FACHBREICH GEOWISSENSCHAFTEN

INSTITUT FÜR GEOGRAPHISCHE WISSENSCHAFTEN

Evaluating the Effect of Objective Functions on Model Calibration

Dissertation

zur Erlangung des Grades

eines Doktors der Naturwissenschaften

(doctor rerum naturalium)

am Fachbereichs Geowissenschaften

der Freien Universität Berlin

vorgelegt von

Qinbo Cheng

Berlin 2014

Erstgutachter:

                Prof. Dr. Achim Schulte

                   Freie Universität Berlin

                   Institut für Geographische Wissenschaften

                   Fachrichtung Angewandte Geographie,

                   Umwelthydrologie und Ressourcenmanagement

Zweitgutachter:

                Prof. Dr. Björn Waske

                   Freie Universität Berlin

                   Institut für Geographische Wissenschaften

                   Fachrichtung Fernerkundung und Geoinformatik

Datum der mündlichen Prüfung: 17.12.2014

Declaration of Authenticity

I hereby certify sole authorship of the thesis "Evaluating the effect of objective functions on model calibration" as submitted to the Department of Earth Sciences of the "Freie Universität Berlin" for the conferral of a doctorate, and that no sources other than those indicated have been used in its preparation. Parts of this thesis that have been drawn on the work of others with regard to contents or by literal quotation have been appropriately marked through indication of source. Where any collaboration has taken place with other researchers, I have clearly stated my own personal share in the investigation.

This work in the same or a similar form has not been submitted to any other examining body.

Erklärung zur Originalität (Ehrenwörtliche Erklärung)

Hiermit versichere ich an Eides statt, dass ich die der Fachrichtung Geowissenschaften der Freien Universität Berlin zur Promotion eingereichte Arbeit "Evaluating the effect of objective functions on model calibration" selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Teile der vorliegenden Arbeit, die anderen Werken wörtlich oder inhaltlich entnommen sind, wurden durch Angaben der Quellen gekennzeichnet. In allen Fällen einer Zusammenarbeit mit anderen Forschenden habe ich meinen persönlichen Arbeitsanteil klar dargestellt.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegt.

Acknowledgements

Contents

List of Figures

List of Tables

ACRONYMS AND ABBREVIATIONS

ABC	Approximate Bayesian Computation
AR	AutoRegressive model
AR(1)	First-order autoregressive model
AR(p)	pth order autoregressive model
BC	Box-Cox transformation method
BC-GED	Error model assuming model residuals after BC follow GED
BC-SGED	Error model assuming model residuals after BC follow SGED
CE	Volume/Cumulative error
CN	Curve Number rainfall-runoff method
CN-ET	Plant evapotranspiration dependence Curve Number method
CN-Soil	Soil water content dependence Curve Number method
DEM	Digital Elevation Model
DREAM	DiffeRential Evolution Adaptive Metropolis MCMC
EDC	Effective Depth Coefficient
E_j	Generalized efficiency
G&A	Green & Ampt infiltration
GA	Genetic Algorithm
GED	Generalized Error Distribution
GIS	Geographic Information System
GLUE	Generalized Likelihood Uncertainty Estimation
GUI	Graphic-User-Interface
GW	Groundwater

HRU	Hydrologic Response Unit
HWSD	Harmonized World Soil Database
I.I.D.	Independent and Identically Distributed
ID	Identify
ITMSE	Inverse Transformed MSE
K	Soil hydraulic conductivity
LTMSE	Log Transformed MSE
MAE	Mean Absolute Error
MCMC	Markov Chain Monte Carlo
MS4E	Mean Quadrupled Error
MSE	Mean Squared Error
MUSLE	Modified Universal Soil Loss Equation
NSE	Nash-Sutcliffe Efficiency coefficient
PDF	Probability Density Function
R^2	Coefficient of determination
RTMSE	Square-Root transformed MSE
SA	Simulated Annealing
SCEM-UA	Shuffled Complex Evolution Metropolis (optimization algorithm)
SCE-UA	Shuffled Complex Evolution (optimization algorithm)
SCS	Soil Conservation Service
SD	Storage Deficit
SGED	Skew Generalized Error Distribution
SLS	Standard Least Squares
STI	Soil Topography Index
SWAT	Soil and Water Assessment Tool
S_y	Specific yield of unconfined aquifer
TI	Topographic Index
TOPMODEL	Topography Model
USDA	United States Department of Agriculture
USGS	United States Geological Survey
USLE	Universal Soil Loss Equation
VE	Volumetric Efficiency
VSA	Variable Source Areas
WB	Water Balance Model
WB-VSA	Water Balance Model with Variable Source Areas
β	Kurtosis or error power
λ	BC transformation parameter
μ	Mean of residuals
ξ	Skewness
σ	Standard deviation of residuals

Summary

Hydrological models are simplified, conceptual representations of the real watershed system. For improvement of the performance of hydrological model, the model parameters are often determined through calibration against the historical record data. In the automatic calibration procedure, the objective functions are used to ascertain the goodness-of-fit of hydrologic model. The results of model calibration depend heavily on the objective function. However, currently used objective functions are largely empirical. In hydrological models, the equifinality is a generic problem. Therefore, the Bayesian approach is widely proposed because it estimates not only the optimal value of model parameters, but also the probability distribution (i.e. uncertainty). Nevertheless, it is difficult to develop a formal likelihood function for Bayesian method.

The objective of this study is to evaluate the effect of objective functions on the model calibration. The primary goal is achieved through two parts: theoretical analysis and comparative study. In the theoretical analysis section, this study develops a formal likelihood function and builds a relationship between the likelihood function and the distance-based objective function. In the comparative study section, this study compares the results of model calibration using different objective functions in three cases: likelihood function comparison, multi-response calibration with the river discharge and sediment load objectives, and rainfall-runoff method comparison. The results are shown in the following.

The Nash–Sutcliffe efficiency coefficient (*NSE*) is proved to be equivalent to a kind of likelihood function with Gaussian, independent and identically distributed (I.I.D.) residuals. The hydrological model with *NSE* simulates high-values (e.g. flood) well, but low-values (e.g. baseflow) badly owing to the assumption of Gaussian error distribution, where the probability of the large error is low, but the small error around zero approximates equiprobability.

The scheme of Box-Cox transformation (BC) with minimum variance constraint is proven to be an effective method to estimate the BC parameter (lambda) for removal of the heteroscedasticity (i.e. inconstant variance) of model residuals. The BC-GED proposed by this study is a formal likelihood function, which assumes the model residuals after BC follow

the generalized error distribution (GED). The hydrological model with BC-GED mimics baseflow well, which is proved in the groundwater level simulation.

According to the field survey and previous related studies, the model parameters estimated by BC-GED are more reasonable than those estimated by *NSE*. The BC-GED objective function can also unify currently used distance-based objective functions (e.g. *NSE*), and reveals the mean absolute error (MAE) can best balance consideration of the high- and low- values. This study found that the volumetric efficiency (VE, i.e. a normalization of MAE) is appropriate for model comparison, because it always balances consideration of flood and baseflow, no matter what the model structure is.

In conclusion, the BC-GED and VE/MAE are more appropriate for automatic calibration than the most widely used *NSE* in this study. However, this conclusion needs to be validated in the other watersheds and other hydrological models in future.

1 Introduction

1.1 Problem Statement

Hydrologists have developed many models based on different theories and concepts, such as SWAT (Neitsch et al., 2005) based on the principle of the hydrologic response unit (HRU) and TOPMODEL (Beven and Kirkby, 1979) based on the topographic index (*TI*). However, because of the hydrologic complexity and especially the hydrologic heterogeneity, these models cannot describe the natural hydrologic processes entirely correctly, and their parameters can be interpreted only to the "effective parameters" which represent the integrated behavior at the model element scale. As it is difficult or impossible to determine the "effective parameters" directly from field measurement, the model parameters (mainly consisting of the coefficients and exponents of model equations) should be determined through calibration against the historical record data (Laloy et al., 2010).

The automatic calibration techniques are widely proposed to estimate the values of model parameters because of convenience and high efficiency (Willems, 2009). For ascertaining the goodness-of-fit/performance of hydrologic model in the automatic calibration program, hydrologists had proposed many statistical measures as efficiency criteria (i.e. objective functions) in the last decades instead of subjective visual judgment (Green and Stephenson, 1986; Dawson et al., 2007; Bennett et al., 2013). Through numerous practical applications of automatic calibration, it has been generally accepted that (1) the results of model calibration depend, to a great degree, on the objective function, and (2) different objective functions are in favor of different hydrographic components (Green and Stephenson, 1986; Legates and McCabe, 1999; Krause et al., 2005; Dawson et al., 2007). However, currently used objective functions are still largely empirical, which results in the difficulties of the objective function chosen. For example, many researchers found that the most widely used objective function — Nash-Sutcliffe efficiency (*NSE*) defined by Nash and Sutcliffe (1970) — puts greater emphasis on high-values/flood at the expense of the low-values/base-flow (Legates and McCabe, 1999; Krause et al., 2005; Guinot et al., 2011; Pushpalatha et al., 2012). But so far, there was no satisfactory explanation to the problem of *NSE*.

In the model calibration processes, owing to the lack of sufficient observation data and the inter-correlation of model parameters, equifinality of parameter sets must be expected instead of a single optimal parameter set (Beven, 2001; Beven and Freer, 2001). Additionally, errors in input data, model structure and measured outcomes are all lumped into a single additive

residual term, and then passed to the model parameters when calibrating the hydrological model. The parameter equifinality and errors, individually or combined, result in parameter uncertainty. So, the calibration of model parameters is being developed to include estimation of the probability distribution of parameters that represents the knowledge about parameter values (Yang et al., 2007b), and the Bayesian approach is popularly proposed (Beven and Binley, 2013).

In Bayesian approach, the likelihood function defined as the joint probability of model residuals is used to measure the deviations between model predictions and observed data. Generally, there are some statistic assumptions of likelihood functions, e.g. Gaussian, independent and identically distributed (I.I.D.) residuals. However, because of complex error-sources of model residuals in hydrology, such as the model input, structural and observed output errors, the model residuals often violate the assumptions of the likelihood function, suggesting that the likelihood function is informal. Therefore, how to develop a formal likelihood function to account for the correlation, heteroscedasticity (i.e. non-constant variance) and non-normality of model residuals is the key problem of the Bayesian method (Beven and Binley, 2013). On the other hand, the likelihood function can be regarded as a special objective function for model calibration. However, there are few studies of the relationship between the likelihood function and the classical objective function.

A success of automatic calibration depends heavily on not only the objective function, but also the model structure (Duan, 2003). The Soil and Water Assessment Tool (SWAT) is widely used to assess water resource and water-related problems all over the world (Gassman et al. 2007). However, there is a structural inadequacy of the surface-runoff generation methods in the SWAT: the widely used curve number method (CN) is an empirical method, which have some limitations in reflecting that soil moisture affects the surface runoff generation, but the performance of the physics-based method (Green & Ampt (G&A)) is often poor (Garen and Moore, 2005; Kannan et al., 2007; Gabellani et al., 2008; White et al., 2011; Han et al., 2012). Therefore, it is necessary to develop a new physics-based rainfall-runoff method to overcome the structural problem of the SWAT model and improve the comparability of the results of model calibration using different objective functions.

1.2   Research Questions and Study Aims

Based on the above analysis of objective functions for the automatic calibration, the following questions arise, which would be dealt with in this study:

1. Why does the *NSE* always put greater emphasis on high values at the expense of the low values?

2. How to account for the non-random components of model residuals?

3. What is the effect of the form of likelihood function on the results of model calibration? Are the results inferred by the formal likelihood function better or more reasonable than those inferred by the informal likelihood function?

4. What's the relationship between the likelihood function and the classical objective function?

5. How to improve the rainfall-runoff method in the SWAT?

Working on these questions, this dissertation will reveal the reason that the objective function emphasizes on specific hydrographic components (i.e. flood or baseflow), and provide guidelines of the objective function chosen. Note that this study only focuses on the most widely used distance-based objective functions.

The aim of this study is to evaluate the effect of objective functions on model calibration. The primary goal is achieved through the theoretical analysis (Section 4) and comparative study (Section 5) methods. The objectives of theoretical analysis are shown as follows:

1. Development of a physics-based rainfall-runoff method to overcome the structural inadequacy of surface-runoff generation methods in the SWAT.

2. Interpretation of the *NSE* from the likelihood function viewpoint.

3. Development of a formal likelihood function to account for the heteroscedasticity and non-normality of model residuals.

4. Extension of the likelihood function to include the river flow discharge and sediment load objectives for multi-response calibration.

The comparative study section includes three case studies of automatic calibration: likelihood function comparison, multi-response calibration, and model comparison. The objectives of the three case studies are separately shown as follows:

1. The case of likelihood function comparison will calibrate the hydrological model using different likelihood functions proposed by this study to evaluate the effect of the form

of likelihood functions on model calibration and discuss the advantages and disadvantages of the different likelihood functions.

2. The case of multi-response calibration will simultaneously calibrate the flow and sediment parameter using the multi-response likelihood functions proposed by this study to improve the performance of sediment simulation and discuss the effect of the formal/informal likelihood function on model calibration.

3. The case of model comparison will separately calibrate SWAT model with different rainfall-runoff methods using the same objective function to compare the performance of different rainfall-runoff methods and discuss the reliability of objective function for model comparison.

Finally, through comparison of the results of model calibration in the three case studies, this study attempts to find out the characteristics of objective functions and the corresponding reasons.

2 Current State of Knowledge

2.1 Classical Model Calibration

Hydrological models attempt to take advantage of digital computers to quantitatively describe the hydrologic processes that take place in a watershed, which are popularly used for a variety of applications, such as water resources assessment, flood forecast and hydrologic engineering design (Schaake, 2003). Generally, the parameters of hydrological model should be specified to guarantee that the hydrological model properly simulates the hydrological processes in watershed (Duan, 2003; Savenije, 2009). Sorooshian and Gupta (1995) pointed out that the model parameters can be classified into physical- and process- parameters. The physical parameters mainly consist of the values of the watershed characteristics, like watershed area, river length and slope of hillslope, which can be measured directly in the field. Usually, the values of physical parameters are extracted via the geographic information system (GIS) (Sorooshian and Gupta, 1995; Duan, 2003).

The process parameters mainly consist of the coefficients and exponents of model equations, such as the decay coefficient of baseflow, the mean soil hydraulic conductivity and the effective soil depth (Moradkhani and Sorooshian, 2009). The process parameters are difficult or impossible to be measured in the field, which should be inferred by indirect methods (Gupta et al., 1998; Moradkhani and Sorooshian, 2009). The performance of watershed model depends, to a great extent, on how quality estimate of these process parameters (Schaake, 2003). Therefore, in order to improve the performance of the hydrological models, these process parameters are usually tuned or calibrated until the model simulation results closely matched the observed results for some historical period where data have been collected (Gupta et al., 1998; Duan, 2003). This procedure of tuning model parameters is referred to as model calibration. In the specialized literature, model calibration is also termed as "parameter optimization/estimation", "inverse modelling", "inverse problem" or "history matching" (Zambrano-Bigiarini and Rojas, 2013). Thus, these terms are used interchangeably hereafter.

2.1.1 Classical Model Calibration Algorithm

The problem of hydrological model calibration has always troubled the hydrologists since the very beginning of the digital hydrological model era (Duan, 2003). The basic approach of

model calibration is the trial and error method: Firstly, the users choose a set of model parameters, then run hydrological model with model parameters to obtain model simulation results, and finally compare the model simulation results with the observed data. If the simulated results closely match the observed data, the parameter set will be accepted. Otherwise, the users change the values of model parameters according to their experiences and repeat the above steps, until obtaining a satisfactory parameter set. The above method is the so called manual-calibration. There are some merits of manual-calibration approach:

1. Its procedure is simple and easy to operate, and the users do not need the special computer-skill.
2. The users can judge the model simulation results, according to a lot of model performance measures accompanied by interactive graphic-user-interface (GUI) software, for avoiding the effect of noises in calibration data (Biondi et al., 2012).

However, the manual-calibration methods are cumbersome and labor-intensive (time-consuming) because of inefficiency of model parameter optimization. And the success of the manual calibration depends, to a great degree, on the experience and knowledge level of the users (Muleta and Nicklow, 2005). Therefore, the automatic calibration method is widely purposed (Schaake, 2003; Duan, 2003; Moradkhani and Sorooshian, 2009).

The automatic calibration approach, just as the name implies, automatically generates the candidate model parameters and judges the goodness-of-fit/performance of the hydrological model according to the model performance measures defined by objective functions (Duan, 2003; Moradkhani and Sorooshian, 2009). The automatic approach can take full advantage of digital computers and free the users from the bored model-calibration. The commonly processes of the automatic calibration at each iteration are shown in Figure 2.1:

1. Generating a candidate parameter set based on the value of objective function and state parameter set (i.e. the optimal parameter set at the moment) by optimization tool/algorithm;
2. Driving the hydrological model with the candidate parameter set to obtain the model predictions;
3. Estimating the value of objective function for measurement of the deviations between the model predictions and observed data;
4. Deciding whether the program stops or continues according to the new value of objective function.

Generally, if the value of objective function is less than the tolerance given by users, the program accepts the candidate parameter set as the optimal values and stops searching. Otherwise, the optimization tool/algorithm will continue to search the optima of model parameters.
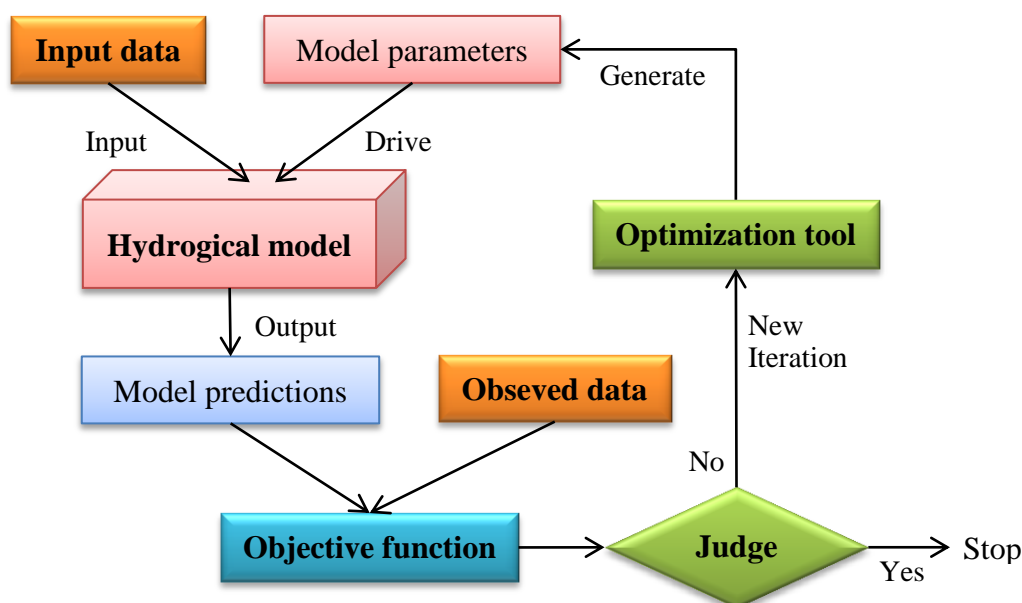
Figure 2.1 Schematic of the automatic calibration procedure. The objective function is used to ascertain the goodness-of-fit/performance of hydrologic models. The "Judge" is based on the value of objective function. The optimization tool is used to generate the candidate parameter set. In general, the observed data are the river discharges.

Figure 2.1 shows the automatic calibration program generally consists of four parts: optimization tool and objective function, hydrological model and observed data (Duan, 2003). In the last decades, the optimization method/algorithm was always the research hotspot. Researchers proposed a lot of optimization algorithms (Duan, 2003; Zambrano-Bigiarini and Rojas, 2013), such as genetic algorithm (GA), particle swarm optimization (PSO) and Shuffled Complex Evolution (SCE-UA). In general, these optimization algorithms can be classified into local- and global- search optimization methods (Duan, 2003). The local search methods (e.g. the large family of Newton and Quasi-Newton methods) generate the candidate parameter set around the initial values given by users, of which the advantage is the high convergence rate. However, the calibration result of the local search algorithm is easily trapped around secondary optima, which is so-called mis-calibration (Andreassian et al., 2012). The surface reason is the non-smooth (roughness) and discontinuity of the objective

function response surface. The reason behind it is that: (1) all hydrological models are process-based or conceptual, rather than purely physically based; and (2) the model parameters are non-unique and non-identifiable because of inter-correlation and lacking physical meaning (Beven and Binley, 1992; Montanari and Koutsoyiannis, 2012).

By contrast, the global search methods can randomly generate the candidate parameter set and get rid of the impact of initial values. Among these optimization methods, the SCE-UA method developed by Duan et al. (1992) is most widely known, which combines the global search method (Genetic Algorithm) with the local search method (Simplex method). The SCE-UA was a highly effective, efficient and robust global search method proved by hundreds of applications (Duan, 2003; Khakbaz et al., 2012). Schaake (2003) even optimistically indicated that since the birth of the SCE-UA, the optimization part of model calibration problem has been no longer a major limiting factor. However, it may be over-optimistic as the convergence rate (time-consuming) of SCE-UA is still an outstanding issue (Zambrano-Bigiarini and Rojas, 2013). Therefore, the new or improved global optimization method is often released, such as dynamically dimensioned search algorithm (DDS; Tolson et al., 2007) and hydroPSO (Zambrano-Bigiarini and Rojas, 2013). In general, however, the problem of mis-calibration is basically solved for hydrological model today (Andreassian et al., 2012).

As the global search methods are widely used, the problem of over-calibration is more apparent (Andreassian et al., 2012). The overcalibration refers to that the mathematically optimum parameter set over the calibration period does not remain mathematically optimum over different periods (Andreassian et al., 2012). The miscalibration is essentially a numerical problem. By contrast, the overcalibration is a purely hydrological problem. It is mainly caused by over-parameterization (i.e. an excessive number of (redundant) parameters) possibly because of the lack of sufficient information, e.g. observed data (Andreassian et al., 2012; Zambrano-Bigiarini and Rojas, 2013). Changing the objective function used in calibration is a possible solution to avoid overcalibration, because different objective functions are in favor of different hydrographic components (e.g. flood or baseflow) (Green and Stephenson, 1986; Legates and McCabe, 1999; Krause et al., 2005; Dawson et al., 2007; Andreassian et al., 2012).

2.1.2 Classical Objective Function

The objective function is used to evaluate the hydrological model simulation results (Figure 2.1). In contrast with the manual calibration approach, the automatic calibration approach inappropriately adopts the visual inspection of similarities and differences between the model simulations and observations. So, hydrologists proposed many statistical measures as efficiency criteria (i.e. objective functions) to ascertain the goodness-of-fit of hydrologic models in the automatic calibration program in the last decades (Green and Stephenson, 1986; Dawson et al., 2007; Bennett et al., 2013). These objective functions can be generally classified into two types (Table 2.1): distance-based objective function (also called absolute error measures by Legates and McCabe (1999), or residual methods by Dawson et al. (2007) and Bennett et al. (2013)) and weak form-based objective function (Guinot et al., 2011; Bennett et al., 2013).

The weak form-based objective function, such as the coefficient of determination (R^2, Table 2.1, ID 1) and the volume/cumulative error (CE, Table 2.1, ID 2), is used to estimate the statistical properties of model residuals (i.e. deviations) between the model predictions and observed data (Guinot et al., 2011). The distance-based objective function, such as mean squared error (MSE, Table 2.1, ID 3) and mean absolute error (MAE, Table 2.1, ID 4), is defined as the "distance" (similar to the spatial distance) between model predictions and observed data (Figure 2.1), which is most widely proposed in the model calibration. There are many variations of distance-based objective functions for different purposes. For example, in order to emphasize on special runoff components (e.g. flood or baseflow), the distances between the model predictions and observed data are often multiplied with user-defined weights in the distance-based objective function (Green and Stephenson, 1986). In order to balance consideration of flood and baseflow, the relative error is often used instead of the original error/residual, which is the ratio between model residual and corresponding observed data (Bennett et al., 2013; Wu and Liu, 2014). And for comparison of the performance of different hydrological models, the normalizations of distance-based objective functions are proposed, e.g. the Nash-Sutcliffe efficiency (*NSE*; Table 2.1, ID 9, in the case of j=2) defined by Nash and Sutcliffe (1970) that is the normalization of MSE (Table 2.1, ID 3).

Table 2.1 Classical objective functions.

ID	Categories	Objective functions	Formula [1]	Characteristics	References
1	Weak form-based objective function	Coefficient of Determination (R^2)	$\dfrac{\left(\sum_1^n (obs_i - \overline{obs})(sim_i - \overline{sim})\right)^2}{\sum_1^n (obs_i - \overline{obs})^2 \sum_1^n (sim_i - \overline{sim})^2}$	Inappropriate for model-performance assessment; emphasize on high flows	Legates and McCabe (1999)
2		Volume/cumulative error (CE)	$\sum_1^n (obs_i - sim_i)$	Monotony; cannot be used alone	Guinot et al. (2011)
3	Distance-based objective function	Mean squared error (MSE)	$\dfrac{1}{n}\sum_1^n (obs_i - sim_i)^2$	Most common; emphasize on high flows; neglect the low flows	McCuen et al. (2006); Krause et al. (2005)
4		Mean absolute error (MAE)	$\dfrac{1}{n}\sum_1^n |obs_i - sim_i|$	Balance consideration of the high- and low- flows	Legates and McCabe (1999); Krause et al. (2005)
5		Mean quadrupled error (MS4E)	$\dfrac{1}{n}\sum_1^n (obs_i - sim_i)^4$	Put greater emphasis on high flows	Baratti et al. (2003); Bennett et al. (2013)
6		Square-root transformed MSE (RTMSE)	$\dfrac{1}{n}\sum_1^n (\sqrt{obs_i} - \sqrt{sim_i})^2$	Put equal emphasis on high- and low- flows; focus on mean flows	Oudin et al. (2006); Pushpalatha et al. (2012)
7		Log transformed MSE (LTMSE)	$\dfrac{1}{n}\sum_1^n (\log(obs_i) - \log(sim_i))^2$	Put greater emphasis on low flows but still show sensitivity to high flows	Oudin et al. (2006); Pushpalatha et al. (2012)
8		Inverse transformed MSE (ITMSE)	$\dfrac{1}{n}\sum_1^n \left(\dfrac{1}{obs_i} - \dfrac{1}{sim_i}\right)^2$	Focus on low flows totally	Pushpalatha et al. (2012)
9		Generalized efficiency (E_j)	$1 - \dfrac{\sum_1^n |obs_i - sim_i|^j}{\sum_1^n |obs_i - \overline{obs}|^j}$	Modelers define the power (j)	Legates and McCabe (1999)

[1] where obs_i and sim_i are the observed and the simulated data at time step i, respectively, $\overline{obs}$ and $\overline{sim}$ are the mean observed and simulated data, respectively, n is the length of data, j is the error power, $\log(x)$ is the logarithm function evaluated at x, Σ is the summation operator.

2.1.2.1 Characteristics of Objective Functions

Many researches discussed the characteristics of different objective functions (Table 2.1). Legates and McCabe (1999) pointed out that the coefficient of determination (R^2) is inappropriate for model quantification because it is oversensitive to high flow but insensitive to additive and proportional differences between model simulations and observations. They recommended the distance-based objective functions (e.g. MSE) as the model evaluation tools. Among all objective functions, the Nash-Sutcliffe efficiency (*NSE*) is most widely used (Beven and Binley, 2013). The *NSE* is calculated by subtracting the ratio between the MSE and the variance of the observations from one, thus the *NSE* is dimensionless and ranges from minus infinity to one in theory (Gupta et al., 2009). If the *NSE* is one, the hydrologic model is perfect. However, if the *NSE* is less than zero, the hydrologic model is not better than a predictor using the mean of the observations (Seibert, 2001; Schaefli and Gupta, 2007; Gupta et al., 2009). It is commonly accepted that if the *NSE* is greater than 0.65, the hydrologic model is acceptable. Otherwise, it is unsatisfactory (Moriasi et al., 2007; Ritter and Muñoz-Carpena, 2013).

The NSE/MSE can be decomposed into three distinctive components: the correlation (coefficient of determination, R^2), the conditional bias and the unconditional bias (i.e. the standardized bias between model predictions and observations) (Gupta et al., 2009). Namely, the NSE/MSE combines R^2 with the standardized bias and overcomes the deficiency of R^2, i.e. insensitive on the model bias, so the NSE/MSE is appropriate to assess the goodness-of-fit of hydrological models. However, the NSE/MSE also inherits some deficiencies of R^2, e.g. the NSE/MSE also puts greater emphasis on the high values, whereas it neglects the low values (Legates and McCabe, 1999; Krause et al., 2005; McCuen et al., 2006).

Legates and McCabe (1999) attributed the drawbacks of the NSE/MSE to square the differences between model predictions and observations, so they proposed a new objective function, i.e. generalized efficiency (Table 2.1, ID 9) where the users can choose the power of differences/errors by themselves. As an example, Legates and McCabe (1999) set the power (*j*) of differences/errors as one (i.e. *j*=1), and then the generalized efficiency (Table 2.1, ID 9) becomes a normalization of the mean absolute error (MAE). They found the MAE is slightly preferred over NSE/MSE when extreme values are present. Krause et al. (2005) pointed out the MAE can balance consideration of the high- and low- flows. Wu and Liu (2014) also indicated the MAE emphasizes neither high- nor low- values. Criss and Winston (2008) further proposed another normalization of MAE, i.e. 'volumetric' efficiency (VE) that is

calculated by subtracting the ratio between the MAE and the mean of the observations from one, to replace *NSE*. Similar to the *NSE*, the VE is dimensionless and ranges from minus infinite to one. However, Bennett et al. (2013) pointed out the MAE is worse than MSE, because MAE has a kink (i.e. no smoothness) at zero value, which may produce a non-smooth operator when used in optimization, whereas the MSE is a smooth function of model residuals.

Although the explanation of the deficiencies of *NSE* proposed by Legates and McCabe (1999) provides a good guideline to choose the distance-based objective function, it still cannot explain why the R^2 (Table 2.1, ID 1) also puts emphasis on the high values and the *NSE* always neglects the low values (Krause et al. 2005).

Another method to overcome the deficiency of the *NSE* is the transformation of model residuals. Vandewiele et al. (1992) and Xu et al. (1996) concluded that the square root transformation (Table 2.1, ID 6) focuses on mean flow. Oudin (2006) recommended square root transformation (Table 2.1, ID 6) and log transformation (Table 2.1, ID 7) to emphasize low flows. However, Pushpalatha et al. (2012) pointed out the square root transformation puts equal emphasis on high- and low- flows, logarithmic transformation puts greater emphasis on low flows, but still shows sensitivity to high-flows, and further proposed the inverse transformation (Table 2.1, ID 8) which totally focuses on low-flows.

Generally, the weak form-based objective functions cannot be used as a single criterion for model calibration (Legates and McCabe, 1999; Guinot et al., 2011). Guinot et al. (2011) compared the weak form-based objective functions with the distance-based objective functions, and concluded that although the distance-based objective functions have the advantage to search an identifiable model-parameter set, they may cause the local extremes in the response surface and lead to mis-calibration i.e. being trapped around secondary optima. By contrast, the weak form-based objective functions (e.g. CE (Table 2.1, ID 2)) are more monotone than the distance-based objective functions.

2.1.2.2 Multi-Objective Approaches

After comparison of many objective functions, almost all hydrologists drew the same pessimistic conclusion: objective function ultimately chosen should depend on the objective of the modeling exercise, because different objective functions are in favor of different hydrographic components (Green and Stephenson, 1986; Legates and McCabe, 1999; Krause

et al., 2005; Dawson et al., 2007). In other words, we would better try the objective functions one by one or combine them using multi-objective optimization scheme when we calibrate our hydrologic model, because no one knows which objective function will be appropriate to our study, and a single goodness-of-fit measure (objective function) may be inappropriate (Legates and McCabe, 1999).

A very good example of the multi-objective approaches is that the NSE/MSE as a most widely used objective function can be decomposed into three weak form-based objective functions (Gupta et al., 2009). It results in that the multi-objective approaches were widely proposed (Gupta et al., 1998; Gupta et al., 2009). The multi-objective approaches can be formulated on the basis of the three types of information (Efstratiadis and Koutsoyiannis, 2010):

1. Multi-response sources that are the different observed fluxes, such as river discharges, sediment loads and groundwater levels.

2. Multi-site data that are the same kind of data but observed from a number of gauges within a watershed, e.g. the river discharges are observed in the different tributaries.

3. Multi-objective functions that are independent objective functions accounting for various aspects of a single hydrological process, typically river discharge process.

Generally, the multi-objective approach refers in particular to the optimization procedure with the multi-objective function, i.e. the third case of information (Efstratiadis and Koutsoyiannis, 2010).

Although some researchers have pointed out the importance of the multi-objective optimization exercise (Gupta et al., 1998; Andréassian et al., 2012), the single-objective optimization procedure is still the mainstream in the practical applications (Guinot et al., 2011), because of some drawbacks of the multi-objective optimization:

1. The multi-objective optimization may cause the serious discontinuity of the response surface and result in the trouble of model calibration and more time-consuming of the optimization procedure (Madsen, 2000; Parajka et al., 2007).

2. With the "Pareto front" strategy, the multi-objective approaches transform the model-parameter optimization into the decision-making based on many candidate parameter sets (Schaake, 2003; Reed et al., 2013), which increases the subjectivity of the model calibration.

3. The theoretical foundation of the multi-objective calibration is that there are significant trade-offs among different objective functions (Efstratiadis & Koutsoyiannis, 2010). However, Kollat et al. (2012) pointed out the meaningful multi-objective trade-offs are less after application of a four-objective calibration strategy in the 392 watersheds across the United States, and only when there is obvious hydrological model deficiencies, the meaningful precision of trade-offs do exist.

In short, currently used objective functions are largely empirical. There is too much messy information about objective functions, which results in the difficulties of the objective function chosen. Therefore, it is necessary to find a theory to unify these objective functions for facilitating their utilization.

2.2 Bayesian Inference for Model Calibration and Uncertainty Analysis

The efficiency of automatic calibration depends heavily on the optimization method, the objective function, the model structure and the observed data (Figure 2.1; Duan, 2003). In practices, numerous studies found there are many equivalent/similar model predictions with different model parameter sets that are difficult to distinguish by the values of objective function, which is the so-called equifinality (Beven and Binley, 1992; Beven, 1993). In the context of optimization problem, it is the problem of non-uniqueness or non-identifiability of model parameters, because of the multiple optima in the response surface (Montanari and Koutsoyiannis, 2012). In essence, it results from the inevitable deficiencies of the hydrological models and the inter-correlation of the model parameters. Therefore, the equifinality is a generic problem and inevitable in the hydrological models (Schaake, 2003; Beven and Binley, 2013). In order to account for the equifinality of hydrological models and parameter sets, Beven and Binley (1992) proposed to estimate the uncertainty distribution of model parameters instead of only searching the optimum parameter set. The Bayesian approach is widely adopted to estimate the posterior distributions of model parameters (Beven and Binley, 2013).

2.2.1 Bayesian Approach

The Bayesian theorem describes the relationship between the conditional distribution for model parameter set (θ) given observed outcomes (*obs*) and the joint distribution of θ and *obs*. The Bayesian theorem states:

$$\pi\left(\theta|obs\right) = \frac{\pi\left(\theta\right)f\left(obs|\theta\right)}{\int \pi\left(\theta\right)f\left(obs|\theta\right)d\theta} \propto \pi\left(\theta\right)f\left(obs|\theta\right) \qquad (2.2\text{:}1)$$

where

θ is the model parameter set,

obs is the observed outcomes,

$\pi(\theta)$ is the set of prior distributions for parameter set (θ),

$f(obs|\theta)$ is the likelihood function,

$\int \pi\left(\theta\right)f\left(obs|\theta\right)d\theta$ is the marginal likelihood (usually sets to unknown constant),

$\pi(\theta|obs)$ is the posterior distribution for θ given obs.

The Bayesian approach tries to separate the observations (e.g. river discharges) into two parts: a deterministic component and a random component describing residuals (Schoups and Vrugt, 2010). The deterministic component is determined by the hydrologic model. The joint probability of the random component, i.e. residuals/errors between observations and simulations generated by hydrologic model with a particular parameter set, is estimated by a likelihood function. By augmenting the likelihood function with prior knowledge of model parameters, the posterior distribution of model parameters is estimated.

However, it is difficult to obtain the analytical solution (even the numerical solution) for the posterior distribution (Eq. (2.2:1)) because of the unknown formulation of hydrologic models and too many model parameters (Yustres et al., 2012). The Markov Chain Monte Carlo (MCMC) scheme as a stochastic simulation approach provides a simple and effective way around the computational difficulties for the posterior distribution. The aim of MCMC scheme is to generate samples of the parameter set based on constructing a Markov chain that has the posterior distribution as its equilibrium distribution (Marshall et al., 2005).

The Differential Evolution Adaptive Metropolis (DREAM) Markov chain Monte Carlo (MCMC) sampling is superior to other adaptive MCMC sampling approaches in the presence of high-dimensionality and multimodality optimization problems, because DREAM scheme follows up on the Shuffled Complex Evolution Metropolis global optimization algorithm (SCEM-UA that replaces the Simplex method in the SCE-UA global optimization algorithm by the Simulated Annealing (SA) method (Vrugt et al., 2003)), runs multiple different chains simultaneously for global exploration, and maintains detailed balance and ergodicity (Vrugt

et al., 2009b). The flow chart of the DREAM scheme is shown in Figure 2.2. Compared with the Figure 2.1, the Figure 2.2 adds a database to store the state parameter set.
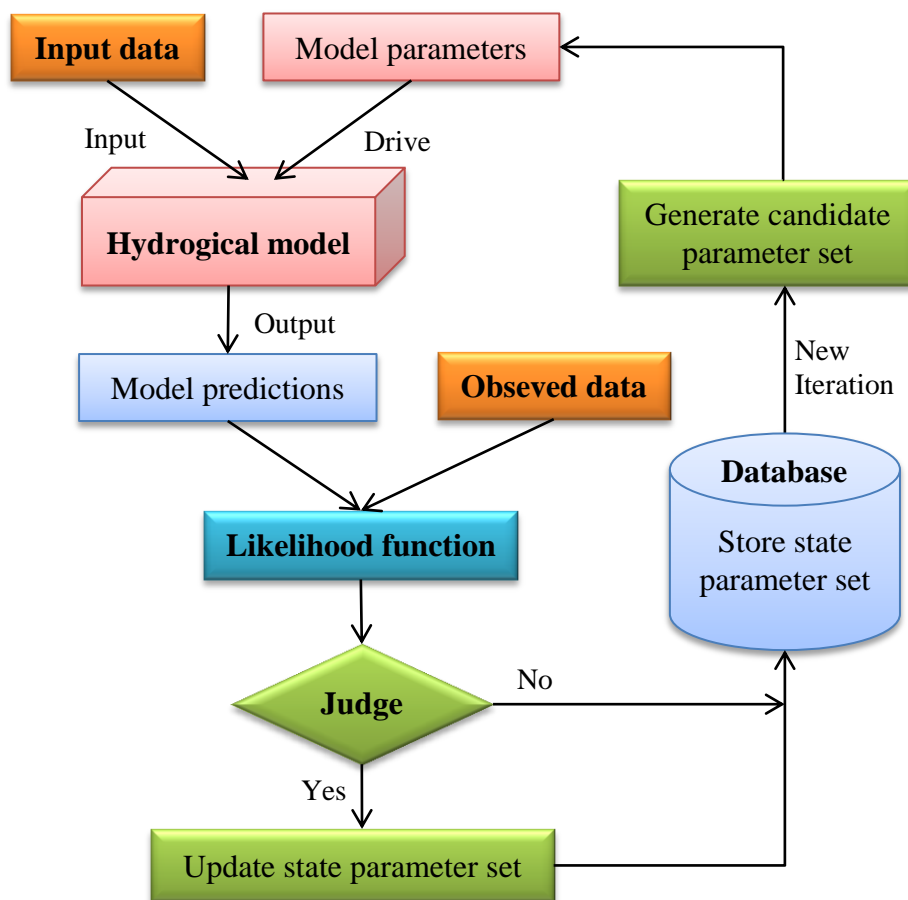
Figure 2.2 Flow chart of the stochastic simulation approach — DREAM (MCMC). The likelihood function also termed error model is used to estimate the joint probability of model residuals. The "Judge" is based on the value of likelihood function. The Database only stores the state parameter set at each iteration, where the probability distribution of model parameter is the corresponding posterior parameter distribution. The candidate parameter set is generated based on the value of objective function and state parameter set.

The main procedures of DREAM at each MCMC iteration are as follows:

1. Generating randomly a candidate parameter set based on the value of objective function and state parameter set;

2. Driving the hydrological model with the candidate parameter set to obtain model predictions;

3. Estimating the value of likelihood function, i.e. the joint probability of the model residuals between the model predictions and observed data;

4. Deciding whether the candidate parameter set replaces the state parameter set or not, based on the value of likelihood function;

5. Storing the state parameter set into database.

The probability distribution of model parameter stored in Database is the corresponding posterior parameter distribution. More MCMC iterations will lead to better convergence of the posterior parameter distribution (Andrieu et al., 2003). As an example for calculation of the posterior parameter distribution, Figure 2.3 (a) shows the value of model parameter stored in database at each MCMC iteration (Figure 2.2), and Figure 2.3 (b) estimates the probability distribution of model parameter by histogram method. In order to avoid the effect of noises in Markov chain, the kernel density estimation is popularly used to smooth the histogram of model parameter (red line in Figure 2.3 (b)). The sharper the shape (or the narrower of the range) of posterior parameter distribution, the more sensitive the model parameter is.

(a) Markov chain of model parameter   (b) Posterior parameter distribution
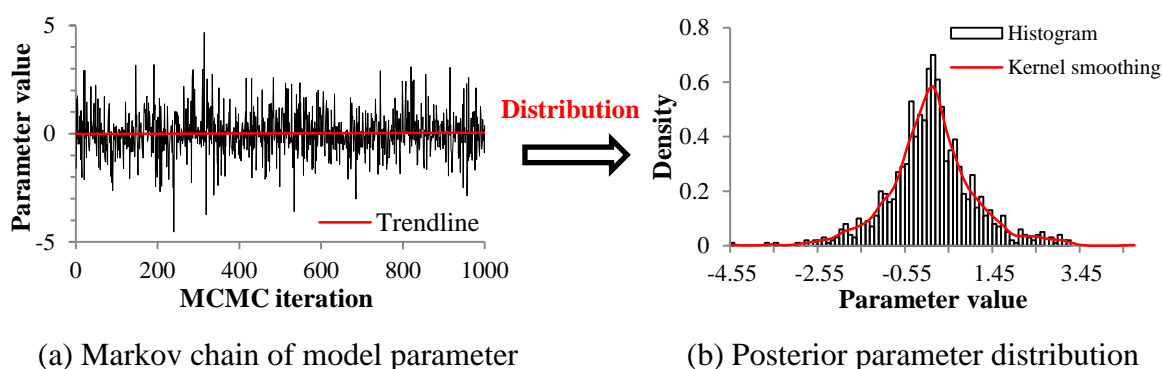
Figure 2.3 Schematic of the posterior parameter distribution. The kernel smoothing is used to reduce the effect of noises in Markov chain.

Recently, DREAM has become a generally known tool for model calibration and uncertainty analysis. For example, McMillan and Clark (2009) used a modified *NSE* as an informal likelihood function to calibrate model parameters. Laloy et al. (2010) used the DREAM for model parameter optimization and uncertainty analysis. Schoups and Vrugt (2010) adopted the DREAM to analyze the uncertainty of the model parameters and predictions. Smith et al. (2010) studied the effect of different likelihood functions on model calibration in the ephemeral watershed by DREAM. Montanari and Koutsoyiannis (2012) used the DREAM with a standard Gaussian likelihood function to calibrate hydrological model.

In Bayesian inference (Eq. (2.2:1)), the prior distributions $\pi(\theta)$ (i.e. prior knowledge) of parameter set are defined by the users to reduce the uncertainty of model parameters. As a result, many kinds of prior distributions are used in Bayesian approach. For example, Reichert and Shuwirth (2012) assumed the prior parameter distribution of hydrological model is the lognormal distribution but that of error model (i.e. likelihood function; Figure 2.2) is the exponential distribution. By contrast, Gelman et al. (2004) suggested the Gamma distribution as the prior distribution of the standard deviation of model residuals. Sikorska et al. (2012) proposed the Gamma distribution for the parameters of error model but the normal and lognormal distribution for the parameters of hydrological model. Therefore, the prior parameter distributions lead to serious subjectivity of Bayesian method. In order to avoid subjective judgment, the uniform distribution as a non-informative prior distribution is widely used as the prior parameter distribution in hydrology, which means that we do not have the prior knowledge about model parameters (Vrugt et al., 2009a and b; Yang et al., 2007a and b). The default of prior parameter distribution in the DREAM is the uniform distribution.

2.2.2 Likelihood Function

The likelihood function of a set of model parameter values (θ) given observed outcomes (*obs*) is equal to the joint probability of the observed outcomes given the parameter values. To put it simply, the likelihood function is the joint probability of model residuals between model predictions and observed data. Similar to the objective function, the likelihood function can also measure the deviations between model predictions and observed data. The maximum likelihood function is often used to estimate the parameters of statistical model. The model-residual/error (e_i) is defined as the difference between the observed and simulated data (Figure 2.4).

If the model residual (e_i) randomly distributes around the observed data according to the probability density function (PDF) of $p(e)$, then the probability density of model residual (e_i) is $p(e_i)$ (Figure 2.4). Further, if the model residuals (e.g. e_{i-1} and e_i; Figure 2.4) are independent and identically distributed (I.I.D.), then the joint probability of the model residuals is the product of error probabilities:

$$P = \prod_{i=1}^{n} p(e_i) \qquad (2.2:2)$$

where

e_i is the model residuals/errors at the time step of i,

n is the length of model residuals,

p(x) is the probability density function of model residuals evaluated at x,

∏ is the product operator.

Usually, we used the logarithmic value of the joint probability (*P*):

$$l = \ln(P) = \ln(\prod_{i=1}^{n} p(e_i)) = \sum_{i=1}^{n} \ln(p(e_i)) \qquad (2.2:3)$$

where

P is the joint probability of the model residuals,

ln(x) is the logarithm function evaluated at x,

Σ is the summation operator.
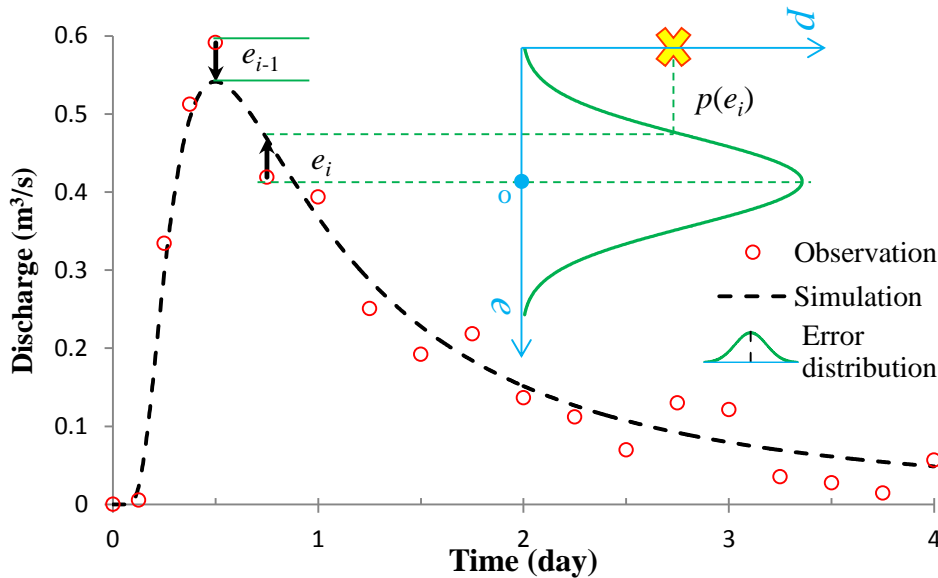
Figure 2.4 Schematic of the likelihood function with I.I.D. residuals. e_i is the model residual between observation and simulation and randomly distributes around the observed data (red circle) according to the error distribution $p(e)$ (green curve line). $p(e_i)$ is the probability density of e_i.

The Eq. (2.2:3) is the likelihood function of the hydrological model parameters. One should keep in mind that in Eq. (2.2:3), the model residuals (e_i) are independent and identically distributed (I.I.D.) according to the error distribution $p(e)$ (usually assumed Gaussian distribution). If the model residuals (e) fulfill these statistical assumptions of the likelihood function, the likelihood function (Eq. (2.2:3)) is termed as the formal likelihood function. Otherwise, it is an informal likelihood function. As an example, Figure 2.5 (a) shows the independent and identically distributed residuals, and Figure 2.5 (b) inspects the Gaussian

error assumption of the likelihood function via plotting the histogram of model residuals versus the assumed error distribution (e.g. Gaussian distribution) with the optimized parameters.

(a) Series of model residuals (b) Test of the distribution of residuals
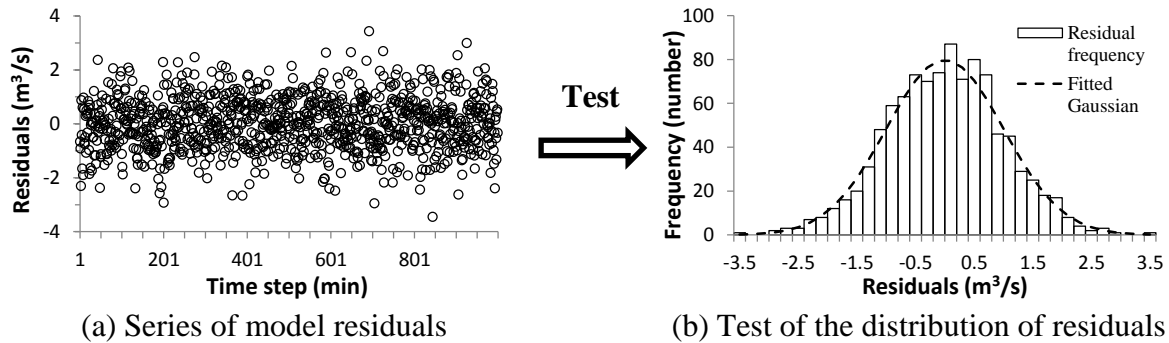
Figure 2.5 Inspection of the statistical assumptions of likelihood function.

The likelihood function with the Gaussian I.I.D. error assumption (Figure 2.5 (b)) is widely used in the Bayesian analysis (Feyen et al., 2007; Yustres et al., 2012; Beven et al., 2012). However, in most cases, the model residuals do not fulfill the Gaussian I.I.D. error assumption (Schoups and Vrugt, 2010; Beven et al., 2012). Beven et al. (2012) pointed out that the model residuals include not only the aleatory/random component but also epistemic/cognitive component that is from the model structural and input errors. The epistemic errors result in the correlated, heteroscedastic and non-Gaussian residuals. Auto-correlation means the model residuals are not independent. And heteroscedasticity (i.e. inconstant variance) means the model residuals are not from the same error distribution. Therefore, the correlation and heteroscedasticity of model residuals completely violate the I.I.D. error assumption of the likelihood function (Eq. (2.2:3)).

However, Evin et al. (2013) indicated that the heteroscedasticity of model residuals is associated with larger rainfalls and streamflow, and the correlation is related to the "memory" of hydrological model. The difference between the viewpoint of Beven et al. (2012) and Evin et al. (2013) is whether we can develop a formal likelihood function to account for the correlated, heteroscedastic and non-Gaussian residuals. Beven et al. (2012) believed it is difficult or impossible to find the formal likelihood function because of the inherent defectiveness of human being. But Evin et al. (2013) attempted to develop a method to account for the heteroscedasticity and correlation of model residuals.

2.2.2.1 Formal Likelihood Function

Many researchers proposed kinds of formal likelihood functions (also termed error models) to account for the correlation, heteroscedasticity and non-normality of model residuals (Table 2.2). For example, Bates and Campbell (2001) used the autoregressive model (AR) and the Box-Cox transformation method (BC) to remove errors' correlation and heteroscedasticity, respectively. Yang et al. (2007a) proposed a modified first-order autoregressive (AR(1)) model (termed additive continuous time autoregressive model) that can deal with the discrete time series, and combined the new autoregressive model with the BC method. Schoups and Vrugt (2010) proposed a generalized error model that firstly removes the errors' correlation by the autoregressive model, and then estimates the error standard deviation (i.e. heteroscedasticity) by a linear function of simulated streamflow, and finally used the skewed generalized error distribution (SGED) to fit the distribution of model residuals. However, Evin et al. (2013) pointed out that applying autoregressive model to raw heteroscedastic residuals could result in unstable error models with poor predictive performance. Therefore, they proposed a new error model that applies the autoregressive model after removing the heteroscedasticity of residuals. Due to many zero inflations in the ephemeral rivers, Smith et al. (2010) developed a formal likelihood function that classifies model residuals into zero error component and non-zero error component to improve the Bayesian inference. For simplifying the error model, Pianosi and Raso (2012) ignored the autocorrelation of the model residuals but allowed for the residual variance to change in time.

In short, almost all the error models (likelihood formulation) separately account for the error's correlation, heteroscedasticity and non-normality (Table 2.2). The first-order autoregressive (AR(1)) scheme and the Box-Cox transformation method (BC) are popularly used to remove errors' correlation and heteroscedasticity, respectively (Bates and Campbell, 2001; Yang et al., 2007a, b; Vrugt et al., 2009a; Laloy et al., 2010; Li et al., 2011).

The Box-Cox transformation method needs to estimate transformation parameter (λ). Most studies (Vrugt et al., 2009a; Engeland et al., 2010; Li et al., 2011) fixed the value of λ (lambda), and some others (Yang et al., 2007a, b; Laloy et al., 2010) treat λ as an inference parameter (Table 2.2). Obviously, it is more effective to remove the errors' heteroscedasticity when λ varied as simulated outcomes. Unfortunately, almost all inference results touch the boundary of λ ($0 \leq \lambda \leq 1$), such as the result (λ) of Yang et al. (2007b) approaches to zero, and λ of Laloy et al. (2010) approaches to one. The boundary value means the extreme situation, e.g. when $\lambda = 1$, the BC is ineffective, i.e. no transformation of model residuals, and

the BC becomes the log transformation when $\lambda = 0$, although it rarely occurs. Therefore, it is necessary to build a new efficient method to estimate the transformation parameter (λ).

Although Gaussian distribution is widely used as the probability distribution of the residuals, there are many cases of non-Gaussian errors in fact (Thiemann et al., 2001; Yang, 2007a, b; Schoups and Vrugt, 2010; Li et al., 2013). The generalized error distribution (GED) (Thiemann et al., 2001; McMillan and Clark, 2009) and the skew generalized error distribution (SGED) (Schoups and Vrugt, 2010) are more flexible and could replace the Gaussian error distribution (Table 2.2).

Table 2.2 Scheme of the currently used formal likelihood functions in the Hydrologic Literature to account for the correlation, heteroscedasticity and non-normality of model residuals.

Likelihood Reference	Correlation [1]	Heteroscedasticity [2]	Non-normality [3]
Bates and Campbell (2001)	AR(p)	BC (lambda=0.0)	—
Thiemann et al. (2001)	—	BC (lambda=0.5)	GED
Yang et al. (2007a)	AR(1)	BC (lambda estimated)	—
Yang et al. (2007b)	AR(1)	BC (lambda=0.0)	Student(t-) distribution
Vrugt et al. (2009a)	AR(1)	BC (lambda=0.3)	—
Schoups and Vrugt (2010)	AR(p)	Linear function of simulated streamflow [4]	SGED
Smith et al. (2010)	—	BC (lambda=0.0)	—
Laloy et al. (2010)	AR(1)	BC (lambda=1.0)	—
Engeland et al. (2010)	AR(1)	BC (lambda=0.2)	Normal Quantile Transformation
Li et al. (2011)	AR(1)	BC (lambda=0.6)	—
Pianosi and Raso (2012)	—	Non-linear function of time [5]	—
Evin et al. (2013)	AR(1)	Linear function of simulated streamflow	—

[1] AR(p) is the pth order autoregressive model; AR(1) is the first-order autoregressive model.

[2] BC is the Box-Cox transformation method; lambda (λ) is the parameter of BC.

[3] GED is the generalized error distribution; SGED is the skew generalized error distribution. The dash line ("—") means that the corresponding component is neglected by the authors.

[4] The standard deviation of errors is modeled by a linear function of simulated streamflow.

[5] The standard deviation of errors is modeled by a non-linear function of time.
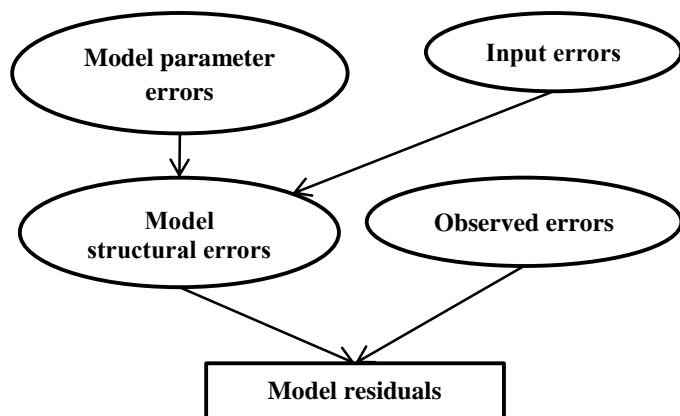
Figure 2.6 Relationship among different error components of model residuals.

The above error models are still limited to Bayesian total error analysis. Reichert and Mieleitner (2009) pointed out that the model residuals come from three parts: the model input error, the model parameter error and the observed output error. The relationship among these error components is shown in Figure 2.6. This figure shows that the model parameter errors and input errors are all lumped into the model structural errors. Reichert and Schuwirth (2012) proposed an error model to separate the model structural error and the observed error. Their error model revealed that the correlated components of model residuals are mainly from model structural errors, and the observed errors only contribute the heteroscedastic components. Vrugt et al., (2009a) also pointed out the effect of the autoregressive scheme (AR) in the error model is to account for the model structural error (Table 2.2). Bennett et al. (2013) indicated that the significant autocorrelation of model residuals demonstrate un-modelled behavior, i.e. the hydrological model is wrong. And the study of Evin et al. (2014) shows the likelihood function with AR(1) model for removing the autocorrelation of model residuals has disadvantages on the estimated parameter uncertainty in some cases. Therefore, Thiemann et al. (2001), Smith et al. (2010) and Pianosi and Raso (2012) all neglected the autocorrelation of the model residuals in error models (Table 2.2), which assumes that there is no significant inadequacy of model structure.

2.2.2.2 Argument between the Formal and Informal Likelihood Function

Although numerous authors proposed many error models, these models are still based on assumptions and restrictive in some practical applications (Montanari and Koutsoyiannis, 2012). Recently, some hydrologists recognized that (1) it is difficult to find the strictly formal likelihood function, and (2) the model residuals always violated the statistical assumption, e.g. the I.I.D. error assumption (Montanari and Koutsoyiannis, 2012; Beven and Binley, 2013; Vrugt and Sadegh, 2013; Sadegh and Vrugt, 2013). However, the Bayesian approach for the model uncertainty analysis requires the formal likelihood functions (Beven and Binley, 2013;

Vrugt and Sadegh, 2013). In order to avoid the bondage of the formal Bayesian approach, Vrugt and Sadegh (2013) proposed an approximate Bayesian computation (ABC) method, i.e. likelihood-free inference. Surprisingly, the Generalized Likelihood Uncertainty Estimation (GLUE) methodology proposed by Beven and Binley (1992) can be interpreted as a form of approximate Bayesian computation method (Nott et al., 2012; Sadegh and Vrugt, 2013).

Beven and Binley (2013) pointed out that the essential difference between GLUE and formal Bayesian approach is simply whether or not the model residuals must be treated by the explicit error model (i.e. the formal likelihood function). The implicit handling of model residuals (i.e. the informal likelihood function) can be used in the GLUE. The advantage of formal Bayesian approach is that it can expand uncertainty information to bracket model predictions. By contrast, the GLUE is also appropriate for the failure modelling either for model structural or observed data error reasons. Beven et al. (2012) and Beven and Binley (2013) concluded the GLUE is more practical for the uncertainty analysis of hydrological model, because of the epistemic errors and the non-existence of generalized likelihood functions that are appropriate for all model structures. However, Beven and Binley (2013) also agreed that the debate between GULE and formal Bayesian approach or between the formal and informal likelihood function is no sign of real resolution recently, because there is no right answer to the problem of epistemic errors.

Most studies within the GLUE framework used the classical objective functions (Table 2.1) as the informal likelihoods, and most commonly, the *NSE* was used, with a threshold value to define the set of model behaviors (Beven and Binley, 2013). Since the GLUE is an approximate Bayesian computation (ABC) method, there should be an equivalence relationship between the classical objective function and the likelihood function. Similar to the likelihood function, numerous researchers pointed out the objective function of mean squared error (MSE) also implies statistical assumptions that the model residuals should be independent and identically distributed (I.I.D.) according to a Gaussian distribution with zero-mean and a constant variance (Clarke, 1973; Xu, 2001; Jain and Sudheer, 2008). Stedinger et al. (2008) indicated that the standard least squares (SLS), equivalent to maximizing *NSE*, is a kind of likelihood function with the Gaussian I.I.D. error assumption. This theoretical derivation, however, is non-strict because of fixing the standard deviation of residuals/errors. Pande (2013) indicated that minimum mean absolute error (MAE) estimator is equivalent to assuming a Laplace likelihood function. The NSE/MSE and MAE are all distance-based objective functions (Table 2.1), so there could be a generalized equivalence-

relation between the distance-based objective function and the likelihood function with special assumptions.

2.3   Surface-Runoff Generation Method in SWAT

The Soil and Water Assessment Tool (SWAT) developed by the Agricultural Research Service (ARS) of the United States Department of Agriculture (USDA) is computationally efficient and capable of continuous simulation over long time periods, which has proved to be an effective tool for assessing water resource and water-related problems over a wide range of scales and environmental conditions (Gassman et al. 2007). Surface runoff occurs along a sloping surface. The traditional SWAT model provides two methods to estimate overland runoff: the Soil Conservation Service curve number procedure (CN) and the Green & Ampt infiltration method (G&A) (Neitsch et al., 2005).

The CN runoff equation is basically an empirical model which came into common use in the 1950s (SCS, 1972). It was developed to investigate rainfall-runoff relationships in small rural watersheds, but it is easily applicable to large areas (Gabellani et al., 2008). Because the curve number (CN) is a function of the soil type, the land use and the antecedent moisture condition, the CN model can be used to estimate the amounts of runoff under varying land use and soil types (Rallison and Miller, 1982). However, the CN model has been found to have structural limitations. For example, the CN is based only on total rainfall volume, but rainfall intensity and duration are not considered (King et al., 1999). The original purpose of CN model is to compute streamflow volume (minus base flow) for a storm, so treating the curve number runoff (that may include some interflow) as only overland flow by SWAT would result in overestimation of the overland flow (Garen and Moore, 2005; Arnold et al., 2011). The unavailability of detailed and descriptive information on soil properties in CN model often limits the implementation of complete infiltration schemes (Gabellani et al., 2008). To overcome these limitations, many efforts to modify the CN schematization were made (e.g. Yu, 1998; Mishra and Singh, 2003; Gabellani et al., 2008).

The Green & Ampt (G&A) equation was developed to predict infiltration when the excess water was on the surface at all times. That equation assumes that the soil profile is homogeneous, the antecedent moisture is uniformly distributed in the soil profile, and the pattern of soil water movement is the saturated plug/piston flow. The physics-based Horton-like methods present parameters that, though measurable to a point, make it difficult to

achieve a reliable estimate at the catchment scale. To overcome this limitation, Gabellani et al. (2008) developed an approach that maps the CN values onto the initial infiltration rate of a modified Horton's equation, which obviously outperforms SCS-CN method on long, multi-peak events.

Both the CN and the Green & Ampt approaches are based on the principle of the infiltration excess (Horton overland flow; Jury et al. 1991). Horton overland flow occurs when the rainfall rate exceeds the infiltration capacity of the surface soil (Loague et al. 2010). Kannan et al. (2007) pointed out that streamflow results were more accurate by using the CN approach as compared to the G&A approach because the G&A approach holds more water in the soil profile and predicts a lower peak runoff rate. Han et al. (2012) found both the CN method and the G&A method in the SWAT have some limitations in reflecting that soil moisture updates into the surface runoff generation.

Another type of surface runoff process is Dunne overland flow (saturation excess). Dunne overland flow occurs when the rainfall rate is less than the infiltration capacity of the surface soil with small soil-saturation deficit (Dunne 1978; Loague et al. 2010). For Dunne overland flow, surface runoff is contributed by only a portion of a watershed. This concept is often referred to as variable source areas (VSA). Although the CN approach is derived from the principle of Horton overland flow, Steenhuis et al. (1995) revealed that the CN approach can be revisited for variable source runoff areas. In order to further improve the physical basis of the CN and identify the location of the contributing runoff areas, soil-topographic index was incorporated into CN-based watershed models, which is termed SWAT-VSA (Lyon et al., 2004; Schneiderman et al., 2007; Easton et al., 2008). However, White et al. (2011) found that the CN-based SWAT-VSA cannot account for the extreme antecedent moisture conditions that are found in monsoonal climates. They incorporated the water balance model (WB) with the spatial adjustments by using a soil wetness index to replace the CN approach and developed a physics-based SWAT model termed SWAT-WB. They found that SWAT-WB was useful in cases where rainfall intensity is generally less than the soil infiltration capacity. In SWAT-WB, the effective depth of soil profile is a very important parameter, which is used to reflect the effect of topography on soil saturation deficit. However, SWAT-WB needs to determine the parameter of effective soil depth in each soil wetness class, which brings many inter-correlated parameters and results in difficulties in model calibration.

2.4   Soil Erosion

Soil loss has caused critically environmental, ecological and economic problems around the world (Pimentel et al., 1995; Yang et al., 2003; Wu and Chen, 2012). Soil erosion denudes land surface, destroys fragile ecosystem and results in stony desertification. Erosion of the soil particles also removes nutrients from land surface and leads to land impoverishment because the land surface maintains the highest organic matter and nutrients of soil profile (Neitsch et al., 2005). Moreover, in intensive agricultural land-use areas, owing to sorb the nutrients and pesticides by the soil particles, erosion of the soil particles has become the pollution sources of the receiving water bodies (Wu and Chen, 2012). In channel network, the sediment deposition raises the riverbed and the sediment degradation changes the channel morphology. For reservoir construction, soil erosion from the upstream deposits in the bottom of the reservoir, which lowers the reservoir's water holding capacity and consequently its usefulness for hydroelectric power generation, water supply, flood control, etc. (Neitsch et al., 2005). Therefore, it is significant to model the spatial and temporal patterns of soil erosion both for understanding and predicting the soil erosion and sediment transport processes as well as the watershed-scale management of sediments and nonpoint-source pollutants (Wu and Chen, 2012; Mukundan et al., 2013).

However, watershed soil erosion is a very complex process and difficult to be modelled because it not only involves the erosion of soil particles from slopes, but also the sediment transport in channel network. Studies show only a small proportion of the soil eroded within a watershed transports out of the watershed, and a considerable portion accumulates on the relative flat regions of the watershed, such as the foot of the slopes, the valley bottoms, the small tributaries, and the floodplains bordering the main channel system (Walling, 2000). The Soil and Water Assessment Tool (SWAT) as a distributed hydrologic model is popularly used to assess water resource and soil erosion problems all over the world (Gassman et al., 2007; Oeurng et al., 2011; Betrie et al., 2011; Lu et al., 2011; van Griensven et al., 2013; Gebremicael et al., 2013; Mukundan et al., 2013). Soil erosion from slopes in SWAT is simulated by the modified universal soil loss equation (MUSLE) (Williams, 1995), and the simplified Bagnold equation (Bagnold, 1977; Williams, 1980) is used to simulate the sediment transport in the channel network. The original Universal Soil loss Equation (USLE) is a semi-physically based erosion estimation method (Sun et al., 2002), and widely used to estimate the average annual gross erosion from sheet and rill. The MUSLE replaced the rainfall energy factor in the USLE with runoff factor that incorporates both total storm runoff

volume and peak runoff rate. Comparing with the USLE, the MUSLE improves the sediment yield prediction, eliminates the need for delivery ratios, and can be applied to individual storm events (Erskine et al., 2002; Chen and Mackay, 2004; Neitsch et al., 2005). The simplified Bagnold equation assumes the sediment transport capacity of channel is the power function of the peak flow velocity, and the channel maintains equilibrium sediment transport. In other words, when sediment from upstream is greater than the sediment transport capacity, the sediment will deposit in the channel, otherwise the channel will be denuded. Therefore, the sediment loads in the watershed outlet are closely related to the surface runoff of slope and the flow velocity of channel.

Unfortunately, most previous studies separately calibrate flow and sediment objectives, following the calibration procedure of first flow then sediment, for simplifying the model calibration processes (Betrie et al., 2011; Oeurng et al., 2011; Arnold et al., 2012; van Griensven et al., 2013; Gebremicael et al., 2013). Obviously, this way neglects the trade-off information between the river flow and sediment objectives, decreases the prediction precision of sediment loads and increases the uncertainty of model parameters (Lu et al., 2011). Therefore, the multi-response/objective calibration method is more suitable to optimize parameters of SWAT for improvement of the performance of sediment simulation.

The simplest way to solve multi-response/objective optimization problem is to convert the multiple objectives into a single objective. The most common method is the weighted sum principle that the objectives are multiplied with user-defined weights and then added together to form a single objective (Abbaspour et al., 2007; Rostamian et al., 2008). In this method, however, it is equivocal to choose the user-defined weights (Efstratiadis and Koutsoyiannis, 2010). Lu et al. (2010) used an evolutionary algorithm to search the "Pareto front" (i.e. a set of special points where we cannot further improve one objective without impairing other objective) for calibrating parameters of SWAT with both the river flow and sediment objectives. Although the evolutionary algorithms are popularly used in multi-objective optimization (Coello Coello, 2006; Bekele and Nicklow, 2007), there are still some disadvantages (Efstratiadis and Koutsoyiannis, 2010):

1. It can only obtain the "Pareto front", rather than best fit results, so it still needs to convert the multi-objective values into a single value using the weighted sum principle for determination of the best fit result.

2. The evolutionary algorithm cannot estimate the parameter uncertainty because most parameter uncertainty analysis methods (e.g. Generalized Likelihood Uncertainty

Estimation (GLUE)) are based on single objective (Efstratiadis and Koutsoyiannis, 2010; Reichert and Schuwirth, 2012).

Reichert and Schuwirth (2012) developed a scheme for linking statistical bias description to multi-response calibration. In this frame, all model residuals are lumped into a single additive residual term that is measured by a likelihood function under the assumption that the residuals of different objectives are independent of each other. This scheme avoids the trouble of choosing the weights and can combine with Bayesian approach to analyze parameter uncertainty. However, Reichert and Schuwirth (2012) neglected to inspect whether or not the model residuals fulfill the statistical assumptions of the likelihood function.

3 Study Area

3.1 Location and Geomorphology

The Baocun watershed locates in the eastern Jiaodong Peninsula that is a part of Shandong Province in the eastern China (Figure 3.1 (a)). The north of Jiaodong Peninsula connects the Bohai Sea, and the south borders the Yellow Sea. The Jiaodong Peninsula includes three districts: Qindao, Yantai and Weihai. The Baocun watershed belongs to Weihai district. The longitude of Baocun hydrometric station (the red star in Figure 3.1 (b)) is 122°21 18 ´E and the latitude is 37°07 45 ´N. The Baocun watershed is the headstream of Guhe River that feeds into the Yellow Sea.

(a)   (b)
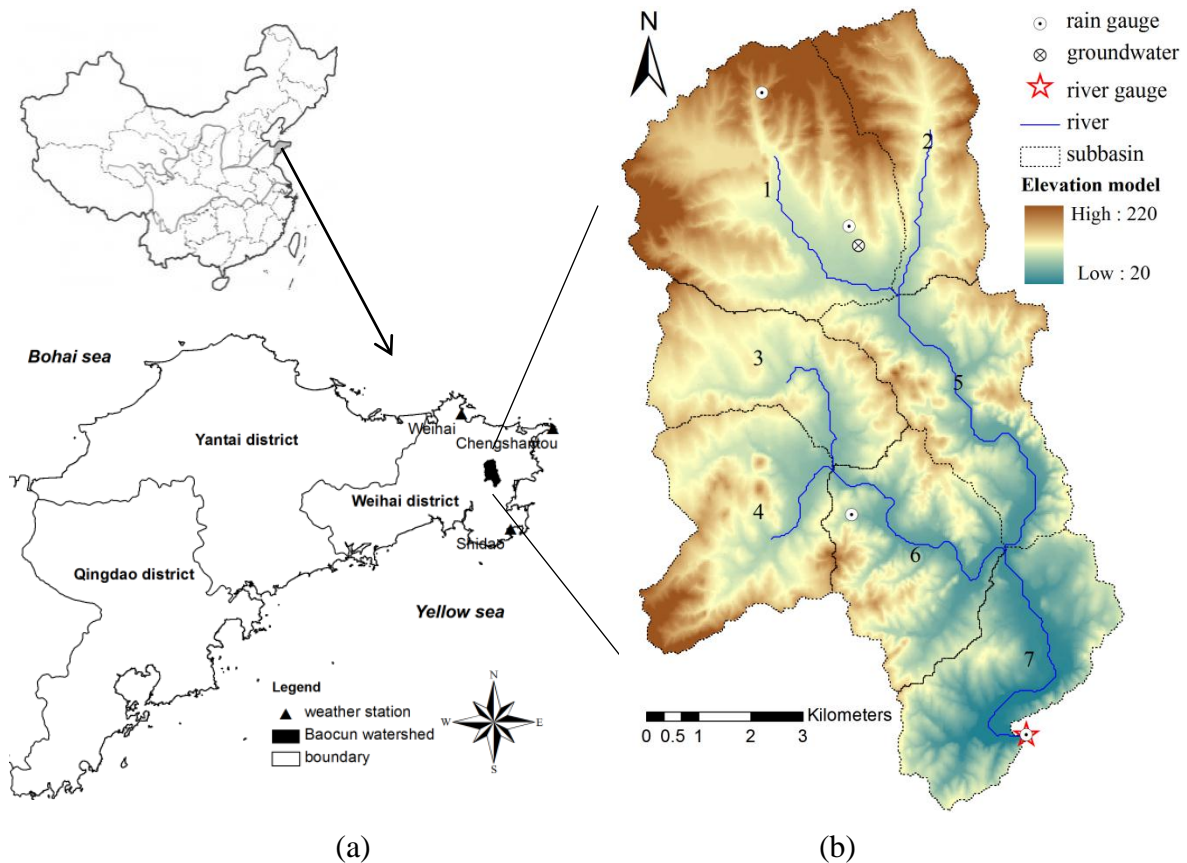
Figure 3.1 Location and topography of the Jiaodong peninsula (a) and the Baocun watershed
(b) as well as position of the gauging stations. Data base: 1:10 000 scale
topographic map. The digit is the river number.

The Baocun watershed is a typical hilly/mountainous area in the Jiaodong Peninsula. The land generally slopes southward. The elevation of the watershed ranges from 20 m at the watershed outlet to 220 m above mean sea level at the headwatershed (northern mountain

area). The watershed area amounts to 86.7 km^2. The length of watershed is 16.1 km, the average width 5.38 km and the average slope 8.2‰. The length of trunk stream is 18.5 km, the average slope 3.5‰, and the meander coefficient 1.29. The Baocun watershed can be divided into seven sub-basins according to the river system. The river number is shown in Figure 3.1 (b). River (1 – 4) are the tributary channels and river (5 – 7) are the main channels, where the river (5 – 6) belong to the second level river and river (7) belongs to the third level river according to the Strahler stream order (Strahler, 1957).

3.2   Geology and Soils

According to the study of Wang et al. (2002), the geological characteristics of the Baocun watershed are shown as follows:

1. The watershed locates in the eastern part of the Jiao-Liao uplift-fault block.
2. The exposed rocks are mainly of the metamorphic and igneous origin of Proterozoic era.
3. The area was always uplifting and suffering from erosion from Early Proterozoic era to Neogene Period.
4. Sediments and weathered residuals deposited in this region until Cenozoic- Quaternary- Pleistocene.
5. The rock stratum is fragmented because of frequent geologic-activities.

Generally, the geology of Baocun watershed is uniform: the Archean and Proterozoic strata (gneiss, diorite and marble) almost cover the whole watershed and the igneous rocks (monzonitic granite) account for a small proportion.

The soil in the Baocun watershed all derived from weathering residues of bedrock such as granite, diorite and gneiss (Wang et al., 2002), so the soil layer is generally shallow, of which the soil texture is coarse. The dominant soil types in the Baocun watershed are Regosols, Luvisols and Fluvisols covering about 94% areas according to the Harmonized World Soil Database (HWSD; FAO/IIASA/ISRIC/ISSCAS/JRC, 2009; Figure 3.2 (a)). The HWSD is from digitizing the 1:1 million scale Soil Map of China. The three main soil types (Regosols, Luvisols and Fluvisols) account for 50, 29 and 15% of the total area, respectively (Figure 3.2 (a)).

The soil map needs to be redrawn in the Baocun watershed due to too coarse spatial-resolution of the world soil database with 1 km resolution (Figure 3.2 (a)). According to field investigations, Regosols are mainly located on mountain tops with steep hillslopes and a low

topographic index (TI, and $TI = \ln\left(\frac{A}{tan\beta}\right)$ where A is upslope contributing area and $tan\beta$ is slope), Luvisols on mountainsides and Fluvisols at mountain-foot/fluvial areas with gentle hillslopes and a high TI. It results from that the soil in the closed mountain area is only from the weathered bedrock or the deposition of upstream sediments (Heimsath et al., 1997; Wang et al., 2002; Norton et al., 2003). In addition, the Baocun watershed is a small area with the uniform geology conditions and the deposition of upstream sediments totally depends on the slope (Prosser, 2000). Therefore, the soil types are closely related to the slope and we can classify the soil types according to the TI values. In fact, the TI/slope values have been widely used by many researchers to estimate the soil thickness, such as S-model (slope model; Saulnier et al., 1997), Sexp-model (slope exponent model; De Rose 1996; Salciarini et al., 2006; Godt et al., 2008) and TI-model (topographic index model; Lee and Ho, 2009).

(a) Harmonized World Soil Database   (b) Redrawn soil map
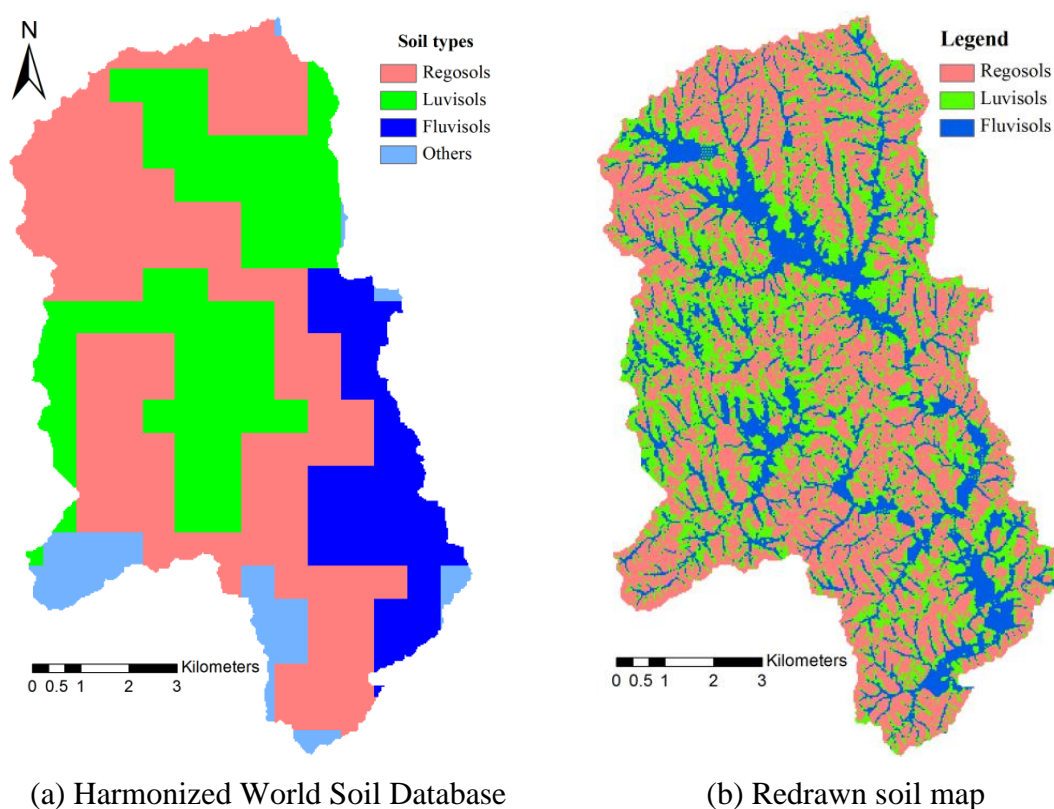
Figure 3.2 Soil type map of the Baocun watershed.

Because the Chinese soil map used by HWSD is charted by merging many fine resolution soil maps based on the major soil types, the area proportion among different soil types is credible. In this study, TI values within the ranges of 3.2~7.5, 7.5~9.5 and 9.5~24.3 correspond to Regosols, Luvisols and Fluvisols, which account for 50, 30 and 20% of the total area that approximate the proportions of soil types in Baocun watershed according to the

HWSD (FAO/ IIASA/ ISRIC/ ISSCAS/ JRC, 2009). The redrawn soil map is shown in Figure 3.2 (b).

In the HWSD, the properties of soil profile are simplified into two layers with the depths of 30cm and 100cm, respectively, which cannot fulfill the requirements of hydrology research. In order to obtain more detailed properties of soil profile, we investigated the soil types in the Baocun watershed from the headwatershed to the watershed outlet, and collected soil samples from 26 points. These sampling points are nearly uniformly distributing from the mountain-top to mountain-foot. In every sampling point, we generally took 4 soil-cores (from the upper layer to the lower layer) to analyze soil properties. The analyzed results include the soil particle size (weight based proportions) measured by laser equipment, saturation soil water content, bulk density and soil hydraulic conductivity (K) measured by the falling head method (Johnson et al., 2005; Cheng et al., 2011). The summary survey results of Regosols, Luvisols and Fluvisols are shown in Table 3.1. This table shows the soil layer of Regosols and Luvisols can be obviously separated into two layers: the upper layer is farming/root layer, and the lower layer is the weathered layer. By contrast, the soil layer of Fluvisols is relatively uniform. For all soil types, the grains of sand are the largest part of soil particles, which account for over 50%; the grains of gravel are the second; and the grains of clay and silt generally account for less than 10%. For Regosols and Luvisols, the saturation water content of upper layer is generally larger than that of lower layer, but the bulk density of upper layer is smaller; and the K of upper layer is much greater than that of lower layer, where the ratio between the K of the upper and lower layer is close to 10. However, all the K of Regosols and Luvisols are much less than that of Fluvisols.

Table 3.1 Summary of soil properties from field survey data.

Categories	Depth [a] (m)	Clay <2um	Silt 2~50um	Sand 50~1000um	1~2mm	Gravel 2-4mm	>4mm	Saturated moisture	Bulk density (g/cm^3)	K [b] (m/s)
Regosols	0.3	0.91%	8.12%	47.80%	13.79%	12.09%	17.29%	0.47	1.35	1.90E-05
	0.7	0.56%	5.41%	48.84%	13.06%	13.02%	19.10%	0.37	1.59	1.90E-06
Luvisols	0.4	1.21%	9.17%	47.98%	15.66%	12.59%	13.39%	0.42	1.45	1.35E-05
	2.6	0.51%	4.14%	38.65%	17.42%	17.53%	21.76%	0.42	1.5	1.93E-06
Fluvisols	2.0	0.64%	4.85%	43.41%	18.53%	16.37%	16.20%	0.46	1.57	3.08E-05

[a] Note that the soil depth is only an approximate value.

[b] K is the soil hydraulic conductivity.

3.3 Vegetation and Landuse

According to the USGS satellite data (MODIS/Terra+Aqua Land cover type yearly L3 global 500 m SIN grid product (short name: MCD12Q1); Figure 3.3 (a)), the main land use is cropland that covers more than 99% area of Baocun watershed. The field survey shows that the dominant crops are peanuts, corn and winter wheat. The agricultural lands are farmed three times every two years, which is termed crop rotation. The detailed schedule of planting crops is peanuts in May, winter wheat in October and corn in June next year. Therefore, in fact, the croplands cyclically vary. To overcome the problem of land utilization change, the average annual land use map is charted. The detailed steps are as follows: firstly, extract the fixed land use from the 1:10 000 scale topographic map, such as water body, forest (mostly are apple orchard) and residential area; secondly, randomly fill the rest area by peanut, winter wheat and corn with the proportion of 0.25, 0.5 and 0.25, respectively, according to the growth time of crops. The average annual land use map is shown in Figure 3.3 (b).

(a) MODIS/Terra+Aqua Land cover type   (b) Average annual land use map
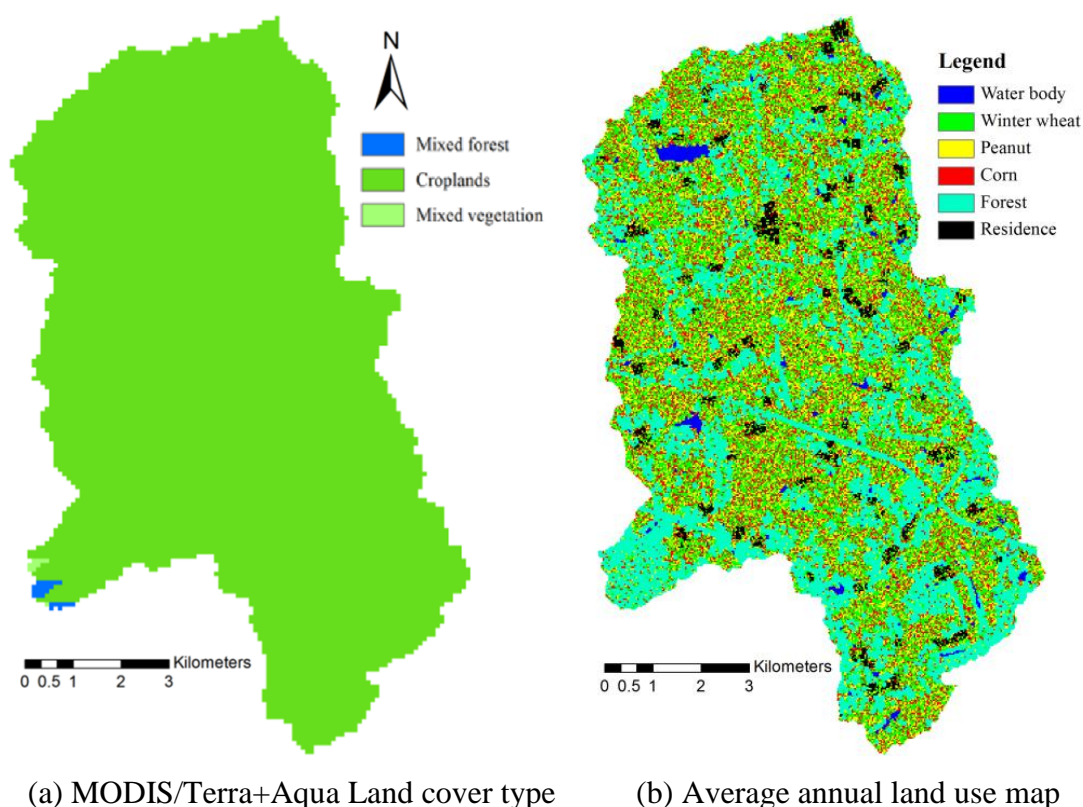
Figure 3.3 Land cover map of the Baocun watershed.

The Baocun watershed as a rural area was affected by human activities intensively. For cultivation of crops, almost all lands were transformed into terraced fields, even the steep slopes in the mountaintop (Figure 3.4). The terraced fields resemble steps, where the sloping

surfaces are cut into a series of successively receding flat surfaces or platforms (Figure 3.4). The terraced fields can delay the overland flow and facilitate the rainfall infiltration; and ultimately decrease the surface runoff and increase the soil water. As a result, terraced fields benefit not only the cultivation of crops, but also the soil and water conservation. Therefore, the terrace farming is widely used on hilly/mountainous terrain in China.

Figure 3.4 Landscape of terraced fields in the Baocun watershed. Terraced fields benefit not only the cultivation of crops, but also the soil and water conservation. The red arrow indicates a small reservoir in the valley.

3.4   Climate

The climate of Baocun watershed belongs to the warm temperate zone in a humid monsoon region with an annual average precipitation of 806 mm. The variation of inter-annual precipitation is extremely large. Over the past fifty years, the maximum was observed in 2003 with an annual precipitation of 1219.7 mm, and the minimum occurred in 1999 when only 383.7 mm was recorded. The intra-annual variability of precipitation is also very large: nearly 70% of the total rainfall occurs in the wet season from June to September (termed flood season; Figure 3.5). Meanwhile, typhoons often bring heavy rainfall, e.g. the daily rainfall reaches 301 mm in 8/19/1997. The annual average potential evaporation is 899.0 mm (measured by pan evaporation equipment termed E601). The average monthly temperature ranges from -0.8 ℃ in January to 24.4 ℃ in August (Figure 3.5). Figure 3.5 shows that the monthly variations of precipitation are nearly the same to those of temperature, where the warm months correspond with the moist months, and vice versa. By contrast, the monthly variations of potential evaporation are different from those of the temperature. The maximum

potential evaporation occurs in May, rather than August (i.e. the warmest and moistest month). It may result from a number of rainy or cloudy days in August, because the large relative-humidity (i.e. water vapour) can reduce/inhibit the evaporation and the clouds can block the solar radiation.
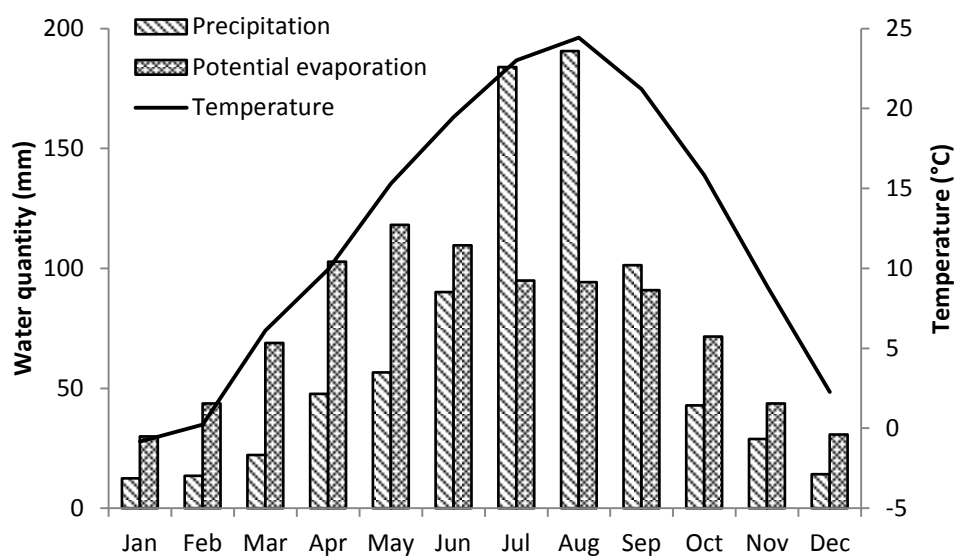
Figure 3.5 The average monthly precipitation, potential evaporation and temperature in Baocun watershed during 1953 - 2008. Data base: Precipitation and Potential evaporation collected from Baocun hydrometric station (Figure 3.1 (b)); and Temperature collected from national weather stations (Figure 3.1 (a)).

3.5 Hydrography/Hydrology

3.5.1 River Discharges

River discharge is the volume rate of water flow transported through the cross-sectional area of hydrometric station. In this study, the daily river discharges are measured at the Baocun hydrometric station (red star in Figure 3.1 (b)). The 19 years (from 1993 to 2011) daily river discharges and the corresponding average rainfall over watershed are shown in Figure 3.6. In this figure, the log-plot is to highlight baseflow and the discontinuous log-river-discharges mean no-river-flow in the corresponding period, such as 1999, 2000 and 2001. The rainfall data are collected from four rainfall gauges inside Baocun watershed (circles in Figure 3.1 (b)). The characteristics of rainfall-runoff in Baocun watershed are summarized as follows:

1. The average runoff depth is 295 mm and the runoff yield rate of rainfall-runoff (i.e. average annual runoff coefficient) reaches 37%.

2. Most (about 80%) of the river discharges occur in the flood season from June to September, where 70% of the total rainfall occurs.

3. The delay in response of floods to rainfall can be neglected possibly because of the short flow length (watershed radius of 5 km) in Baocun watershed and the daily rainfall-runoff process.

4. Floods present impulse form in daily river discharge chart, which reflects the small storage capacity of Baocun watershed owing to the small watershed area (86.7 km^2) with shallow soil layer (Table 3.1).

5. There is a considerable temporal variance of the river discharges: (1) the average of the observed river discharges is 0.810 m^3/s, but the standard deviation is 4.77 m^3/s and the maximum flow is 205 m^3/s in 2007; and (2) the flow less than 1.0 m^3/s accounts for 89% of total flow and the proportion of the no-river-flow reaches 10%, possibly because of the extremely non-uniform precipitation in temporal distribution owing to the great intensity of typhoon rainfall (Figure 3.5).
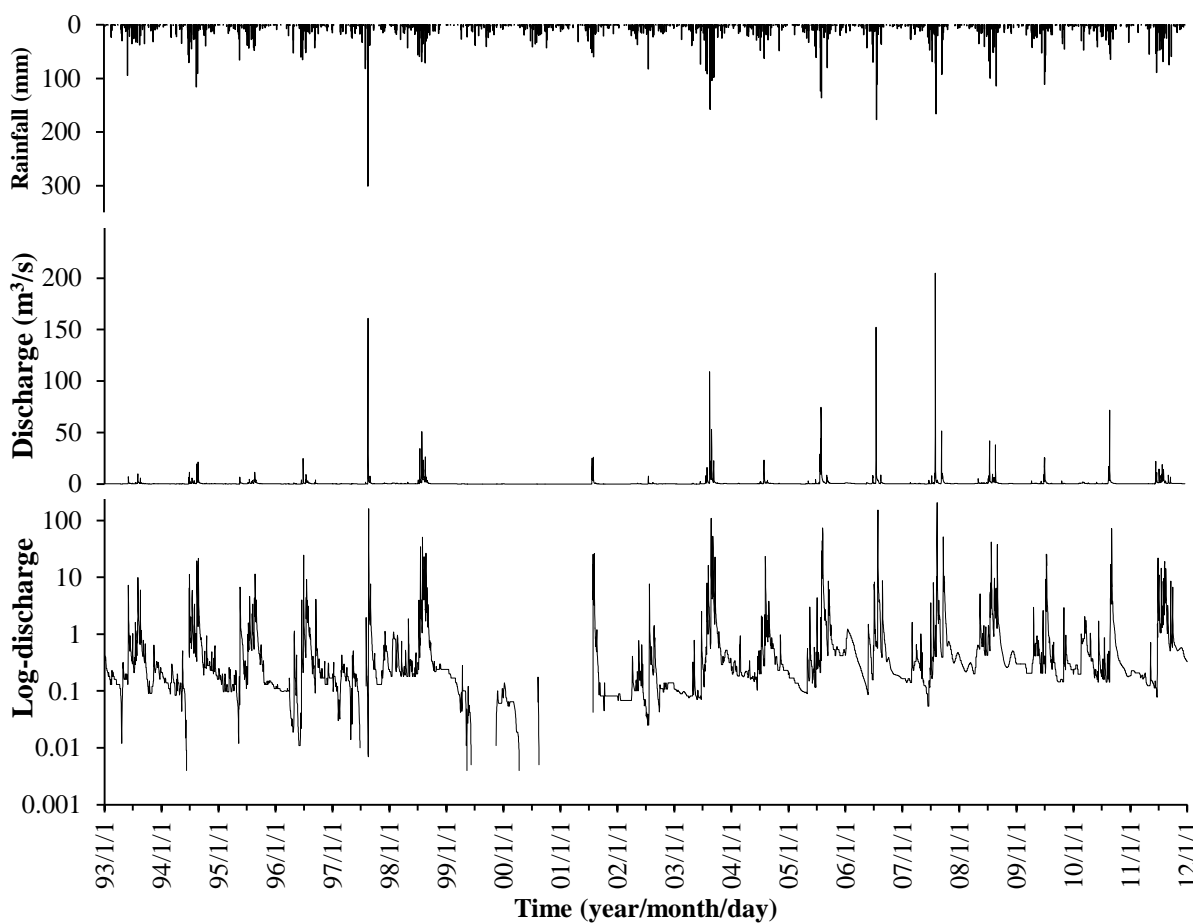
Figure 3.6 The observed river discharges in the Baocun hydrometric station. The log-plot is to highlight baseflow.

3.5.2   Irrigation Constructions

The Baocun watershed is an agricultural area with abundant rainfall. However, the water resources of Baocun watershed are extremely non-uniform in temporal distribution because of uncertain monsoon. Therefore, for agricultural irrigation, farmers built some small reservoirs/ponds in the mountain area (e.g. the small reservoir in the Figure 3.4). In the fluvial area, local residents also built some small dams across the main channel for water supply, irrigation and transportation (Figure 3.7). The effect of small dams in main channel on the daily river discharges is small because of small reservoir-capacity (Figure 3.7). However, those small dams significantly affect the sediment transport by reducing the velocity of river discharge and the sediment transport capacity of flow, which result in the sediment deposition. During the spring ploughing season, there are a lot of irrigations, which result in that the river discharges often reduced suddenly, even dried at the beginning of rainy season, such as 1994 and 1997 (Figure 3.6).

Figure 3.7 Small dam across the main channel in the Baocun watershed. The small dam affects the daily river discharge slightly, but the sediment load significantly, which intercepts sediments.

3.5.3 Groundwater

There are two main features of the groundwater in the Baocun watershed:

1. The groundwater resources of most area (i.e. mountain area) are in the fracture of original rocks and valueless to be mined because of poor storages.

2. The fluvial areas store abundant groundwater resources, but the fluvial area accounts for less than 20% of watershed areas (Figure 3.2).

In summary, because of the shallow soil layer and steep slopes, the available groundwater resources are deficient in Baocun watershed.

As has shown in Figure 3.1 (b), the location of the groundwater gauge lies in the flood plain (i.e. fluvial area) and started working in June 2007. The borehole material shows that the soil can be separated into three layers: i) 0.0 - 1.5 m, filled with loose soil, ii) 1.5 - 3.5 m, silty and coarse sand, and iii) below 3.5 m, coarse sand. The average groundwater table depth is 2 m. The groundwater table depth varies between 0.8 m and 3.5 m during the monitoring period (Figure 3.8). In Figure 3.8, the value of the groundwater level is the elevation above mean sea level. Figure 3.8 shows that the groundwater levels often rise quickly but fall slowly. And the delay in response of groundwater levels to rainfall can be neglected. These demonstrate that the soil water quickly recharges the groundwater owing to the high hydraulic conductivity of soil layer (Table 3.1) and the shallow depth of groundwater table.
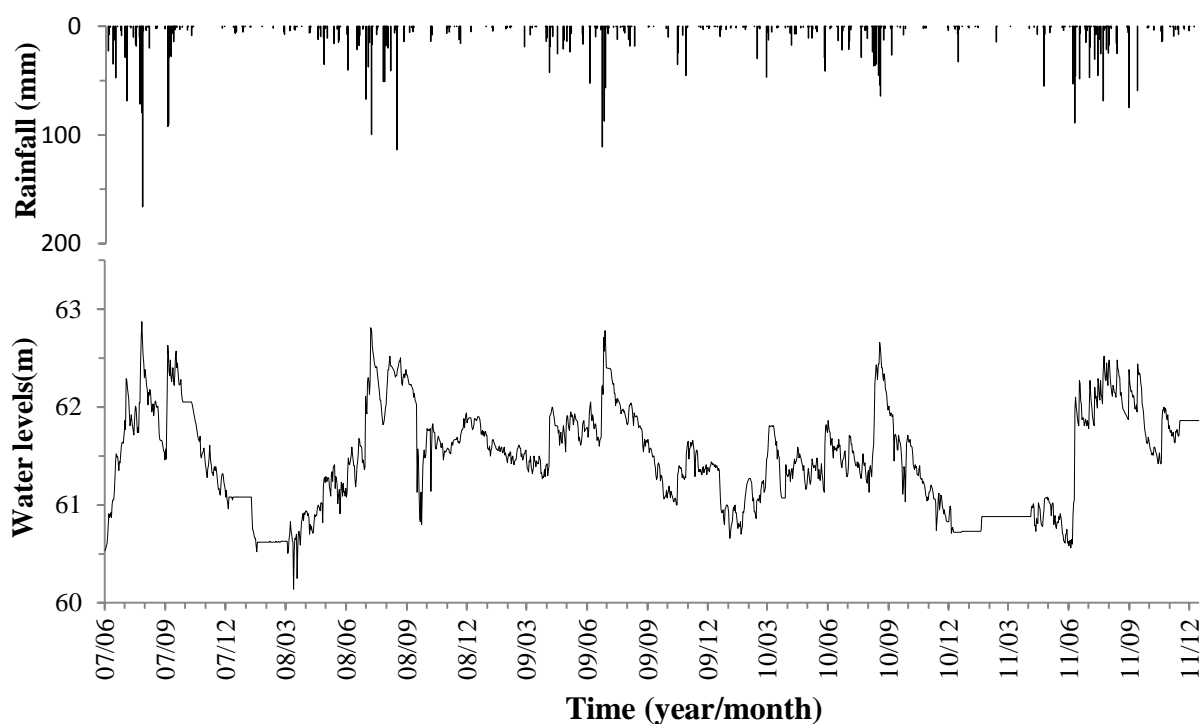
Figure 3.8 Observed rainfall and groundwater levels in the fluvial area. The response of groundwater levels to rainfall is very rapid, especially to rainstorm.

Because of the shallow water table depth, the large variation of groundwater level (Figure 3.8) and the sandy soil-layers with high hydraulic conductivity around the groundwater gauge (Table 3.1), the storage capacity of the unsaturated soil layer could be neglected on the daily time step and there should be a close relationship (i.e. linear relationship) between the groundwater table level and the total soil water volume (Qiao et al., 2012).

3.5.4 Sediment Transport

After the development of hundreds of years, almost all lands of Jiaodong peninsula had been transformed into terraced fields for agriculture (Figure 3.4). The farmlands damage the vegetative cover, make the surface soil loose and result in the soil losses. Additionally, the soils in this area derived from residues of weathered bedrock with many sandy particles (Table 3.1), which are loose and eroded easily. Therefore, there is a serious soil loss problem in the Jiaodong peninsular. The Baocun watershed is a test watershed for studying the soil and water conservation in the Jiaodong peninsular. In this study, the river sediment transport data are collected from the Baocun hydrometric station for studying the soil loss problem. The observed data show that from 1993 to 1999, the annual erosion rate of the watershed is 160 t/(km^2·a). The daily sediment loads during the flood season (from June to September) of 1993 - 1999 are shown in Figure 3.9. There are totally 854 daily data, but the number of non-zero data is too few (92), accounting for less than 11%. The characteristics of sediment transport in the Baocun watershed are summarized as follows:

1. Sediment loads heavily depend on the river discharges, especially flood.
2. Sediment loads are often suddenly close to zero.
3. Sediment transport seems isolated event, probably because the small dams in the main channel intercept sediments after flood (Figure 3.7).

In summary, the Baocun watershed is a rural mountain watershed with the monsoon climate, where 70% of the total rainfall and 80% of the river discharges occur in the flood season from June to September. The uneven temporal distribution of water resources causes water shortages in Baocun watershed. The heavy rains brought by typhoons often cause flood hazard and serious soil loss problem. Meanwhile, the human activities intensively affected the hydrological processes. For example, the terraced fields changed the micro-topography and affected the surface-runoff generation. And the small dams across the main channel

reduced the sediment transport capacity of flow and significantly affected the river sediment loads.
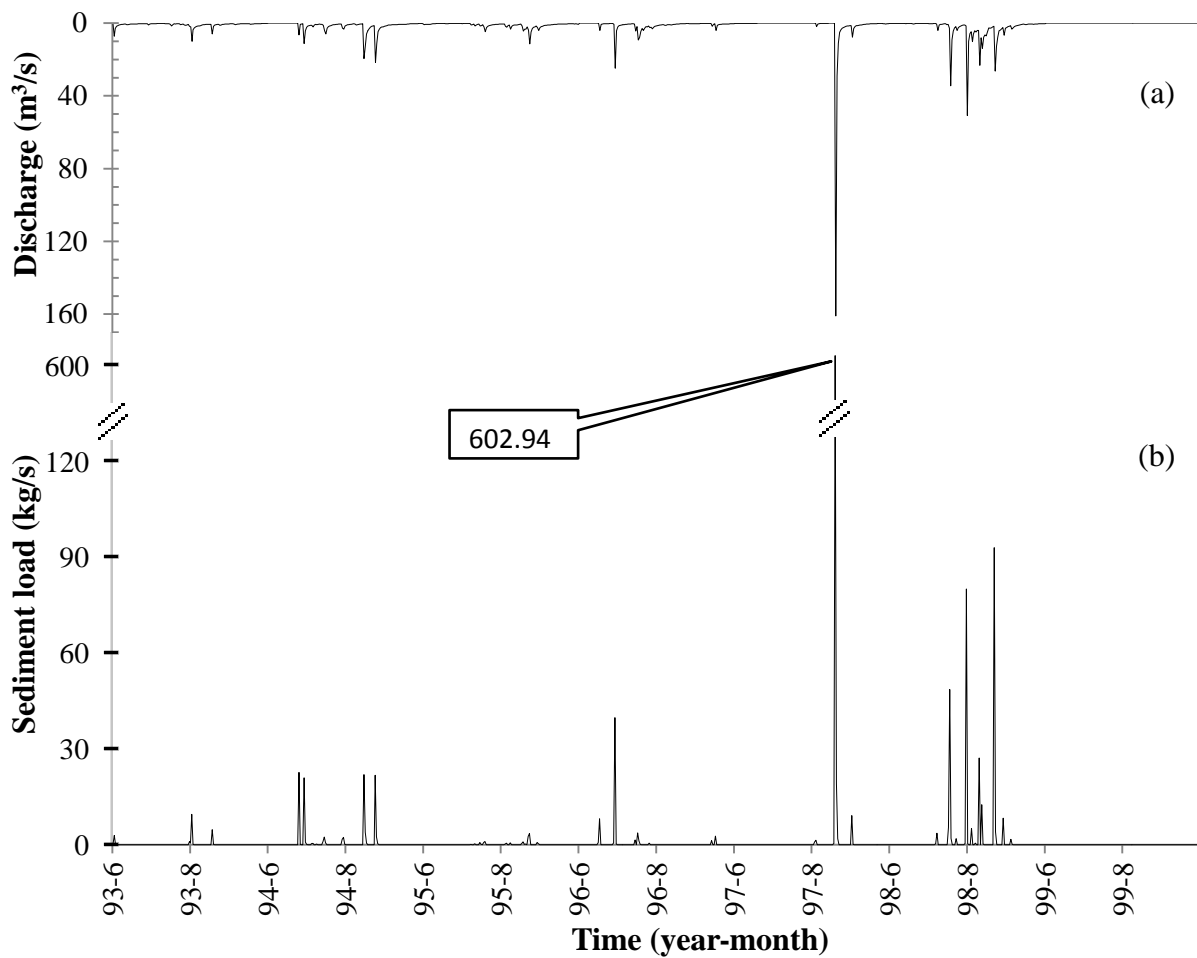
Figure 3.9 River discharges (a) and sediment loads (b) on the daily time step during the flood
season (from June to September) of 1993 - 1999. The values of sediment loads
often suddenly fall to zero, possibly because of small dams across the main
channels.

4   Methods

Hydrological models are popularly used for a variety of applications. However, the performance (goodness-of-fit) of the hydrological model depends on how quality estimate of model parameters (Schaake, 2003). Therefore, for improvement of the performance of hydrological model, the model parameters are often determined through calibration against the historical record data (Duan, 2003). Recently, the automatic calibration techniques are popularly used because of convenience and high efficiency (Willems, 2009). In the automatic calibration procedure, the objective functions are used to ascertain the goodness-of-fit of hydrologic model instead of the subjective visual judgement.

This study attempts to evaluate the effect of the objective functions on model calibration through three cases: likelihood function comparison, multi-response calibration and model comparison. This chapter firstly introduces the hydrological model ― SWAT (Section 4.1), and then presents the procedures of the automatic calibration setup (Section 4.2). Next, it improves and compares the likelihood functions (Section 4.3.1), and then extends the likelihood functions for the multi-response calibration with the river flow and sediment objectives (Section 4.3.2). Finally, it introduces an objective function to compare different rainfall-runoff methods in SWAT model (Section 4.3.3).

4.1   Hydrological Model

Hydrologic models can be classified into various categories based upon the modelling approaches used (Daniel et al., 2011). For example, based on the nature of the employed algorithms, the hydrological models can be classified into empirical, conceptual and physically-based models (Wong and Koh, 2008).

1. *Empirical models*. Empirical models consist of approximate/empirical functions used of fit available data (e.g. river discharges), which is also called black-box models, e.g. the regression model. The empirical models cannot provide detailed information about hydrological processes, e.g. flow paths.

2. *Physically-based models*. Purely physically-based models use the basic physics equations (conservation of mass, momentum and energy) to describe the hydrologic processes. However, because of the hydrologic complexity and especially the hydrologic heterogeneity, the performance (goodness-of-fit) of the purely physically-based models is often poor in the large watershed (Savenije, 2009; Montanari and

Koutsoyiannis, 2012). Montanari and Koutsoyiannis (2012) pointed out that essentially, there are no purely physically based models in the large hydrological system.

3. *Conceptual models*. Conceptual/process models are proposed for the balance between the physical basic/meaning and the model performance. In fact, conceptual/process models are the simplified physically-based models. Therefore, sometimes the conceptual/process distributed models are termed as the physically-based models in contrast to empirical models (White et al., 2011; Montanari and Koutsoyiannis, 2012).

Based on the spatial representation methods, the hydrological models can be classified into lumped and distributed models. The lumped models represent the watershed by a single, lumped hydrologic element. By contrast, the distributed models represent the watershed by a set of spatially distributed elements. Schaake (2003) pointed out that essentially, the smallest element of all distributed models is a lumped model.

The physically based distributed hydrological models not only simulate the river discharges in the watershed outlet, but also mimic the spatial dynamics of hydrological processes, such as flow path and groundwater. More importantly, their parameters have physical meanings, which can be surveyed in the field. Therefore, the physically based distributed hydrological models are widely proposed (Smith et al., 2004).

The Soil and Water Assessment Tool (SWAT) as a distributed hydrological model is popularly used to water resource management all over the world (Gassman et al., 2007). SWAT describes the spatial distribution of hydrological processes by dividing a watershed into multiple sub-basins, which are then further subdivided into hydrologic response units (HRUs) consisting of homogeneous land use, soil characteristics and slope. HRU is the smallest element/unit of SWAT. The HRU represents percentages of the HRU area on the sub-basin area rather than spatially identified locations. HRUs are independent of each other, which mainly include four water storages (surface water, soil water and shallow/deep aquifer) and five pathways for water movement (evapotranspiration, overland flow, interflow/lateral-flow, groundwater return flow/baseflow and percolation from soil layer to groundwater aquifer). The model structure of one HRU in the SWAT model is shown in Figure 4.1, where the shallow GW means the shallow groundwater aquifer, and the arrow is the direction of water movement. In SWAT model, the runoff concentration includes three pathways (Figure 4.1): overland/slope flow concentration, channel flow concentration and main channel routing. Detailed descriptions of these hydrological processes in SWAT are shown below.
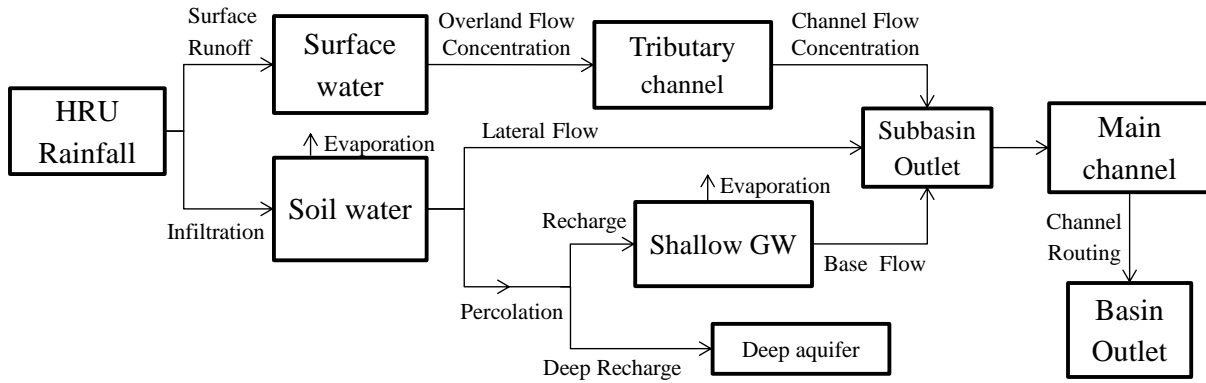
Figure 4.1 Schematic of main pathways for water movement in SWAT. The arrow indicates the direction of water movement. HRU is the smallest element of SWAT.

4.1.1 Surface Runoff

Surface runoff or overland flow occurs along a sloping surface. The traditional SWAT model provides two methods of estimating surface runoff. One is the Soil Conservation Service curve number procedure (CN), including the soil water content dependence (CN-Soil) and the plant evapotranspiration dependence (CN-ET) retention parameter methods. The second option is the Green & Ampt infiltration method (G&A) (Neitsch et al., 2005).

4.1.1.1 Traditional Rainfall-Runoff Methods

Curve Number approach (CN). On the basis of the hypothesis that the ratio between potential runoff volume and runoff volume (Q_{surf}) equals the ratio between potential maximum retention (S) and infiltration volume (F) by the SCS curve number approach (Figure 4.2), a simplified mass balance equation can be obtained:

$$\frac{R_{day}-I_a}{Q_{surf}} = \frac{S}{F} = \frac{S}{R_{day}-I_a-Q_{surf}} \tag{4.1:1}$$

where

Q_{surf} is the accumulated overland flow (mm),

R_{day} is the daily rainfall (mm),

S is the retention parameter (mm),

F is the infiltration volume (mm),

I_a is the initial abstraction (mm) that is commonly approximated as 0.2 S.

Figure 4.2 Principle of the CN approach (cited from Patel (2009)).

By solving Q_{surf} from Equation (4.1:1), we get:

$$Q_{surf} = \frac{(R_{day}-I_a)^2}{R_{day}-I_a+S} = \frac{(R_{day}-0.2S)^2}{R_{day}+0.8S} \qquad (4.1:2)$$

The retention parameter (S) is defined as:

$$S = 25.4(\frac{1000}{CN} - 10) \qquad (4.1:3)$$

where

> CN is the curve number for the day (CN2, Table 4.2), which is a function of the soil type, land cover and antecedent moisture condition.

When the retention parameter varies with soil profile water content (CN-Soil approach), the following equation is used (Neitsch et al., 2005):

$$S = S_{max}(1 - \frac{SW}{SW+e^{(w_1-w_2 \cdot SW)}}) \qquad (4.1:4)$$

where

> S_{max} is the maximum value of the retention parameter (mm),
>
> SW is the soil water content of the entire profile (mm),
>
> w_1 and w_2 are shape coefficients, which depend on the amount of water in the soil profile at field capacity.

When the retention parameter varies with plant evapotranspiration (CN-ET approach), the following equation is used (Neitsch et al., 2005):

$$S = S_{prev} + E_0 \times \exp(\frac{-cncoef \times S_{prev}}{S_{max}}) - R_{day} + Q_{surf} \qquad (4.1:5)$$

where

S_{prev} is the retention parameter for the previous day (mm),

E_0 is the potential evapotranspiration (mm/d),

cncoef is the weighting coefficient (CNCOEF, Table 4.2).

Green-Ampt approach (G&A). The Green & Ampt infiltration equation (G&A approach) assumes that the soil profile and antecedent moisture are all homogeneous, the soil above the wetting front is completely saturated and there is a sharp break in moisture content at the wetting front (Figure 4.3). In SWAT, the Green & Ampt infiltration rate is defined as (Neitsch et al., 2005):

$$f_{inf,t} = K(1 + \frac{\psi_{wf} \cdot \Delta\theta_v}{F_{inf,t}})$$
(4.1:6)

where

$f_{inf,t}$ is the infiltration rate at time *t* (mm/d),

K is the hydraulic conductivity (mm/d, SOL_K, Table 4.2),

ψ_{wf} is the wetting front matric potential (mm),

$\Delta\theta_v$ is the change in volumetric moisture content across the wetting front,

$F_{inf,t}$ is the cumulative infiltration at time *t* (mm).

The surface runoff is the rainfall minus the infiltration rate:

$$Q_{surf} = R_{day} - f_{inf,t}$$
(4.1:7)
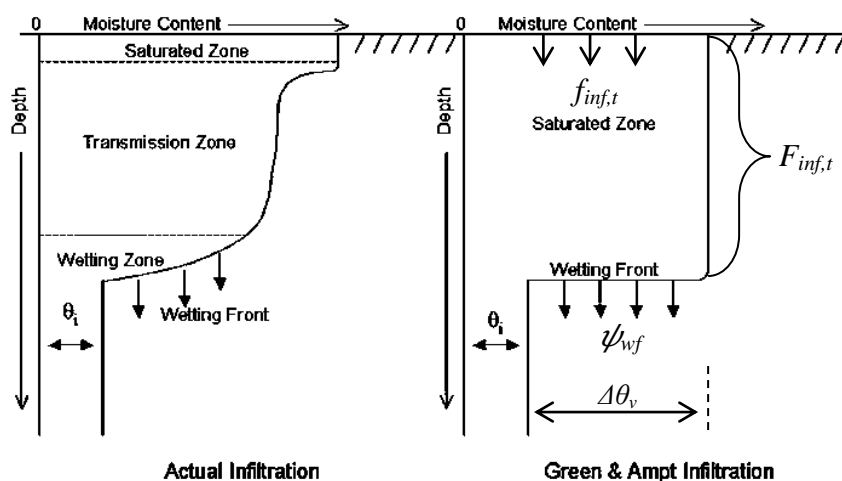
Figure 4.3 Principle of the G&A approach (cited from Neitsch et al. (2005)).

4.1.1.2 Improvement of Rainfall-Runoff Methods

White et al. (2011) proposed a kind of SWAT model termed SWAT-WB that incorporated the physically-based rainfall-runoff approach (i.e. the water balance method, WB; Figure 4.4) into the SWAT for replacing the empirical rainfall-runoff approach (i.e. the curve number method, CN; Figure 4.2). The WB model assumes that (1) there is no surface runoff generation in the unsaturated soil layer, and (2) the rainfall all infiltrates into the soil profile until the soil layer is saturated (Figure 4.4). The surface runoff in the WB model is the rainfall minus the available soil moisture storage of the soil profile (Figure 4.4):

$$Q_{surf} = \begin{cases} R_{day} - \tau & R_{day} > \tau \\ 0 & others \end{cases} \tag{4.1:8}$$

where

τ is the available soil moisture storage (mm).

τ is also termed as the saturation deficit in the soil profile, which is defined as:

$$\tau = \eta' - \theta = EDC \times \eta - \theta \tag{4.1:9}$$

where

η' is the effective porosity of soil profile (mm),

θ is the volumetric soil moisture (mm),

EDC is the effective depth of the soil profile (unitless, ranging from 0 to 1), which reflects the effect of topography on soil saturation deficit (EDC, Table 4.2),

η is the total soil profile porosity (mm).
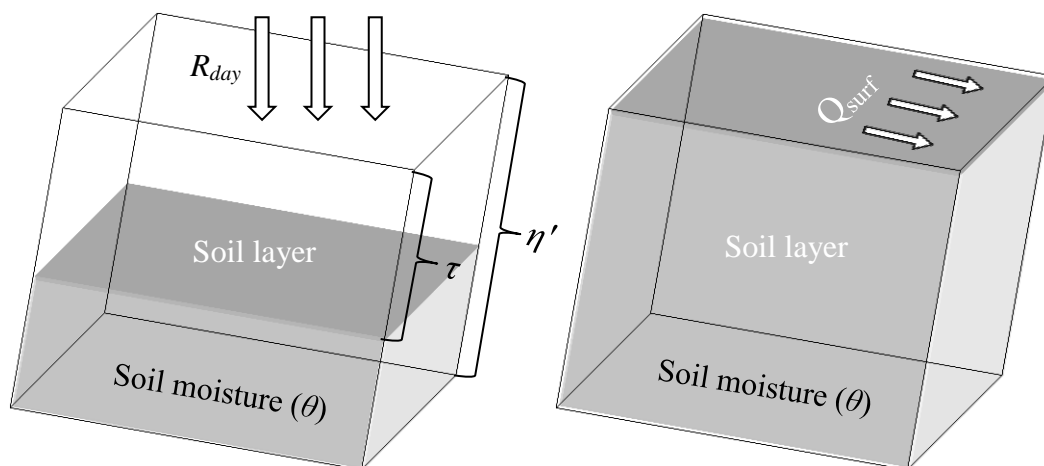
(a) Unsaturated soil layer (no surface runoff)   (b) Saturated soil layer (surface runoff yield)

Figure 4.4 Principle of the Water Balance (WB) model.
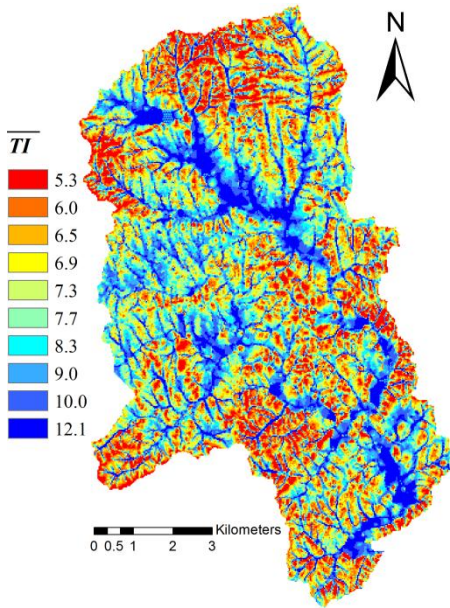
Figure 4.5 Soil-topographic
index map in the
Baocun watershed.

The method, based on simple water balance Eqs. (4.1:8) and (4.1:9), is termed the WB approach. To further express the spatial variation of the rainfall-runoff mechanism influenced by topography in addition to soil type in the SWAT, the soil topography index (*STI*) map is substituted for the soil type map, which combines the soil type map with the topographic index map. The soil topographic index (*STI*) is reclassified into wetness classes of equal area. In the Baocun watershed, because soil types depend on *TI* (as detailed in Section 3.2; Figure 3.2 (b)), soil-topographic indexes are classified into ten units in terms of ten *TI* classes of equal area shown in Figure 4.5. In this figure, the value of *TI* in the legend is the averaged *TI* across

each topographic index class. The higher the value of *TI*, the flatter the sloping surface is. White et al. (2011) computed surface-runoff using Equation (4.1:8) in each wetness class which has its own independence parameter of *EDC*, i.e. EDC_i where i is the order of the wetness class. EDC_i is spatially varied in such a way that low values are assigned to areas with a high likelihood of saturation, and higher *EDCs* are used for areas where not much surface runoff is generated via saturation excess.

This assignment of *EDC* values by White et al. (2011) results in difficulties in model calibration because too many *EDC* values need to be optimized. In fact, *EDC* is closely associated with the topographic index (*TI*): *EDC* is small in an area with a large *TI* and vice versa. Easton et al. (2011) built a method for the estimation of EDC_i based on the assumption of the inversely proportional relationship between EDC_i and TI_i. However, their method is complex and arbitrary. Actually, the spatial variation characteristics of EDC_i are similar to those of storage deficit (SD_i) which is a famous variable in TOPMODEL and a linear relationship with TI_i (Beven and Kirkby, 1979; Chen et al., 2010). In TOPMODEL, the SD_i is calculated as:

$$SD_i = \overline{SD} + m \times (\overline{TI} - TI_i) = (\overline{SD} + m \times \overline{TI}) \times (1 - \frac{TI_i}{\overline{TI}} \times \frac{m \times \overline{TI}}{\overline{SD} + m \times \overline{TI}}) \tag{4.1:10}$$

where

SD_i is the storage deficit of the *i-th* wetness class (mm),

49

$\overline{SD}$ is the catchment average SD_i (mm),

TI_i is the topographic index of the i-th wetness class,

$\overline{TI}$ is the catchment average TI_i,

m is a coefficient of the exponential (m).

By dividing the $(\overline{SD} + m \times \overline{TI})$ in Equation (4.1:10), we get:

$$\frac{SD_i}{\overline{SD}+m\times\overline{TI}} = \begin{cases} 1 - \frac{TI_i}{\overline{TI}} \times (1 - \frac{\overline{SD}}{\overline{SD}+m\times\overline{TI}}) & \frac{TI_i}{\overline{TI}} \times (1 - \frac{\overline{SD}}{\overline{SD}+m\times\overline{TI}}) < 1 \\ 0 & others \end{cases} \qquad (4.1:11)$$

Because $\overline{SD}$ and $\overline{TI}$ are constant in the watershed, the spatial variation of SD_i is a linear relationship with TI_i. This relation is transplanted by us to describe the spatial variation of EDC_i:

$$EDC_i = \begin{cases} 1 - \frac{TI_i}{\overline{TI}} \times (1 - EDC) & \frac{TI_i}{\overline{TI}} \times (1 - EDC) < 1 \\ 0 & others \end{cases} \qquad (4.1:12)$$

where

EDC is the catchment average EDC_i, which is a constant for every HRU.

The calculation method directly links the water balance (WB) with the variable source area (VSA) and is termed the WB-VSA approach in this study. The description of saturation deficit and overland flow by the WB-VSA approach is the same as in the WB approach proposed by White et al. (2011), except that the EDC_i value is estimated by Equation (4.1:12).

In summary, this section presents five rainfall-runoff methods: CN-Soil, CN-ET, G&A, WB and WB-VSA approaches. Comparison of the surface-runoff generation mechanisms among the five approaches is shown in Table 4.1. The curve number (CN) approaches (CN-Soil and CN-ET) are the empirical method. By contrast, the G&A, WB and WB-VSA are the physically based methods. The CN and G&A approaches are all based on the principle of the infiltration excess overland flow. However, CN can be revisited for variable source runoff areas (VSA; Steenhuis et al., 1995). By contrast, WB and WB-VSA approaches are based on the principle of saturation excess overland flow. The distinction between the CN-Soil and CN-ET approaches is whether the retention parameter (S) is estimated according to the soil profile water content (Eq. (4.1:4)) or the potential evapotranspiration (Eq. (4.1:5)). The CN-ET is developed for the watershed with shallow soil layer, where the CN-Soil was predicting too much runoff (Arnold et al., 2011). The difference between the WB-VSA and WB

approaches is whether the principle of variable source areas (VSA) is incorporated into water balance model (WB) or not.

Table 4.1 Comparison of the surface-runoff generation mechanisms among CN-Soil, CN-ET, G&A, WB and WB-VSA approaches.

Characteristics [*]	CN-Soil	CN-ET	G&A	WB	WB-VSA
Physically based method	×	×	√	√	√
Infiltration excess overland flow	√	√	√	×	×
Saturation excess overland flow	√	×	×	√	√
Variable Source Areas	√	×	×	×	√
Soil water dependence	√	×	√	√	√

[*] "√" ("YES") means the approach includes corresponding "characteristic";

"×" ("NO") means the approach does not include corresponding "characteristic".

4.1.2 Other Hydrological Processes

Infiltration.   Infiltration refers to that the rainfall enters into a soil profile from the soil surface. Among the five rainfall-runoff methods (CN-Soil, CN-ET, G&A, WB and WB-VSA) introduced above, only the G&A method does directly model infiltration (Eq. (4.1:6); Figure 4.3). However, the performance of G&A method, at the daily time-step, is often poor (Neitsch et al., 2005; Kannan et al., 2007; White et al. 2011).

Lateral flow.   Lateral flow or interflow, originates below the surface but above the saturated zone. In SWAT, the kinematic storage model is used to model lateral flow in each soil layer, which accounts for variation in conductivity, slope and soil water content. In the model, the rate of lateral flow linearly depends on the soil hydraulic conductivity (SOL_K, Table 4.2).

Percolation.   Percolation occurs when the field capacity of the soil layer is exceeded. The percolation rate is governed by the saturated conductivity of the soil layer (SOL_K, Table 4.2). In SWAT, the lag between the time that water (percolating through soil layer) exits the soil profile and enters the shallow aquifer is modeled by the Exponential-Decay model with the delay time constant (GW_DELAY, Table 4.2).

Recharge.   Recharge is that water percolating through soil layer enters the groundwater aquifer. SWAT partitions the total daily recharge between shallow and deep aquifer according to a partition coefficient (RCHRG_DP, Table 4.2) defined by users (Figure 4.1).

Baseflow.   Baseflow or return flow, originates from groundwater. In SWAT, the shallow, unconfined aquifer contributes return flow to the main channel. The groundwater return flow

is calculated via an Exponential-Decay model with the baseflow recession constant (ALPHA_BF, Table 4.2).

Runoff concentration.   Runoff concentration is the runoff aggregation to basin outlet through sloping surface and channel network (Figure 4.1). The subbasin time of runoff concentration includes overland flow time and the tributary channel flow time, which are estimated using Manning's Formula with the Manning roughness constant of overland (OV_N, Table 4.2; Eq.(4.1:16)) and tributary channel (CH_N1, Table 4.2; Eq.(4.1:17)), respectively. The main channel routing is calculated by the variable storage routing method with the Manning roughness constant of main channel (CH_N2, Table 4.2; Eq.(4.1:20)).

Evapotranspiration. SWAT computes evaporation from soils and plants separately. Soil water evaporation is estimated by layered evaporation mode, which depends on the soil depth and water content. Plant transpiration is simulated as a linear function of potential evapotranspiration and leaf area index. In this study, the potential evapotranspiration (PET) is measured by pan evaporation equipment. In SWAT, the evaporation from the shallow aquifer is termed as "revap", which is estimated by a linear function of potential evapotranspiration with the conversion coefficient of GW_REVAP (Table 4.2).

4.1.3   Sediment Simulation

Sediment simulation involves two aspects: the erosion of soil particles from slopes and the sediment transport in the channel network. SWAT used the modified universal soil loss equation (MUSLE) to estimate the soil erosion from the HRU, and the simplified Bagnold equation to simulate the sediment transport in the channel network (Neitsch et al., 2005).

Soil erosion. The modified universal soil loss equation (MUSLE) is:

$$sed = 11.8 \times (Q_{surf} \times q_{peak} \times area_{hru})^{0.56} \times \text{USLE\_}K \times C \times P \times LS \times CFRG \qquad (4.1:13)$$

where

        sed is the sediment yield (t),

        q_{peak} is the peak runoff rate (m^3/s),

        $area_{hru}$ is the area of the HRU (ha),

        USLE_*K* is the soil erodibility factor (USLE_K, Table 4.2),

        C is the cover and management factor,

        P is the support practice factor,

LS is the topographic factor,

$CFRG$ is the coarse fragment factor.

The peak runoff rate (q_{peak}) is that the daily runoff rate multiplies by a adjustment factor:

$$q_{peak} = adj\_pkr \times \frac{Q_{surf} \times area_{hru}}{3.6 \times t_{conc}}$$

(4.1:14)

where

adj_pkr is peak rate adjustment factor (ADJ_PKR, Table 4.2),

t_{conc} is the time of concentration for the subbasin (h).

The time of concentration (t_{conc}) includes the overland flow time (t_{ov}) and the tributary channel flow time (t_{ch}) that are all estimated using Manning's equations (Figure 4.1):

$$t_{conc} = t_{ov} + t_{ch}$$

(4.1:15)

$$t_{ov} = \frac{L_{slp}^{0.6} \times n_{ov}^{0.6}}{18 \times slp^{0.3}}$$

(4.1:16)

$$t_{ch} = \frac{0.62 \times L_{ch} \times n_{ch1}^{0.75}}{area^{0.125} \times slp_{ch}^{0.375}}$$

(4.1:17)

where

t_{ov} is the overland flow time (h),

t_{ch} is the tributary channel flow time (h),

L_{slp} is the subbasin slope length (m),

slp is the average slope in the subbasin,

n_{ov} is Manning's roughness coefficient of the subbasin slope (OV_N, Table 4.2),

L_{ch} is the channel length from the most distant point to the subbasin outlet (m),

n_{ch1} is Manning's roughness coefficient of the tributary channel (CH_N1, Table 4.2),

$area$ is the subbasin area (ha),

slp_{ch} is the channel slope.

The amount of sediment released to the main channel (sed´ i.e. sediment lag in the surface runoff) is calculated by the Exponential-Decay model:

$$sed' = (sed + sed_{stor,i-1}) \times (1 - \exp(\frac{-surlag}{t_{conc}}))$$

(4.1:18)

where

sed´ is the sediment lag in the surface runoff (t),

$sed_{stor,i-1}$ is the sediment stored or lagged from the previous day (t),

$surlag$ is the surface runoff lag coefficient (SURLAG, Table 4.2).

Sediment transport. The simplified Bagnold equation is used to estimate sediment transport capacity (sed_{ch}) in the main channel, which assumes (1) the sediment transport capacity of channel is the power function of the peak flow velocity, and (2) the channel maintains equilibrium sediment transport:

$$sed_{ch} = spcon \times (prf \times \frac{q_{ch}}{A_{ch}})^{spexp} \times q_{ch} \qquad (4.1:19)$$

where

sed_{ch} is the sediment transport capacity of flow in the main channel (t/m^3),

$spcon$ is a coefficient defined by the user (SPCON, Table 4.2),

prf is the channel peak rate adjustment factor (PRF, Table 4.2),

A_{ch} is the cross-sectional area of flow in the channel (m^2),

$spexp$ is an exponent defined by the user (SPEXP, Table 4.2),

q_{ch} is the average rate of flow (m^3/s), which is calculated by Manning's equation:

$$q_{ch} = \frac{A_{ch} \times R_{ch}^{2/3} \times slp_{ch}^{1/2}}{n_{ch2}} \qquad (4.1:20)$$

where

R_{ch} is the hydraulic radius for a given depth of flow (m),

n_{ch2} is Manning's roughness coefficient of the main channel (CH_N2, Table 4.2).

When the suspended sediment from upstream is greater than the sediment transport capacity, the transport excess sediment will deposit in the main channel. Otherwise, the channel will be denuded to compensate the suspended sediment of the main channel for meeting the sediment transport capacity of flow Eq. (4.1:19).

4.2   Automatic Calibration Setup

The general procedure of automatic calibration is shown in Figure 2.1. The automatic calibration includes five components: hydrological model, input data, calibrated model-parameters, optimization tool, observed data and objective function. The structure of SWAT model used in this study has been introduced above. This section separately introduces input data, calibrated model parameters, optimization tool and observed data.

4.2.1   Input Data for SWAT Model

SWAT model is a distributed hydrological model, of which the smallest element is the HRU consisting of homogeneous land use, soil type and topography. For driving the SWAT model, a wide range of information, especially spatial data, such as topographic data, soil types and land use, is required. In the following, a brief description of the data collected for the Baocun watershed is given.

1. *DEM data.* The digital elevation model (DEM) with 30 m resolution is from the resampling of the 1:10 000 scale topographic map with 2.5 m contour interval (Figure 3.1 (b)). The watershed was divided into seven sub-basins according to the river system (Figure 3.1 (b)).

2. *Soil topographic index data/Soil data*. In the SWAT-WB-VSA, the soil type map is replaced by the soil topography index map to reflect the effect of topography on runoff. In the Baocun watershed, the distribution of soil types depends on topographic index (*TI*) (Figure 3.2 (b)), so soil-topographic indexes can be classified into ten units in terms of ten *TI* classes of equal area shown in Figure 4.5. The soil property data adopt field survey results, as detailed in Section 3.2.

3. *Land use data*. Although the land use is only cropland in the Baocun watershed, the crops cyclically vary because of the crop rotation method with farming three times in every two years. Therefore, the average annual land use map is charted to overcome the problem of land utilization change, according to the growth time of crops (Figure 3.3 (b)). The property data of crops/vegetation are extracted from the database by the GIS software (ArcSWAT).

4. *Weather data*. The daily data (from 1990 to 2011) of the maximum and minimum air temperature, wind speed, relative humidity and solar radiation are collected from three national weather stations (Weihai, Chengshantou and Shidao) that distance from the Baocun watershed about 35 km (black triangles in Figure 3.1 (a)). The precipitation data are observed at four rainfall gauges during 1990-2011 inside Baocun watershed (circles in Figure 3.1 (b)). The daily potential evapotranspiration (PET) uses the data measured by pan evaporation equipment at the Baocun hydrometric station (red star in Figure 3.1 (b)).

The ArcSWAT is a pre-processing software for SWAT model, which provides a graphical support for the disaggregation of watershed. This study used the ArcSWAT to organize the

input data and prepare the input files for running/driving the SWAT model. The detailed steps are shown as follows:

1. *Watershed Delineation.* Based on the DEM data, ArcSWAT automatically delineates subbasins, extracts river information (e.g. stream network) and estimates the subbasin parameters, such as the length and slope of channel and the areas of the subbasin.

2. *HRU analysis.* Based on the landuse, soil and slope map, ArcSWAT generates the HRUs via spatial analysis. The properties of soil and landuse can be specified by users or extracted from the database of ArcSWAT.

3. *Weather data definition.* The weather data can be generated by the weather generator or use the observed data.

4. *Output of SWAT input files.* ArcSWAT finally writes the GIS data into the SWAT input files in text format.

The SWAT model can be driven by these input files and output the model predictions. In this study, the SWAT input-files/input-data are the same, no matter what the form of model calibration is, such as the single-/multi-response calibration and the model calibration with different rainfall-runoff methods.

4.2.2   Selection of Model Parameters for Calibration

The parameters of hydrological model (SWAT) can be classified into physical and process parameters (Sorooshian and Gupta, 1995). The physical parameters (such as the areas of the subbasin and the length of channel) could be extracted by GIS (e.g. ArcSWAT). By contrast, the process parameters (mainly consisting of the coefficients and exponents of model equations) should be determined through model calibration. According to the previous studies (Wang et al., 2003; Hosseini et al., 2011; Güngör and Göncü, 2013), 19 hydrological (flow) parameters and 8 sediment parameters are selected to be identified by model calibration, uncertainty analysis and validation (Table 4.2). However, the parameters of surface runoff generation are different among the five rainfall-runoff methods (CN-Soil, CN-ET, G&A, WB and WB-VSA; Section (4.1.1)) in the SWAT, as detailed in the last column of Table 4.2. Note that the *EDC* (Table 4.2; Eq. (4.1:9)) of the WB approach is defined as the same value in all HRUs to facilitate model calibration in this study, which means the effects of topography on runoff are neglected. The parameters in Table 4.2 involve the following main hydrologic processes of SWAT except snow melting: evapotranspiration, surface runoff, soil

water movement, groundwater movement, tributary/main channel routing, soil erosion
(MUSLE) and sediment transport.

Table 4.2 Description of model parameters and their ranges.

Categories	Parameter [1]	Range		Equation	Definition
		Min	Max		
Evapotrans piration	v__ESCO	0.01	1	—	Soil evaporation compensation factor
	v__EPCO	0.01	1	—	Plant uptake compensation factor
Surface water	r__CN2	-0.9	1	(4.1:3)	Initial curve number for moisture condition II (only in the CN-soil, CN-ET and G&A approaches)
	v__CNCOEF	0.1	10	(4.1:5)	Plant ET curve number coefficient (only in the CN-ET approach)
	v__EDC [2]	0	1	(4.1:9)	Effective depth of the soil profile (only in the WB and WB-VSA approaches)
	v__OV_N	0.005	0.5	(4.1:16)	Roughness coefficient of sloping surface
	v__SURLAG	0	24	(4.1:18)	Surface runoff lag coefficient
Soil water	r__SOL_Z	-0.9	2	—	Soil thickness (mm)
	r__SOL_BD	-0.6	1	—	Moist bulk density (g cm^{-3})
	r__SOL_AWC	-0.99	3	—	Available water capacity of the soil layer
	r__SOL_K	-0.99	10	(4.1:6)	Saturated hydraulic conductivity (mm/d)
Ground water	v__GW_DELAY	0	60	—	Groundwater delay time (d)
	v__ALPHA_BF	0	1	—	Baseflow recession constant
	v__GWQMN	0	1000	—	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)
	v__RCHRG_DP	0	1	—	Deep aquifer percolation fraction
	v__REVAPMN	0	1000	—	Threshold depth of water in the shallow aquifer for revaporization (mm)
	v__GW_REVAP	0.02	0.2	—	Groundwater revaporization coefficient
Tributary/ main channel	v__CH_N1	0.005	0.15	(4.1:17)	Roughness coefficient of tributary channels
	v__CH_N2	0.005	0.15	(4.1:20)	Roughness coefficient of main channels
MUSLE	v__USLE_K1	0	1	(4.1:13)	Regosols erodibility factor
	v__USLE_K2	0	1	(4.1:13)	Luvisols erodibility factor
	v__USLE_K3	0	1	(4.1:13)	Fluvisols erodibility factor
	v__ADJ_PKR	0	10	(4.1:14)	HRU peak rate adjustment factor
	v__PRF	0	10	(4.1:19)	Main channel peak rate adjustment factor
Sediment transport	v__SPCON	0.0001	0.1	(4.1:19)	Linear coefficient in sediment transport
	v__SPEXP	0.0001	6	(4.1:19)	Exponential coefficient in sediment transport
	v__CH_EROD	0	1	—	Channel erodibility factor

[1] "v__" indicates the active value replaces the initial value; "r__" indicates a relative change

to the initial value, i.e. the active value adds one and then multiplies the initial value.

[2] Note that the *EDC* of the WB approach is defined as the same value in all HRUs.

Because the soil-related parameters in different soil types or soil layers need to be calibrated independently, resulting in a large number of parameters, the so-called aggregated parameters will be selected instead of the original parameters. In this way, full use can be made of information about spatial variation, the number of parameters can be reduced, and the calibration process simplified (Yang et al., 2007a). Aggregated parameters are formed by adding a modification term which includes two types: v__ and r__, referring to a replacement and a relative change to the initial parameter, respectively. The aggregate parameters constructed in this study are shown in the second column of Table 4.2. For example, the v__ESCO is the value of parameter ESCO, and the r__SOL_K is the relative change to the initial SOL_K: $r\_\_SOL\_K = SOL\_K_{now}/ SOL\_K_{initial} - 1$, where SOL_K_{now} and $SOL\_K_{initial}$ are the active and the initial soil hydraulic conductivity (SOL_K), respectively.

In Table 4.2, the "min" and "max" columns are the maximum and minimum parameter values, respectively, where the ranges of the aggregate parameters are mainly based on previous studies (e.g. Abbaspour, 2009; Güngör and Göncü, 2013). The characteristics of some parameters (which are very important but easily misunderstood) are described in detail as follows:

v__ESCO affects the depth distribution of soil evaporation. The smaller the value of v__ESCO, the more the potential soil evaporation is extracted from lower soil layers.

v__EPCO affects the plant uptake from the soil layer. The larger the value of v__EPCO, the more the potential plant transpiration is uptaken from lower soil layers.

r__CN2 is the relative change of the initial curve number for moisture condition II (CN2), which varies with different soil types and land use (i.e. HRUs). The ArcSWAT provides the reference values of CN2 for each HRU. In the curve number approaches (CN-Soil and CN-ET), the larger the value of r__CN2, the more the surface water is yielded. In the G&A approach, the CN2 is only used to calculate the initial soil moisture.

v__EDC is the catchment average value of the effective depth of soil profile, which ranges from 0 to 1. When v__EDC approaches 1, the effect of topography on runoff can be neglected. The larger the value of v__EDC, the less the surface runoff is.

v__GW_DELAY is the value of the delay time of groundwater recharge from the soil layer to the groundwater aquifer. The smaller the value of v__GW_DELAY, the shorter the delay time of groundwater recharge is.

v__ALPHA_BF is the value of the decay constant of Exponential-Decay model for baseflow. The larger the value of v__ALPHA_BF, the faster the groundwater decays.

v__RCHRG_DP is the value of the coefficient of percolation from the shallow aquifer to the deep aquifer. The larger the value of v__RCHRG_DP, the more the amount of percolation is.

v__USLE_K is the value of soil erodibility factor, which is the soil characteristic parameter and varies as soil types. The larger the value of v__ USLE_K, the stronger the soil erodibility is.

v__CH_EROD is the value of channel erodibility factor, which affects the soil erosion from the main channel. The larger the value of v__CH_EROD, the more the amount of sediment re-entrained from main channels is.

In SWAT, the river sediment loads depend on the surface runoff and river discharge (Section 4.1.3). In Table 4.2, the four flow parameters (i.e. OV_N, SURLAG, CH_N1 and CH_N2) directly reflect the effect of the surface runoff and river discharge on the river sediment loads.

4.2.3   Optimization Tool

In this study, a number of model parameters (19 flow parameters and 8 sediment parameters shown in Table 4.2) need to be calibrated. Although the Shuffled Complex Evolution (SCE-UA) method is widely used for model calibration, the convergence rate of SCE-UA would reduce rapidly with the increase of the number of model parameters (Duan, 2003; Tolson and Shoemaker, 2007; Vrugt et al., 2009b; Zambrano-Bigiarini and Rojas, 2013). Additionally, the SCE-UA was only designed for searching the global optimization solution, but neglects the equifinality and uncertainty of hydrological model parameters that are the generic problems in the hydrological models (Beven and Binley, 2013). Therefore, Vrugt et al. (2003) proposed a Shuffled Complex Evolution Metropolis global optimization algorithm (SCEM-UA) that replaces the local search method (Simplex method) in the SCE-UA by the Simulated Annealing (SA) method. Further, based on the Bayesian theory, Vrugt et al. (2009b) developed the Differential Evolution Adaptive Metropolis (DREAM) Markov chain Monte Carlo (MCMC) algorithm (Section 2.2.1).

Figure 2.2 shows the flow chart of the DREAM scheme. Compared with the classical calibration method (SCE-UA), the DREAM can not-only efficiently handle problems of model calibration involving high-dimensionality (i.e. numerous model parameters) and

nonlinearity (Vrugt et al., 2009b), but also estimate the posterior distribution (i.e. uncertainty) of model parameters. Recently, DREAM has become a widely used tool for model calibration and model uncertainty analysis (McMillan and Clark, 2009; Laloy et al., 2010; Schoups and Vrugt, 2010; Smith et al., 2010; Montanari and Koutsoyiannis, 2012).

In this study, the DREAM is used as the optimization tool for model calibration and parameter uncertainty analysis. We select eight parallel chains and a total of 40,000 model evaluations (i.e. MCMC iterations) for the DREAM algorithm parameterization on R platform, and run them on the "Soroban" High-Performance Computing System at Freie Universität Berlin. Note that the prior distributions (i.e. prior knowledge) of model parameters in Bayesian equation ($\pi(\theta)$; Eq. (2.2:1)) are all set to the uniform distributions, i.e. non-informative prior distributions.

4.2.4   Observed Data

The river discharge data within the period of 1993-2011 (Figure 3.6; Section 3.5.1) and sediment load data within the flood season (from June to September) of 1993-1999 (Figure 3.9; Section 3.5.4) collected from the Baocun hydrometric station (red star in Figure 3.1 (b)) are used as observed data/outcomes for Bayesian inference (or model calibration). A warm-up period is recommended by SWAT in order to initialize and then obtain reasonable starting values for model variables. The model running period is from 1990 to 2011, and the warm-up period is from 1990 to 1992. The model can be validated using groundwater level data measured in the watershed (Figure 3.1 (b)) from the period of June 2007 to December 2011 (Figure 3.8; Section 3.5.3).

4.3   Objective Functions

This study focuses on three cases of automatic calibration: likelihood function comparison, multi-response calibration and model comparison. The likelihood function comparison attempts to evaluate the effect of the form of likelihood functions on model calibration. The multi-response calibration simultaneously optimizes the flow and sediment parameters to improve the performance of sediment simulation and compare the effect of different likelihood functions. The model comparison evaluates the performance of five rainfall-runoff method in SWAT after model calibration using the same objective function. Comparison of the automatic calibration scheme (Figure 2.1) among the three case studies is shown in Table

4.3. Note that the SWAT input data and the optimization tool are the same among the three case studies.

In Table 4.3, the hydrological model, calibrated parameter and observed data have been introduced in Section 4.2. The next sections will separately introduce the objective functions of the three case studies.

Table 4.3 Automatic calibration scheme of the different case studies.

Cases	Hydrological models	Calibrated parameters	Observed data	Objective function
Likelihood function comparison	SWAT-WB-VSA	Flow	River discharges	*NSE*, BC-GED, BC-SGED
Multi-response calibration	SWAT-WB-VSA	Flow; Sediment	River discharges; Sediment loads	*NSE*, BC-GED
Model comparison	CN-Soil, CN-ET, G&A, WB, WB-VSA	Flow	River discharges	VE

4.3.1   Improvement and Comparison of Likelihood Functions

The likelihood function of a set of model parameter values (θ) is equal to the joint probability of the observed outcomes (Section 2.2.2; Figure 2.4). In the scheme of DREAM, the likelihood function is used to ascertain the goodness-of-fit of hydrologic model, which is the special objective function for automatic calibration. This section attempts to build a relationship between the likelihood function and the most widely used objective function (i.e. Nash–Sutcliffe efficiency coefficient (*NSE*) defined by Nash and Sutcliffe (1970)), and develop a formal likelihood function.

4.3.1.1   Interpretation of the *NSE* from the Likelihood Function Viewpoint

The errors/residuals (e) between the observed and simulated outcomes are treated as random variables:

$$e_i = obs_i - sim_i \qquad\qquad (4.3{:}1)$$

where

obs_i and sim_i are the observed and simulated outcomes at time step i, respectively.

If the error e_i is assumed to be independent and identically distributed (I.I.D.) according to Gaussian distribution with zero mean and a constant variance (Figure 2.4), the probability density function (PDF) of e_i is:

$$p\left(e_i|\theta\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{e_i^2}{2\sigma^2}\right) \tag{4.3:2}$$

where

σ is the standard deviation.

Then the logarithmic likelihood function of parameter set (θ) can be expressed as (Eq.(2.2:3)):

$$l\left(\theta|obs\right) = \ln(\prod_{i=1}^{n} p\left(e_i|\theta\right)) = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2\right) - \sum_{i=1}^{n}\frac{e_i^2}{2\sigma^2} \tag{4.3:3}$$

where

n is the length of time series of errors.

When

$$\sigma^2 = \sum_{1}^{n} e_i^2 / n \tag{4.3:4}$$

The likelihood function (Eq. (4.3:3)) reaches the maximum value. So:

$$l\left(\theta|obs\right)_{max} = -\frac{n}{2}\ln\left(2\pi\varepsilon\right) - \frac{n}{2}\ln\left(\sum_{1}^{n}\frac{e_i^2}{n}\right) \tag{4.3:5}$$

where

ε is the base of the natural logarithm, ≈ 2.718.

NSE (Nash and Sutcliffe, 1970) is:

$$NSE = 1 - \frac{\sum_{1}^{n}(sim_i - obs_i)^2}{\sum_{1}^{n}\left(obs_i - \overline{obs}\right)^2} = 1 - \frac{\sum_{1}^{n}e_i^2}{\sum_{1}^{n}\left(obs_i - \overline{obs}\right)^2} \tag{4.3:6}$$

where

$\overline{obs}$ is the mean observed outcomes.

From Eq. (4.3:6) we can obtain:

$$\sum_{1}^{n}\frac{e_i^2}{n} = \left(1 - NSE\right)\frac{n-1}{n}\sigma_{obs}^2 \tag{4.3:7}$$

where

σ_{obs} is the standard deviation of observed outcomes.

By substituting Eq. (4.3:7) into Eq. (4.3:5), the logarithmic likelihood function is transformed to:

$$l(\theta|obs)_{max} = -\frac{n}{2}\ln\left(2\pi\varepsilon\frac{n-1}{n}\sigma_{obs}^2\right) - \frac{n}{2}\ln(1-NSE) \qquad (4.3:8)$$

In Eq. (4.3:8), the $-\frac{n}{2}\ln\left(2\pi\varepsilon\frac{n-1}{n}\sigma_{obs}^2\right)$ is constant, so the likelihood function is equivalent to *NSE*. In other words, the *NSE* is equivalent to a likelihood function under the assumption that the errors between observed and simulated outcomes follow the Gaussian error distribution with zero mean and a constant variance. In the *NSE* approach, the standard deviation of model residuals (σ) is estimated by the unbiased equation (Eq. (4.3:4)) that is also used by Vrugt et al. (2009a) and Laloy et al. (2010).

4.3.1.2   Formal Likelihood Functions

4.3.1.2.1   Removal of the Heteroscedasticity of Model Residuals

Errors/residuals between observed and simulated river discharges typically exhibit considerable heteroscedasticity, autocorrelation and non-normality (Evin et al., 2013). These error characteristics (correlation, heteroscedasticity, etc.) need to be explicitly accounted for before calculation of the likelihood function. In the Baocun watershed, the floods present impulse form in daily river discharge chart shown in Figure 3.6, because flood concentration time is only two hours, and rainfall events are short but very intense owing to typhoon storms. This implies that the autocorrelation of river discharges at daily time step is relatively weak, or equivalently, the autocorrelation of errors between daily observation and simulation should be weak for a good predictive performance. A graphic check of the autocorrelation of model residuals confirms this assumption (results are shown in Appendix A). Therefore, the errors' autocorrelation will be ignored in this study, which is adopted by many authors (Thiemann et al., 2001; Smith et al., 2010; Pianosi and Raso, 2012). Evin et al. (2014) pointed out that the likelihood function incorporating autocorrelation has disadvantages on the estimated parameter uncertainty in some cases. However the errors' heteroscedasticity resulted from larger rainfalls and streamflows in Figure 3.6 should be accounted for in the Baocun watershed (Evin et al., 2013).

The Box-Cox transformation (BC) method is widely used to remove heteroscedasticity of errors (Yang et al., 2007a, b; Smith et al., 2010; Li et al., 2011, 2013). One-parameter Box-Cox transformation function (Box and Cox, 1964) is:

$$g(y) = \frac{y^\lambda - 1}{\lambda} \qquad 0 < \lambda \leq 1 \tag{4.3:9}$$

where

λ is the Box-Cox transformation parameter (lambda),

y is simulated or observed outcomes.

The errors/residuals (e) between the observed and simulated outcomes after Box-Cox transformation are expressed as:

$$e_i = \frac{obs_i^\lambda - sim_i^\lambda}{\lambda} \qquad 0 < \lambda \leq 1 \tag{4.3:10}$$

λ in Eq. (4.3:10) heavily affects the efficiency of the Box-Cox transformation method. According to previous studies (Bates and Campbell, 2001; Yang et al., 2007a, b), the likelihood function should be changed for estimating λ:

$$l(\theta|obs) = l'(\theta|obs) + (\lambda - 1)\sum_1^n \ln(obs_i) \tag{4.3:11}$$

where

$l'(\theta|obs)$ is the original likelihood function (Eq. (2.2:3)).

Eq. (4.3:11) indicates that if λ approaches zero, the likelihood function reaches the maximum value (positive infinity) for the small observed outcomes (obs_i), e.g. zero. In other words, the inference result of λ has always approached zero (Yang et al., 2007b), no matter what the parameter sets are. So the constraint (Eq. (4.3:11)) of λ is invalid, probably because the Eq. (4.3:11) incorrectly transplanted the Jacobian determinant from the original BC transformation (Eq. (4.3:9)) method to the hydrological BC transformation (Eq. (4.3:10)) method, as detailed in Appendix B. The Jacobian determinant of the transformation from the model residual after BC transformation (Eq. (4.3:10)) to the raw model residual (Eq. (4.3:1)) may be non-existent.

However, if there is no constraint on the value of λ (i.e. totally treating λ as a hydrological model parameter), the inference result of λ will always approach one (i.e. no transformation of model residuals) when there are many small observed outcomes (Laloy et al., 2010). It probably results from that the mode (i.e. the highest probability point) of errors/residuals is

zero, and the no-transformation of model residuals (closed to zero) contributes to the maximization of the likelihood function. Therefore, treating λ as an additional inference parameter of the likelihood function cannot yield an effective Box-Cox transformation parameter for removal of the heteroscedasticity of model residuals at the moment.

Another idea is the independent calculation of λ with each MCMC iteration, where the simulated outcomes were yielded by hydrologic model with fixing parameters. For this purpose, a constraint is introduced for calculation of λ to minimize the variance of the time series of errors/residuals:

$$\min\left(variance\left(e_i\right)\right) = \frac{1}{n-1}\sum_1^n \left(\frac{obs_i^\lambda - sim_i^\lambda}{\lambda} - \mu\right)^2 \big|_{min} \tag{4.3:12}$$

where

μ is the mean of errors after Box-Cox transformation (Eq. (4.3:10)).

4.3.1.2.2 Error Distribution Model

For the non-Gaussian distribution, the generalized error distribution (GED) and the skew generalized error distribution (SGED) are selected for comparison.

The generalized error distribution (GED) (Thiemann et al., 2001; McMillan and Clark, 2009) is expressed as:

$$p\left(e_i | \theta\right) = \frac{\omega(\beta)}{\sigma} \exp\left(-\left|c(\beta)\frac{e_i - \mu}{\sigma}\right|^\beta\right) \tag{4.3:13}$$

where

$$c\left(\beta\right) = \sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}},$$

$$\omega\left(\beta\right) = \frac{\beta c\left(\beta\right)}{2\Gamma[1/\beta]},$$

β termed kurtosis is a parameter of the probability density function of GED ($\beta > 0$),

$\Gamma[\mathrm{x}]$ is the gamma function evaluated at x.

Note that the formulation of GED (Eq. (4.3:13)) is a little difference from that of the exponential power distribution proposed by Thiemann et al., (2001) and McMillan and Clark (2009), which replaces the "β" in GED by the "$2/(1+\beta')$". GED is more flexible than Gaussian distribution (Figure 4.6):

1. When $\beta = 2$, GED becomes the Gaussian distribution;

2. When $\beta = 1$, GED is a Laplace distribution;

3. GED approaches a uniform distribution as β approaches infinity.

However, GED is still a symmetric error distribution. Schoups and Vrugt (2010) used a more flexible error distribution— skew generalized error distribution (SGED) — which is developed from the GED:

$$p\left(e_i|\theta\right) = \frac{2\omega(\beta)\sigma_\xi}{\sigma(\xi + \xi^{-1})} \exp(- \left| c(\beta)\xi^{-\text{sign}(\mu_\xi + \sigma_\xi \frac{e_i - \mu}{\sigma})} (\mu_\xi + \sigma_\xi \frac{e_i - \mu}{\sigma}) \right|^{\beta}) \qquad (4.3\text{:}14)$$

where

$$\mu_\xi = M_1(\xi - \xi^{-1}),$$

$$\sigma_\xi = \sqrt{\left(1 - M_1^2\right)\left(\xi^2 + \xi^{-2}\right) + 2M_1^2 - 1},$$

$$M_1 = \frac{\Gamma[2/\beta]}{\sqrt{\Gamma[3/\beta]\Gamma[1/\beta]}},$$

ξ termed skewness is a parameter of SGED ($\xi > 0$).

SGED is positively (negatively) skewed for $\xi > 1$ ($\xi < 1$), and when $\xi = 1$, SGED is symmetric and becomes the GED. The fGarch package in R provides the functions of the GED and SGED (Würtz et al., 2013). The shapes of probability density of SGED with several of the kurtosis (β) and skewness (ξ) parameters are shown in Figure 4.6.
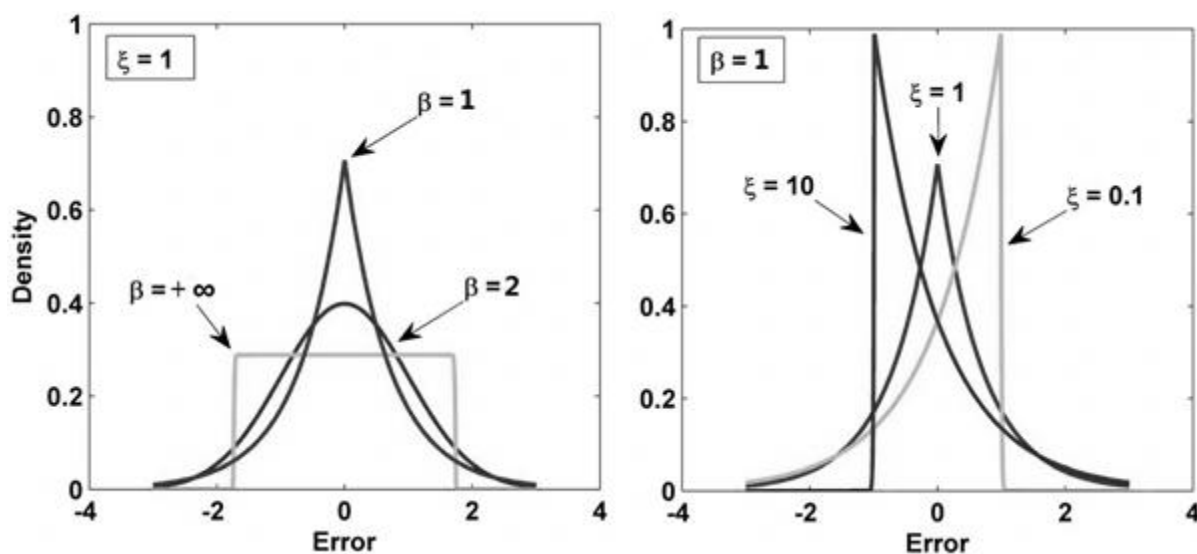
Figure 4.6 Probability density of the SGED with zero-mean and unit standard deviation for several of the kurtosis (β) and skewness (ξ) parameters (modified from Schoups and Vrugt (2010)).

In summary, this section introduced three likelihood functions (also termed error models): *NSE*, generalized error distribution with Box-Cox transformation (BC-GED) and skew generalized error distribution with Box-Cox transformation (BC-SGED) approaches. All the likelihood functions have special statistical assumptions (Section 2.2.2). The statistical assumptions of the three likelihood functions are summarized in Table 4.4. This table shows all the likelihood functions assume the model residuals are independent. The *NSE* and BC-GED are the special cases of BC-SGED. *NSE* is also a special case of BC-GED. However, the number of error model parameters of BC-SGED (λ, μ, σ, β and ξ) is more than that of BC-GED (λ, μ, σ and β), and more than that of *NSE* (parameter-free).

Table 4.4 Comparison of the statistical assumptions of likelihood function among *NSE*, BC-GED and BC-SGED approaches

Likelihood function	Statistical assumption	Errors' distribution [1]
NSE	Independent and homoscedastic	Gaussian with zero mean (= SGED(μ=0, β=2, ξ=1))
BC-GED	Independent and homoscedastic after Box-Cox transformation	GED (= SGED(ξ=1)) after Box-Cox transformation
BC-SGED	Independent and homoscedastic after Box-Cox transformation	SGED after Box-Cox transformation

[1] μ is the mean of SGED; β is the kurtosis coefficient; ξ is the skewness coefficient.

4.3.2   Multi-response Likelihood Function

The sediment loads in the watershed outlet are closely related to the surface runoff of slopes and the flow velocity of channels. However, most previous studies separately calibrate flow and sediment objectives, following the calibration procedure of first flow then sediment, for simplifying the model calibration processes. Obviously, this way impacts the prediction precision of sediment loads. This section extends the likelihood functions (Table 4.4) to include the river flow and sediment objectives for multi-response calibration based on the assumption of independence between the river flow and sediment model residuals.

4.3.2.1   Combination of the Flow and Sediment Objectives into a Likelihood Function

Assuming the errors/residuals (e_i) between observed and simulated river flow discharges follow the distribution of $p(e_{i,flow}|\theta)$, the logarithmic/log- likelihood function of flow can be expressed:

$$l_{flow}\left(\theta\,|\,obs_{flow}\right) = \ln(\prod_{i=1}^{n} p\left(e_{i,flow}\,|\,\theta\right)) = \sum_{1}^{n} \ln(p\left(e_{i,flow}\,|\,\theta\right)) \qquad (4.3:15)$$

where

$e_{i,flow}$ is the error/residual for flow at time step i,

n is the length of time series of river flow discharges.

Similarly, if the errors/residuals (e_i) for sediment follow the distribution of $p'(e_{i,sed}\,|\,\theta)$, the log-likelihood function of sediment is:

$$l_{sed}\left(\theta\,|\,obs_{sed}\right) = \ln(\prod_{i=1}^{m} p'\left(e_{i,sed}\,|\,\theta\right)) = \sum_{1}^{m} \ln(p'\left(e_{i,sed}\,|\,\theta\right)) \qquad (4.3:16)$$

where

$e_{i,sed}$ is the error/residual between the observed and the simulated sediment load (i.e. sediment) at time step i,

m is the length of time series of sediment.

If the errors of sediment ($e_{i,sed}$) are independent of the errors of flow ($e_{i,flow}$), the log-likelihood function of flow and sediment (i.e. the joint probability of the flow and sediment errors) can be described as:

$$l\left(\theta\,|\,obs\right) = l_{flow}\left(\theta\,|\,obs_{flow}\right) + l_{sed}\left(\theta\,|\,obs_{sed}\right) \qquad (4.3:17)$$

where

$l(\theta|obs)$ is the log-likelihood function of flow and sediment,

$l_{flow}(\theta|obs_{flow})$ is the log-likelihood function of flow,

$l_{sed}(\theta|obs_{sed})$ is the log-likelihood function of sediment.

4.3.2.2 Case with the *NSE* Approach

The Nash–Sutcliffe efficiency coefficient (Nash and Sutcliffe 1970) is equivalent to a kind of likelihood function with the assumption that the errors follow a Gaussian distribution with zero mean, as detailed in Section 4.3.1.1. By substituting *NSE* likelihood function (Eq. (4.3:8)) into Eq. (4.3:17), the log-likelihood function is transformed to:

$$\begin{aligned} l\left(\theta\,|\,obs\right) = &-\frac{n}{2}\ln\left(2\pi\varepsilon\frac{n-1}{n}\sigma_{obs,flow}^2\right) - \frac{n}{2}\ln\left(1 - NSE_{flow}\right) \\ &-\frac{m}{2}\ln\left(2\pi\varepsilon\frac{m-1}{m}\sigma_{obs,sed}^2\right) - \frac{m}{2}\ln\left(1 - NSE_{sed}\right) \end{aligned} \qquad (4.3:18)$$

where

ε is the base of the natural logarithm, ≈ 2.718,

$\sigma_{obs,flow}$ and $\sigma_{obs,sed}$ are the standard deviation of the observed river flow discharges and sediment loads, respectively,

NSE_{flow} and NSE_{sed} are the Nash–Sutcliffe efficiency coefficient for flow and sediment, respectively.

4.3.2.3 Case with the Formal Likelihood Function

The BC-GED approach firstly uses the Box-Cox (BC) transformation method (Eq. (4.3:10)) to remove the heteroscedasticity of errors, and then uses the GED to fit the probability distribution of the errors after BC transformation (Eq. (4.3:13)), as detailed in the Section 4.3.1.2. Unlike the flow series (Figure 3.6) in the Baocun watershed (humid area), the sequences of sediment loads is discrete (Figure 3.9), of which the sediment transports only occur in the flood period. As a result, the errors/residuals between the observed and the simulated sediment loads in the non-flood state are always minus values. The characteristic of sediment loads in Baocun watershed is similar to that of river discharges in the arid region, where the river often dries up and the time series of river discharges is discrete (Smith et al., 2010). Smith et al. (2010) pointed out that the model residuals of different flow states (e.g. zero/non-zero flow state) are from different sampling methods and should be treated separately.

Smith et al. (2010) separated model residuals into zero error and non-zero error components based on zero/non-zero flow state. However, for sediment loads in the Baocun watershed, the number of the sediment data in the non-zero state is too few, which is not enough for the statistical analysis of model residuals (Section 3.5.4). Therefore, this study assumes:

1. The residuals/errors of sediment loads in the flood state (including some residuals in the zero-sediment state) are independent and identically distributed (I.I.D).
2. The probability of the residuals of sediment loads in the non-flood state can be estimated by the error distribution concluded using the residuals in the flood state.

Then, the log-likelihood function for sediment can be expressed:

$$l_{sed}\left(\theta \mid obs\right) = \sum_{1}^{m_1} \ln(\mathrm{GED}'\left(e'_{i,sed} \mid \theta\right)) + \sum_{1}^{m_2} \ln(\mathrm{GED}'\left(e^*_{i,sed} \mid \theta\right)) \qquad (4.3{:}19)$$

where

$e'_{i,sed}$ and $e*_{i,sed}$ are the residuals of sediment loads after Box-Cox transformation (Eq. (4.3:10)) in the flood and the non-flood state, respectively,

GED'(x) is the generalized error distribution (GED) function (evaluated at x) (Eq. (4.3:13)) estimated using the residuals in the flood state,

m_1 and m_2 are the number of residuals in the flood and the non-flood state, respectively.

In this study, the flood state is defined that the river discharge is greater than 1.5 m³/s. Otherwise, it is non-flood state. By augmenting the log-likelihood function for sediment (Eq. (4.3:19)) with the log-likelihood function for flow, the log-likelihood function (Eq. (4.3:17)) for flow and sediment is become:

$$l\left(\theta \mid obs\right) = \sum_{1}^{n} \ln\left(\mathrm{GED}\left(e'_{i,flow} \mid \theta\right)\right) + l_{sed}\left(\theta \mid obs\right) \tag{4.3:20}$$

where

$e'_{i,flow}$ is the residuals of river discharges after Box-Cox transformation (Eq. (4.3:10)),

GED(x) is the generalized error distribution (GED) function evaluated at x for flow,

l_{sed} $(\theta|obs)$ is the log-likelihood function of sediment model residuals (Eq. (4.3:19)).

4.3.3 Efficiency Coefficient for Model Comparison

The purpose of model comparison is to evaluate the performance of different models. The Nash–Sutcliffe efficiency coefficient (*NSE*) is popularly used as the performance criterion for model comparison (Nash and Sutcliffe 1970). However, the *NSE* as the objective function always puts greater emphasis on high flow at the expense of the low values (Legates and McCabe, 1999; Krause et al., 2005; Guinot et al., 2011; Pushpalatha et al., 2012). In order to overcome the problem of *NSE*, Criss and Winston (2008) proposed a normalization of MAE (Table 2.1, ID 4), i.e. 'volumetric' efficiency (VE), to replace *NSE*:

$$VE = 1 - \frac{\sum_i |e_i|}{\sum_i obs_i} = 1 - \frac{\sum_i |obs_i - sim_i|}{\sum_i obs_i} \tag{4.3:21}$$

where

obs_i and sim_i are the observed and simulated outcomes at time step i, respectively,

e_i is the errors/residuals between the observed and simulated outcomes,

Σ is the summation operator.

Similar to the *NSE*, the VE is dimensionless and ranges from minus infinite to one in theory. If the VE is one, the hydrologic model is perfect. However, if the VE is less than zero, the hydrologic model is no better than a predictor using zeros.

Although Criss and Winston (2008) claimed that the VE is better than *NSE*, the VE lacked the practical applications and theoretical support. In this study, the VE is used as the objective function to calibrate the five rainfall-runoff methods (Table 4.1) in SWAT for model comparison and the validation of VE.

5 Results

The results of model calibration depend on the objective function. Different objective functions are in favor of different purposes of the modeling exercise (Green and Stephenson, 1986). This Chapter presents the model calibration results of three cases for different hydrological modeling applications: (1) the results of Bayesian inference with three likelihood functions (*NSE*, BC-GED and BC-SGED; Section 4.3.1) for likelihood function comparison (Section 5.1), (2) the results of multi-response calibration with two likelihood functions (*NSE* and BC-GED; Section 4.3.2) for simultaneous optimization of flow and sediment parameters (Section 5.2) and (3) the results of model calibration with the efficiency coefficient (VE; Section 4.3.3) for model comparison (Section 5.3).

5.1   Bayesian Inference with Different Likelihood Functions

The equifinality of different model parameter sets is a generic problem of hydrological model (Beven and Binley, 2013). For accounting for the equifinality problem, the Bayesian approach is proposed to automatically calibrate hydrological model (Beven and Binley, 2013), which estimates not only the optimal value of model parameters, but also the probability distribution (i.e. uncertainty) of model parameters. The likelihood function, defined as the joint probability of model residuals, can be treated as a special objective function in the DREAM.

The Bayesian inferences (or calibration results) of Bayesian approach with three likelihood functions (*NSE*, BC-GED and BC-SGED approaches that are proposed in Section 4.3.1 and compared in Table 4.4) are presented in this Section. In the automatic calibration procedure (Figure 2.1), SWAT-WB-VSA (developed in this study; Section 4.1) is used as the hydrological model tool, the river discharges (Section 3.5.1) as the observed data/outcomes, the DREAM (Section 2.2.1) as the Bayesian (optimization) tool and the flow parameters (Table 4.2) as the calibrated parameters, as detailed in Table 4.3 for the case of likelihood function comparison.

DREAM is a kind of Markov Chain Monte Carlo (MCMC) program that uses the stochastic simulation algorithm to solve Bayesian function (Eq. (2.2:1); Section 2.2.1; Figure 2.2). However, the calculation methods of likelihood function among the *NSE*, BC-GED and BC-SGED approaches (also termed error model) are different.

NSE approach. The NSE approach first calculates the value of *NSE* (Eq. (4.3:6)) for the model parameter set (θ), and then substitutes *NSE* in Eq. (4.3:8) to calculate the value of the likelihood function at each MCMC iteration.

BC-GED approach. The BC-GED approach first uses model predictions yielded by SWAT-WB-VSA model to calculate the value of Box-Cox (BC) transformation parameter (λ/lambda) by the least squares method based on the minimum variance constraint (Eq. (4.3:12)), then calculates errors/residuals after Box-Cox transformation (Eq. (4.3:10)), and finally uses the generalized error distribution (GED) function (Eq. (4.3:13)) to calculate the logarithmic likelihood value ($l(\theta|obs)$) of errors at each MCMC iteration.

BC-SGED approach. The calculating procedure of the BC-SGED approach is similar to that of the BC-GED approach, and the only difference being that the probability distribution of errors in the BC-GED approach is replaced by the skew generalized error distribution (SGED) (Eq. (4.3:14)).

The Bayesian inferences (or calibration results) are presented from two aspects: the simulation results and the posterior distribution of model parameters.

5.1.1 Simulation Results

The simulation results are produced by hydrological model (SWAT-WB-VSA) with the optimal values of model parameters corresponding to the maximum value of the likelihood function (Eq. (2.2:3)). For each likelihood function, firstly the performance (goodness-of-fit) of simulation results is presented, and then the statistical characteristics (mainly consisting of heteroscedasticity (i.e. non-constant variance) and error distribution) of model residuals (between the model predictions and observed data) are inspected.

5.1.1.1 Model Performance

Comparison of the observed and simulated river discharges on the daily time step is shown in Figure 5.1. The daily river discharges from 1993 to 2011 are collected from the Baocun hydrometric station located at the outlet of Baocun watershed (Figure 3.1 (b)). The simulated data are produced by SWAT-WB-VSA using the optimal values of model parameters inferred by NSE/BC-GED/BC-SGED approach. In Figure 5.1, *NSE* and R-square (R^2) are the Nash–

Sutcliffe efficiency coefficient and the coefficient of determination, respectively, which are widely used to assess the predictive power of hydrological models (Beven and Binley, 2013).
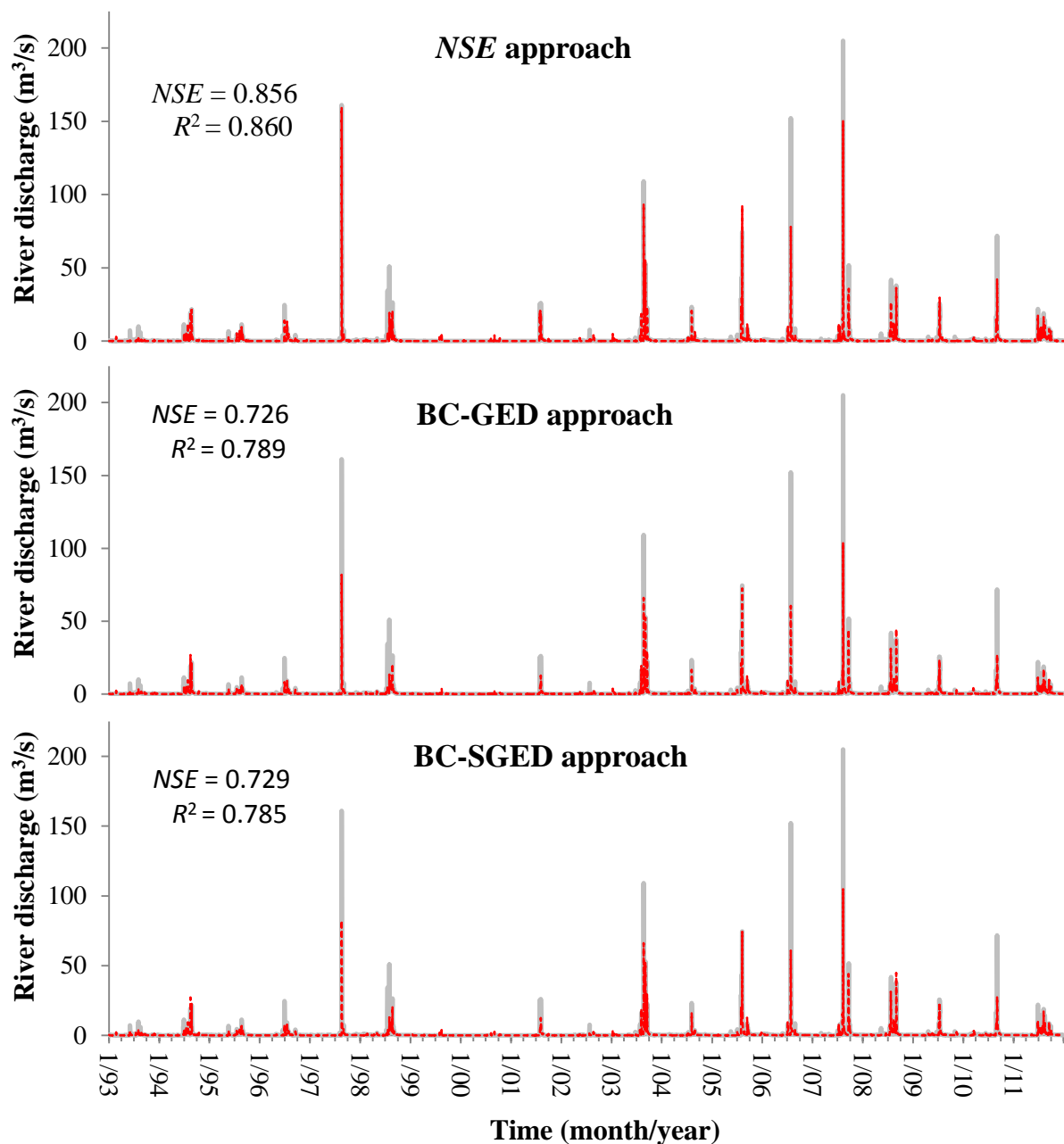
Figure 5.1 Comparison of the observed (black solid line) and simulated (red dot line) river discharges for *NSE*, BC-GED and BC-SGED approaches.

Figure 5.1 shows that, with the *NSE* approach, the hydrological model mimics the observed river discharges best, which reproduces most major flood events. For the BC-GED approach, the simulated results can reproduce most flood events, but in some cases the flood peaks are smaller than those of observed values. The simulated results of the BC-SGED approach are similar to those of the BC-GED approach.

Because of the considerable variance of the values of observed river discharges in Baocun watershed (Section 3.5.1; Figure 3.6), Figure 5.1 is difficult to reflect the baseflow simulation results. Therefore, for highlighting baseflow, the observed and simulated river discharges are plotted in Figure 5.2 with logarithmic vertical/y-axis (base 10).
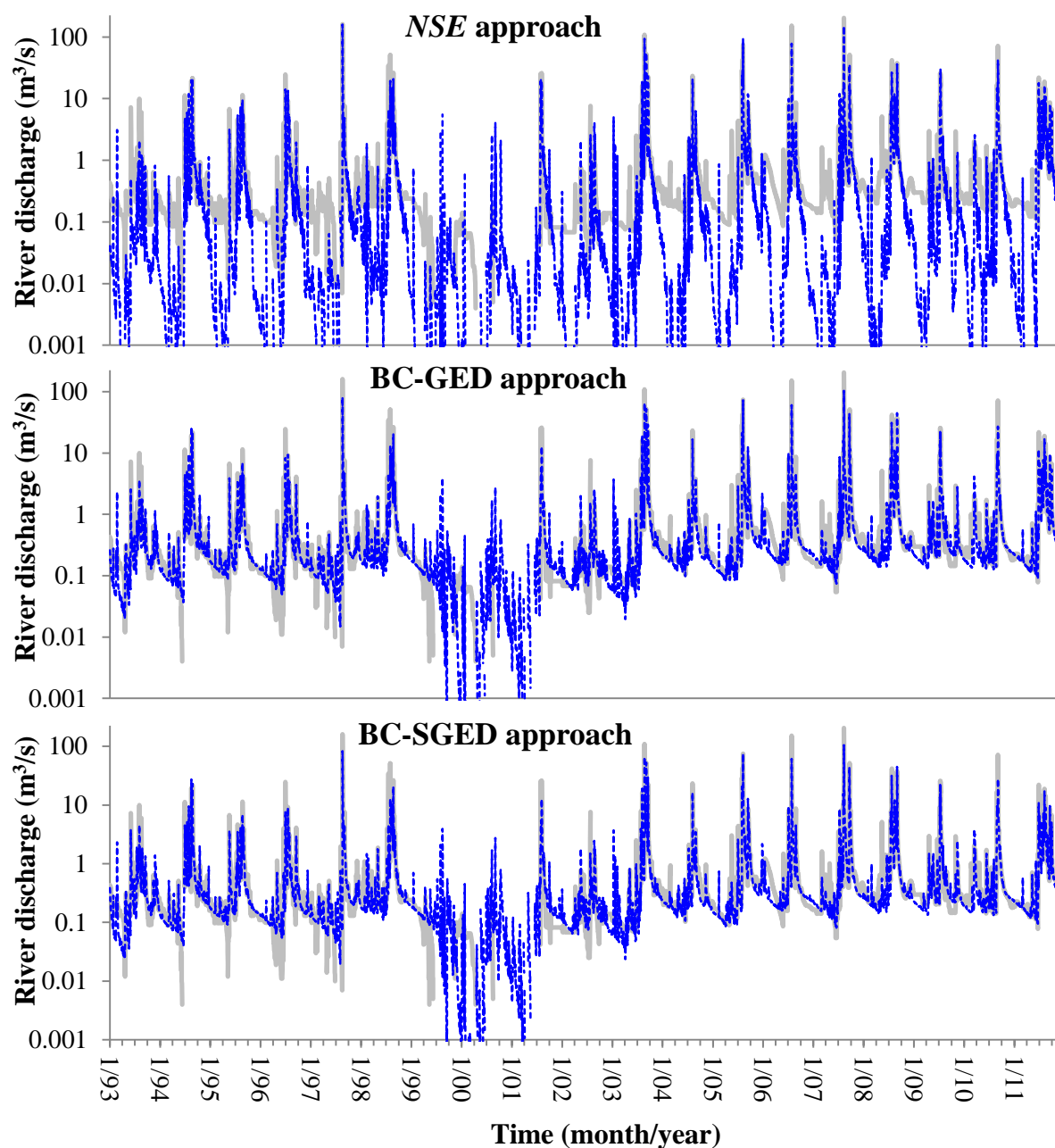
Figure 5.2 Comparison of the observed (black solid line) and simulated (blue dot line) river discharges on the logarithmic y-axis for *NSE*, BC-GED and BC-SGED approaches.

Figure 5.2 shows that, with the *NSE* approach, the hydrological model cannot mimic the baseflow, which was also found in many previous studies (Krause et al., 2005; Pushpalatha et

al., 2012). Surprisingly, the BC-GED and BC-SGED approaches can make the SWAT-WB-VSA model mimic the baseflow well.

5.1.1.2 Test of Statistical Assumptions

The *NSE*, BC-GED and BC-SGED approaches all have the statistical assumptions, e.g. the *NSE* approach assumes the model residuals follow the Gaussian error distribution with zero mean and a constant variance (Section 4.3.1.1), as detailed in Table 4.4. Generally, model residuals exhibit considerable autocorrelation, heteroscedasticity and non-normality (Section 2.2.2; Evin et al., 2013; Beven and Binley, 2013). Because the correlation of model residuals can be ignored in this study (as detailed in Section 4.3.1.2.1 and Appendix A), this section focuses on testing the heteroscedasticity and error distribution of model residuals.

The heteroscedasticity means that the variances of model residuals are inconstant in the time series. In other words, the model residuals are not identically distributed. The model residuals versus observed outcomes (e.g. river discharges) plot is widely used to visually inspect the heteroscedasticity of model residuals (Schoups and Vrugt, 2010). If there is a significant trend between model residuals and observed outcomes (e.g. river discharges), the model residuals exhibit heteroscedasticity. Figure 5.3 separately inspects the heteroscedasticity of model residuals for *NSE*, BC-GED and BC-SGED approaches. In Figure 5.3, the range of y-axis approximates $[\mu-2\sigma, \mu+2\sigma]$ (where μ and σ are the mean and standard deviation of model residuals, respectively); the dash line is the mean of residuals; and the solid line is used to highlight the trend. The model residuals of *NSE* approach (Figure 5.3 (a)) are estimated using observed river discharges minus model predictions directly (Eq. (4.3:1)). However, the model residuals of BC-GED and BC-SGED approaches (Figure 5.3 (b) and (c)) are estimated by the Box-Cox transformation (BC) function (Eq. (4.3:10)), where the BC transformation parameter (λ/lambda) is estimated via the minimum variance constraint (Eq. (4.3:12)).

Figure 5.3 (a) shows that the variance of residuals obviously increases with river discharges, suggesting heteroscedasticity that violates the stationary assumption of residuals in the *NSE* approach (Table 4.4). By contrast, the scatter points in Figure 5.3 (b) and (c) almost fill the whole panel space, showing that after Box-Cox transformation the variances of residuals are nearly the same among different river discharges. In other words, the time series of residuals after Box-Cox transformation are nearly homoscedastic.

(a) *NSE* approach   (b) BC-GED approach   (c) BC-SGED approach
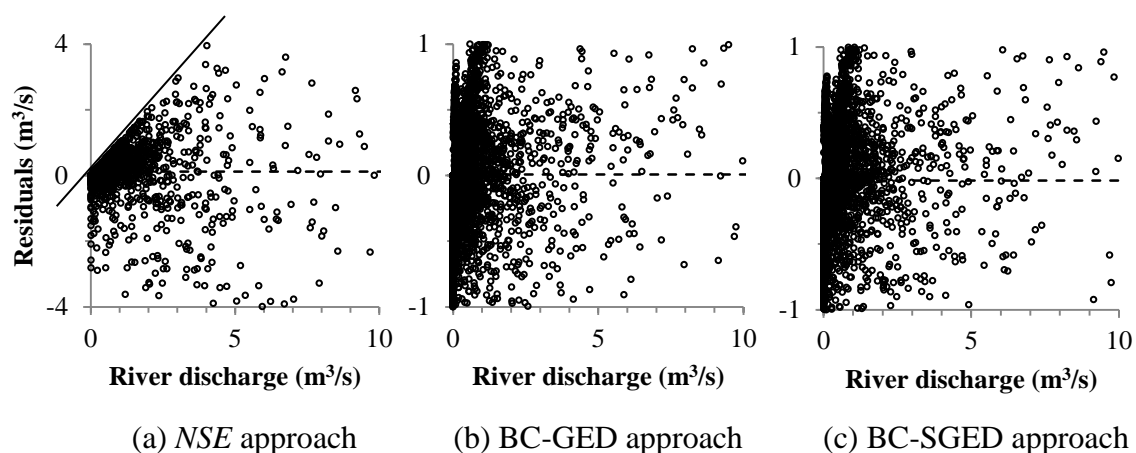
Figure 5.3   Model residuals as a function of observed river discharges for heteroscedasticity diagnostics. The solid line is used to highlight the trend.

The error distribution function is used to estimate the probability of model residuals/errors (Figure 2.4; Eq. (2.2:3)). The *NSE*, BC-GED and BC-SGED approaches assumed the model residuals follow the Gaussian distribution (Eq. (4.3:2)), generalized error distribution (GED) (Eq. (4.3:13)) and skew generalized error distribution (SGED) (Eq. (4.3:14)), respectively. Their parameters (i.e. the parameters of error model) are estimated in the automatic calibration procedure. The empirical error distribution of the time series of model residuals can be estimated by the histogram (Figure 2.5 (b)). Plotting the histogram of model residuals versus the hypothesis error distribution with the optimized parameters can visually inspect the error distribution of model residuals (Schoups and Vrugt, 2010). The histogram matches the hypothesis error distribution, meaning that the hypothesis error distribution can correctly estimate the probability of model residuals. Figure 5.4 separately inspects the assumption of error distribution of *NSE*, BC-GED and BC-SGED approaches.

In Figure 5.4, "error density" is the empirical probability density of model residuals estimated by histogram; and "Inferred error distribution" means the parameters of the hypothesis error distribution of likelihood function are optimized by DREAM. Figure 5.4 (a) shows that the error histogram is substantially different from the assumed Gaussian distribution of *NSE* approach. By contrast, Figure 5.4 (b) and (c) show that the inferred error distribution (GED and SGED) matches the empirical distribution of model residuals well.

In summary, the violations (heteroscedasticity (Figure 5.3 (a)) and non-Gaussian error distribution (Figure 5.4 (a))) demonstrate that SWAT-WB-VSA with the *NSE* approach cannot produce the model residuals that fulfill the assumptions of the approach (Table 4.4) in

the Baocun watershed. By contrast, the formal likelihood functions (BC-GED and BC-SGED) can guarantee the model residuals fulfill their assumptions (Table 4.4).

(a) *NSE* approach   (b) BC-GED approach   (c) BC-SGED approach
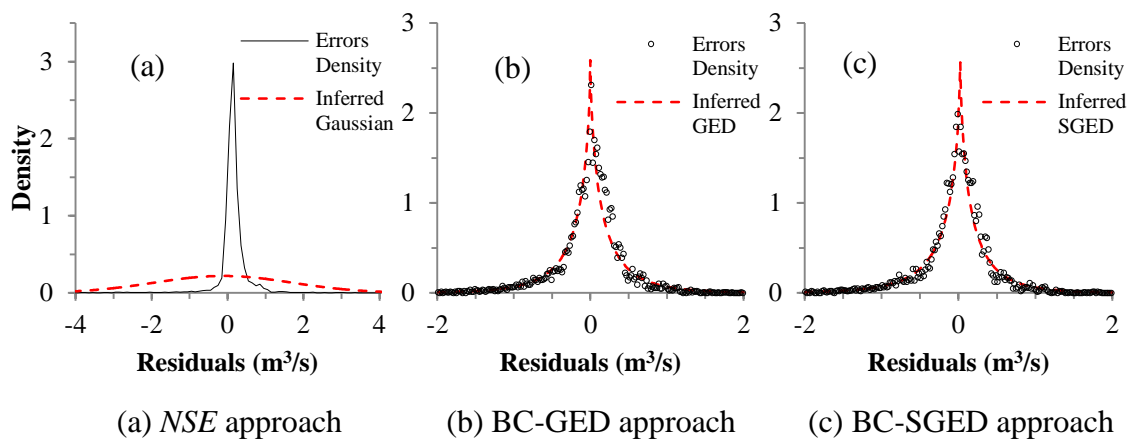
Figure 5.4 Empirical probability density of model residuals versus the assumed error distribution (red dash line) of the likelihood function. The empirical probability density is estimated by the histogram of model residuals (Figure 2.5 (b)).

5.1.2 Posterior Parameter Distribution

As shown in Figure 2.2, the DREAM stores the state parameter set (i.e. model parameters) into database at each MCMC iteration. The distribution of parameter sets stored in the database is the posterior distribution of model parameters (Figure 2.3). In this study, the number of parameter sets is 40,000, i.e. the number of MCMC iterations (Section 4.2.3). The posterior parameter distributions can reflect the uncertainty/equifinality of model parameters.

Optimization results of model parameters are shown in Table 5.1. In this table, the "95% confidence" columns present the 95% confidence interval of the posterior parameter distributions; and the "Best values" columns show the optimal value of model parameters that corresponds to the maximum value of the likelihood function. The parameter is regarded as sensitive if its 95% confidence interval in Table 5.1 is significantly narrower than its initial range in Table 4.2. So in the *NSE* approach, the parameters (related to soil evaporation (ESCO), effective soil depth (EDC), soil water storage and movement, and ground water movement) are very sensitive, whereas Manning roughness coefficients (OV_N, CH_N1 and CH_N2), surface flow lag (SURLAG) and groundwater storage (GWQMN and REVAPMN) parameters are insensitive. In the BC-GED approach, the most parameters are sensitive except the parameters of shallow aquifer revaporization/evaporation (REVAPMN and

GW_REVAP) and Manning roughness coefficients (OV_N, CH_N1 and CH_N2). In the BC-SGED approach, the sensitivity and optimal value of most parameters are nearly the same as in the BC-GED approach.

Table 5.1 Optimized parameters and 95% confidence interval of posterior parameter distributions for *NSE*, BC-GED and BC-SGED approaches.

Categories	Parameter	*NSE*			BC-GED			BC-SGED		
		95% confidence		Best	95% confidence		Best	95% confidence		Best
		Min	Max	values	Min	Max	Values	Min	Max	values
Evapotrans piration	v__ESCO	0.014	0.394	0.013	0.012	0.199	0.071	0.012	0.386	0.121
	v__EPCO	0.185	0.918	0.882	0.937	0.999	0.996	0.928	0.999	0.999
Surface water	v__EDC	0.731	0.895	0.757	0.730	0.774	0.750	0.718	0.768	0.744
	v__OV_N	0.025	0.479	0.032	0.013	0.490	0.242	0.038	0.493	0.483
	v__SURLAG	4.209	23.648	19.613	0.238	1.568	0.954	0.375	1.525	1.289
Soil water	r__SOL_Z	0.251	0.313	0.255	0.810	0.866	0.848	0.792	0.873	0.862
	r__SOL_BD	0.304	0.601	0.358	0.194	0.337	0.283	0.162	0.341	0.274
	r__SOL_AWC	-0.150	0.203	-0.091	0.906	1.456	1.085	0.870	1.599	1.101
	r__SOL_K	-0.769	-0.648	-0.725	-0.262	0.029	-0.046	-0.305	0.110	-0.031
Ground water	v__GW_DELAY	0.009	0.268	0.019	0.478	1.933	0.656	0.474	3.765	2.193
	v__ALPHA_BF	0.963	0.999	0.999	0.481	0.988	0.549	0.529	0.999	1.000
	v__GWQMN	50.8	941.8	845.6	273.6	526.4	391.9	311.3	561.3	376.5
	v__RCHRG_DP	0.001	0.056	0.011	0.152	0.333	0.257	0.062	0.273	0.147
	v__REVAPMN	27.3	911.1	819.6	328.3	981.7	838.0	118.4	968.7	376.9
	v__GW_REVAP	0.174	0.200	0.199	0.026	0.194	0.103	0.021	0.200	0.198
Tributary/main channel	v__CH_N1	0.009	0.145	0.005	0.010	0.147	0.136	0.015	0.148	0.111
	v__CH_N2	0.019	0.147	0.143	0.006	0.138	0.015	0.005	0.050	0.014
Box-Cox	λ (lambda)				0.432	0.445	0.440	0.432	0.445	0.439
Probability density function	μ (mean)				-0.006	0.048	0.007	-0.054	0.045	-0.018
	σ (sigma)				0.520	0.531	0.523	0.520	0.531	0.520
	β (beta)				0.662	0.692	0.672	0.656	0.687	0.667
	ξ (xi)							0.888	0.958	0.938

Because the histograms of model parameters stored in database are usually rough (non-smooth) owing to too many model parameters and insufficient MCMC iterations, the kernel smoothing method is commonly used to smooth the posterior parameter distribution (Figure 2.3). The kernel smoothing densities of the posterior parameter distributions are shown in Figure 5.5, where the lambda (λ) is the Box-Cox transformation parameter (Eq. (4.3:10)), and the mean (μ), sigma (σ), beta (β) and xi (ξ) are the parameters of error distribution function (Table 4.4; Table 5.1). The detailed descriptions of model parameters are shown in Section 4.2.2 and Table 4.2.

In Figure 5.5 for the *NSE* approach, values of ESCO gather to zero, meaning that the evaporation of soil layers is large; values of SOL_K are very small, resulting in less soil interflow; the value of GW_DELAY approaches zero and the value of ALPHA_BF approaches one, indicating that the groundwater declines very fast, which leads to simulate baseflow badly (Figure 5.2); values of RCHRG_DP are close to zero, revealing that little groundwater percolates to deep aquifer; values of GW_REVAP gather to the right boundary (i.e. 0.2), showing strong groundwater revaporization/evaporation rates (Section 4.2.2). In summary, simulated results of the *NSE* approach show that the main way of groundwater loss is revaporization/evaporation, and the main runoff component is the return flow of groundwater with rapid recession.
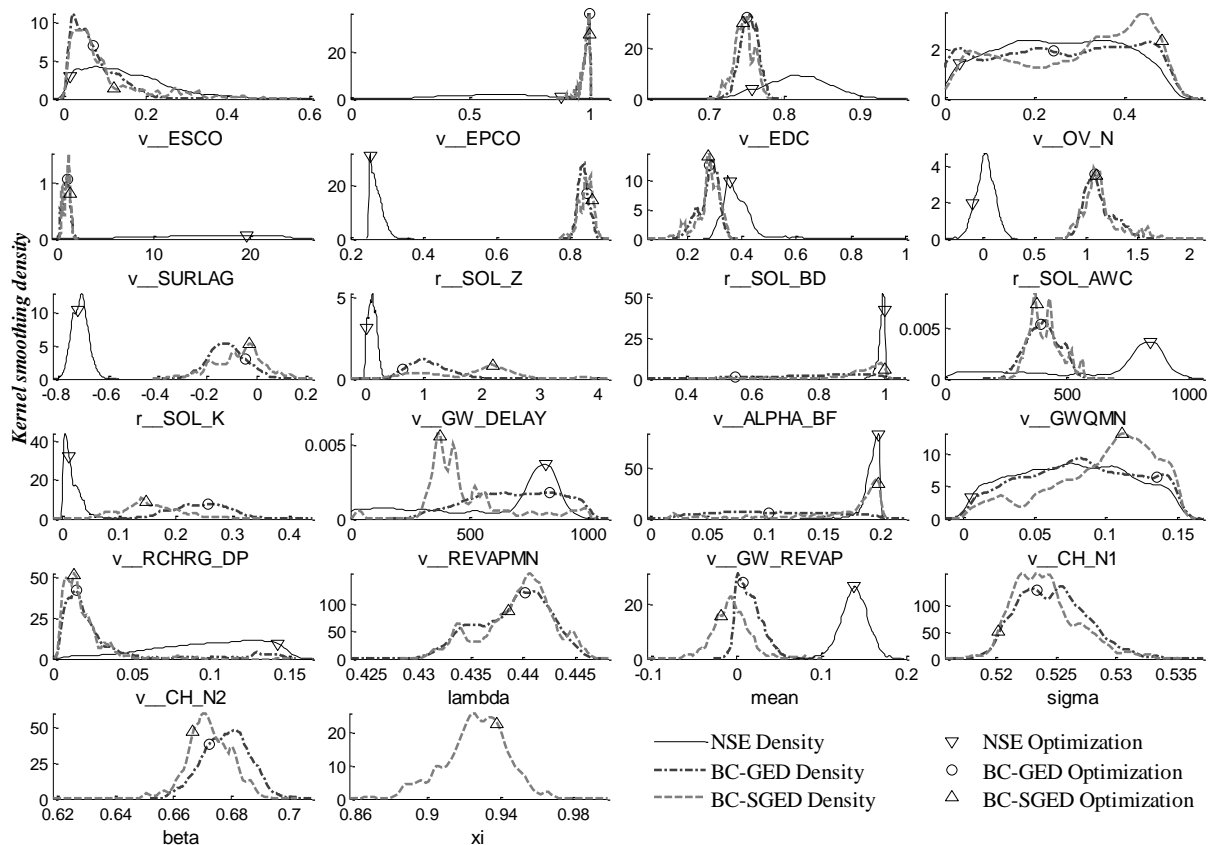
Figure 5.5 The kernel smoothing densities of posterior parameter distributions (lines) and the optimized parameters (points) for *NSE*, BC-GED and BC-SGED approaches. The posterior parameter distribution is the probability distribution of parameter sets stored in the database (Figure 2.2 and Figure 2.3). The lambda, mean, sigma, beta and xi are the parameters of error model (Table 5.1). The detailed descriptions of model parameters are shown in Section 4.2.2 and Table 4.2.

In Figure 5.5 for the BC-GED approach, values of ESCO also gather to zero and values of EPCO approach one, indicating the high soil and plant evapotranspiration; the value of CH_N2 is close to zero, meaning that the storage capacity of main channel is small; the mean errors gather to zero, demonstrating that the water balance is maintained in the BC-GED approach. In summary, the simulation results of the BC-GED approach show that the main pathway of water loss is through soil and plant evapotranspiration, the main channel routing is quick, and the bias of model residuals is small.

Figure 5.5 shows that the most posterior parameter distributions of the BC-SGED approach are similar to those of the BC-GED approach. Figure 5.4 (c), Figure 5.5 and Table 5.1 all demonstrate that the model residuals produced by the BC-SGED approach are symmetric, and the skewness (ξ/xi) of the BC-SGED approach is very close to one, i.e. no-skewness, that is assumed in the BC-GED approach (Eqs. (4.3:13) and (4.3:14)).

5.1.3 Model Validation Using Groundwater Data

In this study, the groundwater data observed in the groundwater gauge (Figure 3.1) are used for model validation. The time series of observed groundwater level is shown in Figure 3.8.

In each hydrologic response unit (HRU), SWAT separately simulates the soil water of soil profile and the groundwater of shallow aquifer, and neglects the interactions between soil water and groundwater (Figure 4.1). Therefore, SWAT cannot mimic the groundwater table level directly. Vazquez-Amábile and Engel (2005) and White et al. (2011) proposed a method that converts the model-predicted/simulated soil water to an equivalent groundwater depth by dividing by soil porosity. This method depends heavily on the measurement precision of soil porosity. However, it is difficult or impossible to accurately measure soil porosity in the field because of heterogeneity.

In Baocun watershed, because of the shallow water table depth and the sandy soil layer with high hydraulic conductivity around the groundwater gauge, there is a close relationship between the groundwater table level and the total soil water volume (Figure 3.8; Section 3.5.3). Therefore, in order to avoid the measurement of soil porosity, we can directly test the linear relationship between the observed groundwater level and the simulated soil water volume of soil profile and shallow aquifer in the HRU where the groundwater gauge fell into, for model validation. The coefficient of determination (R^2, Table 2.1) can be used as the

model performance (goodness-of-fit) indicator to reflect the linear correlation between observed groundwater level and simulated soil water volume.

Comparison of the observed groundwater level and the simulated soil water volume for three likelihood functions (i.e. *NSE*, BC-GED and BC-SGED approaches) is shown in Figure 5.6. In this figure, the soil water (including the soil water of soil profile and the groundwater of shallow aquifer) is produced by SWAT-WB-VSA using optimized parameters (Table 5.1); and S_y is the specific yield of unconfined aquifer and defined as:

$$S_y = \frac{\Delta GW}{\Delta h} \approx \frac{\Delta SW}{\Delta h} \tag{5.1:1}$$

where

ΔGW and ΔSW are the change of groundwater (*GW*) and soil water (*SW*) volumes per unit area, respectively,

Δh is the change of groundwater level.

According to the studies by Fetter (1994), the experiential S_y of silt soil is between 0.03 and 0.19 with the mean of 0.18, of fine sand is between 0.10 and 0.28 with the mean of 0.21, and the S_y value of the groundwater gauge approximates to 0.195.
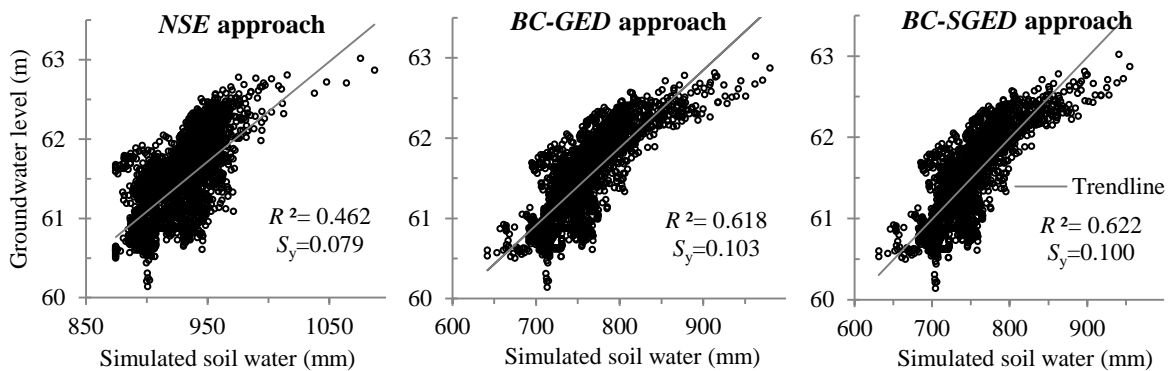
Figure 5.6 The observed groundwater table level versus the simulated soil water volume for *NSE*, BC-GED and BC-SGED approaches. The larger the value of R^2, the better linear relationship (i.e. model performance) is.

Figure 5.6 shows that the R^2 of the BC-GED and the BC-SGED approaches are larger than that of the *NSE* approach, and their S_y are also much closer to the experiential value (Fetter, 1994). So, the SWAT-WB-VSA with the BC-GED or BC-SGED approach mimics the groundwater levels better than the *NSE* approach. However, there is little difference between the BC-GED approach and the BC-SGED approach.

5.2 Multi-response Calibration with Different Likelihood Functions

The purpose of the multi-response calibration approaches is to account for the trade-offs among different objectives/response, e.g. the river discharges and the sediment loads in the watershed outlet (Efstratiadis and Koutsoyiannis, 2010). Based on the assumption of independence between the flow and sediment model residuals, a kind of multi-response likelihood function is proposed in Section 4.3.2. Compared to the traditional multi-response calibration methods, such as the weighted sum method (Abbaspour et al., 2007) and evolutionary algorithm (Lu et al., 2010), the multi-response likelihood approach avoids the trouble of choosing the weights of different objectives and can be incorporated into Bayesian approach for uncertainty analysis.

The calibration approach only using river discharge objective (Section 5.1) is termed as the single-response calibration approach, such as single-response *NSE* and BC-GED approaches, in contrast to multi-response calibration approach using both the river flow discharge and sediment load objectives. The procedures of multi-response calibration method are similar to those of single-response calibration method (Section 5.1), and the only difference being that the multi-response method calibrates the flow and sediment parameters (Section 4.2.2; Table 4.2) simultaneously.

In the automatic calibration procedure (Figure 2.1), SWAT-WB-VSA (Section 4.1) is used as the hydrological model tool, the river discharges (Section 3.5.1) and sediment loads (Section 3.5.4) as the observed data, the DREAM (Section 2.2.1) as the Bayesian (optimization) tool and the flow and sediment parameters (Table 4.2) as the calibrated parameters, as detailed in Table 4.3 for the case of multi-response calibration.

The multi-response calibration results of Bayesian approach with two multi-response likelihood functions (*NSE* and BC-GED approaches proposed in the Section 4.3.2.2 and 4.3.2.3, respectively) are separately presented in this Section. For each likelihood function, the simulation results and the posterior parameter distributions are introduced separately.

5.2.1 Multi-response *NSE* Approach

The procedure of DREAM (MCMC) method for solution of the Bayesian equation (Eq. (2.2:1)) is shown in Figure 2.2. At each MCMC iteration, the multi-response *NSE* approach firstly calculates the Nash–Sutcliffe efficiency coefficient (*NSE*) for flow (NSE_{flow}) and

sediment (NSE_{sed}) after running the SWAT-WB-VSA model with model parameter set (θ), and then substitutes the NSE values in the Eq. (4.3:18) to estimate the value of the likelihood function (Figure 2.2).

5.2.1.1 Simulation Results

Comparison of the observed and the simulated river discharges and sediment loads on the daily time step during the flood season (from June to September) in the period of 1993 - 1999 are shown in Figure 5.7. The simulated data are produced by SWAT-WB-VSA with optimized parameters corresponding to the maximum likelihood value of multi-response NSE approach (Eq. (4.3:18)).
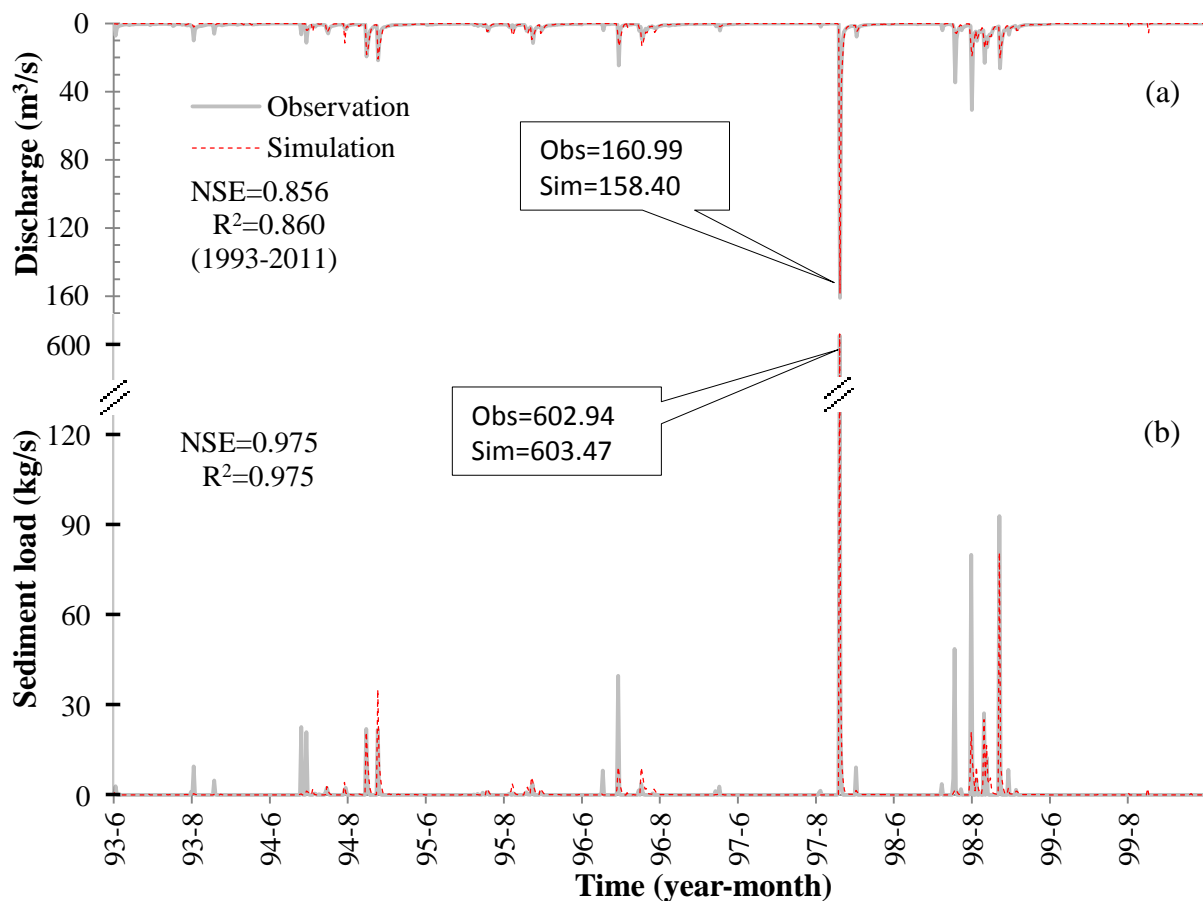
Figure 5.7 Optimal simulation results of the *NSE* approach during the flood season (from June to September) in 1993-1999: (a) comparison of the observed and simulated river discharges; and (b) comparison of the observed (black solid line) and simulated (red dot line) sediment loads.

In Figure 5.7 (a), NSE and R^2 reflect the performance of river-discharge simulation from 1993 to 2011. In Figure 5.7 (b), NSE and R^2 reflect the model performance of sediment-load simulation during the flood season in the period (1993 – 1999). Figure 5.7 shows that, with the *NSE* approach, SWAT-WB-VSA mimics the observed river discharges and sediment loads well, which captures most flood events, especially the large river flow discharge and sediment transport, e.g. in 1997.

Because of large differences between the high- and the low- values of the river discharge and sediment load data, Figure 5.7 hardly shows the low values. Therefore, the amount of river discharges (flow) and sediment loads (sediment) during the flood period in each year are shown in Table 5.2. In this table, "Observation" represents observed data and "Simulation" represents optimal simulation results. Table 5.2 shows in high flow years (such as 1994 and 1997), SWAT-WB-VSA with *NSE* approach can maintain small relative-errors between the observation and simulation of flow/sediment. In low flow years, however, the model yields a large relative-error especially of sediment. For example, in 1995 and 1999, the model obviously over-estimates the sediment loads, but under-estimates in 1993. Surprisingly, the bias of total sediment loads in the period of 1993 - 1999 is almost zero (i.e. no-bias).

Table 5.2 Comparison of the simulated and the observed river flow and sediment amounts during flood season in each year for *NSE* approach.

Categories	Methods [2]	1993	1994	1995	1996	1997	1998	1999	Total
Flow ($\times 24 \times 60 \times 60$ m^3) [1]	Observation	105.7	201.2	133.6	142.2	258.8	376.7	0.2	1218.4
	Simulation	23.5	203.7	127.7	147.3	275.4	292.1	10.2	1079.9
Sediment (t)	Observation	1634.3	8988.7	1081.4	5205.8	54928.7	25444.9	0.0	97283.8
	Simulation	251.8	10138.0	3659.9	4784.6	57388.2	20788.1	302.4	97313.0

[1] The "$\times 24 \times 60 \times 60$" is used to convert the unit of "m^3/s" to the unit of "m^3/d".

[2] The"Observation" is the observed data; "Simulation" is the optimal simulation results.

Figure 5.8 inspects the heteroscedasticity of model residuals, using the model residuals versus observed data plot. The model residuals are estimated using observed data minus simulated data directly (Eq. (4.3:1)). In this figure, the dash line is the mean of residuals; and the black line is used to highlight the trend. Figure 5.8 shows the variance of model residuals obviously increases along with the river-discharges/sediment-loads, suggesting heteroscedasticity that violates the stationary assumption (i.e. the same variance) of the *NSE* approach (Table 4.4).

Further, Figure 5.9 inspects the assumption of independence between the flow and sediment model residuals (Section 4.3.2), using the sediment model residual (Error$_{sed}$) versus flow model residual (Error$_{flow}$) plot. The solid line is the trend line. Figure 5.9 shows the sediment model residual obviously increases with the flow model residual, suggesting correlation that violates the independence assumption of the multi-response likelihood function (Section 4.3.2).

| (a) Flow | (b) Sediment |
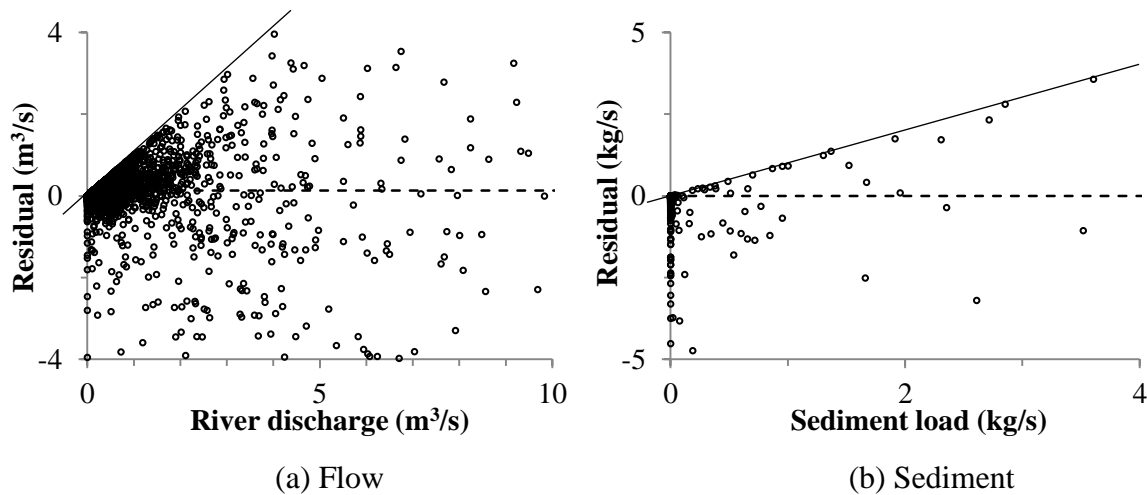
Figure 5.8 Model residuals of the *NSE* approach as a function of observed data for heteroscedasticity diagnostics. The dash line is the mean of residuals. The solid line is used to highlight the trend.
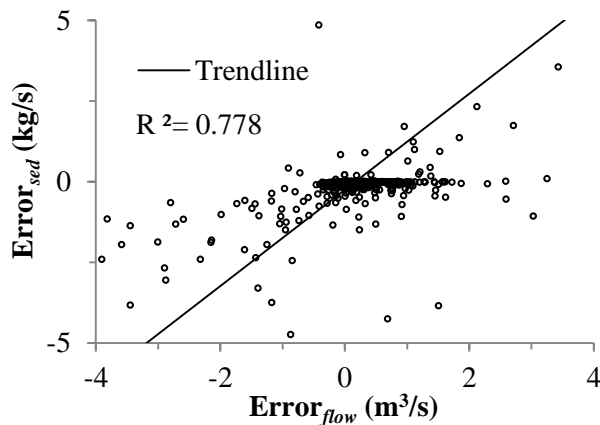
Figure 5.9 Diagnosis of the independence between the flow and sediment model residuals for the *NSE* approach.

Figure 5.10 inspects the assumption of Gaussian error distribution of the *NSE* approach. In this figure, "μ" and "σ" are the mean and standard deviation of model residuals, respectively, which are the parameters of the Gaussian distribution (Eq. (4.3:2)). The σ is estimated by the unbiased equation (Eq. (4.3:4)) and the μ is assumed to be zero by *NSE* approach. Empirical probability density (i.e. errors density) of model residuals is estimated by the histogram. Figure 5.10 shows that the error histograms of both flow and sediment are all substantially different from their assumed Gaussian distribution with zero-mean (Table 4.4).

(a) Flow   (b) Sediment
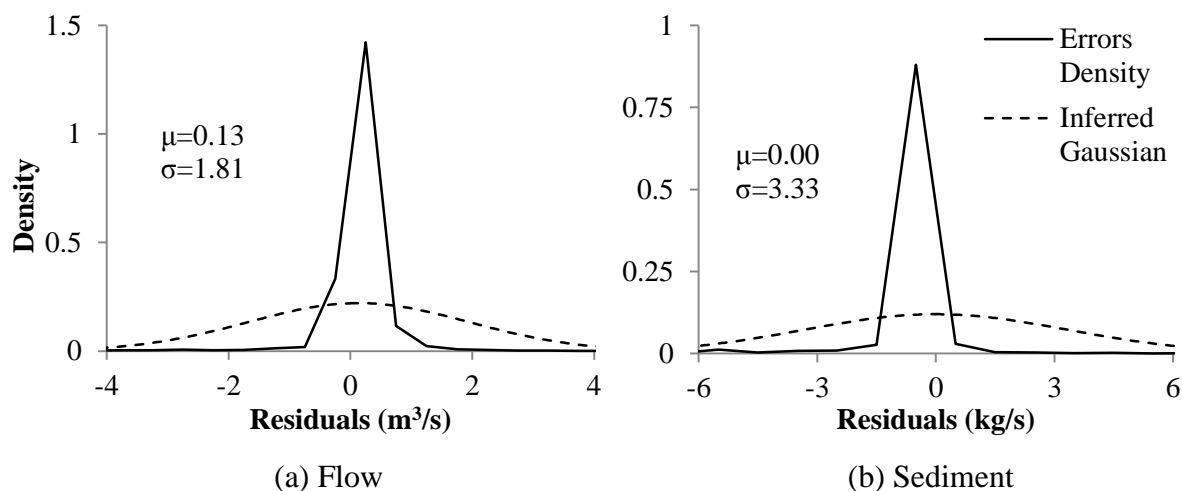
Figure 5.10 Empirical probability density of model residuals (solid line) versus the assumed
Gaussian distribution (dash line) of *NSE* approach. Empirical probability density
is estimated by the histogram of model residuals. The "μ" and "σ" are the mean
and standard deviation of model residuals, respectively.

Such violations (heteroscedasticity, correlation and non-normality) demonstrate that multi-response *NSE* approach cannot guarantee that the model residuals of both flow and sediment fulfill the assumptions of the approach (Table 4.4; Section 4.3.2.1). Therefore, the *NSE* approach is an informal likelihood function in this study.

Computed by SWAT-WB-VSA with the optimum parameter set, the average annual components of simulated river flow and sediment (1993 – 2011) are shown in Table 5.3. The detailed descriptions of runoff components are shown in Section 4.1. For evaluating the effect of multi-response calibration method on runoff, Table 5.3 also includes the average annual runoff components inferred by single-response method (Section 5.1). Table 5.3 shows for *NSE* approach the runoff components are nearly the same between multi- and single-response calibration method, where the main runoff components are groundwater return flow and the main pathway of groundwater loss is revaporization/evaporation.

The sediment simulation includes two processes: the erosion of soil particles from slopes and the sediment transport in the channel network. In Table 5.3, the "Total slope erosion" represents the amount of soil erosion from slopes/HRUs; and the "Total river erosion" represents the amount of soil erosion from channels. Soil erosions from HRUs and Channels in total are the suspended sediment loads in the watershed outlet. The main channels are river 5, 6 and 7 shown in Figure 3.1. In Table 5.3 for sediment, the positive value means erosion and the negative value means deposition. The sediment loads inferred by *NSE* approach

demonstrate the main channels in the Baocun watershed are all eroded/denuded, and more than 36% of sediment loads at the watershed outlet come from the main channels.

Table 5.3 Average annual components of simulated river flow and sediment during 1993 – 2011 for *NSE* and BC-GED approaches.

Objective [1] Categories	*NSE* approach		BC-GED approach	
	Flow	Flow+Sed.	Flow	Flow+Sed.
Flow (mm/yr) Evapotranspiration	489.50	492.90	520.60	521.20
Overland flow	37.04	37.10	65.09	63.17
Interflow	12.49	13.17	90.82	90.39
Return flow	199.84	197.43	102.47	107.18
Revaporization [2]	76.73	74.31	0.00	12.94
Deep percolation [3]	3.34	3.55	38.82	25.73
Sediment [4] (t/yr) Total slope erosion [5]		18101.2		25185.9
Total river erosion [6]		10246.1		-4445.9
Level_2 river_5 erosion		2863.1		-1567.7
Level_2 river_6 erosion		1925.5		-1276.1
Level_3 river_7 erosion		5457.5		-1602.0

[1] "Flow+Sed." represents multi-response calibration using both river discharge and sediment load objectives; "Flow" represents single-response calibration only using river discharge objective.

[2] The evaporation from groundwater.

[3] The groundwater losses from the shallow aquifer to the deep aquifer.

[4] The positive value means erosion and the negative value means deposition; and river 5, 6 and 7 are the main channels (Figure 3.1 (b)), where the river 5 and 6 belong to the second level river and river 7 belongs to the third level river.

[5] The amount of soil erosion from slopes/HRUs.

[6] The amount of soil erosion from main channels.

5.2.1.2 Posterior Parameter Distribution

The posterior parameter distribution (after kernel smoothing; Figure 2.2 and Figure 2.3) and optimal value of flow parameters estimated by the multi- and single- response *NSE* approaches are shown in Figure 5.11. This figure shows almost all the posterior distribution of flow parameters estimated by multi-response *NSE* approach are sharper than those estimated by single-response method, especially the Manning's coefficients of sloping surface (OV_N) and main channel (CH_N2). It demonstrates the multi-response approach reduces the parameter uncertainty. However, some posterior parameter distributions (e.g. the

soil property parameters (SOL_Z, SOL_BD and SOL_K)) of multi-response *NSE* approach exhibit considerable multimodality, possibly because the multi-response calibration approach optimizing both the flow and sediment parameters (Table 4.2) causes more local optimal solutions than the single-response method owing to over-parameterization (Efstratiadis and Koutsoyiannis, 2010; Andreassian et al., 2012). In Figure 5.11, the optimal values of most flow parameters of the multi-response *NSE* approach, such as the soil property parameters and the groundwater storage and movement parameters (ALPHA_BF, GWQMN and RCHRG_DP), are similar to those of the single-response method, which corresponds to the results in Table 5.3, where average annual runoff components of both methods are nearly the same.
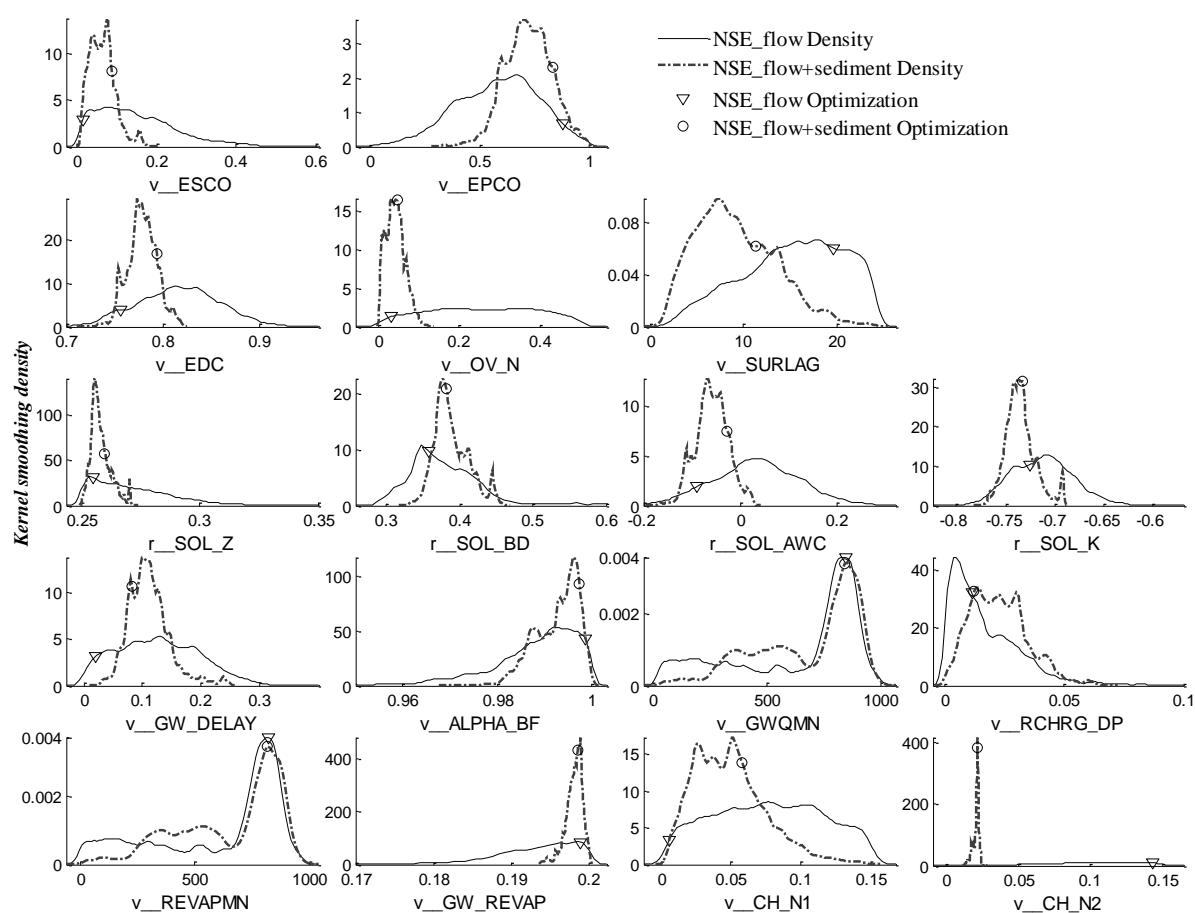
Figure 5.11 Comparison of the posterior distribution (after kernel smoothing) and optimal value (point) of flow parameters estimated by the multi- (dash line) and single-response (solid line) *NSE* approaches. The posterior parameter distribution is the probability distribution of parameter sets stored in the database (Figure 2.2 and Figure 2.3). The detailed descriptions of flow parameters are shown in Section 4.2.2 and Table 4.2.

The posterior distribution and optimal value of sediment parameters estimated by the multi-response *NSE* approach are shown in Figure 5.12. This figure shows the soil USLE erodible factor of Luvisols (USLE_K2) is greater than that of Fluvisols (USLE_K3), and greater than that of Regosols (USLE_K1), i.e. the value of USLE_K2 (Luvisols) > USLE_K3 (Fluvisols) > USLE_K1 (Regosols). The posterior distribution of USLE_K1 (Regosols) is sharper than that of USLE_K2 (Luvisols), and sharper than that of USLE_K3 (Fluvisols), i.e. the uncertainty of USLE_K1 (Regosols) < USLE_K2 (Luvisols) < USLE_K3 (Fluvisols). The SPEXP is the power reflecting the exponential relationship between the peak flow velocity and the sediment transport capacity of channel (Eq. (4.1:19); Table 4.2). The 95% confidence interval of the posterior distribution of SPEXP is [1.33, 2.40]. That interval is close to the range ([1.0, 2.0]) recommended by Neitsch et al. (2005).
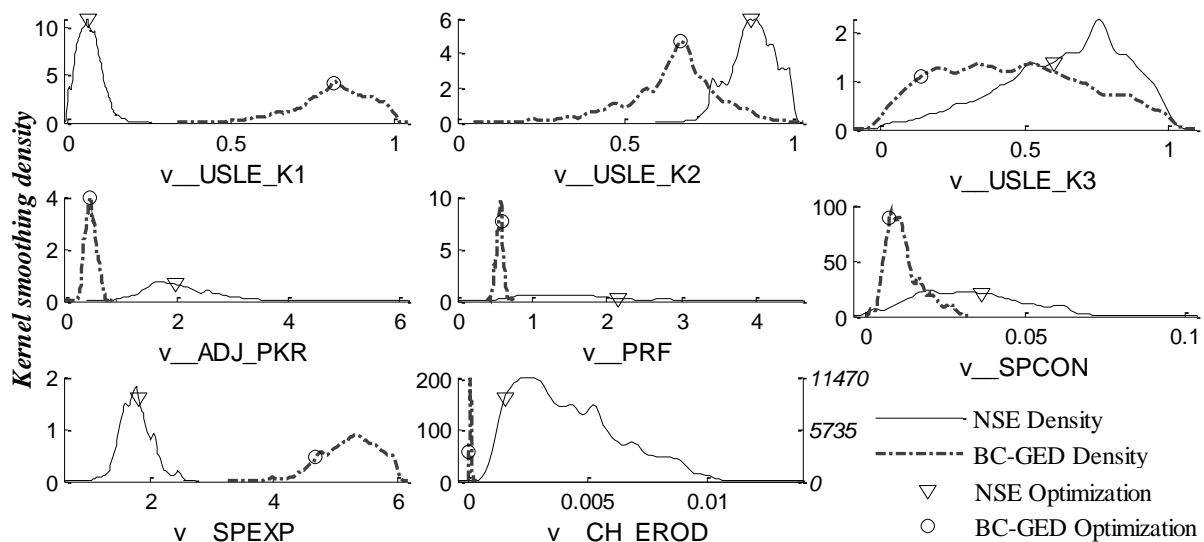
Figure 5.12 Comparison of the posterior distribution (after kernel smoothing) and optimal value (point) of sediment parameters estimated by the multi-response *NSE* and BC-GED approaches. The detailed descriptions of sediment parameters are shown in Section 4.2.2 and Table 4.2.

5.2.2 Multi-response BC-GED Approach

The procedure of DREAM (MCMC) for different likelihood approaches is the same (Figure 2.2), and the only difference being the estimate of the likelihood values of model residuals. At each MCMC iteration, the (single-response) BC-GED approach firstly uses the observed and simulated data to estimate the value of Box-Cox (BC) transformation parameter

(lambda/λ) based on the minimum variance constraint (Eq. (4.3:12)), and then calculates the model residuals after BC (Eq. (4.3:10)). Next, it estimates the parameters (such as the mean (mu/μ), standard deviation (sigma/σ) and kurtosis coefficient (beta/β)) of the error model (i.e. GED; Eq. (4.3:13)) using the model residuals after BC. Finally, this approach calculates the likelihood value of model residuals ($l(\theta|obs)$, Eq. (2.2:3)).

The multi-response BC-GED approach first separately calculates the likelihood values of the river flow and sediment model residuals using the single-response approach. After this, the two likelihood values are added together (Eq. (4.3:20)) at each MCMC iteration. However, unlike the flow series, the sequence of sediment loads is discrete, and the parameters of the error model can only be estimated using the flood-state data, because the sediment model residuals in the non-flood state are not independent and identically distributed (I.I.D.) (Section 4.3.2.3). Additionally, the number of the sediment data (greater than zero) is too few (Figure 3.9). Therefore, for simplification and reduction of uncertainty, the kurtosis coefficient of the GED (Eq. (4.3:13)) for sediment model residuals is set to a constant value of 0.672 that is the optimal inference result of single-response BC-GED approach (Table 5.1). The sediment model residuals in the flood state are used to estimate other parameters (such as the mean (μ) and standard deviation (σ)) of GED function (Eq. (4.3:13)).

5.2.2.1 Simulation Results

Comparison of the observed and the simulated river discharges and sediment loads on the daily time step during the flood season (from June to September) in the period of 1993 - 1999 are shown in Figure 5.13. The simulated data are produced by SWAT-WB-VSA with the optimal values of model parameters estimated by multi-response BC-GED approach. Figure 5.13 (a) shows that the simulated results can reproduce most flood events, but the flood peaks are smaller than those of observed values in some cases, especially in extreme flood events e.g. in 1997. Figure 5.13 (b) shows that SWAT-WB-VSA with BC-GED approach also mimics the sediment loads well.
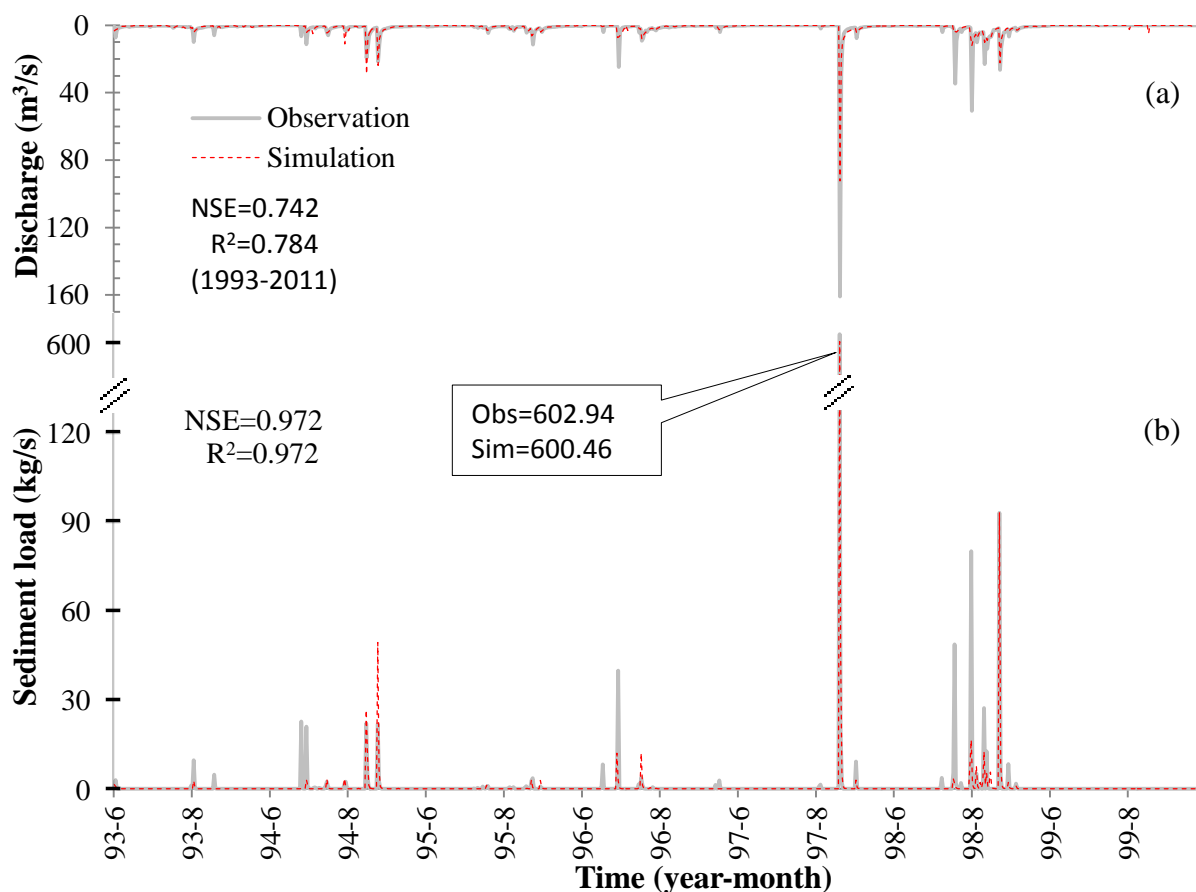
Figure 5.13 Optimal simulation results of the BC-GED approach during the flood season
(from June to September) in 1993-1999: (a) comparison of the observed and
simulated river discharges; and (b) comparison of the observed (black solid line)
and simulated (red dot line) sediment loads.

The amount of river discharges (flow) and sediment loads (sediment) during the flood period in each year are shown in Table 5.4. In this table, "Observation" is the observed data and "Simulation" is the optimal simulation results corresponding to the maximum likelihood value of multi-response BC-GED approach (Eq. (4.3:20)). Table 5.4 shows the amount of total sediment loads of simulation is little less than that of observation. The largest relative error of sediment loads occurs in 1993 and the largest bias of sediment loads occurs in 1996. However, the relative errors of river discharges are much less than those of sediment loads in both 1993 and 1996. It is noticed that in the 1999 the sediment load is very close to zero (i.e. observed value), whereas the amount of simulated river discharges is not zero, i.e. the main channel is not dry. It may demonstrate that the characteristic of the main channel inferred by the BC-GED approach is closely non-erodible.

5 Results

Table 5.4 Comparison of the simulated and the observed river flow and sediment amounts
during flood season in each year for BC-GED approach.

Categories	Methods [2]	1993	1994	1995	1996	1997	1998	1999	Total
Flow ($\times 24 \times 60 \times 60$ m^3) [1]	Observation	105.7	201.2	133.6	142.2	258.8	376.7	0.2	1218.4
	Simulation	91.9	213.5	117.0	135.9	198.6	255.6	20.8	1033.4
Sediment (t)	Observation	1634.3	8988.7	1081.4	5205.8	54928.7	25444.9	0.0	97283.8
	Simulation	513.2	9312.0	718.1	2606.3	56632.3	15452.1	2.4	85236.4

[1] The "$\times 24 \times 60 \times 60$" is used to convert the unit of "m^3/s" to the unit of "m^3/d".

[2] The "Observation" is the observed data; "Simulation" is the optimal simulation results.

Figure 5.14 (a) and (b) inspect the error's heteroscedasticity of flow and sediment, respectively, using the model residuals versus observed data plot. In Figure 5.14, the model residuals of both flow and sediment are calculated by BC function with corresponding BC transformation parameters (Eq. (4.3:10)). In Figure 5.14 (a), the scatter points almost fill the whole panel space, which means the BC transformation method well removes the heteroscedasticity of flow model residuals. Compared with the Figure 5.8 (b), the Figure 5.14 (b) shows the BC method obviously reduces the heteroscedasticity of sediment model residuals. Because too few sediment data are greater than zero, the scatter points only fill the partial panel-space in Figure 5.14 (b).

(a) Flow   (b) Sediment
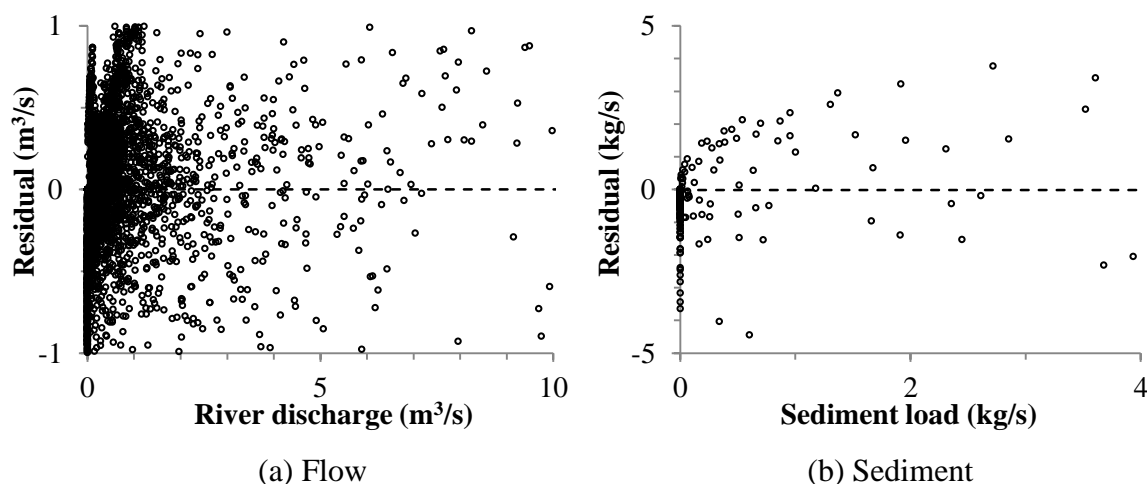
Figure 5.14 Model residuals of the BC-GED approach as a function of observed outcomes
for heteroscedasticity diagnostics. The dash line is the mean value of residuals.

Further, inspecting the assumption of independence (Section 4.3.2.1) between the flow (Error$_{sed}$) and sediment (Error$_{flow}$) model residuals after BC transformation is shown in Figure 5.15. This figure shows Error$_{sed}$ are obviously independent of Error$_{flow}$, which also pass the

94

Spearman test with 5% significance level for "independence" hypothesis (Hollander and Wolfe 1973). In Figure 5.15, the number of the minus values of sediment model residuals is more than that of the positive values because the number of zero data of sediment loads is more than that of positive data (Figure 3.9).
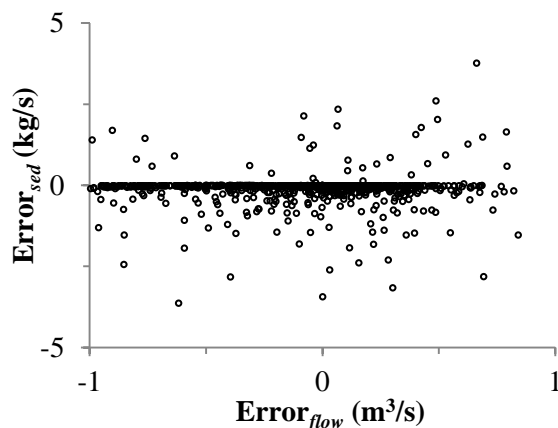
Figure 5.15 Diagnosis of the independence between flow and sediment model residuals for the BC-GED approach.

Inspecting the error distribution assumption of the BC-GED approach is shown in Figure 5.16, via the empirical error distribution versus the generalized error distribution (GED). Because the sediment model residuals in non-flood state are not I.I.D. (Section 4.3.2.3), only the probability distribution of model residuals in the flood state is inspected in Figure 5.16 for sediment. In this figure, the "λ", "μ", "σ" and "β" are the parameters of error model inferred by BC-GED approach, where "λ" is BC transformation parameter (lambda), and "μ", "σ" and "β" are the mean, standard deviation and kurtosis coefficient of GED function, respectively. It is noted that the flow and the sediment model residuals use the different parameter set of error model.

Figure 5.16 (a) shows that the inferred error distribution (GED) matches the empirical distribution of residuals well. Figure 5.16 (b) shows that the inferred GED generally over-estimates the probability of positive residuals, but under-estimates that of minus residuals. Therefore, the empirical error distribution of sediment seems to be skewed. In fact, however, the error distribution of sediment is impossible to be skewed because the mean/bias of residuals is zero ("μ"; Figure 5.16 (b)), and a number of residuals are zero, i.e. the mode of residuals is zero (Figure 5.16 (b) and Figure 5.15), which can guarantee the symmetry of the error distribution with zero-mean and zero-mode.

Therefore, the main reason of the deviation between the empirical and hypothetical error distribution of sediment (Figure 5.16 (b)) is due to statistical bias, which owes too few samplings (Section 4.3.2.3; Figure 3.9). In order to increase the number of statistical

samplings, the absolute residuals of sediment are used instead of the original residuals for estimation of the empirical error distribution. The distribution of absolute residuals of sediment is inspected in Figure 5.16 (c). This figure shows the inferred GED well matches the empirical error distribution.

In summary, the BC-GED approach can guarantee that the model residuals fulfill its statistical assumptions well (Table 4.4; Section 4.3.2.1).

| (a) Flow | (b) Sediment | (c) Absolute residuals of sediment |
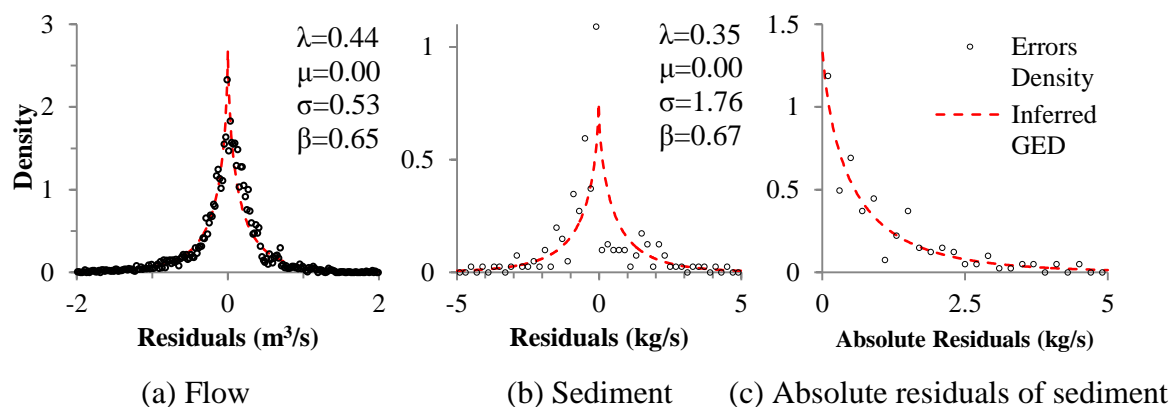
Figure 5.16 Empirical probability density (black circle) of model residuals versus the inferred GED (red dash line). Empirical probability density is estimated by the histogram of model residuals. The λ is BC transformation parameter, and the μ, σ and β are the mean, standard deviation and kurtosis coefficient of GED function (Eq. (4.3:13)), respectively.

Computed by SWAT-WB-VSA with the optimized parameters, the average annual components of river flow and sediment during 1993-2011 for BC-GED approach are shown in Table 5.3. This table shows the runoff components are nearly the same between the multi- and single-response BC-GED methods, where most (about 76%) of the runoff is from underground flow, and the amount of groundwater return flow is close to that of interflow, but greater than that of overland flow. However, the pathways of groundwater loss are bit different. The multi-response method concludes that some groundwater loss occurs due to revaporization/evaporation. The component analysis of sediment loads in the main channel indicates that all the main channels have sediment deposition problem, and nearly 18% sediments eroded from hillslopes/HRUs deposit in the main channel, which corresponds to the results in Table 5.4, where the amount of simulated sediment loads is close to zero in 1999.

5.2.2.2 Posterior Parameter Distribution

The posterior parameter distributions (after kernel smoothing; Figure 2.2 and Figure 2.3) and optimal values of flow parameters estimated by the multi- and single- response BC-GED approaches are shown in Figure 5.17.
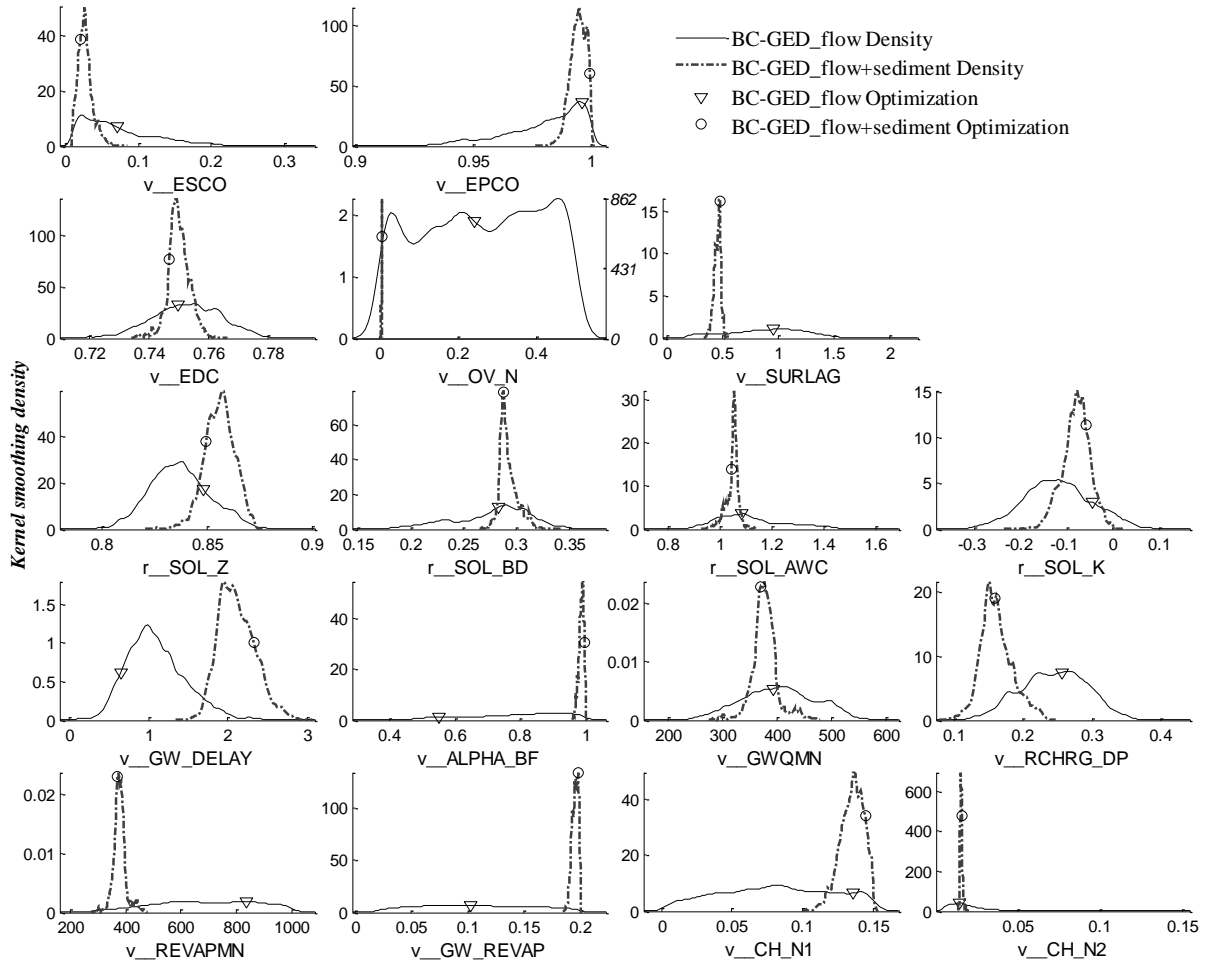
Figure 5.17 Comparison of the posterior distribution (after kernel smoothing) and optimal value (point) of flow parameters estimated by the single- (solid line) and multi-response (dash line) BC-GED approaches. The detailed descriptions of flow parameters are shown in Section 4.2.2 and Table 4.2.

Figure 5.17 shows the posterior distributions of flow parameters estimated by the multi-response BC-GED approach are obviously sharper than that estimated by the single-response method, and even some parameters (such as OV_N and CH_N2) nearly become the constant values. However, the optimal parameter values of the multi-response BC-GED approach, such as the effective soil depth (EDC) and the soil property parameters (SOL_Z, SOL_BD, SOL_AWC and SOL_K), are close to those of the single-response method, which

97

corresponds to the results in Table 5.3, where the average annual runoff components are nearly the same between the multi- and single- response BC-GED methods.

In Figure 5.17, however, the optimal values of groundwater parameters are different between the multi- and single- response BC-GED methods. For example, the values of the recharge partition coefficient (RCHRG_DP) of the multi-response method are obviously less than those of single-response method. But the values of the groundwater evaporation conversion coefficient (GW_REVAP) are greater, which corresponds to the results in Table 5.3 for BC-GED approaches. In that table, the amount of groundwater evaporation of the multi-response method is greater than that of single-response method, but the amount of deep aquifer recharge is less. It may result from the inter-correlation and equifinality of different parameters that lack physical meaning, e.g. the total groundwater loss (including the evaporation and deep percolation of shallow aquifer) is nearly the same between the multi- and single- response methods (Table 5.3).

The posterior parameter distributions and optimal values of sediment parameters estimated by multi-response BC-GED approach are shown in Figure 5.12. This figure shows the optimal value of USLE erodible factor of Regosols (USLE_K1) is greater than that of Luvisols (USLE_K2), and greater than that of Fluvisols (USLE_K3), i.e. the value of USLE_K1 (Regosols) > USLE_K2 (Luvisols) > USLE_K3 (Fluvisols). The posterior distribution of USLE_K1 (Regosols) is sharper than that of USLE_K2 (Luvisols), and sharper than that of USLE_K3 (Fluvisols), i.e. the uncertainty of USLE_K1 (Regosols) < USLE_K2 (Luvisols) < USLE_K3 (Fluvisols). The values of channel erodible factor (CH_EROD) approach zero, which means the main channels are nearly non-erodible. It corresponds to the result in the Table 5.4, where the sediment deposition occurs in all the main channels.

5.3   Comparison of Rainfall-Runoff Methods in SWAT

The five rainfall-runoff methods (CN-Soil, CN-ET, G&A, WB and WB-VSA) in SWAT are introduced in this study, as detailed in Section 4.1. In the Sections 5.1 and 5.2, the physics-based SWAT-WB-VSA developed in this study was employed in the automatic calibration procedure for likelihood function comparison and multi-response calibration. This section focuses on comparing the model performance of WB-VSA with other rainfall-runoff

methods, and studying the effect of the 'volumetric' efficiency (VE) proposed by Criss and Winston (2008) on model calibration with different model structures.

SWAT with the five rainfall-runoff methods are separately calibrated by automatic calibration method. In the automatic calibration procedure (Figure 2.1), the river discharges (Section 3.5.1) are used as the observed data, the DREAM (Section 2.2.1) as the optimization tool, the flow parameters (Table 4.2) as the calibrated parameters and the VE as the objective function, as detailed in Table 4.3 for the case of model comparison.

The results of the model calibration using river discharge data and the model validation using groundwater data are presented separately in the following.

5.3.1   Model Calibration

The optimal values of model parameters (corresponding to the maximum value of VE) and the corresponding model-performance indicators are listed in Table 5.5. The detailed descriptions of model parameters are shown in Section 4.2.2 and Table 4.2. In Table 5.5, "Calibration" performance is the maximum value of the objective function VE, and the *NSE* and R^2 are the widely used performance indicators (Legates and McCabe, 1999). Table 5.5 shows the optimal values of most parameters are distinct differences among the five approaches mainly because of the different rainfall-runoff structures. However, the values of some physically meaningful parameters (such as soil depth (SOL_Z), soil bulk density (SOL_BD) and GWQMN) are close among the physics-based rainfall-runoff methods (G&A, WB and WB-VSA). Among the five rainfall-runoff approaches, the largest value of model efficiency criterion (VE) is from the simulation of WB-VSA approach developed in this study (Table 5.5). However, the model performance indicators (e.g. VE and R^2) of the CN-Soil and WB approaches are close to those of the WB-VSA. Namely, after calibration, the performances of river discharge simulation of the CN-Soil, WB and WB-VSA approaches are nearly the same. By contrast, the G&A and CN-ET approaches, based on the infiltration excess and the evaporation-dominant mechanism of rainfall-runoff generation, respectively, give the poorest results.

Table 5.5 The optimized parameters of five rainfall-runoff approaches and the corresponding model-performance indicators.

Categories [1]	Parameters	CN-Soil	CN-ET	G & A	WB	WB-VSA
Evapotranspiration	v__ESCO	0.515	0.040	0.052	0.468	0.198
	v__EPCO	0.021	0.769	0.991	0.258	0.128
	r__CN2	-0.394	-0.415	-0.191	—	—
	v__EDC	—	—	—	0.765	0.781
Surface water	v__CNCOEF	—	0.192	—	—	—
	v__OV_N	0.351	0.106	0.117	0.404	0.412
	v__SURLAG	0.994	0.510	6.548	1.783	2.550
Soil water	r__SOL_Z	1.403	0.819	0.886	0.865	0.890
	r__SOL_BD	0.835	0.227	0.428	0.354	0.405
	r__SOL_AWC	0.717	1.918	0.004	2.061	1.834
	r__SOL_K	-0.839	-0.311	0.328	-0.185	-0.195
Ground water	v__GW_DELAY	0.210	0.392	0.373	0.924	0.859
	v__ALPHA_BF	0.182	0.995	0.973	0.970	0.936
	v__GWQMN	297.803	382.189	438.765	582.048	548.907
	v__RCHRG_DP	0.447	0.031	0.001	0.075	0.061
	v__REVAPMN	749.327	829.650	397.525	561.083	527.690
	v__GW_REVAP	0.172	0.060	0.199	0.198	0.183
Tributary/main channel	v__CH_N1	0.047	0.018	0.127	0.006	0.120
	v__CH_N2	0.007	0.056	0.017	0.013	0.031
Calibration [2]	VE	0.6071	0.5716	0.5777	0.6059	0.6075
Model performances	NSE	0.7992	0.6025	0.6258	0.8188	0.8224
	R^2	0.8482	0.6935	0.7182	0.8481	0.8486

[1] The dash means the rainfall-runoff method without corresponding model parameter.

[2] The efficiency coefficient of model calibration.

Comparison of the observed and simulated river discharges on the daily time step in the period of 1993 – 2011 for five rainfall-runoff approaches is shown in Figure 5.18. In this figure, discharge results of the five rainfall-runoff approaches are produced by SWAT with the corresponding optimum parameter set in Table 5.5. The black column (top) is the average rainfall over watershed. Figure 5.18 shows all the five rainfall-runoff methods after calibration can reproduce most flood events. However, CN-Soil, WB and WB-VSA approaches can capture the floods better than the G&A and CN-ET approaches. It is also reflected by the value of NSE: the NSE values of CN-Soil, WB and WB-VSA approaches are obviously greater than that of G&A and CN-ET approaches.
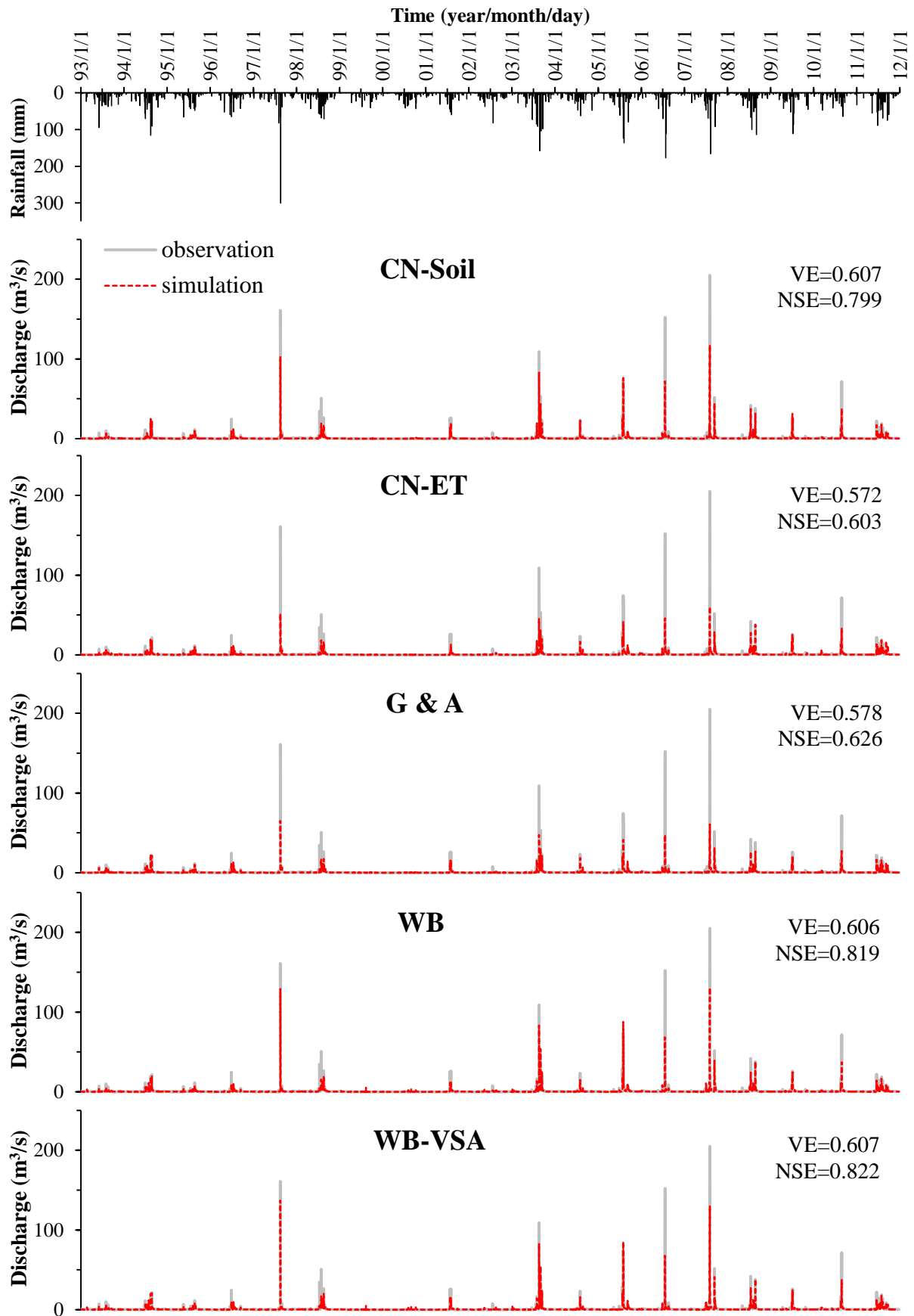
Figure 5.18 Comparison of the simulated and observed river discharges for CN-Soil, CN-ET, G&A, WB and WB-VSA approaches.
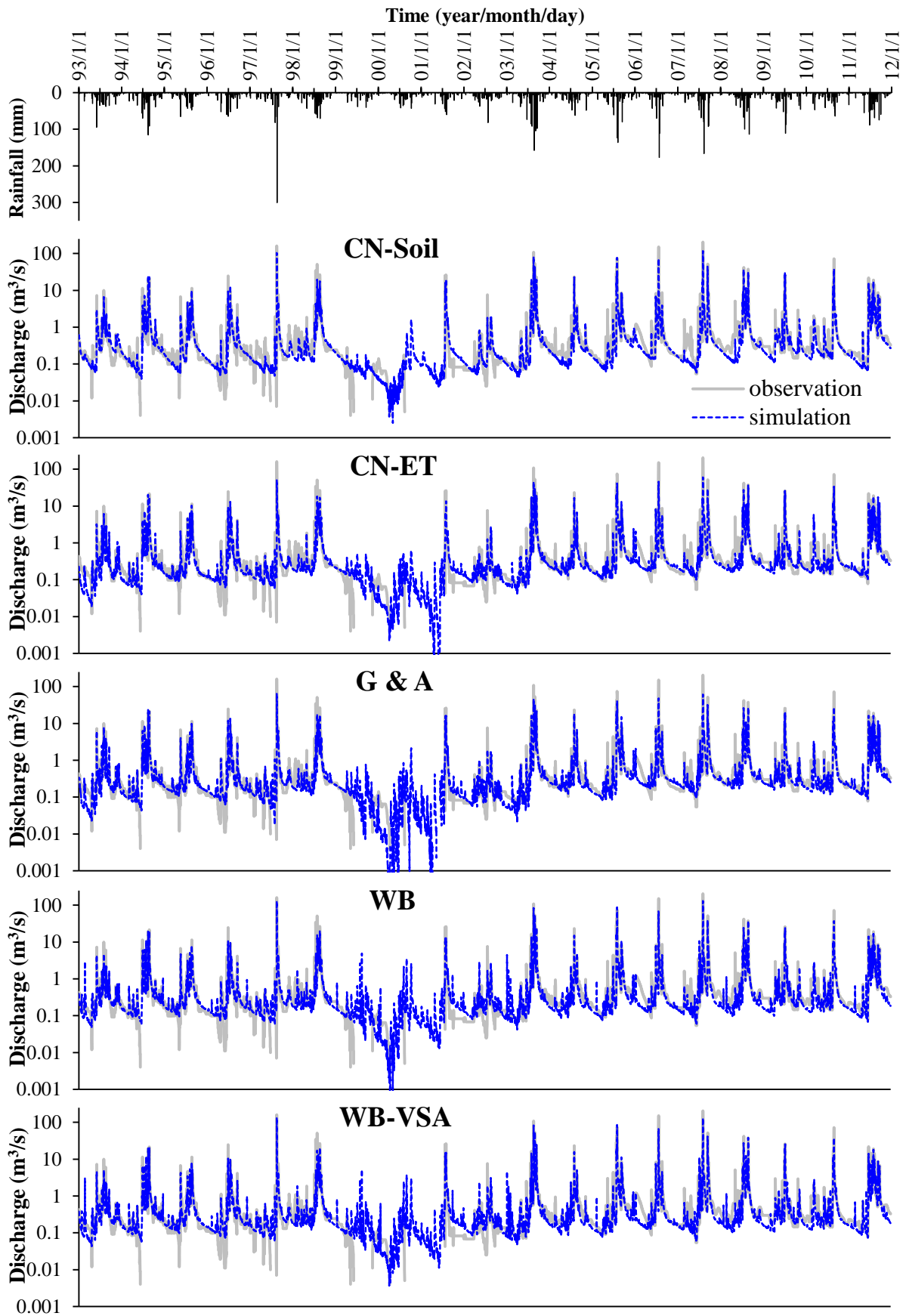
Figure 5.19 Comparison of the simulated and observed river discharges on the logarithmic y-axis for CN-Soil, CN-ET, G&A, WB and WB-VSA approaches.

In order to highlight baseflow, the observed and simulated river discharges are plotted in Figure 5.19 with logarithmic vertical/y-axis (base 10). This figure shows all the rainfall-runoff approaches, after calibration against observed river discharges using the objective function of VE, can mimic the river discharges (especially the baseflow) well. However, these approaches cannot capture extremely small flow closed to zero at the beginning of rainy season, especially in the driest years such as 1999 and 2000, probably because of neglecting the effect of agricultural irrigation on the river discharges during the spring ploughing season in this study (Section 3.5.2).

Five rainfall-runoff methods of SWAT with corresponding optimum parameter set (Table 5.5) are separately used to simulate the hydrologic processes in Baocun watershed. The average annual components of model-predicted runoff during 1993 - 2011 are shown in Table 5.6. The detailed descriptions of runoff components are introduced in Section 4.1.

Table 5.6 Comparison of the average annual runoff components during 1993 - 2011 among CN-Soil, CN-ET, G&A, WB and WB-VSA approaches.

Categories (mm/yr)	CN-Soil	CN-ET	G & A	WB	WB-VSA
Evapotranspiration	490.3	539.5	476.1	476.8	485.3
Overland flow	112.38	7.5	0.0	44.35	45.74
Interflow	62.92	91.04	119.27	117.76	119.92
Return flow	82.27	155.82	134.76	106.25	104.63
Revaporization [1]	0.0	0.0	79.34	50.99	45.29
Deep percolation [2]	68.37	5.17	0.22	13.6	10.37

[1] The evaporation of groundwater.

[2] The groundwater losses from the shallow aquifer to the deep aquifer.

Table 5.6 shows that the overland flow substantially differs among the five approaches: the amount of overland flow simulated by the CN-Soil is the largest, which corresponds to the viewpoint of Arnold et al. (2011) that the CN-Soil predicted too much overland runoff in the shallow-soil area. By contrast, the amount of overland flow is negligible in the CN-ET and G&A approaches. However, the amount of subsurface flow (including interflow and groundwater return flow) accounts for the largest proportion of runoff components, i.e. 56%, 97%, 100%, 83%, and 83% for CN-Soil, CN-ET, G&A, WB and WB-VSA, respectively. The evapotranspiration is the main pathway of water losses from the watershed for all five approaches. However, among the five methods, the shallow groundwater losses are different: for the CN-Soil, the deep recharge is the main way of groundwater losses, whereas the

revaporization/evaporation is the main way for three physics-based approaches (G&A, WB and WB_VSA). For the CN-ET, the groundwater losses are negligible and the smallest among the five approaches, but the evapotranspiration is the largest.

5.3.2   Model Validation

As analyzed in Sections 3.5.3 and 5.1.3, for model validation, we can test the linear relationship between the observed groundwater level, and the simulated soil water volume of soil profile and shallow aquifer in the hydrologic response unit (HRU) where the groundwater gauge fell. Comparison of the observed groundwater level and the simulated soil water volume for the five rainfall-runoff approaches are presented in Figure 5.20. The time series of observed groundwater level is shown in Figure 3.8. In Figure 5.20, the simulated soil water volume including the soil water of soil profile and the groundwater of shallow aquifer is produced by SWAT with the five rainfall-runoff approaches using the corresponding optimum parameter set in Table 5.5. The performance indicator is the coefficient of determination (R^2), which reflects the linear correlation between the observed groundwater level and the simulated soil water volume.

Figure 5.20 shows, among the five rainfall-runoff methods, the empirical rainfall-runoff approach of CN-Soil mimics the groundwater levels worst. Even the value of R^2 of CN-ET, of which the surface runoff generation is independent of soil water content, is much greater than that of CN-Soil. By contrast, the physics-based rainfall-runoff approach (WB-VSA) developed in this study mimics the groundwater levels best.
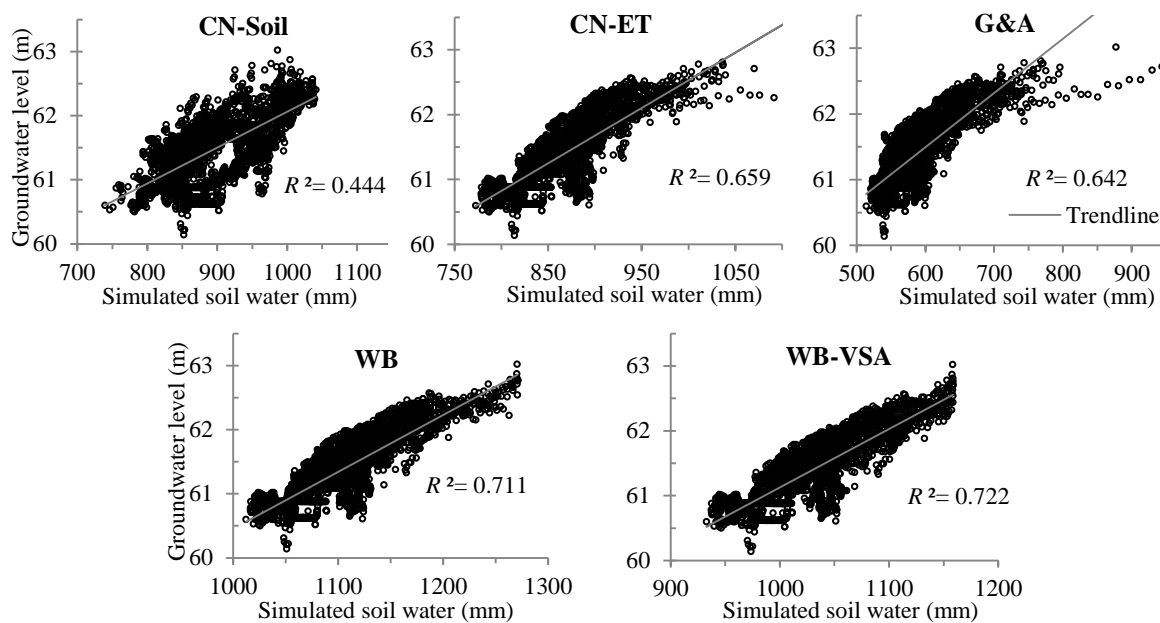
Figure 5.20 The observed groundwater levels versus the simulated soil water volume for
CN-Soil, CN-ET, G&A, WB and WB-VSA approaches. The larger the value of
R^2, the better the performance of groundwater simulation is.

6 Discussion

Chapter 5 presents the model calibration results with different likelihood/objective functions. This Chapter attempts to evaluate the effects of different likelihood/objective functions on model calibration based on the results shown in Chapter 5, and find out the characteristics of objective functions and the corresponding reasons. This Chapter firstly analyzes the structure of different likelihood functions proposed by this study (Section 6.1); next, compares the effects of likelihood functions on model predictions (Section 6.2), and then on the posterior parameter distributions (Section 6.3); after that, interprets the relationship between the distance-based objective function and the likelihood function (Section 6.4); and finally discuss the performances of the different rainfall-runoff methods after calibration using the objective function VE (Section 6.5).

6.1 Form of Likelihood Functions

The difference between the formal and inform likelihood function is whether the model residuals fulfill the statistical assumptions of the likelihood function or not (Section 2.2.2). Formal likelihood function is very helpful. Beven and Binley (2013) pointed out only the formal likelihood function is valuable for uncertainty analysis of model predictions in the formal Bayesian approach. Otherwise the approximate Bayesian computation, e.g. the GLUE, is more practical to uncertainty analysis (Section 2.2.2.2). The informal likelihood function, as an objective function in the automatic calibration procedure, may cause the fallacious/unreasonable optimization results (Clarke, 1973; Xu, 2001; Jain and Sudheer, 2008). This study presents three likelihood functions (also termed error models): *NSE*, BC-GED and BC-SGED (Table 4.4). The results of both single- (Section 5.1) and multi- (Section 5.2) response calibration demonstrate that model residuals inferred by *NSE* approach exhibit heteroscedasticity (Figure 5.3 and Figure 5.8) and non-normality (Figure 5.4 and Figure 5.10), which obviously violate the statistical assumptions of *NSE* approach (Table 4.4). Therefore, *NSE* approach is an informal likelihood function. By contrast, BC-GED and BC-SGED approaches make model-residuals fulfill their statistical assumptions (Table 4.4) much better than *NSE* approach (Figure 5.3, Figure 5.4, Figure 5.14 and Figure 5.16).

From the viewpoint of the strictly/explicitly formal likelihood function (Section 2.2.2), however, the BC-GED and BC-SGED error-models may lose the connection between residual assumptions and the likelihood of residuals because they estimate the Box-Cox

transformation (BC) parameter (i.e. lambda/λ in Eq. (4.3:10)) by the minimum variance constraint (Eq. (4.3:12)), rather than the likelihood function (Eq. (4.3:11)). The reason why the hydrological BC parameter (Eq. (4.3:10)) cannot be inferred by likelihood function has been analyzed in the Section 4.3.1.2.1 and the Appendix B. The key reason is that the likelihood function (Eq. (4.3:11)) does not seem correct for hydrological BC, because the formula of hydrological BC (Eq. (4.3:10)) is different from that of original BC (Eq. (4.3:9)), and we cannot simply transplant the Jacobian determinant from the original BC to the hydrological BC. Here, we discuss why the BC parameter (lambda/λ) can be estimated by the minimum variance constraint (Eq. (4.3:12)).

In fact, Box-Cox transformation (BC) should be treated as an implicit model to remove the heteroscedasticity (i.e. inconstant variance) of model residuals, in contrast to the explicit statistical model, e.g. the error standard deviation is modeled as a linear function of simulated streamflow (Schoups and Vrugt, 2010) or a function of time (Pianosi and Raso, 2012). The parameters of the explicit error-model can be inferred by the likelihood function, because the explicit error-model has the Jacobian determinant of transformation (Schoups and Vrugt, 2010).
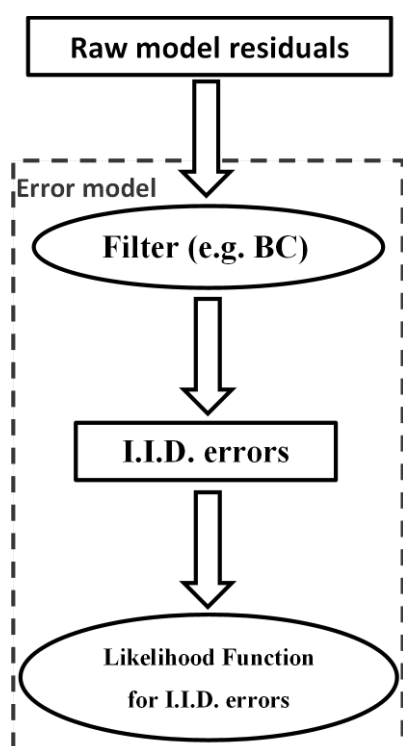
Figure 6.1 Structure of implicit error-model.

The structure of implicit error-model is shown in Figure 6.1. The implicit error model (Figure 6.1) first transforms the raw model residuals (between model predictions and observed data; Figure 2.2) into independent and identically distributed (I.I.D.) errors by implicit error-transformation approaches (e.g. Box-Cox transformation (BC) method), and then estimates the likelihood of I.I.D. errors using error distribution function (e.g. GED, Eq. (4.3:13)). Fixed transformation parameters of the filter/BC (Figure 6.1) by most studies obviously impaired the effectiveness of the filter (Bates and Campbell, 2001; Vrugt et al., 2009; Engeland et al., 2010). Therefore, for improving the performance of the filter, the transformation parameters (e.g. lambda of BC) of the filter could be estimated on the basis of specific constraints (e.g. Eq. (4.3:12)).

The application of this study shows the range of the inferred values of Box-Cox transformation parameter (lambda) (i.e. [0.432, 0.445] in Table 5.1) is very small, which means that the scheme of BC with minimum variance constraint (Eq. (4.3:12)) approximates the scheme of BC with fixed lambda. However, the scheme of BC with Eq. (4.3:12) increases the flexibility and effectiveness of BC (i.e. the filter in Figure 6.1) to remove the heteroscedasticity of model residuals.

In the implicit error model (Figure 6.1), we do not need to know the explicit expression of the probability density function (PDF) of raw model residuals. Actually, in most cases, it is difficult or impossible to get the explicit expression, because of complex error-sources of raw model residuals in hydrology, such as the model input, structural and observed output errors (Figure 2.6; Beven and Binley, 2013). Therefore, the implicit error model is more practical and robust than the explicit error-model.

Figure 6.1 and Figure 2.2 show the filter of BC with minimum variance constraint (Eq. (4.3:12)) is not isolated from the likelihood of model residuals. The connection between the filter and likelihood is implicit: the filter changes the transformation parameters to affect the likelihood, and the likelihood affects the filter via model predictions (Figure 2.2).

6.2 Effects of Likelihood Functions on Simulation Results

6.2.1 Why does the *NSE* always put greater emphasis on High Values?

Comparison of best simulated river discharges corresponding to the maximum likelihood value among *NSE*, BC-GED and BC-SGED approaches shows that the SWAT-WB-VSA with *NSE* approach can mimic flood events well (Figure 5.1), but simulate the baseflow badly (Figure 5.2). By contrast, the SWAT-WB-VSA with the BC-GED or BC-SGED approach simulates the baseflow much better (Figure 5.2). With the multi-response *NSE* approach, the SWAT-WB-VSA also captures extreme flood events well (Figure 5.7). For example, in 1997, the simulated peak flow and sediment are 158.40 m^3/s and 603.47 kg/s, respectively, which are nearly equal to the observed results that are 160.99 m^3/s and 602.94 kg/s, respectively. However the relative errors of baseflow and low sediment load are high, e.g. in 1999 (shown in Table 5.2). The defectiveness of the *NSE* approach is also pointed out by many authors (Legates and McCabe, 1999; Krause et al., 2005; Pushpalatha et al., 2012; Beven and Binley, 2013): *NSE* tends to put more weight on the higher values than on the lower values. One

possible explanation is that *NSE* squares the differences between the observed and simulated values (Legates and McCabe, 1999; Krause et al., 2005).

Similarly, from the likelihood viewpoint, it attributes the statistical assumption that errors/residuals follow Gaussian distribution with zero mean, which can be further explained by the relationship between the error interval and its probability in Figure 6.2, where the σ is the standard deviation of errors. There are high probabilities (68.3%) of errors/residuals nearly uniformly distributed in the error interval [-σ, σ] instead of approaching to zero because of the evenly distributed Gaussian error density in Figure 5.4 (a) and the linear trend for the relationship between the error intervals and the corresponding probabilities in Figure 6.2. It leads to a large relative error of the low-flow/baseflow as shown in Figure 5.2. The probability of error interval [-2σ, 2σ] is 95.5%, and that of [-3σ, 3σ] is 99.7% in Figure 6.2. In other words, the probability of the absolute error more than 3σ is only 0.3%. So there are a few very large errors resulting in the small relative error of high-flow/flood (Figure 5.1).

In summary, the low probability of the large error and the equiprobability of the small error in the Gaussian distribution assumed by *NSE* result in the small relative error of flood and the large relative error of baseflow, respectively.
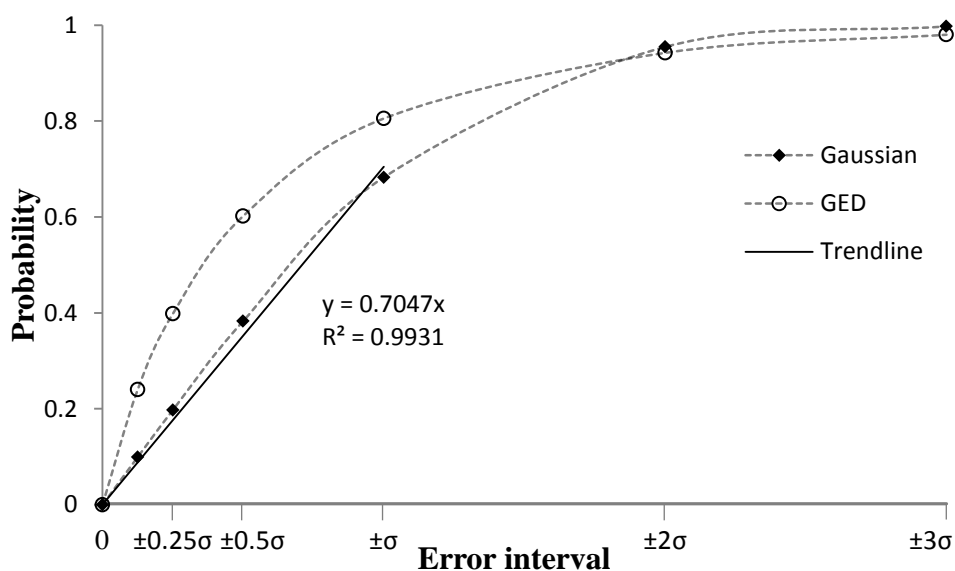
Figure 6.2 Probabilities of different error intervals in Gaussian distribution and GED. The Y-value (in the vertical axis) is the probability of error interval, e.g. the probability of error within the interval [-σ, σ] (i.e. ±σ) is about 68% for Gaussian distribution.

By contrast, the error distribution of GED sharply concentrates on zero error (Figure 5.4 (b)), and the ratio between the probability and the error interval quickly increases with the

narrowing of the error interval (Figure 6.2) in this study. Therefore, the optimization results of the likelihood function with GED guarantee that most errors/residuals approach zero. On the other hand, the Box-Cox transformation of the BC-GED approach leads to amplification of the error of low-flow/baseflow and mitigation of the error of high-flow/flood. So the BC-GED approach generates very small errors of low flows (Figure 5.1) but relatively large errors of high flows (Figure 5.2). For the BC-SGED approach, because the model-error distribution is similar to that of the BC-GED approach (Figure 5.4), of which the skewness is insignificant and close to one, i.e. non-skewness (Table 5.1 and Figure 5.5), the BC-SGED approach also puts greater emphasis on low-flow/baseflow. Nonetheless the results of the BC-SGED approach easily fall into the local optima because of over-parameterization (i.e. the skewness coefficient (ξ/xi); Eq. (4.3:14)) of the error model.

The above analysis demonstrates that SWAT-WB-VSA cannot fully capture flow processes in the study area in terms of error distribution of the three likelihood functions, possibly because of the extremely nonuniform temporal distribution of rainfall in the Baocun watershed owing to the great intensity of typhoon rainfall. Meanwhile, only using observations of the river discharges may be not enough to evaluate the three approaches (*NSE*, BC-GED and BC-SGED) owing to too many parameters in the SWAT model. Comparatively, the formal likelihood approaches (BC-GED and BC-SGED) can mimic baseflow and groundwater level better (Figure 5.6) because of the stronger linear relationship between the observed groundwater level and the simulated soil water volume, and the more reasonable specific yield of unconfined aquifer (Eq. (5.1:1)). It is confirmed that the formal likelihood approaches estimate the soil layer properties better than the *NSE* approach.

6.2.2   Effects of Multi-Response Likelihood Functions

Runoff. Comparison of the best simulation results (corresponding to the maximum value of likelihood function) between the multi- and single- response *NSE*/BC-GED approaches demonstrates the average annual runoff components (Table 5.3) and the most optimal flow-parameters (Figure 5.11 and Figure 5.17) are nearly the same, which means that the effect of the multi-response *NSE*/BC-GED approach on the river discharges is small. It may result from two reasons:

1. Parameter inter-correlation and equifinality occur in SWAT model. For example, the soil erosions from HRUs (4.1:13) depend on not only the surface runoff but also the

soil erodibility factor (USLE_K). In addition, USLE_K as a characteristic parameter of soil type varies with different soil types (Table 4.2). As a consequence, it is unnecessary to tune flow parameters (related to surface runoff) to improve model predictions of sediment loads when calibrating the flow and sediment parameters simultaneously.

2. The flow parameters related directly to the sediment transport (such as manning roughness coefficients of sloping surface (OV_N; Eq. (4.1:16)) and main channel (CH_N2; Eq. (4.1:20))) are insensitive to river discharges (Table 5.1 and Figure 5.5), although they are very sensitive to sediment transport (Figure 5.11 and Figure 5.17). Therefore, changes of OV_N and CH_N2 affect runoff components and river discharges slightly.

However, the small effect of multi-response calibration method on river discharges does not mean that it is unnecessary to calibrate the flow and sediment parameters simultaneously. On the contrary, it reveals some advantages of the multi-response calibration method: (1) it can improve the performance of sediment simulation without impairing the performance of flow simulation; and (2) it can reduce the uncertainty of flow parameters, especially flow concentration parameters (such as OV_N and CH_N2).

Soil erosion. The annual erosion rate of HRUs estimated by the BC-GED approach (290.5 $t/(km^2 \cdot a)$, i.e. about 0.22 (mm/a) with the soil bulk density of 1.35 g/cm^3; Table 5.3) is a little larger than that estimated by the *NSE* approach (207.9 $t/(km^2 \cdot a)$, $\approx$ 0.15 (mm/a); Table 5.3). According to the China standards for soil erosion published by China Ministry of Water Resources (2008), both the simulated results all indicate the Baocun watershed belongs to the mild erodible gradation that ranges from 200 to 2500 $t/(km^2 \cdot a)$ in the classification of the earth and stone area of northern China.

However, the inferred erodibility of the main channel is distinct differences between the *NSE* and BC-GED approaches (Table 5.3). For the *NSE* approach, more than 36% of sediment loads in the watershed outlet are from erosion of the main channel. By contrast, nearly 18% sediments eroded from slopes/HRUs deposit in the main channel for the BC-GED approach. Field survey shows local residents have built some small dams in the main channel for water supply, irrigation and transportation (Figure 3.7; Section 3.5.2). These "reservoirs" have small effect on the daily runoff because of very small storage capacity, but obviously affect sediment transport because of reducing flow velocity (except of extreme flood owing to very low dam shown in Figure 3.7) that impacts on the sediment transport capacity of flow and

results in sediment deposition. Therefore, small dams intercepted many sediments in the main channel, which is also reflected in the recorded sediment-load data that suddenly fell to zero values after the flood peak (Figure 3.9).

The *NSE* approach fails to make the SWAT model capture the characteristic of sediment transport in Baocun watershed because the *NSE* always puts greater emphasis on the high values but neglects the small observed data (e.g. zero values of sediment loads; Section 6.2.1). By contrast, the BC-GED approach as a formal likelihood function can infer the non-erodibility of main channel, which may result from the amplification of the model residual of small sediment loads and the greater emphasis on residuals around zero (Section 6.2.1; Figure 6.2). This difference between *NSE* and BC-GED approaches is also reflected in the inferred posterior distribution of channel erodibility factor (CH_EROD; Figure 5.12) that determines the erodibility of main channel. The CH_EROD estimated by BC-GED approach almost equals zero, which corresponds to the non-erodible channel. By contrast, the CH_EROD estimated by *NSE* approach has much larger value and wider range (i.e. higher uncertainty).

6.3 Effects of Likelihood Functions on Posterior Parameter Distribution

6.3.1 Single-Response Likelihood Function Approach

Most posterior parameter distributions are different between the *NSE* and the formal-likelihood (BC-GED and BC-SGED) approaches as shown in Figure 5.5. But between the BC-GED and the BC-SGED approaches, posterior parameter distributions are nearly the same, except those of some parameters relating to the groundwater movement, probably because of the dependence and the equifinality of different parameters that lack physical meaning. For example, although the BC-GED and the BC-SGED approaches all inferred the slow groundwater-recession processes (Figure 5.2), they attributed the smallest baseflow recession coefficient (ALPHA_BF) for the BC-GED approach and the longest groundwater delay time (GW_DELAY) for the BC-SGED approach. By contrast, the *NSE* approach inferred an unreliable result (Figure 5.2), i.e. the rapid recession of groundwater with the largest ALPHA_BF and the shortest GW_DELAY.

For evapotranspiration processes, the plant transpiration is strong (large value of EPCO in Table 5.1 and Figure 5.5) for the formal-likelihood (BC-GED and BC-SGED) approaches, which possibly result from good vegetation/crop cover in the Baocun watershed.

For surface flow processes, the surface-runoff generation capacity is small and nearly the same for different topography profiles because of the large EDC (Eq. (4.1:9)) for all approaches (Table 5.1). It corresponds to the terrace cultivation method in the Baocun watershed because the terraced fields weaken the effect of topography while improve infiltration capacity (Figure 3.4).

For soil water, the results of the formal likelihood approaches show larger soil thickness (SOL_Z), porosity (which is inversely proportional to the bulk density (SOL_BD)), available water capacity (SOL_AWC) and hydraulic conductivity (SOL_K) than those of the *NSE* approach. This means that the soil layer of the formal likelihood approaches maintain more water and generate more lateral flow, and the parameter values of soil property for the formal-likelihood approaches agree better with the field-investigated results, especially SOL_K of which the relative change to the field-investigated value is close to zero.

For main channel routing, the channel storage capacity (CH_N2) of the *NSE* approach is much larger than that of the formal-likelihood approaches, indicating that the formal-likelihood approaches are more reasonable in neglecting the storage function of the main channel because of only two hours concentration time of flood in the Baocun watershed modeled by daily hydrological model.

The mean errors of the *NSE* approach are greater than zero. By contrast, those of the formal-likelihood (BC-GED and BC-SGED) approaches close to zero. Therefore, the formal-likelihood approaches maintain water balance better than the *NSE* approach.

In conclusion, the posterior parameter distributions estimated by the formal likelihood (BC-GED and BC-SGED) approaches are more reasonable than those estimated by the informal-likelihood approach (*NSE*). For the two formal-likelihood approaches, the results of Bayesian inferences (including the best simulated river discharges and soil water volumes, and the posterior parameter distributions) are all nearly the same, because the model residuals after BC are all symmetric (Figure 5.4) and the SGED degrades into the GED. Meanwhile, it demonstrates the assumption of skewness of the error model may be unnecessary, because the ideal model-residuals should be unbiased (i.e. the mean is zero) and most of them should be zero (i.e. the mode (the highest probability point) is zero), which will result in the symmetry of the error model with zero-mean and zero-mode.

6.3.2 Multi-Response Likelihood Function Approach

Flow parameters. The posterior distributions of flow parameters inferred by the multi-response *NSE*/BC-GED approach (with river flow and sediment objectives) are all sharper than those inferred by the single-response *NSE*/BC-GED approach (only with river flow objective) (Figure 5.11 and Figure 5.17), especially parameters related to surface runoff and river flow velocity (such as OV_N, SURLAG, CH_N1 and CH_N2; Table 4.2). It demonstrates that there are clearly trade-offs between the flow and the sediment objectives, and the multi-response likelihood approaches (Eq. (4.3:17)) decrease the uncertainty of model parameters. It is confirmed that improving information can reduce the uncertainty of model parameters (Goodman, 2002). But the multi-response approaches increase the dimensionality of model calibration, which leads to numerous secondary optima (e.g. the multimodality of posterior distributions of flow parameters in Figure 5.11) and increases the difficulty of the global optimization (Duan, 2003).

Sediment parameters. The soil USLE erodibility factors (USLE_K) inferred by *NSE* approach are completely different to those determined by BC-GED approach. Driessen et al. (2001) summarized the major soils of the world and pointed out that:

1. Regosols in sloping areas is prone to erosion because of low coherence of the matrix material.
2. Luvisols usually has a stable blocky structure except high silt content of surface soil that may be sensitive to slaking and erosion.
3. Fluvisols erosion/deposition depends on its location: usually the erosion occurs in the upper reach and the deposition occurs in the lower reach.

According to the study of Driessen et al. (2001), the USLE_K estimated by BC-GED approach may be more reasonable (Figure 5.12): value of USLE_K1 (i.e. USLE erodible factor of Regosols) is the largest, value of USLE_K2 (Luvisols) is less, and value of USLE_K3 (Fluvisols) is the least with highest uncertainty owing to the wide range slope (i.e. 9.5~24.3 *TI*) of Fluvisols in Baocun watershed, where the erosion and the deposition occur together.

The parameters of PRF, SPON and SPEXP are used to calculate the channel sediment transport capacity (Eq. (4.1:19)). Figure 5.12 shows values of PRF and SPON estimated by the BC-GED approach are smaller than those estimated by the *NSE* approach, but the value of SPEXP is obviously larger. As a result, in the high flow, channel sediment transport capacity

estimated by the BC-GED approach is larger than that estimated by the *NSE* approach, but smaller in the low flow. It may result from that the SWAT with BC-GED approach under-estimates some flood peaks (Figure 5.13), and there are large differences between the high- and low- values of sediment loads (Figure 3.9).

Although Neitsch et al. (2005) recommended the closed interval [1, 2] as the range of the SPEXP that reflects the power of the flow velocity transporting sediments, the simplified Bagnold equation (Eq. (4.1:19)) is a half-theoretical and half-empirical formula (Neitsch et al., 2005), and even the study of Guo (2002) showed the power of the flow velocity in Yangtze river can reach 4.5. So the value of SPEXP (power) depends on the morphology of the channel and the SPEXP estimated by BC-GED approach may be also reasonable.

In summary, the erodibility of soil and main channel inferred by the multi-response formal likelihood (BC-GED) approaches are more reasonable than that inferred by the multi-response *NSE* approaches, according to the field survey and previous studies. Comparing with the single-response likelihood (*NSE*/BC-GED) approaches, the multi-response approaches reduce the uncertainty/equifinality of the model parameters, but increase the difficulty of the global optimization.

6.4 Relationship between Classical Objective Function and Likelihood Function

6.4.1 Unifying Distance-Based Objective Functions

Both the classical objective function and the likelihood function can be used to ascertain the performance/goodness-of-fit of hydrological model. However, they are based on different principles. The distance-based objective function uses the spatial distance method to measure the differences between model predications and observation data (Table 2.1). The likelihood function is based on the statistical theory to estimate the joint probability of model residuals (Eq. (2.2:3)), of which the basic assumption is that the model residuals are independent and identically distributed (I.I.D.). In Section 4.3.1.1, we proved that the *NSE* is equivalent to a likelihood function with Gaussian I.I.D. residuals, and in Section 6.2.1, we used the Gaussian assumptions of *NSE* to successfully interpret the reason why the *NSE* always puts greater emphasis on high values but less on the small values.

By extending the derivation method of *NSE* in the Section 4.3.1.1 to the BC-GED error model that assumes the model residuals after Box-Cox transformation (BC) follow the generalized error distribution (GED) with zero-mean (Figure 6.2), we get an equivalent objective function of BC-GED error model (as detailed in Appendix C):

$$\text{BC-GED} = \frac{\Gamma[1/\beta]}{\beta} \sqrt[\beta]{\varepsilon\beta \frac{\sum_{1}^{n}|e_i|^{\beta}}{n}} = \frac{\Gamma[1/\beta]}{\beta} \sqrt[\beta]{\varepsilon\beta \frac{\sum_{1}^{n}|g(obs_i) - g(sim_i)|^{\beta}}{n}} \quad (6.4:1)$$

where

BC-GED is a distance-based objective function,

e_i is the error/residual between observed (obs_i) and simulated (sim_i) outcomes at time step i,

β is the power of errors,

$\Gamma[\text{x}]$ is the gamma function evaluated at x,

ε is the base of the natural logarithm (≈ 2.718),

$g(x)$ is the Box-Cox transformation function (Eq. (4.3:9)) evaluated at x,

n is the length of time series of errors.

The maximum likelihood (i.e. BC-GED error model) estimation of model parameters is equivalent to the minimization of BC-GED (objective) function (Eq. (6.4:1)) estimation, as detailed in Appendix C. The BC-GED (objective) function has two parameters (Eq. (6.4:1)): the Box-Cox transformation parameter λ (Eq. (4.3:9)) and the kurtosis coefficient β (also termed the error power) of generalized error distribution (GED; Eq. (4.3:13)). The distance-based objective function proposed by other researchers (Table 2.1) can be unified by the BC-GED (objective) function (Eq. (6.4:1)), and their corresponding parameter values of BC-GED function are shown in Table 6.1.

Table 6.1 reveals that all the distance-based objective functions imply the statistical assumptions, e.g. the mean squared error (MSE) and the mean absolute error (MAE) imply statistical assumptions that the model residuals follow the Gaussian distribution and the Laplace distribution with zero-mean, respectively. Table 6.1 also shows the square-root, log and inverse transformation methods are all the special cases of the Box-Cox transformation (BC) method, of which the goal is to remove the heteroscedasticity (i.e. inconstant variance) of model residuals via amplification of the model residual of low values (e.g. baseflow) and mitigation of the model residual of high values (e.g. flood).

Table 6.1 Unifying the distance-based objective functions by BC-GED error model.

Objective functions	Formula [1]	Characteristics	BC-GED λ [2]	β [3]
Mean squared error (MSE)	$\frac{1}{n}\sum_1^n (obs_i - sim_i)^2$	Equivalent to *NSE*	1.0	2.0
Mean absolute error (MAE)	$\frac{1}{n}\sum_1^n |obs_i - sim_i|$	Balance consideration of the high- and low- flows	1.0	1.0
Mean quadrupled error (MS4E)	$\frac{1}{n}\sum_1^n (obs_i - sim_i)^4$	Put greater emphasis on high flows	1.0	4.0
Square-root transformed MSE (RTMSE)	$\frac{1}{n}\sum_1^n (\sqrt{obs_i} - \sqrt{sim_i})^2$	Put equal emphasis on high- and low- flows	0.5	2.0
Log transformed MSE (LTMSE)	$\frac{1}{n}\sum_1^n (\log(obs_i) - \log(sim_i))^2$	Put greater emphasis on low flows but still shows sensitivity to high flows	0.0	2.0
Inverse transformed MSE (ITMSE)	$\frac{1}{n}\sum_1^n (\frac{1}{obs_i} - \frac{1}{sim_i})^2$	Focus on low flows totally	-1.0	2.0
Generalized efficiency (E_j) [4]	$1 - \frac{\sum_1^n |obs_i - sim_i|^j}{\sum_1^n |obs_i - \overline{obs}|^j}$	Modelers define the power (j)	1.0	j

[1] where obs_i and sim_i are the observed and the simulated outcomes at time step i, respectively, n is the length of outcomes, $\overline{obs}$ is the mean observed outcomes, j is the power.

[2] λ is the Box-Cox (BC) transformation parameter (Eq. (4.3:9)).

[3] β is the kurtosis coefficient of generalized error distribution (GED) and also termed the power of model residuals (Eq. (6.4:1)).

[4] Generalized efficiency is only a normalization of j-th power distance-based objective functions.

6.4.2 Why are Objective Functions in favor of different Hydrographic Components?

After numerous practical applications of automatic calibration using different objective functions, it is widely acknowledged that the different objective functions are in favor of different hydrographic components (Green and Stephenson, 1986; Legates and McCabe, 1999; Krause et al., 2005; Dawson et al., 2007). The characteristics of different objective functions are summarized in Table 2.1 and Table 6.1. These tables show that the bigger the value of error power (β), the greater emphasis is on high values, e.g. the mean quadrupled error (MS4E; $\beta=4$) put greater emphasis on high flow/values than MSE ($\beta=2$), and greater than MAE ($\beta=1$). And the transformation methods (e.g. square-root- or log- transformation) can make the objective functions put greater emphasis on low-values/baseflow.

In Section 6.2.1, we discuss the effect of Gaussian error distribution assumed by *NSE* on the runoff components. Here, we attempt to reveal the effect of the kurtosis coefficient (β) of GED (Table 6.1) on hydrolographic components.

The formulation of Generalized Error Distribution (GED) is shown in Eq. (4.3:13). The densities of the GED with zero-mean for various values of the kurtosis (β) are shown in Figure 6.3. In this figure, "$\beta = 2.0$" is Gaussian distribution; "$\beta = 1$" is Laplace distribution; and "$\beta = +\infty$" is uniform distribution. Figure 6.3 shows the larger the value of kurtosis (β), the lower probability of large model residual is, e.g. when β approaches infinity, all model residuals are limited into the interval $[-\sqrt{3}\sigma, \sqrt{3}\sigma]$ (where σ is the standard deviation of model residuals). As a consequence, the relative error of high-flow/flood is smaller. Therefore, the greater emphasis on high values is put by the objective function with high value of β, e.g. $\beta = 2$ for MSE/NSE (Section 6.2.1; Table 2.1 and Table 6.1) and $\beta = 4$ for MS4E (Table 6.1).
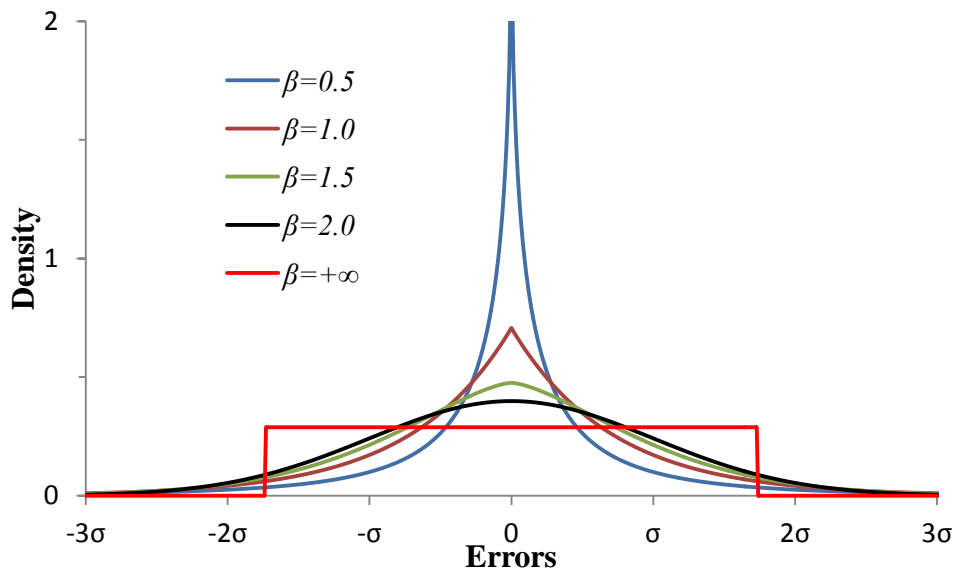
Figure 6.3 Probability density functions of the GED with zero-mean for various values of the kurtosis (β).

Figure 6.3 also shows that the peakedness of the density of GED decreases along with the value of kurtosis/error-power (β), i.e. the smaller the value of β, the sharper the density function of GED is. And the derivative of GED probability density function (PDF) at zero error varies with the value of kurtosis/error-power (β) (Figure 6.3):

$$\left.\frac{\partial \text{GED}(x)}{\partial x}\right|_{x=0} \begin{cases} =0 & \beta>1 & \text{smooth} \\ \neq 0 & \beta \leq 1 & \text{non-smooth} \end{cases} \tag{6.4:2}$$

where

GED(x) is the PDF of GED (Eq. (4.3:13)) evaluated at *x*.

$\dfrac{\partial \mathrm{GED}(x)}{\partial x}$ is the derivative of the PDF of GED

β is the kurtosis coefficient of GED and also termed the power of errors.

In Section 6.2.1, we concluded that the reason why the *NSE* always neglects the low flow is that the small error around zero approximates equiprobability because of the smoothness of Gaussian density function at zero error ($\beta = 2$; Figure 6.2 and Figure 6.3). Therefore, the distance-based objective function can be separated into two classes with the boundary of the kurtosis/error-power (β) of one ($\beta = 1$). The objective function with the error-power greater than one ($\beta > 1$; e.g. *NSE*) neglects the low flow (Figure 5.2). By contrast, the objective function with the error-power less than or equal to one ($\beta \leq 1$; such as BC-GED and MAE) puts greater emphasis on low values (Figure 5.2 and Figure 5.19 for WB-VSA). Note that the VE (Section 4.3.3) is equivalent to MAE, which is the normalization of MAE.

In summary, similar to the likelihood function, the distance-based objective function (Table 2.1 and Table 6.1) also imply the statistical assumptions. The effects of the objective function on runoff components result from the characteristics of error distribution implied in the objective function. The larger the value of the kurtosis/error-power (β), the greater emphasis on high values (e.g. flood) is put by the distance-based objective function. The objective function with the error-power less than or equal to one ($\beta \leq 1$) puts greater emphasis on low values (e.g. baseflow). Otherwise, it neglects the low values. Because the kurtosis/error-power (β) of MAE/VE is one ($\beta=1$), i.e. the boundary of smooth/non-smooth (Eq. (6.4:2)), MAE/VE can best balance consideration of the high- and low- values (Table 6.1; Figure 5.18 and Figure 5.19 for WB-VSA).

As an example, in this study, we have separately calibrated the SWAT-WB-VSA model using three objective functions: *NSE*, VE and BC-GED. Their performances of simulation results are shown in Table 6.2. Values of kurtosis/error-power (β) of *NSE*, VE and BC-GED are 2, 1 and 0.67, respectively. Because of the value of error power (β) of *NSE* > VE > BC-GED, i.e. $2 > 1 > 0.67$, *NSE* makes SWAT-WB-VSA capture the flood better than VE, and better than BC-GED (Figure 5.1 and Figure 5.18 for WB-VSA). Because of the value of error power (β) of *NSE* > 1, SWAT-WB-VSA with *NSE* mimics baseflow badly (Figure 5.2). By contrast, because of the value of error power (β) of VE/BC-GED ≤ 1, SWAT-WB-VSA with VE/BC-GED mimics baseflow well (Figure 5.2 and Figure 5.19 for WB-VSA).

Table 6.2 Comparison of the model performance of SWAT-WB-VSA
among *NSE*, VE and BC-GED approaches.

Objective function [1]	β [2]	Error distribution	Flood	Baseflow
NSE	2.0	Gaussian	Very good	Poor
VE	1.0	Laplace	Good	Very good
BC-GED	0.67	GED	Fair	Very good

[1] *NSE* and VE are equivalent to MSE and MAE, respectively (Table 6.1);
BC-GED is equivalent to BC-GED (objective) function (Eq. (6.4:1)).

[2] β is the power of model residuals.

The BC transformation method (Eq. (4.3:9)) is another method to balance consideration of the flood and baseflow (Pushpalatha et al. 2012; Table 6.1). Actually, the BC transformation method redefines the observed and simulated data to reduce/adjust the differences between the flood and baseflow (i.e. heteroscedasticity). As analyzed in Section 6.2.1, the big differences between the values of flood and baseflow result in the different relative-error of simulation results. After BC transformation (Figure 6.1), the model residuals are closely homoscedastic/I.I.D., so the effect of the kurtosis coefficient (β) of GED on model calibration is significantly weakened (results are shown in Appendix E).

In summary, the BC-GED (objective) function can unify currently used distance-based objective functions. By modifying the parameters (i.e. BC transformation parameter λ and GED kurtosis coefficient β) of BC-GED error model, we can obtain different forms of distance-based objective functions to emphasize on distinct hydrographic components, as shown in Appendix E. Among these objective functions, the MAE is the best to balance consideration of the flood and baseflow; and the BC-GED is the best to guarantee that the model residuals fulfill the statistical assumptions of error model (Figure 5.3 (b) and Figure 5.4 (b)), which may infer the most reasonable model-parameters (Bates and Campbell, 2001; Yang et al., 2007b; Schoups and Vrugt, 2010; Evin et al., 2013).

6.4.3   Advantage and Disadvantage of different Objective Functions

Comparison of the advantage and dis-advantage of *NSE*, BC-GED and VE are shown as in Table 6.3, based on the model calibration results in this study.

Table 6.3 Comparison of the advantage and disadvantage of *NSE*, BC-GED and VE.

Objective	Advantage	Disadvantage
NSE/MSE	Most widely used efficiency criterion; put greater emphasis on high values.	Informal likelihood method; infer some unreasonable results.
BC-GED	Formal likelihood method; unify currently used distance-based objective functions.	Need to calibrate extra parameters for error model (e.g. BC and GED).
VE/MAE	Best balance consideration of the high and low values.	May infer unreasonable values of model parameters.

NSE. *NSE* is the most widely used efficiency criterion (objective function) in hydrological model calibration (Beven and Binley, 2013), of which the advantage is that it puts greater emphasis on high values (e.g. flood) and can make hydrological model capture the high values (e.g. floods and large sediment loads) well (Figure 5.1 and Figure 5.7). However, the *NSE* not only neglects the low values (e.g. baseflow; Figure 5.2), but also infers some fallacious/unreasonable results:

1. Some parameters estimated by the *NSE* approach are unreasonable. For example, the optimized soil hydraulic conductivity (SOL_K) of Regosols and Luvisols accounting for more than 80% of watershed area is less than 30% of field survey values (Table 5.1). In other words, the optimized SOL_K approximates $30\% \times 2 \times 10^{-6}$ m/s ($\approx$ 0.05 m/day; Table 3.1). However the optimized groundwater delay time (GW_DELAY that is the delay time of recharge from the soil layer to shallow groundwater) is the shortest (i.e. 0.02; Table 5.1), which is impossible in actual soil profile with 0.05 m/day hydraulic conductivity. Another example is that in Figure 5.12 for the *NSE* approach, the values of the soil erodibility factor of Regosols (USLE_K1) are unreasonably much less than that of Luvisols (USLE_K2) and Fluvisols (USLE_K3). In fact, the Regosols is usually more prone to erosion owing to the low coherence of the matrix material (Driessen et al., 2001).

2. Runoff components concluded by *NSE* approach are absurd. In Table 5.3 for *NSE* approach, the amount of groundwater return flow is very large, which accounts for about 80% of total runoff, but the amount of interflow is unreasonably very small, model predictions/simulations in Figure 5.2 cannot mimic the baseflow, and the model validation results using groundwater levels in Figure 5.6 are worst. These demonstrate

that the main component of flood is the groundwater with rapid recession, rather than the surface runoff or interflow.

3. The source of sediment loads concluded by *NSE* approach is inconsistent with the fact. The multi-response *NSE* approach concluded that considerable sediment-loads are from the main channel erosion (Table 5.3). Nonetheless, the field survey shows there are some small dams across the main channels built by local residents for irrigation and transportation (Figure 3.7), and these dams intercept a lot of suspended sediment.

These fallacious results may result from the informal likelihood method of *NSE* (Clarke, 1973; Xu, 2001; Jain and Sudheer, 2008). *NSE* assumes the model residuals (i.e. observed data minus model predictions) follow the Gaussian distribution (Table 4.4). However, model residuals of both river flow and sediment concluded by *NSE* approach are obviously heteroscedastic (Figure 5.3 (a) and Figure 5.8), non-Gaussian (Figure 5.4 (a) and Figure 5.10) and dependent (Figure 5.9), which completely violate the statistical assumptions of *NSE* approach (Sections 4.3.1.1 and 4.3.2.2).

BC-GED. The BC-GED approach firstly uses the Box-Cox transformation (BC) method to remove the heteroscedasticity of model residuals, and then adopts the generalized error distribution (GED) to estimate the likelihood of the model residuals after BC, which not only guarantees that the model residuals fulfill its statistical assumptions, such as homoscedasticity (Figure 5.3 (b) and Figure 5.14), the error distribution of GED (Figure 5.4 (b) and Figure 5.16) and independence between the flow and sediment model residuals (Figure 5.15), but also concluded more reasonable results than *NSE* approach in this study.

1. The SWAT-WB-VSA with the BC-GED approach mimics baseflow well (Figure 5.2). And the multi-response BC-GED approach concluded the non-erodibility of main channel only based upon the river sediment load data (Table 5.3).

2. The model parameters estimated by BC-GED approach agree with the field-investigated results and previous studies better than that estimated by *NSE* approach, such as soil properties (especially soil hydraulic conductivity), storage function of main channel and erodibility of soil and main channel (Section 6.3.1 and 6.3.2). The results inferred by BC-GED approach match the characteristics of river flow and sediment in Baocun watershed: the storage function of main channel is small on the daily time step (Figure 3.1; Section 3.1); the soil erodibility of Regosols is largest (Driessen et al., 2001); and a lot of suspended sediment deposit in the main channel due to the small dams across channels (Figure 3.7; Section 3.5.2).

The BC-GED function (equivalent to BC-GED likelihood function; Eq. (6.4:1)) also can unify currently used distance-based objective functions (Section 6.4.1 and Table 6.1). However, BC-GED approach needs to calibrate extra parameters for error model (such as the BC parameter (lambda/λ) and kurtosis coefficient (beta/β) of GED), which may cause some troubles during model calibration, e.g. falling into the local optimum because of over-parameterization. Therefore, the BC-GED approach lacks robustness for comparison and assessment of the different models including the empirical (e.g. CN) and physically-basis (e.g. WB) structures. On the other hand, the BC-GED error model filters some errors (probably including model structural errors; Figure 6.2), which results in that the objective function (BC-GED) is not sensitive enough to the structure of hydrological model. Therefore, the BC-GED approach is difficult to reflect the structural differences among different hydrological models.

VE. VE is a normalization of the mean absolute error (MAE; Table 6.1), which is equivalent to MAE. The MAE/VE can best balance consideration of the high and low values (Figure 5.18 and Figure 5.19), as detailed in Section 6.4.2. Similar to the *NSE*, the MAE/VE does not have extra parameters for error model. Therefore, the MAE/VE may be more practical than BC-GED, if we only focus on the performance of the hydrological model, e.g. mimic of the river discharges. In this study, the calibration results of VE are similar to those of BC-GED approach, and the main difference being the available water capacity of soil layer (SOL_AWC). The value of SOL_AWC estimated by VE in Table 5.5 for WB-VSA approach is obviously greater than that estimated by BC-GED in Table 5.1. As a result, the amount of interflow (Table 5.6 for WB-VSA) is greater than that of BC-GED (Table 5.3), but the amount of overland flow (surface runoff) is less. By contrast, the value of soil porosity (which is inversely proportional to the bulk density (SOL_BD); Table 5.5 for WB-VSA) is less than that estimated by BC-GED (Table 5.1). Therefore, the calibration results of VE may overestimate the available water capacity of soil layer (Neitsch et al., 2005).

Figure 5.19 shows that all the five rainfall-runoff methods (CN-Soil, CN-ET, G&A, WB and WB-VSA) after calibration using the volumetric efficiency (VE) can mimic the baseflow well, which demonstrates that no matter what the model structure is, VE always balances consideration of flood and baseflow. Therefore, VE is an effective and robust objective function for model comparison.

6.5   Performances of different Rainfall-Runoff Approaches

Table 4.1 shows that the five rainfall-runoff methods (CN-Soil, CN-ET, G&A, WB and WB-VSA) are based on different surface-runoff generation mechanisms: WB and WB-VSA approaches are based on the principle of saturation excess overland flow, but G&A is based on the principle of the infiltration excess overland flow. In Table 5.5, the model performance indicators (VE, NSE and R^2) of water balance methods (WB-VSA and WB) are much better than those of G&A, which may reveal that the saturation excess overland flow method is more appropriate for the Baocun watershed than the infiltration excess overland flow method on the daily time step. That may attribute to the terraced fields (Figure 3.4) in the whole Baocun watershed, because the terraced fields can delay the overland flow, facilitate the rainfall infiltration and ultimately result in no surface runoff generation until the soil layer is saturated.

Because both the NSE and the R^2 put emphasis on the high flow (Krause et al. 2005), actually it is that the flood simulation results of the water balance methods (WB-VSA and WB) are much better than those of G&A method (Table 5.6 and Figure 5.18). In G&A, the very large value of soil hydraulic conductivity (SOL_K) estimated by the automatic optimization program (Table 5.5) results in the very large soil infiltration capacity and no infiltration excess overland flow occurring (Table 5.6; Kannan et al., 2007).

Because the CN-Soil implies the principle of saturation excess overland flow (Table 4.1; Steenhuis et al. 1995), it gets the similar model performance (i.e. values of VE, NSE and R^2) on river discharges with the water balance methods (WB-VSA and WB) in the Baocun watershed (Table 5.5). However the CN-Soil is an empirical method that assumes two empirical relationships: one between the retention (rainfall not converted into runoff) and runoff properties of the watershed and the rainfall (Eq. (4.1:2)), and the other between the retention parameter and the soil water content (Eq. (4.1:4)), so it mimics the groundwater levels worst (Figure 5.20). The limitation is also pointed out by Han et al. (2012): the CN-Soil cannot well reflect that soil moisture affects the surface runoff generation. Surprisingly, the groundwater simulation results of the CN-ET are much better than those of the CN-Soil (Figure 5.20), although the two approaches are all based on the curve number method (CN). It may result from two reasons: First, the CN-ET simulates the surface flow and the subsurface flow separately, because the retention parameter of CN-ET is estimated by the potential evaporation, rather than the soil water content (Eq. (4.1:5)); and Second, the automatic optimization program with the VE makes the CN-ET approach emphasize the

performance of baseflow simulation. These reasons also result in that the CN-ET simulates the flood worse than the CN-Soil (Figure 5.18).

The TOPMODEL under the steady-state condition and exponential decline of transmissivity with depth assumptions has been widely accepted by hydrologists (Beven, 1997). Its key contribution is building a linear relationship between the soil moisture deficit and the topographic index (Eq. (4.1:11); Beven, 1997). In WB, White et al. (2011) introduced a parameter (effective soil depth (EDC)) to reflect the effect of topography on soil moisture deficit. The WB-VSA further improved WB by transplanting the linear expression of soil moisture deficit in the TOPMODEL to the EDC in WB. However, in the Baocun watershed, because the terraced fields change the micro-topography and weaken the effects of topography on soil saturation deficit, the effective soil depth (EDC) is close to the original soil depth (i.e. 1.0) in both WB and WB-VSA (Table 5.5 and Eq. (4.1:9)). As a consequence, the river discharge simulation results of the WB-VSA only improved a little comparing with the WB (Table 5.5). But some others still confirmed the theory of the WB-VSA: First, the EDC as the catchment average EDC_i in the WB-VSA (Eq. (4.1:12)) approximates to the constant EDC of all HRUs assumed by the WB (Table 5.5). In other words, the WB is only a special case of the WB-VSA in the Baocun watershed because of small effects of topography on soil saturation deficit. Second the groundwater simulation result of the WB-VSA is obviously better than the WB because of the stronger linear relationship between the observed groundwater level and the simulated soil water volume (Figure 5.20).

In summary, the saturation excess overland flow approaches (WB-VSA, WB and CN-Soil) capture the floods much better than the infiltration excess overland flow approach (G&A) and the evaporation-dependent curve number approach (CN-ET). However, the CN-Soil as an empirical method simulates the groundwater levels worst. By contrast, the WB-VSA as a physics-based method developed by this study simulates the groundwater levels best.

7 Conclusions

By strict derivation, the Nash–Sutcliffe efficiency coefficient (*NSE*) is proved to be equivalent to a kind of likelihood function with the assumption that the model residuals follow a Gaussian distribution with zero mean. However, these assumptions cannot be satisfied totally, as is confirmed by the application of the *NSE* approach in the Baocun watershed, so the *NSE* is an informal likelihood function. Comparison of the observed and simulated river discharges indicates that SWAT-WB-VSA model with the *NSE* approach can simulate flood well, but baseflow badly owing to the assumption of Gaussian error distribution, where the probability of the large error is low, but the error around zero approximates equiprobability. The Gaussian error assumption also results in that the multi-response *NSE* approach with the river flow and sediment objectives neglects the zero-values of sediment load. As a consequence, the *NSE* concludes an unrealistic result that more than 36% of sediment loads in the watershed outlet are from the main channel erosion.

After review of the Box-Cox transformation (BC) method in the hydrologic literature, we found that most authors fixed the value of BC parameter (lambda), and others all inferred unreasonable result of lambda. Therefore, we concluded that the BC method is an implicit method. Then we proposed a new estimation method of BC parameter to improve the effective elimination of the heteroscedasticity of model residuals. The practical applications of this study proved that the scheme of BC with minimum variance constraint is an effective method to estimate the BC parameter (lambda).

The BC-GED and BC-SGED likelihood functions proposed by this study assume the model residuals after BC follow the generalized error distribution (GED) and skew generalized error distribution (SGED), respectively. In this study, BC-GED and BC-SGED are the formal likelihood functions, as model residuals after BC are homoscedasticity and fitted by GED/SGED well. The SWAT-WB-VSA with the BC-GED/BC-SGED approach mimics baseflow well, which is proved in the groundwater level simulation. The multi-response BC-GED approach also successfully inferred the non-erodibility of main channel only based on the sediment-load data in the watershed outlet, which estimated that nearly 18% sediments eroded from hillslopes/HRUs deposit in the main channel. All the results of the BC-SGED are nearly the same to those of BC-GED, because the skewness coefficient of BC-SGED is close to one, i.e. no-skewness, and BC-SGED degenerates into BC-GED. Therefore, the assumption of skewness of the error model (i.e. likelihood function) may be unnecessary.

According to the field survey and previous related studies, the model parameters estimated by the formal likelihood function (BC-GED) are more reasonable than those estimated by the informal likelihood function (*NSE*), such as the properties of soil profile, the storage function of main channel, and the erodibility of soil and main channel. However, because the BC-GED filters some model-structural errors, it is not sensitive enough to the structure of hydrological model and inappropriate to be treated as the objective function for comparison of different hydrological models.

By extending the derivation method of *NSE* likelihood function to the BC-GED error model, this study found currently used distance-based objective functions can be unified by the BC-GED (objective) function, which demonstrates the distance-based objective functions imply the statistic assumptions. For example, the mean squared error (MSE) and the mean absolute error (MAE) imply statistical assumptions that the model residuals follow the Gaussian and Laplace distribution with zero-mean, respectively. BC-GED also reveals the MAE (i.e. the kurtosis/error-power of one) can best balance consideration of the high and low values (e.g. the flood and baseflow). In this study, the volumetric efficiency (VE, i.e. a normalization of MAE) always balances consideration of the flood and baseflow, no matter what the model structure is. Therefore, VE is an effective and robust objective function for model comparison.

The calibration results of the five rainfall-runoff methods in SWAT by using VE demonstrate that all methods mimic the baseflow well, but the saturation excess overland flow approaches (WB-VSA, WB and CN-Soil) capture the floods much better than the infiltration excess overland flow approach (G&A) and the evaporation-dependent curve number approach (CN-ET). However, the soil-water-dependent curve number method (CN-Soil) as an empirical method mimics the groundwater levels worst. By contrast, the Water Balance model with Variable Source Area (WB-VSA) approach as a physics-based method mimics the groundwater levels best. Therefore, WB-VSA developed by this study is the most right owing to reflection of the spatial variation of runoff generation affected by topography and soil properties.

8 References

Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. Journal of Hydrology 333, 413– 430.

Andréassian, V., Le Moine, N., Perrin, C., Ramos, M.H., Oudin, L., Mathevet, T., Lerat, J., Berthet, L., 2012. All that glitters is not gold: the case of calibrating hydrological models. Hydrological Processes 26, 2206 – 2210.

Andrieu, C., de Freitas, N., Doucet, A., Jordan, M., 2003. An introduction to MCMC for machine learning. Machine Learning 50(1-2), 5–43.

Arnold, J.G., Kiniry, J.R., Srinivasan, R., Williams, J.R., Haney, E.B., Neitsch, S.L., 2011. Soil and Water Assessment Tool Input/Output File Documentation Version 2009. Texas Water Resources Institute Technical Report No. 365. PP: 94.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: model use, calibration and validation. Trans. ASABE 55(4), 1491-1508.

Bagnold, R.A., 1977. Bedload transport in natural rivers. Water Resources Research 13, 303-312.

Baratti, R., Cannas, B., Fanni, A., Pintus, M., Sechi, G.M., Toreno, N., 2003. River flow forecast for reservoir management for neural networks. Neurocomputing 55, 421-437.

Bates, B.C., Campbell, E.P., 2001. A Markov Chain Monte Carlo Scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. Water Resources Research 37(4), 937–947.

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

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environmental Modelling & Software 40, 1-20.

Betrie, G.D., Mohamed, Y.A., van Griensven, A., Srinivasan, R., 2011. Sediment management modelling in the Blue Nile Basin using SWAT model. Hydrology and Earth System Sciences 15, 807–818.

Beven, K., Kirkby, M., 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 303-325.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological processes 6: 279–298.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16, 41–51.

Beven, K., 1997. TOPMODEL: a critique. Hydrological Processes 11(9), 1069–1085.

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrology and Earth System Sciences 5 (1), 1–12.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in echanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249 (1–4), 11–29.

Beven, K., Smith, P., Westerberg, I., Freer, J., 2012. Comment on: "Pursuing the method of multiple working hypotheses for hydrological modeling" by P. Clark et al. Water Resources Research 48 (11), W11801.

Beven, K., Binley, A., 2013. GLUE: 20 years on. Hydrological processes. Doi: 10.1002/hyp.10082.

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., Montanari, A., 2012. Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. Physics and Chemistry of the Earth 42–44, 70–76.

Chen, X., Cheng, Q.-B., Chen, Y.D., Smettem, K., Xu, C.-Y., 2010. Simulating the integrated effects of topography and soil properties on runoff generation in hilly forested catchments, South China. Hydrological Processes 24, 714–725.

Chen, E., Mackay, D.S., 2004. Effects of distribution-based parameter aggregation on a spatially distributed agricultural nonpoint source pollution model. Journal of Hydrology 295, 211–224.

Cheng, Q.-B., Chen, X., Chen, X., Zhang, A., Ling, M., 2011. Water infiltration underneath single-ring permeameters and hydraulic conductivity determination. Journal of Hydrology 398, 135–143.

Clarke, R.T., 1973. A review of some mathematical models used in hydrology, with observations on their calibration and use. Journal of Hydrology 19, 1-20.

Coello Coello, C.A., 2006. Evolutionary multi-objective optimization: A history view of the field. IEEE Computational Intelligence Magazine 28–36.

Criss, R.E., Winston, W.E., 2008. Do Nash values have value? Discussion and alternate proposals. Hydrological Processes 22, 2723–2725.

Daniel, E.B., Camp, J.V., LeBoeuf, E.J., Penrod J.R., Dobbins, J.P., Abkowitz M.D., 2011. Watershed Modeling and its Applications: A State-of-the-Art Review. The Open Hydrology Journal 5, 26-50.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modelling & Software 22, 1034-1052.

De Rose R.C., 1996. Relationships between slope morphology, regolith depth, and the incidence of shallow landslides in eastern Taranaki hill country. Zeitschrift für Geomorphologie Supplementband 105, 49–60.

Duan, Q.-Y, Sorooshian S., Gupta V.K., 1992. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. Water Resources Research 28(4), 1015-1031.

Duan, Q.-Y., 2003. Global Optimization for Watershed Model Calibration. Calibration of watershed models, Q.-Y., Duan, et al. (Eds.). American Geophysical Union: Washington, DC; 89–104.

Dunne, T., 1978. Field studies of hillslope flow processes. In Hillslope Hydrology, Kirkby, M.J. (ed.). Wiley: Chichester.

Driessen, P., Deckers, J., Spaargaren, O., Nachtergaele, F., 2001. Lecture notes on the major soils of the world. Report of the World Soil Resources. Food and Agriculture Organization, Rome.

Easton, Z.M., Walter, M.T., Fuka, D.R., White, E.D., Steenhuis, T.S., 2011. A simple concept for calibrating runoff thresholds in quasi-distributed variable source area watershed models. Hydrological Processes 25, 3131-3143.

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

Engeland, K., Renard, B., Steinsland, I., Kolberg, S., 2010. Evaluation of statistical models for forecast errors from the HBV model. Journal of Hydrology 384 (1–2), 142–155.

Erskine, W.D., Mahmoudzadeh, A., Myers, C., 2002. Land use effects on sediment yields and soil loss rates in small basins of Triassic sandstone near Sydney, NSW, Australia. Catena 49, 271 – 287.

Evin, G., Kavetski, D., Thyer, M., Kuczera, G., 2013. Pitfalls and improvements in the joint inference of heteroscedasticity and autocorrelation in hydrological model calibration. Water Resources Research 49, 1-7.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resources Research 50, 2350–2375.

FAO/IIASA/ISRIC/ISSCAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Feyen, L., Vrugt, J.A., Nuallain, B.O., van der Knijff, J., Roo, A.D., 2007. Parameter optimization and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model. Journal of Hydrology 332, 276–289.

Fetter C.W. (1994), Applied Hydrogeology (3rd ed.), Upper Saddle River, NJ: Prentice Hall, Inc.

Gabellani, S., Silvestro, F., Rudari, R., Boni, G., 2008. General calibration methodology for a combined Horton-SCS infiltration scheme in flash flood modeling. Natural Hazards and Earth System Sciences 8, 1317–1327.

Garen, D.C., Moore, D.S., 2005. Curve Number Hydrology in Water Quality Modeling: Uses, Abuses and Future Directions. Journal of the American Water Resources Association 41, 377–388.

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. American Society of Agricultural and Biological Engineers 50(4), 1211–1250.

Gebremicael, T.G., Mohamed, Y.A., Betrie, G.D., van der Zaag, P., Teferi, E., 2013. Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and landuse maps. Journal of Hydrology 482 (4), 57-68.

Gelman, A., Varlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. 2nd edition. Chapman & Hall/CRC, Boca Raton.

Godt, J.W., Baum, R.L., Savage, W.Z., Salciarini, D., Schulz, W.H., Harp, E.L., 2008. Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework. Engineering Geology 102(3–4), 214–226.

Goodman, D., 2002. Extrapolation in Risk Assessment: Improving the Quantification of Uncertainty, and Improving Information to Reduce the Uncertainty. Human and Ecological Risk Assessment: An International Journal 8(1), 177-192.

Green, I.R.A., Stephenson, D., 1986. Criteria for comparison of single event models. Hydrological Sciences Journal 31(3), 395-411.

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

Guo, J., 2002. Logarithmic matching and its applications in computational hydraulics and sediment transport. Journal of Hydraulic research 40(5), 555-565.

Gupta, V.K., Sorooshian, S., Yapo, P.O., 1998. Towards improved calibration of hydrological models: multiple and noncomensurable measures of information. Water Resources Research 34, 751–763.

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

Güngör, Ö, Göncü, S., 2013. Application of the soil and water assessment tool model on the Lower Porsuk Stream Watershed. Hydrological Processes 27: 453–466

Han, E., Merwade, V., Heathman, G.C., 2012. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. Journal of Hydrology 416, 98–117.

Heimsath, A .M., Dietrich, W.E., Nishiizumi, K., Finkel, R.C., 1997. The soil production function and landscape equilibrium. Nature 388, 358– 361.

Hosseini, M., Amin, M.S.M., Ghafouri, A.M., Tabatabaei, M.R., 2011. Application of Soil and Water Assessment Tools Model for Runoff Estimation. American Journal of Applied Sciences 8(5): 486-494.

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

Johnson, D.O., Arriaga, F.J., Lowery, B., 2005. Automation of a falling head permeameter for rapid determination of hydraulic conductivity of multiple samples. Soil Science Society of America Journal 69, 828–833.

Jury, W.A., Gardner, W.R., Gardner, W.H., 1991. Soil physics. 5thed. John Wiley & Sons, New York.

Kannan, N., White, S.M., Worrall, F., Whelan, M.J., 2007. Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modeling in SWAT2000. Journal of Hydrology 332(34), 456–466.

Khakbaz, B., Imam, B., Hsu, K., Sorooshian, S., 2012. From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. Journal of Hydrology 418–419, 61–77.

King, K.W., Arnold, J.G., Bingner, R.L., 1999. Comparison of Green – Ampt and curve number methods on Goodwin Creek watershed using SWAT. American Society of Agricultural Engineers 42(4), 919–925.

Kollat, J. B., Reed, P. M., Wagener, T., 2012. When are multiobjective calibration trade-offs in hydrologic models meaningful? Water Resources Research 48, W03520.

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

Laloy, E., Fasbender, D., Bielders, C.L., 2010. Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach. Journal of Hydrology 380(1–2), 82–93.

Lee, K.T., Ho, J.-Y., 2009. Prediction of landslide occurrence based on slope-instability analysis and hydrological model simulation. Journal of Hydrology 375, 489–497.

Legates, D.R., McCabe, Jr. G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Researches 35(1), 233-241.

Li, L., Xu, C.-Y., Xia, J., Engeland, K., Reggiani, P., 2011. Uncertainty estimates by Bayesian method with likelihood of AR (1) & Normal model and AR (1) & Multi-normal model in different time-scales hydrological models. Journal of Hydrology 406, 54–65.

Li, L., Xu, C.-Y., Engeland, K., 2013. Development and comparison in uncertainty assessment based Bayesian modularization method in hydrological modeling. Journal of Hydrology 486, 384–394.

Loague, K, Heppner, C.S., Ebel, B.A., VanderKwaak, J.E., 2010. The quixotic search for a comprehensive understanding of hydrologic response at the surface: Horton, Dunne, Dunton and the role of concept-development simulation. Hydrological Processes 24, 2499–2505.

Lu, S.-L., Kayasthab, N., Thodsena, H., van Griensvenb, A., Andersena, H.E., 2011. Multiobjective calibration for comparing channel sediment routing models in the soil and water assessment tool. J. Environ. Qual. 43(1), 110-120.

Lyon, S.W., Walter, M.T., Gerard-Marchant, P., Steenhuis, T.S., 2004. Using a topographic index to distribute variable source area runoff predicted with the SCS curve-number equation. Hydrological Processes 18, 2757–2771.

Madsen, H., 2000. Automatic calibration of a conceptual rain-fall–runoff model using multiple objectives. Journal of Hydrology 235, 276–288.

Marshall, L., Nott, D., Sharma, A., 2005. Hydrological model selection: A Bayesian alternative. Water Resources Research 41, W10422.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash-Sutcliffe efficiency index. Journal of hydrologic engineering 11, 597-602.

McMillan, H., Clark, M., 2009. Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. Water Resources Research 45, W04418.

Ministry of Water Resources of the People's Republic of China, 2008. Standards of classification and gradation of soil erosion (SL 190-2007). China Water Power Press, Beijing.

Mishra, S.K., Singh, V.P., 2003. Soil Conservation Service Curve Number (SCS-CN) Methodology. Kluwer Academic Publishers: Dordrecht, Netherlands.

Montanari, A., Koutsoyiannis D., 2012. A blueprint for process-based modeling of uncertain hydrological systems. Water Resources Research 48, W09555.

Moradkhani, H., Sorooshian, S., 2009. General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis. Hydrological Modelling and the Water Cycle, S. Sorooshian et al. (Eds.). Springer Science+Business Media B.V; 1–24

Moriasi, D.N., Arnold, J.G., van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900.

Mukundan, R., Pradhanang, S.M., Schneiderman, E.M., Pierson, D.C., Anandhi, A., Zion, M.S., Matonse, A.H., Lounsbury, D.G., Steenhuis, T.S., 2013. Suspended sediment source areas and future climate impact on soil erosion and sediment yield in a New York City water supply watershed, USA. Geomorphology 183, 110–119.

Muleta, M.K., Nicklow J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of Hydrology 306, 127–145.

Nash, J.E., Sutcliffe, J.V., 1970. River Flow Forecasting through Conceptual Models 1: A Discussion of Principles. Journal of Hydrology 10(3), 282–290.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2005. Soil and Water Assessment Tool Theoretical Documentation and User's Manual, Version 2005. GSWR Agricultural Research Service & Texas Agricultural Experiment Station, Temple Texas.

Norton, J.B., Sandor, J.A., White, C.S., 2003. Hillslope soils and soil organic matter dynamics within a native American agroecosystem on the Colorado plateau. Soil Science Society of America 67, 225-234.

Nott, D.J., Marshall, L., Brown, J., 2012. Generalized Likelihood Uncertainty Estimation (GLUE) and approximate Bayesian computation: what's the connection? Water Resources Research 48, W12602.

Oeurng, C., Sauvage, S., Sánchez-Pérez, J.-M., 2011. Assessment of hydrology, sediment and particulate organic carbon yield in a large agricultural catchment using the SWAT model. Journal of Hydrology 401, 145–153.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., 2006. Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations. Water Resources Research 42(7), W07410.

Pande, S. 2013. Quantile hydrologic model selection and model structure deficiency assessment 2: Applications. Water Resources Research, 49, 5658–5673.

Parajka, J., Merz, R., Blöschl, G., 2007. Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments. Hydrological processes 21(4), 435–446.

Patel, K.P., 2009. Watershed modeling using HEC-RAS, HEC-HMS, and GIS models: a case study of the Wreck Pond Brook Watershed in Monmouth County, New Jersey. Rutgers University.

Pianosi, F., Raso, L., 2012. Dynamic modeling of predictive uncertainty by regression on absolute errors. Water Resources Research 48, W03516.

Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., Kurz, D., McNair, M., Crist, S., Shpritz, L., Fitton, L., Saffouri, R. & Blair, R., 1995. Environmental and economic costs of soil erosion and conservation benefits. Science 267, 1117–23.

Prosser, L.P., 2000. Sediment transport capacity relations for overland flow. Progress in Physical Geography 24(2), 179–193.

Pushpalatha, R., Perrin, C., Moine, Le N., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology 420–421, 171–182.

Qiao, L., Herrmann, R. B., Pan, Z.-T., 2012. Parameter uncertainty reduction for SWAT using grace, streamflow, and groundwater table data for lower Missiouri river basin. Journal of the American water resources association 49(2), 343-358.

Rallison, R.E., Miller, N., 1982. Past, present, and future SCS runoff procedure. In Rainfall–Runoff Relationship, Singh VP (ed.). Water Resources Publications: Littleton, CO; 353–364.

Reed, P. M., Hadka, D., Herman, J., Kasprzyk, J., Kollat, J., 2013. Evolutionary multiobjective optimization in water resources: The past, present, and future. Advances in Water Resources 51, 438-456.

Reichert, P., Mieleitner, J. 2009. Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters. Water Resources Research 45, W10402.

Reichert, P., and Schuwirth, N., 2012. Linking statistical bias description to multiobjective model calibration. Water Resources Research 48, W09543.

Ritter, A, Muñoz-Carpena, R., 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. Journal of Hydrology 480, 33–45.

Rostamian, R., Jaleh, A., Afyuni, M., Mousavi, S.F., Heidarpour, M., Jalalian, A., Abbaspour, K.C., 2008. Application of a SWAT model for estimating runoff and sediment in two mountainous basins in central Iran. Hydrological Sciences Journal 53(5), 977-988.

Sadegh, M., Vrugt, J. A., 2013. Bridging the gap between GLUE and formal statistical approaches: approximate Bayesian computation. Hydrology and Earth System Sciences 17, 4831-4850.

Salciarini, D., Godt, J.W., Savage, W.Z., Conversini, R., Baum, R.L., Michael, J.A., 2006. Modelling regional initiation of rainfall-induced shallow landslides in the eastern Umbria Region of central Italy. Land-slides 3, 181–194.

Saulnier, G.M., Beven, K., Obled, C., 1997. Including spatially variable effective soil depths in TOPMODEL. Journal of Hydrology 202, 158–172.

Savenije, H.H.G., 2009. HESS Opinions: "The art of hydrology". Hydrology and Earth System Sciences 13, 157–161.

Schaake, J.C., 2003. Global Introduction. Calibration of watershed models, Q.-Y., Duan, et al. (Eds.). American Geophysical Union: Washington, DC; 1–7.

8 References

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

Schneiderman, E.M., Steenhuis, T.S., Thongs, D.J., Easton, Z.M., Zion, M.S., Neal, A.L., Mendoza, G.F., Walter, M.T., 2007. Incorporating variable source area hydrology into a curve-number-based watershed model. Hydrological Processes 21, 3420–3430.

Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research 46, W10531.

SCS, 1972. SCS National Engineering Handbook, Section 4. Hydrology, Soil Conservation Service, US Department of Agriculture: Washington, DC.

Seibert, J., 2001. On the need for benchmarks in hydrological modeling. Hydrological Processes 15, 1063–1064.

Sikorska, A.E., Scheidegger, A., Banasik, K., Rieckermann, J., 2012. Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models. Hydrology & Earth System Sciences 16, 1221–1236.

Smith, M.B., Seo, D., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. Journal of Hydrology 298, 4–26.

Smith, T., Sharma, A., Marshall, L., Mehrotra, R., Sisson, S., 2010. Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments. Water Resources Research 46, W12551.

Sorooshian, S., Gupta, V.K., 1995. Model calibration. In: Singh, V. P. (Ed.), Computer Models of Watershed Hydrology, Water Resources Publications, Highlands Ranch, CO, pp. 23–67.

Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44, W00B06.

Steenhuis, T., Winchell, M., Rossing, J., Zollweg, J., Walter, M., 1995. SCS Runoff Equation Revisited for Variable-Source Runoff Areas. Journal of Irrigation and Drainage Engineering 121(3), 234–238.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Trans. American Geophysical Union 38 (6), 913–920.

Sun, H., Cornish, P.S., Daniell, T.M., 2002. Contour-based digital elevation modeling of watershed erosion and sedimentation: Erosion and sedimentation estimation tool (EROSET). Water Resources Research 38(11), 1233.

Thiemann, M., Trosset, M., Gupta, H., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. Water Resources Research 37 (10), 2521-2535.

Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research 43, W01413.

Vandewiele, G.L., Xu, C.-Y., Win, N.-L., 1992. Methodology and comparative study of monthly water balance models in Belgium, China and Burma, Journal of Hydrology 134, 315–347.

van Griensven, A., Popescu, I., Abdelhamid, M.R., Ndomba, P.M., Beevers, L., Betrie, G.D., 2013. Comparison of sediment transport computations using hydrodynamic versus hydrologic models in the Simiyu River in Tanzania. Phys. Chem. Earth 61–62, 12–21.

Vazquez-Amábile G.G., Engel B.A., 2005. Use of SWAT to compute groundwater table depth and streamflow in the Muscatatuck River watershed. Trans ASABE 48(3):991–1003.

Vrugt, J.A., Gupta H.V., Bouten W., Sorooshian S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39, 1201.

Vrugt, J.A., Ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009a. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment 23 (7), 1011-1026.

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009b. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation 10(3), 273-290.

Vrugt, J. A., Sadegh M., 2013. Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resources Research 49, 4335–4345.

Walling, D.E., 2000. Linking land use, erosion and sediment yields in river basins. Hydrobiologia 410, 223-240.

Wang, L., Song, M., Wang, P., 2002. Study development on Jiaonan - Weihai tectonic belt and discussion of some important geological problems. Geology of Shandong 18(3-4), 78–83 (in Chinese).

Wang, G.X., Liu, C.M., Huang, Y.B., 2003. The theory of SWAT model and its application in Heihe Basin. Process Geography 22: 79–86.

White, E.D., Easton, Z.M., Fuka, D.R., Collick, A.S., Adgo, E., McCartney, M., Awulachew, S.B., Selassie, Y.G., Steenhuis, T.S., 2011. Development and application of a physically based landscape water balance in the SWAT model. Hydrological Processes 25, 915–925.

Willems, P., 2009. A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. Environmental Modelling & Software 24, 311–321.

Williams, J.R., 1980. SPNM, a model for predicting sediment, phosphorus, and nitrogen yields from agricultural basins. Water Resources Bulletin 16, 843-848.

Williams, J.R., 1995. Chapter 25: The EPIC model. In V.P. Singh (ed.) Computer models of watershed hydrology. Water Resources Publications, Littleton, pp. 909-1000.

Wong, T.S.W., Koh, X.C., 2008. Which model type is best for deterministic rainfall-runoff modelling? In: Robinson, L.N. (Ed.), Water Resources Research Progress. Nova Science Publishers, New York, pp. 241–260.

Wu, Y.-p, Chen, J., 2012. Modeling of soil erosion and sediment transport in the east river basin in southern china. Sci. Total Environ. 44, 159-168.

Wu, Y.-P., Liu, S.G., 2014. A suggestion for computing objective function in model calibration. Ecological Informatics 24, 107–111.

Würtz, D., et al., 2013. Rmetrics—autoregressive conditional heteroscedastic modelling, http://cran.r-project.org/web/packages/fGarch/index.html.

Xu, C.-Y., Seibert, J., Halldin, S., 1996. Regional water balance modelling in the NOPEX area: development and application of monthly water balance models, Journal of Hydrology 180, 211-236.

Xu, C.-Y., 2001. Statistical analysis of parameters and residuals of a conceptual water balance model – methodology and case study. Water Resources Management 15, 75–92.

Yang, D.-W., Kanae, S., Oki, T., Koike, T. & Musiake, K. 2003. Global potential soil erosion with reference to land use and climate changes. Hydrological processes 17, 2913 – 2928.

Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007a. Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference. Journal of Hydrology 340, 167-182.

Yang, J., Reichert, P., Abbaspour, K.C., 2007b. Bayesian uncertainty analysis in distributed hydrological modelling: a case study in the Thur River basin (Switzerland). Water Resources Research 43, W10401.

Yu, B.-F., 1998. Theortical Justification of CSC method for runoff estimation. Journal of irrigation and drainage engineering 124, 306–310.

Yustres, á, Asensio, L., Alonso, J., Navarro, V., 2012. A review of Markov Chain Monte Carlo and information theory tools for inverse problems in subsurface flow. Computational Geosciences 16, 1-20.

Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent Particle Swarm Optimisation software for model calibration. Environmental Modelling & Software 43, 5-25.

Appendices

Appendix A: Diagnosis of the Autocorrelation of Model Residuals

Figure A1 inspects the autocorrelation of model residuals for *NSE*, BC-GED and BC-SGED approaches. This figure shows that the absolute value of first-order auto-correlation coefficient is much less than one, and other order auto-correlation coefficients are close to zero for all approaches in the Baocun watershed. Therefore, the autocorrelation of model residuals is weak, which means the residuals are closely independent. So, it is reasonable to neglect the procedure for removing the errors' autocorrelation in this study.
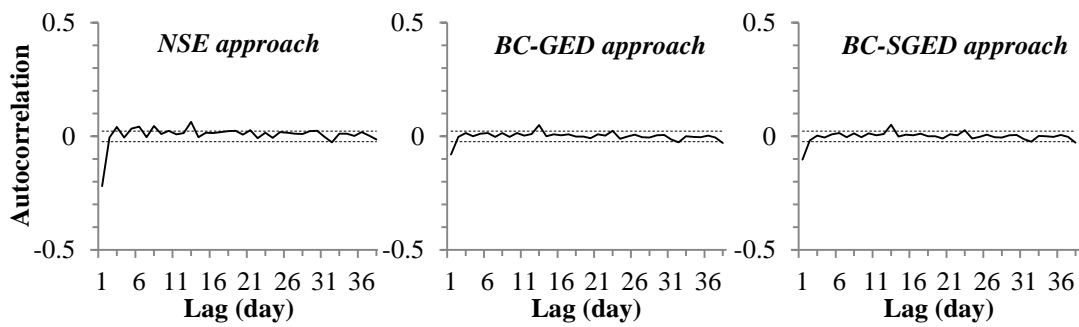
Figure A1 Diagnosis of the autocorrelation of model residuals for *NSE*, BC-GED and BC-SGED approaches.

Appendix B: Re-interpretation of Box-Cox Transformation Method

In fact, the Box-Cox transformation in the literature of hydrology (Eq. (4.3:10)) is different from the original Box-Cox transformation (Eq. (4.3:9)). Actually, the original Box-Cox transformation of model residuals is

$$g(e) = \frac{e^\lambda - 1}{\lambda} = \frac{(obs - sim)^\lambda - 1}{\lambda} \tag{B1}$$

where e is the error between observed (*obs*) and simulated (*sim*) outcome.

If the model residuals after BC transformation (Eq. (B1)) follow the distribution of $f(g)$ (usually assumed as the Gaussian distribution), the raw model residuals (Eq. (4.3:1)) will follow the error distribution of $p(e)$:

$$p(e) = J_{g \to e} f(g) \tag{B2}$$

where $J_{g \to e}$ is the Jacobian determinant of the transformation from g to e:

Appendices

$$J_{g\to e} = \frac{dg}{de} = e^{\lambda-1} = \exp((\lambda-1)\ln(e)) \tag{B3}$$

The Eqs. (B1, B2 and B3) reveal the aim of original BC transformation is to convert the non-Gaussian error to the Gaussian error. However, the hydrological BC transformation (Eq. (4.3:10)) does not have the Jacobian determinant. If we simply transplant the Jacobian determinant (Eq. (B3)) from the original BC transformation (Eq. (4.3:9)) to the hydrological BC transformation (Eq. (4.3:10)) (Eq. (4.3:11); Bates and Campbell, 2001; Yang et al., 2007a, b):

$$J_{G\to e} = \frac{dG}{de} = \frac{d(g(obs)-g(sim))}{de} \approx \left.\frac{d(g(y))}{dy}\right|_{y=obs} = \exp((\lambda-1)\ln(obs)) \tag{B4}$$

The estimate of lambda must be zero. In other words, the constraint (Eq. (4.3:11)) of the lambda is incorrect.

Actually, the hydrological BC transformation (Eq. (4.3:10)) redefines the errors for removal of the heteroscedasticity of model residuals, i.e. amplification of the model residual of low flow and mitigation of the model residual of high flow. So the lambda essentially belongs to the hydrological model parameters, rather than the error model parameters. But it is also independent of other hydrological model parameters, and if there is no constraint of lambda (λ) (i.e. totally treating lambda as the hydrological model parameter), the inference result of lambda (λ) will always approach one (i.e. no transformation of model residuals) when there are many small observed outcomes (Laloy et al., 2010). It probably results from that the mode (i.e. the highest probability point) of errors is zero, and the no-transformation of errors (closed to zero) contributes to the maximization of the likelihood function. Therefore, a constraint of the lambda is necessary to yield an effective BC transformation parameter for removal of the heteroscedasticity of model residuals.

Appendix C: Unifying Distance-based Objective Functions by BC-GED

The BC-GED error model assumes the model residuals after BC transformation follow the GED with zero-mean, of which the logarithmic likelihood function can be expressed as:

$$l(\theta) = \sum_{1}^{n} \ln(\text{GED}(e_i)) = n\ln\omega(\beta) - n\ln\sigma - \frac{c(\beta)^{\beta}}{\sigma^{\beta}} \sum_{1}^{n} |e_i|^{\beta} \tag{C1}$$

where $c(\beta) = \sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}$, $\omega(\beta) = \frac{\beta c(\beta)}{2\Gamma[1/\beta]}$, σ termed standard deviation and β termed kurtosis are parameters of the probability density function of GED ($\sigma, \beta > 0$; Eq. (4.3:13)), n

140

is the length of model residuals, e_i is the model residual at time step i, $\Gamma[x]$ is the gamma function evaluated at x, and GED(x) is the generalized error distribution function evaluated at x.

The maximum likelihood function is frequently used to estimate model parameters, which is termed maximum likelihood estimation (MLE). When

$$\sigma = c(\beta)\sqrt[\beta]{\beta}\sqrt[\beta]{\frac{\sum_1^n |e_i|^\beta}{n}} = \sqrt[\beta]{\beta}\sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}\sqrt[\beta]{\frac{\sum_1^n |e_i|^\beta}{n}} \tag{C2}$$

The likelihood function equation (C1) reaches the maximum value (as detailed in **Appendix D**):

$$l(\theta)_{max} = -n\ln\frac{\sigma\sqrt[\beta]{\varepsilon}}{\omega(\beta)} = -n\ln\frac{2\Gamma[1/\beta]}{\beta}\sqrt[\beta]{\varepsilon\beta\frac{\sum_1^n |e_i|^\beta}{n}} \tag{C3}$$

where ε is the base of the natural logarithm, ≈ 2.718.

Because n is a constant, the likelihood ($l(\theta)_{max}$) in equation (C1) can be simplified as:

$$\text{BC-GED} = \frac{\Gamma[1/\beta]}{\beta}\sqrt[\beta]{\varepsilon\beta\frac{\sum_1^n |e_i|^\beta}{n}} = \frac{\Gamma[1/\beta]}{\beta}\sqrt[\beta]{\varepsilon\beta\frac{\sum_1^n |g(obs_i) - g(sim_i)|^\beta}{n}} \tag{C4}$$

where BC-GED is a unified distance-based objective function termed BC-GED model, β is the power of model residuals, and $g(x)$ is the Box-Cox transformation function evaluated at x (Eq. (4.3:10)).

Appendix D: Maximum Likelihood of BC-GED

The derivative of logarithmic likelihood function at maximum point should equal to zero, so from the likelihood function equation (Eq. (C1)) of BC-GED error model, we can derive:

$$\frac{\partial(l(\theta))}{\partial\sigma} = -\frac{n}{\sigma} + \frac{c(\beta)^\beta \beta}{\sigma^{\beta+1}}\sum_1^n |e_i|^\beta = 0 \tag{D1}$$

where $c(\beta) = \sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}$, σ termed standard deviation and β termed kurtosis are parameters of the probability density function of GED (σ, $\beta > 0$; Eq. (4.3:13)), n is the length of model residuals, e_i is the model residual at time step i, and $\Gamma[x]$ is the gamma function evaluated at x.

By solving σ from Eq. (D1), we get:

$$\sigma = c(\beta)\sqrt[\beta]{\beta}\sqrt[\beta]{\frac{\sum_1^n |e_i|^\beta}{n}} = \sqrt[\beta]{\beta}\sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}\sqrt[\beta]{\frac{\sum_1^n |e_i|^\beta}{n}} \tag{D2}$$

Eq. (D2) is a method to estimate the standard deviation of model residuals. Its mathematic expectation can be expressed as:

$$\mathrm{E}(\sigma) = \sqrt[\beta]{\beta}\sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}\sqrt[\beta]{\mathrm{E}(\frac{\sum_1^n |e_i|^\beta}{n})} \tag{D3}$$

If e_i follows the generalized error distribution (GED) with zero-mean, the Eq. (D3) can be rewritten:

$$\begin{aligned}
\mathrm{E}(\sigma) &= \sqrt[\beta]{\beta}\sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}\sqrt[\beta]{\int_{-\infty}^{+\infty} |x|^\beta \frac{\omega(\beta)}{\sigma}\exp(-c(\beta)^\beta \frac{|x|^\beta}{\sigma^\beta})dx} \\
&= \sqrt[\beta]{\beta}\sqrt{\frac{\Gamma[3/\beta]}{\Gamma[1/\beta]}}\sqrt[\beta]{\frac{1}{\beta}(\sigma\sqrt{\frac{\Gamma[1/\beta]}{\Gamma[3/\beta]}})^\beta} \\
&= \sigma
\end{aligned} \tag{D4}$$

where $\omega(\beta) = \dfrac{\beta c(\beta)}{2\Gamma[1/\beta]}$.

Eq. (D4) means the Eq. (D2) is the unbiased estimation of the standard deviation of model residuals when the model residuals follow the GED with zero-mean.

Appendix E: Calibration Results Using the BC-GED (objective) Functions

In order to assess the effect of the parameters of BC-GED model on model calibration, two classifications of distance-based objective functions were selected:

1. BC transformation is not used, i.e. BC parameter λ is imposed equal to 1.0. But the kurtosis coefficient β is imposed equal to 2.0, 1.5, 1.0 and 0.5, respectively, which were marked GED2.0 (equivalent to the MSE), GED1.5, GED1.0 (equivalent to the MAE) and GED0.5, respectively. Or β is calculated from a constraint that minimizes the Eq. (6.4:1), which is marked GED.

2. BC transformation is used (Figure 6.1), where λ is estimated based on the minimum variance constraint (Eq. (4.3:12)), and the kurtosis coefficient β is set by the same method as the first classification.

In the automatic calibration procedure (Figure 2.1), SWAT-WB-VSA (developed in this study; Section 4.1) is used as the hydrological model tool, the river discharges (Section 3.5.1) as the observed data, the DREAM (Section 2.2.1) as the optimization tool and the flow parameters (Table 4.2) as the calibrated parameters. The model calibration results of the two classifications of BC-GED objective functions are shown in Table E1 and E2, respectively.

Table E1 Comparison of calibration results using different objective functions without Box-Cox transformation.

Categories	Parameter	MSE	GED1.5	MAE	GED0.5	GED
BC-GED [1]	λ	1	1	1	1	1
	β	2	1.5	1	0.5	0.307
River-flow performance [2]	NSE	0.857	0.854	0.822	0.667	0.648
	R^2	0.860	0.859	0.849	0.778	0.769
	VE	0.486	0.503	0.607	0.582	0.571
Ground-water performance [3]	R^2	0.486	0.460	0.722	0.608	0.604
	S_y	0.081	0.078	0.111	0.100	0.104
Average annual runoff components (mm/yr) [4]	Evapotranspiration	494	488.8	485.3	523.5	524.8
	Overland flow	35.02	37.85	45.74	60.38	53.41
	Interflow	13.02	17.64	119.92	82.43	86.82
	Return flow	198.88	180.19	104.63	96.07	93.74
	Revaporization	76.58	73.62	45.29	14.41	13.45
	Deep recharge	0.09	19.14	10.37	44.34	41.41
Evapo-transpiration [5]	v__ESCO	0.021	0.064	0.198	0.154	0.215
	v__EPCO	0.720	0.734	0.128	0.997	0.998
Surface water	v__EDC	0.788	0.776	0.781	0.764	0.764
	v__SURLAG	21.268	20.974	2.550	0.490	0.558
	r__SOL_Z	0.253	0.286	0.890	0.833	0.833
Soil water	r__SOL_BD	0.345	0.465	0.405	0.243	0.200
	r__SOL_AWC	-0.021	-0.071	1.834	1.229	1.407
	r__SOL_K	-0.710	-0.684	-0.195	-0.175	-0.075
	v__GW_DELAY	0.083	0.100	0.859	3.592	3.631
	v__ALPHA_BF	1.000	1.000	0.936	0.996	0.876
Ground water	v__GWQMN	865.628	838.449	548.907	294.973	442.131
	v__RCHRG_DP	0.000	0.067	0.061	0.277	0.257
	v__REVAPMN	836.881	813.388	527.690	295.347	442.473
	v__GW_REVAP	0.200	0.200	0.183	0.196	0.193
Main channel	v__CH_N2	0.059	0.057	0.031	0.012	0.010

[1] The parameters of BC-GED objective function (Eq. (6.4:1)): λ is the BC transformation parameter (Eq. (4.3:9)); and β is the kurtosis coefficient of GED (Eq. (6.4:1)).

^{2} NSE, R^2 and VE are Nash–Sutcliffe efficiency coefficient (Eq. (4.3:6)), coefficient of determination (Table 2.1, ID 1) and volumetric efficiency (VE; Eq. (4.3:21)), respectively, for ascertaining the performance of river flow simulation. The greater the value of NSE/R^2, the better the performance of flood simulation is. The greater the value of VE, the better the balanced consideration of the flood and baseflow is.

^{3} R^2 and S$_y$ are the coefficient of determination and the specific yield of unconfined aquifer (Eq. (5.1:1)), respectively. The larger the value of R^2, the better the performance of groundwater simulation is.

^{4} Runoff components are produced by SWAT-WB-VSA with optimum parameter set.

^{5} Optimal values of model parameters correspond to the minimization of BC-GED (objective) functions. The detailed descriptions of model parameters are shown in Section 4.2.2 and Table 4.2.

Table E1 shows: (1) the NSE/R^2 clearly increases with the value of kurtosis β, which means that the larger the value of β, the greater emphasis is on flood; and (2) the objective functions can be clearly separated into two classes by the kurtosis β: the class $\beta > 1$ including MSE and GED1.5 and the class $\beta \leq 1$ including MAE, GED0.5, and GED. The calibration results are substantial differences between the two classes of objective functions but tiny differences among the same class. Values of VE and groundwater performance indicators (R^2 and S$_y$) of the class $\beta \leq 1$ are obviously greater than that the class $\beta > 1$. It demonstrates the results of class $\beta \leq 1$ mimic the groundwater levels and the baseflow obviously better than that of class $\beta > 1$.

In Table E2, the performance indicators (NSE/R^2) of river discharges still increase with the value of kurtosis β. Compared with the Table E1, however, the Table E2 shows that the effect of the kurtosis β on hydrological components is obviously weakened after using the filter of BC transformation (Figure 6.1). The bigger values of VE and groundwater performance indicators (R^2 and S$_y$) in Table E2 indicate all objective functions with Box-Cox transformation can make SWAT-WB-VSA well simulate the baseflow and groundwater levels.

In summary, the different objective functions are in favor of different hydrographic components. Among these objective functions, the MAE is the best to balance consideration of the flood and baseflow; and the BC-GED is the best to guarantee that the model residuals fulfill the statistical assumptions of error model (Figure 5.3 (b) and Figure 5.4 (b)).

Table E2 Comparison of calibration results using different objective functions with Box-Cox transformation.

Categories	Parameter	BC-MSE	BC-GED1.5	BC-MAE	BC-GED0.5	BC-GED [1]
BC-GED	λ	0.444	0.437	0.437	0.437	0.439
	β	2.0	1.5	1.0	0.5	0.697
River-flow performance	NSE	0.749	0.729	0.729	0.728	0.728
	R^2	0.801	0.788	0.786	0.776	0.785
	VE	0.5935	0.5922	0.5922	0.5856	0.5911
Ground-water performance	R^2	0.621	0.629	0.625	0.625	0.629
	S_y	0.114	0.115	0.113	0.111	0.113
Average annual runoff components (mm/yr)	Evapotranspiration	519.1	519.8	518.6	520.7	520.5
	Overland flow	63.86	66.23	66.69	78.11	65.08
	Interflow	81.88	86.07	89.83	89.02	93.68
	Return flow	118.14	108.5	107.13	105.81	107.34
	Revaporization	9.5	11.25	12.61	12.46	12.12
	Deep recharge	26.61	27.76	25.48	9.29	21.46
Evapotranspiration	v__ESCO	0.010	0.146	0.133	0.127	0.086
	v__EPCO	0.700	1.000	1.000	1.000	1.000
Surface water	v__EDC	0.753	0.753	0.748	0.742	0.747
	v__SURLAG	0.570	0.552	0.589	0.523	0.593
Soil water	r__SOL_Z	0.857	0.868	0.862	0.843	0.864
	r__SOL_BD	0.304	0.273	0.277	0.260	0.276
	r__SOL_AWC	0.997	1.103	1.089	1.192	1.094
	r__SOL_K	-0.239	-0.169	-0.084	-0.020	-0.007
Ground water	v__GW_DELAY	1.645	2.013	2.221	2.417	2.021
	v__ALPHA_BF	1.000	1.000	1.000	1.000	1.000
	v__GWQMN	401.652	372.358	371.801	472.508	371.963
	v__RCHRG_DP	0.164	0.179	0.167	0.066	0.145
	v__REVAPMN	402.732	373.055	372.277	472.935	372.433
	v__GW_REVAP	0.200	0.200	0.200	0.200	0.200
Main channel	v__CH_N2	0.015	0.014	0.014	0.014	0.014

[1] BC-GED objective function (i.e. formal likelihood function) first estimates λ based on the minimum variance constraint (Eq. (4.3:12)) for most effectiveness of removing the heteroscedasticity of model residuals, and then calculates β from a constraint that minimizes the Eq. (6.4:1) for best fitting the distribution of model residuals.

Curriculum Vitae

For reasons of data protection,

the curriculum vitae is not included in the online version.

For reasons of data protection,

the curriculum vitae is not included in the online version.