# The impact of digital elevation model and land use spatial information on Hydrologic Simulation Program-FORTRAN – predicted stream flow and sediment uncertainty

Huiliang Wang, Xuyong Li, Wenzan Li and Xinzhong Du

## ABSTRACT

The Hydrologic Simulation Program-FORTRAN (HSPF) model is widely used to develop management strategies for water resources. The spatial resolution of the input data used to parameterize the HSPF model may lead to uncertainty in model outputs. In this study, we evaluated the impact of the spatial resolution of the digital elevation model (DEM) and land use data on uncertainty in HSPF-predicted flow and sediment. The resolution of DEMs can affect stream length, watershed area, and average slope, while the resolution of land use data can influence the distribution of land use information. Results showed that DEMs and land use maps with finer resolutions generated higher flow volumes and sediment loads. There was a non-linear relationship between changes in resolution of the DEM and land use data and changes in the uncertainty of predicted flow and sediment loads. Relative error was used to describe model uncertainty and the probability density function was used to estimate these uncertainties. The best-fit distributions of uncertainty in modeled flow and sediment related to DEM and land use data resolution were the generalized Pareto distribution and the Johnson SB distribution, respectively. The results of this study provide useful information for better understanding and estimating uncertainties in the HSPF model.

**Key words** | digital elevation models, Hydrologic Simulation Program-FORTRAN model, land use, spatial resolution, uncertainty

**Huiliang Wang**
**Xuyong Li** (corresponding author)
**Wenzan Li**
**Xinzhong Du**
State Key Laboratory of Urban and Regional Ecology,
Research Center for Eco-Environmental Sciences,
Chinese Academy of Sciences,
Beijing 100085,
China
E-mail: *xyli@rcees.ac.cn*

**Huiliang Wang**
**Wenzan Li**
**Xinzhong Du**
Graduate School,
Chinese Academy of Sciences,
Beijing 100049,
China

## INTRODUCTION

Watershed models have been developed to examine watershed-scale processes and to evaluate the hydrologic effect of various management scenarios (Beven 2001). The lack of data at appropriate spatial and temporal scales, particularly in China and some African countries, is a major limitation to the use of watershed models for predicting flow and sediment, and this has affected watershed management. The spatial information in the input data has been identified as a key issue in hydrological modeling (Wilson *et al.* 1996; Quiroga *et al.* 2013). Wagenet & Hutson (1996) pointed out that while the use of geographic information systems (GIS) has greatly improved the simulation of watershed processes, the scale at which GIS data, including soil surveys, digital elevation models (DEMs), and land use, should be collected and used is an issue that needs further study. Fine-scale DEMs and detailed or high-resolution land use and soil maps (e.g., 1:25,000) generate accurate estimations. However, researchers cannot always acquire such detailed information because of the high costs involved. It is therefore important to understand the effect of spatial data resolution on the accuracy of watershed models and to estimate the resolution-induced uncertainties. Recently, a new method was developed to estimate uncertainty related to the resolution of input data, using probability density functions (*pdfs*) and cumulative distribution functions (*CDFs*) (Patil *et al.* 2011; Patil & Deng 2012). The new method uses non-symmetric distributions of uncertainty in the prediction of water quality variables, which simplify

doi: 10.2166/hydro.2013.227

weather and water quality monitoring and consequently reduce the cost of flow and water quality monitoring.

The spatial data for watershed model include land use, soil, and DEMs for which the accuracy varies depending on the scale of the watershed and mapping methods. In a study using 71 watersheds and TOPMODEL, Wolock & Price (1994) found that increasing DEM coarseness from $30 \times 30$ m to $90 \times 90$ m increased the ratio of overland flow to total flow. Cho & Lee (2001) found that the Soil and Water Assessment Tool (SWAT) simulated flow was higher for a DEM with a resolution of $30 \times 30$ m than for a DEM with a $90 \times 90$ m resolution. Similar conclusions were drawn by Chaubey *et al.* (2005) and Chaplot (2005). Cotter *et al.* (2003) and Luzio *et al.* (2005) showed that coarse DEM and land use resolution not only reduced predicted flow but also reduced predicted sediment. Although these studies highlighted the impact of spatial input data resolution on simulated flow and sediment, they did not examine the effects of spatial data resolution on model output given the model structure and parameters. Furthermore, after the researchers quantified the uncertainties, only a few studied the estimated *pdfs* and *CDFs* of these uncertainties, information that is very important to the estimation of total uncertainties using the Bayesian approach and similar methods (Patil *et al.* 2011; Patil & Deng 2012). In this paper, we investigate the impact of the mesh size of the DEM and the land use map scale on simulated runoff and sediment derived using the HSPF model.

The overall aim of this study was to estimate the impact of spatial information (DEM and land use) on uncertainties in HSPF model-predicted stream flow and sediment. The particular objectives of this study were to quantify the uncertainties in simulation flow and sediment related to spatial input data resolution and to estimate these uncertainties using *pdf*s and *CDF*s.

## MATERIALS AND METHODS

### Study area and available data

The study area is a tributary of the Luan River, located in northeastern China, and it covers an area of about 2,200 km$^2$ (Figure 1). The watershed is a typical sub-basin of the Luan River basin, which behaves as an ecological

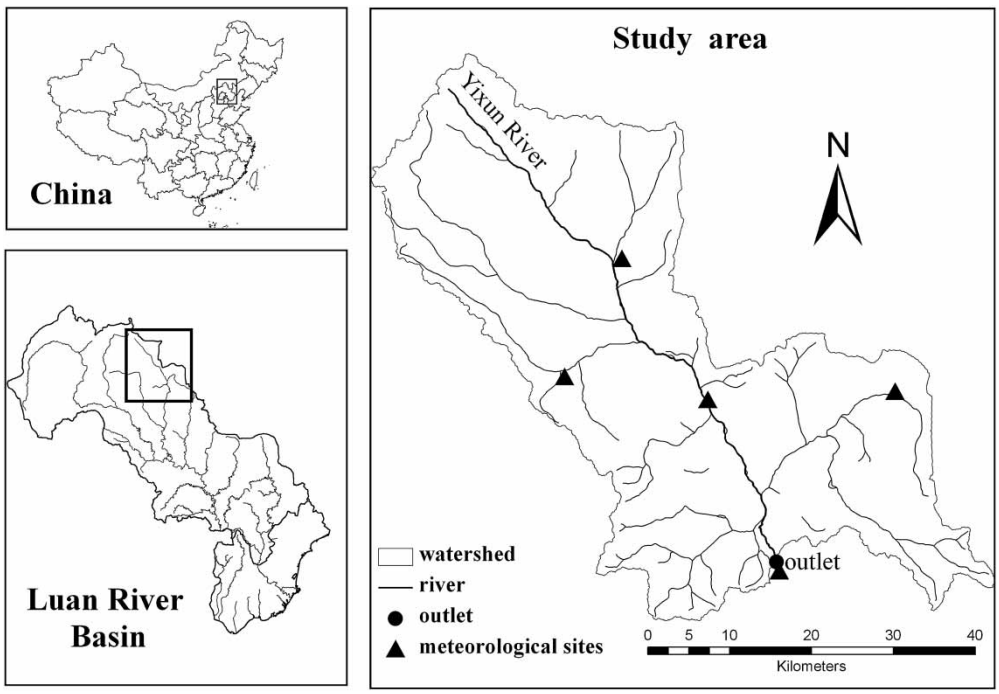


**Figure 1** | The upper Yixun River, Luan River basin, China.

barrier between Tianjin and Beijing and the entire North China zone. The elevation varies between 663 and 1,814 m and the average slope is about 24%. Most of the watershed (70% of the total area) is covered by forest, while agricultural land accounts for about 15% of the total area. The predominant soil in the watershed (about 75% of the total) is brown forest soil. The average annual precipitation and evaporation from the water surface are about 450 and 470 mm, respectively, and the average annual runoff and sediment are about $0.9 \times 10^8$ m$^3$ and $3.5 \times 10^8$ kg, respectively.

The DEM (1:24,000) for the watershed was downloaded from the International Scientific Data Service Platform, Chinese Academy of Sciences, and had a $30 \times 30$ m horizontal resolution. The land use data were obtained from Spot5 images acquired during 2007 at a resolution of 20 m (three images from July 27, 2007) by the Data Sharing Infrastructure of Earth System Science. The land use data were classified into 18 groups according to the standards regulated by the State Land Management Industry of China. Soil data (1:1,000,000) were downloaded from the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences. Hydrology and sediment data for the watershed outlet from 2007 to 2009 were collected from the Chengde Branch of the Hebei Provincial Survey Bureau of Hydrology & Water Resources (Table 1), and the meteorological data (precipitation, evaporation, temperature, wind speed, solar radiation, dew-point temperature, and cloud cover) were collected from the China Meteorological Data Sharing Service System.

## HSPF model description

The HSPF model is a distributed, continuous time watershed-scale model developed to simulate water quantity and quality at any point in a watershed (Bicknell *et al.* 2001). The HSPF model was originally developed from the Stanford Watershed Model and is an extension of several previously developed models: the Stanford Watershed Model (SWM), the Hydrologic Simulation Program (HSP) including HSP Quality, the Agricultural Runoff Management (ARM) model, and the Nonpoint Source Runoff (NPS) model (Johnson *et al.* 2003). In the HSPF model, a watershed is represented in terms of land segments and river reaches or reservoirs. Each land segment is referred to as a hydrologic response unit (HRU). Conceptually, runoff from the watershed moves laterally to a downslope segment or to a river reach or reservoir. The HSPF model uses simple storage-based (non-linear reservoir) equations for flow routing (Johnson *et al.* 2003). The equations consist of the spatially uniform and temporally variable continuity equation and a flow equation expressed in terms of channel (or plane) roughness and geometry, such as Manning's equation. Other processes simulated on the watershed include evapotranspiration (ET), interception, percolation, interflow, and groundwater movement. Channel routing is computed using storage routing or kinematic wave routing (Borah & Bera 2003).

Within HSPF, Module section SEDMNT simulates the erosion processes and the removal of sediment from a pervious land segment. The basic process in the simulation of sediment is soil detachment and the removal of detached sediment. In hillslope routing, rainfall splash detachment and wash off of the detached sediment are based on transport capacity as a function of water storage and outflow plus scour from flow using a power relationship with water storage and flow. In channel routing, non-cohesive (sand) sediment transport is modeled using a user-defined relationship with flow velocity, the Toffaleti method or the Colby method, and cohesive (silt, clay) sediment transport is based on critical shear stress and settling velocity (Borah & Bera 2003).

**Table 1** | Measured data (average annual) from upper Yinxun watershed

| | 2007 | 2008 | 2009 |
|---|---|---|---|
| Rainfall (mm) | 435 | 580 | 519 |
| Flow ($\times 10^8$ m$^3$) | 1.046 | 1.447 | 0.748 |
| Sediment (kg/m$^3$) | 1.04 | 4.46 | 1.91 |

## Model identification

The HSPF model parameters that affect flow and sediment in the upper Yixun River were calibrated using monthly measured hydrologic and sediment data for the period 2007–2009 (Table 1). The model was calibrated according to the standard procedure described by Donigian *et al.* (1999) and evaluated using the observed data for the watershed outlet. The objective function used in model

calibration was the Nash–Sutcliffe coefficient ($R^2_{NS}$), defined as follows:

$$R^2_{NS} = 1\frac{\sum_{t=1}^{n}\left(y^{sim}-y^{avg}\right)^2}{\sum_{t=1}^{n}\left(y^{sim}-y^{avg}\right)^2} \quad (1)$$

where $y^{sim}$ is the monthly simulated output, $y^{obs}$ is the monthly observed output, $y^{avg}$ is the average output of interest, and $i$ is month.

High-resolution spatial data (i.e., 30 × 30 m resolution DEM and 20 × 20 m resolution land use) were used to simulate the *base scenario*. Flow and sediment were simulated using 15 additional DEM resolutions (60 × 60 m, 90 × 90 m, 120 × 120 m, 150 × 150 m, 180 × 180 m, 210 × 210 m, 240 × 240 m, 270 × 270 m, 300 × 300 m, 330 × 330 m, 360 × 360 m, 390 × 390 m, 420 × 420 m, 450 × 450 m, 480 × 480 m), and 16 additional land use data resolutions (60 × 60 m, 120 × 120 m, 180 × 180 m, 240 × 240 m, 300 × 300 m, 360 × 360 m, 420 × 420 m, 480 × 480 m, 540 × 540 m, 600 × 600 m, 660 × 660 m, 720 × 720 m, 780 × 780 m, 840 × 840 m, 900 × 900 m, 960 × 960 m). Spatial data at coarser resolutions were obtained by resampling the raw data (DEM at 30 × 30 m resolution and land use data at 20 × 20 m resolution) using the nearest neighbor method.

## Model uncertainty due to DEM and land use resolution

Changes in the model performance related to variations in spatial scale were quantified using the relative error (RE) given as:

$$RE = \frac{y^{sim}-y^{base}}{y^{base}} \quad (2)$$

where $y^{base}$ is the output from the *base scenario* (DEM at 30 × 30 m resolution and land use data at 20 × 20 m resolution), and $y^{sim}$ is the simulated variable at different resolutions.

## Estimation of DEM and land use resolution-induced uncertainty

The *pdf* and the *CDF* can adequately describe model uncertainty (Andronova & Schlesinger 2001; Xiu & Karniadakis 2003; Diaz-Ramirez *et al*. 2007; Patil *et al*. 2011; Patil & Deng 2012). Several hydrological distributions (i.e., generalized extreme value (GEV), generalized Pareto, normal, log-logistic (3P), Johnson SB, and uniform) were used to fit the distribution of the RE. The optimum *pdf* was selected based

**Table 2** | Parameters of generalized extreme value (GEV), generalized Pareto, Johnson SB, log-logistic (3P), and uniform distribution[a]

| Distribution | *Pdf* | *CDF* |
|---|---|---|
| Generalized extreme value | $f(x)=\frac{1}{\sigma}\exp\left[-(1+kz)^{-1/k}\right](1+kz)^{1/k};\ k\neq 0$ | $F(x)=\exp\left(-(1+kz)^{-1/k}\right)^{b}$ |
| Generalized Pareto | $f(x)=\frac{1}{\sigma}\left(1+k\frac{(x-\mu)}{\sigma}\right)^{-1-1/k};\ k\neq 0$ | $F(x)=1-\left(1+k\frac{(x-\mu)}{\sigma}\right)^{-1/k};\ k\neq 0$ |
| Johnson SB | $f(x)=\frac{\delta}{\lambda\sqrt{2\pi}z(1-z)}\bullet \exp\left(-\frac{1}{2}\left(\gamma+\delta\ln\left(\frac{z}{1-z}\right)\right)^2\right)$ | $F(x)=\phi\exp\left(\gamma+\delta\ln\left(\frac{z}{1-z}\right)\right)^{c}$ |
| Log-logistic (3P) | $f(x)=\frac{\alpha}{\beta}\left(\frac{x-y}{\beta}\right)^{\alpha-1}\left(1+\left(\frac{x-y}{\beta}\right)^{\alpha}\right)^{-2}$ | $F(x)=\left(1+\left(\frac{\beta}{x-\gamma}\right)^{\alpha}\right)^{-1}$ |
| Uniform | $f(x)=\frac{1}{b-a}$ | $F(x)=\frac{x-a}{b-a}$ |

[a]$\mu$, $\xi$, and $\alpha$ are location parameters; $\lambda$, $\sigma$, and $\beta$ are scale parameters; and $\alpha_1$,$\alpha_2$ , $k$, $\delta$, and $\gamma$ are shape parameter; $a$ and $b$ are boundary parameters.
[b]Where, $z=x-\mu/\sigma$.
[c]Where, $z\equiv x-\zeta/\lambda$.

on the Kolmogorov–Smirnov (KS) statistic, which serves as a goodness-of-fit test and has been widely used in hydrological studies (e.g., Haan & Skaggs 2003). The KS statistic is based on the maximum vertical difference between the theoretical and the empirical *CDF* and the latter is given by:

$$F_n(x_1, x_2, x_3 \ldots x_n) = \frac{1}{n}N \tag{3}$$

where $(x_1, x_2, x_3 \ldots x_n)$ is a set of random samples and $N$ is the number of observations $\leq x$.

The KS statistic is given by:

$$\mathrm{KS} = \max_{1 \leq i \leq n}\left[F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i)\right] \tag{4}$$

The five distributions used in this study are shown in Table 2. The corresponding statistical parameters were estimated depending on the distribution fit (lowest KS value).

## RESULTS AND DISCUSSION

### Model performance

The potential of the model to predict observed flow and sediment data was assessed first. Results of flow calibration for the period 2007–2009 are shown in Figure 2(a). The simulated flow followed an annual pattern similar to the observed values and characterized by low volume in

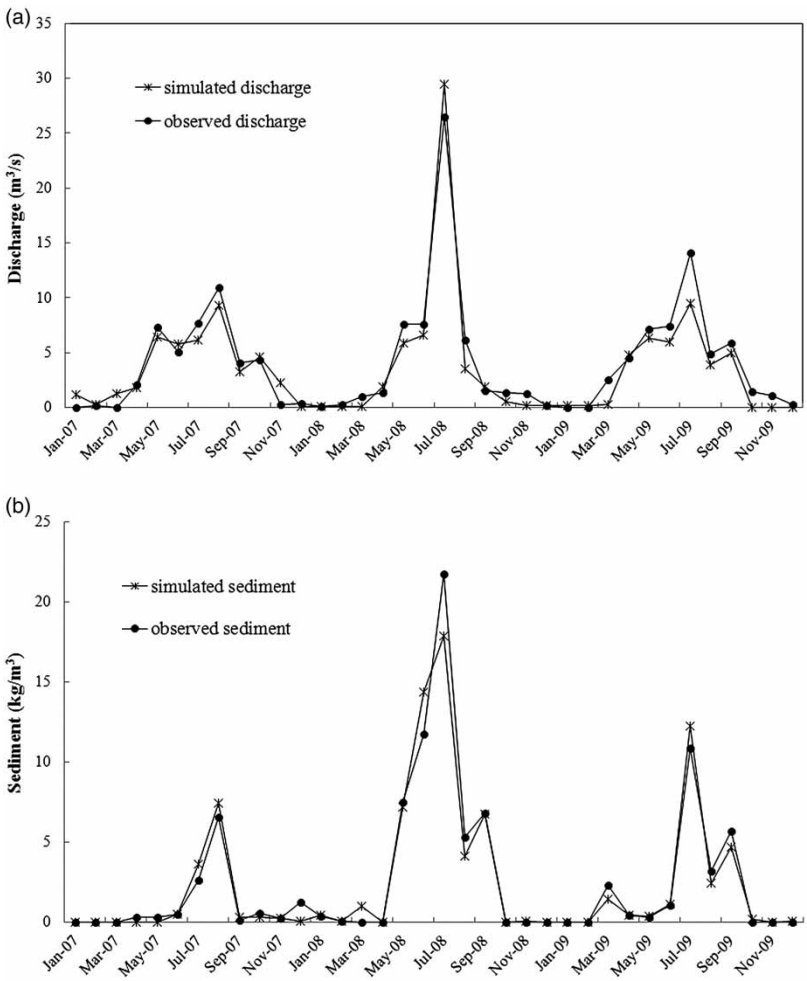


**Figure 2** | Calibration of (a) monthly flow and (b) monthly sediment simulation and observed data based on DEM 30 m and land use map 20 m.

winter and high volume in summer. The $R^2_{NS}$ of the monthly flow between simulated and observed values was 0.83, implying that the simulated flow fitted reasonably well with observed values. Figure 2(b) shows the observed and simulated time series for monthly sediment for the calibration period. The monthly sediment calibration yielded an $R^2_{NS}$ of 0.79, which showed that the simulated values were acceptable. Although the focus of this study was model calibration, the period between 2010 and 2011 was used for model validation. The $R^2_{NS}$ of monthly flow and sediment during the validation period were 0.78 and 0.72, respectively.

## Estimation of DEM resolution-induced uncertainty

Table 3 presents the effects of DEM resolution on the watershed delineation, stream network, average slope, and channel slope drop. As the DEM resolution decreased, total computed watershed area, average watershed slope, total stream length, and channel slope drop also decreased. However, against the trend, when the DEM resolution was 90 m, the value for the watershed area was larger than that simulated using the 30 m resolution DEM, and the value for the total stream length was larger than for the 60 m DEM and similar to the value for the 30 m DEM. The modeled stream network became consistently less accurate at coarser resolutions. These results were a direct consequence of the loss of topographic detail at the coarser DEM resolutions, resulting in the inability of the model to correctly predict the stream flow and sediment.

The data from the 30 m resolution DEM were used to simulate the *base scenario*. Flow and sediment were also simulated using 15 additional DEM resolutions ($60\times60$ m, $90\times90$ m, $120\times120$ m, $150\times150$ m, $180\times180$ m, $210\times210$ m, $240\times240$ m, $270\times270$ m, $300\times300$ m, $330\times330$ m, $360\times360$ m, $390\times390$ m, $420\times420$ m, $450\times450$ m, $480\times480$ m). The REs for the average simulated stream flow and sediment between each of the DEMs at the 15 resolutions listed above and the *base scenario* are shown in Table 4.

The average annual flow and sediment (2007–2009) predicted by the model decreased as the DEM resolution decreased. When the resolution of the DEM decreased from 30 m to 60 m, 90 m, 120 m, 240 m, and 480 m, the

**Table 3** | DEM resolution effects on watershed characteristics

| DEM resolution (m) | Watershed area (km²) | Total stream length (km) | Max stream length (km) | Average slope | Channel slope drop (m) |
|---|---|---|---|---|---|
| 30 | 2,266.00 | 87.22 | 31.98 | 0.24 | 84.00 |
| 60 | 2,262.70 | 85.50 | 31.61 | 0.24 | 84.40 |
| 90 | 2,267.30 | 86.67 | 31.84 | 0.24 | 70.83 |
| 120 | 2,261.50 | 85.55 | 31.58 | 0.23 | 70.00 |
| 150 | 2,259.65 | 85.63 | 31.51 | 0.22 | 63.42 |
| 180 | 2,258.26 | 85.11 | 31.41 | 0.21 | 67.86 |
| 210 | 2,246.87 | 84.59 | 31.32 | 0.20 | 62.30 |
| 240 | 2,238.20 | 84.97 | 31.22 | 0.19 | 69.17 |
| 270 | 2,231.81 | 84.32 | 31.12 | 0.18 | 64.21 |
| 300 | 2,224.24 | 84.02 | 31.03 | 0.18 | 67.14 |
| 330 | 2,216.66 | 83.73 | 30.93 | 0.17 | 66.09 |
| 360 | 2,212.09 | 83.43 | 30.83 | 0.16 | 69.05 |
| 390 | 2,209.51 | 83.14 | 30.73 | 0.15 | 70.00 |
| 420 | 2,191.94 | 82.84 | 30.64 | 0.14 | 70.95 |
| 450 | 2,193.37 | 82.54 | 30.54 | 0.14 | 71.91 |
| 480 | 2,194.20 | 81.79 | 32.05 | 0.14 | 71.83 |

**Table 4** | Average annual HSPF model simulations at varying DEM resolution (2007–2009)

| | Flow | | Sediment | |
|---|---|---|---|---|
| DEM resolution (m) | Volume (m³) | RE (percent) | Mass loss (kg) | RE (percent) |
| 30 | 114,269,065 | | 305,082,255 | |
| 60 | 114,253,529 | −0.014 | 304,321,836 | −0.249 |
| 90 | 114,241,339 | −0.024 | 299,487,322 | −1.834 |
| 120 | 114,238,633 | −0.027 | 283,061,377 | −7.218 |
| 150 | 114,103,375 | −0.145 | 279,912,941 | −8.250 |
| 180 | 113,782,279 | −0.426 | 276,270,260 | −9.444 |
| 210 | 113,069,240 | −1.050 | 272,655,035 | −10.62 |
| 240 | 111,524,990 | −2.401 | 269,741,081 | −11.58 |
| 270 | 111,389,485 | −2.520 | 266,827,127 | −12.539 |
| 300 | 111,021,538 | −2.842 | 265,360,519 | −13.020 |
| 330 | 110,950,691 | −2.904 | 263,380,536 | −13.669 |
| 360 | 110,587,316 | −3.222 | 261,354,790 | −14.333 |
| 390 | 110,481,045 | −3.315 | 259,963,615 | −14.789 |
| 420 | 110,229,654 | −3.535 | 259,322,942 | −14.999 |
| 450 | 110,076,533 | −3.669 | 258,947,691 | −15.122 |
| 480 | 109,720,033 | −3.981 | 258,575,434 | −15.244 |

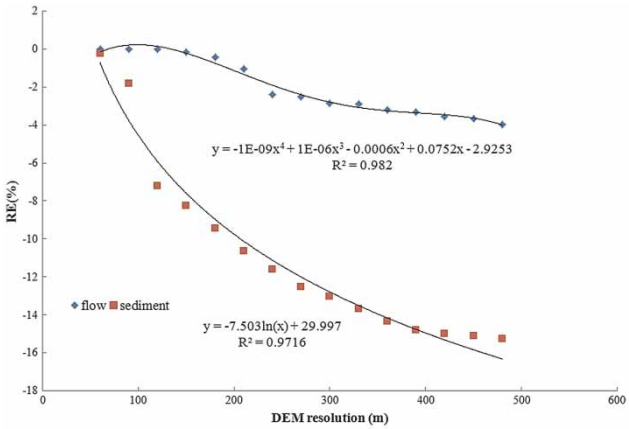


**Figure 3** | Average relative error (RE) (2007–2009) for flow and sediment using different DEM.

simulated flow decreased by 0.014, 0.024, 0.027, 2.401 and 3.981%, respectively, and the simulated sediment decreased by 0.249, 1.834, 7.218, 11.584 and 15.244%, respectively, indicating that predicted flow and sediment are substantially affected by DEM resolution. Coarser resolutions also reduced the average slope and the channel slope drop of the modeled watershed, which are important in the calculation of water balance in the HSPF model. The average slope of the watershed was reduced from 24% predicted using the 30 m resolution DEM to 14% predicted using the 480 m resolution DEM, which in turn resulted in less flow because watershed characteristics affect the flow generation derived from the Manning formula. Sediment predictions followed the same trend with

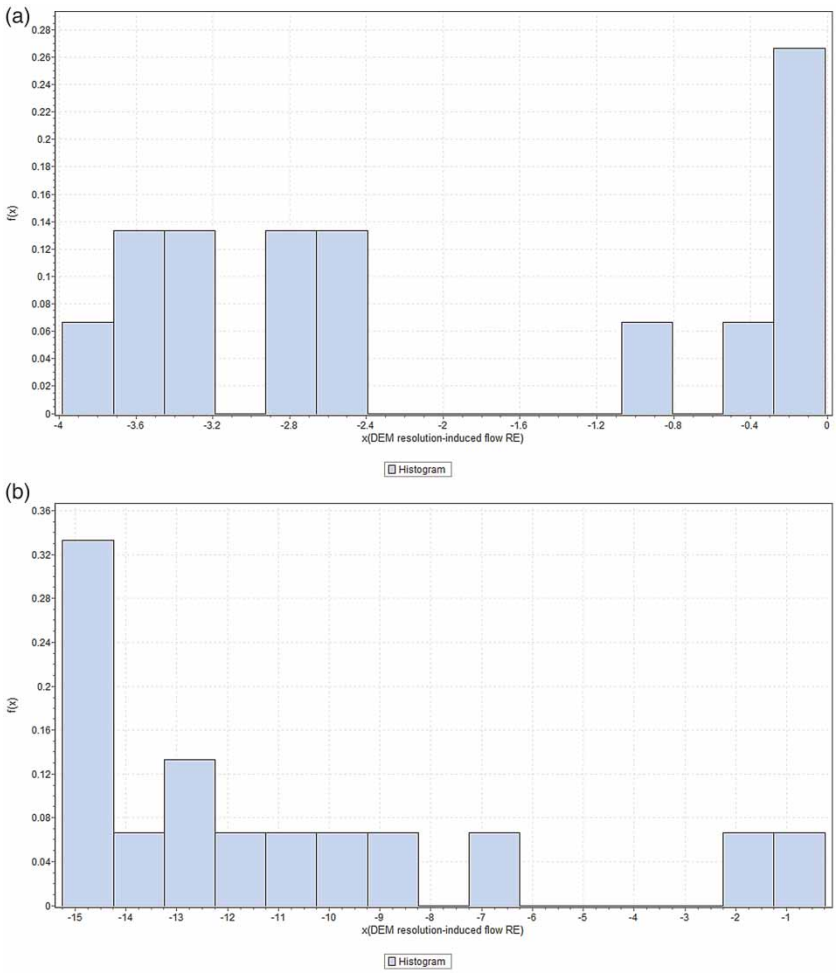


**Figure 4** | Probability density function for DEM resolution-induced uncertainty in: (a) flow *pdf*, (b) sediment *pdf*.

coarser DEMs resolution reducing the predicted sediment values, probably because of the decreased slope and stream length at coarser DEM resolutions. Watershed characteristics also affected the coefficient in the detached sediment washoff equation applied during sediment simulation. Our results showed that the resolution of the DEM input data affected sediment prediction more than flow prediction, agreeing with the findings of Cotter *et al.* (2003) and Chaplot (2005) using the SWAT model. It should be noted that the results for DEM resolution-induced uncertainty (RE) illustrated in Figure 3 may not be directly applicable to other river basins, but the method presented here should be suitable for use in the investigation of other basins.

The variance in RE for the model output of the simulated time series from 2007–2009 using different DEM resolutions compared with the simulated time series using the *base scenario* is illustrated in Figure 3. There is a non-linear relationship between the change in resolution of the DEM and the change in the uncertainty of predicted flow and sediment. The RE of predicted stream flow and sediment, representing the uncertainty related to DEM resolution, followed a logarithmic trend line and a multinomial trend line, respectively.

The *pdf*s of flow and sediment uncertainties induced by DEM resolution are shown in Figure 4.

From the *pdfs* for DEM resolution-induced uncertainty in flow prediction, it can be seen that as DEM resolution

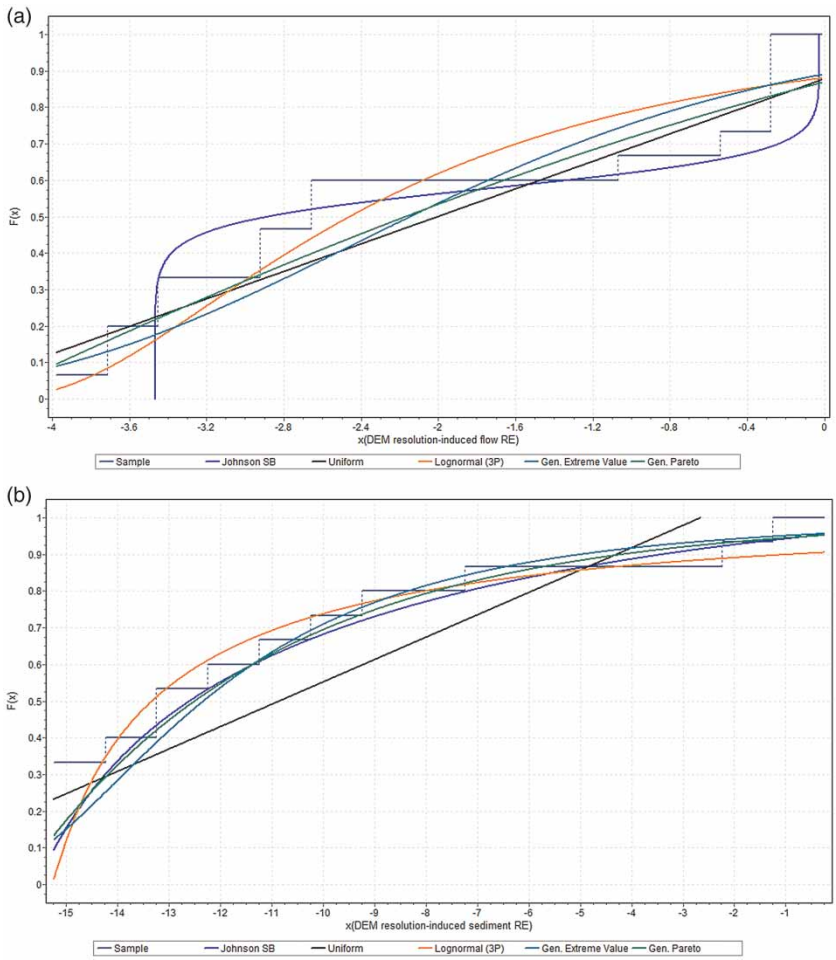


**Figure 5** | Probability density function for DEM resolution-induced uncertainty in: (a) flow *CDFs*, (b) sediment *CDFs*.

**Table 5** | Distributions fits and parameters estimates for data resolution-induced uncertainty in flow

| | KS statistic | | Parameters | |
|---|---|---|---|---|
| **Distribution of RE for flow** | **RE due to DEM** | **RE due to land use** | **RE due to DEM** | **RE due to land use** |
| Generalized extreme value | 0.176 | 0.122 | $k=-0.135$<br>$\sigma=1.417$<br>$\mu=-2.654$ | $k=0.261$<br>$\sigma=2.746$<br>$\mu=-13.393$ |
| Generalized Pareto | **0.147** | 0.133 | $k=-0.683$<br>$\sigma=3.975$<br>$\mu=-4.366$ | $k=0.036$<br>$\sigma=4.904$<br>$\mu=-15.948$ |
| Johnson SB | 0.229 | **0.091** | $\gamma=0.193$<br>$\delta=0.121$<br>$\lambda=3.441$<br>$\xi=-3.468$ | $\gamma=1.107$<br>$\delta=0.615$<br>$\lambda=21.485$<br>$\xi=-15.655$ |
| Log-normal (3P) | 0.170 | 0.142 | $\alpha=2.026$<br>$\beta=1.761$<br>$\gamma=-4.2476$ | $A=0.858$<br>$B=1.857$<br>$\Gamma=-15.244$ |
| Uniform | 0.175 | 0.233 | $a=-4.656$<br>$b=0.646$ | $A=-19.066$<br>$B=-2.657$ |

decreased from 30 to 480 m, the data were distributed in two blocks, in the zone between −1.2 and 0, and in the zone between −4.0 and −2.4. In the zone (bin) between −2.4 and −1.2 there were no data. This break in the data suggests that as the resolution of the DEM changed, there were resolutions at which the RE changed considerably. This is also evident in Figure 3, which illustrates that when the DEM resolution decreased from 180 to 210 m and from 210 to 240 m, the RE for predicted flow increased to a greater extent than when the DEM resolution decreased from 30 to 180 m and from 240 to 480 m. In other words, when the resolution of the DEM was decreased to resolutions coarser than 180 × 180 m, there was a large degree of uncertainty. Therefore, in practical terms, a DEM resolution finer than 180 × 180 m should be selected. In the distribution of the RE for predicted sediment, there were no data in the zone between −6 and −2, indicating a similar break in the data as seen for flow prediction. Figure 3 shows that when the DEM resolution decreased from 90 to 180 m, there

**Table 6** | Land use change with different resolution

| Land use resolution (m) | Agricultural areas ($km^2$) | Forest areas ($km^2$) | Grass areas ($km^2$) | Orchard areas ($km^2$) | Urban and built-up areas ($km^2$) | Water ($km^2$) | Total area ($km^2$) |
|---|---|---|---|---|---|---|---|
| 20 | 313.75 | 1,676.45 | 115.70 | 9.92 | 103.68 | 38.59 | 2,258.09 |
| 60 | 311.18 | 1,687.31 | 109.59 | 10.02 | 101.80 | 37.62 | 2,257.52 |
| 120 | 305.85 | 1,699.95 | 102.91 | 9.68 | 102.10 | 36.82 | 2,257.31 |
| 180 | 302.36 | 1,711.40 | 96.61 | 9.63 | 100.95 | 35.91 | 2,256.86 |
| 240 | 290.25 | 1,710.62 | 104.16 | 10.99 | 104.41 | 37.26 | 2,257.69 |
| 300 | 287.93 | 1,717.99 | 100.98 | 10.57 | 102.77 | 35.93 | 2,256.17 |
| 360 | 284.35 | 1,723.33 | 100.37 | 10.75 | 102.83 | 35.49 | 2,257.12 |
| 420 | 283.77 | 1,728.66 | 98.76 | 10.93 | 102.89 | 35.05 | 2,260.07 |
| 480 | 283.3 | 1,715.07 | 100.54 | 11.03 | 97.14 | 46.87 | 2,253.94 |
| 540 | 284.07 | 1,720.41 | 100.83 | 11.24 | 100.43 | 40.62 | 2,257.61 |
| 600 | 289.71 | 1,730.06 | 85.32 | 11.42 | 100.10 | 40.00 | 2,256.62 |
| 660 | 275.36 | 1,739.72 | 92.81 | 11.59 | 99.78 | 40.38 | 2,259.62 |
| 720 | 281.00 | 1,739.37 | 87.29 | 11.76 | 99.45 | 40.76 | 2,259.63 |
| 780 | 292.64 | 1,738.03 | 77.78 | 11.93 | 99.12 | 41.14 | 2,260.64 |
| 840 | 282.28 | 1,737.69 | 85.26 | 12.10 | 98.80 | 41.52 | 2,257.65 |
| 900 | 287.92 | 1,738.34 | 77.75 | 12.27 | 98.47 | 41.90 | 2,256.66 |
| 960 | 269.96 | 1,740.42 | 104.62 | 7.37 | 94.38 | 43.55 | 2,260.30 |

was a large increase in uncertainty (RE) in sediment prediction.

The DEM resolution-induced uncertainty distributions are illustrated in Figure 5, and results from the distribution-fit analysis are presented in Table 5. For flow, the generalized Pareto distribution was the best fit for DEM resolution-induced uncertainty (RE), with the lowest KS value of 0.147 (Table 5). Therefore, the DEM resolution-induced uncertainty (RE) in flow ($f(D_f)$) simulation can be described with a generalized Pareto distribution using the parameter estimates of Table 5.

$$f((D_f)) = \frac{1}{3.975}\left(1 - 0.683 \bullet \frac{(x + 4.366)}{3.975}\right)^{-1+\frac{1}{0.683}} \quad (5)$$

For sediment, the lowest KS value for the DEM resolution-induced uncertainty (RE) was 0.091 for a Johnson SB distribution (Table 5). The *pdf* of sediment uncertainty (RE) related to DEM resolution ($f(D_s)$), which followed a Johnson SB distribution, is defined in Equation (6) using the parameter estimates of Table 5.

$$f((D_s)) = \frac{0.121}{3.441\sqrt{2\pi}z(1-z)} \bullet \exp\left(-\frac{1}{2}\left(0.193 + 0.121 \bullet \ln\left(\frac{z}{1-z}\right)\right)^2\right),$$
$$z = \frac{x + 3.468}{3.441} \quad (6)$$

However, Figure 5 shows that the best-fit distribution (theoretical *CDF*) could not be clearly defined by these uncertainties (empirical *CDF*). This is mainly because of the limited number of resolutions of DEM data used in this study. A larger number of RE values derived from more DEM resolutions would improve the distribution fit. A dataset with more resolutions would enable better identification of the distribution fit and smoother functions of DEM resolution-induced uncertainty.

## Estimation of land use data resolution-induced uncertainty

We considered land use classes at resolutions ranging from 20 m (finest) to 960 m (coarsest). Table 6 shows an insignificant change in the total area for resolutions of 20 m up to 960 m. However, the area of forest increased, the area of agricultural land decreased, and the areas of other land use classes varied erratically. The spatial resolution was resampled based on the nearest neighbor method in which the area is redistributed in accordance with the class of clusters of neighboring pixels. The data were then used to simulate flow and sediment for different sets of land use maps. The relative distribution of urban and built-up land, agricultural land, and forest areas within a watershed can affect the prediction of flow and sediment response when land use characteristics are used to derive model parameters. When land use was redistributed from agricultural land to forest, the evaporation losses (including potential, interception, upper zone, lower zone, baseflow, and active groundwater) increased, resulting in lower predicted surface runoff. This led to a decrease in predicted flow and sediment transport from the watershed.

The land use data at a $20 \times 20$ m resolution were used to simulate the *base scenario*. Flow and sediment were

**Table 7** | Average annual HSPF model simulations at varying land use resolution (2007–2009)

| | Flow | | Sediment | |
|---|---|---|---|---|
| Land use resolution (m) | Volume ($m^3$) | RE (percent) | Mass loss (kg) | RE (percent) |
| 20 | 114,269,065 | | 305,082,255 | |
| 60 | 113,964,135 | −0.267 | 302,624,552 | −0.806 |
| 120 | 113,673,341 | −0.521 | 302,001,435 | −1.010 |
| 180 | 113,570,767 | −0.611 | 301,692,791 | −1.111 |
| 240 | 113,480,152 | −0.690 | 303,028,668 | −0.673 |
| 300 | 113,314,918 | −0.835 | 301,354,150 | −1.222 |
| 360 | 112,885,267 | −1.211 | 298,977,559 | −2.001 |
| 420 | 112,556,172 | −1.499 | 296,942,660 | −2.668 |
| 480 | 112,027,540 | −1.962 | 295,405,858 | −3.172 |
| 540 | 111,870,557 | −2.099 | 294,816,237 | −3.365 |
| 600 | 111,785,220 | −2.174 | 293,217,606 | −3.889 |
| 660 | 111,530,890 | −2.396 | 292,540,323 | −4.111 |
| 720 | 111,276,560 | −2.619 | 291,899,651 | −4.321 |
| 780 | 111,022,229 | −2.841 | 290,532,882 | −4.769 |
| 840 | 110,767,899 | −3.064 | 289,825,091 | −5.001 |
| 900 | 110,513,569 | −3.287 | 288,748,151 | −5.354 |
| 960 | 109,864,770 | −3.854 | 287,246,666 | −5.846 |

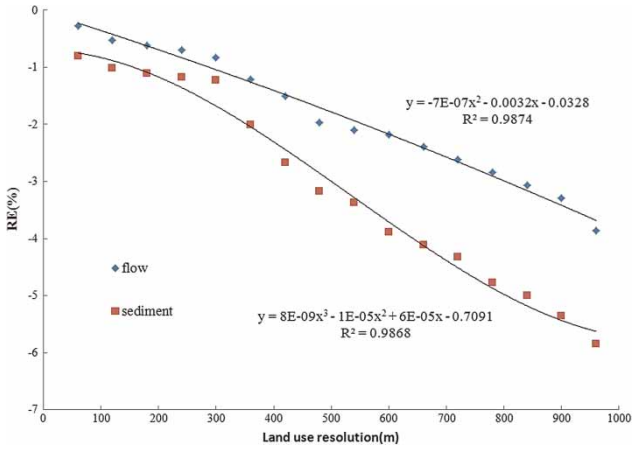


**Figure 6** | Average relative error (RE) (2007–2009) for flow and sediment using different land use.

simulated using 16 land use data resolutions (60 × 60 m, 120 × 120 m, 180 × 180 m, 240 × 240 m, 300 × 300 m, 360 × 360 m, 420 × 420 m, 480 × 480 m, 540 × 540 m, 600 × 600 m, 660 × 660 m, 720 × 720 m, 780 × 780 m, 840 × 840 m, 900 × 900 m, 960 × 960 m). The REs of the average simulated stream flow and sediment between each resolution and the *base scenario* are listed in Table 7. As the resolution decreased, especially when the resolution was coarser than 240 m, the predicted flow and sediment also gradually decreased. When the spatial resolution decreased from 20 m to 60 m, 120 m, 240 m, 480 m, and 960 m, the predicted flow decreased by 0.267, 0.512, 0.690, 1.962 and 3.854%, respectively, and the predicted sediment decreased by 0.806, 1.010, 0.673, 3.172, and 5.846%, respectively

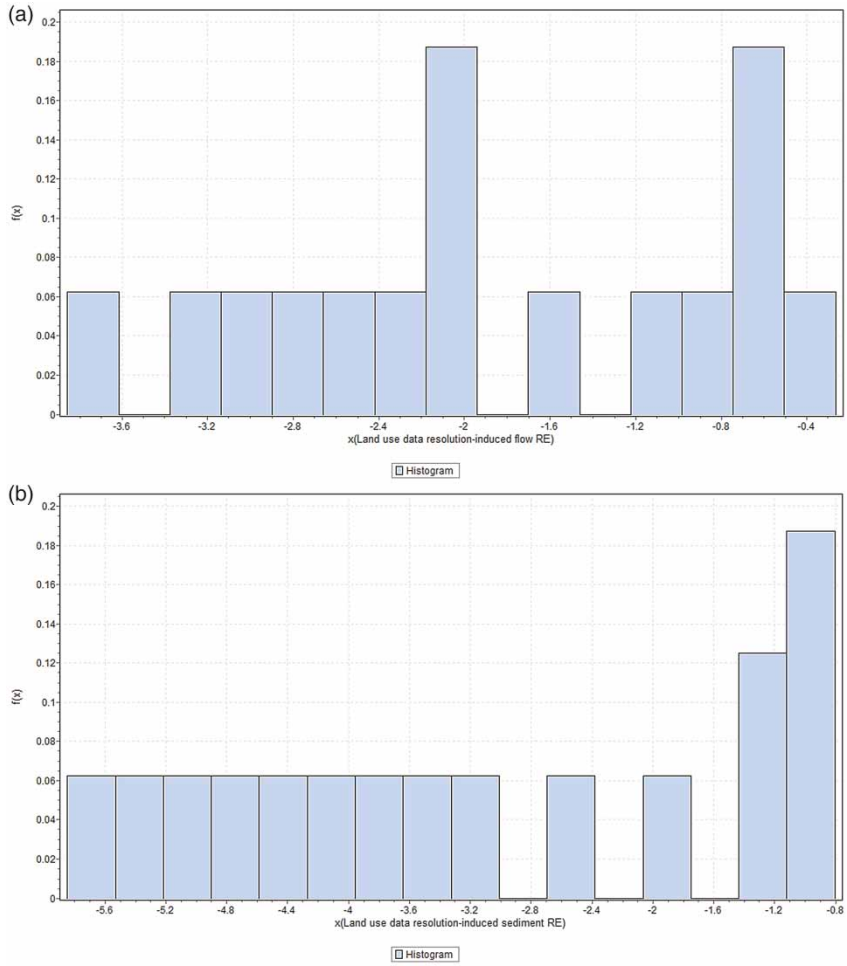


**Figure 7** | Probability density function for land use data resolution-induced uncertainty in: (a) flow *pdfs* (b) sediment *pdfs*.

(Figure 6). The spatial resolution of land use data affected sediment prediction more than flow prediction. These results support previous findings by Luzio *et al.* (2005) using the SWAT model, in which they showed that, although the land use input did not affect runoff estimates, sediment yields were significantly biased.

Figure 6 shows the RE variance in model output related to land use resolution for the calibration period. Similar to the DEM resolution-induced uncertainty, the relationships between land use data resolution and the predicted flow and predicted sediment were non-linear.

As for the RE related to DEM resolution, the RE related to land use resolution was analyzed using *pdfs* and *CDFs*. The *pdfs* for land use resolution-induced uncertainties in predicted values of flow and sediment are shown in Figure 7 and the *CDFs* for land use resolution-induced uncertainties in predicted values of flow and sediment are shown in Figure 8.

Results of the distribution fit for RE related to land use resolution are presented in Table 8. Based on the lowest KS values for flow and sediment of 0.099 and 0.096, respectively (Table 8), the *pdfs* of land use data resolution-induced uncertainty (RE) in simulated flow ($f(L_f)$) and sediment ($f(L_s)$), as described in Equation (7) and Equation (8) using the parameter estimates in Table 8, followed a generalized Pareto distribution and a Johnson SB distribution, respectively. The determination of the distribution fit was restricted by the limited resolutions of land use data available for this study (Figure 8). The accuracy of the results would be improved if land use data at a greater range of

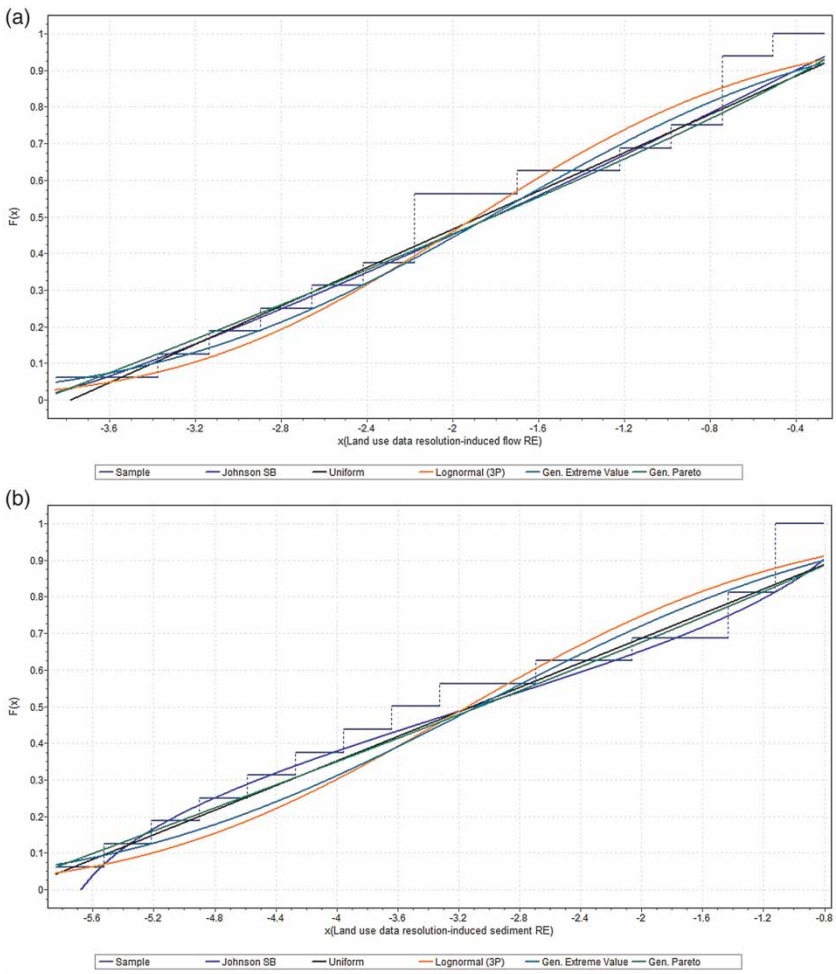


**Figure 8** | Probability density function for land use data resolution-induced uncertainty in: (a) flow *CDFs*, (b) sediment *CDFs*.

**Table 8** | Distributions fits and parameters estimates for data resolution-induced uncertainty in sediment

| | KS statistic | | Parameters | |
|---|---|---|---|---|
| **Distribution of RE for sediment** | **RE due to DEM** | **RE due to land use** | **RE due to DEM** | **RE due to land use** |
| Generalized extreme value | 0.117 | 0.149 | $k = -0.351$ | $k = -0.291$ |
| | | | $\sigma = 1.191$ | $\sigma = 1.843$ |
| | | | $\mu = -2.240$ | $\mu = -3.723$ |
| Generalized Pareto | **0.099** | 0.113 | $k = -1.153$ | $k = -1.016$ |
| | | | $\sigma = 4.432$ | $\sigma = 6.324$ |
| | | | $\mu = -3.929$ | $\mu = -6.219$ |
| Johnson SB | 0.100 | **0.096** | $\gamma = -0.126$ | $\gamma = 0.021$ |
| | | | $\delta = 0.691$ | $\delta = 0.429$ |
| | | | $\lambda = 4.180$ | $\lambda = 5\ .257$ |
| | | | $\xi = -4.098$ | $\xi = -5.676$ |
| Log-normal (3P) | 0.142 | 0.161 | $\alpha = 4.929$ | $\alpha = 10.551$ |
| | | | $\beta = 3.205$ | $\beta = 11.079$ |
| | | | $\gamma = -3.205$ | $\gamma = -14.272$ |
| Uniform | 0.100 | 0.117 | $a = -3.784$ | $a = -6.136$ |
| | | | $b = 0.043$ | $b = -0.029$ |

resolutions were used in future.

$$f((L_f)) = \frac{1}{4.432}\left(1 - 1.153 \bullet \frac{(x + 3.929)}{4.432}\right)^{-1+\frac{1}{1.153}}; k \neq 0 \quad (7)$$

$$f((L_s)) = \frac{0.429}{5.257\sqrt{2\pi}z(1-z)} \bullet \exp\left(-\frac{1}{2}\left(0.021 + 0.429 \bullet \ln\left(\frac{z}{1-z}\right)\right)^2\right),$$
$$z = \frac{x + 5.676}{5.257} \quad (8)$$

This study examined input data-induced uncertainty for the HSPF model only in terms of spatial data resolution. However, previous studies have shown that the spatial–temporal distribution of rainfall can significantly affect model uncertainty (Segond *et al.* 2007; Younger *et al.* 2009; Pechlivanidis *et al.* 2010). A limitation of our study is that we neglected the impact of the spatial resolution of rainfall at fine temporal resolutions (i.e., hourly time steps). Although we applied disaggregated hourly data in the HSPF model, a loss of information resulting from real temporal rainfall distribution was expected. Finally, from a water quality management perspective, because the resolution of DEM and land use data affects flow and sediment generation, we believe that the resolution of DEM and land use data will also affect pollution load (e.g., $NH_3$-N, phosphate, total nitrogen, and total phosphorus) simulated using the HSPF model. Further studies are required to investigate this proposition.

## CONCLUSIONS

The results of our study highlighted DEM and land use data resolution-induced uncertainty in stream flow and sediment predicted using the HSPF model, and demonstrated the discriminative power of the method of *pdfs* in quantifying these uncertainties. Input data from fine-scale DEMs and land use resolution maps generated higher flow volumes and sediment loads compared with coarser-scale counterparts. The resolution of DEMs can affect the stream length, watershed area, and average slope, which can in turn affect the uncertainty in simulated flow and sediment. Land use data resolution can affect infiltration and evaporation loss,

which in turn affects flow and sediment detachment and washoff, with an impact on sediment generation. Our results indicate that every effort must be made to collect spatial data at a fine resolution to minimize uncertainty in model simulations.

Digital elevation models and land use *pdf*s can be used to determine the spatial data resolution-induced uncertainty in simulated flow and sediment. Using the *pdf*s, we identified a non-linear relationship between changes in the resolution of DEM and land use data and changes in the uncertainty of predicted flow and sediment. In this study, we selected the best-fit distributions (theoretical *CDF*) to represent flow and sediment uncertainty (empirical *CDF*). However, the best-fit distributions (theoretical *CDF*) could not adequately represent the uncertainties (empirical *CDF*) because of the shortage of a large number of resolutions in DEM and land use data. The distribution fit would be improved by including DEM and land use data for a wider range of resolutions. The distributions of DEM and land use resolution-induced uncertainty (RE) could provide useful information for estimating the total model uncertainties using methods like the Bayesian approach, and the parameter estimates could provide insights into watershed model uncertainty. However, the use of different distributed models and datasets of different resolutions will result in variations in parameter estimations.

## ACKNOWLEDGEMENTS

Funding was supported by One Hundred Talents Program of the Chinese Academy of Sciences, the National Natural Science Foundation of China (No. 41071323 and 40971271), and the CAS/SAFEA International Partnership Program for Creative Research Teams of 'Ecosystem Processes and Services' (Grant No. KZCX2-YW-T13). The authors would like to thank the Chengde Branch of Hebei Provincial Survey Bureau of Hydrology and Water Resources for providing hydrological data. Finally, the authors would like to express their sincere gratitude to the anonymous reviewers for their constructive comments and the editor of the journal. Their detailed suggestions have resulted in an improved paper.

## REFERENCES

Andronova, N. G. & Schlesinger, M. G. 2001 Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.* **106** (D19), 22605–22611.

Beven, K. 2001 *Rainfall-Runoff Modelling. The Primer*. John Wiley and Sons, Chichester, UK, 360 pp.

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

Borah, D. K. & Bera, M. 2003 Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases. *T. ASABE.* **46** (6), 1553–1566.

Chaplot, V. 2005 Impact of DEM mesh size and soil map scale on SWAT runoff, sediment, and $NO_3$- N loads predict ions. *J. Hydrol.* **312** (1–4), 207–222.

Chaubey, I., Cotter, A. S., Costello, T. A. & Soerens, T. S. 2005 Effect of DEM data resolution on SWAT output uncertainty. *Hydrol. Process.* **19** (3), 621–628.

Cho, S. M. & Lee, M. W. 2001 Sensitivity considerations when modeling hydrologic processes with digital elevation model. *J. Am. Water Resour. Assoc.* **37** (4), 931–934.

Cotter, A. S., Chaubry, I., Costello, T. A., Soerens, T. S. & Nelson, M. A. 2003 Water quality model output uncertainty as affected by spatial resolution of input data. *J. Am. Water Resour. Assoc.* **39** (4), 977–986.

Diaz-Ramirez, J. N., McAnally, W. H. & Martin, J. L. 2007 Evaluation of HSPF uncertainty bounds using a probabilistic point estimate method. Watershed Management to Meet Water Quality Standards and TMDLS (Total Maximum Daily Load) Proceedings of the Fourth Conference 10–14 March 2007. San Antonio, Texas, USA, pp. 161–169, ASABE Publication 701P0207.

Donigian Jr., A. S., Imhoff, J. C. & Kittle Jr, J. L. 1999 *HSPF-Parm: An interactive database of HSPF model parameters*. Version 1.0. EPA-823-R-99-004. US Environmental Protection Agency, Athens, GA.

Haan, P. K. & Skaggs, R. W. 2003 Effect of parameter uncertainty on DRAINMOD predictions: I Hydrology and Yield. *T. ASABE.* **46** (4), 1061–1067.

Johnson, M. S., Coon, W. F., Mehta, V. K., Steenhuis, T. S., Brooks, E. S. & Boll, J. 2003 Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. *J. Hydrol.* **284** (1–4), 57–76.

Luzio, M. D., Arnold, J. G. & Srinivasan, R. 2005 Effect of GIS data quality on small watershed stream flow and sediment simulations. *Hydrol. Process.* **19** (3), 629–650.

Patil, A. & Deng, Z. 2012 Input data measurement-induced uncertainty in watershed modeling. *Hydrolog Sci. J.* **57** (1), 118–133.

Patil, A., Deng, Z. & Malone, R. F. 2011 Input data resolution-induced uncertainty in watershed modeling. *Hydrol. Process.* **25** (2), 2302–2312.

Pechlivanidis, I. G., McIntyre, N. R. & Wheater, H. S. 2010 Calibration of the semi-distributed PDM rainfall-runoff model in the Upper Lee catchment, UK. *J. Hydrol.* **386** (1–4), 198–209.

Quiroga, V. M., Popescu, I., Solomatine, D. P. & Bociort, L. 2013 Cloud and cluster computing in uncertainty analysis of integrated flood models. *J. Hydroinform.* **15** (1), 55–70.

Segond, M. L., Wheater, H. S. & Onof, C. 2007 The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. *J. Hydrol.* **347** (1–2), 116–131.

Wagenet, R. J. & Hutson, J. L. 1996 Scale-dependency of solute transport modeling/GIS applications. *J. Environ. Qual.* **25** (3), 499–510.

Wilson, J. P., Inskeep, W. P., Wraith, J. M. & Snyder, R. D. 1996 GIS-based solute transport modeling applications: scale effects of soil and climate data input. *J. Environ. Qual.* **25** (3), 445–453.

Wolock, D. M. & Price, C. V. 1994 Effect of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resour. Res.* **30** (11), 3041–3052.

Xiu, D. & Karniadakis, G. E. 2003 Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.* **187** (1), 137–167.

Younger, P. M., Freer, J. E. & Beven, K. J. 2009 Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. *Hydrol. Process.* **23** (14), 1988–2003.

First received 21 November 2012; accepted in revised form 26 September 2013. Available online 13 November 2013