Journal of Environmental Informatics 38(1) 16-26 (2021)

Journal of Environmental Informatics

www.iseis.org/jei

# Development of A Universal Calibration Platform for Watershed Models Using Global Optimization

L. Chen[1], W. Z. Wang[1], G. Y. Wei[1], K. Wang[1], H. B. Liu[2], L. M. Zhai[2], and Z. Y. Shen[1 *]

[1] *State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, P.R. China*
[2] *Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Key Laboratory of Non-point Source Pollution Control, Ministry of Agriculture, Beijing 100081, P.R. China*

Received 15 November 2018; revised 26 December 2019; accepted 24 February 2020; published online 21 April 2021

**ABSTRACT.** Calibration is an essential part of watershed models, and a universal calibration platform based on advanced genetic algorithms is needed. In this study, a universal platform was constructed for different watershed models by transferring the configuration files of models and incorporating the Non-dominated Sorted Genetic Algorithm-II (NSGA-II). It was tested in two real cases studies by using two commonly used models, including the Hydrological Simulation Program-FORTRAN (HSPF) and the Storm Water Management Model (SWMM). For HSPF, the results showed that the goodness-of-fit indicators, in terms of NSE and $R^2$, were 0.82, 0.83 and 0.66, 0.67 during the calibration and validation period, respectively. For SWMM, NSE ranged from 0.854 to 0.920 and $R^2$ ranged from 0.737 to 0.912. The results indicated that this universal platform provided good model calibrations for both two models and it could be extended to other watershed models and other catchments as an effective and robust method for model calibration.

*Keywords:* watershed model, calibration, parameter, universal platform, NSGA-II, SWMM, HSPF

## 1. Introduction

A model is a formalized expression that abstracts some practical problems or objective rules, and a watershed model represents a mathematical platform for simulating the entire hydrological process in a watershed (Clarke, 1973). And watershed models have become vital tools in dealing with many practical and challenging issues that arise in watershed management (Chen et al., 2016; Liu et al., 2016; Chen et al., 2017c). It can characterize the statistics of flow, sediment, and pollution cycles (Ouyang et al., 2009; Chen et al., 2014; Chen et al., 2017a, 2017b). According to their basic calculation unit, watershed hydrological models can be divided into lumped models and distributed models, which have become commonly used tools due to their descriptions of spatial variation of climate input as well as underlying processes (Freeze and Harlan, 1969). At present, numerous watershed models have been developed such as the Hydrological Simulation Program-FORTRAN (HSPF), the Storm Water Management Model (SWMM), the Soil and Water Assessment Tool (SWAT), and the Variable Infiltration Capacity (VIC) model (Becknell et al., 1993; Arnold and Allen, 1996; Cherkauer et al., 2003; Ross-man, 2009).

Model parameters are variables used for describing specific hydrological processes, functions or equations. Generally, parameters can be categorized as physical, geometric and empirical parameters. Physical parameters represent the specific description of watershed function and most can be directly measured but they are often estimated because of large human efforts and the amount of time involved. Therefore, model calibration has become the most important task for nearly all modelling studies (Liang et al., 2011) whose purpose is to find a set of optimal parameters through searching to make the best fit between simulations and observations.

Typically, parameter calibration can be categorized as manual or automatic. During manual calibration, model parameters are determined by manual debugging, which could require much knowledge and experience of modelers. Moreover, many parameters exist for the distributed models, such as the SWAT model that contains over 200 parameters, which requires unbearable time and effort during the calibration process. Recently, automatic calibration has become more popular due to the detailed understanding of watershed processes as well as the development of computers and artificial intelligence (Gupta et al., 1998). Several optimization methods have been used for parameter calibration such as Particle Swarm Optimization (PSO), the Genetic Algorithm (GA), Shuffled Complex Evolution (SCE-UA), Gauss-Marquardt-Levenberg (GML), Shuffled Frog Leaping Algorithm (SFLA), and so on (Duan et al., 1993; Shi and Eberhart, 2002; Doherty and Johnston, 2010; Sahoo et al., 2010; Eusuff and Lansey, 2015). To date several watershed models have incorporated an automatic optimization algorithm for a better search of optimal parameters. For example, Song et al. (2012) coupled a hydrological meta-modelling approach

* Corresponding author. Tel.: + 86 10 58804733; fax: + 86 10 58804733
*E-mail address:* zyshen@bnu.edu.cn (Z. Y. Shen).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.202100454

with the SCE-UA algorithm for model calibration. The SLFA algorithm, which combines two kinds of intelligent optimization algorithms, named Memetic Algorithm (MA) and PSO, can also be applied to the actual water resources allocation problem (Chutima and Olanviwatchai, 2010; Fang et al., 2018). Nonetheless, conventional optimization algorithms face difficulties in achieving global optimization when searching an entire parameter space, especially for the distributed models. To solve this problem, researchers have noted that GA, an adaptive global optimization algorithm, could solve the Multi-objective Optimization Problem (MOP) in the calibration of watershed models (Goldberg, 1989; Wang et al., 2018). As an improved multi-objective genetic algorithm, the Nondominated Sorted Genetic Algorithm-II (NSGA-II) has been widely and efficiently applied in various disciplines (Bekele and Nicklow, 2007; De Vos and Rientjes, 2008; Shafii and De Smedt, 2009; Chen et al., 2016). Ercan and Goodall (2016) have proven that NSGA-II can be applied to the calibra-tion of the SWAT model. Because of its easier modification and better searching ability, NSGA-II can be used for better parameter calibration, especially for complex physical-distributed models. Moreover, as far as we know, many models, such as HSPF model, do not have any auto-calibration platform, which would cost much effort in their complex operations. Even some models with their own calibration platforms still face difficulties when they are used for comparison and ensemble prediction of multiple models because of the lack of a universal calibration platform.

Moreover, some calibration platforms are very expensive, which may hinder the automatic calibration of these models. Some free calibration software did exist, such as the SWAT Calibration and Uncertainty Programs (SWAT-CUP) and the Parameter Estimation Software (PEST), but they could not serve as a universal calibration platform. Specifically, SWAT-CUP, which is developed by Abbaspour, is based on the interface for SWAT and only enables calibration of SWAT models (Abbaspour et al., 2007; Liu et al., 2017). Instead, PEST, developed by Doherty (2008), has been developed as independent parameter estimation software based on the GML algorithm. However, it still needs complex operations and should improve its global optimization ability for more accurate calibration. In this sense, a universal and free calibration platform based on the better optimization algorithm is needed for the development of hydrological studies.

In this study, a calibration platform based on NSGA-II with an elitist strategy was designed to search for the optimal parameters of models, and the universal calibration platform was designed to adapt to most models. In this article, firstly, the flow chart of this universal calibration platform was presented, and then the setting of the universal interface as well as the connection between the universal calibration platform and NSGA-II were described. Then, two actual applications of this platform were demonstrated, calibrating the HSPF model in a typical small agricultural watershed in the Three Gorges Reservoir area of China and the SWMM model of a typical catchment community in Beijing City, China.

## 2. Design of the Universal Calibration Platform

### 2.1. Flow Chart of the Universal Calibration Platform

The main chart of the calibration platform is presented in Figure 1. The configuration of most watershed models' functions based on specific input and output files and can be executed by a command line programme. The characteristics of input and output files are as follows: first most files are stored in txt and binary format, which makes the design of the universal calibration platform possible by creating configuration files of different watershed models; second, model parameter values can be adjusted by finding the specific file of the calibrated parameters. To calibrate most watershed models, the universal calibration platform only needs to deal with the path and the format of the input and output files instead of hundreds of parameters, which constitutes the universal and easy features of this new calibration platform. In this study, a universal interface was created for realizing the calibration function for most watershed models, and Table 1 explains how this universal interface was structured.

### 2.2. The Global Optimization Engine

In this study, the designed calibration platform was based

**Table 1.** Design of Universal Interface for the Calibration Platform

| Parameter Rewriting Interface (Input File Rewriter) | |
|---|---|
| Type | Requirements |
| Text type | Requires row, column (character inch), format (% format notation) |
| Binary type | Requires offset and type (type tabulation) |
| Each write value can be specified by the value of the formula to flexibly respond to the correlation between different parameters. | |
| Simulation Results Interface (Output File Reader) | |
| Type | Requirements |
| Fixed text type | Requires row, column (character inch), format (% format notation), and applies to a text format output table aligned with space |
| Separator text type | Requires specification of rows, partitions, columns (separated by symbols), and applies to text format files dividing columns with separators (such as CSV file format) |
| Binary type | Requires offset and type (type tabulation) |
| Model running interface | Calls the model run through the command line (bash in Linux) command and specifies the parameter formula in the incoming parameters at the same time |

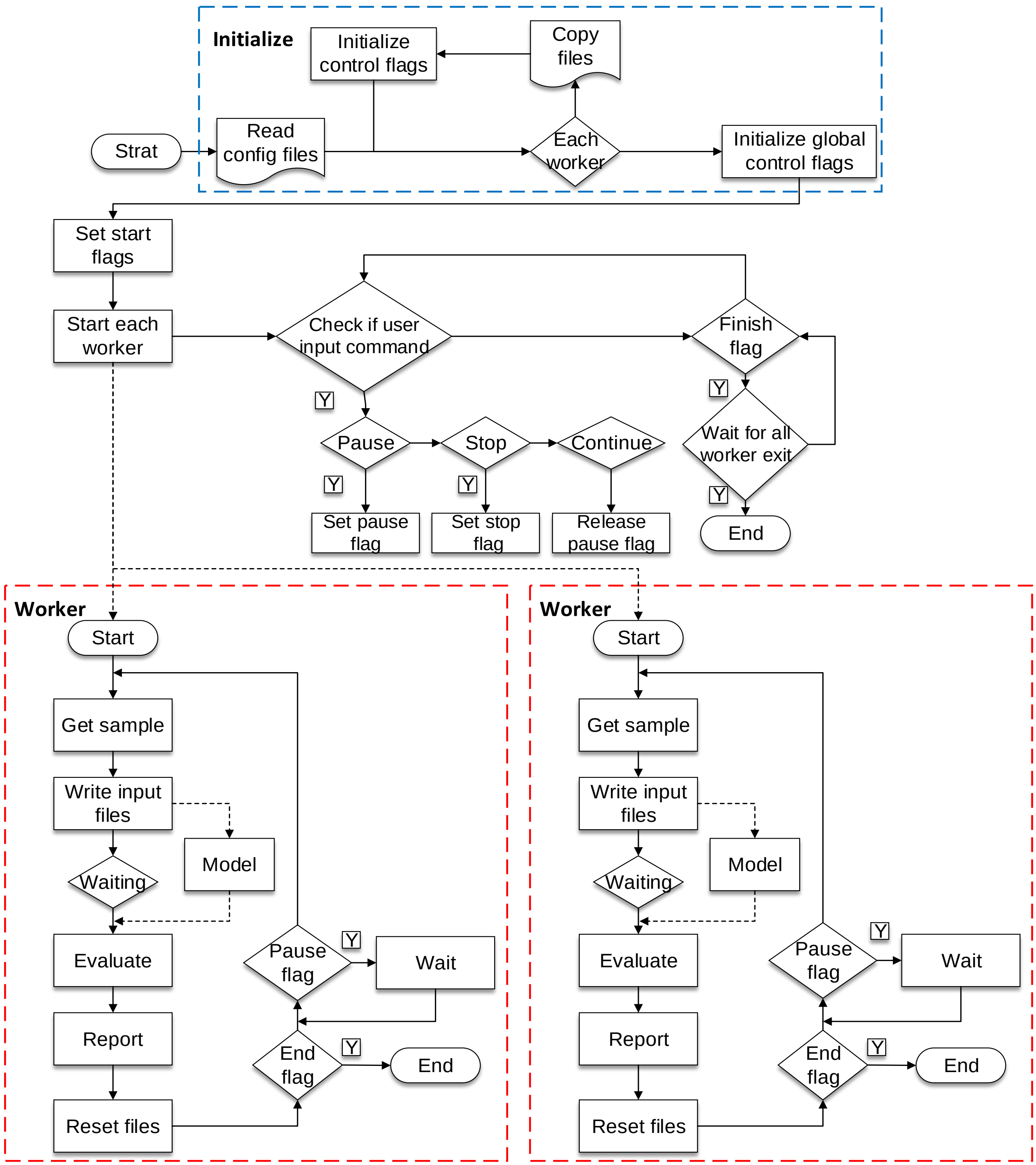


**Figure 1.** The design and implementation of the calibration platform.

on NSGA-II, which was presented as an improvement on the NSGA algorithm in 1995 by N. Srinivas and K (Srinivas and Deb, 1994). Compared with the traditional algorithms, NSGA-II uses the non-dominated dominance ranking method and the Pareto optimal frontier to sort the solution of parent and offspring (Chen et al., 2015). Recently, NSGA-II has been used in many watershed studies because it has overcome high computational complexity, the lack of elitism, as well as the need for specifying the sharing parameter (Deb et al., 2002). The Pareto optimal solution obtained by this algorithm has uniform distribution, good convergence and robustness, and a specific effect for multi-objective optimization problems that could benefit the optimization of model parameters (Zhang et al. 2012). In the universal platform, the mapping relation between the NSGA-II and most models was established, and a solution of the NSGA-II represented the model parameter set. Each gene stood for a group of model parameters that existed in a parameter set and a chromosome of a gene represents each single parameter (Ercan and Goodall, 2016). In the universal platform, the possible value range of each parameter which could be manually derived from the model or defined by the modelers was encoded, and the required parameter set was transformed into digital coding of the NSGA-II. Specifically, the gene value represented each possible value of the parameter, and the solution space for each parameter was generated according to the upper and lower boundaries of its parameter values. Therefore, each parameter set was treated as one specific chromosome with information on the decision variables.

During model calibration, the universal platform generates a group of populations (parameter sets) and evaluates their fitness between the simulation results and the observation values based on specific goodness-of-fit indicators (shown in section 2.3). These generated parameter sets are returned in binary form in the input file of the model, the objective function is extracted and the non-dominated fitness level is assigned based on the fitness between simulation and observation (Westenbroek et al., 2012). This process is based on the Time Series Processor (TSPROC.exe) Software, which is another general coding of this platform to assist in translating model calibration results in the form of text. Then, the new generation of chromosomes (parameter sets) is generated according to the fitness evaluation, and the chromosomes will not evolve until the maximum algebra is reached. Finally, a new offspring population is generated by the basic operation of the genetic algorithm, and so on, until the condition of the end of the model performance is satisfied.

### 2.3. Objective Function and Evaluation Indicator

For the universal platform, the observation values should be provided for model evaluation, and several objective functions were incorporated for better model calibration conducted by a regression measure, most commonly the point-to-point pairs (a series of single data pairs) of predicted and measured data. Several commonly used evaluation functions have been provided in the platform system (Table 2): 1) Mean-square error (MSE), 2) Root-mean-square error (RMSE) (Zhou and Bovik, 2009), 3) Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), 4) Absolute value of peak error (PFE), 5) Absolute value of total error (VE), and 6) Pearson correlation coefficient ($R^2$) (Tarald O. Kvålseth, 1985). Specifically, rainfall event simulation has become more popular during recent years, and several specific indicators are thus also set such as 1) the variance of average flow per day (SEF), 2) the variance of different flow levels (SEEF), and 3) the variance of the storm runoff of elected individual rainfall (SESSV). Modelers could also add other goodness-of-fit indicators as the objective functions of the global optimization and the evaluation index of the optimization results.

## 3. Case study

In this study, the universal calibration platform was tested for two different watershed models. The first case focused on the application of the HSPF model for a typical agricultural catchment in the Three Gorges Reservoir area of China. The second case used the SWMM model for a typical urban catchment in Beijing, China. Detailed information related to model construction and the calibration results are now provided for a clear comparison.

### 3.1. Case One

#### 3.1.1. Calibration Process of the HSPF

In the first case, the famous HSPF model (Bicknell et al., 1997), a semi-distributed watershed model, is selected because it is one of the most commonly used watershed models and can accurately describe the hydrological characteristics of a watershed. The HSPF is available through a download from https://pubs.usgs.gov/sir/2005/5099/. The hydrological part of the HSPF model is based on the Stanford model, which can simulate runoff, soil erosion, and water quality processes. This model has been widely applied to quantify the impacts of climate change and land use on watersheds among different climatic zones and countries, especially for the northern part of the United States. The HSPF was calibrated in a typical small agricultural watershed in the Three Gorges Reservoir area of China (Figure 2). The study area, with a total basin area of 1.62 $km^2$, has a subtropical monsoon climate with annual average temperature and rainfall of approximately 16.8 ℃ and 1200 mm, respectively. The soil type of the watershed is dominated by yellow brown loam soil, and the land use type is dominated by farmland, economic tea forest and woodland.

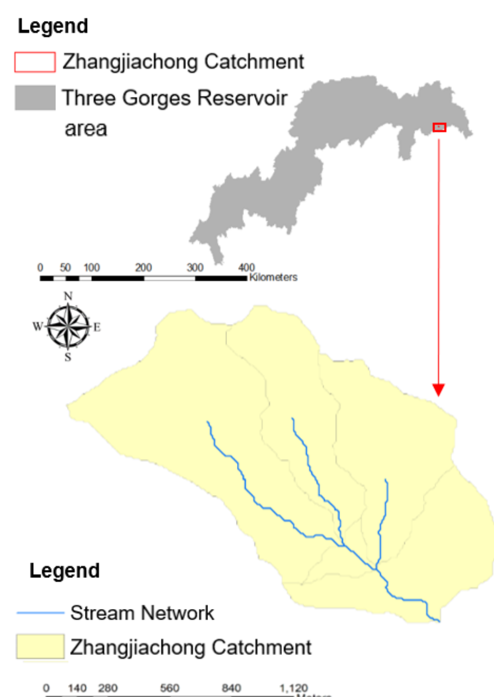


**Figure 2.** Location of the typical agricultural catchment in the Three Gorges Reservoir area of China.

The spatial data, such as the digital elevation model and land use map, were obtained from the local water and soil

**Table 2.** The Calculation Formulas of the Evaluation Functions

| Classification | Function | Calculation Formulas | Illustration |
|---|---|---|---|
| Conventional Index | Mean-square error (MSE) | $MSE = \sum_{i=1}^{N}\left(O_i - M_i\right)^2$ | where $Q$ represents the average daily runoff, $V$ represents runoff for selected rainstorms, $N$ is the total simulation time step, $i$ represents the simulation of each time point order period, $p$ represents the peak, $V$ represents total flow, $m$ represents the simulation results, $o$ represents the observed values for the selected st screenings, and 10, 50 and 90% represent the three flow levels. |
| | Root-mean-square error (RMSE) | $RMSE = \sqrt{\sum_{i=1}^{N}\left(O_i - M_i\right)^2}$ | |
| | Nash-Sutcliffe efficiency (NSE) | $NSE = 1 - \left[\frac{\sum_{i=1}^{N}\left(O_i - M_i\right)^2}{\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2}\right]$ | |
| | Absolute value of peak error (PFE) | $PFE = \left[\frac{Q_{O,p} - Q_{M,p}}{Q_{O,p}}\right] \times 100\%$ | |
| | Absolute value of total error (VE) | $VE = \left[\frac{V_O - V_M}{V_O}\right] \times 100\%$ | |
| | Pearson correlation coefficient($R^2$) | $R^2 = \left\{\frac{\sum_{i=1}^{N}\left(O_i - \bar{O}\right)\left(M_i - \bar{M}\right)}{\left[\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2\right]^{0.5}\left[\sum_{i=1}^{N}\left(M_i - \bar{M}\right)^2\right]^{0.5}}\right\}^2$ | |
| Individual rainfall Index | The variance of average flow per day (SEF) | $SEF = \sum_{i=1}^{N}\left(Q_{M,i} - Q_{O,i}\right)^2$ | |
| | The variance of different flow levels (SEEF) | $SEF = \sum_{i=1}^{N_{10\%}}\left(Q_{M,10\%,i} - Q_{O,10\%,i}\right)^2 + \sum_{i=1}^{N_{50\%}}\left(Q_{M,50\%,i} - Q_{O,50\%,i}\right)^2 + \sum_{i=1}^{N_{90\%}}\left(Q_{M,90\%,i} - Q_{O,90\%,i}\right)^2$ | |
| | The variance of the storm runoff of elected individual rainfall (SESSV) | $SESSV = \sum_{i=1}^{Nst}\left(V_{M,i} - V_{O,i}\right)^2$ | |

conservation bureau, and the soil type map was obtained by field investigation and the China Soil Scientific Database (Ouyang et al., 2012). The meteorological data, such as rainfall and temperature, were obtained from a local weather monitoring station, and other necessary management practices were investigated by discussion with local farmers. Then, the collected data were imported into the configuration file (the input file) of the model. Specifically, we monitored several rainfall events from January 1, 2010 to December 31, 2014 at the catchment outlet, and hydrological data were obtained by an automatic water level monitor (WGZ-1). Finally, the flow and water quality data were used for the calibration of the HSPF model. The calibration period and the validation period were selected as 01/01/2010 ~ 12/31/2011 and 01/01/2013 ~ 12/31/2014, respectively. And the spin up period for the continuous modeling was one-year (01/01/2010 ~ 12/31/2010). More information about the study area can be found in our previous study (Xie et al., 2017).

### 3.1.2. Calibration Results

Figure 3 shows the link of the HSPF and NSGA-II process for a better explanation of how the universal calibration platform is connected to the HSPF model. In this case, the initial population number and the maximum algebra were set to 600 and 2200, respectively. The goodness-of-fit indicators used in this study are derived from the HSPEXP model (Lumb et al., 1994), which is a manual assessment index recommended by the United States Environmental Protection Agency (EPA). Three traditional indicators were used to evaluate the effectiveness: NSE, $R^2$ and RMSE. Additionally, we chose SEF, SEEF and SESSV as three objective functions and the global optimal solution (best overall performance, OP) and the optimal solution interval (optimized parameter range) for the parameters.

**Table 3.** Criteria of the HSPEXP Model

| Measure | Criteria, % error |
|---|---|
| Total runoff | ±10% |
| Highest 10% flows | ±10% |
| Lowest 50% flows | ±15% |
| Seasonal volume | ±10% |
| Storm peak | ±15% |
| Summer storm volume | ±15% |

Table 4 lists the calibrated parameters and the calibration results for the HSPF model. The global optimal solution was based on the Pareto optimal frontier and the origin of the shortest calculated Euclidean distance (Hallema et al., 2013). The optimal solution of the interval of each parameter was calculated according to the following two points: 1) selecting parameters to meet the group of 6 criteria in the HSPEXP model (Table 3) (Lumb et al., 1994); and 2) the original interval of each parameter group is normalized between 0 and 1, and the parameter interval in step 1 is obtained. The preliminary screening of step 1 helps the modelers to find a more reasonable parameter range and to improve the reliability of the subsequent comparative analysis.

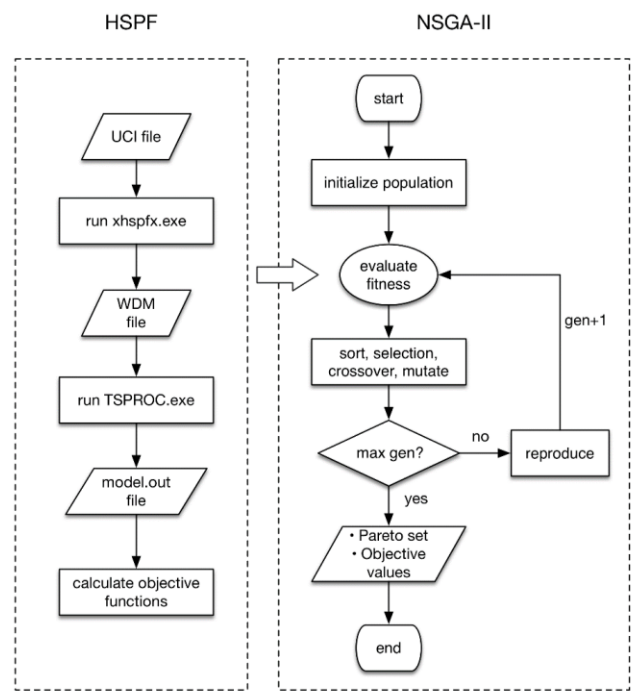


**Figure 3.** Flow diagram of HSPF/NSGA-II.

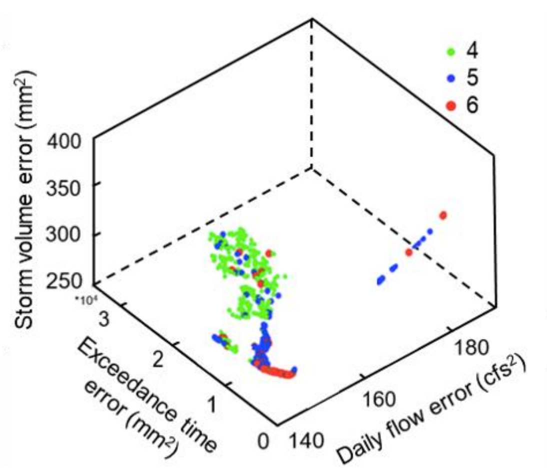


**Figure 4.** Pareto sets achieved by multi-objective calibration.

Then, the multi-target calibration ran 1,320,000 times over 183 hours on a desktop personal computer (Centrino Duo processor running at 2.8 GHz), and the Pareto frontier, which contained 600 parameters, was finally obtained, as shown in Figure 4. There were 345, 175, and 80 parameter groups that can satisfy criteria 4, 5, and 6 in the HSPEXP model, and the green, blue, and red dots represent these three solutions, respectively. Figure 4 also showed the evolution trend of model parameters, in which closeness to the origin of the coordinates indicates a better approximation of the optimal solution. Finally, the optimal parameter set can be reached in the Pareto frontier.

The calibrated results can be found in Table 4. In general, a good fit between simulation and observation was obtained in both the dry and high flow periods. In the process of calibration, INFILT, which was an index of the mean soil infiltration rate, was adjusted in range from the default value 0.003 to the optimal value 0.135, and the interflow inflow parameter (INTFW) was calibrated to the optimal value of 1.000. During the calibration and validation period, the values of NSE, $R^2$ and RMSE were 0.82, 0.83 and 0.76 mm and 0.66, 0.67 and 2.37 mm, respectively, which indicates a very good performance of the universal calibration platform. The model performance was judged as good compared with that reported in previous literature (Chung et al., 2011; Fonseca et al., 2014; Chounghyun et al., 2015; Hayashi et al., 2015), which indicates that this universal calibration platform could be applied to those agricultural catchment predictions, which used the HSPF and other similar models.

### 3.2. Case Two

#### 3.2.1. Model Calibration for SWMM

In the second case, SWMM, a dynamic rainfall runoff simulation tool developed by the United States Environmental Protection Agency in 1971, was selected. SWMM has become one of the most widely used models in simulating urban floods, water quality, drainage network design, and low impact development (LID) (Huber et al., 1995; Rossman, 2009; Dai et al., 2018). The SWMM was tested in a typical urban catchment (Beijing Normal University) in Beijing, China (Figure 5). The SWMM is available from https://www.epa.gov/water-research/storm-water-management-model-swmm. And in this study, the PCSWMM 2014 was used, and the engine was SWMM 5.1.007. The study area is located in a typical semi-humid continental monsoon climate, with an annual average temperature range of 10 to 12 ℃. Precipitation is mainly concentrated in several heavy rainstorms in late July and early August. The input data were mainly from local departments and field monitoring. Specifically, detailed land use data (Mxd format) were downloaded from the information network centre of Beijing Normal University, and the rainwater pipe network data came from the logistics management department. The DEM data were derived based on elevation data of all rainwater nodes (2048 points), as well as elevation data for 622 points from a local survey. The meteorological data, such as rainfall, relative humidity, wind speed, solar radiation and other data at 5-min intervals, came from a local HOBO weather station. The catchment area was divided into 749 sub-watershed areas for a detailed simulation. During the study period, eight rainfall events were monitored, and the first and latter four rainfall events were selected for model calibration and validation (Table 5).

For a genetic algorithm, the search space increases exponentially with the number of calibration parameters, and the result of calibration will also be affected by the existence of many fixed parameters. Therefore, only the important parameters were analysed by sensitivity analysis and then calibrated. For the SWMM model, surface depression storage (Dstore-imperv) is the only parameter that determines the impervious

**Table 4.** Parameter Optimization Results and Calibration Results for the HSPF Model by the Multi-Objective Genetic Algorithm NSGA-II

| | Calibration | Validation | |
|---|---|---|---|
| NSE | 0.82 | 0.66 | |
| $R^2$ | 0.83 | 0.67 | |
| RMSE | 0.76 | 2.37 | |
| Parameter | Description | Possible Range | Calibrated Values |
| AGWETP | Fraction of remaining evapotranspiration that be met from active groundwater storage | 0.0 ~ 0.20 | 0.000 |
| AGWRC | Groundwater recession rate | 0.85 ~ 0.999 | 0.987 |
| BASETP | Fraction of potential evapotranspiration which fulfilled only as outflow exists | 0.0 ~ 0.20 | 0.000 |
| CEPSC | Interception storage capacity | 0.01 ~ 0.40 | 0.010 |
| DEEPFR | Fraction of infiltrating water which enters deep aquifers | 0.0 ~ 0.50 | 0.005 |
| INFILT | Index to mean soil infiltration rate | 0.001 ~ 0.50 | 0.003 ~ 0.135 |
| INTFW | Interflow inflow parameter | 1.0 ~ 10.0 | 1.000 |
| IRC | Interflow recession parameter | 0.3 ~ 0.85 | 0.300 |
| LZETP | Index to lower zone evapotranspiration | 0.1 ~ 0.9 | 0.137 ~ 0.900 |
| LZSN | Low zone nominal soil moisture storage | 2.0 ~ 15.0 | 2.002 ~ 10.143 |
| KVARY | Parameter to describe non-linear groundwater recession rate | 0.0 ~ 0.50 | 3.643 |
| UZSN | Nominal upper zone soil moisture storage | 0.05 ~ 2.0 | 0.33 ~ 0.97 |

**Table 5.** Hydrometeorological Data for the Selected Events, Rainfall Date, Rainfall Duration, Rainfall Peak Intensity, Rainfall Depth, Runoff Duration, Peak Flow Rate (*Qp*), Runoff Volume (*V*), and Times to Peak (*Tp*)

| Events (y/mm/dd) | Rainfall Date (min) | Rainfall Peak Intensity (mm/h) | Rainfall Depth (mm) | Runoff Duration (min) | Peak Flow Rate ($m^3$/s) | Runoff Volume ($m^3$) | Times to Peak (min) |
|---|---|---|---|---|---|---|---|
| Calibration events | | | | | | | |
| 2014/07/29 | 400 | 43.28 | 35.7 | 940 | 0.5375 | 4740 | 170 |
| 2014/08/04 | 260 | 9.45 | 5.9 | 595 | 0.06908 | 490.9 | 90 |
| 2014/08/09 | 120 | 24.08 | 7.2 | 275 | 0.1401 | 499.9 | 45 |
| 2014/08/23 | 45 | 38.4 | 10.4 | 235 | 0.5032 | 989.4 | 15 |
| Validation events | | | | | | | |
| 2014/08/30 | 105 | 69.6 | 29 | 200 | 1.022 | 3159 | 30 |
| 2014/08/31 | 165 | 86.2 | 70.76 | 330 | 0.9653 | 7205 | 15 |
| 2014/09/01 | 1880 | 72 | 33.6 | 1875 | 0.7034 | 4098 | 1745 |
| 2014/09/26 | 20 | 50.4 | 7.8 | 160 | 0.3958 | 783 | 10 |

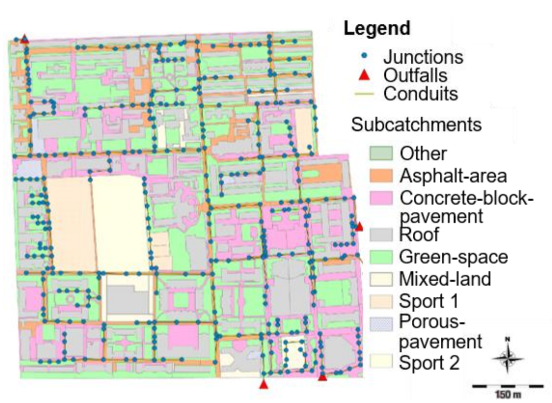


**Figure 5.** The catchment discretization for the study area (blue dots indicate the sewer inlets, yellow lines represent the stormwater sewer network, and grey lines are sub catchment boundaries).

surface losses after the impermeable rate is fixed, and this parameter was thus calibrated (Tsihrintzis and Hamid, 2015). All parameters were divided into 6 categories, marked D1-D6, based on the land use distribution of Beijing Normal University. The Manning's roughness for conduit (N-c), which represents the roughness of the inner wall of a pipe and determines the hydraulic load of a pipe, was also considered a key parameter (Krebs et al., 2014). Three objective functions, the NSE, PFE and VE, were used for model calibration.

3.2.2. Calibration Results

For the universal calibration platform, the crossover rate, the initial population number and the maximum algebra were set to 0.9, 100 and 200, respectively. The entire calibration took 40 hours on the desktop computer, with 290,000 runs through the calibration platform. The calibration results and the related NSE, PFE and VE values are shown in Table 6.

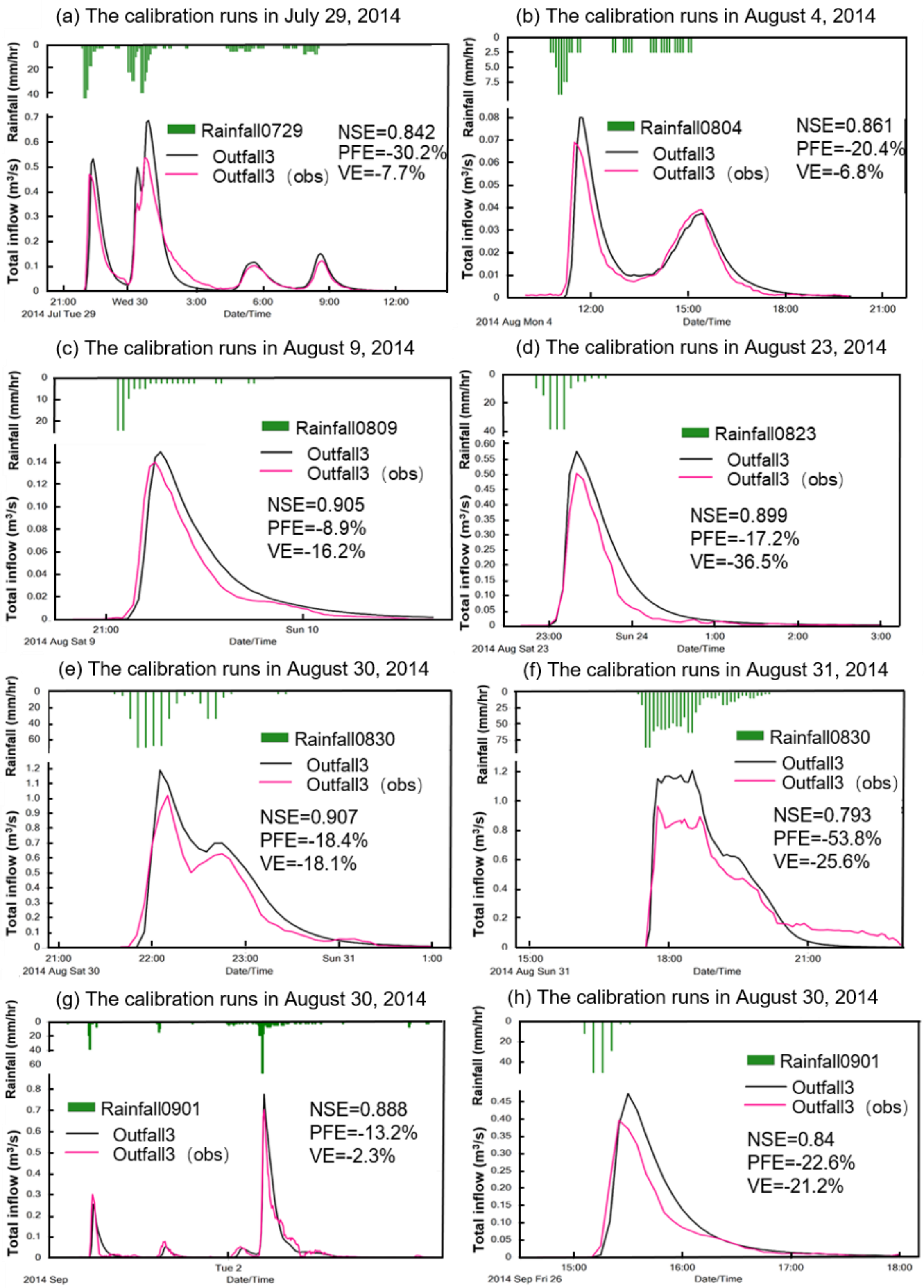


**Figure 6.** Comparison of predicted and measured hydrographs from (a) the calibration runs in July 29th, 2014; (b) the calibration runs in August 4th, 2014; (c) the calibration runs in August 9th, 2014; (d) the calibration runs in August 23th, 2014; (e) the validation runs in August 30th, 2014; (f) the validation runs in August 31th, 2014; (g) the validation runs in September 1st, 2014; (h) the validation runs in September 26th, 2014.

**Table 6.** Parameter Optimization Results and Calibration Results for the SWMM Model using the Multi-Objective Genetic Algorithm NSGA-II

| | Calibration | Validation | |
|---|---|---|---|
| NSE | 0.854 ~ 0.920 | 0.737 ~ 0.912 | |
| PFE | -30.2 ~ 8.9% | -53.8 ~ 13.2% | |
| VE | -36.5 ~ 6.8% | -2.3 ~ 25.6% | |
| **Parameter** | **Description** | **Possible Range** | **Calibrated Values** |
| N-c | Manning's roughness conduit | 0.011 ~ 0.024 | 0.166 |
| Dstore-imperv | Depression storage impervious | 1.000 ~ 5.000 | 3.500 |
| D1 | Asphalt area | 1 ~ 2.5 | 1.151 |
| D2 | Roof | 1 ~ 2.5 | 1.225 |
| D3 | Concrete block pavement | 1 ~ 2.5 | 1.344 |
| D4 | Sport I | 1 ~ 2.5 | 1.77 |
| D5 | Sport II | 1 ~ 2.5 | 1.684 |
| D6 | Mixed land | 1 ~ 2.5 | 2.216 |
| Width | Width coefficient of flood in sub-watershed | 0.2 ~ 5 | 5.000 |
| N-perv | Manning's roughness pervious | 0.02 ~ 0.8 | 0.800 |
| Dstore-perv | Depression storage pervious | 3 ~ 10.2 | 10.200 |
| Horton-Max | Horton's maximum infiltration rate | 50 ~ 200 | 150.000 |
| Horton-Min | Horton's minimum infiltration rate | 0 ~ 20 | 20.000 |
| Horton-d | Horton's decay rate | 2 ~ 7 | 2.000 |
| N-imperv | Manning's roughness impervious | 0.011 ~ 0.033 | 0.012 |

Figure 6 is a comparison of the simulated and monitored values for model calibration. Of the four rainfall patterns, 0809 and 0823 were single-peak, 0804 was double-peak and 0729 was multi-peak. From the results, the manning coefficient of conduit was calibrated to the optimal value of 0.015, the NSE during the calibration process was from 0.854 to 0.920, and PFE and VE were -30.2% to 8.9% and -36.5% to 6.8%, respectively. During the validation period, NSE, PFE and VE ranged from 0.737 to 0.912, -53.8% to 13.2%, -2.3% to 25.6%, respectively. The calibration results of the universal calibration platform can be judged as good compared to other studies, such as the research by Rosa et al. (2015) with value of NSE were 0.413 of total nitrogen (TN) and 0.134 of TP in the traditional watersheds. The second case study indicates that the universal calibration platform could provide good calibration of the SWMM and other similar models in those typical urban catchments.

Comparing these two different cases, the performances of the universal calibration platform were different, such as different operating speeds of calibration process and simulation results. The SWMM model ran less time and times to achieve the desired results than the HSPF model. And the simulation results in SWMM model also were better than the HSPF model though the universal calibration platform.

## 4. Conclusions

In this paper, a universal calibration platform based on the NSGA-II algorithm was constructed to provide the global optimization of parameters for different watershed models. This platform was then tested on two real applications using the HSPF and the SWMM models. The results indicated that the calibration platform performed well in both cases; optimal parameter sets could be obtained by interaction between the calibration platform and the watershed model. In general, the new platform developed in this study can easily be extended to any other watershed model and to other environmental models such as fluid mechanics models, groundwater models, soil mechanics models, structural mechanics models and so on.

In practice, there could be several concerns regarding the application of this universal platform. On one hand, we do not claim that this universal platform is always the optimal way for all models across catchments or computational budgets. Further studies are still needed to speed up the calibration process (183 and 40 hours for the two case studies). In the future, the cloud technique or other advanced techniques are needed for this purpose. At the same time, the platform can be improved in the future by using multi-algorithm methods such as multi-algorithm Genetic Adaptive multi-objective (AMALGAM) method (Vrugt and Robinson, 2007), which uses self-adaptive offspring creation to combine the advantages of a variety of optimization algorithms with the calibration platform to create better calibration results. In addition, a very robust and parsimonious method is also needed for better calibrations, and more uncertainty analysis functions should be incorporated in the future.

**Acknowledgements.** This research was funded by the Fund for the Innovative Research Group of the National Natural Science Foundation of China (No.51721093), the National Natural Science Foundation of China (No.51779010), and the Interdisciplinary Research Funds of Beijing Normal University.

## References

Abbaspour, K.C., Vejdani, M., and Haghighat, S. (2007). SWAT-CUP calibration and uncertainty programs for SWAT, *Modsim International Congress on Modelling & Simulation Land Water & Environ-*

*mental Management Integrated Systems for Sustainability*, 364, 1603-1609. https://doi.org/10.1109/MICRO.2007.30.

Arnold, J.G. and Allen, P.M. (1996). Estimating hydrologic budgets for three Illinois watersheds, *J. Hydrol.*, 176, 57-77. https://doi.org/10.1016/0022-1694(95)02782-3.

Becknell, B.R., Imhoff, J.C., Kittle, J.L., Donigian, A.S., and Johanson, R.C. (1993). *Hydrological Simulation Program-FORTRAN user's manual for release 12*, Us Epa.

Bekele, E.G. and Nicklow, J.W. (2007). Multi-objective automatic calibration of SWAT using NSGA-II. *J. Hydrol.*, 341, 165-176. https://doi.org/10.1016/j.jhydrol.2007.05.014

Bicknell, B.R., Imhoff, J.C., Kittle, J.L.J., Donigian, A.S.J., and Johanson, R.C. (1997). *Hydrological Simulation Program-Fortran: User's manual for version 11*.

Chen, L., Li, S., Zhong, Y.C. and Shen, Z.Y. (2017a). Improvement of model evaluation by incorporating prediction and measurement uncertainty, *Hydrol. Earth Syst. Sci.*,1-24, https://doi.org/10.5194/hess-22-4145-2018.

Chen, L., Qiu, J.L., Wei, G.Y., and Shen, Z.Y. (2015). A preference-based multi-objective model for the optimization of best management practices, *J. Hydrol.*, 520, 356-366. https://doi.org/10.1016/j.jhydrol.2014.11.032.

Chen, L., Shen, Z.Y., Yang, X.H., Liao, Q., and Yu, S.L. (2014). An Interval-Deviation Approach for hydrology and water quality model evaluation within an uncertainty framework, *J. Hydrol.*, 509, 207-214. https://doi.org/10.1016/j.jhydrol.2013.11.043.

Chen, L., Sun, C., Wang, G.B., Xie, H., and Shen, Z.Y. (2017b). Event-based nonpoint source pollution prediction in a scarce data catchment, *J. Hydrol.*, 552, 13-27. https://doi.org/10.1016/j.jhydrol.2017.06.034.

Chen, L., Sun, C., Wang, G.B., Xie, H., and Shen, Z.Y. (2017c). Modeling Multi-Event Non-Point Source Pollution in a Data-Scarce Catchment Using ANN and Entropy Analysis, *Entropy*, 19. https://doi.org/10.3390/e19060265.

Chen, L., Wei, G.Y., and Shen, Z.Y. (2016). Incorporating water quality responses into the framework of best management practices optimization, *J. Hydrol.*, 541, 1363-1374. https://doi.org/10.1016/j.jhydrol.2016.08.038.

Cherkauer, K.A., Bowling, L.C., and Lettenmaier, D.P. (2003). Variable infiltration capacity cold land process model updates, *Global Planetary Change*, 38, 151-159. https://doi.org/10.1016/S0921-8181(03)00025-0.

Chounghyun, S., Younggu, H., and Brian, B. (2015). Automatic calibration tool for hydrologic simulation program-FORTRAN using a shuffled complex evolution algorithm, *Water*, 7, 503-527. https://doi.org/10.3390/w7020503.

Chung, E.S., Park, K., and Lee, K.S. (2011). The relative impacts of climate change and urbanization on the hydrological response of a Korean urban watershed, *Hydrol. Process.*, 25, 544-560. https://doi.org/10.1002/hyp.7781.

Chutima, P. and Olanviwatchai, P. (2010). Mixed-model U-shaped assembly line balancing problems with coincidence memetic algorithm, *J. Software Eng. Appl.*, 3, 347-363. https://doi.org/10.4236/jsea.2010.34040.

Clarke, R.T. (1973). A review of some mathematical models used in hydrology, with observations on their calibration and use, *J. Hydrol.*, 19, 0-20. https://doi.org/10.1016/0022-1694(73)90089-9.

Dai, Y., Chen, L., Hou, X.S., and Shen, Z.Y. (2018). Effects of the spatial resolution of urban drainage data on nonpoint source pollution prediction, *Environ. Sci. Pollut. Res. Int.*, 25, 1-14. https://doi.org/10.1007/s11356-018-1377-8.

De Vos, N.J. and Rientjes, T.H.M. (2008). Multiobjective training of artificial neural networks for rainfall-runoff modeling, *Water Resour. Res.*, 44, 134-143. https://doi.org/10.1029/2007WR006734.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, 6, 182-197. https://doi.org/10.1109/4235.996017.

Doherty, J. and Doherty, J. (2008.) *PEST, model independent parameter estimation user manual*, Watermark Computing, Corinda, Australia.

Doherty, J. and Johnston, J. M. (2010). Methodologies for Calibration and Predictive Analysis of a Watershed Model, *Jawra J. Am. Water Resour. Assoc.*, 39, 1563-1564. https://doi.org/10.1111/j.1752-1688.2003.tb04381.x.

Duan, Q.Y., Gupta, V.K., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization, *J. Optimiz. Theory Appl.*, 76, 501-521. https://doi.org/10.1007/BF00939380.

Ercan, M.B. and Goodall, J.L. (2016). Design and implementation of a general software library for using NSGA-II with SWAT for multi-objective model calibration, *Environ. Model. Software*, 84, 112-120. https://doi.org/10.1016/j.envsoft.2016.06.017.

Eusuff, M.M. and Lansey, K.E. (2015). Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm, *J. Water Resour. Plan. Manag.*, 129, 210-225. https://doi.org/10.1061/(ASCE)0733-9496(2003)129:3(210).

Fang, G., Yuxue, G., Xin, W., Xiaomin, F., Xiaohui, L., Yu, T., and Ting, W. (2018). Multi-Objective Differential Evolution-Chaos Shuffled Frog Leaping Algorithm for Water Resources System Optimization, *Water Resour. Manag.*, 32, 3835-3852. https://doi.org/10.1007/s11269-018-2021-6.

Fonseca, A., Ames, D.P., Yang, P., Botelho, C., Boaventura, R., and Vilar, V. (2014). Watershed model parameter estimation and uncertainty in data-limited environments, *Environ. Model. Software*, 51, 84-93. https://doi.org/10.1016/j.envsoft.2013.09.023.

Freeze, R.A. and Harlan, R.L. (1969). Blueprint for a physically-based, digitally-simulated hydrologic response model, *J. Hydrol.*, 9, 237-258. https://doi.org/10.1016/0022-1694(69)90020-1.

Goldberg, D.E. (1989). Genetic Algorithm in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, Massachusetts. xiii. https://doi.org/10.1111/j.1365-2486.2009.02080.x.

Gupta, H.V., Sorooshian, S., and Yapo, P.O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34, 3663-3674. https://doi.org/10.1029/97WR03495.

Hallema, D.W., Moussa, R., Andrieux, P., and Voltz, M. (2013). Parameterization and multi-criteria calibration of a distributed storm flow model applied to a Mediterranean agricultural catchment, *Hydrol. Process.*, 27, 1379-1398. https://doi.org/10.1002/hyp.9268.

Hayashi, S., Murakami, S., Xu, K.Q., and Watanabe, M. (2015). Simulation of the reduction of runoff and sediment load resulting from the Gain for Green Program in the Jialingjiang catchment, upper region of the Yangtze River, China, *J. Environ. Manag.*, 149, 126-137. https://doi.org/10.1016/j.jenvman.2014.10.004.

Huber, W.C., Rossman, L.A., and Dickinson, R.A. (1995). *EPA Storm Water Management Model SWMM 5.0. Computer Models of Watershed Hydrology*, EPA, US.

Krebs, G., Kokkonen, T., Valtanen, M., Setälä, H., and Koivusalo, H. (2014). Spatial resolution considerations for urban hydrological modelling, *J. Hydrol.*, 512, 482-497. https://doi.org/10.1016/j.jhydrol.2014.03.013.

Liang, Z.M., Li, B.Q., Yu, Z.B. and Chang, W.J. (2011). Application of Bayesian approach to hydrological frequency analysis, *Sci. China*, 54, 1183-1192. https://doi.org/10.1007/s11431-010-4229-4.

Liu, R.M., Wang, Q.R., Xu, F., Men, C., and Guo, L.J. (2017). Impacts of Manure Application on SWAT Model Outputs in the Xiangxi River Watershed, *J. Hydrol.*, 555. https://doi.org/10.1016/j.jhydrol.2017.10.044

Liu, R.M., Xu, F., Zhang, P.P., Yu, W.W., and Men, C. (2016). Identifying non-point source critical source areas based on multi-factors at a basin scale with SWAT, J. Hydrol., 533, 379-388. https://doi.org/10.1016/j.jhydrol.2015.12.024.

Lumb, A.M., Mccammon, R.B., and Jr, J.L.K. (1994). *Users manual* *for an expert system (HSPEXP) for calibration of the hydrological simulation program; Fortran*, Water-Resources Investigations Report.

Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I - A discussion of principles, *J. Hydrol.*, 10, 282-290. https://doi.org/10.1016/0022-1694(70)90255-6.

Ouyang, W., Huang, H., Hao, F., Shan, Y., and Guo, B. (2012). Evaluating spatial interaction of soil property with non-point source pollution at watershed scale: The phosphorus indicator in Northeast China, *Sci. Total Environ.*, 432, 412-421. https://doi.org/10.1016/j.scitotenv.2012.06.017.

Ouyang, W., Wang, X.L., Hao, F.H., and Srinivasan, R. (2009). Temporal-spatial dynamics of vegetation variation on non-point source nutrient pollution, *Ecol. Model.*, 220, 2702-2713. https://doi.org/10.1016/j.ecolmodel.2009.06.039.

Rosa, D.J., Clausen, J.C., and Dietz, M.E. (2015). Calibration and Verification of SWMM for Low Impact Development, *Jawra J. Am. Water Resour. Assoc.*, 51, 746-757. https://doi.org/10.1111/jawr.1 2272.

Rossman, L.A. (2009). *Storm Water Management Model User's Manual*.

Sahoo, D. Smith, P.K., and Ines, A.V.M. (2010). Autocalibration of HSPF for Simulation of Streamflow Using a Genetic Algorithm, *Trans. Asabe*, 53, 75-86. https://doi.org/10.13031/2013.29504.

Shafii, M. and De Smedt, F. (2009). Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, 13, 2137-2149. https://doi.org/10.5194/ hess-13-2137-2009.

Shi, Y. and Eberhart, R.C. (2002). *Empirical study of particle swarm optimization*.

Song, X.M., Zhan, C.S., and Xia, J. (2012). Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model, *Sci. Bull.*, 57, 3397-3403. https://doi.org/10.1007/s11434-012-5305-x

Srinivas, N. and Deb, K. (1994). *Muiltiobjective optimization using nondominated sorting in genetic algorithms*. https://doi.org/10.1162/evco.1994.2.3.221

Kvålseth, T.O. (1985). Cautionary note about $R^2$, *Am. Stat.*, 39, 279-285. https://doi.org/10.1080/00031305.1985.10479448.

Tsihrintzis, V.A. and Hamid, R. (2015). Runoff quality prediction from small urban catchments using SWMM, *Hydrol. Process.*, 12, 311-329.https://doi.org/10.1002/(SICI)1099-1085(199802)12:2<3 11::AID-HYP579>3.0.CO;2-R.

Vrugt, J.A. and Robinson, B.A. (2007). Improved evolutionary optimization from genetically adaptive multimethod search, *Proceedings of the National Academy of Sciences of the United States of America*, 104, 708-711, 2007. https://doi.org/10.1073/pnas. 0610471104.

Wang, Q.R., Liu, R.M., Men, C., and Guo, L.J. (2018). Application of genetic algorithm to land use optimization for non-point source pollution control based on CLUE-S and SWAT, *J. Hydrol.*, 560. https://doi.org/10.1016/j.jhydrol.2018.03.022.

Westenbroek, S.M., Doherty, J., Walker, J.F., Kelson, V.A., Hunt, R.J., and Cera, T.B. (2012). *Approaches in highly parameterized inversion: TSPROC, a general time-series processor to assist in model calibration and result summarization, Techniques and Methods*, Reston, VA, U.S.G. Survey. https://doi.org/10.3133/tm 7C7.

Xie, H., Shen, Z.Y., Chen, L., Qiu, J.L., and Dong, J.W. (2017). Time-varying sensitivity analysis of hydrologic and sediment parameters at multiple timescales: Implications for conservation practices, *Sci. Total Environ.*, 598, 353-364. https://doi.org/10.1016/j.scitotenv.2017.04.074.

Zhang, X.S., Izaurralde, R., Zong, Z., Zhao, K.G., and M. Thomson, A. (2012). Evaluating the efficiency of a multi-core aware multi-objective optimization tool for calibrating the SWAT Model, *Trans. Asabe*, 55, 1723-1731. https://doi.org/10.13031/2013.42363.

Zhou, W., and Bovik, A.C. (2009). Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures, *IEEE Signal Process. Mag.*, 26, 98-117. https://doi.org/10.1109/MSP.2008.930649.