Sensitivity Analysis-Based Automatic Parameter Calibration of the VIC Model for Streamflow Simulations Over China

Jiaojiao Gou¹, Chiyuan Miao¹, Qingyun Duan¹, Qiuhong Tang², Zhenhua Di¹, Weihong Liao¹, Jingwen Wu¹, and Rui Zhou⁴

¹State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing, China, ²Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, ³State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, China, ⁴China Three Gorges International Corporation, Beijing, China

Abstract
Model parameter calibration is a fundamentally important stage that must be completed before applying a model to address practical problems. In this study, we describe an automatic calibration framework that combines sensitivity analysis (SA) and an adaptive surrogate modeling-based optimization (ASMO) algorithm. We use this framework to calibrate catchment-specific sensitive parameters for streamflow simulation in the variable infiltration capacity (VIC) model with a 0.25° spatial resolution over 10 major river basins of China from 1960 to 1979. We found that three parameters—the infiltration parameter (B) and two of the soil depth parameters (D₁, D₂)—are highly sensitive in most basins, while other parameter sensitivities are strongly related to the dynamic environment of the basin. Compared with directly calibrating the seven parameters recommended for the default calibration procedure, our framework not only reduced the computing time by two thirds through opting out of insensitive parameters (type I error) but also improved the Nash-Sutcliffe model efficiency coefficient (NSE) for optimized results when it identified a missing sensitive parameter (type II error) in the case study river basins. Results show that the SA-based ASMO framework is an effective and efficient model-optimization technique for matching simulated streamflow with observations across China. The NSE for monthly streamflow ranged from 0.75 to 0.97 and from 0.71 to 0.97 during the validation and calibration periods, respectively. The calibrated parameters can be applied directly in streamflow simulations across China, and the proposed calibration framework holds important implications for relevant simulation studies in other regions.

1. Introduction
Streamflow modeling plays an important role in many practical hydrological problems (Sorooshian & Chu, 2013), such as water-related hazards forecasting (Wang et al., 2018), water resource management (Liu et al., 2018; Zheng et al., 2019), and management of ecosystem services (Wenger et al., 2010). Additionally, evidence from recent years has shown that the issue of global warming is becoming more serious in nature (IPCC, 2014; Niraula et al., 2017), and there is a need to use regional hydrological modeling tools to identify changes in the magnitude and timing of associated hydrological processes (Niraula et al., 2017; Ran et al., 2017; Wang et al., 2018). However, models are just approximations of natural systems, and there remains substantial discrepancy between model results and reality (Sorooshian & Chu, 2013). Model parameter calibration helps to match model predictions to corresponding observations (Gupta et al., 1999; Shi et al., 2008; Yapo et al., 1998). Strengthening the model parameter calibration procedure improves model performance and therefore is critical for addressing the abovementioned practical water problems for water resource supply and management institutions involving energy companies, resource management agencies, risk reduction agencies, and municipalities.

With increased computing power available and growing demand for the representation of spatial heterogeneity and physical features of river basins, more complex hydrological models are being developed (Nijssen & Bastidas, 2006; Sorooshian & Chu, 2013). This increase in model complexity has been accompanied by a large increase in the number of tunable parameters, which further exacerbates the model parameter...
calibration burden (Muleta & Nicklow, 2005; van Griensven et al., 2006). Therefore, parameter sensitivity analysis (SA) is required to screen out the most important parameters and reduce parameter dimensionality before parameter estimation (Gong et al., 2016; Muleta & Nicklow, 2005). In order to simplify model use and reduce frustration from attempts at performing SA, the modeler generally chooses “sensitive” parameters to be calibrated based on expert judgment or the model builders’ recommendations (Choi et al., 2002; Rosero et al., 2009; Wang et al., 2018; Zhang et al., 2015). However, the sensitivity of model results to parameters varies from watershed to watershed, and which parameters are most important depends on the combined impacts of climate and watershed conditions in a given basin (Demaria et al., 2007; Donigian & Imhoff, 2009). Using the recommended sensitive parameters without regard to the basin characteristics within the calibration procedure may result in involving unnecessary parameters (hereinafter “type I error”) or missing the important parameters (hereinafter “type II error”), which ultimately decreases the efficiency and accuracy of the calibration process. Therefore, SA is a prerequisite for narrowing parameter uncertainty in order to strengthen the effectiveness of hydrological model calibration.

In addition to the issue of how best to identify the sensitive parameters, the reliability of parameter estimation approaches for model calibration is also questionable. Many users perform hydrologic calibration of the model manually using a trial-and-error procedure to tune the parameters (Gupta et al., 1999). However, manual procedures for calibration can be extremely time-consuming, labor-intensive, and subjective, and their success highly depends on the modeler’s experience level (Muleta & Nicklow, 2005). Therefore, various studies since the 1960s have been devoted to the development of automatic model calibration methods with the goal of creating high-speed and objective calibration processes, such as the downhill simplex method (Press et al., 1992), genetic algorithms (Holland, 1992), and the shuffled complex evolution algorithm (Duan et al., 1992). But some researchers felt that these automatic optimization algorithms are prone to producing a single optimized parameter set (Yapo et al., 1998; Price et al., 2012). Bayesian parameter estimation methods have been used in hydrologic models to overcome this problem (Duan & Phillips, 2010; Thiemann et al., 2001) and are designed to provide not only “best fit” calibrated parameter sets but also confidence bands, ensembles, and likelihood distributions. However, those traditional optimization methods are either not able to handle the high dimensionality of distributed hydrological models or are impractical because they require a large number of model runs (Duan et al., 2017). As a recent successful attempt to combine Gaussian processes (GPs) and the shuffled complex evolution algorithm, an adaptive surrogate modeling-based optimization (ASMO) algorithm (Wang et al., 2014) was developed that facilitates searches for optimal parameters of large, complex models using a low number of true model runs.

The variable infiltration capacity (VIC) model was developed as a physically based distributed macroscale hydrological model by Liang et al. (1994), and it has been widely used in streamflow simulations around the world, including China (Zhang et al., 2014), the United States (Niraula et al., 2017), Europe (Greuell et al., 2018), India (Chawla & Mujumdar, 2018), and other regions. VIC is able to capture transient basin discharge and has 46 or more tunable parameters (Bennett et al., 2018). Although we know that parameter sensitivities respond to the dynamic environment in a basin and automatic calibration performs better than manual calibration, the practice of manually calibrating the seven expert-recommended parameters (hereinafter referred to as the seven default parameters) is still largely used in VIC model calibration (Niraula et al., 2017; Ran et al., 2017; Wang et al., 2018; Xie et al., 2007; Zhang et al., 2014). The use of the expert-recommended parameters in the process of calibration precludes our ability to quantify the potential impact of uncertainties in parameters or climate data and consequently makes it impossible to quantify the associated uncertainties in model performance. Considering that past model calibration practices have significantly impacted streamflow forecasting (Shi et al., 2008; Zhang et al., 2014) and have the potential for more impacts in the future, there is a need to develop an advanced framework to improve the efficiency and accuracy of the VIC calibration process.

In addition, China has serious water scarcity problems and has only a quarter of the world average in per capita water resources (Ge et al., 2011). Moreover, the remarkable topographic gradients and monsoon climate make for an uneven distribution of water resources, which intensifies the water crisis in China (Miao et al., 2016; Piao et al., 2010). Thus, there is a need to use hydrological model tools to forecast the amount and distribution of water resources at the country scale. At present, there are many water resources modeling studies for China, but many of them are confined to one river basin, such as the Yangtze River (Zhou et al., 2013), the Yellow River (Sheng et al., 2017), and the Pearl River (Wang et al., 2018). The different
methods used by different researchers in those studies lead to an inability to share the hydrological information and modeling output between river basins, thus the results cannot be compared directly. We are working to provide a large-scale comprehensive simulation of surface runoff in the whole of China; such large-scale modeling has rarely been attempted in previous studies, but it is necessary for China’s water resources management. In support of this goal, in this study, we propose a SA-based ASMO calibration framework for calibrating catchment-specific parameters for streamflow simulations in the VIC model over 10 major river basins of China from 1960 to 1979. In this study, we attempt to address the following three questions:

1. Are there any type I or type II errors in the list of seven default parameters recommended for VIC calibration across China?
2. Assuming that both types of error exist, how well does the SA-based ASMO calibration framework work to avoid those errors?
3. How does our framework perform over China?

2. VIC Model, Parameters, and Data Sets

2.1. VIC Model and Its Parameters

The VIC model is a macroscale hydrological model that represents surface and subsurface hydrologic processes on spatially distributed grid cells (Demaria et al., 2007). VIC accounts for 1-D variably saturated infiltration and includes a decoupled surface routing model (Lohmann et al., 1996) that is able to capture transient basin discharge (Liang et al., 1994). The VIC model can operate in both an energy balance and a water balance mode. For this study, we ran the VIC model version 4.2 in water balance mode at a daily time step. We calculated only the water balance because our interest focused on streamflow generation mechanisms and in determining how to optimize calibration with fewer iterations and faster execution speed. The VIC model divides the soil column of each grid cell into three layers, and the surface flow generated from the upper two soil layers is simulated based on the variable soil moisture capacity curve (Bao et al., 2011). That curve is expressed as

\[ W = W_{mm} \left(1 - (1-A)^{1/B}\right) \]  

(1)

where \( W \) and \( W_{mm} \) are the soil moisture capacity at a grid point and the maximum soil moisture capacity at that point, respectively; \( A \) is the fraction of area within the grid cell for which the soil moisture capacity is less than \( W \); and \( B \) is the soil moisture capacity shape parameter. The surface flow, \( Q_s \), is calculated as

\[
Q_s = \begin{cases} 
PE - (W_m - W_0), & PE + W \geq W_{mm} \\
PE - (W_m - W_0) + W_m \left(1 - \frac{PE + W}{W_{mm}}\right)^{1+B}, & PE + W < W_{mm}
\end{cases}
\]

(2)

where \( PE \) is effective precipitation, which equals precipitation minus evapotranspiration; \( W_m \) is soil moisture capacity of the upper two soil layers; and \( W_0 \) is initial soil moisture.

The slow response runoff, or baseflow, is only generated from the third layer. Using the nonlinear ARNO model, baseflow, \( Q_b \), is modeled as

\[
Q_b = \begin{cases} 
\frac{D_s D_m \theta_3}{W_s \theta_{3,S}}, & 0 \leq \theta_3 \leq W_s \theta_{3,S} \\
D_s \frac{D_m}{W_s} \theta_3 + \left( D_m - \frac{D_s D_m}{W_s} \right) \left( \frac{\theta_3 - W_3 \theta_{3,S}}{W_s - W_3 \theta_{3,S}} \right)^2, & \theta_3 > W_s \theta_{3,S}
\end{cases}
\]

(3)

where \( D_m \) is the maximum velocity of baseflow, \( D_s \) and \( W_s \) are the fraction of \( D_m \) and maximum soil moisture content of the third soil layer (\( \theta_{3,S} \)), respectively; and \( \theta_3 \) is the current soil moisture of the third layer. The baseflow recession curve is linear below a threshold (\( W_s \theta_{3,S} \)) and nonlinear above that threshold.

The relationship between the unsaturated hydraulic conductivity and the soil moisture content is modeled using a power law following Brooks and Corey (1964), while drainage between the three layers is represented as a gravity-driven process (Liang et al., 1994):
where $Q_{i+1}$ is the vertical drainage between soil layers $i$ and $i+1$, $W_i$ is the soil moisture content of layer $i$, $\text{Exp}_i$ is a function of the pore size distribution, $K_i$ is the saturated hydraulic conductivity for layer $i$, and $\theta_{r,i}$ is the residual moisture content of layer $i$. The maximum soil moisture content of each layer is described by the following equation (Demaria et al., 2007):

$$W_i^{\text{max}} = D_i \times \phi_i$$

where $D_i$ is the thickness of layer $i$, and $\phi_i$ is the porosity of each layer, which is considered to be the same for the three soil layers.

The 13 tunable streamflow-related parameters in the above equations were selected for study to determine which ones were sensitive parameters in the model (Table 1). Even though the VIC model has 46 or more tunable parameters (Bennett et al., 2018), we chose only 13 parameters for this study, because the values of these 13 parameters are typically subject to calibration rather than direct measurement. The parameters include the variable soil moisture capacity curve parameter, $B$ (equations (1) and (2)); the three baseflow parameters, $Ds$, $Ws$, and $Dm$ (equation (3)); the saturated hydraulic conductivities of the three layers, $K_1$, $K_2$, and $K_3$ (equation (4)); the exponents of the Brooks-Corey relationship, $\text{Exp}_1$, $\text{Exp}_2$, and $\text{Exp}_3$ (hereinafter $E_1$, $E_2$, and $E_3$; equation (4)); and the thicknesses of the three soil layers, $D_1$, $D_2$, and $D_3$ (equation (5)). The parameters selected for the SA also include the seven recommended default parameters—$B$, $D_1$, $D_2$, $D_3$, $Ds$, $Ws$, and $Dm$—which are the ones suggested by the VIC model developers as being the most sensitive parameters for most climatic, edaphic, and physiographic watershed settings.

### 2.2. Data Sets

In this study, three kinds of input data were used to drive the VIC model and the routing model for streamflow generation: (1) meteorological forcing data; (2) vegetation and soil data; and (3) topographical data. For the meteorological forcing, we used observed daily precipitation data covering the period 1960–1979, obtained from the National Meteorological Information Center of the China Meteorological Administration. This precipitation data, with a spatial resolution of 0.25° × 0.25°, is constructed from observations of 2,419 gauge stations (Figure S1 in the supporting information) across China (Shen et al., 2010; Sun et al., 2017). The maximum and minimum daily temperatures from 2,481 gauge stations across China were acquired from the China Meteorological Data Sharing Service System (http://cdc.nmic.cn/home.do). The temperature data also cover the period from 1960 to 1979, and they were interpolated to 0.25° × 0.25° using the inverse distance weighting method (Sun & Miao, 2018). The daily wind speed from 756 gauge stations across China ranged from 0.3 to 21.5 m s⁻¹ and was used in the VIC model. The wind speed data were also interpolated using the inverse distance weighting method (Sun & Miao, 2018).

The daily wind speed from 756 gauge stations across China ranged from 0.3 to 21.5 m s⁻¹ and was used in the VIC model. The wind speed data were also interpolated using the inverse distance weighting method (Sun & Miao, 2018).

The global soil and vegetation data used to run the VIC model are the same as those used by Zhang et al. (2014). Soil texture is based on the global 5-arcmin data set from the Food and Agriculture Organization of the United Nations (FAO, 1998). The 1 km land cover data are from the University of Maryland (http://glcfapp.glcf.umd.edu:8080/esdi/index.jsp). The leaf area index (LAI) is based on the gridded (0.25° × 0.25°) monthly global LAI database of Myneni et al. (1997). In this study, the monthly LAI is specified for each grid cell and for each vegetation class, with no interannual variation included (Zhang et al., 2014). The 1 km digital elevation model data set used to extract the channel network was obtained from the Cold and Arid Regions Sciences Data Center at Lanzhou (http://westdc.westgis.ac.cn).

This study focuses on parameter calibration improvement in 10 major river basins of Mainland China (hereinafter “China”, Figure 1) which together cover 63.7% of China’s area and feed 97.7% of China’s population. For the large-scale run, a total of 15,775 grid cells with a 0.25° horizontal grid mesh were used. The 30 hydrological gauge stations located in these 10 basins (obtained from the hydrological yearbook of China and local water resources departments) with monthly time series for naturalized streamflow (hereinafter “streamflow”) from 1960 to 1979 were used to calibrate and validate the VIC hydrological model. The details of...
the gauge stations are shown in Figure 1 and Table S1. The period 1960–1979 was chosen for this study because streamflow in China since the 1980s has been greatly altered by human activities such as dam construction, urbanization, and land use changes (Wang et al., 2012). Note that we performed model calibration and validation for Huangjiagang station using data only from 1960 to 1967 due to the operation of the Danjiangkou Reservoir for water storage starting in 1968 (Xu, 1996). The construction of the Danjiangkou Reservoir is likely to have interfered with the natural hydrological systems at Huangjiagang station (Chen, 1985). The naturalized streamflow without water management effects (e.g., irrigation and reservoir regulation) was reconstructed by the Bureau of Hydrology of the Chinese Ministry of Water Resources based on the water balance principle. The details of the natural streamflow reconstruction methods are given in the supporting information. The observed streamflows at two stations, Changdu and Datong, were used in this study, because these regions had been less affected by human activities before 1980, and so the observed streamflow may be seen as natural.

| Parameter | Brief description | Unit | Range | Note |
|-----------|------------------|------|-------|------|
| $B$       | The shape of the variable infiltration capacity curve controlling surface runoff | N/A  | 0.001–0.4 | Default parameter |
| $D_1$     | Thickness of soil layer 1 (uppermost) | m   | 0.01–0.5 | Default parameter |
| $D_2$     | Thickness of soil layer 2 | m   | 0.05–1.0 | Default parameter |
| $D_3$     | Thickness of soil layer 3 (lowermost) | m   | 0.5–2.3 | Default parameter |
| $D_s$     | The fraction of maximum velocity of baseflow | N/A | 0.001–1 | Default parameter |
| $D_m$     | The maximum velocity of baseflow | mm/day | 5–20 | Default parameter |
| $W_s$     | The fraction of maximum soil moisture content of third soil layer | N/A | 0.1–1 | Default parameter |
| $E_1$     | Exponent of the Brooks-Corey drainage equation in soil layer 1 | N/A | 8–30 | — |
| $E_2$     | Exponent of the Brooks-Corey drainage equation in soil layer 2 | N/A | 8–30 | — |
| $E_3$     | Exponent of the Brooks-Corey drainage equation in soil layer 3 | N/A | 8–30 | — |
| $K_1$     | Saturated hydraulic conductivity in soil layer 1 | mm/day | 163–4,765 | — |
| $K_2$     | Saturated hydraulic conductivity in soil layer 2 | mm/day | 163–4,765 | — |
| $K_3$     | Saturated hydraulic conductivity in soil layer 3 | mm/day | 163–4,765 | — |

"Default parameter" in the Note column indicates that the parameter is among those suggested by the VIC model developers as being the most sensitive parameters. "Source: Shi et al. (2008)." "Source: Demaria et al. (2007)." "Source: Bennett et al. (2018)."

Figure 1. The 10 major river basins in China that were studied and associated hydrological stations. I, Songhua River; II, Liao River; III, Hai River; IV, Yellow River; V, Hei River; VI, Yangtze River; VII, Huai River; VIII, Southeast River; IX, Pearl River; X, Southwest River drainage.
3. Methods

3.1. SA

SA is a statistical analysis tool to ascertain how a given model (numerical or otherwise) depends on its input factors and then screen out the most important factors in the model (Di et al., 2017; Saltelli et al., 1999). SA has been widely used to reduce the parameter dimensionality of function optimization in the field of hydrological model calibration (Muleta & Nicklow, 2005; van Griensven et al., 2006). Qualitative methods provide a heuristic score to intuitively represent the relative sensitivity of parameters, while quantitative methods tell how sensitive the parameter is by computing the impact of the parameter on the total variance of model output (Gan et al., 2014). The results of quantitative methods are more reliable and can be used to validate the results of qualitative methods (Di et al., 2017).

In this study, we applied three qualitative SA methods and one quantitative SA method to do parameter screening for optimization in each river basin, since using multiple SA methods is expected to lead to more robust conclusions (Li et al., 2013; Neumann, 2012). The SA methods are briefly described below.

3.1.1. SOT

The Sum-of-Trees (SOT) model is a qualitative SA method that is fundamentally an additive model with multivariate components and which belongs to the class of random tree-based methods (Chipman et al., 2010). The SOT model can be more explicitly expressed as (Chipman et al., 2010)

$$Y = \sum_{j=1}^{m} g(x; T_j; M_j) + \epsilon, \quad \mathcal{E} \sim N(0, \sigma^2)$$  \hspace{1cm} (6)

where for each binary regression tree $T_j$ and its associated terminal node parameters $M_j$, $g(x; T_j; M_j)$ is the function that assigns $u_{ij} \in M_j$ to $x$. When the number of trees $m > 1$, each $u_{ij}$ here is merely a part of $E(Y | x)$. Furthermore, each such $u_{ij}$ will represent a main effect when $g(x; T_j; M_j)$ depends on only one component of $x$ (i.e., a single variable) and will represent an interaction effect when $g(x; T_j; M_j)$ depends on more than one component of $x$ (i.e., more than one variable). In this study we explore the main effects only. The total number of splits (with scaling at each level) for each input variable in the model stands for the importance of this variable; that is, the variable with the most splits in the model is considered to be the most important one (Gan et al., 2014; Li et al., 2013).

3.1.2. MARS

The Multivariate Adaptive Regression Splines (MARS) technique is another qualitative SA method we adopted for determining which model input parameters contribute most substantially to a model output—in our case, streamflow. MARS is an extension of linear models that make use of linear regression, the mathematical construction of splines, and binary recursive partitioning (Friedman, 1991; Gan et al., 2014). The MARS model can be more explicitly expressed in the form

$$\hat{f}(x) = a_0 + \sum_{k=1}^{m} f_i(x_i) + \sum_{k=2}^{m} f_{ij}(x_i, x_j) + \cdots$$  \hspace{1cm} (7)

where $a_0$ is the coefficient of the constant basis function and the first sum is over all basis functions that involve only a single parameter, that is, the main effects (Friedman, 1991). The second sum is over all basis functions that involve exactly two variables, representing (if present) two-variable interactions.

3.1.3. DT

The Delta test (DT), a qualitative SA method, was devised by Pi and Peterson (1994) for identifying parameter dependencies in continuous functions. It was first applied to parameter screening by Eirola et al. (2008). The DT model is a method based on nearest neighbors for estimating the variance of the residuals (Gan et al., 2014). The nearest neighbors of a point are defined as the (unique) point that minimizes a distance metric to that point in the input space:

$$N(i) = \arg \min \|x_i - x_l\|^2$$  \hspace{1cm} (8)

where $x_i$ is the input point for the subset $S \subseteq \{x_1, \cdots, x_m\}$. Hence, the DT criterion of a variable can be computed as
where \( y_{N(i)} \) is the function value corresponding to \( N(i) \); \( y_i \) is the function value corresponding to \( x_i \); and \( M \) is the sample size. The variable subset \( S \) with the smallest DT criterion corresponds to the most important subset of variables, that is, the most sensitive parameters (Li et al., 2013).

### 3.1.4. Metamodel-Based Sobol’

The Sobol’ method is a reliable quantitative SA algorithm, as it computes the precise contribution ratio of each parameter to the total variance of model output (Li et al., 2013). We adopted the more efficient metamodel-based Sobol’ method owing to the time-consuming nature of the original Sobol’ method, which needs tens of thousands of model runs to obtain accurate approximation of the variances (Wang et al., 2016).

We built a metamodel with a support vector machine using initial input samples, then ran the Sobol’ analysis on the metamodel. For more details on the support vector machine, refer to Chang and Lin (2011). The Sobol’ method asserts that an integrable function can be decomposed into summands of different dimensions (Sobol, 1990). The Sobol’ method decomposes the variance of the output as (Li et al., 2013; Sobol, 1990)

\[
V = \sum_{i=1}^{n} V_i + \sum_{1 \leq i < j \leq M} V_{ij} + \cdots + V_{1,2,\ldots,n}
\]

where \( n \) denotes the total number of variables, \( V_i \) represents the part of the variance that can be explained by the \( i \)th variable only, \( V_{ij} \) represents the parts of the variance that can be explained by the interaction of the \( i \)th and \( j \)th variables, and \( V_{1,2,\ldots,n} \) represents the parts of the variance that can be explained by the interaction of all the variables. The Sobol’ global sensitivity index is defined as

\[
S_{i_1,\ldots,i_s} = \frac{V_{i_1,\ldots,i_s}}{V}
\]

where \( V_{i_1,\ldots,i_s} \) denotes the variance corresponding to \( (i_1, \ldots, i_s) \), and the integer \( s \) is called the order or the dimension of the index. All the \( S_{i_1,\ldots,i_s} \) are nonnegative, and their sum is equal to 1. The importance of the variables can be seen from this index.

### 3.2. ASMO

An adaptive surrogate modeling-based optimization (ASMO) algorithm resulting from previous research from our team members (Wang et al., 2014) was used in this study to calibrate the catchment-specific sensitive parameters of the VIC model. ASMO has been successfully applied in several complex model calibration studies (Duan et al., 2017; Gong et al., 2016; Li et al., 2018). The main procedure for the ASMO algorithm, which is composed of four steps, is briefly described as follows (Wang et al., 2014).

1. **Initial sampling.** This step uses a specific sampling method to create random parameter samples that are evenly distributed within the physical variability ranges of the parameters. As suggested by Wang et al. (2014), low-discrepancy quasi-Monte Carlo methods are the most suitable initial sampling designs for ASMO, and a sample of 15–20 times the number of parameters under consideration may be the proper initial sample size. Hence, in this study, the Sobol’ sequence (Sobol, 1967), one of the quasi-Monte Carlo sampling methods, was used in initial sampling and the initial sample size was set equal to 20 times the number of the catchment-specific sensitive parameters being evaluated. The specific sensitive parameters for each catchment come from the results of sensitivity analyses by the DT, SOT, MARS, and Sobol’ methods addressed in section 3.1.

2. **Building a surrogate model.** This step constructs an error response surface by using the catchment-specific parameter samples from the previous step. GPs, which can easily integrate reinforcement learning, have been demonstrated to outperform other surrogate methods for ASMO (Gong et al., 2016; Wang et al., 2014). In this study, we adopted a GP method to build a surrogate model among the initial sample points.

3. **Optimizing the surrogate model.** This step runs an objective optimization algorithm on the GP surrogate models built in the previous step, and the optimal solutions are obtained. The goal of optimization is to find the minimum of an error response surface in the multiparameter space. A global optimization algorithm, shuffled complex evolution (Duan et al., 1992), was used in this study.
4. Adaptive sampling. The fourth step is to refine the GP response surface iteratively by using an adaptive sampling strategy that places more parameter samples in the promising parameter space based on information already gained on the existing response surface. Once this iterative process converges, the final response surface is updated as the new surrogate model.

Steps (3) and (4) are repeated until the convergence criteria for parameter optimization of the real physical model are met. The optimal solution of this surrogate model should approximate the optimal solution of the real model. In this study, the convergence criteria were defined as either the objective function value of the VIC simulation remaining unchanged after a number of searches equal to 20 times the dimensionality of the parameters, or the number of searches reaching the maximum number of model runs allowed, $N_{\text{max}}$ (in this study $N_{\text{max}} = 500$, excluding initial samples).

3.3. Experimental Design

In this study, we focus mainly on the evaluation of the SA-based ASMO calibration framework for streamflow simulations over China. The VIC model is used to simulate the streamflow; four SA methods are used to screen parameters of the physical model; and an ASMO algorithm is used to autocalibrate specified model parameters for each river basin. Three stages of experiments were designed to progressively address the questions outlined in section 1.

Stage 1: Sensitivity analysis (SA). Sensitivities for 13 streamflow-related parameters in VIC (Table 1) were calculated using four SA methods (section 2.2) in each of the 14 selected catchments, each associated with a station, which are distributed across the 10 major river basins (Figure 1 and Table S1). The 14 stations were chosen for this SA because they cover relatively small subdrainage areas within their respective basins, which reduced computational costs while ensuring equal spatial distribution of the stations for full expression of the hydroclimate characteristics across China. Note that we performed a multisection SA for the basins that span multiple climate zones, such as the Yellow River basin, the Yangtze River basin, and the Southwest River drainage (Figure 1 and Table S1). We ran 6,000 training simulations for the period from 1960 to 1979 (from 1960 to 1967 for Huangjiagang station) based on samples of parameter combinations obtained from the Sobol’ sequence for each of the 14 selected catchments. We define the performance of streamflow simulation (the output variable of interest), by calculating the Nash-Sutcliffe model efficiency coefficient (NSE) of monthly streamflow from 1961 to 1979 (where the first year of each simulation served as the spin-up period). We use the SA results of the selected catchment in each basin to represent the parameter sensitivity of the whole basin for streamflow, because runoff regimes within a basin are assumed to be similar between catchments, following the principle of parameter regionalization over continental China (Xie et al., 2007).

Stage 2: Evaluation of SA-based ASMO calibration framework. To demonstrate the effectiveness and efficiency of our calibration framework, two case study basins with two kinds of parameter error (type I and type II) were selected to compare the optimal solutions and corresponding search times between an optimization based on catchment-specific parameters and optimization based on the default parameters. For each case basin, two optimization runs from 1960 to 1979 were conducted to maximize the NSE of the streamflow simulation optimization by using different parameter sets: catchment-specific sensitive parameters and the seven default parameters. We employed the ASMO method for the parameter optimization. We set the initial sample sizes equal to 20 times the number of parameters being evaluated.

Stage 3: Optimization over China. In this stage, we tuned the catchment-specific sensitive parameters in the VIC model to match the observations at 30 hydrological stations during the calibration period (1961–1969). Then the tuned parameters were validated for the validation period (1970–1979). For the same reason addressed in section 2.2, we chose periods before the 1980s to minimize the effects of human activities on calibration and validation of the VIC model. For Huangjiagang station in particular, we set the calibration and the validation periods to before 1968 to avoid potential human interference with the hydrology due to the operation of the Dangjiangkou Reservoir; 1961–1964 and 1965–1968 were the calibration and validation periods, respectively. The model parameters for uncalibrated catchments were set to be identical to those of a neighboring calibrated catchment. In this study, neighbors are defined by Thiessen polygons.
4. Results and Discussion

4.1. Sensitivity Analysis

Using the Stage 1 experimental design described above, we ran the VIC model using 6,000 randomly generated parameter sets based on the Sobol’ sequence, for each of the 14 selected catchments of the 10 major river basins. Then we computed the sensitivity scores of 13 streamflow-related parameters for all catchments, using both qualitative and quantitative SA methods. We found that the three qualitative SA methods—DT, SOT, and MARS—produced consistent sensitivity score results for all 13 tunable parameters for theNSE of monthly streamflow, which is visible as well-matched distributions of lines in parallel coordinates plots for most catchments (Figure 2). Thus, for each panel in Figure 2, the parameters with mean scores above 0.5 for all three qualitative SA methods are identified as sensitive parameters for our streamflow simulation. The quantitative Sobol’ method further confirms that the qualitative SA results are reasonable (Figure 3 and Table 2). For the Sobol’ sensitivity index, we set the threshold for ruling out insensitive parameters to 0.1, as recommended by Hou et al. (2015) and Tang et al. (2007). The sensitive parameters identified by the Sobol’ method are consistent with the qualitative SA results for 64% of catchments. We adopted an integrated strategy to synthesize the SA results from the qualitative and quantitative SA methods described above to ensure the reliability of the parameter screening results (Li et al., 2013; Neumann, 2012). As presented in Table 2, for each of the 14 catchments, a parameter is considered sensitive for the streamflow simulation if it has been identified by either the Sobol’ method or the qualitative SA methods.

The parameters identified as sensitive for modeling streamflow varied spatially across 14 catchments across China and are poorly matched to the seven default parameters (Table 2 and Figure S2). Compared with the seven default parameters ($B$, $D_1$, $D_2$, $D_3$, $D_4$, $D_5$, and $D_6$), the number of parameters our screening identified as important was greatly reduced except for that in the upper sections of the Southwest River drainage (Table 2). Therefore, using the seven default parameters without SA screening within the calibration procedure results in including unnecessary parameters (type I error). For example, only three parameters are sensitive in the Liao River basin and in the Songhua River basin only four are (Table 2), and therefore there is no need to calibrate all seven recommended parameters. This result supports the finding by Demaria et al. (2007) and Bao et al. (2011) that only a small number of parameters is highly sensitive in the VIC model. In addition, indiscriminately calibrating only the default sensitive parameters also can lead us to miss some other important parameters (type II error). Even though most of the sensitive parameters can be found on the list of seven default parameters, in some cases other sensitive parameters are not among those seven default parameters. For instance, streamflow simulations of the Songhua River, the Hai River, the Middle Yellow River, and the Huai River basin are sensitive to the second soil drainage parameter ($E_2$), which is not among the seven default parameters (Table 2).

Next, we examined the SA results and found that three parameters—the infiltration parameter ($B$) and two of the soil depth parameters ($D_1$ and $D_2$)—are highly sensitive in most basins, while other parameter sensitivities strongly related to the dynamic environment surrounding the basin, including the climate, vegetation, soil, and topography characteristics. As shown in Table 2 and Figure S2c, streamflow simulation was highly sensitive to the $B$ parameter in all basins except the Huai River basin. The $B$ parameter controls the shape of the variable capacity curve and effectively dictates the partitioning of rainfall into infiltration and surface runoff (Demaria et al., 2007; Liang et al., 1994). A higher value of $B$ gives lower infiltration and yields higher surface runoff (Shi et al., 2008; Xie et al., 2007). The identification of parameter $B$ as sensitive in this study indicates that this parameter is playing a key role in the generation of direct runoff, and this finding is consistent with a previous study by Atkinson et al. (2002).

The parameters for thickness of the upper two soil layers, $D_1$ and $D_2$, show high sensitivity to streamflow in all catchments (Table 2 and Figures S2d and S2e). In general, for runoff generation, thicker soil layers slow down seasonal peak flows and increase water loss due to evapotranspiration (Xie et al., 2007). Vegetation root allocation in our model was designed as a two-zone distribution; and most of the vegetation is located in the upper two soil layers. Therefore, changes in the $D_1$ and $D_2$ parameters affect the soil evapotranspiration and the soil moisture that percolates to deeper layers, which further influences the surface and subsurface flows. The $D_1$ parameter is the thickness of the third soil layer, which contributes to the baseflow (Bao et al., 2011). As shown in Table 2 and Figure S2f, streamflow simulations in the Upper Yellow River basin, the Middle Yangtze River basin, and the Upper and Middle Southwest River drainage were sensitive to the
Figure 2. The qualitative parameter sensitivity results from the DT, SOT, and MARS methods in (a–n) 14 catchments of 10 major river basins across China. The sensitivity scores are normalized to [0, 1]; 1 means most sensitive and 0 means least sensitive. The lines in gray show the range of parameter sensitivity across all catchments.

Figure 3. Parameter sensitivity rankings of Sobol' total effect analysis methods in 14 catchments of 10 major river basins across China. The gray bar with gradient denotes sensitivity index values between [0, 1]; 1 means most sensitive and 0 means least sensitive.
Table 2

| Basin                  | Section | Station    | Qualitative SA | Sobol’ | Final parameter used |
|------------------------|---------|------------|----------------|--------|----------------------|
| Songhua River          | —       | Guchengzi  | B, D1, D2, E2  | B, D1, D2 | B, D1, D2, E2        |
| Liao River             | —       | Tonghua    | B, D1, D2     | B, D1, D2 | B, D1, D2, E2        |
| Hai River              | —       | Guantai    | D1, D2, Ds, E2| B, D1, Ws | B, D1, D2, Ds, Ws    |
| Yellow River           | Upper   | Tangnaihai | D1, D2, Ds, Ds| B, D1, D2, Ds, Ws, Ds, Ws | B, D1, D2, Ds, Ws    |
|                        | Middle  | Hekou      | B, D1, D2, E2  | B, D1, D2 | B, D1, D2, E2        |
| Hei River              | —       | Yingluoxia  | D1, D2, Ds, Ws| B, D1, Ws | B, D1, D2, Ds, Ws    |
| Yangtze River          | Middle  | Chishui    | B, D1, D2, Ds, Ws | B, D1, D2, Ds, Ds, Ws | B, D1, D2, Ds, Ws    |
|                        | Lower   | Huangqiagang| B, D1, D2, Ds | B, D1, D2 | B, D1, D2, Ds        |
| Huai River             | —       | Zhongdu    | D1, D2, Ds    | B, D1, Ws | B, D1, D2, Ds, Ws    |
| Southeast River        | —       | Tunxi      | D1, D2, Ds, Ws| B, D1, D2, Ds, Ws | B, D1, D2, Ds, Ws    |
| Pearl River            | —       | Pingla     | D1, D2, Ds, Ws| B, D1, D2, Ds, Ws | B, D1, D2, Ds, Ws    |
| Southwest River drainage| Upper  | Lasa       | B, D1, D2, Ds, Ws, Dm | B, D1 | B, D1, D2, Ds, Ds, Ws, Dm |
|                        | Middle  | Changdu    | B, D1, D2, Ds, Ws | B, D1, D2 | B, D1, D2, Ds, Ws    |
|                        | Lower   | Yale       | B, D1, D2, Ds  | B, D1, D2, Ds, Ws | B, D1, D2, Ds, Ws    |

$D_3$ parameter. A possible reason for the identification of the $D_3$ parameter in those river basins is that the catchments used for the SA (Tangnaihai, Lasa, and Changdu) are all located in a dry environment where plant available water (the difference between soil moisture content at field capacity and at wilting point) is extremely low (areas of low temperature and small annual precipitation shown in Figures S2a and S2b). At the same time, baseflow in these basins is relatively large (Huang et al., 2019), because the amount of soil moisture that can be used effectively by the vegetation for evapotranspiration is small. Therefore, the $D_3$ parameter that controls the baseflow generation is particularly important in water-stressed environments.

The sensitivities of the three baseflow parameters, $D_s$, $D_m$, and $W_s$, differ between the 14 catchments (Table 2). The streamflow is sensitive to the $D_s$ and $W_s$ parameters in most basins except the four northern basins: Songhua, Liao, Huai, and the Middle Yellow River (Table 2 and Figures S2g and S2h). $D_s$ is the fraction of maximum baseflow velocity, and $W_s$ is the fraction of maximum soil moisture content of the third layer, where nonlinear baseflow occurs. These two baseflow parameters determine how quickly the water stored in the third layer is evacuated as baseflow (Liang et al., 1994), and they control the nonlinear part of the baseflow generation function (equation (3)). The probable reason for not identifying the $D_s$ and $W_s$ as sensitive parameters in the four northern basins listed above is that those basins are located within semiarid or semiwet hydroclimatic zones (Figures S2a and S2b). The precipitation is highly concentrated in the summer, and the soil layers are deep in those four basins, especially in the Middle Yellow River section. The third soil layer in these basins usually does not reach saturation; thus, the streamflow simulation in those areas is insensitive to the $D_s$ and $W_s$ values because the nonlinear part of the baseflow generation function only activates when the moisture storage in the third layer is above a threshold ($W_sD_3S$, equation (3)). By contrast, streamflow simulation was sensitive to the $D_m$ parameter only in one basin: the Southwest River drainage in the southern Tibet region (Table 2 and Figure S2i). $D_m$ is daily maximum baseflow velocity, which can be estimated by the horizontal pressure gradient that increases with altitude (Bao et al., 2011). Therefore, this parameter might be playing a key role in the generation of baseflow for higher-altitude regions.

For the three drainage $E$ parameters, $E_1$, $E_2$, and $E_3$, only $E_2$ shows high sensitivity in streamflow simulation (Table 2). The $E_2$ parameter is an exponent of the Brooks-Corey relationship, and it controls the hydraulic conductivity between the second and third soil layers and, therefore, affects baseflow generation. $E_2$ was identified as a sensitive parameter for streamflow in four basins: Songhua, Hai, Huai, and the Middle Yellow River basin (Table 2 and Figure S2j). A possible reason for the importance of the $E_2$ parameter in these cases relates to the semiarid hydroclimatic regimes in those basins (Figures S2a and S2b). In humid areas, drainage water will always penetrate to the second soil layer due to the abundant precipitation; while by contrast, in hyperarid areas, drainage water usually cannot reach the second layer. Therefore, streamflow
simulations are not very sensitive to the $E_2$ parameter in either humid or hyperarid areas. In semiarid areas, however, such as the four basins mentioned above, drainage water simulation is mainly controlled by the value setting of $E_2$, and thus streamflow simulation is more sensitive to $E_2$ for these basins.

Contrary to our findings for the drainage parameter $E_2$, we found that streamflow was only slightly sensitive to saturated hydraulic conductivity parameters in all soil layers within the VIC model ($K_1$, $K_2$, and $K_3$), and these parameters are not among those identified as sensitive (Table 2). The $K$ parameter is the hydraulic conductivity when the soil moisture reaches saturation (Yang et al., 2014). The finding that streamflow is not sensitive to the $K$ parameter is consistent with previous studies (Bennett et al., 2018; Demaria et al., 2007).

### 4.2. Evaluation of SA-Based ASMO Calibration Framework

According to SA results discussed above, the parameters we identified as sensitive parameters for modeling streamflow in the 10 major river basins did not match the seven default parameters ($B, D_1, D_2, D_3, Dm, Ws, Ds$). Rather than using these default parameters in model calibration (Ran et al., 2017; Shi et al., 2008; Zhang et al., 2014), we can eliminate calibration of the unnecessary parameters (those resulting from type I error) and instead add the sensitive parameters missing from the default calibration procedure (type II error). So, how does the SA-based ASMO calibration framework avoid those two errors? To address this question, we conducted optimization comparison experiments (Stage 2 in section 3.3) to explore the effectiveness and efficiency of the SA-based ASMO calibration framework for VIC streamflow simulation in two case studies: the Liao River and the Huai River.

Figure 4 shows the convergence process for two optimization procedures in the Tonghua station of the Liao River basin: a SA-based optimization (in which only three important parameters were calibrated: $B, D_1$, and $D_2$) and an optimization with calibration of all seven default parameters. According to Figure 4, SA-based optimization has a clear advantage over the optimization with all seven default parameters in terms of search efficiency. For the SA-based optimization runs, the optimization process converges to the optimal function value (represented as NSE) of 0.9023 after 153 model runs (including 60 initial samples and 93 additional adaptive samples). By contrast, the optimization with all seven default parameters requires up to 450 model runs to reach the function value of 0.9048, which is approximately 3 times the number of runs required for the SA-based optimization (Figure 4). A possible reason is that the smaller number of sensitive parameters has a smaller searching space for optimal function value, which allows the procedure to focus on searching the region that is most likely to contain the optimum and thus be more efficient. Moreover, the difference between the optimal function values found by the two optimization procedures is negligible ($\Delta \text{NSE} = -0.0025$, where $\text{NSE}$ equals the optimal function value found by SA-based optimization minus the optimal function value found by optimization of the seven default parameters). Therefore, using a smaller number of sensitive parameters to calibrate the model also finds the ideal optimal value and does it with greatly improved efficiency. The Liao River case study demonstrated that the likelihood of overoptimization in the VIC model can be reduced by using the SA-based optimization framework.

Figure 5 shows the effect of including an additional sensitive parameter, the second soil drainage parameter ($E_2$), on the VIC model calibration for the Zhongdu station of the Huai River basin. What is most notable in this figure is that the optimal function value found by SA-based optimization (which included $E_2$) is larger than the value found by using the default seven-parameter optimization ($\Delta \text{NSE} = 0.0032$), even though the SA-based optimization runs optimized for only three parameters (Figure 5a). In addition, the efficiency of the model calibration was greatly improved. For the SA-based optimization runs, the optimization process converges to the optimal function value of 0.8440 after a total sample of 145 runs, while the default optimization process requires up to 300 model runs to reach the function value of 0.8408 (Figure 5a). Therefore, using the SA method to screen for important parameters before model calibration can save at least half the time and also improve the optimal solution, as we have demonstrated at least for the Huai River basin case.
To further explore the influence of the \( E_2 \) parameter on model results, we compared the model calibration results of the default seven-parameter optimization with and without \( E_2 \). The inclusion of the \( E_2 \) parameter in the optimization improves the NSE of the streamflow simulations by approximately 0.0314 compared to optimization of the seven default parameters alone (Figure 5b). This means that some additional parameters, like \( E_2 \) in the Huai River basin, may have a significant influence on streamflow simulation, and those parameters should not automatically be viewed as unimportant.

In summary, the SA-based ASMO calibration framework performed well to reduce the errors and computational costs that would result from use of the default seven-parameter optimization. Sensitivity analysis not only eliminates unnecessary parameters from the calibration space but also adds important parameters that are not included among the seven default parameters, further improving the efficiency and effectiveness of VIC streamflow calibration.

### 4.3. SA-Based ASMO Calibration Over China

#### 4.3.1. Calibration

Given the successful performance of our SA-based ASMO calibration framework, we next applied this framework to the VIC streamflow calibration over all of China. The comparison between the naturalized streamflow and modeled monthly streamflow for the calibration period across China is shown in Figures 6 and S3. The skill scores (expressed as NSE and \( R^2 \)) for 30 hydrological stations over the calibration period show that the SA-based ASMO calibration framework performed reasonably well in matching observations over China (Figure 6). The correlation coefficient \( (R^2) \) between the naturalized streamflow and simulated monthly streamflows under calibration mode varies between 0.72 and 0.97 at the \( p = 0.001 \) level, with an average \( R^2 \) of 0.90. The NSE for monthly streamflow ranges from 0.75 to 0.97, with an average of 0.86 under calibration mode at the 30 hydrological stations. About 80% of hydrological stations have an NSE greater than 0.80. The timing and amount of simulated streamflow peaks and valleys at different stations for all 10 basins closely matches the naturalized streamflow (Figures 6 and S3), which means that the VIC model is able to capture both magnitude and phase of the monthly streamflow signal in all hydroclimatic zones. Specifically, the model performs particularly well for the basins in areas with a wet hydroclimatic regime, for example, Wuzhou station in the Pearl River basin and Weiren station in the Southeast Rivers basins (Figure 6). The low flows were largely underestimated in semiarid basins, such as Tangnaihai and Huayuankou station in the Yellow River basin, leading to lower NSE values (0.76 and 0.75, respectively, Figure S3). The high flows were slightly overestimated for arid basins in the Hei River basin. Even if the model does not perform well in arid and semiarid areas, the results overall show a well-simulated output for the SA-based ASMO framework for most hydrological stations. These stations across all hydroclimatic zones demonstrate that the present calibration framework, the SA-based ASMO optimization adopted in this study, is robust across a wide range of hydroclimatic regimes with different climates and geophysical features across China.

#### 4.3.2. Validation

The performance of the SA-based ASMO optimization for the validation period is shown in Figures 7 and S4. The figures show well-simulated output overall for the SA-based ASMO framework for most hydrological stations. The skill score scatterplot shows high values of both NSE and \( R^2 \) for all 30 hydrological stations, from 0.71 to 0.97 and from 0.78 to 0.97, respectively (Figures 7 and S4). The streamflow simulation in basins of all hydroclimatic regimes, from arid (e.g., Yingluoxia), to semiarid (e.g., Lanzhou), to moderate (e.g., Changdu), and to wet (e.g., Liuzhou), closely matches the naturalized streamflow (Figures 7 and S4). The NSE and \( R^2 \) values averaged over all 30 hydrological stations are 0.88 and 0.83, respectively, which demonstrates that our parameter optimization results performed reasonably well in representing the multiple flow regimes across China. About 83% of the hydrological stations have an NSE greater than 0.75, and about 53% have an NSE greater than 0.80. The NSE and \( R^2 \) values were consistent with the calibration period at most
stations, however, large discrepancies appear in the Hei, Huai, and Southwest River basins, where the NSE value decreased by 0.08–0.19 during the validation period (Figures 7 and S4). Note that the Hei and Huai River basins are located in arid and semiarid hydroclimatic regions; the variability of arid zone precipitation and infiltration losses, as well as the general paucity of data on forcing, flow, and soil properties, all add to the difficulty of modeling in those areas (Al-Qurashi et al., 2008; Grayson et al., 1992). As for the Southwest River basin, it originates in the highest part of the Tibetan Plateau, which has the most varied topography in the world and which has a great amount of global-warming-induced glacier melting augmenting the surface-runoff rate (Immerzeel et al., 2010). Despite the poor performance in arid and high-altitude basins, the current framework is still a computationally low-cost and effective method for parameter optimization as compared to manual calibration method for the seven default parameters applied in previous studies in China (e.g., Ran et al., 2017; Wang et al., 2012; Xie et al., 2007).

Figure 6. The ASMO optimization results for streamflow across 10 major river basins in China during the calibration period. Streamflow was simulated by tuning catchment-specific sensitive parameters. The scatterplot at the top illustrates the skill scores (expressed as NSE and $R^2$) of VIC-simulated monthly streamflows for 30 hydrological stations during the calibration period (the station numbers are labeled as in Table S1, and the blue and orange circles indicate the $R^2$ and NSE scores, respectively). The surrounding 10 panels illustrate the time series of monthly streamflows ($10^2$ m$^3$/s) for the best performing hydrological stations in each of the 10 basins during the calibration period 1961–1969 (blue and black lines indicate the simulated and natural streamflows, respectively).
The proposed framework highly improves the efficiency and accuracy for streamflow calibration, and the calibrated parameters provide a leading reference for representing the natural hydrological regime. The calibrated parameters, however, still have certain uncertainties. For example, only 13 streamflow-related soil parameters were involved in sensitivity analysis and optimization, and the effects of the vegetation parameters (e.g., LAI, albedo, roughness length, and canopy fraction) are not taken into account in this study. Previous studies have indicated that changes in runoff are sensitive to albedo (Bennett et al., 2018) and vegetation roots (Demaria et al., 2007) in the VIC model. Therefore, the optimal solution found by our automatic calibration framework is likely to be compensating for errors in some of the other vegetation parameters and model prediction uncertainty sources, including input data and structural uncertainties associated with the underlying model. In addition, some studies have identified a significant role of parameter sensitivity for understanding of model behavior across watersheds (van Werkhoven et al., 2008) and for simplification of hydrologic modeling (Markstrom et al., 2016). Thus, comprehensive parameter sensitivity (including vegetation and other hydrological-variable-related parameters) across a hydroclimatic gradient should be explored further to improve the understanding of the hydrologic processes in the model. Furthermore, in

Figure 7. The ASMO optimization results for streamflow across 10 major river basins in China during the validation period. Streamflow was simulated using the same catchment-specific parameters as in the calibration period. All other information is consistent with Figure 6. Monthly streamflow time series are expressed in \(10^2\) m\(^3\)/s.
this study, a simple postregionalization, that is, the nearest neighbor method, was used to transfer the calibrated parameters to the ungauged catchments, which may introduce physically unreasonable estimates for some of those unknown parameters, especially at the basin boundaries. In future work, a better simultaneous regionalization method, such as the multiscale parameter regionalization technique by Samaniego et al. (2010) will need to be carried out through simultaneous calibration of transfer function parameters by assuming prior relationships between basin predictors (e.g., elevation, slope, soil texture, or vegetation characteristics) and model parameters.

5. Conclusions

This paper describes an automatic calibration framework for the VIC model that integrates a SA and an ASMO algorithm. In this study, a SA was conducted to examine streamflow sensitivities to 13 streamflow-related VIC parameters following the use of four SA methods (three qualitative and one quantitative) across China. Parameter estimation was performed at 30 hydrological stations during the periods 1961–1969 (the calibration period) and 1970–1979 (the validation period). This calibration framework allows us to identify the important parameters for a given basin and avoid both type I and type II errors resulting from using the traditional calibration procedure, which focuses exclusively on the seven parameters \( B, D_1, D_2, D_3, D_s, W_s, \) and \( D_m \) that are recommended by the model builder. Another important aspect of this study is the use of an ASMO algorithm we developed to tune the model parameters automatically, which circumvents the limitations of existing manual calibration techniques.

We identified three parameters, the infiltration parameter \( B \) and two of the soil depth parameters \( D_1 \) and \( D_2 \), that are highly sensitive in most basins, while other parameters sensitivities were strongly related to the dynamic environment surrounding the basin, including the climate, vegetation, soil, and topographic characteristics. In comparison with the default method (calibrating with the seven recommended parameters), our framework reduces time spent on calibration of insensitive parameters; in the case of the Liao River basin, the default method required three times as much time as our framework. Moreover, our framework can identify additional sensitive parameters specific to a basin (parameters that are not among the default suggested parameters). In the case of the Huai River basin, adding a parameter that we identified through SA-based optimization to the default seven-parameter calibration improved the optimal NSE by 0.0314. The SA-based optimization framework effectively avoids type I and type II errors from calibration when compared to the default calibration, which improves the efficiency and accuracy of model calibration.

The calibrated parameters performed reasonably well in representing the natural and observed streamflow at 30 hydrological stations of 10 major river basins. The NSE for monthly streamflow ranged from 0.75 to 0.97 and from 0.71 to 0.97 under the validation and calibration modes, respectively. About 80% and 53% of hydrological stations had an NSE greater than 0.80 during the validation and calibration periods, respectively.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (41622101, 41730645, and 41877155), and the State Key Laboratory of Earth Surface Processes and Resource Ecology. We are grateful to the Ministry of Water Resources of China (http://www.mwr.gov.cn/) for providing the natural and observed streamflow, the China Meteorological Administration (CMA) for providing the climatic data (http://data.cma.cn/), and the Cold and Arid Regions Sciences Data Center for providing the DEM data (http://westdc.westgis.ac.cn). We would like to thank the high-performance computing support from the Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University (https://gda.bnu.edu.cn/).

References

Al-Qurashi, A., McIntyre, N., Wheeler, H., & Unkrich, C. (2008). Application of the Kineros2 rainfall–runoff model to an arid catchment in Oman. Journal of Hydrology, 355(1–4), 91–105. https://doi.org/10.1016/j.jhydrol.2008.03.022

Atkinson, S. E., Woods, R. A., & Sivapalan, M. (2002). Climate and landscape controls on water balance model complexity over changing timescales. Water Resources Research, 38(12), 1314. https://doi.org/10.1029/2002WR001487

Bao, Z., Liu, J., Zhang, J., Fu, G., Wang, G., Yan, X., et al. (2011). Estimation of baseflow parameters of variable infiltration capacity model with soil and topography properties for predictions in ungauged basins. Hydrology and Earth System Sciences Discussions, 8(4), 7017–7053. https://doi.org/10.5194/hessd-8-7017-2011

Bennett, K. E., Blanco, J. R. U., Jonko, A., Bohn, T. J., Atchley, A. L., Urban, N. M., & Middleton, R. S. (2018). Global Sensitivity of Simulated Water Balance Indicators Under Future Climate Change in the Colorado Basin. Water Resources Research, 54, 132–149. https://doi.org/10.1002/2017WR020471

Bohn, T. J., Livneh, B., Oyler, J. W., Running, S. W., Nijssen, B., & Lettenmaier, D. P. (2013). Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. Agricultural and forest meteorology, 176, 38–49. https://doi.org/10.1016/j.agrformet.2013.03.003

Brooks, R. H., & Corey, A. T. (1964). Hydraulic Properties of Porous Media and Their Relation to Drainage Design. Transactions of the ASAE, 7(1), 26–28. https://doi.org/10.13031/2013.40684

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 27. https://doi.org/10.1145/1961189.1961199

Chawla, I., & Mozumdar, P. P. (2018). Partitioning Uncertainty in Streamflow Projections under Nonstationary Model Conditions. Advances in Water Resources, 122, 266–282. https://doi.org/10.1016/j.advwatres.2017.10.013
Chen, N. (1985). Changes in river regime after the construction of upstream reservoirs. *Earth Surface Processes and Landforms*, 10(2), 143–159. https://doi.org/10.1002/esp.3210100207

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266–298. https://doi.org/10.1214/09-aoas285

Choi, J.-Y., Engel, B. A., & Chung, H. W. (2002). Daily streamflow modelling and assessment based on the curve-number technique. *Hydrological Processes*, 16(16), 3131–3150. https://doi.org/10.1002/hyp.1092

Demaria, E. M., Nielsen, B., & Wagener, T. (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research*, 112, D11113. https://doi.org/10.1029/2006jd007334

Di, Z. H., Duan, Q. Y., Gong, W., Ye, A. Z., & Mao, C. Y. (2017). Parametric sensitivity analysis of precipitation and temperature based on multi-uncertainty quantification methods in the Weather Research and Forecasting model. *Science China Earth Sciences*, 60(5), 876–898. https://doi.org/10.1007/s11430-016-9201-6

Donigian, A. S., & Imhoff, J. C. (2009). Evaluation and Performance Assessment of Watershed Models. *Proceedings of the Water Environment Federation*, 2009(6), 12–45. https://doi.org/10.2175/193864709X3958066

Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Yan, Y., et al. (2017). Automatic Model Calibration: A New Way to Improve Numerical Weather Forecasting. *Bulletin of the American Meteorological Society*, 98(5), 959–970. https://doi.org/10.1175/bams-d-15-00164.1

Duan, Q., & Phillips, T. J. (2010). Bayesian estimation of local signal and noise in multimodel simulations of climate change. *Journal of Geophysical Research*, 115, D18123. https://doi.org/10.1029/2009jd013654

Duan, Q. Y., Sorooshian, S., & Gupta, V. (1992). Effective and Efficient Gofal Optimization for Conceptual Rainfall-Runoff Models. *Water Resources Research*, 28(4), 1015–1031. https://doi.org/10.1029/91wr02985

Erola, E., Lilißänen, E., Lendasse, A., Corona, F., & Verleysen, M. (2008). Using the Delta Test for Variable Selection. Paper presented at the 16th European Symposium on Artificial Neural Networks (ESANN 2008), Belgium.

FAO, 1998: Digital Soil Map of the World and Derived Soil Properties. *Land and Water Digital Media Series 1*, Food and Agriculture Organization, CD-ROM.

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. https://doi.org/10.1214/aos/1176347963

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., et al. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling & Software*, 51, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Ge, L., Xie, G., Zhang, C., Li, X., Qi, Y., Cao, S., & He, T. (2011). An evaluation of China’s water footprint. *Water Resources Management*, 25(10), 2633–2647. https://doi.org/10.1007/s11269-011-9830-1

Gong, W., Duan, Q., Li, J., Wang, C., di, Z., Ye, A., et al. (2016). Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models. *Water Resources Research*, 52, 1984–2008. https://doi.org/10.1002/2015wr018230

Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes. *Water Resources Research*, 28(10), 2639–2658. https://doi.org/10.1029/92wr01258

Greuell, W., Fransen, W. H. F., Biemans, H., & Hulse, R. W. A. (2018). Seasonal streamflow forecasts for Europe - Part I: Hindcast verification with pseudo- and real observations. *Hydrology and Earth System Sciences*, 22(6), 3453–3472. https://doi.org/10.5194/hess-22-3453-2018

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of Automatic Calibration for Hydrologic Models: Comparisons With Multilevel Expert Calibration. *Journal of Hydrologic Engineering*, 4(2), 135–143. https://doi.org/10.1061/(asce)1084-0699(1999)4:2(135)

Holland, J. H. (1992). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Ann Arbor: MIT press.

Hou, T., Zhu, Y., Lü, H. S., Sudicky, E., Yu, Z. B., & Ouyang, F. (2015). Parameter sensitivity analysis and optimization of Noah land surface model with field measurements from Huaihe River Basin, China. *Stochastic Environmental Research and Risk Assessment*, 29(5), 1383–1401. https://doi.org/10.1007/s00477-015-1033-5

Huang, Z., Tang, Q., Lo, M., Liu, X., & Lu, H. (2019). The influence of groundwater representation on hydrological simulation and its assessment using satellite-based water storage variation. *Hydrological Processes*, 33, 1218–1230. https://doi.org/10.1002/hyp.13393

Immerzeel, W. W., Van Beek, L. P., & Bleijs, M. F. (2010). Climate change will affect the Asian water towers. *Science*, 328(5984), 1382–1385. https://doi.org/10.1126/science.1183188

Intergovernmental Panel on Climate Change (IPCC) (2014). Summary for policymakers. In C. B. Field, et al. (Eds.), *Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, (pp. 1–32). Cambridge, UK and New York: Cambridge University Press.

Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Mao, C., et al. (2013). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences*, 17(8), 3297–3293. https://doi.org/10.5194/hess-17-3297-2013

Li, J. D., Duan, Q. Y., Wang, Y.-P., Gong, W., Yan, J. Y., & Wang, C. (2018). Parameter optimization for carbon and water fluxes in two global land surface models based on surrogate modelling. *International Journal of Climatology*, 38, e1016–e1031. https://doi.org/10.1002/joc.5428

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, 99(D7), 14415–14428. https://doi.org/10.1029/94jd00483

Liu, D., Guo, S., Shao, Q., Liu, P., Xiong, L., Wang, L., et al. (2018). Assessing the effects of adaptation measures on optimal water resources allocation under varied water availability conditions. *Journal of Hydrology*, 556, 759–774. https://doi.org/10.1016/j.jhydrol.2017.12.002

Lohmann, D., Nolte-Holube, R., & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus Series a-Dynamic Meteorology and Oceanography*, 48(5), 708–721. https://doi.org/10.3402/tellusa.v48i5.9953

Markstrom, S. L., Hay, L. E., & Clark, M. P. (2016). Towards simplification of hydrologic modeling: Identification of dominant processes. *Hydrology and Earth System Sciences*, 20, 4665–4671. https://doi.org/10.5194/hess-20-4665-2016

Miao, C. Y., Su, L., Sun, Q. H., & Duan, Q. Y. (2016). A nonstationary bias-correctation technique to remove bias in GCM simulations. *Journal of Geophysical Research: Atmospheres*, 121, 5718–5735. https://doi.org/10.1002/2015jd024519

Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, 306(1–4), 127–145. https://doi.org/10.1016/j.jhydrol.2004.09.005

Myneni, R. B., Ramakrishna, R., Nemani, R., & Running, S. W. (1997). Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, 35(6), 1380–1393. https://doi.org/10.1109/36.649788
Zheng, H. Y., Miao, C. Y., Wu, J. W., Lei, X. H., Liao, W. H., & Li, H. (2019). Temporal and spatial variations in water discharge and sediment load on the Loess Plateau, China: A high-density study. *Science of the Total Environment, 666*, 875–886. https://doi.org/10.1016/j.scitotenv.2019.02.246

Zhou, F., Xu, Y., Chen, Y., Xu, C. Y., Gao, Y., & Du, J. (2013). Hydrological response to urbanization at different spatio-temporal scales simulated by coupling of CLUE-S and the SWAT model in the Yangtze River Delta region. *Journal of Hydrology, 485*, 113–125. https://doi.org/10.1016/j.jhydrol.2012.12.040

**Reference From the Supporting Information**

Du, J., & Shi, C. X. (2012). Effects of climatic factors and human activities on runoff of the Weihe River in recent decades. *Quaternary International, 262*, 58–65. https://doi.org/10.1016/j.quaint.2012.06.036