Can multi-objective calibration of streamflow guarantee better hydrological model accuracy?

Ruqiang Zhang, Junguo Liu, Hongkai Gao and Ganquan Mao

ABSTRACT

Hydrological models often require calibration. Multi-objective calibration has been more widely used than single-objective calibration. However, it has not been fully ascertained that multi-objective calibration will necessarily guarantee better model accuracy. To test whether multi-calibration was effective in comparison to single-calibration in terms of model accuracy, two strategies were tested out. For these strategies, the objective functions used included the Nash–Sutcliffe efficiency and its logarithmic form, which highlight high flow and low flow, respectively. These two indexes were first used for multi-objective calibration, and then they were separately employed for single-objective calibration. To assess the calibration strategies' accuracy, the simulated streamflow was compared with observed streamflow, particularly high flow and low flow. This study was conducted in the upper stream of the Heihe River basin in northwest China using the FLEX-Topo model and MOSCEM-UA algorithm. The results show that the simulation based on the Nash–Sutcliffe efficiency performed best both in modelling the dynamics and simulating the high flow of the observed streamflow. Thus, it seems that multi-objective calibration does not necessarily lead to better model accuracy. This conclusion might provide useful information for hydrologists in calibrating their models, making their simulations more reliable.

Key words | FLEX-Topo model, model accuracy, MOSCEM-UA algorithm, multi-objective calibration, objective functions

Ruqiang Zhang
Junguo Liu (corresponding author)
School of Nature Conservation,
Beijing Forestry University,
Beijing 100083,
China
E-mail: *junguo.liu@gmail.com*

Junguo Liu
Ganquan Mao
School of Environmental Science and Engineering,
South University of Science and Technology of
 China,
Shenzhen 518055,
China

Hongkai Gao
Julie Ann Wrigley Global Institute of Sustainability,
Arizona State University,
Tempe, AZ 85287,
USA

INTRODUCTION

In order to understand hydrological processes and thus predict the likely influence of global climate change on hydrology, a means of extrapolating into the past and the future has been deemed necessary, mainly due to the limitations in space and time of hydrological measurement techniques and scope (Beven 2012). There are many types of models that can enable us to quantitatively extrapolate, predicting flooding, runoff, the availability of various water resources and nutrient transport (Mimikou *et al.* 2015;

Molina-Navarro *et al.* 2017). However, these models are only simplified representations of the real world, and their accuracy cannot be taken for granted (Muleta 2012). In addition, some model parameters may not physically manifest and therefore cannot be determined through direct measurements. Thus, model calibration is always necessary (Kim & Lee 2014). Model calibration is the process of identifying model parameter values from the available parameter ranges to enable the maximization or minimization of the objective function (a function representing difference between simulated and observed values) (Muleta 2012).

There has been considerable research into various optimization algorithms used to identify optimal model

parameter values. In the past, this particular research has mainly focused on the calibration of a single objective (Vrugt *et al.* 2003a). For example, a widely used algorithm for identifying optimal parameter values with a single-objective function, the shuffled complex evolution, was developed at the University of Arizona (SCE-UA) (Kan *et al.* 2016). This type of calibration algorithm is primarily concentrated with matching one component of the hydrograph produced by the model to the observed data (Boyle *et al.* 2000). However, in regards to the observed data, it has been argued that any single-objective function, no matter how elaborately chosen, is usually inadequate to properly represent all of the characteristics (Vrugt *et al.* 2003a; Sadeghi-Tabas *et al.* 2017). To circumvent this problem, multiple-objective calibration, which represents more than one aspect of a hydrological system and its behaviour, has been proposed as an alternative as it can incorporate more information from the available data (Zhang *et al.* 2015).

In the past, assumptions have been made that multi-objective calibration improves model accuracy (Kim & Lee 2014). Nevertheless, little research has gone into testing this supposition prior to now, which this study undertakes. To carry out this test, the Nash–Sutcliffe efficiency and its logarithmic form were selected as the objective functions for both multi- and single-objective calibration for the FLEX-Topo (Topography-driven FLux EXchange model) used in the upper stream of the Heihe River basin (UHRB) in northwest China. At the basin outlet, streamflow discharge was used for model calibration. Model accuracy was assessed by comparing the simulated streamflow with the observed streamflow data.

## MATERIALS AND METHODS

Study area

The Upper Heihe River Basin (UHRB) is the upper stream of the second largest inland river in Northwest China, the Heihe River (Figure 1(a)). The river originates in the Qilian Mountains. The UHRB has an area of 10,009 km² and stretches for 303 km. The elevation of the UHRB ranges from 1,700 m to 5,000 m (Figure 1(b)). This mountainous region surrounding the UHRB is characterized by a cold desert climate, with an annual temperature of UHRB

2–3 °C (Liu *et al.* 2008). The long-term average annual precipitation is ~430 mm y^{-1} and potential evaporation is ~520 mm y^{-1}. More than 80% of the annual precipitation occurs from May through September. Soil types in this region are predominantly mountain straw and grassland soil, chestnut-coloured soil, chernozemic soil and desert. The land cover includes forest, grassland, bare rock or bare soil, wetland and permanent snow (Figure 1(c)). On average, the UHRB produces ~70% of the total river-borne runoff of the whole Heihe River Basin (Chen *et al.* 2005). As it is the main runoff-producing region in the Heihe River basin, the UHRB is essential for an integrated water management practice, and so the hydrodynamics of the UHRB have garnered a great deal of research interest.

One hydrological station and four meteorological stations are located in and around the UHRB (Figure 1(c)). For this study, these meteorological stations provided input data for the hydrological model. In this model, Thiessen polygons were used to define the monitored area of these meteorological stations. Since the meteorological stations are all located at relatively low elevations, empirical formulas were used to adjust precipitation and temperature data. The streamflow discharge data for model calibration were obtained from the hydrological station (Yingluoxia Station) at the outlet of the UHRB to calibrate the model. All these input data were obtained from the Cold and Arid Regions Science Data Center (http://westdc.westgis.ac.cn/).

FLEX-Topo model

The FLEX-Topo model, a semi-distributed hydrological model, was used as the basis for different calibration strategies. Input data included daily precipitation and daily temperature, and the outputs included streamflow discharge and evapotranspiration. The FLEX-Topo model consists of four identically structured, parallel components representing four hydrological landscapes (*hydrological upland*, *sunny-hillslope*, *shady-hillslope* and *hydrological lowland*) which are classifed through topographic information that included height above the nearest drainage (HAND), elevation, slope and aspect (Gao *et al.* 2014). There are 25 parameters, shown in red in Figure 2. Table 1 shows the definition of the parameters and the ranges of their values determined by experience or references.
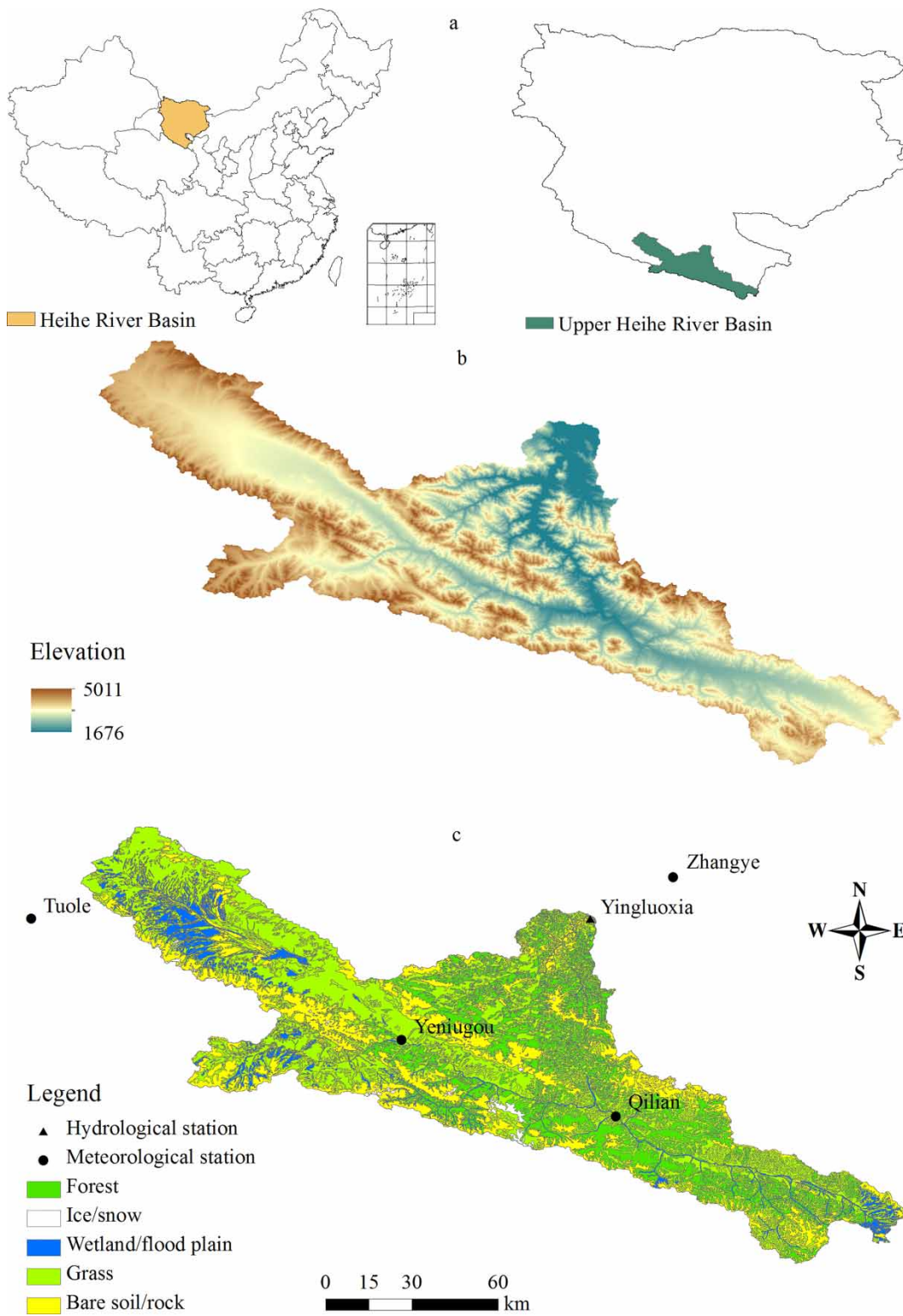
Figure 1 │ (a) Location of the Upper Heihe River basin in China; (b) DEM of the Upper Heihe River basin; (c) land cover map of the Upper Heihe River basin and distribution of hydrology gauges and meteorology gauges in and around the Upper Heihe.
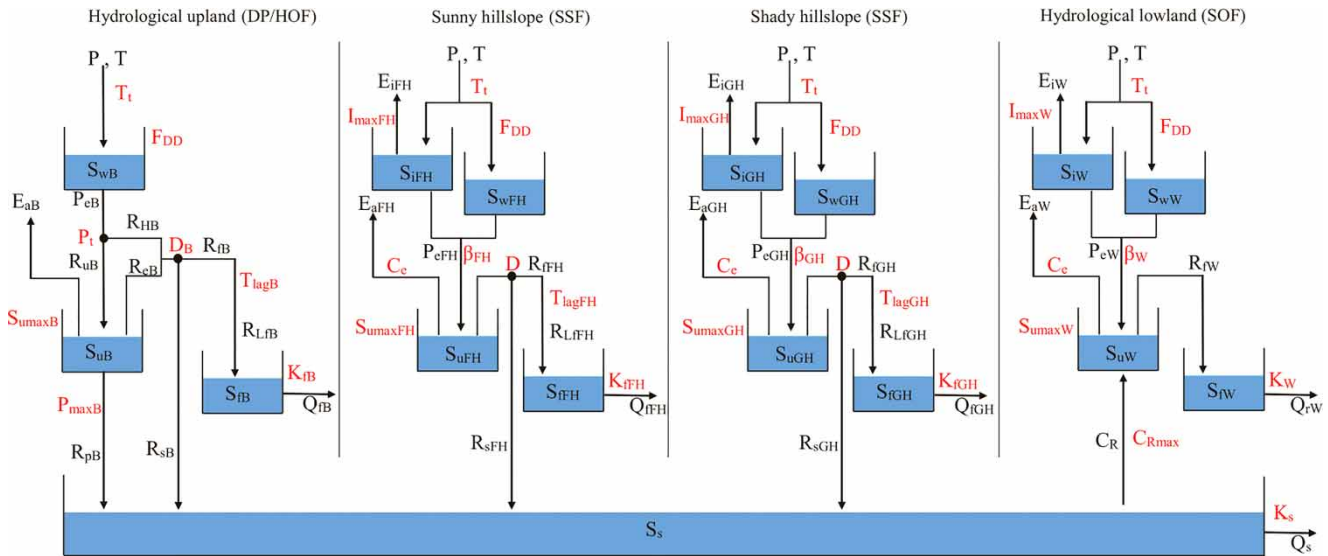
Figure 2 | Structure of FLEX-Topo model (revised from Gao et al. 2014). Please refer to the online version of this paper to see this figure in colour: http://dx/doi:10.2166/hydro.2018.131.

Table 1 | Parameters and ranges of values used in the FLEX-Topo model

Parameter	Range	Description
F_{DD}	[1,8]	Degree day factor defining the melted water per day per Celsius degree above T_t
T_t	[−2.5,2.5]	Air temperature threshold above which snow melts
P_t	[5,35]	Threshold parameter indicating occurrence of HOF in hydrological upland
S_{umaxB}	[5,500]	Maximum soil moisture capacity in the root zone of hydrological upland
P_{maxB}	[0.1,10]	Maximum percolation of hydrological upland
D_B	[0,1]	Splitter to separate preferential flow from recharge in hydrological upland
T_{lag}	[0,5]	Time lag between storm and fast runoff generation
K_{fB}	[2,50]	Timescale of the runoff from fast reservoir in hydrological upland
I_{maxFH}	[1,10]	Daily maximum interception capacity of sunny hillslope
C_e	0.5	A ratio above which the actual evaporation is equal to potential evaporation
S_{umaxFH}	[100,1,000]	Maximum soil moisture capacity in the root zone of sunny hillslope
β_{FH}	[0.1,5]	Parameter used to calculate runoff coefficient in sunny hillslope
D	[0,1]	Parameter separating the generated surface runoff on hydrological upland into the splitter to separate preferential flow from recharge in sunny and shady hillslope
K_f	[2,50]	Timescale of the runoff from fast reservoir in sunny and shady hillslope
I_{maxGH}	[0,10]	Daily maximum interception capacity of shady hillslope
S_{umaxGH}	[50,1,000]	Maximum soil moisture capacity in the root zone of shady hillslope class
β_{GH}	[0.1,5]	Parameter used to calculate runoff coefficient in shady hillslope
I_{maxW}	[0.1,10]	Daily maximum interception capacity of hydrological lowland
S_{umaxW}	[5,1,000]	Maximum soil moisture capacity in the root zone of hydrological lowland
β_W	[0.1,5]	Parameter used to calculate runoff coefficient in hydrological lowland
K_W	[1,9]	Timescale of runoff from fast reservoir in hydrological lowland
K_s	90	Timescale of the runoff from slow reservoir
C_{Rmax}	[0,5]	Parameter indicating a constant amount of capillary rise

The four landscapes are each characterized by various runoff-producing mechanisms, which are embodied in the parallel components of the FLEX-Topo model. The hydrological processes in each landscape are briefly described next. More details on the model parameters, the model structure and the water balance equations have been described by Gao *et al.* (2014).

In the hydrological upland, the dominant hydrological processes are Hortonian overland flows (HOF) and deep percolation (DP). Precipitation (P) can be stored as snow cover (S_{wB}) if the daily temperature is below the threshold value T_t. It is assumed that there is no interception due to a lack of significant vegetation. The sum (P_{eB}) of rainfall and the snowmelt, which is calculated by a degree-day model, flows towards the unsaturated zone reservoir (S_{uB}). When P_{eB} is greater than the infiltration capacity (P_t), Hortonian overland flow (R_{HB}) occurs. Infiltration (R_{uB}) recharges the unsaturated reservoir (S_{uB}). Percolatation (R_{pB}) from S_{uB} that flows into the slow-response reservoir (S_s) is calculated from the relative soil moisture (S_{uB}/S_{umaxB}) and maximum percolation (P_{maxB}). Actual evapotranspiration (E_{aB}) from the unsaturated reservoir is estimated by the relative soil moisture (S_{uB}/S_{umaxB}) and the potential evapotranspiration that is calculated by the Hamon equation (Hamon 1960). Saturation excess overland flow (R_{eB}) occurs if the amount of water in the unsaturated reservoir exceeds the maximum storage capacity (S_{umaxB}). R_{eB} plus R_{HB} is partitioned by a splitter parameter (D_B) into flow (R_{fB}) going into the fast-response reservoir (S_{fB}) and flow (R_{sB}) re-infiltrating into the slow-response reservoir (S_s). R_{fB} turns into R_{LfB} after convolution using the time lag parameter T_{lag}, which represents the time interval between storm and fast runoff generation. Flow (Q_{fB}) from S_{fB} is routed to the stream channel after time K_{fB}.

The major runoff-producing mechanism of shady-hillslope is the storage excess subsurface flow (SSF). The existence of vegetation indicates the necessity for the interception reservoir (S_{iFH}). Evapotranspiration (E_{iFH}) from the interception reservoir is assumed to be equal to potential evapotranspiration if the storage of the interception reservoir is nonzero. The sum (P_{eFH}) of the remainder of rainfall after interception and the snowmelt that is calculated from a degree-day model flow towards the unsaturated reservoir (S_{uFH}). P_{eFH} is partitioned into

runoff and flow that is routed to S_{uFH} through the use of the runoff coefficient that is calculated from the parameter β_{FH} representing the heterogeneity of soil properties. Runoff from P_{eFH} and S_{uFH} is separated by a splitter parameter (D) into flow (R_{fFH}) going into the fast-response reservoir (S_{fFH}) and flow (R_{sFH}) re-infiltrating into S_s. Actual evapotranspiration (E_{aFH}) from the S_{uFH} is estimated by parameter C_e and potential evapotranspiration, the latter of which is calculated by the Hamon equation. C_e is a threshold value. If S_{uFH}/S_{umaxFH} is larger than C_e, actual evapotranspiration is assumed to be equal to potential evapotranspiration. R_{fFH} becomes R_{LfB} after convolution using the time lag parameter T_{lag} and flows into the fast-response reservoir (S_{fFH}). Water (Q_{fFH}) from S_{fFH} is routed to the stream channel with timescale (K_{fH}).

The major runoff-producing mechanism of sunny-hillslope is assumed to be the same as that of shady-hillslope, but the parameters for this landscape take different values.

In hydrological lowland, the dominant hydrological process is saturation excess overland flow (SOF). Interception and snowmelt are the same as they are in shady-hillslope, but soil routine is different. Runoff (R_{fW}) produced from P_{eW} and S_{uW} is directly routed to the fast-response reservoir (S_{fW}) without delay because of its proximity to the stream channel. Due to the shallow groundwater level and limited storage capacity, capillary rise (C_R) occurs. C_R is represented by a parameter (C_{Rmax}) that indicates a constant amount of capillary rise.

Optimization algorithm and objective functions

The MOSCEM-UA (Multi-Objective Shuffled Complex Evolution Metropolis-University of Arizona) algorithm that was developed by Vrugt *et al.* (2003a) was proposed for multi-objective calibration because it maintains a uniform sampling density within the Pareto set of solutions and includes the single-criterion end points. The evolution strategy employed in the MOSCEM-UA algorithm is similar to that of the SCEM-UA algorithm (Vrugt *et al.* 2003b), but the probability ratio concept in the SCEM-UA algorithm was replaced by a multi-objective fitness assignment concept in order to develop the initial population of points toward a set of solutions resulting from a stable distribution. The MOSCEM-UA algorithm has been successfully applied in

a wide range of hydrological and environmental models, demonstrating its reliability and effectiveness (Efstratiadis & Koutsoyiannis 2010). In this study, the MOSCEM-UA algorithm (Vrugt *et al.* 2003a) was adopted to identify the parameters in the model (Figure 3). With these parameters, simulated streamflow is generated. Using the simulated and observed streamflow, objective function values were calculated by a module within the FLEX-Topo model.

Three basic parameters had to be set in order to control the operation of the algorithm: the number of iterations, the number of parallel sequences, and the number of random samples. Kuczera (1997) suggests that the number of parallel sequences be set to be equal to the number of the calibration parameters. The other two algorithmic parameters were set according to the recommended values in Vrugt *et al.* (2003b). In line with these suggestions, these three parameters were
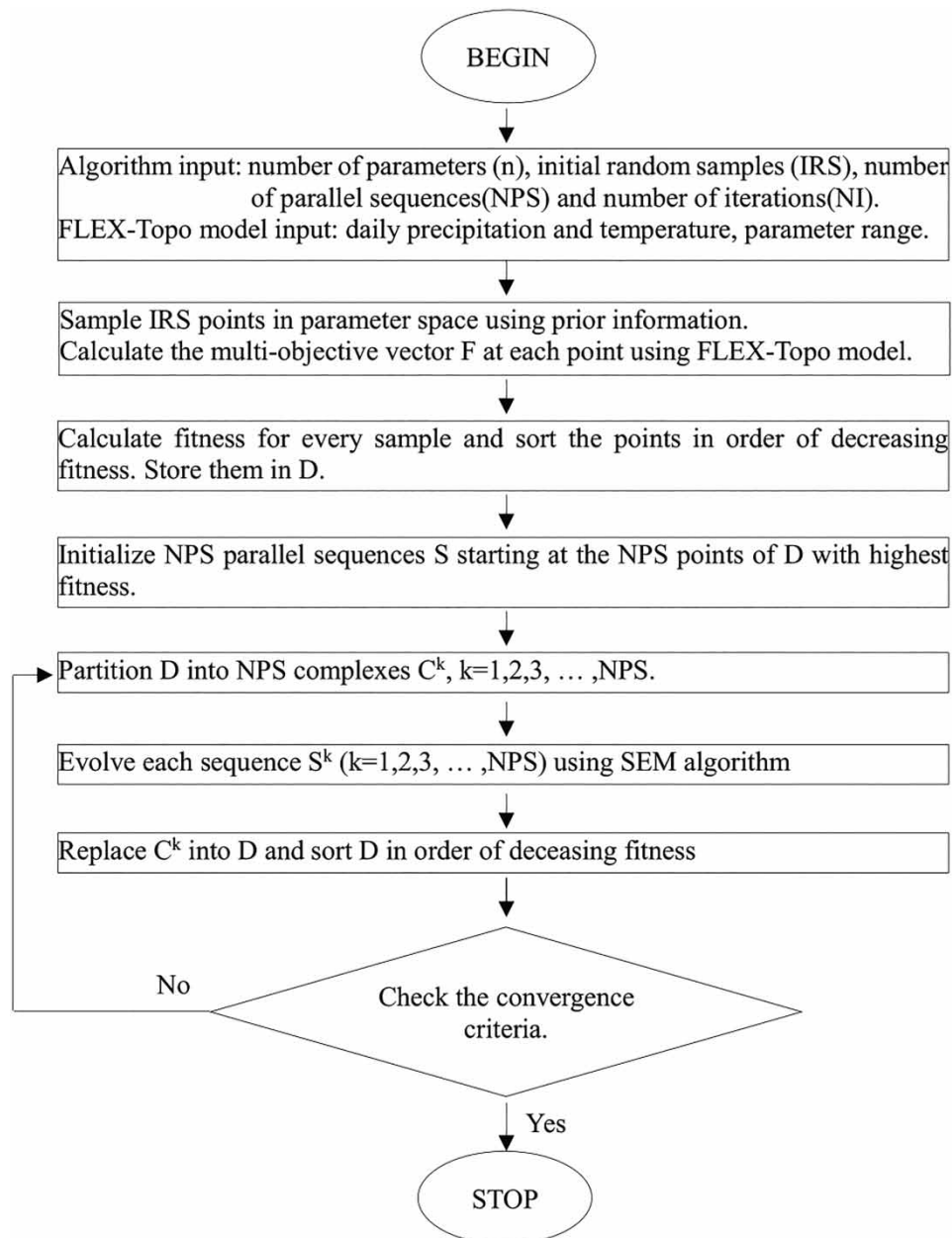

Figure 3 │ Flow chart of MOSCEM-UA algorithm (revised from Vrugt *et al.* 2003a).

set as 100,000, 25 and 2,500, respectively. The 20-year-period from 1959 to 1978 was used for model calibration. The first year was disregarded in order to minimize any influence from the initial conditions. A computer having a 3 GHz Intel Xeon CPU model E5-2687Wv4 was used to run the algorithms.

There are many commonly used objective functions, such as the Nash–Sutcliffe efficiency (Nash & Sutcliffe 1970) and its logarithmic form (Legates & Mccabe 1999; Kim & Lee 2014). The Nash–Sutcliffe efficiency and its logarithmic form highlight high flow and low flow, respectively. Therefore, both indexes were included for multi-objective calibration in order to simultaneously consider both high flow and low flow. Meanwhile, these two indexes were also separately used for single-objective calibration. Since the MOSCEM-UA algorithm generally deals with a minimization procedure, the Nash–Sutcliffe efficiency and its logarithmic form were subtracted from one. Table 2 shows the mathematical formulation of the objective functions. Comparison between different calibration strategies was made to assess the simulation accuracy in modelling observed streamflow, focusing on high flow and low flow.

RESULTS

Objective function values of single- and multi-objective calibrations

For single-objective calibration, one set of parameters was obtained, and for multi-objective calibration the Pareto front consisting of 33 sets of parameters was obtained. Figure 4 shows the results of different calibration strategies.

It is clear that several tradeoffs exist in multi-objective calibration, indicating that no single solution can optimize both objective functions simultaneously. The values of NSE and lnNSE corresponding to solutions in the Pareto front (dots marked in blue) vary widely. The NSE values range from 0.154 to 0.199, while the lnNSE values range from 0.100 to 0.151, indicating good performance of the model.

In regards to the NSE calibrated model (dot marked in red), the NSE value is 0.157, which is very close to the value (0.154) of the corresponding extreme end of the Pareto front. As regards the lnNSE calibrated model (dot marked in green), the lnNSE value is 0.097 and it is close in value to the other extreme end of the Pareto front (0.100). This proximity in value confirms that the Pareto front includes a single-criterion end point, which is consistent with the results of Vrugt *et al.* (2003a). In addition, it can be found that the extreme end of the Pareto front on lnNSE is closer to the single-objective calibration, yet the other extreme end is not proximal to the NSE calibrated solution.

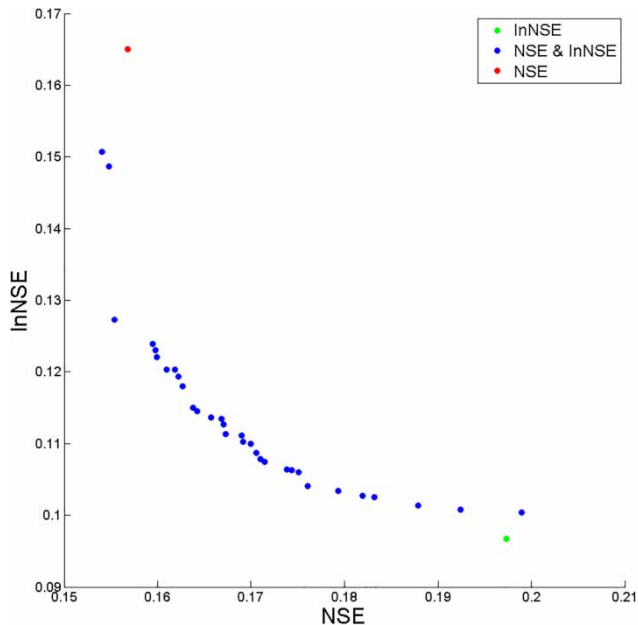### Assessment of single- and multi-objective calibrations

Since all Pareto optimal solutions are equally important, it may be difficult to prefer one solution over another. Therefore, every solution in the Pareto front was used to run the model, obtaining 33 sets of simulated streamflow. The average of these simulated flows was calculated and was denoted as the '*Average*'. The hydrographs of the different objective calibrated simulations are also shown in Figure 5. The simulated flows are respectively marked as Q_{NSE}, Q_{lnNSE} and $Q_{Average}$.

It can be seen that the simulated flows fluctuated synchronously with the observed flow, which indicates that both strategies could capture the dynamics of the observed discharge. It seems, however, that the resulting hydrograph

Table 2 │ Mathematical formulations of the objective functions and evaluation index

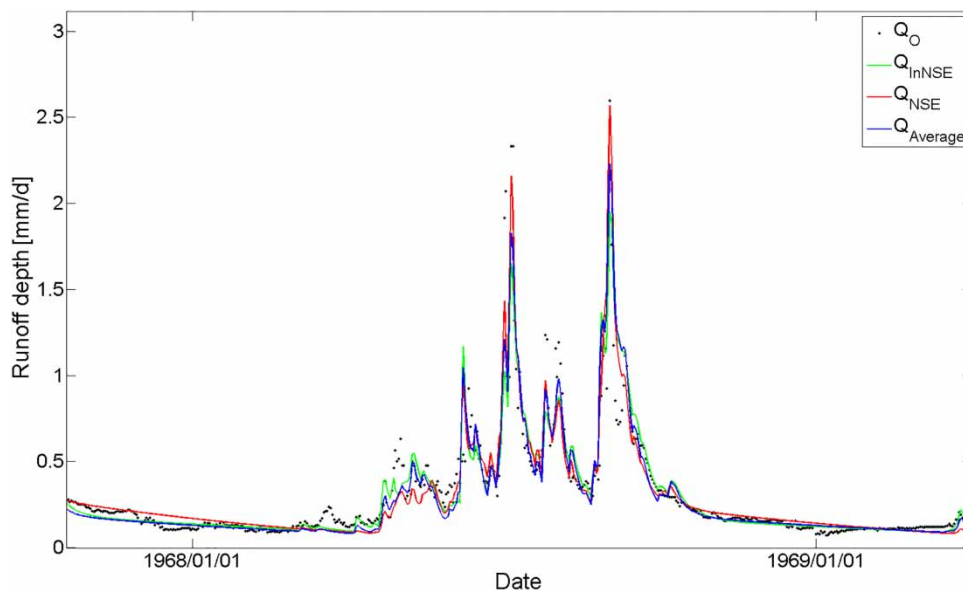| Objective function	Mathematical formulation	Range of values	Reference
NSE	$NSE = \dfrac{\sum_{i=1}^{N}(O_i - S_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$	$(-\infty, 1]$	Nash & Sutcliffe (1970)
lnNSE	$lnNSE = \dfrac{\sum_{i=1}^{N}(\ln O_i - \ln S_i)^2}{\sum_{i=1}^{N}(\ln O_i - \overline{\ln O_i})^2}$	$(-\infty, 1]$	Kim & Lee (2014)

Note: O_i is the observed discharge at time step i; S_i is the simulated discharge; $\bar{O}$ is the mean observed discharge over the entire simulation period of length N; $\bar{S}$ is the mean simulated discharge over the entire simulation period of length N.

Figure 4 | Objective function values of single- and multi-objective calibration solutions. Please refer to the online version of this paper to see this figure in colour: http://dx/doi:10.2166/hydro.2018.131.

tend to show a more centred scatterplot and the distributions of the dots were similar. This pattern indicated that these three simulations performed similarly in modelling low flow, although it was supposed that the lnNSE calibrated simulation would behave better. For higher flows the dots show a rising trend, demonstrating that all three simulations underestimated the high flows but to various degrees. The multi-objective calibrated simulation performed no better than the simulation based on NSE, the latter which presented a relatively low degree of underestimation. Taking the values of NSE into consideration, it is interesting to find that the lower the NSE value, the better the modelling of the high flow. In general, the NSE tends to perform better in modelling high flow and in reproducing the observed streamflow dynamics. There were no apparent differences among these three simulations in simulating low flow.

Influence of optimization algorithm parameters on model performance

based on NSE appeared slightly truer to the observed high flow than the other two resulting hydrographs.

Figure 6 shows a scatterplot of the residuals in regards to observed flow. For low flows, these three simulations all
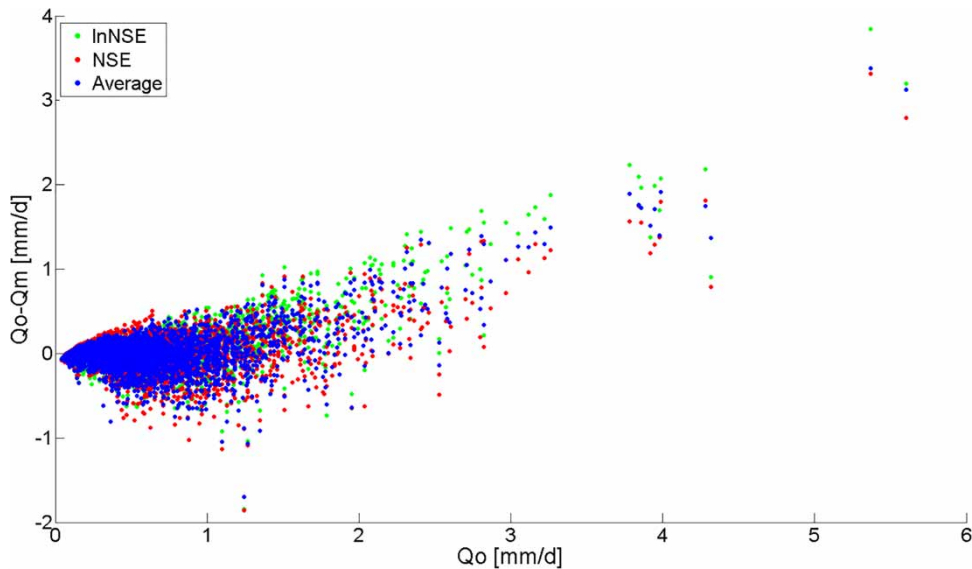
The MOSCEM-UA algorithm is a random sampling approach, thus initial samples may have an impact on the solutions. In addition, three parameters control the behaviour of MOSCEM-UA: the number of iterations (NI), the

Figure 5 | Hydrographs for single- and multi-objective calibrations.

Figure 6 │ Residuals plot of the different simulations corresponding to the multiple- and single-objective calibration.

number of random samples (NRS) and the number of parallel sequences (NPS). To examine the influence of samples and these three parameters on the simulations, various combinations of these algorithm parameters were carried out. The results of these trials are shown in Figure 7. It seems that trials of the changing parameter combinations of the optimization algorithm almost made no difference in modelling high flow and low flow. The results of these trials also prove the simulations based on NSE outperformed the simulations based on multi-objective.

DISCUSSION

This study aimed at testing whether multi-objective calibration would guarantee a better simulation of observed streamflow. Two widely used objective functions, NSE and lnNSE, were selected for model calibration. Since all the settings, including the input data, the hydrological model, the parameter ranges and the optimization algorithm were kept the same (except for the calibration strategies), it can reasonably be assumed that it was the objective functions that influenced model performance.

For high flow, the results show that simulation based on NSE performed slightly better. This may be because NSE weighs error on high flow more than error on low flow.

Because of the use of the logarithmic form, lnNSE weighs more on the error on low flow (Fenicia *et al.* 2007; Kim & Lee 2014). Therefore, the simulation calibrated by NSE outperformed the simulation based on lnNSE in reproducing high flow. Since the multi-objective calibration takes both NSE and lnNSE into consideration, it is not surprising that simulation based on NSE outperformed simulation based on multi-objective calibration in modelling high flow. By contrast, there was no significant difference among the three types of simulations based on different objective function(s), although simulation based on lnNSE was supposed to improve the low flow modelling. These results may be due to two factors. The first is that NSE still reacts to low flows although it highlights high flow. The second factor is that there generally tends to be more measurement errors in the observed low flow values. (Laaha *et al.* 2006). The greater level of error in observed low flow can be confirmed by Figure 6, in which the centred scatter illustrates a nearly symmetrical distribution around the line whose ordinate value is equal to zero.

It can be argued that a great deal of attention should be paid to objective function selection for multi-objective calibration as objective function has impact on model performance (Muleta 2012). It is crucial that performance of objective function be assessed before application in multi-objective calibration.
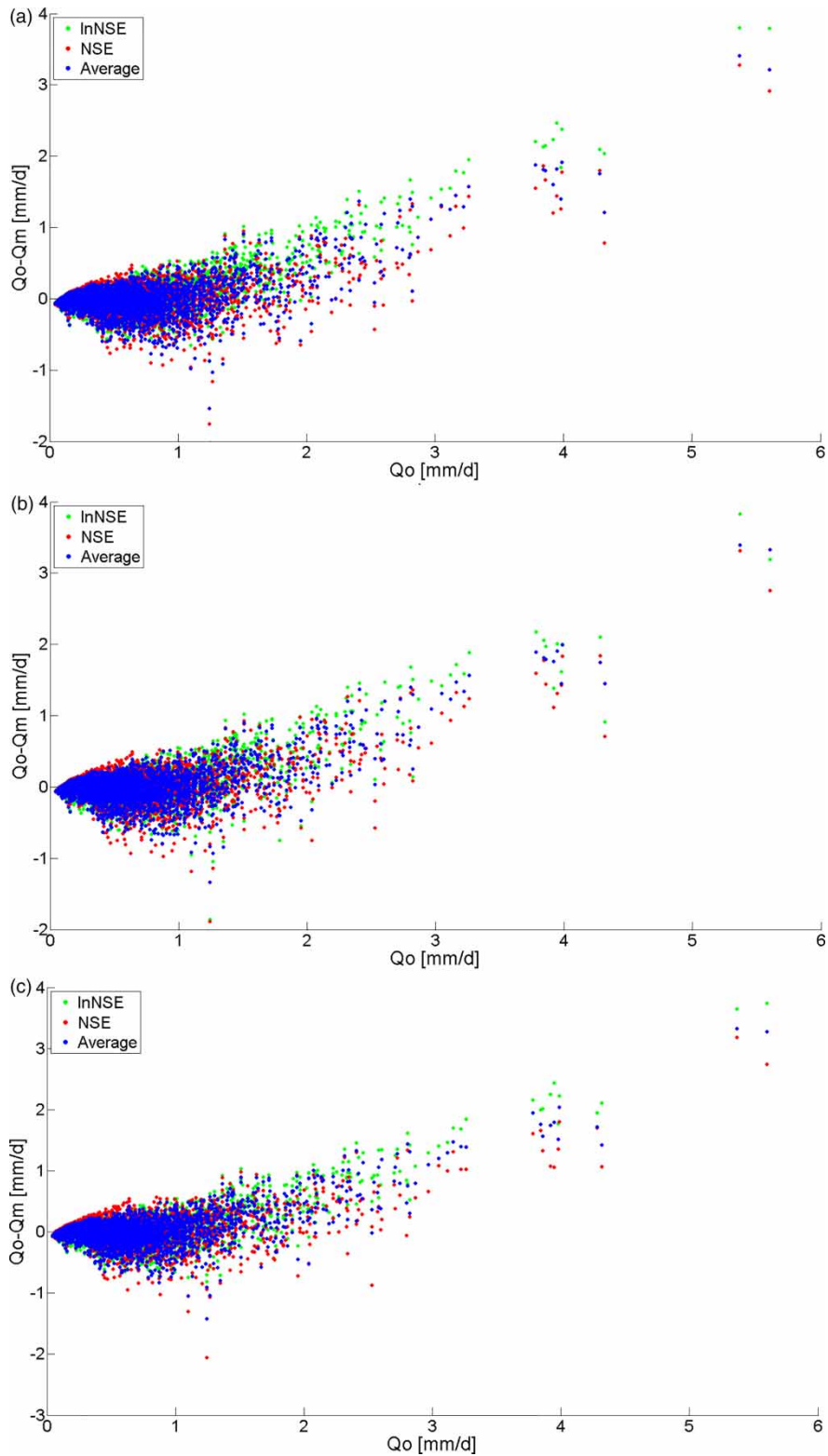
Figure 7 | Residuals plots of the simulations based on different optimization parameter combinations. *Note:* IRS is initial random samples (IRS); NRS is the number of random samples; NI is the number of iterations. (a) IRS = 25, NRS = 25*2, NI = 50,000; (b) IRS = 25, NRS = 25*4, NI = 50,000; (c) IRS = 25*2, NRS = 25*4, NI = 50,000.

It has been suggested that the inclusion of more variables in model calibration might provide additional information about the catchment characteristics (Zhang *et al.* 2016). In this study, the variable was streamflow. However, it is worth mentioning that there are other observed variables that could be used for multi-objective calibration, such as time shift between observed and simulated peak flow (Fenicia *et al.* 2007), energy (Scheerlinck *et al.* 2009) and chemical constituent (Haas *et al.* 2016; Zhang *et al.* 2016). For example, Zhang *et al.* (2016) calibrated both water quantity and water quality simultaneously. This type of multi-objective calibration was not included in this particular study, but it may prove to be an area of interest in the future.

CONCLUSION

This hydrological modelling study aimed at testing the supposition that a multi-objective calibration guaranteed better model accuracy in modelling high flow and low flow. To capture the dynamics of high and low flow, the Nash–Sutcliffe efficiency and its logarithmic form were selected as the objective functions. These two criteria were both included in the multi-objective calibration, while in the single-objective calibration they were separately used. The FLEX-Topo, a semi-distributed hydrological model, was used to simulate the Upper Heihe River basin hydrological processes, and the MOSCEM-UA was used for calibration. Assessment of different calibration strategies was carried out by comparing the simulated streamflow with the observed streamflow, focusing on the high flow and low flow.

It seems that multi-objective calibration, in fact, did not significantly improve model performance compared to single-objective calibration. Our results show that the Nash–Sutcliffe efficiency performed slightly better in modelling high flow while the logarithmic form of the Nash–Sutcliffe efficiency did not significantly improve the low flow simulation and at the same time more seriously underestimated the high flow. Multi-objective calibration led to no better accuracy in modelling high flow and similar simulation in low flow compared to single-objective calibration which was based on the Nash–Sutcliffe

efficiency. Thus, it cannot be taken for granted that multi-objective calibration of streamflow will necessarily guarantee better model accuracy. Indeed, model performance depends on the selected objective function(s). With this in mind, in order to design better and more accurate hydrological simulations of hydrological processes, a serious look at the selection of the objective functions should be undertaken.

ACKNOWLEDGEMENTS

REFERENCES

Beven, K. J. 2012 *Rainfall-runoff Modelling: The Primer*. John Wiley & Sons, Chichester, UK.

Boyle, D. P., Gupta, H. V. & Sorooshian, S. 2000 Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resources Research* **36**, 3663–3674.

Chen, Y., Zhang, D., Sun, Y., Liu, X., Wang, N. & Savenije, H. H. G. 2005 Water demand management: a case study of the Heihe River Basin in China. *Physics & Chemistry of the Earth Parts A/B/C* **30**, 408–419.

Efstratiadis, A. & Koutsoyiannis, D. 2010 One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal* **55**, 58–78.

Fenicia, F., Savenije, H. H. G., Matgen, P. & Pfister, L. 2007 A comparison of alternative multiobjective calibration strategies for hydrological modeling. *Water Resources Research* **43**, 93–99.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S. & Savenije, H. H. G. 2014 Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China. *Hydrology & Earth System Sciences Discussions* **10**, 12663–12716.

Haas, M. B., Guse, B., Pfannerstill, M. & Fohrer, N. 2016 A joined multi-metric calibration of river discharge and nitrate loads with different performance measures. *Journal of Hydrology* **536**, 534–545.

Hamon, W. R. 1960 *Estimating Potential Evapotranspiration. Thesis*, Massachusetts Institute of Technology, Cambridge, MA.

Kan, G., Zhang, M., Liang, K., Wang, H., Jiang, Y., Li, J., Ding, L., He, X., Hong, Y. & Zuo, D. 2016 Improving water quantity simulation & forecasting to solve the energy-water-food nexus issue by using heterogeneous computing accelerated global optimization method. *Applied Energy* **210**, 420–433.

Kim, H. S. & Lee, S. 2014 Assessment of a seasonal calibration technique using multiple objectives in rainfall–runoff analysis. *Hydrological Processes* **28**, 2159–2173.

Kuczera, G. 1997 Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research* **33**, 177–185.

Laaha, G., Blöschl, G., Demuth, S., Gustard, A., Planos, E., Seatena, F. & Servat, E. 2006 Low flow estimation in Austria. In: *Climate Variability & Change–Hydrological Impacts, Proceedings of the Fifth FRIEND World Conference*, Havana, Cuba.

Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35**, 233–241.

Liu, S., Zhang, G., Zhang, C. & Zhou, X. 2008 Water resources system evolution and impact of human activities in Heihe River Basin. *Journal of Jilin University (Earth and Science Edition)* **38**, 806–812.

Mimikou, M., Makropoulos, C. & Papathanasiou, C. 2015 Hydrological modelling for flood forecasting: calibrating the post-fire initial conditions. *Journal of Hydrology* **529**, 1838–1850.

Molina-Navarro, E., Andersen, H. E., Nielsen, A., Thodsen, H. & Trolle, D. 2017 The impact of the objective function in multi-site and multi-variable calibration of the SWAT model. *Environmental Modelling & Software* **93**, 255–267.

Muleta, M. K. 2012 Model performance sensitivity to objective function during automated calibrations. *Journal of Hydrologic Engineering* **17**, 756–767.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – A discussion of principles. *Journal of Hydrology* **10**, 282–290.

Sadeghi-Tabas, S., Samadi, S. Z., Akbarpour, A. & Pourreza-Bilondi, M. 2017 Sustainable groundwater modeling using single and multi objective optimization algorithms. *Journal of Hydroinformatics* **19**, 97–114.

Scheerlinck, K., Pauwels, V. R. N., Vernieuwe, H. & Baets, B. D. 2009 Calibration of a water and energy balance model: recursive parameter estimation versus particle swarm optimization. *Water Resources Research* **45**, 208–225.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S. 2003a Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resources Research* **39**. DOI:10.1029/2002wr001746.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. 2003b A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* **39**. DOI:10.1029/2002wr001642.

Zhang, R., Moreira, M. & Corte-Real, J. 2015 Multi-objective calibration of the physically based, spatially distributed SHETRAN hydrological model. *Journal of Hydroinformatics* **18**, jh2015219.

Zhang, Y., Shao, Q. & Taylor, J. A. 2016 A balanced calibration of water quantity and quality by multi-objective optimization for integrated water system model. *Journal of Hydrology* **538**, 802–816.