Effects of model segmentation approach on the performance and parameters of the Hydrological Simulation Program – Fortran (HSPF) models

Changan Yan and Wanchang Zhang

ABSTRACT

Although models are one of the most powerful tools for watershed management, their effectiveness is limited by prediction uncertainties resulting from not only model input data but also spatial discretization. In this paper, Hydrological Simulation Program – Fortran (HSPF) models were constructed for the Linyi watershed according to three segmentation approaches including model segments based on differences in: (1) sub-watershed, (2) meteorological station, and (3) physical characteristics. Then the static sensitivity method and dynamic sensitivity method were employed to evaluate the effect of the segmentation approach on model performance and parameters of HSPF. The main conclusions were: (1) modeling with 12 segments had the best simulation efficiency and the corresponding estimated parameters had a certain representation within the Linyi watershed; (2) HSPF model performance was significantly affected by the segmentation approach, especially by the model segmentation construction process which considering a meteorological station or not; (3) parameters INTFW (interflow inflow parameter), lower zone nominal storage, and upper zone nominal storage (UZSN) were most affected by the model segmentation approach, while parameter AGWRC (groundwater recession coefficient) changed indistinctly; (4) parameters UZSN and INTFW had the same variation tendency whenever the segmentation approach changed.

Key words | hydrological model, Hydrological Simulation Program – Fortran, parameters, performance, segmentation approach, spatial discretization

Changan Yan
State Key Laboratory of Pollution Control & Resource Reuse,
School of the Environment,
Nanjing University,
Nanjing 210093,
China

Wanchang Zhang (corresponding author)
Key Laboratory of Digital Earth Science,
Institute of Remote Sensing and Digital Earth (RADI),
Chinese Academy of Sciences (CAS),
Beijing 100094,
China
E-mail: *wczhang@ceode.ac.cn*

INTRODUCTION

With the increasingly serious water crisis over the world, water resource management of watershed is extremely important. Watershed models, widely used to simulate watershed-scale water-cycle process and evaluate the hydrological effect of various management scenarios, have become one of the most powerful tools for watershed management in the last two decades (Albek *et al.* 2004). Since the physically based distributed hydrological models considering the spatial variability of hydrological processes predict the change tendency of water resources in watersheds more accurately, an important issue now for modelers is how well a model describes the physical characteristics of watersheds. In other words, how the modeling can obtain the optimal simulation performance and the most reasonable parameter. Therefore, much of current research in modeling is focused on the effect of the spatial-temporal heterogeneity of input data and discretization of watershed on the model simulation results.

Many studies exist on evaluating the impact of spatial-temporal variability of rainfall on the accuracy of model predictions (Sun *et al.* 2000; Andreassian *et al.* 2001; Krajewski & Smith 2002; Moon *et al.* 2004; Chaplot *et al.* 2005; Cho *et al.* 2009; Reichert & Mieleitner 2009; Pechlivanidis *et al.* 2010; Mohamoud *et al.* 2012). Some investigators showed that higher spatial rainfall resolution leaded to an increased efficiency of streamflow predictions for a physically based

doi: 10.2166/nh.2014.182

distributed hydrological model (Sun *et al*. 2000; Pechlivanidis *et al*. 2010; Xu *et al*. 2013). Some investigators reported that model performance had improved because of higher temporal resolution (Krajewski & Smith 2002). And other researchers analyzed the interaction between temporal and spatial variability of rainfall (Reichert & Mieleitner 2009; Mohamoud *et al*. 2012), and pointed out that greater improvements were due to spatial resolution rather than temporal resolution.

Additionally, there have been several articles in the literature conducted to assess the impact of the spatial resolution of DEM (Digital Elevation Model), land use, and soils on the accuracy of the simulated model output (Bosch *et al*. 2004; Chaplot 2005; Chaubey *et al*. 2005; Wang & Melesse 2006; Du *et al*. 2009; Gong *et al*. 2009; Donnelly *et al*. 2013; Li *et al*. 2014). From these, we can conclude that the coarse spatial resolution may reduce the prediction accuracy of streamflow and sediment simulations. Probably because low resolution of DEM leads to a decrease of derived slope resulting in longer travel times and lower peak flows (Du *et al*. 2009). Gong *et al*. (2011) developed a TRG (topographic runoff – generation) algorithm based on high-resolution topography to describe runoff generation at sub-cell and cell levels for global hydrological models. The application results showed that the model using the TRG algorithm performed equally well or slightly better. Hughes *et al*. (2013) made an investigation of spatial scale effects on the application of an approach to quantify the parameter values of a rainfall-runoff model with a relatively large number of parameters. The consequence of this approach was that the greater the diversity of topography and pedology within a land type, the greater the uncertainty in the parameter estimates.

More recently, there have been some investigations into the effect of temporal sampling of data on model performance and model parameters. Bastola & Murphy (2013) revealed a general increase in modeled peaks when moving from longer to shorter time steps, and the sensitivity of these parameters increased when the calibration time step is decreased from daily to 3-hourly. Littlewood & Croke (2013) further reported that the inaccuracy of calibrated parameters in a discrete-time rainfall-streamflow model was caused by data time-step.

Moreover, a number of studies have been reported which examine the impact of spatial discretization of watershed on model simulation. Beldring *et al*. (2003) investigated a typical approach for HRU (hydrologic response unit) delineation which combined with land use and soil types to improve the prediction of streamflow. Xu *et al*. (2012) explored a discretization method considering land use, soil types, the hydrological condition of vegetation and slope to improve the simulation efficiency and determine the CN (curve number) value more accurately. Haverkamp *et al*. (2005) demonstrated that the model efficiency improved when spatial heterogeneity of the watershed was considered.

In summary, it indicated that hydrological simulation is influenced by the spatial discretization of basin physical characteristics and spatio-temporal heterogeneity of meteorological inputs. Therefore, it is expected that spatially distributed hydrologic models outperform their lumped counterpart which do not account for such variabilities (Kumar *et al*. 2010). However, there are many more parameters to estimate in distributed hydrological models which represent hydrological processes and make significant effects on model performance. Meanwhile, the model performance can vary depending on model parameters and result in parameter uncertainty. From the literature review, most of the studies only evaluate the effect of spatial discretization on predictive performance, while the impact of spatial discretization on model parameters, as well as the interaction between model performance and parameters have rarely been examined, particularly in the Hydrological Simulation Program – Fortran (HSPF) model which is a distributed, continuous time watershed scale model developed to simulate water quantity and quality at any point in a watershed. The HSPF model has been successfully applied to the study watershed in our previous research. However, it is difficult for us to know to what extent spatial discretization affects model performance in terms of model parameter values. For the HSPF model, dividing the level of discretization of the model creates model segments and is referred to as segmentation which takes spatial heterogeneity into account including topography, meteorological data, soils variations, etc. It is expected that performance should be in principle improved with an increasing level of spatial discretization, because spatial discretization can be regarded as a tradeoff between reality and a model (Xu *et al*. 2012). However, over-parameterization and scaling problems also emerge at the same time. Thus, it is crucial to investigate an optimal segmentation approach which has main influences on HSPF model performance, the discretized factor which affects HSPF model efficiency significantly, and the parameter which changes markedly in different segmentation approaches.

The main objective of this paper is to assess the effect of the segmentation approach on HSPF performance and parameter estimation for the Linyi watershed which is typical of a semi-humid climate and located on northeast of the Huaihe River Basin, eastern China.

MATERIALS AND METHODS

Study area

Yihe River is a main tributary in Yishusihe River systems of the Huaihe River Basin. It originates from the northern slope of Niujiaoshan Mountain (Gosrnga) in Yiyuan, Shandong province, drains an extensive area with a river channel about 333 km in length and flows across Yishui, Yinan, Linyi, Mengyin, Pingyi, Yancheng counties, and ultimately discharges into the Luoma Lake in Xinyi, Jiangsu province. Because downstream of the Yihe River changes its direction to the Shuhe River, the selected watershed in this study is restricted to the upstream section of the Yihe River that is controlled by the Linyi Hydrometric Station alone. It is referred to as the Linyi watershed. The total drainage of this watershed is about 10,315 km^2 and 223 km length, accounting for about 87.3% of the Yihe watershed. The topography is characterized by high elevation in northwest and getting lower towards the southeast, the elevation of the watershed ranges from 57 to 1,004 m above sea level. The mountainous area and plain area cover about 70% and 30%, respectively, of the studied watershed. It has a typical temperate continental monsoon climate with an annual average temperature of 14.1 °C and an annual average precipitation of 830 mm. Rainfall shows high seasonal variability, with an average precipitation of 616 mm and an average runoff of 36.06 million m^3 between July and September. It is also a typical semi-humid climate region. Most of the watershed area is covered by agricultural land (about 62%), whereas the forest land accounts for about 19%, and other land use types cover small areas of the watershed including urban or built-up land (about 3%) and wetlands/water (about 8%). The dominant soils of the watershed are cinnamon soil (about 36%), brown soil (about 35%), paddy soil (about 13%), moisture soil (about 10%) and Shajiang black soil (about 3%). See Figure 1 for a map of the study area.

Description of HSPF model

HSPF model has been developed by the US Environmental Protection Agency (USEPA) to simulate water quantity and quality at any point in the watershed. It is a distributed, continuous time watershed scale model evolved from the Stanford Watershed Model in which mathematics methodology is applied to a hydrological calculation and forecast. The HSPF model is composed of three application modules: PERLND (Pervious land), IMPLND (Impervious land) and RCHRES (Reach & Reservoir) (Bicknell *et al.* 2001). PERLND and IMPLND modules simulate the hydrologic and water quality processes over pervious land surfaces and impervious land surfaces, respectively. The RCHRES module is utilized to represent hydraulic and water quality processes for streams, well-mixed lakes and reservoirs (Mohamoud & Prieto 2012). Water quality constituents simulated by HSPF include temperature, dissolved oxygen, biochemical oxygen, sediment detachment and transport, sediment routing, nitrate, organic nitrogen, organic phosphorus, orthophosphate, ammonia nitrogen, pesticides, conservatives, phytoplankton and zooplankton (Diaz-Ramirez *et al.* 2011). The hydrologic components of HSPF consist of five storage classes (interception, upper zone, lower zone, baseflow and deep percolation) and each of them allows different types of inflow and outflow. All inflows and outflows are modeled based on the principle of water-balance. The following processes such as interception, evapotranspiration, surface detention, surface runoff, infiltration, shallow subsurface flow (interflow), baseflow and deep percolation (Donigian *et al.* 1995) occur in each pervious land segment (hydrologic response unit). HSPF applies Manning's equation for routing overland flow and kinematic wave method for channel routing (He & Hogue 2012). Detailed information about the structure and theories of HSPF can be found in the HSPF Version 12 user's manual (Bicknell *et al.* 2001). HSPF requires extensive data input and complex procedures in the initial stage. For the

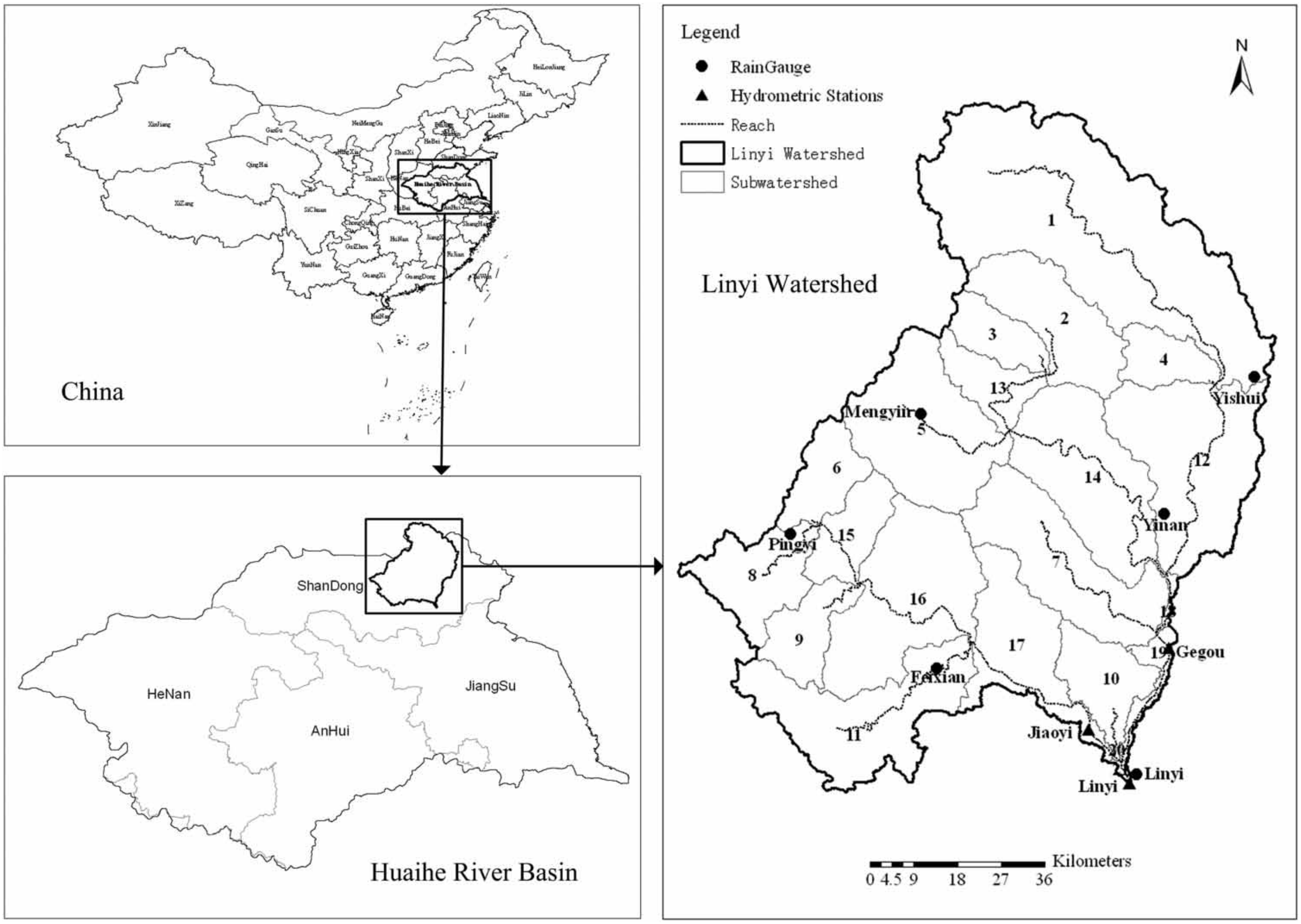

Figure 1 | Geographical location of study area within the Huaihe River Basin, China.

user's convenience, the USEPA has developed the Better Assessment Science Integrating Point and Non-point Sources (BASINS), a watershed management system based on geographical information system (GIS) which integrates hydrological models such as HSPF, Soil and Water Assessment Tool, Pollutant Load model, and Automated Geospatial Watershed Assessment, as well as auxiliary means: Watershed Data Management Utility and Generation and Analysis of Model Simulation Scenarios.

Data collecting

The data required for driving HSPF, including DEM, land use, soil data, and meteorological data, were established by BASINS/HSPF Map-Window interface. The DEM for the study area, with 30×30 m resolution, was downloaded from the International Scientific Data Service Platform, Chinese Academy of Science. The land use data of China with a 1 km resolution in China were obtained from Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences. The land use data were acquired using the method of supervised classification based on remote sensing Landsat ETM images. The soil data was downloaded from the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences. The meteorological data for streamflow simulation (2001–2006) were collected from six weather stations (Figure 1) nearest to the study watershed. These time series of meteorological data, including precipitation, evapotranspiration, air temperature, wind speed, solar radiation, dew-point temperature and cloud cover, were stored in the Watershed Data Management (WDM) program from BASINS.

Besides these data, HSPF also needs time series of observed runoff for model calibration and validation. The observed daily flow data were obtained from the Linyi Hydrometric Station located at the watershed outlet (latitude 35°01′; longitude 118°24′) for the period 2001 to 2006.

SEGMENTATION APPROACH

According to Singh (1997), the following factors representing the spatial heterogeneity of the HSPF model are mainly considered: topography, meteorological data, physical characteristics of soil, reach characteristics, and any other important physical characteristic (infiltration, overland slope, etc.). In HSPF, model segments, the sub-areas of a watershed with uniform parameters that are connected by a reach network, are specially created to account for differences in topography, meteorological data, physical characteristics of soil, etc. Hence, three segmentation approaches used in HSPF will be introduced and examined to assess model performance and parameter estimation.

Model segments based on differences in sub-watershed

Based on the topographic characteristic, the study watershed is divided into 21 segments corresponding to sub-watersheds. Once segmented, it is necessary to assign an appropriate meteorological station to each model segment using Thiessen polygons.

Model segments based on differences in meteorological station

The study watershed consists of 21 sub-watersheds, and there are six meteorological stations near the watershed that are representative of different portions of the drainage. To highlight the variation in meteorological data, the study watershed is partitioned into six segments.

Model segments based on differences in physical characteristics

Considering that heterogeneity of physical characteristics may be the basis for model segmentation, we should segment our model based on differences in physical characteristics in addition to the meteorological conditions. In other words, two different model segments can use the same meteorological data but may have different physical characteristics. In the study watershed, the predominant soil types of 12 sub-watersheds (1, 2, 4, 6, 7, 10, 11, 12, 14, 15, 16, and 17) are silt. Sub-watersheds 3, 8 and 13 represent land where the main soil type is sand. The remaining sub-watersheds (5, 9, 18, 19, 20, and 21) mainly consist of silt and sand. Moreover, the physical characteristics of three soil types which are a representation of each sub-watershed are significantly different. And these differences will influence factors such as subsurface water storage, surface roughness and infiltration. On the basis of the previous six segments, we further divided the study watershed into 12 separate model segments according to three predominant soil types.

Model setup

The study watershed modeled was divided into 21 sub-watersheds with a mean sub-watershed area of 459 km^2, each sub-watershed with one reach (Figure 1), according to DEM of the study area. Then, a number of model segments were created to represent the spatial heterogeneity of the studied watershed based on three segmentation approaches. Six meteorological data were assigned to each model segment by Thiessen polygons. And each model segment has the same input and parameter values. To fully reflect spatial variability of the underlying surface within the study watershed, 19 unique combinations of land use and soil type were obtained by the overlaying analysis of GIS. Therefore, each model segment was further sub-discretized into 19 classes of pervious lands (PERLNDs) and an IMPLND, and they are also referred to as hydrologic response units (HRUs). All geoprocessing operations were performed using the toolkits provided by BASINS (Diaz-Ramirez *et al.* 2013). In the following step, model segments based on three segmentation approaches were specified by the BASINS Model Segmentation tool, and then the HSPF model for Linyi Watershed was built through the BASINS Model Setup tool. For the HSPF modeling of the hydrological processes, the water storage in pervious land of the study watershed can be generalized as 5 parts vertically including interception storage, upper zone, lower zone, groundwater storage and deep percolation storage. All the hydrological processes, snow accumulation and melt, interception, evapotranspiration, surface detention, surface runoff, infiltration, shallow subsurface flow (interflow), baseflow and deep percolation were calculated for each HRU. The initial model parameters representing the land surface hydrological processes are system defaults and need to be calibrated (Table 1); while, initial model parameters describing channel hydraulic processes are obtained by combining DEM with Manning's equation that do not require calibration. HSPF model can simulate the main water balance components such as total runoff (surface flow, interflow, baseflow) and total actual evapotranspiration (ET) (interception ET, upper zone ET, lower zone ET, baseflow ET, active groundwater ET) in daily, hourly, and minute timesteps. However, the objective of this study only focuses on performing long-term continuous streamflow simulations. Therefore, a daily time step was chosen for HSPF model simulation.

Model calibration and validation

Model calibration is a process of adjusting model parameters within a suitable range to achieve agreement between observed and simulated flows. While model validation is a process of evaluating the robustness of calibrated model parameters. Two calibration approaches are commonly used for HSPF calibration: the expert HSPF manual calibration tool known as the decision-support software Expert System for the Calibration of HSPF (HSPEXP) (Lumb *et al.* 1994) and automatic Parameter Estimation software (PEST) (Doherty 2004). Manual calibration is time consuming and tedious (Madsen 2000), and HSPEXP manual calibrations offer only parameter adjustment guidance to model users. While PEST is a model-independent parameter optimization program that uses the Gauss–Marquardt–Levenberg (GML) (Marquardt 1963) method for nonlinear parameter estimation. PEST automatic calibration is less subjective and faster, and more efficient than HSPEXP-assisted manual calibrations. Several studies have performed HSPF calibration with

Table 1 | List of adjusted parameters for calibration of HSPF model based on 12 segments

Parameters	Definition	Unit	Initial value	Calibrated value
LZSN[a]	Lower zone nominal soil storage	inches	6.5	2.000
LZSN[b]			6.5	4.819
LZSN[c]			6.5	2.000
INFILT[a]	Index to infiltration capacity	in/hr	0.16	0.315
INFILT[b]			0.16	4.835
INFILT[c]			0.16	0.140
AGWRC[a]	Groundwater recession rate	1/d	0.98	0.998
AGWRC[b]			0.98	0.973
AGWRC[c]			0.98	0.999
DEEPFR	Fraction of groundwater inflow to deep recharge	none	0.1	8.69E-02
BASETP	Fraction of remaining ET from baseflow	none	0.02	2.11E-02
AGWETP	Fraction of remaining ET from active groundwater	none	0	0.200
UZSN	Upper zone nominal soil storage	inches	1.128	1.807
INFTW	Interflow inflow parameter	none	0.75	10.000
IRC	Inflow recession constant	1/d	0.5	0.381

[a]Forest land.
[b]Wetlands/water.
[c]Agricultural land.

PEST (Doherty & Johnston 2003; Kim *et al*. 2007; Ryu 2009). In this paper, PEST aided by HSPEXP is applied to calibrate the HSPF model. The nine most sensitive parameters (Table 2) for hydrological simulation in HSPF (Donigian *et al*. 1984) were adjusted to acquire a match between simulated and observed flows during the period of calibration (from January 1, 2001 to December 31, 2004). These parameters were selected through a synthesis of our experiences from past HSPF model application in the Linyi watershed, literature survey (Mohamoud *et al*. 2012; Diaz-Ramirez *et al*. 2013; Xie & Lian 2013), and results of model sensitivity analysis. The parameter correlation coefficient matrix of the nine parameters (Table 2) came from PEST and was used to analyse the uncertainty of model parameters. The results confirmed that these nine parameters used to the calibrate model would not cause any problem of over-parameterization due to the lower correlation coefficients.

Model evaluation

The agreement between the observed and simulated flows was quantitatively evaluated using the criteria coefficient of determine (R^2), Nash–Sutcliffe coefficient (*NSE*), and relative error (*RE*), which are defined as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(y^i_{\mathrm{sim}} - \overline{y_{\mathrm{sim}}}\right)\cdot\left(y^i_{\mathrm{obs}} - \overline{y_{\mathrm{obs}}}\right)\right]^2}{\sum_{i=1}^{n}\left(y^i_{\mathrm{sim}} - \overline{y_{\mathrm{sim}}}\right)^2\cdot\sum_{i=1}^{n}\left(y^i_{\mathrm{obs}} - \overline{y_{\mathrm{obs}}}\right)^2} \tag{1}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(y^i_{\mathrm{obs}} - y^i_{\mathrm{sim}}\right)^2}{\sum_{i=1}^{n}\left(y^i_{\mathrm{obs}} - \overline{y_{\mathrm{obs}}}\right)^2} \tag{2}$$

$$RE = \frac{\sum_{i=1}^{n}\left(y^i_{\mathrm{sim}} - y^i_{\mathrm{obs}}\right)^2}{\sum_{i=1}^{n} y^i_{\mathrm{obs}}} \times 100\% \tag{3}$$

where is daily simulated flow in i day, y^i_{sim} is daily observed flow in i day, $\overline{y_{\mathrm{obs}}}$ is average observed flow of the simulated period, $\overline{y_{\mathrm{sim}}}$ is average simulated flow of the simulated period, and i is day.

RESULTS AND DISCUSSION

Impact of three segmentation approaches on model performance

To assess the impacts of segmentation approaches on model performance, the dynamic sensitivity method was employed in this study. It shows that the best model efficiency and corresponding optimized parameters can be obtained through model calibration with different hydrological segmentation approaches, and comparing their R^2, *NSE*, and *RE* to select the best segmentation approach with optimum model efficiency. The static sensitivity method was adopted to explore model sensitivity to the segmentation approach

Table 2 | Correlation coefficient matrix of nine parameters

	LZSN	INFILT	AGWRC	DEEPFR	BASETP	AGWETP	UZSN	INTFW	IRC
LZSN	1.000	– 0.333	0.395	– 0.099	0.109	– 0.305	– 0.120	– 0.041	– 0.074
INFILT	0.333	1.000	– 0.476	0.112	– 0.031	0.576	0.042	0.164	– 0.193
AGWRC	0.395	– 0.476	1.000	– 0.197	0.035	– 0.195	– 0.168	– 0.030	0.132
DEEPFR	– 0.099	0.112	– 0.197	1.000	– 0.180	– 0.404	– 0.074	0.009	– 0.008
BASETP	0.109	– 0.031	0.035	– 0.180	1.000	– 0.144	– 0.558	– 0.281	0.210
AGWETP	– 0.305	0.576	– 0.195	– 0.404	– 0.144	1.000	0.051	0.106	– 0.196
UZSN	– 0.120	0.042	– 0.168	– 0.074	– 0.558	0.051	1.000	0.401	– 0.116
INTFW	– 0.041	0.164	– 0.030	0.009	– 0.281	0.106	0.401	1.000	– 0.381
IRC	– 0.074	– 0.193	0.132	– 0.008	0.210	– 0.196	– 0.116	– 0.381	1.000

by first obtaining a calibration considered to be optimal and then leaving it unchanged (Oudin *et al*. 2006). Model performance with different segmentation approaches and unchanged parameters is assessed by comparing obtained R^2, *NSE,* and *RE*.

The simulated mean monthly and daily streamflow for three segmentation approaches during the calibration and validation period are plotted with the observed values in Figures 2 and 3. All these hydrographs show acceptable agreement and tendency, although obvious errors of peak flows appeared between simulations and observations. In Figure 2, we can see that the modeling with six and 12 segments could better replicate the entire shape of hydrographs throughout the calibration period than 21 segments. However, during the validation period, compared with the observed flows, the modeling with six segments has lower efficiency than 12 segments, especially the flow peaks and recessions, but it still performed better than 21 segments. These observations are confirmed by the evaluation criteria in Table 3, which displayed that the modeling with three segmentation approaches was excellent to simulate the mean monthly streamflow due to excellent agreement, correlation and RE ($NSE = 0.95$, $R^2 = 0.95$, $RE = -5\%$ for 12 segments; $NSE = 0.93$, $R^2 = 0.93$, $RE = -9\%$ for six segments; $NSE = 0.92$, $R^2 = 0.92$, $RE = -15\%$ for

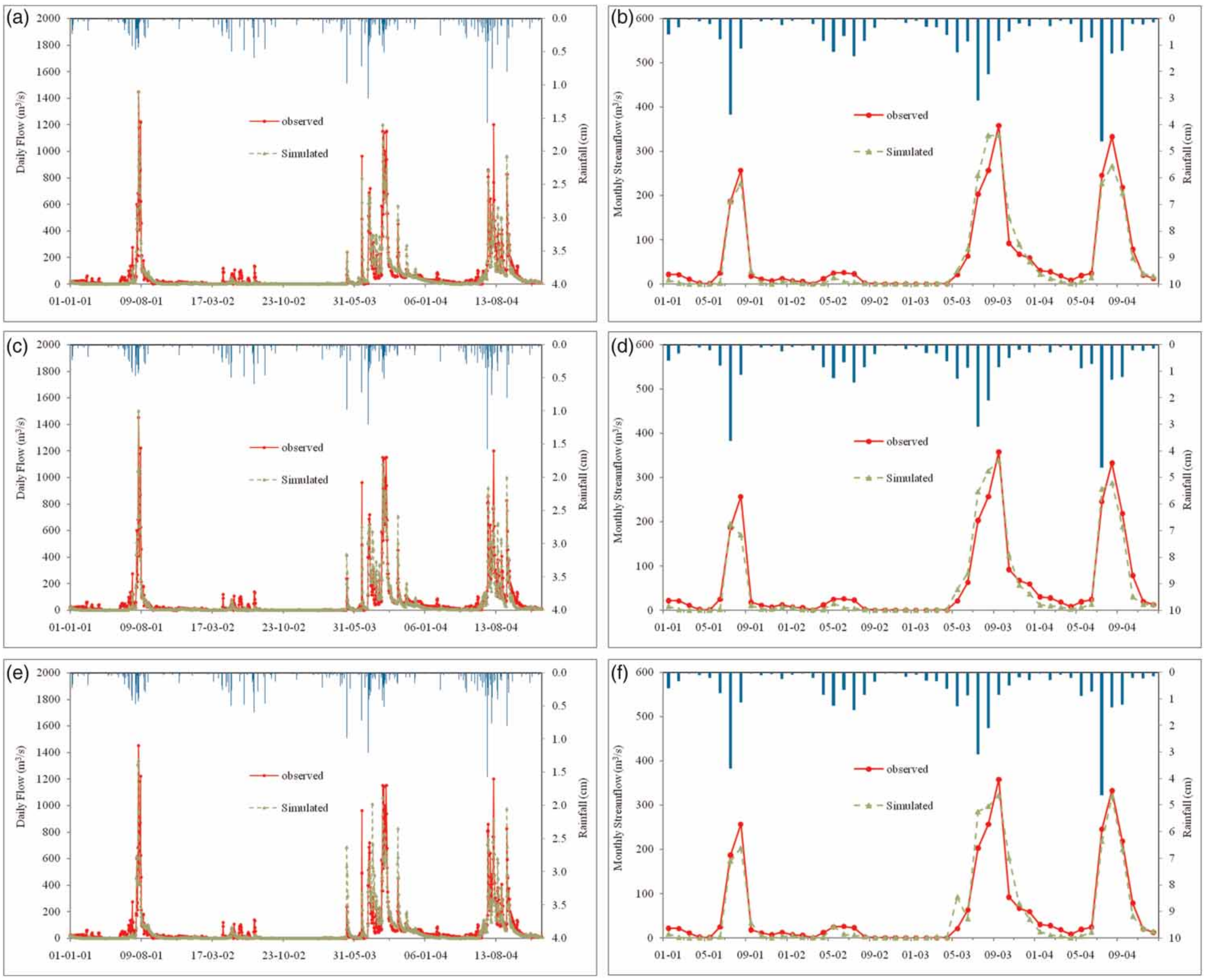

Figure 2 | Daily and monthly streamflow calibration results with 12 segments (a, b), six segments (c, d), and 21 segments (e, f) during 2001 to 2004.

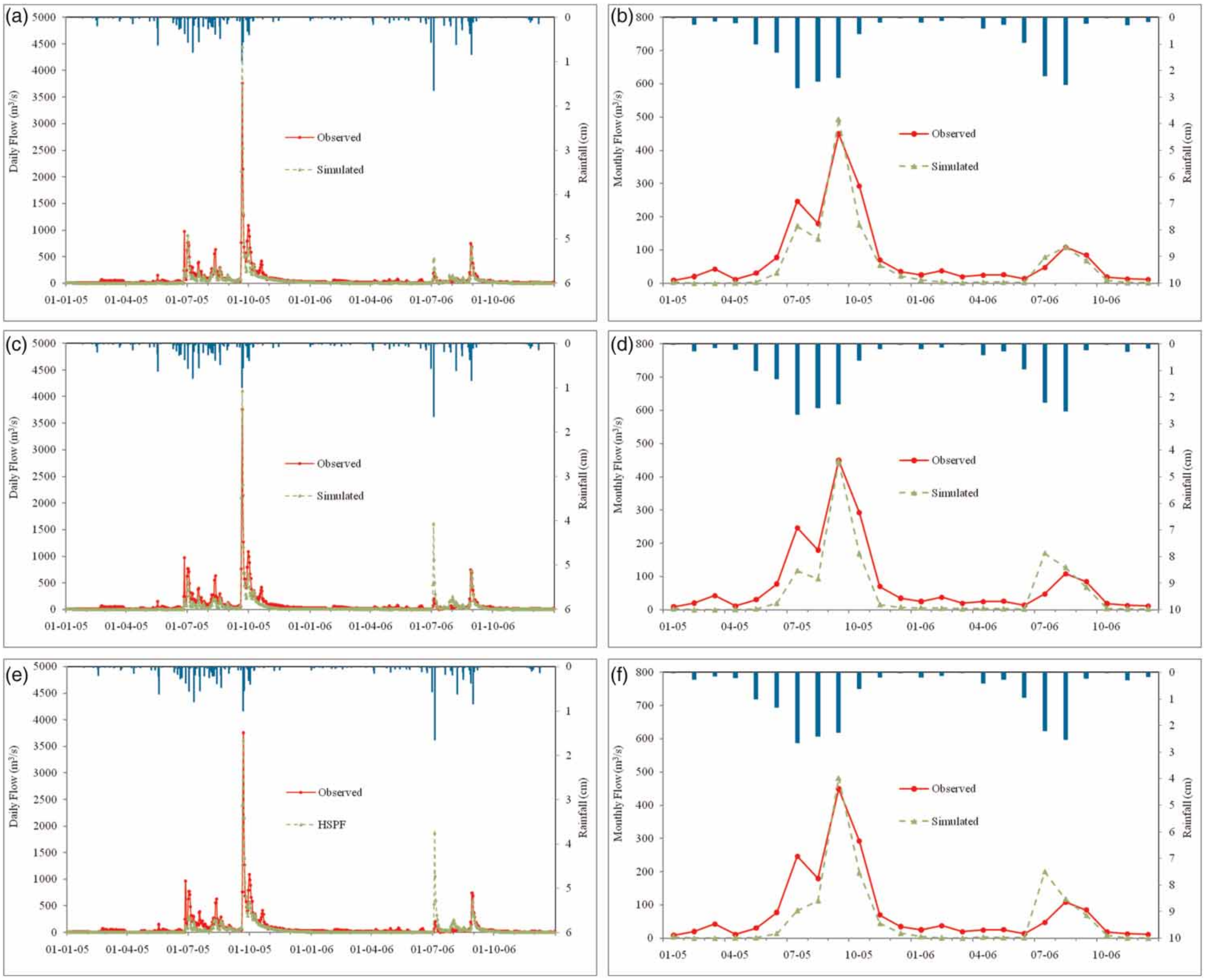

Figure 3 | Daily and monthly streamflow validation results with 12 segments (a, b), six segments (c, d), and 21 segments (e, f) during 2005 to 2006.

Table 3 | Model evaluated statistics for three segmentation approaches during the calibration period (2001–2004) and the validation period (2005–2006)

Segmentation approach		**12**			**6**			**21**		
Evaluation criteria		R^2	**NSE**	**RE (%)**	R^2	**NSE**	**RE (%)**	R^2	**NSE**	**RE (%)**
Calibration Period	Daily	0.85	0.85	– 0.05	0.81	0.80	– 0.08	0.74	0.73	– 0.15
	Monthly	0.95	0.95	– 0.05	0.93	0.93	– 0.09	0.92	0.92	– 0.15
Validation Period	Daily	0.86	0.82	– 0.17	0.77	0.72	– 0.22	0.70	0.64	– 0.28
	Monthly	0.92	0.88	– 0.17	0.81	0.75	– 0.22	0.70	0.72	– 0.28

21 segments) during the calibration period. However, for daily streamflow simulation throughout the calibration period, the modeling with 12 segments showed best results for NSE, R^2, and RE of 0.85, 0.85, and -5%, respectively. The NSE (0.80), R^2 (0.81), and RE (-8%) for the modeling with six segments was better than for the modeling with 21 segments ($NSE = 0.73$, $R^2 = 0.74$, $RE = -15\%$). During the validation period, the evaluation criteria values estimated by the modeling with

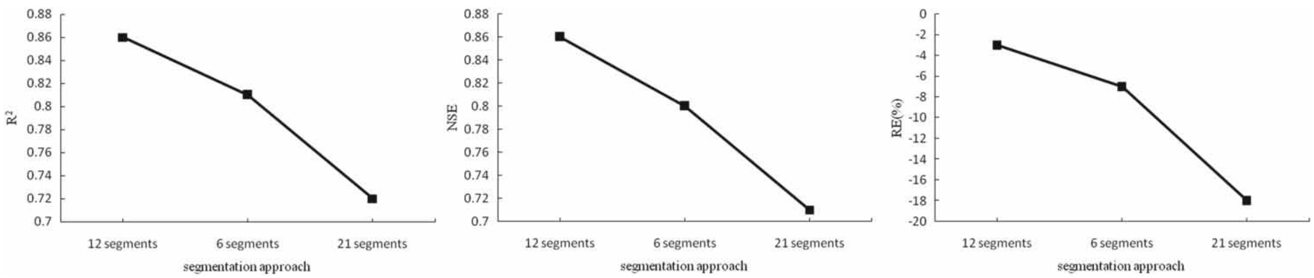

Figure 4 | Model performance in daily streamflow analysis using 12 segments, six segments, and 21 segments.

6 ($NSE = 0.72$, $R^2 = 0.77$, $RE = -22\%$ for daily streamflow simulation; $NSE = 0.75$, $R^2 = 0.81$, $RE = -22\%$ for mean monthly streamflow simulation) and 21 segments ($NSE = 0.64$, $R^2 = 0.70$, $RE = -28\%$ for daily streamflow simulation; $NSE = 0.72$, $R^2 = 0.78$, $RE = -28\%$ for mean monthly streamflow simulation) decreased remarkably, while the results ($NSE = 0.82$, $R^2 = 0.86$ for daily streamflow simulation; $NSE = 0.88$, $R^2 = 0.92$ for mean monthly streamflow simulation) from the modeling with 12 segments had no obvious difference with the exception of RE ($RE = -17\%$ for mean monthly and daily streamflow simulation). These conclusions revealed that the modeling with 12 segments had the best simulation efficiency and their corresponding estimated parameters (Table 1) had a certain representation in studied watershed.

Therefore, the optimal parameter values listed in Table 1 were applied to the modeling with three segmentation approaches during the simulation period (2001–2006). Based on these parameter values, the model simulated daily streamflow using three segmentation approaches and we defined the modeling with six segments, 12 segments and 21 segments as S6, S12, and S21, respectively. Figure 4 illustrates comparisons of model performance in daily streamflow simulations for S6, S12, and S21. In addition, it also shows how model performance is influenced by three segmentation approaches. There was an insignificant decrease in R^2 and NSE when the watershed segmentation varied from 12 ($NSE = 0.86$, $R^2 = 0.86$) to 6 ($NSE = 0.80$, $R^2 = 0.81$). However, NSE and R^2 reduced markedly from 0.80 and 0.81 to 0.71 and 0.72 when segment was changed from 6 to 21. Similarly, the RE of daily streamflow calculated for segmentation 12, 6 and 21 decreased sharply which were -3, -7 and -18%, respectively. Paired-samples T test, which can compare different analysis results from different segmentation approaches and assess whether the difference is significant between these segmentation approaches, is introduced to further reflect the influences of segmentation approach on HSPF model performance. Here, we considered S12 as the optimum, and comparison between S6 and S12, S21 and S12 are shown in Table 4. The value of Sig (2 tailed) was less than 0.05, which revealed that there was significant difference between the compared segmentation approaches. Table 4 indicates that the difference between S6 and S12 is not significant, but the difference between S21 and S12 is obvious and marked. All of these indicate model performance is affected by the segmentation approach. Especially, when model segmentation considers meteorological station and physical characteristics, it could obtain the optimal model performance. Because model segments in HSPF are subareas of a watershed with the same input data and uniform parameters that individually conduct hydrological simulation, the hydrological calculation result of each segment based on meteorological station and physical characteristics of underlying surface is able to reproduce the real situation more accurately.

Impact of three segmentation approaches on model parameters

In the previous section, we made comparisons to display and interpret the results on impacts of different segmentation approach on model performances. In this section, the

Table 4 | Paired-samples T test of daily streamflow simulation of S12, S6 and S21 segment series

Segment series	Std. deviation	Std. error of mean	Sig (2-tail)
S12–S6	0.010	0.006	0.13
S12–S21	0.006	0.003	0.01

dynamic sensitivity method was employed to assess the impacts of segmentation approach on model parameters. We analyzed the evolution of model parameter values when the model was calibrated with three segmentation approaches to obtain optimal parameters. Figure 5 shows the evolutions of parameters set for the HSPF model. The

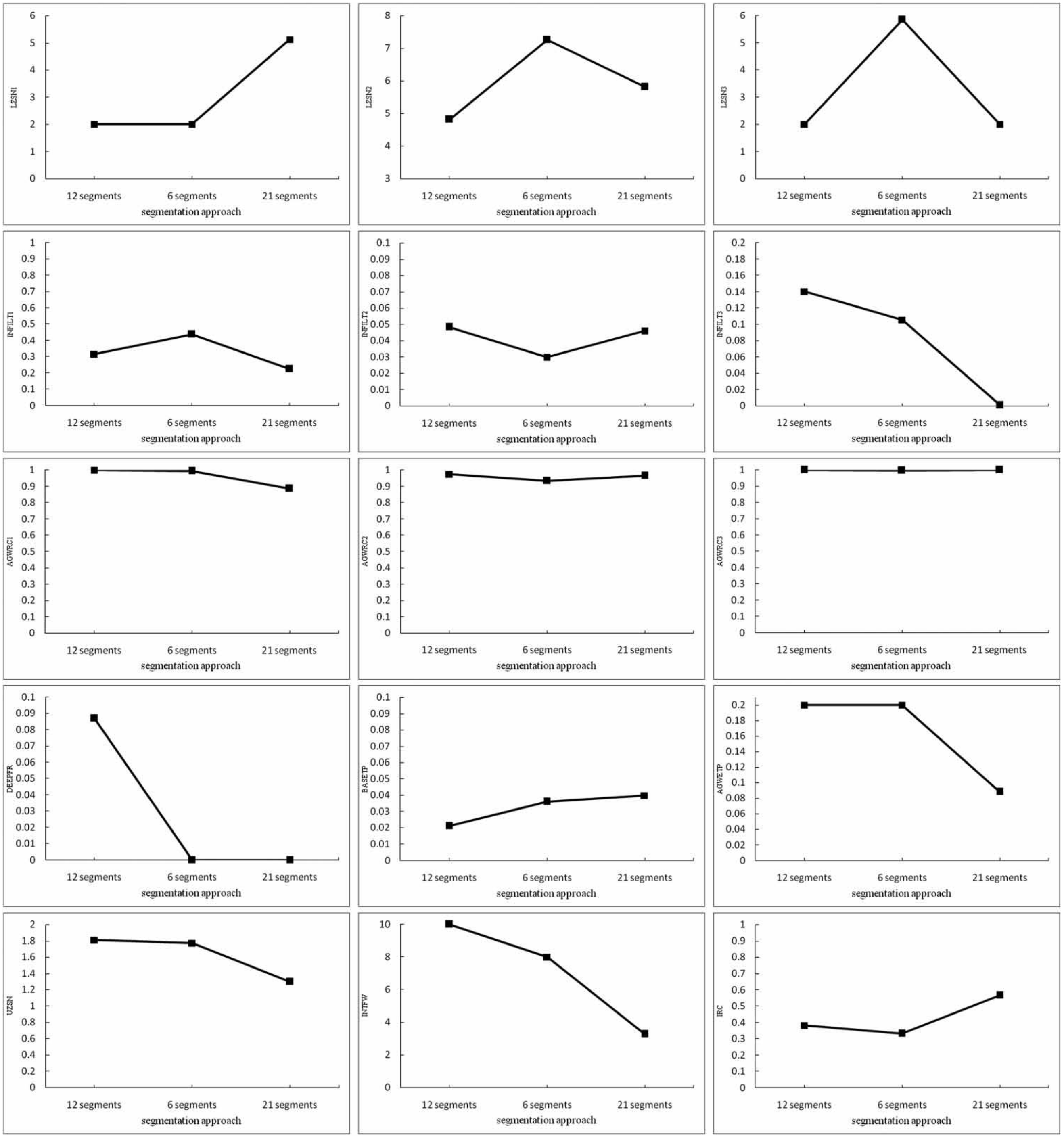

Figure 5 | Comparison of model parameters in daily streamflow analysis using 12 segments, six segments, and 21 segments.

range of these parameters was analyzed by statistical analysis method and shown in Table 5. From Figure 5 and Table 5, it can be seen some parameter values change markedly whenever a different segmentation method is used, whereas other parameter values are very stable. Among these parameters, INTFW, lower zone nominal storage (LZSN), and upper zone nominal storage (UZSN) were sensitive to the segmentation approach, especially parameter INTFW, whose change was more apparent than others. On the contrary, parameters AGWRC and BASETP were insensitive to changes in the segmentation approach.

In HSPF, the water movement into the soil and through it is modeled by dividing the soil into an upper, lower and ground water zones (Albek *et al*. 2004). The water content in upper and lower zones is represented by nominal soil storage parameters referred to as USZN and LZSN. LZSN is related to rainfall and soil characteristics in the watershed. USZN is related to land surface characteristics, topography, and LZSN. This parameter can change the amount of surface runoff, just as increasing USZN can decrease the amount of surface runoff. INTFW is the coefficient that determines the amount of water which enters the ground from surface detention storage and becomes interflow, rather than direct surface runoff and upper zone storage. This process has an influence on the variation of runoff by controlling water allocation between overland flow and interflow. Hence, decreasing the value of INTFW can decrease interflow, and increase overland flow to maintain water balance. AGWRC is the groundwater recession rate.

Table 5 | Statistical analysis for the parameters adjusted by three segmentation approaches

Parameters	Mean	Range	Variance	Std. deviation	Std. error of mean
LZSN[a]	3.0395	3.1185	3.2417	1.8005	1.0395
LZSN[b]	5.9633	2.4373	1.5016	1.2254	0.7075
LZSN[c]	3.2851	3.8553	4.9544	2.2259	1.2851
INFILT[a]	0.3261	0.2134	0.0115	0.1071	0.0618
INFILT[b]	0.0414	0.0185	0.0001	0.0101	0.0058
INFILT[c]	0.0821	0.1390	0.0052	0.0723	0.0418
AGWRC[a]	0.9597	0.1122	0.0040	0.0636	0.0367
AGWRC[b]	0.9577	0.0395	0.0004	0.0211	0.0122
AGWRC[c]	0.9980	0.0030	0.0000	0.0017	0.0010
DEEPFR	0.0290	0.0869	0.0025	0.0502	0.0290
BASETP	0.0323	0.0186	0.0001	0.0099	0.0057
AGWETP	0.1628	0.1117	0.0042	0.0645	0.0372
UZSN	1.6255	0.5088	0.0806	0.2839	0.1639
INTFW	7.0916	6.7226	11.9138	3.4516	1.9928
IRC	0.4270	0.2338	0.0152	0.1234	0.0712

[a]Forest land.
[b]Wetlands/water.
[c]Agricultural land.

We used three segmentation approaches to compare model efficiency and concluded that the modeling with 12 segments had the best simulation efficiency and the optimal parameters in the above section. Herein, we further discussed the variation trend and range of these sensitive parameters. Figure 5 shows that many parameters changed markedly in the 21 segments, which may result from the lowest model efficiency. And the modeling with 21 segments estimated lower values for UZSN, INTFW, AGWETP and a higher value for IRC compared with two other segmentation approaches that both considered meteorological station. It may imply that these parameters mentioned above were closely related to the rainfall distribution of each segment. While for the DEEPE parameter, modeling with 12 segments estimated higher values than two other segmentation approaches. Because DEEPE (the fraction of groundwater inflow that will enter deep groundwater) is proportional to the infiltration capacity of soil, and the infiltration capacity of soil is affected by physical characteristics of land, decreasing the value of DEEPE can increase the amount of runoff generation. Similarly, the same variation trends of UZSN and INTFW are showing in Figure 5. From all these analyses, it is concluded that modeling with six and 21 segments had higher peak flows (Figures 1 and 2) and larger RE (Table 3). When model segments change in different segmentation approaches, the rainfall intensity and soil moisture consequently change in the hydrological calculation of each segment, thus, the water balance of each model segment will redistribute. This may explain why LZSN, UZSN, and INTFW representing land surface hydrological processes of the PERLND module (Linsley *et al*. 1986) are the most sensitive parameters. In summary, INTFW, LZSN, and UZSN were most affected by the model segmentation approach, while AGWRC changed indistinctly across the different segmentation approaches. In addition, UZSN, INTFW, and DEEPE had the same variation tendency whenever the segmentation approach changed, all of them could influence the amount of surface runoff. The variations of three parameters (INTFW, LZSN, and UZSN) reflected that increasing the value of three parameters could improve model efficiency and reduce the RE.

CONCLUSIONS

The objective of this study is to assess whether the model segmentation approach can affect both the model performance and the parameter estimation. For three different model segmentation approaches including model segments based on differences in: (1) sub-watershed, (2) meteorological station, and (3) physical characteristics, the dynamic sensitivity method was employed to obtain the best model performance and corresponding optimized parameters by comparative analysis of streamflow simulation. Results showed that modeling with 12 segments had the best simulation efficiency and the corresponding estimated parameters had a certain representation in the studied watershed. Then the static sensitive method and Paired-samples T test were adopted to investigate whether the difference between these optimal parameter values derived from three segmentation approaches was significant and assess the influences of segmentation approach on HSPF model performance. The results revealed that the difference between six segments and 12 segments was not significant, but the difference between 21 segments and 12 segments was obvious and marked. Therefore, HSPF model performance was affected significantly by the segmentation approach, especially by model segmentation considering meteorological station or not. Finally, the dynamic sensitivity method and statistical method were employed to evaluate the impacts of segmentation approach on model parameters. The results showed that parameter INTFW, LZSN, and UZSN were most affected by the model segmentation approach, while AGWRC changed indistinctly in different segmentation approaches. Parameters UZSN, INTFW, and DEEPE had the same variation tendency whenever the segmentation approach changed, all of them could influence the amount of surface runoff. The variations of three parameters (INTFW, LZSN, and UZSN) reflected that increasing the value of three parameters could improve model efficiency and reduce the RE.

ACKNOWLEDGEMENTS

This work was supported by the National Basic Research Program of China (973, Grant No. 2010CB951404) and the National Natural Science Foundation of China (No. 41175088 and No. 40971024).

REFERENCES

Albek, M., Ögütveren, ü. B. & Albek, E. 2004 Hydrological modeling of Seydi Suyu watershed (Turkey) with HSPF. *Journal of Hydrology* **285**, 260–271.

Andreassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. & Lavabre, J. 2001 Impact of imperfect rainfall knowledge on the efficiency of watershed models. *Journal of Hydrology* **250**, 206–223.

Bastola, S. & Murphy, C. 2013 Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data. *Hydrology Research* **44**, 484–494.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R. & Voksø, A. 2003 Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrology Earth System and Science* **7**, 304–316.

Bicknell, B. R., Imhoff, J. C., Kittle Jr., J. L., Jobes, T. H. & Donigian, A. S. 2001 *Hydrological Simulation Program – FORTRAN (HSPF). User's Manual for Release 12*. U.S. Environmental Protection Agency, Athens, GA.

Bosch, D. D., Sheridan, J. M., Batten, H. L. & Arnold, J. G. 2004 Evaluation of the SWAT model on a coastal plain agricultural watershed. *Transactions ASAE.* **47**, 1493–1506.

Chaplot, V. 2005 Impact of DEM mesh size and soil map scale on SWAT runoff, sediment, and NO3-N loads predictions. *Journal of Hydrology* **312**, 207–222.

Chaplot, V., Saleh, A. & Jaynes, D. B. 2005 Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO_3-N loads at the watershed level. *Journal of Hydrology* **312**, 223–234.

Chaubey, I., Cotter, A. S., Costello, T. A. & Soerens, T. S. 2005 Effect of DEM data resolution on SWAT output uncertainty. *Hydrological Processes* **19**, 621–628.

Cho, J., Bosch, D., Lowrance, R., Strickland, T. & Vellidis, G. 2009 Effect of spatial distribution of rainfall on temporal and spatial uncertainty of SWAT output. *American Society of Agricultural and Biological Engineers* **52**, 1545–1555.

Diaz-Ramirez, J. N., McAnally, W. H. & Martin, J. L. 2011 Analysis of hydrological processes applying the HSPF model in selected watersheds in Alabama, Mississippi, and Puerto Rico. *Applied Engineering in Agriculture* **27**, 937–954.

Diaz-Ramirez, J. N., Johnson, B. E., McAnally, W. H., Martin, J. L., Alarcon, V. J. & Camacho, R. A. 2013 Estimation and propagation of parameter uncertainty in lumped hydrological models: a case study of HSPF model applied to Luxapallila Creek watershed in southeast USA. *Journal of Hydrogeology & Hydrologic Engineering* **2**, 1–9.

Doherty, J. 2004 *PEST: Model-Independent Parameter Estimation*, User manual: 5th edn., Watermark Numer. Comput., Brisbane, Queensland, Australia.

Doherty, J. & Johnston, J. M. 2003 Methodologies for calibration and predictive analysis of a watershed model. *Journal American Water Resource Associate* **39**, 251–265.

Donigian, A. S., Imhoff, J. C., Bicknell, B. R. & Kittle, J. I. Application guide for hydrological simulation program-FORTRAN (HSPF). EPA-600/3-84-065. USEPA, Athens, GA.

Donigian Jr, A. S., Bicknell, B. R. & Imhoff Jr, J. C. 1995 Chapter 12: Hydrological simulation program: FORTRAN (HSPF). In: *Computer Models of Watershed Hydrology* (V. P. Singh, ed.). Water Resources Publications, CO, USA, pp. 395–442.

Donnelly, C., Rosberg, J. & Isberg, K. 2013 A validation of river routing networks for catchment modelling from small to large scales. *Hydrology Research* **44**, 917–925.

Du, J. K., Xie, H., Hu, Y. J., Xu, Y. P. & Xu, C. Y. 2009 Development and testing of a new storm runoff routing approach based on time variant spatially distributed travel time method. *Journal of Hydrology* **369**, 44–54.

Gong, L., Halldin, S. & Xu, C. Y. 2011 Large-scale runoff generation – parsimonious parameterisation using high-resolution topography. *Hydrology and Earth System Sciences* **15**, 2481–2494.

Gong, L., Widén-Nilsson, E., Halldin, S. & Xu, C. Y. 2009 Large-scale runoff routing with an aggregated network-response function. *Journal of Hydrology* **368**, 237–250.

Haverkamp, S., Fohrer, N. & Frede, H. G. 2005 Assessment of the effect of land use patterns on hydrologic landscape functions: a comprehensive GIS-based tool to minimize model uncertainty resulting from spatial aggregation. *Hydrological Processes* **19**, 715–727.

He, M. X. & Hogue, T. S. 2012 Integrating hydrologic modeling and land use projections for evaluation of hydrologic response and regional water supply impacts in semi-arid environments. *Environmental Earth Sciences* **65**, 1671–1685.

Hughes, D. A., Kapangaziwiri, E. & Tanner, J. 2013 Spatial scale effects on model parameter estimation and predictive uncertainty in ungauged basins. *Hydrology Research* **44**, 441–453.

Kim, S. M., Benham, B. L., Brannan, K. M., Zeckoski, R. W. & Doherty, J. 2007 Comparison of hydrologic calibration of HSPF using automatic and manual methods. *Water Resources Research* **43**, W01402.

Krajewski, W. F. & Smith, J. 2002 Radar hydrology: rainfall estimation. *Advances in Water Resources* **25**, 1387–1394.

Kumar, R., Samaniego, L. & Attinger, S. 2010 The effect of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. *Journal of Hydrology* **392**, 54–69.

Li, H., Beldring, S. & Xu, C. Y. 2014 Implementation and testing of routing algorithms in the distributed HBV model for mountainous catchments. *Hydrology Research*. In press, doi: 10.2166/nh.2013.009.

Linsley, R. K., Kohler, M. A. & Paulhus, J. L. H. 1986 *Hydrology for Engineers*. McGraw-Hill Publishing, NY.

Littlewood, I. G. & Croke, B. F. W. 2013 Effects of data time-step on the accuracy of calibrated rainfall-streamflow model parameters: practical aspects of uncertainty reduction. *Hydrology Research* **44**, 430–440.

Lumb, A. M., McCammon, R. B. & Kittle, J. L. 1994 Users manual for an EXPert System (HSPEXP) for calibration on the Hydrological Simulation Program –Fortran. U.S. Geol. Surv. Water Resources Investigations Report 94-4168, U.S. Geological Survey, Reston, VA.

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

Marquardt, D. W. 1963 An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics* **11**, 431–441.

Mohamoud, Y. M. & Prieto, L. M. 2012 Effect of temporal and spatial rainfall resolution on HSPF predictive performance and parameter estimation. *Journal of Hydrologic Engineering* **17**, 377–388.

Mohamoud, M. Y., ASCE, M. & Prieto, M. L. 2012 Effect of temporal and spatial rainfall resolution on HSPF predictive performance and parameter estimation. *Journal of Hydrological Engineering* **17**, 377–388.

Moon, J., Srinivasan, R. & Jacobs, J. H. 2004 Stream flow estimation using spatially distributed rainfall in the Trinity River basin, Texas. *Transactions ASAE* **47**, 1445–1451.

Oudin, L., Perrin, C., Mathevet, T., Andreassian, V. & Michel, C. 2006 Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology* **320**, 62–83.

Pechlivanidis, I. G., Mclntyre, N. R. & Wheater, H. S. 2010 Calibration of the semi-distributed PDM rainfall-runoff model in the Upper Lee catchment, UK. *Journal of Hydrology* **386**, 198–209.

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

Ryu, J. H. 2009 Application of hspf to the distributed model intercomparison project: Case study. *Journal of Hydrology Engineering* **14**, 847–857.

Singh, V. P. 1997 *Kinematic Wave Modeling in Water Resources*. John Wiley & Sons Inc., Publishing, New York.

Sun, X., Mein, R. G., Keenan, T. D. & Elliott, J. F. 2000 Flood estimation using radar and raingauge data. *Journal of Hydrology* **239**, 4–18.

Wang, X. & Melesse, A. M. 2006 Effects of STATSGO and SSURGO as inputs on SWAT model's snowmelt simulation. *Journal of American Water Resources Association* **42**, 1217–1236.

Xie, H. & Lian, Y. Q. 2013 Uncertainty-based evaluation and comparison of SWAT and HSPF applications to the Illinois River Basin. *Journal of Hydrology* **481**, 119–131.

Xu, Y. D., Fu, B. J., He, C. S. & Gao, G. Y. 2012 Watershed discretization based on multiple factors and its application in the Chinese Loess Plateau. *Hydrology and Earth System Sciences* **16**, 59–68.

Xu, H. L., Xu, C. Y., Chen, H., Zhang, Z. X. & Li, L. 2013 Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *Journal of Hydrology* **505**, 1–12.

First received 11 November 2013; accepted in revised form 26 February 2014. Available online 8 April 2014