Sobol′'s sensitivity analysis for TOPMODEL hydrological model: A case study for the Biliu River Basin, China[1]

W. Qi, C. Zhang *, J. Chu and H. Zhou
School of Hydraulic Engineering, Dalian University of Technology, Dalian 116024, China

Abstract: *For an in-depth understanding of TOPMODEL performance and influences of TOPMODEL parameters on flood simulation, the global sensitivity analysis methodology, Sobol′'s method, is used in this paper to assess TOPMODEL parameter sensitivity in terms of individual parameter, interaction of parameters and combination of parameters using four flood prediction matrices. The results show that all the TOPMODEL parameters have little influences individually on Nash–Sutcliffe Efficiency (NSE), Relative Error of Runoff Amount (RERA) and Relative Error of Peak Flow (REPF). However, the effective lateral saturated transmissivity has a noticeable influence on Peak Flow Time Error (PFTE). The combinations of parameters, form of exponential decline in conductivity and unsaturated zone time delay have notable influences on NSE, RERA and REPF. The combination of parameters, form of exponential decline in conductivity and effective lateral saturated transmissivity has remarkable influences on PFTE. These findings are important and will assist in understanding the performance of TOPMODEL and its calibration for flood prediction.*

Keywords: Sensitivity analysis, Hydrological modeling, Sobol′'s method, TOPMODEL, Flood simulation

1 Introduction

Hydrological models have been popularly implemented in basin hydrological prediction which is needed for many aspects of water resources management. Developed by Beven in 1979, TOPMODEL was one of the first attempts to simulate distributed hydrological responses. As to parameter sensitivity analysis of TOPMODEL, the Mento Carlo based Hornberger–Spear–Yong (HSY) local sensitivity method is often used; however, the interaction of parameters and their combination effects are not easy to detect.

Sobol′'s method is a global sensitivity analysis method and is able to identify the influence of each parameter, interaction of parameters and their combination effects on the model outputs (Sobol′, 1993). Recently Sobol′'s method has become increasingly prevalent in hydrological modeling as its ability to integrate consideration of parameters interaction and the relatively straightforward interpretation of the associated results (e.g., Pappenberger et al., 2008; Van Werkhoven et al., 2008; Yang, 2011; Fu et al., 2012). Tang et al. (2007b) comprehensively compared Sobol′'s method with other tools including local analysis, Regional Sensitivity Analysis (RSA), and Analysis of Variance (ANOVA). It found that Sobol′'s method is the most effective approach to characterize single- and multi-parameter interactive sensitivities for lumped watershed models. Furthermore, Tang et al. (2007a) used Sobol′'s method to a distributed hydrological watershed model named as the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM), and the results obtained reveals that the method provides robust sensitivity rankings.

Until now, Sobol′'s approach has not been used to analyze TOPMODEL parameter sensitivity in terms of individual parameter, interaction of parameters and their combination. Therefore, in this paper, Sobol′'s method is implemented to reveal sensitivity for TOPMODEL parameters in the Biliu River basin, China, in terms of four different matrices: Nash–Sutcliffe Efficiency (NSE), Relative Error of Runoff Amount (RERA), Relative Error of Peak Flow (REPF) and Peak Flow Time Error (PFTE). This paper is organized as follows. Section 2 describes some details of

[1] *Paper JHER005 submitted 17/09/2013; accepted for publication after peer review and subsequent revision on 19/10/2013*
**Corresponding author may be contacted at* czhang@dlut.edu.cn

TOPMODEL hydrological model and Sobol′'s method. Section 3 outlines the case study details. Section 4 provides the sensitivity analysis results and discussion. Conclusions are given in section 5.

2 Methodology

2.1 Overview of TOPMODEL hydrological model

TOPMODEL is a physically based, variable contributing area model of basin hydrology which attempts to combine the advantages of simple lumped parameter model with distributed effects (Beven et al., 1979). Fundamental of TOPMODEL's parameterization of the dynamic spatial hydrological response and the utilization of a topographic index of hydrological similarity are three basic assumptions: (1) saturated-zone dynamics can be approximated by successive steady-state representations; (2) hydrological gradients of the saturated zone can be approximated by the local topographic surface slope; and (3) the transmissivity profile with a form of exponentially declining along vertical depth of water table or storage is spatially constant. On the basis of above mentioned assumptions, the index of hydrological similarity is represented as the topographic index, , where is the area per unit contour length and is local slope angle. The greater upslope contributing areas and lower gradients areas are more likely to be saturated. More detailed descriptions of TOPMODEL and its mathematical formulation can be found in Beven et al. (1979).

As the relatively simple model structure, fewer model parameters included and easily implemented, TOPMODEL has been popularly utilized in many research studies (Bastola et al,. 2008, Blazkova et al., 1997, Bouilloud et al., 2010, Cameron et al., 1999, Gallart et al., 2008, Hossain et al., 2005). Take China as a particular example, TOPMODEL has been used in climate change study (Liu et al., 2012), DEM resolution influences on hydrological simulation (SUN Li-qun et al., 2008, Xiong et al., 2004, Xu et al., 2007), and other aspects (Guo et al., 2000, Wang et al., 2007). However, none of the above applications include a global sensitivity analysis to advance the understanding of influences of each parameter and parameters combination on the model performance in terms of different flood simulation matrices.

2.2 Sobol′'s method

Sobol′'s sensitivity is variance-based global quantitative sensitivity analysis method and has been popularly used in many aspects of hydrological modelling (Fu et al., 2012, Hall et al., 2005, Saltelli, 2002, Sobol', 2001, Tang et al., 2007, Zhang et al., 2013). Sobol′'s method (Sobol′, 1993) assumes that a model could be represented in the following functional form:

$$Y = f(X) = f(X_1, \ldots, X_p) \tag{1}$$

where Y is the goodness-of-fit metric of model output, and $X = (X_1, \ldots, X_p)$ is the parameter set. In Sobol′'s method, the total variance of function f, $D(y)$, is decomposed into component variances from individual parameter and their interactions:

$$D(y) = \sum_i D_i + \sum_{i<j} D_{ij} + \sum_{i<j<k} D_{ijk} + \ldots + D_{12\ldots p} \tag{2}$$

where D_i is the amount of variance due to the i th parameter X_i, and D_{ij} is the amount of variance due to the interaction between parameter X_i and X_j. The sensitivity of single parameter or parameters interaction, i.e. Sobol′'s sensitivity indices of different orders, is then assessed based on their percentage contribution to the total variance D:

First-order index $S_i = \frac{D_i}{D}$ (3)

Second-order index $S_{ij} = \frac{D_{ij}}{D}$ (4)

$$S_{Ti} = 1 - \frac{D_{\sim i}}{D} \quad (5)$$

where $D_{\sim i}$ is the amount of variance due to all of the parameters except for X_i, S_i measures the sensitivity from the main effect of X_i, S_{ij} measures the sensitivity from the interactions between X_i and X_j, and S_{Ti} measures the main effect of X_i and its interactions with all the other parameters.

The variances in Eq. 2 can be evaluated using approximate Monte Carlo numerical integrations, particularly when the model is highly nonlinear and complex. The Monte Carlo approximations for D, D_i, D_{ij}, and $D_{\sim i}$ are defined as presented in the following prior studies (Sobol′,1993, 2001; Hall et al. 2005):

$$\hat{f}_0 = \frac{1}{n}\sum_{s=1}^{n} f\left(X_s\right) \quad (6)$$

$$\hat{D} = \frac{1}{n}\sum_{s=1}^{n} f^2\left(X_s\right) - \hat{f}_0^{\,2} \quad (7)$$

$$\hat{D}_i = \frac{1}{n}\sum_{s=1}^{n} f\left(X_s^{(a)}\right) f\left(X_{(\sim i)s}^{(b)}, X_{is}^{(a)}\right) - \hat{f}_0^{\,2} \quad (8)$$

$$\hat{D}_{ij}^{c} = \frac{1}{n}\sum_{s=1}^{n} f\left(X_s^{(a)}\right) f\left(X_{(\sim i,\sim j)s}^{(b)}, X_{(i,j)s}^{(a)}\right) - \hat{f}_0^{\,2} \quad (9)$$

$$\hat{D}_{ij} = \hat{D}_{ij}^{c} - \hat{D}_i - \hat{D}_j \quad (10)$$

$$\hat{D}_{\sim i} = \frac{1}{n}\sum_{s=1}^{n} f\left(X_s^{(a)}\right) f\left(X_{(\sim i)s}^{(a)}, X_{is}^{(b)}\right) - \hat{f}_0^{\,2} \quad (11)$$

where n is the sample size, X_s is the sampled individual in the scaled unit hypercube, and superscripts (a) and (b) represent two different samples. All of the parameters take their values from sample (a) are represented by $X_s^{(a)}$. The variables $X_{is}^{(a)}$ and $X_{is}^{(b)}$ denote that parameter X_{is} uses the sampled values in sample (a) and (b), respectively. The symbols $X_{(\sim i)s}^{(a)}$ and $X_{(\sim i)s}^{(b)}$ represent cases when all of the parameters except for X_{is} use the sampled values in sample (a) and (b), respectively. The symbol $X_{(i,j)s}^{(a)}$ represents parameters X_{is} and X_{js} with sampled values in sample (a). Finally, $X_{(\sim i,\sim j)s}^{(b)}$ represents the case when all of the parameters except for X_{is} and X_{js} utilize sampled values from sample (b).

Sobol′'s sensitivity indices have been shown to be more effective than other approaches in capturing the interactions between a large number of variables for highly nonlinear models (Tang et al., 2007a and b). Building on the recommendations of Tang et al. (2007a), the Latin Hypercube sampling method (McKay et al., 1979) is used for implementing Sobol''s method. By using Sobol′'s improvement method (Saltelli, 2002), computing the first-order, second-order and total-order sensitivity indices only requires $n \times (m+2)$ model evaluations, where n is the number of Latin Hypercube samples and m is the number of parameters being analyzed.

3 Case study

3.1 Biliu River basin

Biliu River basin (2567 km^2), locating in coast region between the Bohai Sea and the Huanghai Sea of China, covers longitudes from 122.29°E to 122.92°E and latitudes from 39.54°N to 40.35°N. This basin is characterized by a temperate monsoon marine climate and summer (June to September) is the major rainfall period. The major land cover types are forest and farmland, and the average annual temperature is 10.6°C. Due to effects from ocean, the seasonal temperature difference is small. The basin average elevation is 240m. The maximal elevation is 985m in northern part, which is a mountainous region, and the minimum elevation is 4.5m in southern part. There are ten rainfall stations and a discharge gauge, and their spatial distributions are shown in Figure 1. The main study region is the upper sub-basin of Biliu reservoir, which accounts for about 74.1% of the total catchment area.

3.2 Data set

The required data sets for TOPMODEL modeling mainly include: precipitation data, observed flow, evaporation, topography index and cumulative area-distance relationship.

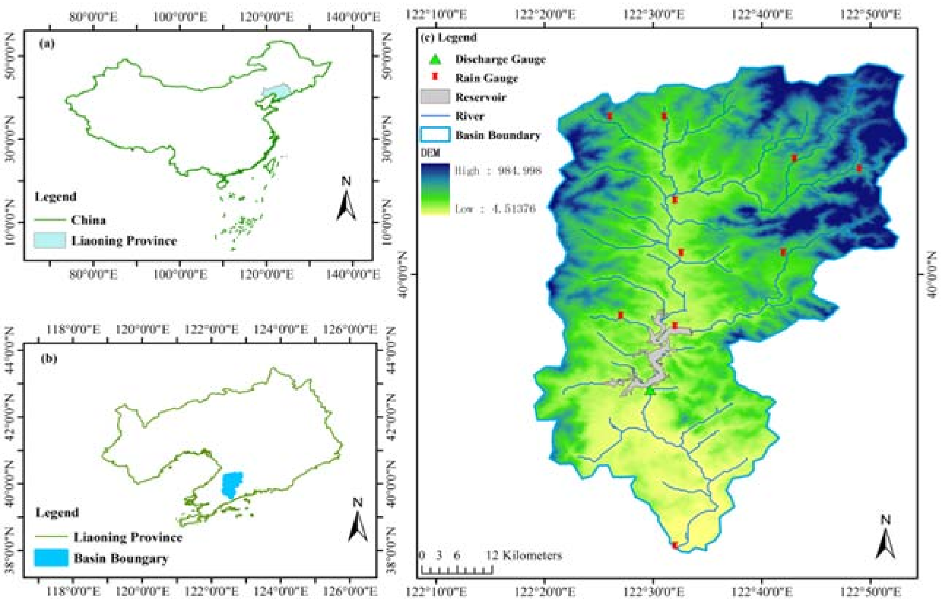

Figure 1 Biliu River Basin

In this paper, the basin average rainfall is calculated using thiessen method, and eight flood events of Biliu River basin are considered, including one, two, three, eight, thirty and forty years Average Recurrence Interval (ARI). Evaporation is estimated with equation 12.

$$E = E_p\left(1 - \frac{SR_0}{SR_{\max}}\right) \tag{12}$$

where SR_0 is the root zone storage deficit; SR_{max} is the maximum capacity of the root zone; and E_p is potential evaporation which can be obtained from meteorological gauge. Topography index and cumulative area-distance relationship are derived with FORTRAN program, and the spatial distribution of topography index is showed in Figure 2. Percentage distribution of topgraphy index and cumulative area-distance relationship are displayed in Figure 3 and Figure 4, respectively.

3.3 Model parameters

Six main parameters are included in TOPMODEL, and their ranges and brief description are listed in Table 1. Usually SR_0 and SR_{max} could be estimated by experience or observation. However, all these methods cannot provide precise parameter values. Therefore, SR_0 and SR_{max} are given by ranges. In this paper, the parameter value ranges are given based on prior research by Liu (2008).

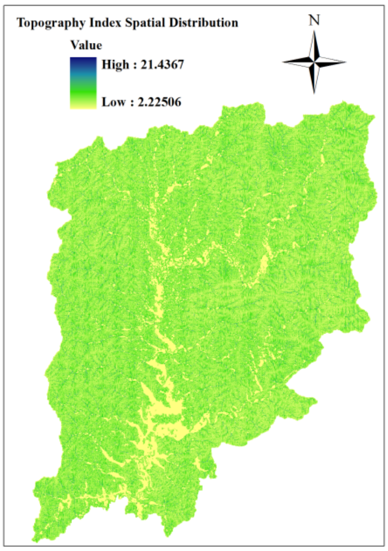

Figure 2 The spatial distribution of topography index

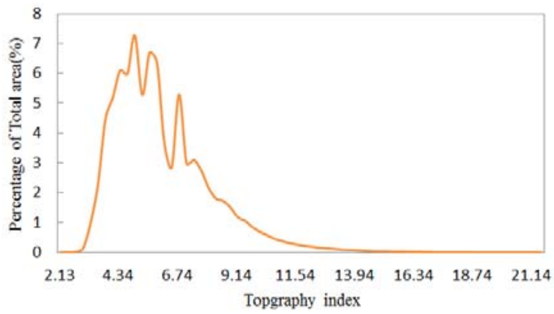

Figure 3 The distribution of topgraphy index

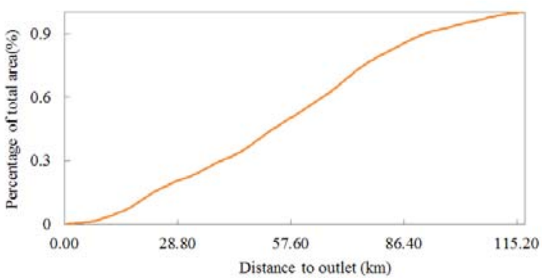

Figure 4 Cumulative area-distance relation

3.4 Goodness-of-fit matrices

The sensitivity analysis for TOPMODEL with Sobol′'s approach considers four goodness-of-fit matrices which are considered as major criteria to evaluate the applicability of flood prediction in China. These four matrices include Nash–Sutcliffe Efficiency (NSE), Relative Error of Runoff Amount (RERA), Peak Flow Time Error (PFTE) and Relative Error of Peak Flow (REPF), and they are calculated using equation 13, 14, 15 and 16, respectively.

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{pi} - Q_{ti}\right)^2}{\sum_{i=1}^{n}\left(Q_{ti} - \overline{Q}_t\right)^2} \tag{3}$$

$$RERA = \left(\frac{\sum_{i=1}^{n} Q_{pi}}{\sum_{i=1}^{n} Q_{ti}} - 1\right) \times 100\% \tag{14}$$

$$PFTE = T(Q_{pp}) - T(Q_{tp}) \tag{15}$$

$$REPF = \left(1 - \frac{Q_{pp}}{Q_{tp}}\right) \times 100\% \tag{16}$$

where Q_{pi} and Q_{ti} are the simulated and measured flows at time i ; n is the total number of flood time; $\overline{Q_t}$ is the mean measured flows during the flood period; $T(Q_{pp})$ is the simulated peak flood time of simulation; $T(Q_{tp})$ is the observed peak flood time; Q_{pp} is the simulated peak flood volume and Q_{tp} is the observed peak flood volume. The NSE should be more close to 1.0 if simulation is better. The lower the RERA, PFTE and REPF values are the better the simulation is performed. RERA, PFTE and REPF values of 0 indicate a perfect fit.

3.5 Sensitivity analysis

Statistical sample size is a vital parameter for Sobol′'s approach. Tang et al. (2007b) employed a sample size of 8192 for Sobol′'s analysis considering 18 model parameters. Fu et al. (2012) adopted a set of 2000 LHS samples for 21 parameters. Tang et al. (2007a) used a sample size of 2000 for 403 variables. On the basis of these research studies, this paper takes a LHS sample size of 3000 into our research which leads to 3000*(6+2) = 24000 model evaluations. The sensitivity analysis for TOPMODEL is based on the average NSE, RERA, PFTE and REPF of eight flood events, which stands for a common sensitivity.

4 Results

4.1 Flood simulation

The best simulations of the eight flood events within total 24000 model runs are shown in Figure 5 and the corresponding four matrices are displayed in Table 2. Table 2 and Figure 5 show that satisfactory flood simulation results are obtained. The aim of this paper is to analyze parameter sensitivity with Sobol′'s approach, i.e., to obtain information concerning how individual parameter and parameters combination influence model performances, which provides references to model parameter calibration.

4.2 First and total order sensitivity

The first and total order sensitivity indices are presented in Figure 6. In each panel, the x-axis represents parameters, and y-axis represents first and total order sensitivity indices. These indices are defined with a threshold of 1%, which is subjective and their ease-of-satisfaction decreases with increasing parameter interactions or numbers of parameters (Tang et al., 2007a.b).

Table 1 Parameter ranges for TOPMODEL

Parameter	Description	Lower bound	Upper bound
SZM [m]	Form of the exponential decline in conductivity	0.005	0.04
LNT0[m^2 h^{-1}]	Effective lateral saturated transmissivity	-25	10
RV[m^2 h^{-1}]	Hillslope routing velocity	3500	8000
SR_{max} [m]	Maximum root zone storage	0.001	0.01
SR_0[m]	Initial root zone deficit	0	0.01
TD [m h^{-1}]	Unsaturated zone time delay per unit deficit	0.5	5

Figure 6 reveals that each parameter of TOPMODEL has few individual influences on NSE, RERA and REPF, and LNT0 has many effects on PFTE. LNT0 mainly controls the relative proportion of surface flow and subsurface flow when surface soil is saturated. Usually surface flow has faster routing speed than subsurface flow. As the assumption of TOPMODEL, surface flow and subsurface flow have the same routing speed, therefore different component percentage of surface flow and subsurface flow could influence PFTE.

4.3 Second order sensitivity

The second order sensitivity indices are displayed in Figure 7. In each panel, the x-axis represents parameter couples, and y-axis represents sensitivity indices. These indices are defined with a threshold of 1%. Figure 7 reveals that the parameter couple, SZM and TD, has massive influences on NSE, RERA and REPF, and parameter couple, SZM and LNT0, has relatively remarkable effects on PFTE. It is mainly because that discharge of TOPMODEL consists of two sources: surface flow and subsurface flow, and SZM, LNT0 and TD have something to do with the saturated groundwater table variation and the delay time of runoff in soil profile.

Therefore, SZM, LNT0 and TD have many influences on runoff volume of every time step and the relative proportion of surface and subsurface flow. Furthermore, as mentioned above, TOPMODEL assumes that surface flow and subsurface flow have the same routing velocity (Beven et al. 2001, Beven et al. (1979), Peters et al., 2003), and hence discriminative proportion of surface and subsurface flow could influence runoff hydrograph.

Overall, it can be found from Figure 6 and Figure 7 that SR_{max} and SR_0 which primarily control evaporation have few influences. In the Biliu River basin, runoff volume is relatively high and flood has a fairly short duration and hence evaporation has only minor effects.

Table 2 Flood simulation results for the Biliu River basin

Flood	NSE	RERA(%)	PFTE	REPF (%)
19840615	0.96	-1.00	1	10.06
19850818	0.92	-9.78	1	7.52
19940815	0.96	-1.43	1	1.39
19950806	0.90	4.65	0	9.73
19960729	0.91	4.83	0	12.11
19980809	0.92	-1.32	1	1.28
20010816	0.85	-0.67	1	-1.69
20110807	0.97	-6.46	0	8.97

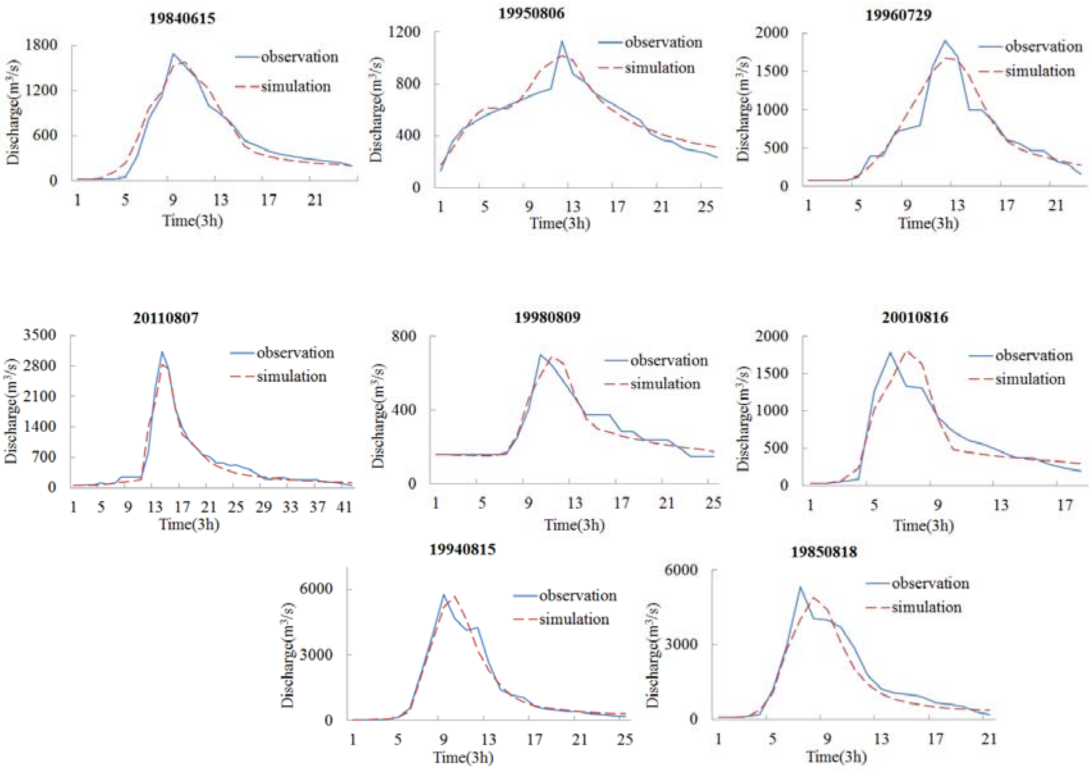

Figure 5 Flood simulation for the Biliu River basin

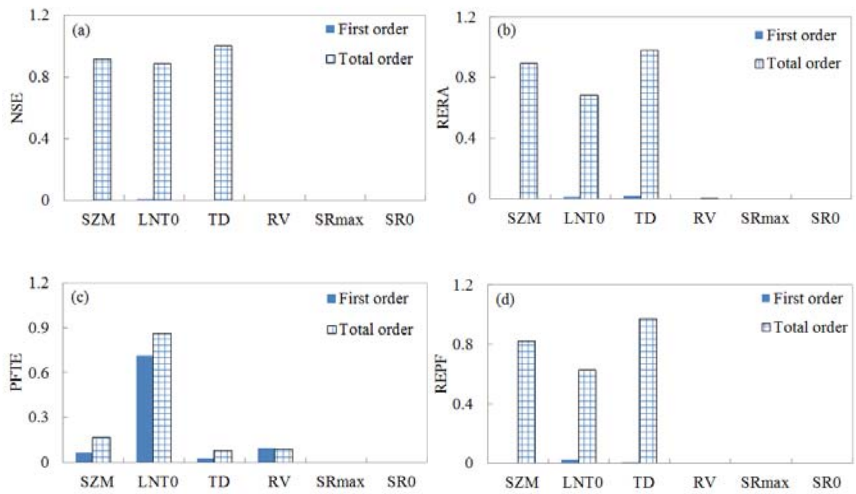

Figure 6 First order and total order sensitivity index of TOPMODEL parameters using the four goodness-of-fit matrices.

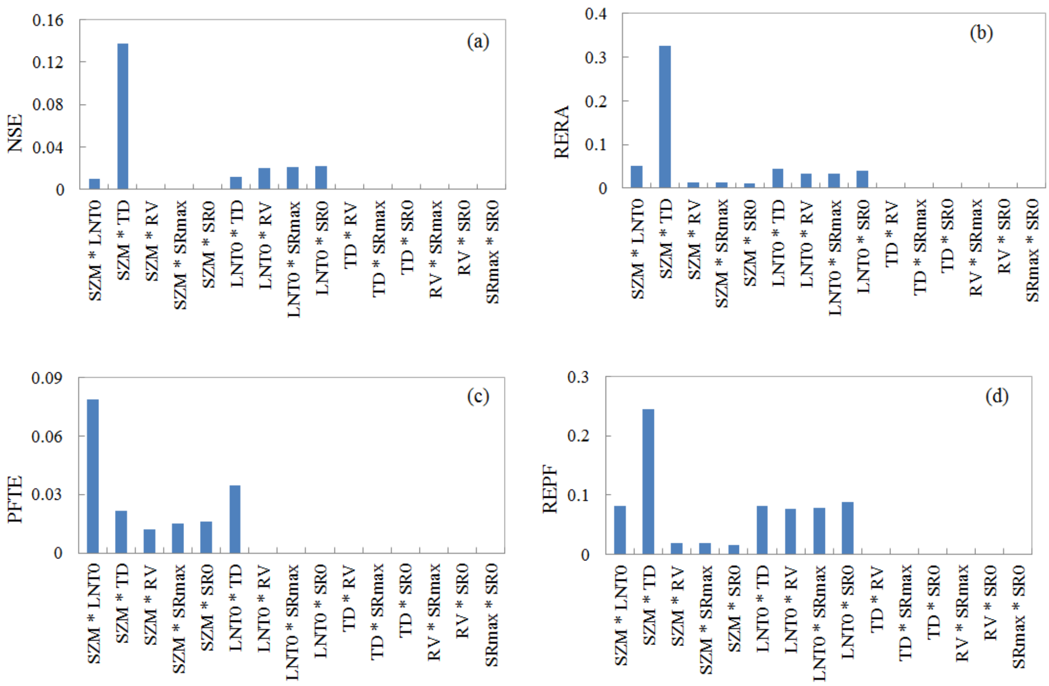

Figure 7 Second order sensitivity index of TOPMODEL parameters using the four goodness-of-fit metrics.

5 Conclusions

This paper firstly provides a variance-based sensitivity analysis for TOPMODEL for the Biliu River basin, China. The analysis reveals the individual effects, interaction effects and parameters combination effects on the model performances in terms of four matrices of flood prediction. The results of this paper can provide valuable references for parameter calibration of flood prediction using TOPMODEL. The main findings are summarized below:

i. Individual parameter of TOPMODEL has few influences on Nash–Sutcliffe Efficiency, Relative Error of Runoff Amount and Relative Error of Peak Flow;
ii. One parameter, effective lateral saturated transmissivity, has great influences on Peak Flow Time Error;
iii. The parameter couple, form of the exponential decline in conductivity and unsaturated zone time delay per unit deficit, has massive effects on Nash–Sutcliffe Efficiency, Relative Error of Runoff Amount and Relative Error of Peak Flow; and

iv. The parameter couple, form of the exponential decline in conductivity and effective lateral saturated transmissivity has many effects on Peak Flow Time Error.

Acknowledgements

The authors are grateful to the reviews whose comments are helpful to improve the quality of this paper. The authors also gratefully acknowledge financial support provided by the China Scholarship Council. This study was supported by the National Basic Research Program of China (Grant No. 2013CB036400), and the National Natural Science Foundation of China (Grant No. 51279021 and 51079014)

References

Bastola S, Ishidaira H, Takeuchi K (2008). Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe, *Journal of Hydrology*, 357(3-4), 188-206.

Beven K, Freer J (2001). A dynamic TOPMODEL, *Hydrological Processes*, 15(10), 1993-2011.

Beven KJ, Kirkby MJ (1979). A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, 24(1), 43-69.

Blazkova S, Beven K (1997). Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL, *Journal of Hydrology*, 195.

Bouilloud L, Chancibault K, Vincendon B (2010). Coupling the ISBA Land Surface Model and the TOPMODEL Hydrological Model for Mediterranean Flash-Flood Forecasting: Description, Calibration, and Validation, *Journal of Hydrometeorology*, 11(2), 315-333.

Cameron DS, Beven KJ, Tawn J (1999). Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), *Journal of Hydrology*, 219.

Fu G, Kapelan Z, Reed P (2012). Reducing the Complexity of Multiobjective Water Distribution System Optimization through Global Sensitivity Analysis, *Journal of Water Resources Planning and Management*, 138(3), 196-207.

Gallart F, Latron J, Llorens P (2008). Upscaling discrete internal observations for obtaining catchment-averaged TOPMODEL parameters in a small Mediterranean mountain basin, *Physics and Chemistry of the Earth*, 33(17-18), 1090-1094.

GUO F, LIU X, REN L (2000). A topography based hydrological model: TOPMODEL and its widened application, *Advances in Water Science,* 11(3): 296-301. (in Chinese)

Hall JW, Tarantola S, Bates PD (2005). Distributed Sensitivity Analysis of Flood Inundation Model Calibration, *Journal of Hydraulic Engineering*, 131.

Hossain F, Anagnostou EN (2005). Assessment of a stochastic interpolation based parameter sampling scheme for efficient uncertainty analyses of hydrologic models, *Computers & Geosciences*, 31(4), 497-512.

Liu Y, Zhang J, Wang G, Liu J, He R, Wang H, Jin J (2012). Evaluation on the influence of natural climate variability in assessing climate change impacts on water resources: I. Model and methodology in baseline period, *Advances in Water Science*, 23(2): 147-155. (in Chinese).

Peters NE, Freer J, Beven K (2003). Modelling hydrologic responses in a small forested catchment (Panola Mountain, Georgia, USA): A comparison of the original and a new dynamic TOPMODEL, *Hydrological Processes*, 17(2), 345-362.

Saltelli A (2002). Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications*, 145.

Sobol' IM (2001). Globalsensitivityindices for nonlinearmathematicalmodels and their MonteCarloestimates, *Mathematics and Computers in Simulation*, 55(1-3).

Sun L, Hu C, Chen G (2008). Effects of DEM resolution on the TOPMODEL, *Advances in Water Science*, 19(5): 699-706. (in Chinese).

Tang Y, Reed P, Van WK (2007). Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resources Research*, 43(6).

Wang L, Li Z, Bao H (2007). Application of hydrological models based on DEM in the Yihe basin, *Journal of Hydraulic Engineering*, S1:417-422. (in Chinese).

Xiong L, Guo S, Tian X (2004). DEM-based distribution hydrological model and its application, *Advances in Water Science*, 15(4): 518-520. (in Chinese).

Xu J, Ren L, Cheng Y, Yuan F (2007). Study on the translation relationship of DEM spatial scale based on TOPMODEL, *Journal of Hydraulic Engineering*, S1: 404-408. (in Chinese).

Zhang C, Chu J, Fu G (2013). Sobol′'s sensitivity analysis for a distributed hydrological model of Yichun River Basin, China, *Journal of Hydrology,* 480, 58-68.