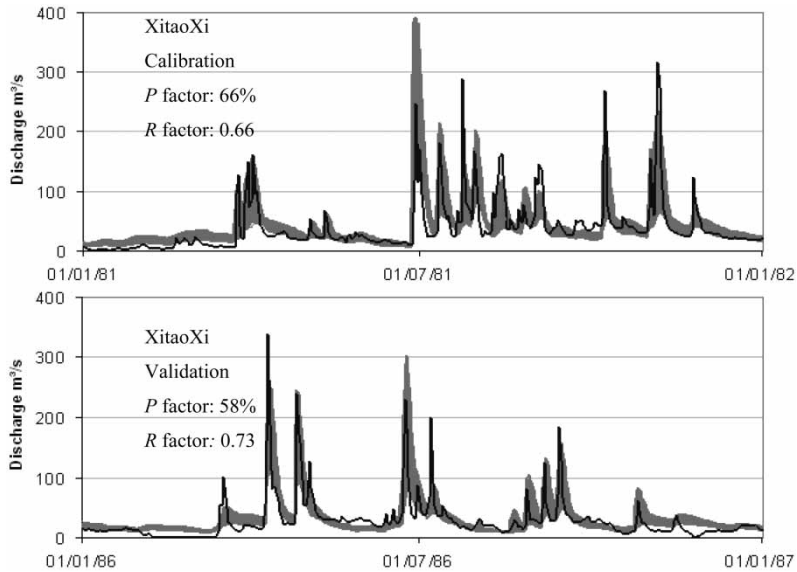

Figure 8 | Calibration and validation results for XitaoXi basins showing the 90% simulation uncertainty intervals (gray shadow) along with the measured discharge (black line).

(P factor > 50%) with acceptable R factor values (R factor < 1). The plot reveals that there is a tendency for small discharges to be underestimated in Kielstau during the summer time, but overestimated in XitaoXi at its spring seasons. For example at the beginning of calibration period in Kielstau, the low runoff is underestimated during the months of July and August. This can be partly a result of the recorded dry season, and partly because of the relatively low water storage modeled in wetland or groundwater aquifers accumulated from previous calibration years, which are supposed to be important baseflow sources in low-rain seasons. Meanwhile in XitaoXi, the river discharges are

slightly overpredicted around the month of March for both calibration and validation. This can happen given the great complexity in water management (like irrigation) of such periods for which plants are growing with low precipitations (Gao *et al.* 2006). However, the global shape of the hydrograph is nevertheless well approximated. When comparing the calibration and validation results for the Kielstau basins, it is noticeable that better *P* factor and *R* factor values are achieved for the validation period. As we can see from Figure 7, the validation year included more peak flows, while a longer period of low flow was presented during the calibration year. The pattern is consistent with the inference that low-flow seasons contribute more to the overall uncertainty than peak-flow seasons (Zhang *et al.* 2011). Regarding the XitaoXi catchment, the total uncertainty bounds are relatively thin despite similar *R* factor values, indicating a good model performance. This could partly be due to the much larger scales in the *y*-axis of the XitaoXi river discharge amount compared with the small Kielstau streamflow. As the calibration and validation results show accepted accuracy for both basins, it clearly indicates that the data are adequate to capture the scale of local hydrologic processes in spite of different resolution data used for each catchment.

Finally, the simulated uncertainty bounds in Figures 7 and 8 using SUFI-2 are mainly associated with parameter estimation but account for all sources of uncertainties such as model structure and input data (Abbaspour *et al.* 2004). The fact that the parameter uncertainty is relatively narrow but does not always bracket the observations indicates that the model structure may need further improvement. However, the advantage of a more complex model is limited and may be outweighed by an increase in model uncertainty with added processes and information.

CONCLUSION

A simple and iterative parameter optimization-uncertainty analysis routine, SUFI-2, has been successfully used to sample the parameter spaces in the KIDS model and to find the associated simulation uncertainty from the non-unique parameter sets. The random sampling scheme adopted in our approach is proved to be adequate and efficient for relatively simple low-dimensional sampling problems, as the applications in two very-differing study basins demonstrates here.

The scatter plots of parameter values versus model performance (NS) clearly exhibit equifinality where many possible parameter sets can provide acceptable simulations of the catchment response. This also implies a large uncertainty from parameter estimations. It ensures the assumptions that a meaningful calibration should include the uncertainty estimation instead of identifying a unique set of parameter values. Further investigation into the posterior parameter distribution functions reveals a clearer distribution pattern for each parameter. Among them, most of their optimal values can be well defined. Two parameters, I*m* and K_g, have a relatively smooth response surface and therefore are not as identifiable for a good model performance. This is consistent with the results of parameter correlations, that these two parameters are not close-correlated with others. Soil and groundwater parameters like ρ and K_g are more sensitive in the Kielstau models. This may reflect the active soil–groundwater interactions in the Kielstau region because of the near-surface groundwater level and the large fraction of wetland area. A strong negative correlation was found between soil parameters S_{fk} and K_c due to their effect on model performance. The validated hydrograph shows an acceptable simulation uncertainty range (*R* factor < 1) for both catchments, which can bracket the observations most of the time (*P* factor > 50%). Despite the fact that some simulations are less accurate during low-flow periods, the global shape of the hydrograph is reasonably well approximated for both basins, and the 90% uncertainty bounds are always narrow. However, the parameter uncertainty bounds do not always cover the measured data during the validation periods. This indicates that further improvements in the model structure may be required for more accurate predictions in future research.

ACKNOWLEDGEMENTS

The authors thank Kiel University for financial support of this program. We appreciate help from our colleagues at the Institute for the Conservation of Natural Resources,

Kiel University, for access to their research data. We thank Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, for the provision of XitaoXi data. Part of this work is sponsored by the National Research Program (No. 2008CB418106) from the Chinese Ministry of Science and Technology; and by the Key project (No. KZCX1-YW-14–6) from the Chinese Academy of Sciences.

REFERENCES

Abbaspour, K., Johnson, C. & van Genuchten, M. 2004 Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone J.* **3**, 1340–1352.

Arnold, J., Srinivasan, R., Muttiah, R. & Williams, J. 1998 Large area hydrologic modeling and assessment-Part I: model development. *J. Am. Water Resour. Assoc.* **34**, 73–89.

Aronica, G., Bates, P. & Horritt, M. 2002 Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrol. Process.* **16**, 2001–2016.

Beven, K. 2007 Towards integrated environmental models of everywhere: uncertainty, data and modeling as a learning process. *Hydrol. Earth Syst. Sci.* **11**, 460–467.

Beven, K. & Binley, A. 1992 The future of distributed models – model calibration and uncertainty prediction. *Hydrol. Process.* **6**, 279–298.

Beven, K., Smith, P. & Freer, J. 2007 Comment on 'Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology' by Pietro Mantovan and Ezio Todini. *J. Hydrol.* **338**, 315–318.

Blasone, R., Vrugt, J., Madsen, H., Rosbjerg, D., Robinson, B. & Zyvoloski, G. 2008 Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Adv. Water Resour.* **31**, 630–648.

Boyle, D., Gupta, H. & Sorooshian, S. 2000 Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods. *Water Resour. Res.* **36**, 3663–3674.

Cheng, C., Ou, C. & Chau, K. 2002 Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall–runoff model calibration. *J. Hydrol.* **268**, 72–86.

Cheng, C., Zhao, M., Chau, K. & Wu, X. 2006 Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *J. Hydrol.* **316**, 129–140.

Choi, H. & Beven, K. 2007 Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *J. Hydrol.* **332**, 316–336.

Chow, V., Maidment, D. & Mays, L. 1988 *Applied Hydrology.* McGraw-Hill, New York, USA, pp. 585.

Doherty, J. & Johnston, J. 2003 Methodologies for calibration and predictive analysis of a watershed model. *J. Am. Water Resour. Assoc.* **39**, 251–265.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.* **28**, 1015–1031.

Eggemann, G., Sterr, H. & Kuhnt, G. 2001 Geomorphologie Schleswig Holsteins. Ergebnisse aus dem Studienprojekt 'Einführung in die Geomorphologische Detailkartierung' im Jahr 2001. Kiel (unpublished), pp. 140.

Feyen, L., Vrugt, J., Ó Nualláin, B., van der Knijff, J. & De Roo, A. 2007 Parameter optimization and uncertainty assessment for large scale streamflow simulation with the LISFLOOD model. *J. Hydrol.* **332**, 276–289.

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.

Gao, J., Lu, G., Zhao, G. & Li, J. 2006 Watershed data model: a case study of XitiaoXi sub-waterhed, Taihu Basin (in Chinese, with English abstract). *J. Lake Sci.* **18**, 312–318.

Hörmann, G., Zhang, X. & Fohrer, N. 2007 Comparison of a simple and a spatially distributed hydrologic model for the simulation of a lowland catchment in Northern Germany. *Ecolog. Model.* **209**, 21–28.

Krzysztofowicz, R. & Kelly, K. 2000 Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.* **36**, 3265–3277.

Kuczera, G. & Parent, E. 1998 Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *J. Hydrol.* **211**, 69–85.

Laloy, E., Fasbender, D. & Bielders, C. 2010 Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach. *J. Hydrol.* **380**, 82–93.

Lamb, R. & Kay, A. 2004 Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain. *Water Resour. Res.* **40**, 1–13.

Li, L., Xia, J., Xu, C., Chu, J. & Wang, R. 2009 Analyse the sources of equifinality in hydrological model using GLUE methodology. *IAHS-AISH Publ.* **331**, 130–138.

Li, X., Weller, D. & Jordan, T. 2010 Watershed model calibration using multi-objective optimisation and multi-site averaging. *J. Hydrol.* **380**, 277–288.

Madsen, H. 2003 Parameter estimation in distributed hydrological catchment modeling using automatic calibration with multiple objectives. *Adv. Water Resour.* **26**, 205–216.

Montanari, A. & Brath, A. 2004 A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour. Res.* **40**, 1–11.

Nash, J. & Sutcliffe, J. 1970 River flow forecasting through conceptual models. Part I: A discussion on principles. *J. Hydrol.* **10**, 282–290.

Schmidtke, K. 1995 *Land im Wind, Wetter und Klima in Schleswig-Holstein.* Wachholtz Verlag, Neumuenster, pp. 116.

Schmitz, O., Karssenberg, D., van Deursen, W. P. A. & Wesseling, C. 2009 Linking external components to a spatio-temporal modeling framework: coupling MODFLOW and PCRaster. *Environ. Modell. Softw.* **24**, 1088–1099.

Schuol, J. & Abbaspour, K. 2006 Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa. *Adv. Geosci.* **9**, 137–143.

Sponagel, H. 2005 *Bodenkundliche Kartieranleitung*. ADHOCARBEITSGRUPPE BODEN der Staatlichen Geologischen Dienste und der Bundesanstalt für Geowissenschaften und Rohstoffe. 5.verbesserte und erweiterte Auflage, Hannover, pp. 438.

Sumner, N., Fleming, P. & Bates, B. 1997 Calibration of a modified SFB model for twenty-five Australian catchments using simulated annealing. *J. Hydrol.* **197**, 166–188.

Tang, Y., Reed, P., Wagener, T. & Werkhoven, K. 2007 Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrol. Earth Syst. Sci.* **11**, 793–817.

Thiemann, M., Trosset, M., Gupta, H. & Sorooshian, S. 2001 Bayesian recursive parameter estimation for hydrological models. *Water Resour. Res.* **37**, 2521–2535.

Uhlenbrook, S. & Sieber, A. 2005 On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model. *Environ. Modell. Softw.* **20**, 19–32.

Van Deursen, W. P. A. 1995 Geographical Information Systems and Dynamic Models: development and application of a prototype spatial modeling language. *Netherlands Geographic Studies*. Koninklijk Nederlands Aardrijkskundig Genntschap, Utrecht, pp. 198.

Vrugt, J., Gupta, H., Bastidas, L., Bouten, W. & Sorooshian, S. 2003a Effective and efficient algorithm for multi-objective optimization of hydrologic models. *Water Resour. Res.* **39**, 1214–1232.

Vrugt, J., Gupta, H., Bouten, W. & Sorooshian, S. 2003b A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* **39**, 1201–1213.

Wan, R., Yang, G., Li, H. & Yang, L. 2007 Simulating flood events in mesoscale watershed: a case study from River Xitiaoxi watershed in the upper region of Taihu Basin. *J. Lake Sci.* **19**, 170–176.

Wang, Q. 1997 Using genetic algorithms to optimize model parameters. *Environ. Modell. Softw.* **12**, 27–34.

Wesseling, C., Karssenberg, D., Burrough, P. & Van Deursen, W. P. A. 1996 Integrated dynamic environmental models in GIS: The development of a Dynamic Modeling language. *Trans. GIS* **1**, 40–48.

Xu, L., Zhang, Q., Li, H., Viney, N., Xu, J. & Liu, J. 2007 Modeling of surface runoff in Xitiaoxi catchment, China. *Water Resour. Manage.* **21**, 1313–1323.

Yapo, P., Gupta, H. & Sorooshian, S. 1998 Multi-objective global optimization for hydrological models. *J. Hydrol.* **204**, 83–97.

Zak, S. & Beven, K. 1999 Equifinality, sensitivity and predictive uncertainty in the estimation of critical loads. *Sci. Total Environ.* **236**, 191–214.

Zhang, X. Y., Hörmann, G. & Fohrer, N. 2007 The effects of different model complexity on the quality of discharge simulation for a lowland catchment in northern Germany. In Heft 20.07 'Einfluss von Bewirtschaftung und Klima auf Wasser- und Stoffhaushalt von Gewässern' (2007), Band 2, *Forum für Hydrologie und Wasserbewirtschaftung* 111–114.

Zhang, X. Y., Hörmann, G. & Fohrer, N. 2008 An investigation of the effects of model structure on model performance to reduce discharge simulation uncertainty in two catchments. *Adv. Geosci.* **18**, 31–35.

Zhang, X. Y., Hörmann, G. & Fohrer, N. 2009 Hydrologic comparison between a lowland catchment (Kielstau, Germany) and a mountainous catchment (XitaoXi, China) using KIDS model in PCRaster. *Adv. Geosci.* **21**, 125–130.

Zhang, X. Y., Hörmann, G., Gao, J. F. & Fohrer, N. 2011 Structural uncertainty assessment in a discharge simulation model. *Hydrol. Sci. J.* **56**, 854–869.

First received 29 June 2011; accepted in revised form 16 December 2011. Available online 22 March 2012