



Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE Model for a large watershed in northern China

S. Wang¹, Z. Zhang², G. Sun³, P. Strauss⁴, J. Guo², Y. Tang², and A. Yao²

¹MOE Key Laboratory of Regional Energy and Environmental Systems Optimization, Resources and Environmental Research Academy, North China Electric Power University, Beijing 102206, China

²Key Laboratory Soil and Water Conservation and Desertification Combating, Ministry of Education, College of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China

³Eastern Forest Environmental Threat Assessment Center, Southern Research Station, USDA Forest Service, Raleigh, NC 27606, USA

⁴Federal Agency for Water Management, 3252 Petzenkirchen, Austria

Correspondence to: Z. Zhang (zhqzhang@bjfu.edu.cn)

Received: 31 March 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 2 May 2012

Revised: 31 October 2012 – Accepted: 1 November 2012 – Published: 6 December 2012

Abstract. Model calibration is essential for hydrologic modeling of large watersheds in a heterogeneous mountain environment. Little guidance is available for model calibration protocols for distributed models that aim at capturing the spatial variability of hydrologic processes. This study used the physically-based distributed hydrologic model, MIKE SHE, to contrast a lumped calibration protocol that used streamflow measured at one single watershed outlet to a multi-site calibration method which employed streamflow measurements at three stations within the large Chaohe River basin in northern China. Simulation results showed that the single-site calibrated model was able to sufficiently simulate the hydrographs for two of the three stations (Nash-Sutcliffe coefficient of 0.65–0.75, and correlation coefficient 0.81–0.87 during the testing period), but the model performed poorly for the third station (Nash-Sutcliffe coefficient only 0.44). Sensitivity analysis suggested that streamflow of upstream area of the watershed was dominated by slow groundwater, whilst streamflow of middle- and down- stream areas by relatively quick interflow. Therefore, a multi-site calibration protocol was deemed necessary. Due to the potential errors and uncertainties with respect to the representation of spatial variability, performance measures from the multi-site calibration protocol slightly decreased for two of the three stations, whereas it was improved greatly for the third station. We concluded that multi-site calibration protocol reached a compromise in term of model performance for the three stations, reasonably representing the hydrographs of all three

stations with Nash-Sutcliffe coefficient ranging from 0.59–0.72. The multi-site calibration protocol applied in the analysis generally has advantages to the single site calibration protocol.

1 Introduction

The importance of spatial variability of land surface characteristics is widely recognized in understanding the physical/hydrological, biological, and other related process in watersheds (Becker and Braun, 1999). It is important to take into account the spatial variability when modeling watershed hydrology and understanding watershed hydrological processes (Beven, 2001; Blöschl et al., 1995; Merz and Bárdossy, 1998; Anquetin et al., 2010). This is particularly true for modeling large-scale watersheds that have more diverse hydrological conditions than small watersheds (Sivapalan, 2003). Also, mountainous watersheds usually exhibit great heterogeneity in geology, topography, soil, vegetation, and climate (e.g. Weingartner et al., 2003; Gurtz et al., 1999; Smerdon et al., 2009), thus they pose significant challenges in hydrologic modeling.

Numerous hydrological models have been developed in either a distributed or a lumped framework to simulate watershed hydrologic processes. Lumped hydrologic models are usually precluded in the case of applications to un-gauged watersheds as a result of the significant changes between

watershed conditions (Knudsen et al., 1986; Sahoo et al., 2006). In contrast, distributed hydrological model provides a comprehensive approach for characterizing spatial variability of watershed, enabling the spatial variability of watershed well characterized (Refsgaard, 1997) by specifying data and parameters for a network of grid of points. During the past two decades, various distributed hydrologic models, such as the SWAT model (Cao et al., 2006; Zhang et al., 2009; White and Chaubey, 2005), HBV model, ModSpa model (Moussa et al., 2007), Wetspa model (Shafii and De Smedt, 2009) and the GWLF model (Li et al., 2010) have emerged. MIKE SHE, as the first generation of distributed hydrologic model, has shown its great appeal for a wide range of hydrologic applications (Graham and Butts, 2005; Lu et al., 2006; Zhang et al., 2008; Dai et al., 2010; Wijesekara et al., 2012), especially for the watershed of large spatial heterogeneities.

Nevertheless, when applying distributed hydrologic model, it is usually required to carry out a careful model calibration in order to obtain an internal consistency of results. Previous studies suggested that hydrologic model calibrated only against the discharge measurements at the watershed outlet can not perform well for the internal variables simulation (such as the groundwater recharge, unsaturated water content, and water table level) (e.g. Ambroise et al., 1995; Refsgaard, 1997). This calls for a rigorous calibration and validation procedure for distributed hydrological models (Freer et al., 2003; Moussa et al., 2007).

Researchers have long recognized the importance of a multi-objective framework (multi-site, multi-variables, as well as multi-criteria) for distributed hydrologic modeling (e.g. Ambroise et al., 1995; Andersen et al., 2001; Khu et al., 2008; Dai et al., 2010). Bergström et al. (2002) suggested that the model calibrated against more measured internal variables rather than streamflow only can greatly increase confidence in the physical relevance of the model. Vázquez et al. (2008) adopted a multi-criteria protocol which included statistical, analytical and visual criteria to calibrate the model. They also suggested that multi-criteria calibration protocol enhanced the physical consistency of model prediction. Khu et al. (2008) stressed the benefits of multi-objective strategy to decision support framework. A summary of the merits of multi-objective framework is found in Madsen (2003). Generally, in addition to reduce the uncertainty and modeling bias (Kuczera and Mroczkowski, 1998; Dai et al., 2010), it was believed that a multi-objective strategy better constrain the calibration process, and is able to unlock the equifinality of distributed hydrological models to a certain degree (e.g. Mroczkowski et al., 1997; Seibert et al., 2000; Andersen et al., 2001; Bergström et al., 2002; Khu et al., 2008; Dai et al., 2010). Khu et al. (2008) applied multi-objective calibration approach to calibrate MIKE SHE model and tested on the Danish Karup catchment. And consequently, parameter values were well defined in that analysis. Andersen et al. (2001) employed MIKE SHE to simulate watershed hydrology as well. Although the nested multi-site

measurements were employed in that research, the model simulated the distributed results well. Feyen et al. (2000) used multi-site measurements for MIKE SHE model evaluation. However, the multi-site measurements were only used for model validation rather than calibration.

This study tested the MIKE SHE model in a large watershed of northern China, Chaohe watershed using a multi-site calibration protocol. The watershed has a mountainous topography, being a headwater of the Miyun Reservoir that supplies near half of the drinking water for Beijing (Jia and Cheng, 2002; Yang et al., 2007). About 60% inflows of Miyun Reservoir was originated from the Chaohe Watershed (Sun et al., 2008). However, due to the increased population and climate change, flow to the Miyun Reservoir has decreased greatly over the past decades (e.g. Li and Li, 2008; Sun et al., 2008; Ma et al., 2010), especially since 1999. Successive dry years have slowed the average annual inflow to only $2.51 \times 10^8 \text{ m}^3$ (Li and Li, 2008). There is an urgent need to understand the hydrologic processes of the watershed and take an adaptive watershed management to cope with the emerging water resource issues in the watershed.

The objectives of this study are to (1) assess the applicability of MIKE SHE model in the large-scale watershed of northern Chian, (2) understand the spatial controls on watershed hydrology of Chaohe watershed, and (3) examine the benefits of multi-site calibration protocol for modeling analysis.

2 Methods

2.1 Watershed characteristics

The Chaohe watershed has an area of around 4854 km² with elevation ranging from 159 m to 2218 m a.s.l. (Fig. 1) with 80% classified as mountainous topography. Two large mountain ranges, Yanshan Mountain and Yinshan Mountain intersect with the watershed. The substrate of the watershed is made of granite, gneiss, and lime rock, which is mainly overlain by brown soil with varied depth. The upstream area of the watershed is adjacent to the Inner-Mongolia Plateau, and thus characterized by deep soil, whereas the middle- and down-stream area is characterized by thin soil. Due to a severe weathering process, the soil across the watershed is represented with a high degree of gravel content. Temperate continental monsoon climate dominates the study region. The average long-term annual precipitation is around 494 mm, 80% of precipitation falling from June to September. The watershed has a high vegetation coverage (80%). Land use is dominated by grassland, shrubland and mixed deciduous broadleaf and evergreen conifer forests. In addition, cropland, residential area, and bare area account for a small percent of the watershed.

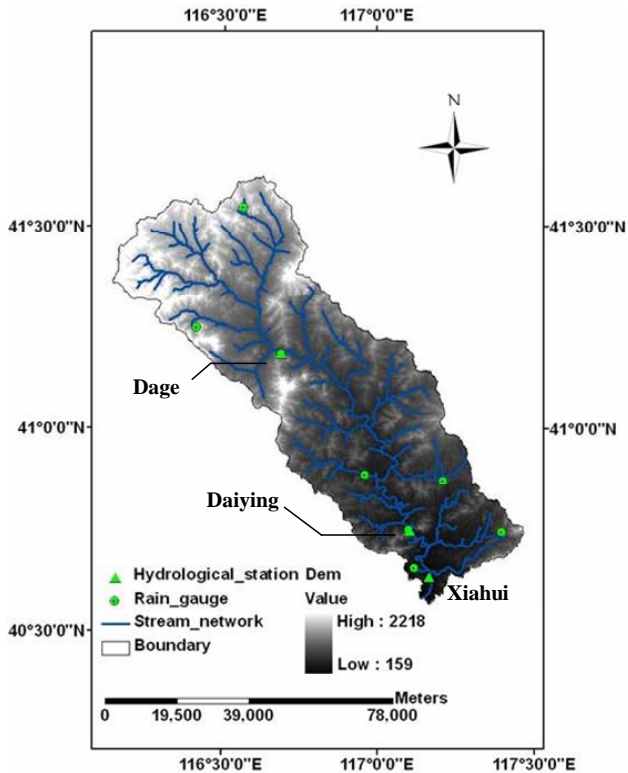


Fig. 1. Topography and the distribution of hydroclimatic stations of Chaohe watershed.

We have acquired daily rainfall records from seven rain gauges (Fig. 1) and daily streamflow records from three hydrological stations of the watershed, i.e. the outlet station (Xiahui) and the other two internal hydrologic stations (i.e. Daiying and Dage) (Fig. 1). Meteorological data is not registered in the watershed, therefore, data records of national meteorological base stations around the watershed were introduced. Potential evapotranspiration was estimated by the Penman-Monteith equation (Allen et al., 1998), and arithmetic average statistic of the estimations was used for modeling analysis. DEM of the watershed (resolution of 30 m) was acquired from International Scientific & Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences (<http://datamirror.csdb.cn>), and land use of 1999 was interpreted according to remote sensing TM images, which were used to represent surface condition of the watershed for the study period. In addition, soil dataset of the watershed was derived from HWSD (The Harmonized World Soil Database). According to FAO-90 classification system, ten soil units were found in the watershed, Calcaric Cambisols (CC_Soil), Haplic Luvisols (HL_Soil) and Calcic Luvisols (CL_Soil) accounting for much of the watershed.

2.2 Model construction and parameterization

MIKE SHE is a fully distributed physically-based hydrological model. It covers the major processes of the hydrologic cycle and includes process models for evapotranspiration, overland flow, unsaturated flow, groundwater flow, and channel flow and their interactions. Detailed model description can refer to the literature (e.g. Refsgaard and Storm, 1995; Graham and Butts, 2005). We have applied MIKE SHE in arid regions in northern China (Zhang et al., 2008).

MIKE SHE uses a network of regular grids to discretize the horizontal plane of watershed. Choice of grid size depends on several factors, such as the degree of heterogeneity of the hydrological parameters and the boundary conditions, and the extent of the flow domain defined by computational limitations (Feyen et al., 2000). Considering the large scale modeling domain of Chaohe watershed, grid size of $2\text{ km} \times 2\text{ km}$ was used in the analysis, which was comparable to the grid size of the researches of Andersen et al. (2001) and Khu et al. (2008). Our preliminary analysis suggested that the application of the grid size of $2\text{ km} \times 2\text{ km}$ was able to greatly reduce computation time but maintaining reasonable representation of the long-term streamflow variation. According to field investigation and experience, we have specified seasonal dynamic variation of each land use with respect to both LAI and rooting depth, however, vegetation growth was ignored in the model. Overland flow is calculated by solving the diffusive wave approximations of the Saint-Venant equations. Parameters of Manning's number (M) and detention storage (D) were all subject to model calibration, and both were specified with uniform spatial distributions. Unsaturated flow was simulated with simplified gravity-flow procedure, which demonstrated the capability of reducing the computation time, while preserving the modeling accuracy to a certain degree. Given that most area of the watershed was of mountainous topography with shallow soil profile, groundwater table for lower boundary of the unsaturated zone (Bnd_UZ) was specified with a uniform depth of 1.5 m. Parameters of saturated hydraulic conductivity in unsaturated zone (K_s) in three dominant profiles were subject to model calibration, whilst parameters of the other profiles were unchanged in calibration process. In order to keep the number of free calibrated parameters as small as possible, a dependent relationship was further specified between the K_s of the three dominant soil profiles (i.e. Calcaric Cambisols, Haplic Luvisols, and Calcic Luvisols). As a result, only parameters of one profile (Calcaric Cambisols) have freedom in calibration process, whereas parameters of the other two (i.e. Haplic Luvisols and Calcic Luvisols) were assigned with a dependent relationship. Due to the lack of data on geo-hydrological information of the watershed, saturated flow of the watershed was simulated by using simple linear reservoir method which accounts for the interflow, the baseflow and the percolation from interflow reservoir. Parameters of specific yield (S_p) and time constant (T) in both interflow and

baseflow reservoirs were all subject to model calibration. Except to the initial values of K_s which was specified according to the estimation of Pedotransfer function by using soil physical properties, the other parameters were initialized according to literature or experience.

2.3 Model calibration, validation, and sensitivity analysis

Parameter adjustments during the calibration process were carried out manually by trial and error method. Two kinds of calibration protocol were considered (i.e. the single-site calibration and multi-site calibration). In the first protocol, the model was calibrated against the discharge measurement of watershed outlet (i.e. Xiahui station). Streamflow data for 1991–1995 and 1996–1999 were used for model calibration and validation, respectively. Both time periods experienced wet, dry and normal climate, representing a wide range of hydrologic conditions. In order to test the capability of the model in simulating internal hydrologic variables (e.g. Refsgaard, 1997; Feyen et al., 2000), a multi-site validation test was introduced as well, in which the previously calibrated model was tested against the discharge measurements of Daiying and Dage stations for 1991 to 1999 for model validation. Table 1 has presented the calibrated parameter set by single-site calibration protocol.

In the second protocol, the model was calibrated against multi-site discharge measurements simultaneously (i.e. Xiahui, Daiying, and Dage). Though few of authors argued that multi-site model calibration should employ independent multi-site measurements (Migliaccio and Chaubey, 2007), the fact is that independent multi-site measurements are rarely available for model testing, which causes that, in most of case, it remains a common practice of using nested multi-site measurements (e.g. Andersen et al., 2001; Moussa et al., 2007). After the model was calibrated using the first protocol, the model was further calibrated against the other two site discharge measurements simultaneously (i.e. Daiying and Dage). The same parameter set as that of single-site protocol was subject to model calibration. The aim of the process was looking for suitable parameter values which were able to derive satisfied modeling results for all hydrological stations. As such, the model was run for 1991 to 1995 for model calibration and 1996 to 1999 for model validation, respectively. Both single-site model calibration and multi-site model calibration was initialized by running the model for 1990 for warming up exercise.

In order to examine the differences of hydrological processes between the different areas of the watershed, sensitivity analysis was carried out by manually altering parameter values in the parameter range which was specified by referring to the literatures and our previous modeling experience. We reckoned that, sensitivity analysis not only enhances model calibration greatly (Beven, 2001; Xevi et al., 1997), but also capable of revealing different hydrological

processes between different areas of watershed. The influence of each parameter on hydrological process was assessed according to the variation of model performance.

2.4 Model performance criteria

Both correlation coefficient (R) and Nash-Sutcliffe coefficient (EF) (Nash and Sutcliffe, 1970) were employed for evaluating the goodness of model performance (Eqs. 1 and 2). Correlation coefficient indicates the strength of a linear relationship between observed and calculated discharge, whilst Nash-Sutcliffe Coefficient measures the ability of the model to simulate variation of the hydrographs for a particular river gauge station. Optimal values for R and EF are all 1. According to the previous studies (e.g. Henriksen et al., 2003; Moussa et al., 2007), four performance levels were defined for EF and R (see Table 2). As EF gives more weight for peak flows, to give an equal weight for all the maximum and minimum, a mean residual (simulation minus measurement) (ME) was calculated after a logarithmic transformation of the discharges were taken, which corresponds to the model bias of the results (El-Nasr et al., 2004).

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

$$EF = 1 - \frac{\sum_i (Q_{s,i} - Q_{o,i})^2}{\sum_i (Q_{o,i} - \bar{Q}_o)^2} \quad (2)$$

$$ME = \log(Q_{s,i}) - \log(Q_{o,i}). \quad (3)$$

In which, $Q_{s,i}$ and $Q_{o,i}$ represents the simulated and observed daily discharge for day i ($\text{m}^3 \text{s}^{-1}$), respectively; \bar{Q}_o is the mean of the observed discharge in test period ($\text{m}^3 \text{s}^{-1}$).

3 Results and discussions

3.1 Single-site model calibration and validation

Generally, the model underestimated the mean runoff of Xiahui station with ME of -0.21 and $-0.41 \text{ m}^3 \text{ s}^{-1}$, for the calibration and validation periods, respectively (Table 3). The simulated streamflow showed a somewhat flashier response than the observed hydrograph (see Fig. 2a), especially during the year of 1999. However, it generally represented the dynamic variation of the streamflow with acceptable EF (0.72 and 0.75) and R (0.85 and 0.87) for the calibration and validation period, respectively. According to the performance criteria (Table 2), both EF and R indicated that the model generally performed well in simulating hydrograph of Xiahui station.

Table 1. The calibrated model parameter by single-site calibration protocol.

Module	Parameter	Unit	Initial value	Final value
Overland flow (OL)	Manning number (M)	$M^{1/3} s^{-1}$	20	25.5
	Detention storage (D)	mm	2	6
Unsaturated zone (UZ)	K_s for CC_Soil (K_{s-1})	$m s^{-1}$	0.68×10^{-7}	2×10^{-7}
	K_s for HL_Soil (K_{s-2})	$m s^{-1}$	$0.92 \cdot K_{s-1}$	$0.92 \cdot K_{s-1}$
	K_s for CL_Soil (K_{s-3})	$m s^{-1}$	$1.543 \cdot K_{s-1}$	$1.543 \cdot K_{s-1}$
Saturated flow (SZ)	Specific yield for interflow ($S_{pinterflow}$)	–	0.1	0.09
	Time constant for interflow ($T_{interflow}$)	Day	5	5.25
	Time constant for percolation ($T_{percolation}$)	Day	5	2.61
	Specific yield for baseflow ($S_{pbaseflow}$)	–	0.5	0.62
	Time constant for baseflow ($T_{baseflow}$)	Day	50	72.21

Table 2. Performance criteria for model evaluation.

Performance indicator	Excellent	Good	fair	Poor
EF	> 0.85	0.65–0.85	0.5–0.65	< 0.5
R	> 0.95	0.85–0.95	0.85–0.75	< 0.75

Table 3. Model performance of MIKE SHE in Chaohe watershed*.

Station	Period	Single site calibration protocol			Multi-site calibration protocol		
		R	EF	ME	R	EF	ME
Xiahui	1991–1995	0.85	0.72	–0.21	0.85	0.69	–0.47
	1996–1999	0.87	0.75	–0.41	0.85	0.72	–0.67
Daiying	1991–1995	0.81	0.65	–0.31	0.82	0.61	–0.57
	1996–1999	0.86	0.73	–0.50	0.84	0.67	–0.76
Dage	1991–1995	0.81	0.44	–0.46	0.80	0.59	–0.72
	1996–1999	0.80	0.52	–0.78	0.77	0.53	–1.01

* R means the correlation coefficient, EF denotes the Nash-Sutcliffe Coefficient, whilst ME represents the mean residual of the simulations. Details explanation on R , EF, and ME were given in the section of “Methods”.

The satisfactory model performance as indicated by EF for the Xiahui station was mostly due to the model’s ability to match peak flows and median flows. EF statistics give more weights for peak flow simulation (Henriksen et al., 2003). Thus, the model runs for large stormflow events can obtain an acceptable EF and R -values much easier than runs for the dry periods with low flow. We found that the model generally underestimated the streamflow of around 1 to $10 m^3 s^{-1}$, whilst over-predicted flow of lower than $1 m^3 s^{-1}$ (Fig. 2). The systemic underestimation for the flows of 1 to $10 m^3 s^{-1}$ suggested that there existed errors with ground water simulation. Refsgaard (1997) reported that underestimation of baseflow could be mainly explained by the biased representation of internal groundwater divide and boundary conditions.

More detailed analysis on modeling errors is given in the section of errors analysis.

The modeling results by single-site calibration protocol for Daiying and Dage stations were presented in Fig. 2b and c. The model underestimated the mean runoff of the two stations as well. The model bias (ME) of Daiying was -0.31 and -0.50 in the calibration and validation period, respectively, and larger for the Dage station (ME being -0.46 and -0.78) (see Table 3). The model did not accurately simulate flow of lower than $10 m^3 s^{-1}$ for the other two stations. However, the model performed reasonably as judged by the hydrograph for the Daiying station. The EF and R -values for Daiying were similar to those for Xiahui station with acceptable EF (0.65 and 0.73) and R (0.81 and 0.86) for the testing period (Table 3). The performance for Dage station was fair in the validation period (EF=0.52), and it was poor in the calibration period (EF=0.44). The R of Dage station was 0.81 and 0.80 for the calibration and validation period, respectively.

The reasonable model performance of Daiying, as judged by EF and R , was mainly due to the fact that Daiying station was very close to Xiahui, and both watersheds almost drain the same area with very similar hydrological processes. Thus it was easier to obtain a reasonable modeling result even though the model was only calibrated against the discharges of Xiahui station. The fair model behavior of Dage station was partly due to the underestimated soil water storage capacity in the up-stream area. Also, the soil depth of the watershed varied greatly across the watershed (Li and Li, 2008). The middle and downstream area was usually characterized by a severe weathering process and a mountainous topography. Whilst the northern part of the watershed is adjacent to the Inner Mongolia Plateau, and commonly characterized by deep loess profiles with high soil water storage capacity.

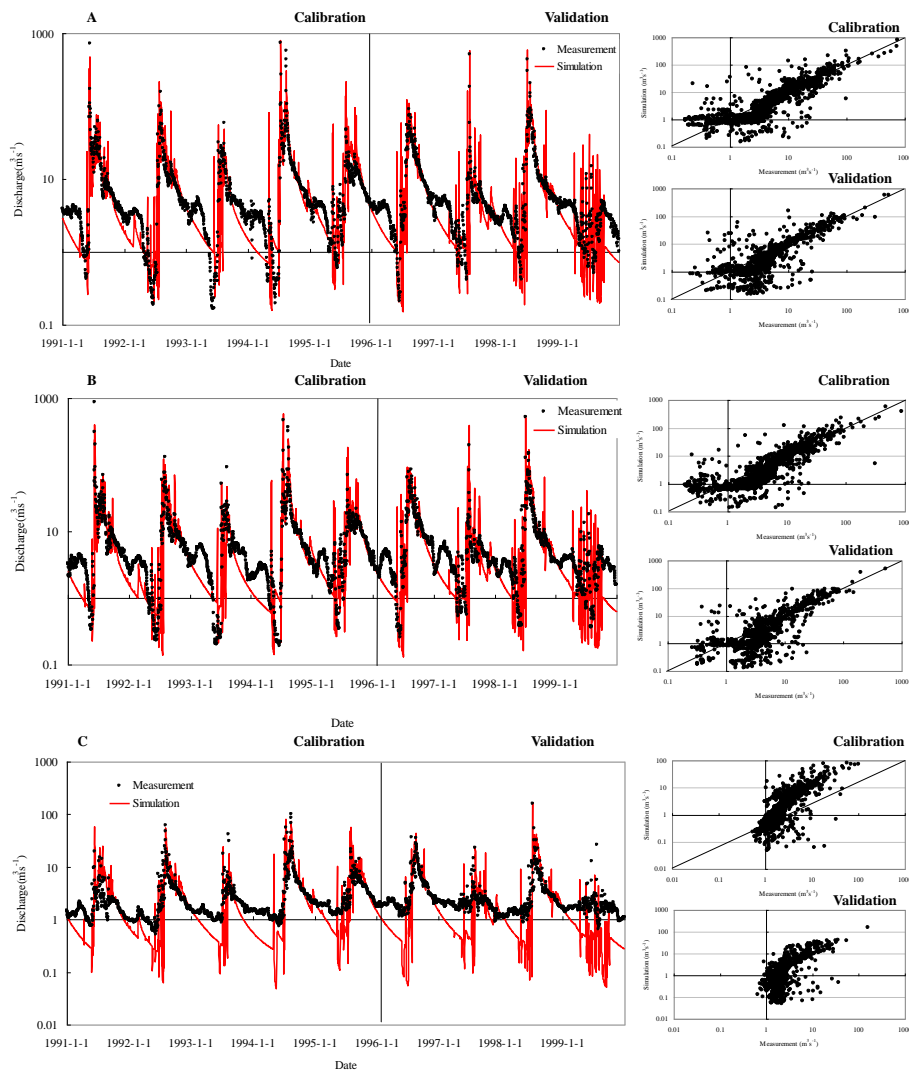


Fig. 2. Comparison of model simulation by single-site calibration procedure with measurement for the testing period. (A) Xiahui station; (B) Daiying station; (C) Dage station.

This caused that much of water for the upstream area was able to be stored in the unsaturated zone, and recharging the groundwater and discharging the river flow subsequently. On the other hand, due to the shortage of detailed information, the model was specified with a thin soil depth with a uniform value of 1.5 m across the watershed, even though spatially varied soil hydraulic properties were taken into account. These resulted in the simulated water storage capacity of the upstream area to be much lower than the reality was, and consequently the simulated runoff larger than the observations, especially during the wet period. The scattering point of Dage station in the calibration period well exhibited the over-prediction of discharge of this station (see Fig. 2c).

3.2 Multi-site model calibration and validation

The unsatisfied modeling results of Dage station by single-site calibration protocol called for a multi-site calibration

(Fig. 3). It was found that, by only changing parameter values of K_s from 2×10^{-6} to $4 \times 10^{-6} \text{ m s}^{-1}$, the streamflow simulation for the upstream area of the watershed (i.e. Dage station) was improved greatly with EF of 0.59 in the calibration period, whilst the EF for both Xiahui and Daiying stations slightly decreased in the testing periods, ranged from 0.61 to 0.72 (see Table 3). Streamflow simulation of 1 to $10 \text{ m}^3 \text{ s}^{-1}$ was not improved for the three stations, the recession limbs for the three stations worse than that of the single-site calibration protocol, and the mean runoff remained underestimated by the model, ME even lower than that of the single-site calibration (ranged from -1.01 to -0.47 , see Table 3). We assumed that both the worse model behavior with respect to recession limb simulations and the underestimated mean runoff was partly associated with the increase of K_s in unsaturated zone. The R-values for the three stations varied slightly ranging from 0.77 to 0.85 (Table 3).

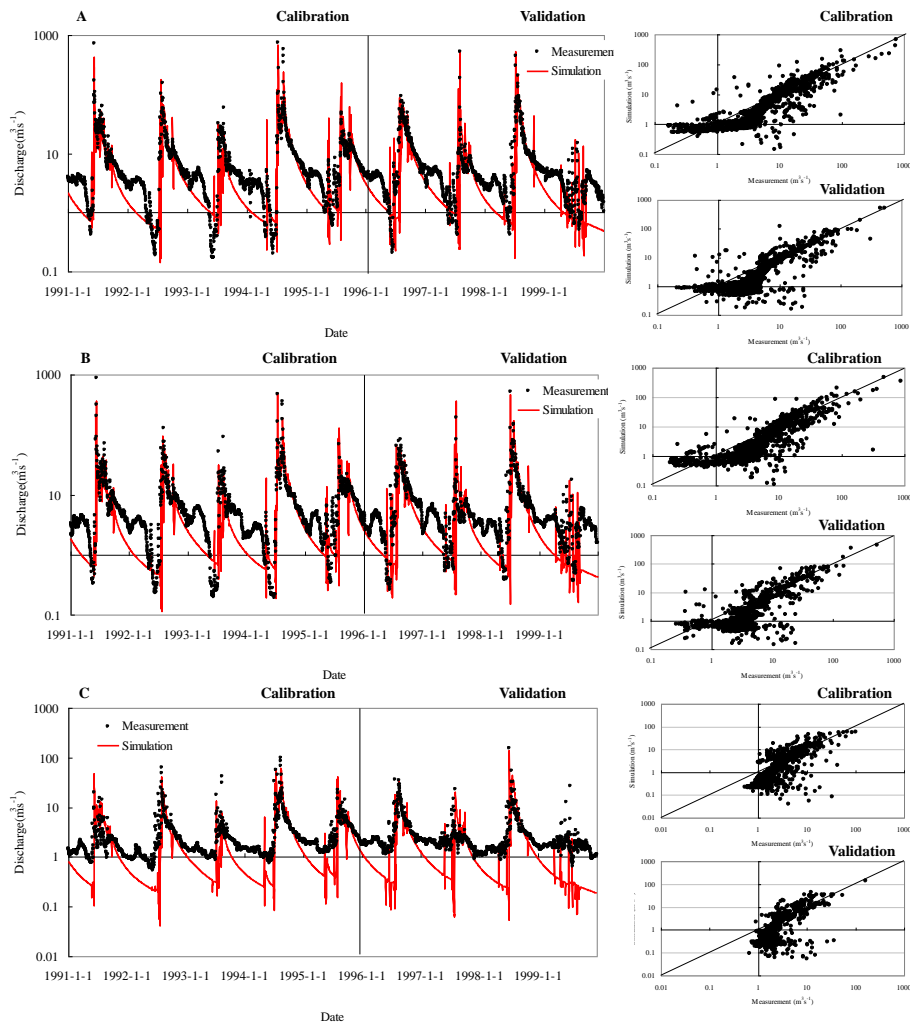


Fig. 3. Comparison of model simulation by multi-site calibration procedure with measurement for the testing period. (A) Xiahui station; (B) Daiying station; (C) Dage station.

The higher EF and better model performance (Fig. 3c) for the Dage station for the calibration period was mainly because that the pseudo lower soil water storage capacity for the up-stream area was improved and that EF was generally enhanced, especially in 1991, 1994 and 1995 with high rainfall of 559, 622, and 499 mm, respectively (Fig. 4). The poor model performance for Xiahui and Daiying stations was partly because that a pseudo higher water storage capacity was specified for the downstream area when K_s was changed from 2×10^{-6} to $4 \times 10^{-6} \text{ m s}^{-1}$. The changes in soil parameters caused that more water was likely to be stored in the unsaturated zone, and both mean runoff and peak flows simulation were reduced and decreased EF, R , and ME accordingly.

3.3 Sensitivity analysis

All parameters considered but S_p , M and $C3$ exhibited certain degree of influence on ME, but few parameters had influence on EF.

The two parameters, $T_{\text{interflow}}$ and $T_{\text{percolation}}$, controlled the discharge of relative quick interflow, and directly affected peak flow simulations. With $T_{\text{interflow}}$ decreased from 5 to 0, i.e., reducing time for water flowing through the interflow reservoir, the model performance for the Dage station was decreased greatly with the percentage change in EF as high as 25%. However, the EF was improved for Xiahui and Daiying stations (Fig. 5a) with the percentage change around 11%. The increase in time for water seeping down into baseflow reservoir (i.e. the increase in $T_{\text{percolation}}$) induced a decrease in EF by 13% for Dage station but a slight increase by 5% for Xiahui and Daiying stations (Fig. 5b). As such, both $S_{p_{\text{baseflow}}}$ and T_{baseflow} influenced the discharge of groundwater and indirectly affected peak flow simulations as

Table 4a. Percentage change in performance measures due to the change in parameters of Chaohe watershed.

Module	Parameter*	Unit	Xiahui			Daiying			Dage		
			ME(%)	R (%)	EF (%)	ME (%)	R (%)	EF(%)	ME (%)	R (%)	EF (%)
SZ	Specific yield for interflow ($S_{pinterflow}$)										
	0.3		0	-1	-3	0	-1	-3	0	-3	-6
	0.2		0	-1	-2	0	-1	-2	0	-1	-3
	0.1	-	0	0	0	0	0	0	0	0	-1
	<i>0.09</i>		-	-	-	-	-	-	-	-	-
	0.01		0	0	1	0	0	1	0	1	2
	Time constant for interflow ($T_{interflow}$)										
	15		-2	-2	-5	-2	-1	-4	-1	-3	-2
	10		-1	-1	-3	-1	0	-3	-1	-2	-1
	5.25	Day	-	-	-	-	-	-	-	-	-
	1		6	2	8	5	1	7	4	3	-9
	0.1		13	3	11	11	1	9	10	4	-24
	Time constant for percolation ($T_{percolation}$)										
	15		7	2	6	6	0	4	5	1	-12
	10		5	2	5	4	0	4	4	1	-8
	2.61	Day	-	-	-	-	-	-	-	-	-
	1		-2	-1	-4	-2	0	-3	-1	-1	0
	0.1		-3	-2	-7	-3	-1	-5	-2	-3	-2
	Specific yield for baseflow ($S_{pbaseflow}$)										
	0.1		23	-1	0	19	-1	0	16	-1	-11
0.2		19	0	0	16	-1	0	13	-1	-6	
0.4	-	11	0	0	9	0	0	7	0	-2	
<i>0.62</i>		-	-	-	-	-	-	-	-	-	
0.8		-13	0	-1	-10	0	-1	-8	0	1	
Time constant for baseflow ($T_{baseflow}$)											
20		20	0	0	17	-1	0	14	-1	-7	
40		13	0	0	11	0	0	9	0	-3	
60	Day	5	0	0	4	0	0	4	0	-1	
72.21		-	-	-	-	-	-	-	-	-	
80		-4	0	0	-3	0	0	-3	0	0	
OL	Manning number (M)										
	10		-2	1	0	-2	-1	-2	-1	0	-1
	25.5		-	-	-	-	-	-	-	-	-
	30	$M^{1/3} s^{-1}$	1	0	0	1	0	0	0	0	0
	40		0	0	0	0	0	0	0	0	0
	50		2	0	0	2	0	1	1	0	0
	Detention storage (D)										
	2		-11	-1	1	-9	0	3	-7	1	-2
	4		-5	0	1	-4	0	1	-3	0	0
	6	mm	-	-	-	-	-	-	-	-	-
8		5	0	-1	4	0	-2	3	-1	-1	
10		8	0	-2	6	0	-3	4	-2	-1	

* The figures in bold and italic are the parameter values applied in multi-site calibration protocol. The percentage changes in performance measures is estimated according to $(S'-S) \cdot 100/S$, in which S' means the simulation results of changed parameter values, and S is the basic results of multi-site calibration protocol.

well. When reducing the volume of water released by aquifer (i.e. altering $S_{pbaseflow}$ from 0.62 to 0.1), the EF of Dage station decreased by 11 %, whilst it was almost no change for Xiahui and Daiying stations (Table 4). The decrease in time for water flowing through the baseflow reservoir (i.e. changing $T_{baseflow}$ from 72.21 to 20) caused a decrease in EF by 7 % at Dage, but no influence for Xiahui and Daiying stations (Table 4). The distinct model behaviors among three stations actually indicated that the hydrological processes were different between Dage and the other two stations. A

slow groundwater release mechanism might play a more important role in controlling hydrological process of up-stream area, and the relative quick interflow might be more relevant for hydrological process of the down-stream area.

Parameters of UZ (unsaturated zone) have significant influence on EF-values. Higher K_s -value represents quickly downward water movement and recharging base flow reservoir subsequently, whilst higher Bnd_UZ-value increased the volume of water stored in unsaturated zone. Both of the changes in K_s and Bnd_UZ would affect the peak flow

Table 4b. Continued.

Module	Parameter*	Unit	Xiahui			Daiying			Dage		
			ME(%)	R (%)	EF (%)	ME (%)	R (%)	EF(%)	ME (%)	R (%)	EF (%)
UZ	Saturated water conductivity (K_{s-1})										
	0.000001	m s^{-1}	12	-5	-18	8	-10	-24	4	-7	-19
	6×10^{-7}		16	-3	-9	13	-4	-11	9	-5	-12
	4×10^{-7}		-	-	-	-	-	-	-	-	-
	6×10^{-8}		-180	-3	-11	-149	-3	-5	-121	-4	-181
	1×10^{-8}		-100	-6	-27	-82	-6	-23	-65	-5	-281
	Lower UZ boundary (Bnd.UZ)										
	-0.5	m	407	1	-19	335	-5	-37	270	-6	-654
	-1		9	1	3	8	0	2	6	0	-4
	-1.5		-	-	-	-	-	-	-	-	-
	-2		31	-10	-25	26	-5	-21	23	-14	-28
	-3		143	-14	-34	118	-7	-29	101	-29	-64
	ET	Cint									
0.5		mm	2	0	-1	2	0	-1	2	-1	-1
0.1			-5	0	0	-4	0	0	-3	0	-1
0.05			-5	0	0	-4	0	0	-3	0	-1
0.01			-2	0	0	-1	0	0	-1	0	0
0.005			-	-	-	-	-	-	-	-	-
C1											
0.1		-	-37	0	1	-30	0	1	-25	-1	-7
0.3			-	-	-	-	-	-	-	-	-
0.5			6	0	0	5	0	0	4	0	1
0.8			7	0	0	6	0	0	5	0	1
1			8	0	0	6	0	0	5	0	1
C2											
0.1		-	-1	0	0	-1	0	0	-1	0	0
0.2			-	-	-	-	-	-	-	-	-
0.5			-2	0	0	-2	0	0	-1	0	-1
0.8			-24	0	0	-20	0	0	-16	-1	-3
1			-160	-2	0	-132	-4	-1	-107	-4	-78
C3											
5		mm day^{-1}	-1	0	0	-1	0	0	-1	0	-1
10			0	0	0	0	0	0	0	0	0
20			-	-	-	-	-	-	-	-	-
30			0	0	0	0	0	0	0	0	0
40			0	0	0	0	0	0	0	0	0
Aroot											
0.1		-	3	0	0	2	0	0	2	0	0
0.5			-	-	-	-	-	-	-	-	-
1	-4		0	0	-4	0	0	-3	0	-1	
2	-18		0	0	-15	0	0	-12	-1	-3	
3	-42		0	1	-35	0	1	-28	-2	-10	

* The figures in bold and italic are the parameter values applied in multi-site calibration protocol. The percentage changes in performance measures is estimated according to $(S'-S) \cdot 100/S$, in which S' means the simulation results of changed parameter values, and S is the basic results of multi-site calibration protocol.

simulations indirectly. As K_s decreased from 4×10^{-7} to $1 \times 10^{-8} \text{ m s}^{-1}$ the EF of Dage station sharply decreased by 281 % (Table 4), whilst it was only decreased by around 27 % for Xiahui and Daiying stations. Again, the change in lower boundary of UZ (Bnd.UZ) from 1.5 to 0.5 caused the EF of Dage to decrease by 654 %, whilst it was only by around 19 % for Xiahui and Daiying stations. The different model behaviors between Dage and the other two stations suggested that it was inappropriate to apply a uniform value with respect to the lower boundary of UZ zone. It was critically important to use a higher soil water storage capacity for accurately simulating hydrological process of up-stream area

of the watershed. This observation was in line with the previous assumption that a high water storage capacity existed in the upstream area of the watershed. It was noted that a linear relationship between UZ parameter and the EF response did not exist in the analysis. This may be explained by the impacts of parameter interaction.

Parameters of ET had no influence on EF-values of downstream area of the watershed. However, this was not true for the up-stream area since a moderate effect was observed for Dage station when altering either C2 or Aroot-values (Table 4). Both C2 and Aroot affected the ratio of soil evaporation to total evapotranspiration (ET). Elevated values

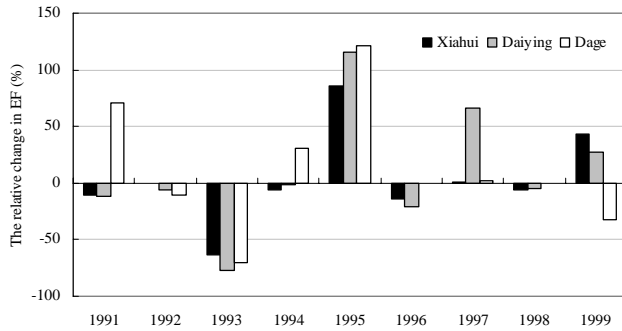


Fig. 4. The relative change in EF when compared the multi-site model calibration protocol with the single-site model calibration protocol. The relative change in EF was estimated according to $(S' - S)/S$, in which S' denotes the multi-site model calibration results, and S means the single-site model calibration results.

either C2 or Aroot corresponded to smaller soil evaporation and smaller ET, resulting in an increased runoff accordingly. The differences in EF sensitivity between Dage and Xiahui were mainly due to most of peak flows for Dage station was significantly over-predicted when increasing either C2 or Aroot, whereas it was almost no change for Xiahui station.

3.4 Potential source of simulation errors

Beven (2001) suggested that model uncertainties mainly were resulted from the errors on the model itself, errors on initial and boundary conditions, and errors on calibration data. In our modeling analysis, even though the model of multi-site calibration protocol was calibrated against the streamflow measured at three stations, model performances from multi-site calibration protocol did not show any improvement except to the EF of Dage station in the calibration period. Two kinds of errors were assumed to be responsible for the unimproved model behavior. This included the errors on the boundary conditions and the errors on model itself.

The errors of boundary conditions mainly referred to the incorrect representation of groundwater divide. Owing to the lack of information on saturated zone, saturated flow in our modeling analysis was only simulated by the simplified linear reservoir method. However, this method did not consider the flux change across the outer boundary of the model. Thus, the watershed was probably represented with a smaller drainage area than the reality, and the estimated groundwater recharge was less than the measurements. This may explain mostly the underestimation of streamflow rate in the range of 1 to $10 \text{ m}^3 \text{ s}^{-1}$, irrespective of which calibration protocol was applied. Streamflow of lower than $1 \text{ m}^3 \text{ s}^{-1}$ was generally over-predicted by the model. We assumed that it was attributed to the artifact of MIKE SHE that does not allow for a river/stream to dry out (Lu et al., 2006; Dai et al., 2010).

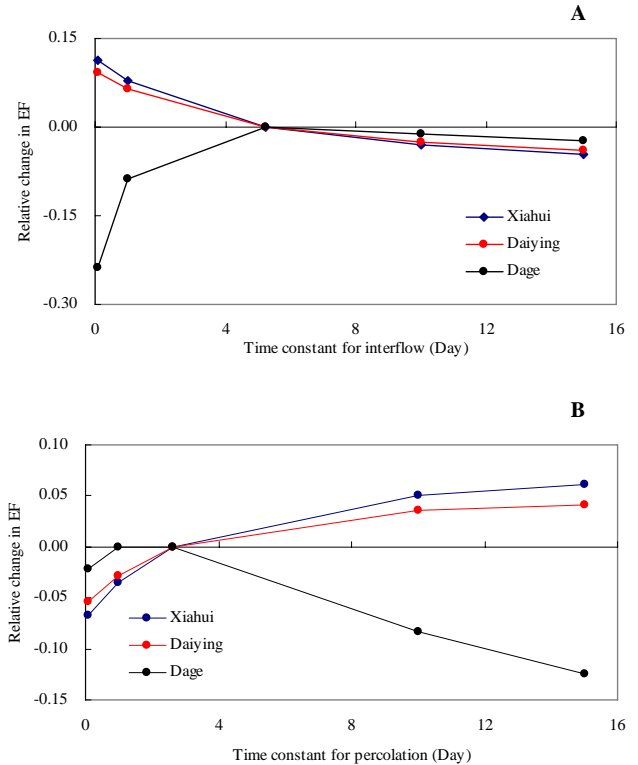


Fig. 5. Relative changes in EF in response to the variation of (A) time constant for interflow; and (B) time constant for percolation. The change was estimated according to $(S' - S)/S$, in which S' denotes the simulation with the changed parameter values, and S means the simulation with the calibrated values.

The errors on model itself mainly referred to the incorrect representation of the spatial distribution of hydro-geologic properties and hydrologic processes. As suggested in the previous discussion, the effects of the influential parameters on model performance varied across the watershed. It was necessary to apply spatially distributed parameter set in the model in order to improve the model performance of each station. Although spatially varied K_s have been taken into account in the model, much of the uncertainty remained with the specified dependent relationship of K_s between the dominant soil profiles of the watershed. In addition, parameters such as $S_{\text{pbaseflow}}$, T_{baseflow} , $T_{\text{interflow}}$, $T_{\text{percolation}}$, and Bnd_UZ that demonstrated a certain degree of influences on model performance, were all specified with uniform values across the watershed due to the absence of the detailed information. This was inconsistent with the reality that the watershed was characterized by varied soil depth, resulted in the representation of spatially varied geo-hydrological properties with much of errors, and consequently impeded improving the model behavior.

Due to the above errors and uncertainties, our modeling results from multi-site calibration protocol was not improved when compared with that of single-site protocol, and the

performance measures for Xiaohui and Daiying stations were, even, slightly decreased. Nevertheless, considering that performance measures from multi-site calibration protocol were generally satisfied with a compromise between the three stations, the multi-site calibration protocol had advantage to the single-site calibration protocol. In addition, it was necessary to give consideration to both hydrological processes of up- and down-stream areas of the watershed by applying multi-site calibration protocol. Model behaviors could be further improved when detailed spatially varied information of geo-hydrological properties was available.

4 Conclusions

We applied a physically-based distributed watershed hydrologic model, MIKE SHE, to investigate hydrologic processes of a large-scale mountainous watershed, Chaohe Basin, in northern China. A comparison study of two calibration protocols, single site vs multi-site, showed that the later method had some advantages over the former. Model sensitivity analysis suggested that hydrological process (e.g., groundwater flow) varied across the Chaohe watershed.

We concluded that it was necessary to use a multi-site calibration protocol for a distributed model to capture the varied hydrological processes governed by the spatially varied geo-hydrological properties. Large uncertainty remained in understanding the hydrologic processes due to lacking spatially explicit model parameters for a large watershed. Future field and modeling studies should focus on key model parameter determination, internal hydrologic flux measurements, and understanding the hydrologic processes in the two distinct areas as identified by the present simulation study.

Acknowledgements. The work is financially supported by special fund for the scientific research of forest public welfare industry (Project No. 201204102) and by the Ministry of Science and Technology, China through Key International Scientific and Technical Cooperation Project (Grant No. 2009DFA92430). Also part of the work was supported by the Program for Innovative Research Team in University (IRT1127), and the Fundamental Research Funds for the Central Universities (12MS68).

Edited by: A. Wei

References

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop Evapotranspiration: Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper No. 56, FAO, Rome, 1998.
- Ambroise, B., Perrin, J. L., and Reutenauer, D.: Multicriterion validation of a semi-distributed conceptual model of the water cycle in the Fecht catchment (Vosges Massif, France), *Water Resour. Res.*, 31, 1467–1481, 1995.
- Andersen, J., Refsgaard, J. C., and Jensen, K. H.: Distributed hydrological modeling of the Senegal River Basin-model construction and validation, *J. Hydrol.*, 247, 200–214, 2001.
- Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J. D., and Manus, C. L.: Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood event, *J. Hydrol.*, 394, 134–147, 2010.
- Becker, A. and Braun, P.: Disaggregation, aggregation and spatial scaling in hydrological modeling, *J. Hydrol.*, 217, 239–252, 1999.
- Bergström, S., Lindström, G., and Pettersson, A.: Multi-variable parameter estimation to increase confidence in hydrological modeling, *Hydrol. Process.*, 16, 413–421, 2002.
- Beven, K. J.: *Rainfall-Runoff Modelling-The Primer*, 360 pp., John Wiley, Hoboken, NJ, 2001.
- Blöschl, G., Grayson, R. B., and Sivapalan, M.: On the representative elementary area (REA) concept and its utility for distributed rainfall runoff modelling, *Hydrol. Process.*, 9, 313–330, 1995.
- Cao, W., Bowden, W. B., Davie, T., and Fenemor, A.: Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability, *Hydrol. Process.*, 20, 1057–1073, 2006.
- Dai, Z., Li, C., Trettin, C., Sun, G., Amatya, D., and Li, H.: Bi-criteria evaluation of the MIKE SHE model for a forested watershed on the South Carolina coastal plain, *Hydrol. Earth Syst. Sci.*, 14, 1033–1046, doi:10.5194/hess-14-1033-2010, 2010.
- El-Nasr, A., Arnold, J. G., Feyen, J., and Berlamont, J.: Modelling the hydrology of a catchment using a distributed and a semi-distributed model, *Hydrol. Process.*, 19, 573–587, 2004.
- Feyen, L., Vázquez, R., Christiaens, K., Sels, O., and Feyen, J.: Application of a distributed physically-based hydrological model to a medium size catchment, *Hydrol. Earth Syst. Sci.*, 4, 47–63, doi:10.5194/hess-4-47-2000, 2000.
- Freer, J., Beven, K., and Peters, N.: Multivariate Seasonal Period Model Rejection Within the Generalised Likelihood Uncertainty Estimation Procedure, in: *Calibration of Watershed Models*, edited by: Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A., and Turcotte, R., AGU, Washington, DC, 69–88, 2003.
- Graham, D. N. and Butts, M. B.: Chapter 10 Flexible Integrated Watershed Modeling with MIKE SHE, in: *Watershed Models*, edited by: Singh, V. P. and Frevert, D. K., CRC Press, Boca Raton, FL, USA, 2005.
- Gurtz, J., Baltensweiler, A., and Lang, H.: Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins, *Hydrol. Process.*, 13, 2751–2768, 1999.
- Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, *J. Hydrol.*, 280, 52–71, 2003.
- Jia, H. and Cheng, S.: Spatial and dynamic simulation for Miyun reservoir waters in Beijing, *Water Sci. Technol.*, 46, 473–479, 2002.
- Khu, S. T., Madsen, H., and di Pierro, F.: Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, *Adv. Water Resour.*, 31, 1387–1398, 2008.
- Knudsen, J., Thomsen, A., and Refsgaard, J. C.: WATBAL: A semidistributed, physically based hydrological modelling system, *Nord. Hydrol.*, 17, 347–362, 1986.

- Kuczera, G. and Mroczkowski, M.: Assessment of hydrologic parameter uncertainty and the worth of multi-response data, *Water Resour. Res.*, 35, 1481–1489, 1998.
- Li, X., Weller, D. E., and Jordan, T. E.: Watershed model calibration using multi-objective optimization and multi-site averaging, *J. Hydrol.*, 380, 277–288, 2010.
- Li, Z. J. and Li, X. B.: Impacts of Engineering Measures for Water Conservancy on Runoff in the Chaohe River Basin Based On an Empirical Statistical Model, *Acta Geographica Sinica*, 63, 958–968, 2008 (in Chinese with an English abstract).
- Lu, J., Sun, G., Amatya, D., Harder, S., and McNulty, S.: Understanding the hydrological response of a coastal plain watershed to forest management and climate change in South Carolina, USA, in: *Hydrology and Management of Forested Wetlands Proceedings of the International Conference*, New Bern, North Carolina, 8–12 April 2006, 231–239, 2006.
- Ma, H., Yang, D., Tan, S. K., Gao, B., and Hu, Q.: Impact of climate variability and human activity on streamflow decrease in the Miyun Reservoir catchment, *J. Hydrol.*, 389, 317–324, 2010.
- Madsen, H.: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, 26, 205–216, 2003.
- Merz, B. and Bárdossy, A.: Effects of spatial variability on the rainfall runoff process in a small loess catchment, *J. Hydrol.*, 212–213, 304–317, 1998.
- Migliaccio, K. W. and Chaubey, I.: Comment on: “Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability” by Cao et al., 2006, *Hydrol. Process.*, 21, 3226–3228, 2007.
- Moussa, R., Chahinian, N., and Bocquillon, C.: Distributed hydrological modelling of a Mediterranean mountainous catchment-Model construction and multi-site validation, *J. Hydrol.*, 337, 35–51, 2007.
- Mroczkowski, M., Raper, G. P., and Kuczera, G.: The quest for more powerful validation of conceptual catchment models, *Water Resour. Res.*, 30, 2325–2335, 1997.
- Nash, J. E. and Sutcliffe, J. V.: Riverflow forecasting through conceptual models. Part 1: a discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.*, 198, 69–97, 1997.
- Refsgaard, J. C. and Storm, B.: *Computer Models of Watershed Hydrology*, Water Resources Publications, Englewood, USA, 809–846, 1995.
- Sahoo, G. B., Ray, C., and De Carlo, E. H.: Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream, *J. Hydrol.*, 327, 94–109, 2006.
- Seibert, J., Uhlenbrook, S., Leibundgut, C., and Haldin, S.: Multiscale calibration and validation of a conceptual rainfall-runoff model, *Phys. Chem. Earth Pt. B*, 25, 59–64, 2000.
- Shafii, M. and De Smedt, F.: Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, 13, 2137–2149, doi:10.5194/hess-13-2137-2009, 2009.
- Sivapalan, M.: Process complexity at hillslope scale, process simplicity at the watershed scale: Is there a connection?, *Hydrol. Process.*, 17, 1037–1041, 2003.
- Smerdon, B. D., Allen, D. M., Grasby, S. E., and Berg, M. A.: An approach for predicting groundwater recharge in mountainous watersheds, *J. Hydrol.*, 365, 156–172, 2009.
- Sun, N., Li, X. B., Li, Z. J., and Yu, X. X.: Simulation of impacts of changes in land use and cover on annual streamflow in the upper reach of Chaohe River Basin. *Journal of Beijing Forestry University*, 30, 23–30, 2008 (in Chinese with an English abstract).
- Vázquez, R. F., Willems, P., and Feyen, J.: Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteria approach, *Hydrol. Process.*, 22, 2159–2179, 2008.
- Weingartner, R., Barben, M., and Spreafico, M.: Floods in mountain areas-an overview based on examples from Switzerland, *J. Hydrol.*, 282, 10–24, 2003.
- White, K. L. and Chaubey, I.: Sensitivity analysis, calibration, and validations for a multisite and multi variables SWAT model, *J. Am. Water Resour. As.*, 41, 1077–1089, 2005.
- Wijesekara, G. N., Gupta, A., Valeo, C., Hasbani, J. G., Qiao, Y., Delaney, P., and Marceau, D. J.: Assessing the impact of future land-use changes on hydrological processes in the Elbow River watershed in southern Alberta, Canada, *J. Hydrol.*, 412–413, 220–232, 2012.
- Xevi, E., Christiaens, K., Espino, A., Sewnandan, W., Mallants, D., Sorensen, H., and Feyen, J.: Calibration, Validation and Sensitivity Analysis of the MIKE-SHE Model Using the Neuenkirchen Catchment as Case Study, *Water Resour. Manag.*, 11, 219–242, 1997.
- Yang, J., Reichert, P., Abbaspour, K. C., and Yang, H.: Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference, *J. Hydrol.*, 340, 167–182, 2007.
- Zhang, X., Srinivasan, R., and Liew, M. V.: On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model, *Hydrol. Process.*, 24, 955–969, 2009.
- Zhang, Z., Wang, S., Sun, G., McNulty, S. G., Zhang, H., Li, J., Zhang, M., Klaghofer, E., and Strauss, P.: Evaluation of the Distributed Hydrologic Model MIKE SHE for Application in a Small Watershed on the Loess Plateau, Northwestern China, *J. Am. Water Resour. As.*, 44, 1108–1120, 2008.