Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions

Searching for an optimized single-objective function matching multiple objectives with automatic calibration of hydrological models

Fuqiang Tian[1]*, Yu Sun[1], Hongchang Hu[1], Hongyi Li[2]

1 Department of Hydraulic Engineering, State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, China

2 Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA

*Corresponding author information:

Email: tianfq@tsinghua.edu.cn

Telephone: +86 010 6277 3396

Fax: +86 010 6279 6971

Submitted to Hydrology and Earth System Sciences

February 22, 2016

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Abstract:

In the calibration of hydrological models, evaluation criteria are explicitly and quantitatively defined as single- or multi-objective functions when utilizing automatic calibration approaches. In most previous studies, there is a general opinion that no single-objective function can represent all of the important characteristics of even one specific kind of hydrological variable (e.g., streamflow). Thus hydrologists must turn to multi-objective calibration. In this study, we demonstrated that an optimized single-objective function can compromise multi-response modes (i.e., multi-objective functions) of the hydrograph, which is defined as summation of a power function of the absolute error between observed and simulated streamflow with the exponent of power function optimized for specific watersheds. The new objective function was applied to 196 model parameter estimation experiment (MOPEX) watersheds across the eastern United States using the semi-distributed Xinanjiang hydrological model. The optimized exponent value for each watershed was obtained by targeting four popular objective functions focusing on peak flows, low flows, water balance, and flashiness, respectively. The results showed that the optimized single-objective function can achieve a better hydrograph simulation compared to the traditional single-objective function Nash-Sutcliffe efficiency coefficient for most watersheds, and balance high flow part and low flow part of the hydrograph without substantial differences compared to multi-objective calibration. The proposed optimal single-objective function can be practically adopted in the hydrological modeling if the optimal exponent value could be determined *a priori* according to hydrological/climatic/landscape characteristics in a specific watershed. This is, however, left for future study.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

Keywords: automatic calibration, single-objective function, multi-objective functions, hydrological model

1. Introduction

Hydrological models are often used to simulate the past and to predict the future hydrological behaviors of catchment. All kinds of models, lumped conceptual models or distributed physically based models, are simplifications of reality. Their parameters usually cannot be directly observed or easily derived from measurable catchment characteristics, but have to be indirectly estimated by some kind of calibration methods (Booij and Krol, 2010; Madsen, 2000; Pokhrel and Gupta, 2010; Vrugt et al., 2002). Calibration means that parameters are adjusted to match model simulation with historically observed data as closely as possible. Generally, it is translated into an optimization problem from the perspective of mathematics, and performed automatically using optimization algorithms (Guinot et al., 2011; Muleta, 2012). In this approach, objective functions are necessary to evaluate the closeness between the simulated and observed variable.

There are ample literatures on model performance evaluation, because it is important not only for calibration but also for model development and intercomparison (Krause et al., 2005; Muleta, 2012; Wagener, 2003). The traditional approach is using a single-objective function. However, many researchers share the concerns about single-objective functions such as the most widely used Nash-Sutcliffe efficiency (NSE) (Jain and Sudheer, 2008; McCuen et al., 2006; Schaefli and Gupta, 2007). The opinion that a single-objective function cannot capture all of the important characteristics of the observed data has been gradually accepted (Vrugt et al., 2003; Wagener, 2003). More and more hydrologists seek to improve the calibration methods to capture various aspects of hydrologic responses simultaneously (Fenicia et al., 2007; Madsen et al., 2002).

Inspired by the excellent studies by Gupta et al. (1998) and Yapo et al. (1998), multi-objective calibration has been considered to be able to extract more information from historical data hence widely used to identify non-dominated or Pareto optimal parameter sets (Gupta et al., 2009; Hall et al., 2005; Matott et al., 2009; van Werkhoven et al., 2009). The progress in multi-objective calibration in recent years was well summarized in a comprehensive review paper by Efstratiadis and Koutsoyiannis (2010).

In general, multi-objective calibration can be categorized into three types, i.e., multiple objectives based on multi-variable measurements, multi-site measurements, and multi-response modes (Madsen, 2003). In this study, multi-objective calibration referred to the third type that measures various responses of the hydrological processes, especially the streamflow hydrograph. High flows and low flows are two important characteristics of the hydrograph and the trade-offs between them have been considerably discussed (Bekele and Nicklow, 2007; Boyle et al., 2000; Gill et al., 2006; Khu et al., 2008; Tang et al., 2007; van Griensven and Bauwens, 2003). In addition, water balance and flow variability are of great importance as well (Kollat et al., 2012; Price et al., 2012; van Werkhoven et al., 2009).

The multi-objective calibration produces a series of parameter sets located on the Pareto front, which provides new perspectives for parameter estimation. However, single-objective calibration has still been widely used, because a unique parameter set is often preferred by decision makers for daily water resources management practices. It is useful to identify a good compromise against the conflicting objectives. For this purpose, in this study we proposed a new methodology to obtain an optimized single-objective function (OSOF) which can simultaneously address multi-response modes for automatic calibration of hydrological models.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences Discussions

This paper was organized as follows. After this brief introduction, in Section 2, we introduced the definition of new single-objective function. Case study areas and data, the applied methods including the hydrological model, the optimization algorithm, the evaluation framework, and the procedure of numerical experiments were presented in Section 3. Then in Section 4, we showed the optimized single-objective function for the study areas, compared with traditional single-objective function NSE and multi-objective calibration. Finally, conclusions were drawn in Section 5.

2. Definition of objective function

Most objective functions used for calibration of hydrological models contain a summation of the error term, i.e. the difference between the simulated and observed variable (Krause et al., 2005). In addition, absolute or square function is introduced to avoid the offset between errors with opposite signs. In order to normalize the objective function, a baseline or benchmark, such as average of observed variables, is often used in many objective functions. As the average of observed variables is a constant value, such linear normalization has no impact on the calibration results. Therefore, only the power function of absolute errors is of major importance for model calibration.

Different exponent value of the objective function leads to emphasizing different hydrological response mode in the calibration. For example, the exponent value of NSE function is 2 (see Eq. (2) below). When we try to maximize NSE to find the best parameter set of a hydrological model, it leads to matching high flow parts of the hydrograph at the expense of low flow parts, because errors related to high flows are amplified and tend to be larger than those related to low flows. Conversely, if the exponent is smaller, errors related to low flows tend to be relatively emphasized and low flows can thus be better replicated.

Based on the above analysis, we proposed a hypothesis that appropriate (optimal) exponent value can balance multi-response modes of the hydrograph. To explore the optimal exponent value (OEV), a general form of the new single-objective function is defined as,

$$C = \sum_{t=1}^{n} \left| Q_{s,t} - Q_{o.t} \right|^{b} \tag{1}$$

where, C is the newly proposed single-objective evaluation criterion, $Q_{s,t}$ is the simulated streamflow at time t, $Q_{o.t}$ is the observed streamflow at time t, n is the length of entire simulation period. Given a specific watershed and a proper hydrological model, there is an optimal exponent b with which the proposed objective function can simultaneously address multi-response modes of the hydrograph. We call the objective function with the exponent value of OEV the optimal single-objective function (OSOF). Practical hydrological modeling experience suggests that NSE is an appropriate objective function for replicating high flows, thus the exponent b was assumed less than or equal to 2 in this study. Here we take streamflow as a demonstrating hydrological variable, but one can easily extend to other variables in the future study.

3. Materials and Methods

3.1 Hydrological model

In this study, the Xinanjiang model (Zhao, 1992) was used as a runoff generation module, the Model for Scale Adaptive River Transport (MOSART) (Li et al., 2013) was used as a routing module. The Xinanjiang model was proposed in 1973 and is based on the concept of runoff formation on repletion of storage. The runoff generation is composed of three components: surface, subsurface, and groundwater, which are calculated based on tension water capacity and free water

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

capacity. Based on soil moisture and potential evapotranspiration, the evapotranspiration is calculated from three vertical layers. This model has been widely used in humid and semi-humid watersheds (Bao et al., 2011; Cheng et al., 2006; Gan et al., 1997; Ju et al., 2009; Li et al., 2009; Zhao, 1992), and cannot be used in watersheds where snow processes is of importance because it has no snow module. Two criteria, therefore, need to be satisfied for the watersheds used in this study: the climate is humid or semi-humid, and snow processes can be ignored.

A runoff routing model, called MOSART, was developed by Li et al. (2013) and has been applied at different spatial resolutions. In this model, surface runoff was assumed to be first routed across hillslopes and then discharged along with subsurface runoff into a sub-network channel before entering the main channel. The sub-network channel is a hypothetical equivalent to all tributaries combined, i.e., with equivalent transport capacity. For the hillslope and sub-network channel routing, the kinematic wave routing method is used. For the main channel routing, both kinematic and diffusion wave routing methods are available, but the former was used in this study.

In summary, the hydrological model has 14 calibrated parameters: 5 parameters related to evaporation, i.e., K, C, WUM, WLM, WDM; 2 parameters related to runoff generation, i.e., B, IMP; 4 parameters related to runoff partition, i.e., SM, EX, KG, KSS; and 3 parameters related to runoff routing, i.e., C_nh, C_nr, C_twidth. The physical meaning of the model parameters and the range of parameter values were given in Table 1.

Table 1. Parameters of the hydrological model

Parameter	Physical meaning	Unit	Range
K	Evaporation pan coefficient	-	0.70-0.99
C	Coefficient of the deep layer, that depends on the proportion of the basin area covered by vegetation with deep roots	-	0.1-0.4
WUM	Averaged soil moisture capacity of the upper layer	mm	5-120

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

WLM	Averaged soil moisture capacity of the lower layer	mm	5-120
WDM	Averaged soil moisture capacity of the deep layer	mm	5-120
B	Representation of the non-uniformity of the spatial distribution of soil moisture storage capacity over the watershed	-	0.1-0.7
IMP	Percentage of impervious and saturated areas in the watershed	-	0-0.05
SM	Areal mean free water capacity of the surface soil layer, which represents the maximum possible deficit of free water storage	mm	1-30
EX	Exponent of the free water capacity curve influencing the development of the saturated area	-	0-2
KG	Outflow coefficients of the free water storage to groundwater relationships	-	0-0.4
KSS	Outflow coefficients of the free water storage to interflow relationships	-	0-0.6
C_nh	Scale factor for Manning's roughness coefficient for hillslope routing	-	0-1
C_nr	Scale factor for Manning's roughness coefficient for channel routing	-	0-1
C_twidth	Coefficient to account for the difference between the hypothetical sub-channel network and real tributary network	-	0.1-10.0

3.2 Study area and data

The Model Parameter Estimation Experiment (MOPEX) watersheds were chosen as the study areas. The MOPEX dataset was described by Duan et al. (2006) and can be downloaded from http://www.nws.noaa.gov/oh/mopex/mo_datasets.htm. Daily mean areal precipitation, potential evaporation, and streamflow are available for 438 watersheds ranging from 67 to 10329 km^2 across the United States. As shown in Kollat et al. (2012), 392 of the MOPEX watersheds have 11 complete years of data from 1 Oct. 1961 to 30 Sep. 1972. Among these watersheds, 196 watersheds across the eastern United States (shown in Figure 1) were selected because of the applicability of the Xinanjiang hydrological model. Daily precipitation and potential evaporation were used to drive the hydrological model. Daily streamflow series were used to calibrate the

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

hydrological model. The periods of 1 Oct. 1961 - 30 Sep. 1962, 1 Oct. 1962 - 30 Sep. 1972, and 1 Oct. 1972 - 30 Sep. 1982 were selected for warm-up, model calibration, and validation respectively.

Faustini et al. (2009) developed the downstream hydraulic geometry relationships for bankfull channel width as a function of drainage area for nine aggregate eco-regions comprising the conterminous United States using 1588 sites from the US Environmental Protection Agency's National Wadeable Streams Assessment. Using these relationships, we calculated the bankfull width for each watershed in this study. The channel slope, sub-channel slope, and drainage density were calculated based on the National Hydrography Plus Dataset (NHDPlus) which is a geo-spatial, hydrologic framework dataset incorporating the National Hydrography Dataset (NDH), the National Elevation Dataset (NED), and the Watershed Boundary Dataset (WBD), and can be downloaded from http://www.horizon-systems.com/NHDPlus/index.php. In this study, the minimum slope was set to 0.005%. The main channel was defined as the channel draining out of the watershed outlet and/or into the downstream watershed. Following Eqns. (6) and (8) in Li et al. [2013], the length and width of hypothetical sub-network channel were estimated based on the drainage density value derived from NHDPlus and the aforementioned hydraulic geometry relationships.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

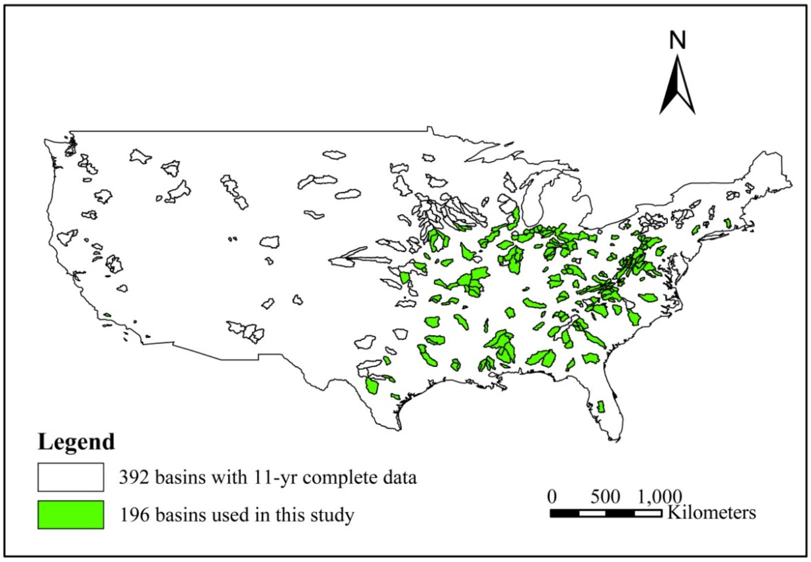


Figure 1. Map of the 196 MOPEX watersheds used in this study. These watersheds were selected from the 392 watersheds with 11 years of data [Kollat et al., 2012].

3.3 Optimization algorithm

The Epsilon Dominance Nondominated Sorted Genetic Algorithm-II (ε-NSGAII) (Kollat and Reed, 2006) was chosen as the optimization algorithm for model calibration in this study. The ε-NSGAII integrates epsilon dominance strategy (Laumanns et al., 2002) and automatic parameterization (Reed et al., 2003) with the NSGA-II (Deb et al., 2002). The algorithm has been frequently used in hydrological modeling and has been demonstrated to be consistently better compared to other state-of-the-art evolutionary algorithms (Kollat and Reed, 2006; Sun et al., 2014; Tang et al., 2006).

3.4 Evaluation framework

According to Kollat et al. (2012), four widely used objective functions focusing on peak flows,

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

low flows, water balance, and flashiness, respectively, were applied in this study. The first one is Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) which emphasizes peak flows, as shown in Eq. (2),

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{s,t} - Q_{o,t})^2}{\sum_{t=1}^{n}(Q_{o,t} - \overline{Q_o})^2} \tag{2}$$

where, $\overline{Q_o}$ is the mean observed streamflow over the entire simulation period of length n.

The second objective is Transformed Root Mean Square Error (TRMSE) which emphasizes low flows, as shown in Eq. (3),

$$TRMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Q'_{s,t} - Q'_{o,t})^2}\text{, where } Q' = \frac{(1+Q)^{\lambda} - 1}{\lambda} \tag{3}$$

where, $Q'_{s,t}$ is the Box-Cox transformed (Box and Cox, 1964) simulated streamflow at time *t*, $Q'_{o,t}$ is the Box-Cox transformed observed streamflow at time *t*, Q' is the Box-Cox transformation of the streamflow Q, λ is a constant ($\lambda = 0.3$).

The third objective is Runoff Coefficient Percent Error (ROCE) which emphasizes water balance, as shown in Eq. (4),

$$ROCE = \frac{1}{Y}\sum_{y=1}^{Y}\left|\frac{\overline{Q_{s,y}}}{\overline{Q_{o,y}}} - 1\right| \times 100\% \tag{4}$$

where, $\overline{Q_{s,y}}$ is the mean annual simulated streamflow, $\overline{Q_{o,y}}$ is the mean annual observed streamflow, Y is the number of years in the simulation period.

The fourth objective is Slope of the Flow Duration Curve (SFDCE) which emphasizes flashiness of the hydrological response, as shown in Eq. (5),

$$SFDCE = \left|\frac{Q_{s,67\%} - Q_{s,33\%}}{Q_{o,67\%} - Q_{o,33\%}} - 1\right| \times 100\% \tag{5}$$

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

where, $Q_{s,67\%}$ and $Q_{s,33\%}$ are the 67th and 33rd percentile of the simulated streamflow, $Q_{o,67\%}$ and $Q_{o,33\%}$ are the 67th and 33rd percentile of the observed streamflow.

Price et al. (2012) provided an aggregated multi-objective functions termed the composite likelihood index (CL) compositing three metrics in order to allow trade-offs in fitting high flow, low flow, and flow variability components. In this study, Price et al.'s principle (2012) was applied to aggregate these four selected objective functions described above for evaluating the proposed objective function with different exponent values. Following the process in Price et al. (2012), we first transformed the four objective functions to a similar scale for aggregation, as shown in Eq. (6),

$$\begin{aligned}
\theta_{NSE} &= \frac{\max(0, NSE_i)}{\sum_{i=1}^{L}\max(0, NSE_i)} \\
\theta_{TRMSE} &= \frac{1-\min(1,|1-TRMSE_i|)}{\sum_{i=1}^{L}\left[1-\min(1,|1-TRMSE_i|)\right]} \\
\theta_{ROCE} &= \frac{1-\min(1,|1-ROCE_i|)}{\sum_{i=1}^{L}\left[1-\min(1,|1-ROCE_i|)\right]} \\
\theta_{SFDCE} &= \frac{1-\min(1,|1-SFDCE_i|)}{\sum_{i=1}^{L}\left[1-\min(1,|1-SFDCE_i|)\right]}
\end{aligned} \tag{6}$$

where, $\theta_{NSE}, \theta_{TRMSE}, \theta_{ROCE}, \theta_{SFDCE}$ are the scaled $NSE, TRMSE, ROCE, SFDCE$, respectively; the subscript i represents each calibration run with different exponent value of the proposed objective function. L is the total number of calibration runs, which is 9 in this study (see Section 3.5). Then CL equally weights these four scaled objective functions, as shown in Eq. (7),

$$CL = mean(\theta_{NSE}, \theta_{TRMSE}, \theta_{ROCE}, \theta_{SFDCE}) \tag{7}$$

3.5 Framework to search for an optimized single-objective function

In order to investigate the possibility of using the proposed single-objective function to

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

compromise the multi-response modes of hydrograph, four experiments were designed in this study. The purpose of the first experiment is to identify the OEV for each of 196 MOPEX watersheds, which is the prerequisite for application of the proposed objective function. In this study, we designed nine numbers quasi-uniformly distributed in the range of exponent values [0-2.0] (see Section 2), 0.1, 0.3, 0.5, 0.7, 1.0, 1.3, 1.5, 1.7, and 2.0. For each MOPEX watershed, 9 automatic calibrations were conducted with the objective function defined using the 9 exponent values respectively. The OEV was then identified as the exponent value corresponding to the optimal value of CL defined in Section 3.4.

According to the identified OEVs, 196 MOPEX watersheds could be grouped into 9 categories. We selected one representative watershed from each category for further exploration. In the context of single-objective calibration, the second experiment was implemented to compare the OSOF with the most widely used objective function NSE. Then in the third experiment, we did a comparative study between single-objective calibration with the OSOF and multi-objective calibration with four objective functions described in Section 3.4. Finally, in the fourth experiment, in order to investigate the robustness of calibration with the OSOF, the simulated streamflow was validated in a period of 10 years. We also compared the hydrograph replicating capability of optimized parameters using single-objective calibration and multi-objective calibration in the fourth experiment.

4. Results and discussion

4.1 Single-objective calibration

Figure 2 shows the results of 196×9 single-objective calibrations in the first experiment. In order

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

to investigate the effect of exponent values on the result of single-objective calibration, four widely used metrics, i.e., NSE, TRMSE, ROCE, and SFDCE, were adopted to evaluate 9 optimized simulations for each of 196 MOPEX watersheds.

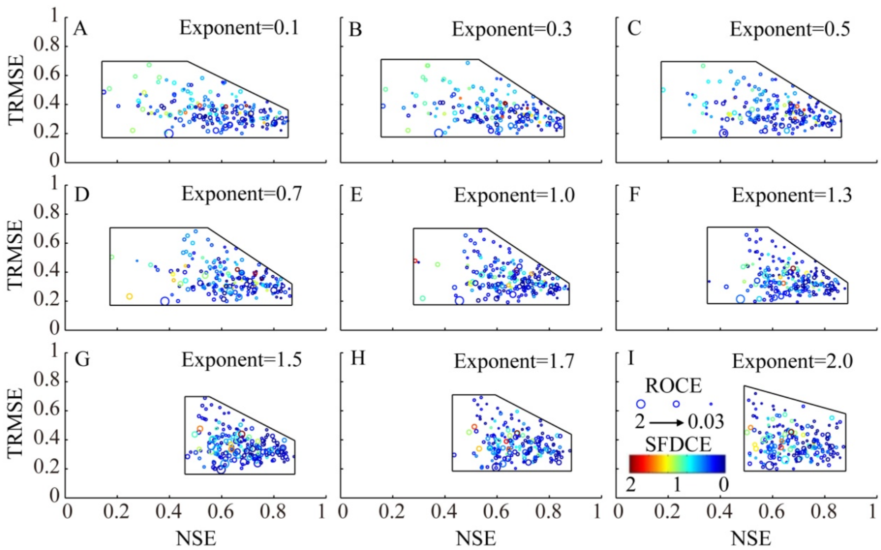


Figure 2. The evaluation merits for single-objective calibration. Each sub-figure represents the results calibrated by one exponent value but for all the 196 watersheds. NSE was plotted on the horizontal axis, TRMSE was plotted on the vertical axis, ROCE was plotted as the diameter of circles, SFDCE was plotted as color. A-I represent results with different exponent values. The shapes enclosed by the black envelop lines show the distribution range of the results.

From Figure 2(I) to Figure 2(A), as the exponent value decreases from 2.0 to 0.1 the distribution range of NSE increases significantly from 0.5-0.9 to 0.15-0.86. The larger the exponent value is, the better (high score and narrow range) the NSE evaluation merit becomes. This indicates that the objective function with larger exponent value tends to match high flows. For the TRMSE (more related to low flow), its distribution ranges are almost unchanged with different exponent values. The average TRMSEs of 196 watersheds for nine exponent cases were

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions

calculated and shown in Figure 2. When the exponent decreases from 2.0 to 0.1, the average TRMSE decreases from 0.386 to 0.351. Also, the proportion of the TRMSE greater than 0.4 (worse performance) was calculated for each figure, which also shows a decreasing trend from 37.3% to 21.8%. These distribution characteristics of TRMSE show that the simulations of low flows are improved when the exponent value decreases. As the TRMSE tends to concentrate on lower portions, this demonstrates that the objective function with lower exponent value emphasizes low flows. With respect to ROCE and SFDCE, there is no obvious trend detected from the figures. The general performances of ROCE and SFDCE are reasonably well for most calibrations in this study.

In Figure 2(I), many results fall in the bottom right corner, which means that they have a sound simulation for both high flows and low flows. In addition, the other two metrics emphasizing water balance and flashiness of hydrological response are also reasonable. For these watersheds, the objective function with exponent of 2.0 is suitable. However, there are many other watersheds with good simulation of high flows at the expense of poor simulation of low flows, which indicates that NSE is not a reasonable objective function for all kinds of watersheds (also see the discussion in Schaefli and Gupta, 2007).

The shapes enclosed by envelop lines of the distribution ranges were also plotted in Figure 2. We found that the right sides of the shapes in Figure 2(A)-(F) are narrower than those in Figure 2(G)-(I), which means that some simulations of the watersheds with sound NSE and poor TRMSE are improved when the exponents are moderately adjusted. Especially in Figure 2(C)-(G), many results concentrate in the bottom right corner, indicating that the hypothesis that a proper exponent value can compromise multi-response modes of the hydrograph be reasonable.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

4.2 The proposed objective function with the optimal exponent value

The results of single-objective calibration discussed above are evaluated by the composite likelihood index to identify the OEV for each watershed. Figure 3 shows the identified OEVs of the proposed objective function for the 196 MOPEX watersheds. Most of the watersheds with lower OEVs locate in the northwest region, while the watersheds with larger OEVs generally locate in the northeast region. According to the identified OEVs, these watersheds were grouped into eight categories, and then eight representative watersheds were arbitrarily selected from these categories as shown in Figure 3.

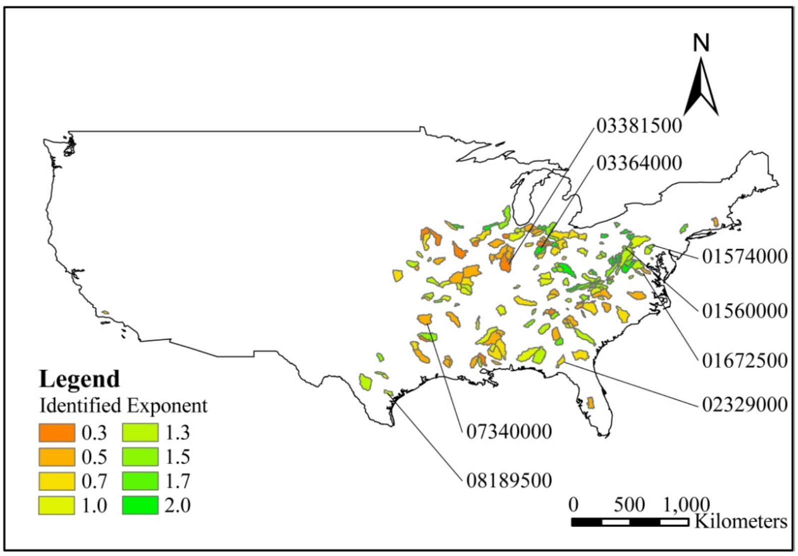


Figure 3. The optimal exponent values of the proposed objective function for the 196 MOPEX watersheds. Labels are provided for eight watersheds selected as representative watersheds to be explored in detail based on the identified OEVs.

Figure 4 shows the histogram of these identified OEVs. Identified OEVs distribute from 0.3 to 2.0. For 22 of the 196 MOPEX watersheds, the objective function with exponent of 2.0 is the optimal, which are all located in the bottom right corner of Figure 2(I) with sound simulations of high flows and low flows, as well as water balance and flashiness of hydrological response. Most watersheds tend to prefer the OEV of 0.5, 0.7, 1.0, 1.3, and 1.5. There are also 12 watersheds with OEV of 0.3. The regional distribution of these watersheds concentrate in northwest as shown in Figure 3.

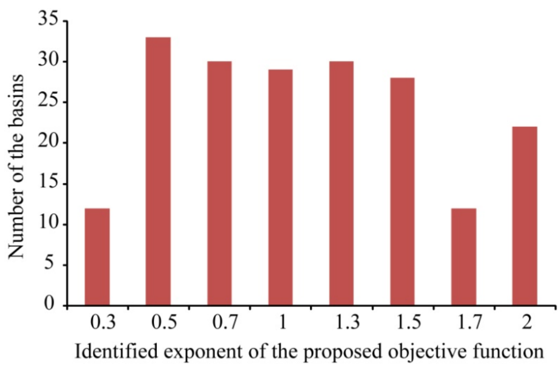


Figure 4. Histogram of identified OEVs.

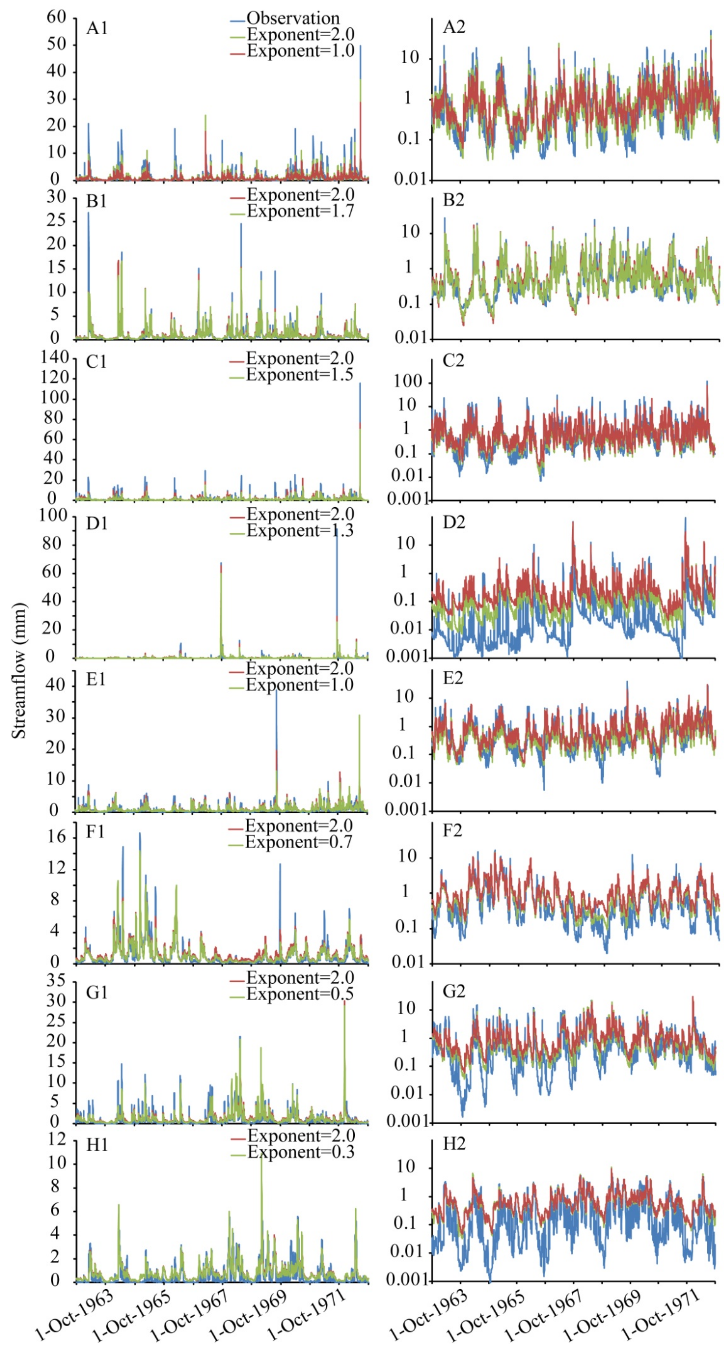
4.3 Comparison between the proposed objective function and NSE

In the context of single-objective calibration, NSE is a traditionally widely used objective function. In order to verify the advantage of the proposed objective function, we compared the OSOF and NSE. Figure 5 shows the results of comparative study at eight representative watersheds. To be noted, the identified OEV is the same as NSE in the watershed 01560000. As shown in Figure 5(A1) and 5(A2), we found that the results using the proposed objective function with exponent of 2.0 (similar to NSE) are much better in both high flows and low flows when compared to that with exponent of 1.0 which is arbitrarily selected. Other figures on the left side show that the simulations of high flows calibrated by the OSOF have little impact on replicating high flows when compared to NSE. However, focusing on the low flows as shown in the right figures with logarithmic vertical axis, the calibration with the OSOF significantly improve the simulations especially in Figure 5(D2) to 5(G2). Four widely used metrics including NSE, TRMSE, ROCE and SFDCE, were applied to quantitatively analyze the difference between the results using two objective functions. Table 2 shows that the OSOF simultaneously improves TRMSE, ROCE, and SFDCE with a slight impairment on NSE at most watersheds, except for watershed 02329000. All the results indicate that there does exist a single-objective function (OSOF here) that can compromise multi-response modes of hydrograph, which is usually not the traditionally used single-objective function NSE.

Table 2. Evaluation merits of the simulations calibrated by the optimal single-objective function (OSOF) and NSE

ID of	OSOF				NSE			
Watersheds	NSE	TRMSE	ROCE	SFDCE	NSE	TRMSE	ROCE	SFDCE
01560000	0.640	0.495	0.093	0.093	0.640	0.495	0.093	0.093
03364000	0.787	0.247	0.121	0.039	0.806	0.277	0.255	0.243
01574000	0.679	0.456	0.092	0.021	0.692	0.504	0.174	0.219
08189500	0.652	0.345	1.510	5.897	0.655	0.388	2.997	7.468
01672500	0.615	0.320	0.118	0.096	0.642	0.385	0.356	0.549
02329000	0.848	0.253	0.186	0.155	0.854	0.335	0.548	0.046
07340000	0.726	0.367	0.088	0.141	0.734	0.438	0.256	0.285
03381500	0.714	0.326	0.471	0.017	0.720	0.341	0.503	0.018

Hydrology and Earth System Sciences
Open Access
Discussions



323

324 Figure 5. Comparison between the results using proposed objective function and NSE in the calibration

325 period for 8 representative watersheds. The vertical axis of the four figures in the right column is

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions

logarithmic. From A-H, the identified OEVs decrease from 2.0 to 0.3, the IDs of these watersheds are 01560000, 03364000, 01574000, 08189500, 01672500, 02329000, 07340000, and 03381500, respectively. Specifically, for A1 and A2, the proposed objective function is same as NSE, we randomly selected the result with exponent of 1.0 for comparison.

4.4 Comparison between single-objective calibration and multi-objective calibration

Figure 6 compares single-objective calibration with the OSOF and multi-objective calibration with four widely used objective functions in the calibration period at eight representative watersheds. At all these watersheds, the simulation of OSOF calibration locates in the hydrograph ranges enclosed by the envelop lines of Pareto results using multi-objective calibration and can be able to capture the major variability of the hydrograph. In Figure 6(A), 6(C), 6(E) and 6(G), the hydrograph ranges are so narrow that the difference between the results of single-objective function and multi-objective functions can be ignored. It means that at these watersheds multi-objective calibration does not extract more information from historical data. Figure 7(A) shows the flow duration curves (FDCs) at watershed 01574000 (the same as Figure 6(C)). The uncertainty of the streamflow frequency distribution is also obvious smaller than that in the other two watersheds in Figure 7. In Figure 6(B), the simulation using the OSOF calibration is near to the upper bound of uncertainty zone at low flow portions, and the observed hydrograph especially for low flows is generally slightly greater than the simulations. Similarly, in Figure 7(B), the simulated FDC using single-objective calibration almost replicates the observed FDC. It also means that the uncertainty bound using multi-objective calibration does not provide useful information in this case. On the contrary, in Figure 6(D), 6(F), and 6(H), results of multi-objective calibration contain some simulations which can better capture the variability of the streamflow, but the best simulations in the hydrograph ranges are also far away from the observation especially for low flows. Corresponding to Figure 6(H), Figure 7(C) shows the similar results from the perspective of frequency distribution. For example, the 80th percentile of the lower bound of Pareto results is 0.09 mm, while that of the observed streamflow is 0.02 mm. Overall, single-objective calibration with the OSOF can compromise multi-response modes of the hydrograph to obtain a relatively sound simulation, which is comparable to the result of multi-objective calibration.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

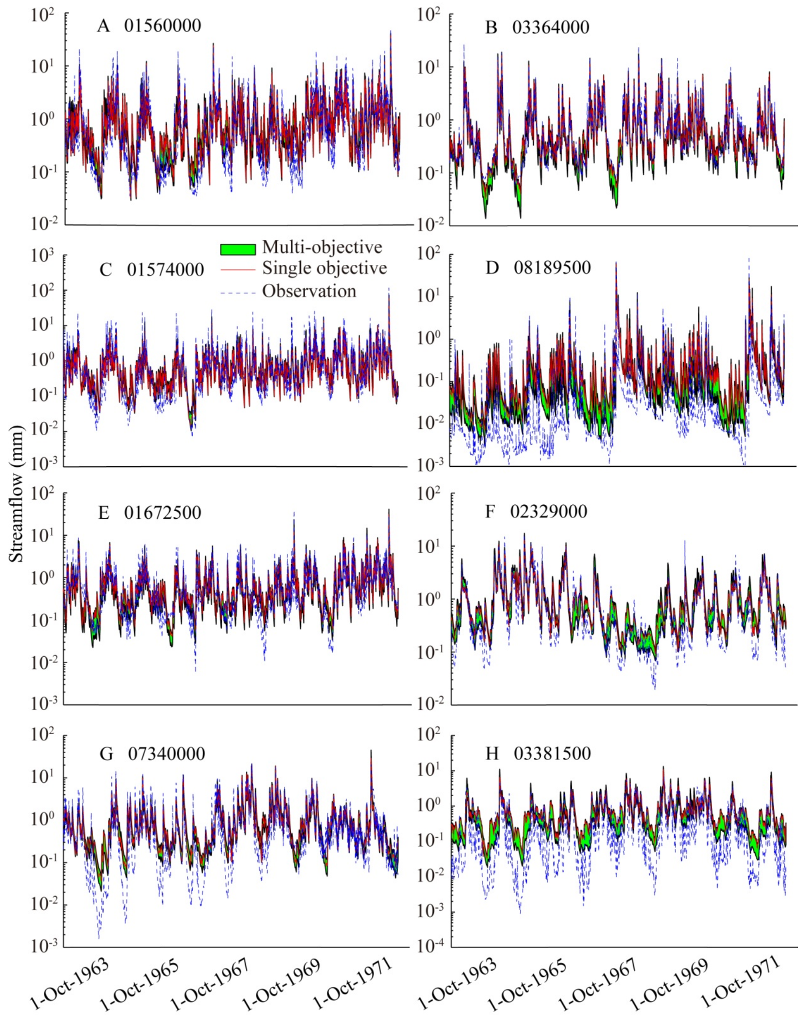


Figure 6. Comparison between the results using single objective calibration with the optimal single-objective function and multi-objective calibration with four widely used objective functions in the calibration period for 8 representative watersheds. The vertical axis of the figure is logarithmic. From A-H, the identified optimal exponent value decreases from 2.0 to 0.3.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Discussions
Open Access

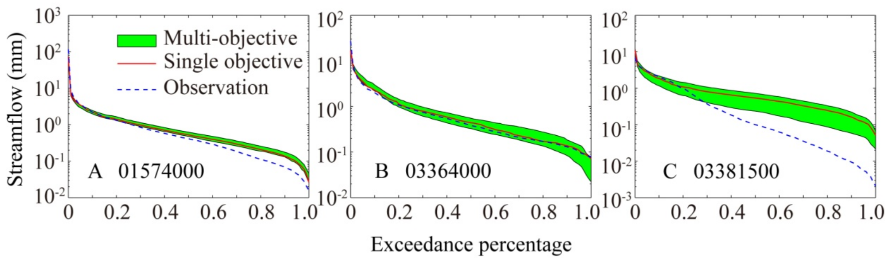
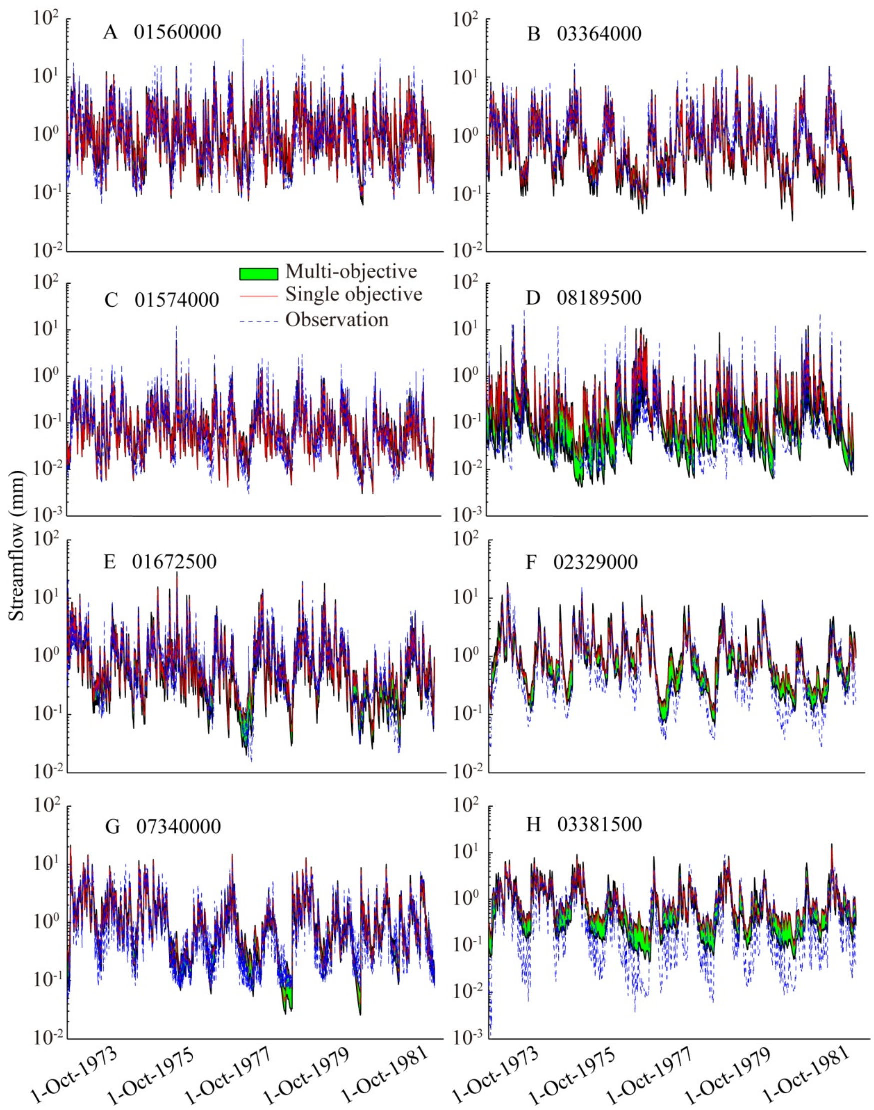


Figure 7. Observed and simulated flow duration curves (FDCs) for 3 representative watersheds. The vertical axis of the figure is logarithmic.

4.5 Model validation

The above analyses demonstrate that the proposed OSOF is effective during the calibration period. In this section, we aimed to test its robustness by comparing the performance of single-objective calibration with OSOF to that of multi-objective calibration during the validation period. The simulations using OSOF calibration can also capture the major patterns of the hydrograph during the validation period as shown in Figure 8. Similar to the results during the calibration period (Figure 6), in Figure 8(A), 8(B), 8(C), 8(E), and 8(G), the hydrograph ranges in the validation period are also narrow, which means that not only in the calibration period can the results of OSOF calibration be comparable to that of multi-objective calibration, but also in the validation period at these watersheds. At the other three watersheds, multi-objective calibration provides more useful information about the hydrograph, but there is also a wide gap between the best simulations in the hydrograph ranges and the observations. In general, parameter set identified by the proposed optimal single-objective function has a robust performance in hydrological modeling.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016



377

378 Figure 8. Comparison between the results using single objective calibration with the proposed objective
379 function and multi-objective calibration with four widely used objective functions in the validation
380 period for 8 representative watersheds. The vertical axis of the figure is logarithmic. From A-H, the
381 identified optimal exponent values decrease from 2.0 to 0.3.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Hydrology and Earth System Sciences
Open Access
Discussions

5. Discussion and Conclusions

In this study, we proposed a new methodology that can compromise multi-response modes (i.e., multi-objective functions) of the hydrograph by using an optimized single-objective function. The single-objective function is generally defined as a power function of the absolute error between observed and simulated streamflow, and the exponent of power function is then optimized when applying to a specific watershed. The methodology was applied to 196 model parameter estimation experiment (MOPEX) watersheds across the eastern United States using Xinanjiang model. The optimal exponent value for each watershed was obtained by targeting four popular objective functions focusing on peak flows, low flows, water balance, and flashiness, respectively. The results show that the optimal single-objection function can achieve a better hydrograph simulation compared to traditional Nash-Sutcliffe efficiency for most watersheds, as well as can balance high flow part and low flow part of the hydrograph to be comparable to multi-objective calibration. Our work demonstrates the potential of single-objective function to simultaneously address multi-response modes of the hydrograph.

Multi-objective calibration is based on the concept of Pareto optimal, which defines the improvement strictly as that given an initial solution, a change to a better solution results in making at least one objective function better without making any other objective function worse. Actually, as discussed in Tekleab et al. (2011), modelers may be willing to accept suboptimal performance of one aspect of the streamflow series in order to improve accuracy in one or more other aspects. The proposed optimal single-objective function in this study essentially relaxes the strict definition of Pareto optimal. For example, when compared to NSE, the proposed OSOF improves two or three objective functions which represent corresponding modes of hydrological responses at a slight expense of the simulation of high flows, which is not Pareto improvement. Kollat et al. (2012) found that the meaningful multi-objective trade-offs are far less frequent than prior literature has suggested. They introduced the concept of epsilon-dominance, a way to relax the Pareto optimal, to obtain meaningful results, and identified only one solution at many MOPEX watersheds. While multi-objective calibration is helpful to address multi-modes, single-objective calibration with a proper objective function can also identify a good compromise solution against multiple criteria which would be applicable for hydrological planning and management.

In addition, automatic calibration (whether single-objective or multi-objective) of hydrological models is usually implemented with the same objective function or combination of objective functions for different watersheds. However, we may expect different objective functions for watersheds with significant differences in terms of climatic condition, vegetation cover, land use/cover, soil texture, etc. This study demonstrates this idea by identifying a specific objective function for a specific watershed.

Theoretically, the proposed methodology utilizes the trade-off on the exponent value of power function to substitute for the trade-offs on multiple objectives. It could be practically adopted in the hydrological modeling if the optimized exponent could be determined *a priori* according to hydrological/climatic/landscape characteristics in a specific watershed. In the following study, an empirical equation is expected to be established to relate the optimal exponent value of the proposed objective function with some watershed and hydrograph indices or a combination of these indices.

Acknowledgements

This work has been funded by the National Science Foundation of China (NSFC 51190092, U1202232), the foundation of State Key Laboratory of Hydroscience and Engineering of Tsinghua University (2016-KY-03). H. Li was supported by the Office of Science of the U.S. Department of Energy as part of the Earth System Modeling Program. Their supports are greatly appreciated. The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RLO1830.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

References

Bao, H.J. et al., 2011. Coupling ensemble weather predictions based on TIGGE database with Grid-Xinanjiang model for flood forecast. Advances in Geosciences, 29: 61-67.

Bekele, E.G., Nicklow, J.W., 2007. Multi-objective automatic calibration of SWAT using NSGA-II. Journal of Hydrology, 341(3-4): 165-176.

Booij, M.J., Krol, M.S., 2010. Balance between calibration objectives in a conceptual hydrological model. Hydrological Sciences Journal, 55(6): 1017-1032.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B, 26(2): 211-252.

Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research, 36(12): 3663-3674.

Cheng, C.-T., Zhao, M.-Y., Chau, K.W., Wu, X.-Y., 2006. Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. Journal of Hydrology, 316(1-4): 129-140.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2): 182-197.

Duan, Q. et al., 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology, 320(1-2): 3-17.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Hydrological Sciences Journal, 55(1): 58-78.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Faustini, J.M., Kaufmann, P.R., Herlihy, A.T., 2009. Downstream variation in bankfull width of wadeable streams across the conterminous United States. Geomorphology, 108(3-4): 292-311.

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(3): 93-108.

Gan, T.Y., Dlamini, E.M., Biftu, G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. Journal of Hydrology, 192(1-4): 81-103.

Gill, M.K., Kaheil, Y.H., Khalil, A., McKee, M., Bastidas, L., 2006. Multiobjective particle swarm optimization for parameter estimation in hydrology. Water Resources Research, 42(7): 257-271.

Guinot, V., Cappelaere, B., Delenne, C., Ruelland, D., 2011. Towards improved criteria for hydrological model calibration: theoretical analysis of distance- and weak form-based functions. Journal of Hydrology, 401(1-2): 1-13.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2): 80-91.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research, 34(4): 751-763.

Hall, J., W., Tarantola, S., Bates, P., D., Horritt, M., S., 2005. Distributed sensitivity analysis of flood inundation model calibration. Journal of Hydrologic Engineering, 131(2): 117-126.

Jain, S.K., Sudheer, K.P., 2008. Fitting of hydrologic models: a close look at the Nash-Sutcliffe index. Journal of Hydrologic Engineering, 13: 981-986.

Ju, Q. et al., 2009. Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. Neurocomputing, 72(13-15): 2873-2883.

Khu, S.-T., Madsen, H., di Pierro, F., 2008. Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering. Advances in Water Resources, 31(10): 1387-1398.

Kollat, J.B., Reed, P.M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. Advances in Water Resources, 29(6): 792-807.

Kollat, J.B., Reed, P.M., Wagener, T., 2012. When are multiobjective calibration trade-offs in hydrologic models meaningful? Water Resources Research, 48(3): 221-239.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5: 89-97.

Laumanns, M., Thiele, L., Deb, K., Zitzler, E., 2002. Combining convergence and diversity in evolutionary multiobjective optimization. Evolutionary Computation, 10(3): 263-282.

Li, H. et al., 2013. A Physically Based Runoff Routing Model for Land Surface and Earth System Models. Journal of Hydrometeorology, 14(3): 808-828.

Li, H., Zhang, Y., Chiew, F.H.S., Xu, S., 2009. Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. Journal of Hydrology, 370(1-4): 155-162.

Madsen, H., 2000. automatic calibration of a conceptual rainfall-runoff model. Journal of Hyrology, 235: 276-288.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

automatic calibration with multiple objectives. Advances in Water Resources, 26: 205-216.

Madsen, H., Wilson, G., Ammentorp, H., Christian, 2002. Comparison of different automated strategies for calibration of rainfall-runoff models. Journal of Hydrology, 261: 48-59.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resources Research, 45(6): 735-748.

McCuen, R.H., Snyder, W.M., Cutter, A.G., 2006. Evaluation of the Nash-Sutcliffe efficiency index. Journal of Hydrologic Engineering, 11(6): 597-602.

Muleta, M.K., 2012. Model Performance Sensitivity to Objective Function during Automated Calibrations. Journal of Hydrologic Engineering, 17(6): 756-767.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I. A discussion of principles. Journal of Hydrlogy, 10(3): 282-290.

Pokhrel, P., Gupta, H.V., 2010. On the use of spatial regularization strategies to improve calibration of distributed watershed models. Water Resources Research, 46(1): 1-17.

Price, K., Purucker, S.T., Kraemer, S.R., Babendreier, J.E., 2012. Tradeoffs among watershed model calibration targets for parameter estimation. Water Resources Research, 48(10): 136-151.

Reed, P., Minsker, B.S., Goldberg, D.E., 2003. Simplifying multiobjective optimization: An automated design methodology for the nondominated sorted genetic algorithm-II. Water Resources Research, 39(7): 257-271.

Schaefli, B., Gupta, H.V., 2007. Do Nash values have value? Hydrological Processes, 21: 2075-2080.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 3 March 2016

Sun, Y., Tian, F., Yang, L., Hu, H., 2014. Exploring the spatial variability of contributions from climate variation and change in catchment properties to streamflow decrease in a mesoscale basin by three different methods. Journal of Hydrology, 508: 170-180.

Tang, Y., Reed, P., Wagener, T., 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? Hydrology and Earth System Sciences, 10: 289-307.

Tang, Y., Reed, P.M., Kollat, J.B., 2007. Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications. Advances in Water Resources, 30(3): 335-353.

Tekleab, S. et al., 2011. Water balance modeling of Upper Blue Nile catchments using a top-down approach. Hydrology and Earth System Sciences, 15(7): 2179-2193.

van Griensven, A., Bauwens, W., 2003. Multiobjective autocalibration for semidistributed water quality models. Water Resources Research, 39(12): 1348-1357.

van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources, 32(8): 1154-1169.

Vrugt, J.A., Bouten, W., Gupta, H.V., Sorooshian, S., 2002. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. Water Resources Research, 38(12): 1312-1324.

Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resources Research, 39(8): 1214-1232.

Wagener, T., 2003. Evaluation of catchment models. Hydrological Processes, 17(16): 3375-3378.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology, 204: 83-97.

Zhao, R.-J., 1992. The Xinanjiang model applied in China. Journal of Hydrology, 135: 371-381.