A comprehensive evaluation of input data-induced uncertainty in nonpoint source pollution modeling

L. Chen[1], Y. Gong[2], and Z. Shen[1,*]

1. State Key Laboratory of Water Environment, School of Environment, Beijing Normal University, Beijing 100875, P.R. China

2. Key Laboratory of Urban Stormwater System and Water Environment, Ministry of Education, Beijing University of Civil Engineering and Architecture, Beijing, China 100044.

Corresponding author: Zhenyao Shen Tel/fax: +86 10 58800398.

*E-mail address:*zyshen@bnu.edu.cn; chenlei1982bnu@bnu.edu.cn

Abstract

Watershed models have been used extensively for quantifying nonpoint source (NPS) pollution, but few studies have been conducted on the error-propagation from different input data sets to NPS modeling. In this paper, the effects of four input data, including rainfall, digital elevation models (DEMs), land use maps, and the amount of fertilizer, on NPS simulation were quantified. A systematic input-induced uncertainty was investigated using watershed model for phosphorus load prediction. Based on the results, the rain gauge density resulted in the largest model uncertainty, followed by DEMs, whereas land use and fertilizer amount exhibited limited impacts. The simulation errors, in terms of coefficient of variation, related to single rain gauges-, multiple gauges-, ASTER GDEM-, NFGIS DEM-, land use-, and fertilizer amount information was 0.390, 0.274, 0.186, 0.073, 0.033 and 0.005, respectively. The use of specific input information, such as key gauges, is also highlighted to achieve the required model accuracy. In this sense, these results provide valuable information to other model-based studies for the control of prediction uncertainty.

1. Introduction

Nonpoint source (NPS) pollution has become the major obstacle in sustaining high-quality water supplies in developed countries, such as the United States, as well as in developing countries, such as China (Zheng et al., 2011). Hydrological models, such as the Agricultural Non-Point Source Model (AGNPS) and Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998), provide important tools for quantifying NPS loads and understanding their perturbations to water quality. Nevertheless, due to the complexity of watershed systems and substantial requirements for input data, uncertainty becomes an inevitable part of model-based research and thus management plans (Beven, 2006; Xue et al., 2014). Typically, model uncertainty comes from its structure, parameter choice and input data. Structure uncertainty results from incomplete knowledge of watershed processes or different assumptions during model setup, whereas parameter uncertainty arises due to the imprecise representation of parameter ranges and distributions. In addition, input uncertainty is generated from simplification in natural randomness and temporal-spatial data variability and would be inevitably propagated to model output errors.

Model inputs typically include spatial data, such as spatial precipitation input, digital elevation models (DEMs), land use maps and soil maps, as well as attribute data, such as fertilizer amount (Shen et al., 2013a). The uncertainty of spatial data, typically in the forms of GIS maps, is derived from many factors, including the quantity of available images, the resolution for the data that were captured, and the choice of interpolation techniques (Wu et al., 2005). Rainfall plays a crucial role in runoff production and mass transport so its reliability has been considered as major factor for the accuracy of hydrological models (Andréassian et al., 2001; McMillan et al., 2011). Traditionally, the rain station is the fundamental tool for representing spatial distribution of rainfall within a watershed (Andréassian et al., 2001). Designing the proper location, number and density of rain-gauge stations is important to hydrological research (Duncan et al., 1993). Studies have explored the impact of

heterogeneous rainfall data on parameter estimation and model outputs and concluded that large bias could be expected if detailed variations in the rainfall data are not considered (Strauch et al., 2012).

As another important GIS data, a DEM is used to extract surface characteristic parameters, such as watershed boundary, slope, and thus flow direction, so its resolution influences model outputs (Lin et al., 2013; Wellen et al., 2014). Studies have noted that coarser DEMs smooth watershed slope and thereby reduce the simulated peak flow or sediment yields (Zhang et al., 2014). It is also shown that nitrogen output decreased with the decreased DEM resolution, while a decreased DEM resolution does not always resulted in decreased total phosphorus (TP) (Chaubey et al., 2005). In this sense, the question about whether higher-resolution data would always lead to better model performance should be considered first (Shen et al., 2013). One of the interesting results from Chaplot's (2005a) work is that there exists a spatial resolution saturation level, beyond which further refinements to resolution do not improve model performance. In the meantime, GIS data may be available from alternative sources; therefore, another question is which specific data set should be used. For example, land use maps could be obtained from federal, state and local government agencies, whereas county and local governments are developing detailed datasets (Shen and Zhao, 2010; Han et al., 2014). Land use maps for a specific point in time, typically obtained by interpreting remote sensing data, are often used, and possible changes in land uses during that specific period are not considered (Mango et al., 2011; Pai and Saraswat, 2013). These statistical or modeling analyses have demonstrated that the land use changes affect hydrological characteristics, which further alter the occurrence of soil erosion and transport of the NPS pollutants.

Despite the research progress described above, input-induced uncertainty remains a significant challenge due to various input data, which largely limits the applicability of watershed models. For example, model-based programs, such as Total Maximum Daily Loads (TMDLs), are often criticized for their inadequate consideration of input uncertainty (Chen et al., 2012). First, there is relatively more uncertainty research about hydrological processes (Beven, 2006; Balin et al., 2010; Vrugt et al., 2008) but

less on NPS pollution (Chaplot et al., 2005a; Chaplot, 2005b; Gassman et al., 2007; Wellen et al., 2015). These studies have showed the input uncertainty is propagated through the watershed model, to some extent, to sediment modeling and then carry-over and magnify into pollutant simulation. Uncertainty is currently considered as one of the core dilemmas in watershed studies, especially in the field of NPS modeling. Second, the sensitivity of watershed models also depends on how well attribute data aggregation describes the relevant characteristics of human management. For example, the SWAT assumed P could be added onto the soil in the form of fertilizer or manure, and specific attribute data include the timing of fertilization, the type and amount of fertilizer/manure, and the distribution of the soil layer. Thus, it is useful to understand the assumptions of these attribute data and how these assumptions will likely impact the model results. Third, previous studies have not evaluated the relative contribution of each input data set so a strategy on how to reduce input uncertainty cannot be formulated in a cost-effective manner (Munoz-Carpena et al., 2006).

The main objective of this paper is to conduct a comprehensive assessment of input-induced uncertainty in TP modeling. Four key types of input data, i.e., rainfall, topography, land use and fertilizer amount, are analysed, and their uncertainties are quantified. The uncertainties related to these input data are then compared.

2. Materials and Methods

2.1 The description of the study area

The Upper Daning River Watershed, which is located in the Three Gorges Reservoir Area of China, was selected as the studied watershed (**Fig. 1**). This watershed, covering an area of 2,421 km^2, is characterized as being located in a northern subtropical monsoon climate with an annual mean rainfall of 1,182 mm (ranging from 761 mm to 1,356 mm). This watershed is very mountainous with elevations ranging from 200-2605 m. The primary land uses in this watershed are forest (61.8%), arable land (25.3%), and pasture (12.5%), and yellow-brown earths (26.5%), yellow-cinnamon soils (16.9%) and purplish soils (14.5%) are the dominant soil types.

115 More information about the study area are referred to Shen et al., (2012a, 2013a, b).
116 Based on the characteristics of the river system, the studied watershed was broken
117 into six drainage regions: Dongxi river, Xixi river, Baiyang river, upper region of the
118 Wuxi hydrological gauge, Houxi river, and upper region of the county boundary
119 (watershed outlet). As illustrated in **Fig. 1**, the corresponding outlets of are referred to
120 as DX, XX, BY, WX, HX, and CF, respectively. In this study, TP was evaluated as P
121 was recognized as the key limiting factor of eutrophication in this region.

122 ## 2.2 Model description

123 In this study, the SWAT model, as a commonly-used watershed model, was used for
124 NPS-TP modeling. The studied watershed was partitioned into 22 sub-watersheds
125 from a constructed DEM and each sub-watershed is then divided into hydrologic
126 response units (HRUs) by designing their homogeneous slope, soil, and land use. The
127 SWAT-CUP software (Abbaspour et al., 2007) was applied for model calibration and
128 validation. The measured water quality and flow data were obtained from the
129 Changjiang Water Resources Commission as well as local government. Thereafter,
130 the SWAT model was calibrated and validated using the initial input data (Shen et al.,
131 2012a), and the propagation error from input data to model outputs was quantified by
132 changing the available datasets while keeping the calibrated parameters fixed. The
133 model outputs were simulated flow amount, sediment load, and TP load, which were
134 predicted at a monthly step because only monthly measured TP were available in this
135 area.

136 ## 2.3 Generation of input-induced uncertainty

137 Errors introduced by rainfall data, DEMs and land use maps were analyzed. The
138 influence of soil type maps was not analyzed, because only one soil map data (coarse
139 resolution at 1:1000000) was available for the study region. These GIS data are the
140 most frequently used in hydrology and NPS modeling in the Yangtze River
141 Watershed and other areas of China. The errors related to fertilizer amount were also
142 investigated due to the lack of detailed farm-scale data.

2.3.1 Spatial data 1: Rainfall data

In this study, rainfall datasets were collected from twelve rain gauges located within the watershed boundary and two outside stations that were within approximately 10 km of the watershed boundary were also used (**Fig. 1**). The rain gauge falling within a given sub-catchment is identified using the GIS software. The annual mean rainfall recorded by these rain gauges is listed in **Table 1**. Previous studies have demonstrated rainfall uncertainty comes from the lack of representative rain gauges and then the need to interpolate the rainfall data between rain gauges (Andréassian et al., 2001; McMillan et al., 2011). Our previous study (Shen et al., 2012a) has already focused on the impact of interpolation methods on the spatial rainfall heterogeneity so we focused on the representativeness of rainfall stations. In this sense, rainfall data-induced uncertainty was analyzed in two steps: 1) the dataset of each rain gauge was used as inputs for the SWAT model separately, and the model performances were ranked based on the Nash–Sutcliffe efficiency coefficient (E_{NS}) values for single gauge simulations; 2) random combinations of m rain gauges (m ranged from 2 to 12) were generated and used as SWAT inputs. The expected rainfall spatial distributions were only generated by the centroid method was selected because it was the current approach incorporated into the current version of SWAT model and the easiest to apply (Shen et al., 2012a).

2.3.2 Spatial data 2: DEMs

In this watershed, two DEM sets were available for NPS modeling: 1) the National Fundamental Geographic Information System of China DEM (NFGIS DEM) and 2) the ASTER GDEM. Specifically, the NFGIS DEM was acquired in 1998 from a topographic map with a resolution of 90 m, whereas the ASTER GDEM was created by a satellite-borne image that covered the surface land at a resolution of 30 m (Shen et al., 2013a). To study the impact of data resolution on NPS simulations, both DEMs were converted to coarser ones using the resample function of ArcMap. Finally, four NFGIS DEM maps (90*90 m, 120*120 m, 150*150 m and 180*180 m), and ten

ASTER GDEM maps (30*30 m, 40*40 m, 50*50 m, 60*60 m, 70*70 m, 80*80 m, 90*90 m, 120*120 m, 150*150 m and 180*180 m) were obtained.

2.3.3 Spatial data 3: Land use maps

As discussed above, land use data available for the modeling effort will likely come from numerous sources; therefore, an assessment of available land use data and the time period covered by these data should be made. In this study, land use data were obtained from the 1980s (1980–1989), 1995, 2000, and 2007. Specifically, maps from the 1980s, 1995 and 2000 were interpreted from MSS/TM/ETM images by the Chinese Academy of Sciences, whereas the land use map for 2007 was created from a TM image. To substantiate the impacts of land use maps, an analytical framework was developed in two steps. Firstly, the characteristics of land use distribution during each period were analyzed according to land use type of each map. The land use statistics are shown in **Table 2**. Second, these four land use maps were used as model inputs and their impacts were estimated respectively using the calibrated SWAT model. In our previous study (Shen et al., 2013a), the resolution of land use data was shown to have only a slight influence on simulated NPS-P for the study region; therefore, the land use map was not resampled in this study.

2.3.4 Attribute data: amount of fertilizer

Traditional potato-sweet potato rotation was the most popular cropping system in the agricultural area under the slope of 15-degree, while the duration of rotations were typically half year-half year. Besides, most of the growers on the higher area (>15-degree) panted corn, which is becoming more and more popular due to higher returns under recent market conditions. In our analysis, we studied the impacts of fertilizer and did not attempt to change the rotation pattern or introduce alternative crops. Attribute data, including crop planting time, irrigation, fertilization, and tillage, were mainly obtained from the agricultural bureau and local farmers; therefore, these data only reflect the average information at an average level. In this sense, there were inevitable differences in management practices among farmers;

199 therefore, the use of this average information might result in fertilizer amount errors.
200 In this analysis, the errors in the recorded amount of fertilizer applied was also treated
201 as input uncertainty. Based on our limited local investigation, the initial annual
202 applied urea and compound fertilizer was set as 450kg/ha and 300kg/ha for the
203 potato-sweet potato rotation, while 150kg/ha and 225 kg/ha for the corn system,
204 respectively. A survey conducted by local agricultural administration revealed that the
205 error or standard deviation in the record fertilizer amount was ±5%, which was based
206 on a statistical analysis of historical fertilizer data. Because there was not enough
207 information available regarding the distribution of the fertilizer, normal distribution
208 was used in this study. Using the Monte Carlo technique, these errors were generated
209 by sampling stochastically from a normal distribution expressed as $X \sim N(\mu, \sigma^2)$, where
210 μ and σ are the recorded amount of fertilizer and the standard deviation (*SD*),
211 respectively. The Latin Hypercube sampling technique, which employs a constrained
212 sampling scheme instead of random sampling, was applied to ensure a sufficient
213 precision of sampling. To cover 99.7% of the error range, the sampling range was
214 designated as ±15% from the initial amount of fertilizer and 5,000 model runs were
215 conducted.

216 ## 2.4 Analysis of the model results

217 This study focused on error-propagation from input data to NPS-TP predictions (the
218 sum of organic P and mineral P) at the WX for the period from 2000 to 2007. First,
219 the sensitivity of simulated TP to each input data was quantified in the form of
220 summary statistics, such as the *SD* and the coefficient of variation (*CV*). Specifically,
221 the *CV*, which is a normalized measure of dispersion of a probability distribution, is
222 defined as a dimensionless number by quantifying the ratio of the *SD* to the *MV*.
223 Compared to *SD*, the *CV* is more appropriate for comparing different data sets;
224 therefore, it was used as the main approach for expressing uncertainty in this study.

225 $$s = \sqrt{\frac{1}{m} \sum_{j=1}^{m} \left(x_j - \bar{x} \right)^2} \quad (1)$$

$$c = \frac{s}{\bar{x}} \quad (2)$$

where s and c represents the *SD* and the *CV*, respectively, x_j represents simulated data point j, $\bar{x} = \frac{1}{m}\sum_{j=1}^{m} x_j$ indicates the mean value of simulated data, and m is the number of simulated data.

The E_{NS} was used as the goodness-of-fit indicator to evaluate the model performance.

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n}\left(x_{sim,i} - x_{mea,i}\right)^2}{\sum_{i=1}^{n}\left(x_{mea,i} - \bar{x}_{mea}\right)^2} \quad (3)$$

where $x_{mea,i}$ and $x_{sim,i}$ is the simulated and measured data for the i_{th} pair, respectively, $\bar{x}_{mea}$ represents the mean value of the measured values, and n is the total number of paired values.

In this study, model structure was fixed and model uncertainty will stems predominantly from input errors. Based on the performance ratings by Moriasi et al. (2007), 0.5 was judged as a reasonable E_{NS} value for TP simulation so a threshold of $E_{NS} \geqslant 0.5$ was defined to select acceptable SWAT runs (Liu and Gupta, 2007). In the next step, behavior input data ($E_{NS} \geq 0.5$), which refer to the phenomenon of equifinality and can be representative of a watershed system ($E_{NS} \geq 0.5$), were grouped to express the prediction uncertainty by using a multi-input ensemble method. Finally, input-induced model uncertainty was generated via sampling from the output distributions that are generated from these effective input datasets.

3. Results

3.1 Calibration and validation

As shown in **Table 3**, for the flow simulation, the E_{NS} were 0.66 and 0.89 in the calibration and validation periods, respectively. The E_{NS} values were 0.73 and 0.67 for sediment during the calibration and validation periods, and 0.75 and 0.46 for TP. More details about the final SWAT parameters can be found in our previous studies

(Shen et al., 2012a; Shen et al., 2013a). Compared to the SWAT performances complied by Moriasi et al (2007), the accuracy of flow prediction could be judged as very good, while the sediment and TP simulations were judged to be satisfactory.

3.2 Sensitivity of each input dataset

To determine the sensitivity of each input dataset, the degree of uncertainty of simulated TP was illustrated in **Fig. 2**. As shown in **Fig. 2a**, the annual mean CV ranged from 0.284 (2006) to 0.587 (2003), indicating there were significant uncertainties if only the dataset of single rain gauge was used as model inputs. The E_{NS} values for each rain gauge are 0.70 for XN, 0.49 for LM, 0.39 for TF, 0.38 for SY, 0.31 for WX2, 0.07 for WX, 0.06 for WG, 0.02 for XJB, -0.07 for ZL, -0.12 for CA, -0.68 for GL, and -2.87 for JL. This indicates that most of the E_{NS} values were low, especially for ZL, CA, GL and JL because no rainfall data were recorded in these gauges for the period from 2000 to 2003. These rainfall stations were ranked based on the E_{NS} values, and combinations of m rain gauges (m ranged from 2 to 12) were used as SWAT inputs. As shown in **Fig. 2b**, using data from multiple rain gauges as inputs, the CVs ranged from 0.098 (2006) to 0.433 (2000), suggesting that TP simulations are sensitive to the density of rain gauges. The model performance, in terms of E_{NS}, improved when the number of rain gauges increased from 2 to 5. However, a plateau was reached at approximately 6 gauges.

Using NFGIS DEMs (**Fig. 2c**), the CV values were found to be low with an annual mean CV of 0.026–0.119, but the CV values were higher using ASTER DEMs (**Fig. 2d**), with CV values ranging from 0.105 to 0.383. **Fig. 2e** shows the statistical analysis using different land use maps. Compared to the input data presented above, the annual mean CV values, which ranged from 0.009 to 0.036, were relatively low. Besides, as shown in **Fig. 2f**, the simulated TP showed only slight variation related to the errors in the amount of fertilizer, with mean CV values of 0.003-0.008.

Finally, a multi-input ensemble method was used for a comprehensive evaluation of input-induced model uncertainty. As shown in **Table 4**, the annual CV values of simulated TP ranged from 0.101 to 0.271, indicating a temporal variation for the

period from 2000 to 2007. The ensemble of input-induced outputs was also determined for all six given outlets. As illustrated in **Fig. 3**, the annual mean CV values were 0.190 for XX, 0.088 for DX, 0.206 for HX, 0.162 for BY, 0.168 for WX and 0.135 for CF.

4. Discussion

4.1 Comparison between different input data-induced uncertainty

Table 4 gives a clear comparison between different types of input data. For the given catchment and rainfall characteristics, rainfall input is identified as the most important factor in NPS simulation, whereas rain gauge density is the most important source contributing to the overall uncertainty. The results from the statistical analysis are reasonable as rainfall is the major driving force of runoff generation and therefor the transportation of NPS pollutants (Andréassian et al., 2001; McMillan et al., 2011). As shown in **Table 1**, rainfall data varied substantially among different gauges, with a 933-mm difference between the highest and lowest annual rainfalls. This finding agrees with previous research (Strauch et al., 2013) in which the rainfall input was averaged across the watershed by a single rain gauge, but failed to adequately reflect spatial rainfall variations. This can be attributed to the SWAT rule for quantifying the sub-watershed rainfall, in which rainfall data from the closest gauge is selected as inputs for each sub-watershed. In cases where a sub-watershed contains no rain gauges, the centroid is used to find the nearest gauge and its data are substituted for the sub-watershed rainfall. Another reason might be the use of the same parameter set in all simulations. Bardossy and Das (2008) found that fewer gauge simulations might produce similar results when compared with those obtained by more rain gauges due to the compensation effect from calibration. If the model had been re-calibrated to each perturbed input set, the calibrated parameters would likely have compensated somewhat for the perturbed inputs in an effort to reproduce the observed data. However, even with the best calibration process, there is always parameter uncertainty in the model predictions due to the imprecise representation of parameter ranges and distributions; therefore, recalibration was not conducted in this study (Van

309 Griensven et al., 2006). It should be noted that comparison using un-recalibrated
310 models is useful to evaluate the differences in model predictions because calibration
311 masks the differences that may occur as a result of the input data sets. In addition, the
312 un-recalibrated model results can show how good each dataset predicts stream flow
313 before calibration, which would indicate the effort required for calibration when using
314 each data set.
315 **Fig. 2b** illustrates that there were reductions in the CV values compared with the
316 single-gauge simulations, which clearly showed that the ensemble of multi-gauge
317 simulations outperformed the single-gauge simulations. However, no clear
318 relationship existed between the E_{NS} values and the rain gauge location, which is also
319 inconsistent with a previous study. Schuurmans and Bierkens (2007) found greater
320 model errors if gauges outside the watershed were used, but this is not the case for the
321 present study because the outside gauges were relatively close (10 km) to the
322 watershed boundary. **Fig. 2b** indicates that the use of these key gauges appear to be
323 more informative in constraining spatial rainfall variations but simulation efficiency
324 did not always improve when additional gauges are added. This demonstrates that the
325 information content in rainfall spatial variation is reached after a relatively small
326 number of key gauges are used as model input (Seibert and Beven, 2009). It is
327 encouraging that a small number of gauges distributed more optimally and perform
328 well for logistical reasons (Bárdossy and Das, 2008; McMillan et al., 2011). In reality,
329 there might not be many dense rain gauge networks similar to those used for this
330 study; therefore, the fact that spatial rainfall variation is a function of key gauges
331 rather than all gauges would indicate a wider range of applicability. For this study
332 area (2,421 km^2), the optimal number of gauges were identified as 6 beyond which
333 improvements to the model predictions would not be found.
334 As illustrated in **Fig. 2c** and **2d**, the second highest uncertainty was caused by DEMs,
335 and the ASTER GDEM-induced uncertainty was higher than by uncertainty induced
336 by NFGIS DEM. These higher values could be due to the following two reasons: first,
337 NFGIS DEM was already validated in many places in China, which was not the case
338 for ASTER GDEM (Wu et al., 2007; Dixon and Earls, 2009). In fact, ASTER GDEM

contains systematic errors; i.e., a significant number of anomalies attributable to cloud disturbances, the algorithm used to generate the final GDEM, and not applying inland water mask. Second, the initial resolution of NFGIS DEM (90*90m) was lower than that of ASTER GDEM (30*30m). In reality, those high resolution DEMs might provide better simulations, but sometimes a moderate one would be more suitable due to the nonlinearity of erosion processes and its subsequent effect on P processes (Chaplot et al., 2005a). Given the nature of ASTER GDEM, the greater degree of averaging has occurred by adding shallower slopes, and the predicted TP would be lower by increasing more infiltration and deposition of NPS-TP. In this sense, it is important to select an appropriate data source because DEMs are generated at different scales and a number of the implied watershed processes are scale-dependent (Brazier et al., 2005). Care must be taken in DEMs data resolution because their resolutions cannot be up-scaled directly. In theory, topography exerts some level of control on surface flow and thus NPS loads. Therefore, the smoothing of the landscape shape induced by coarser DEMs could result in a biased estimation of TP outputs (Dixon and Earls, 2009). It was worthwhile to parameterize the SWAT model with the extreme slopes, as these slopes controlled the fluxes of NPS-TP. However, our previous study has also demonstrated that the TP simulations would not be improved if certain resolution was reached (Shen et al., 2013a). In this sense, some balance must be found between improving the DEMs resolutions and reducing the complexity of the model utility.

In contrast, land use maps and fertilizer amount resulted in low uncertainties. The result differ from those of Payraudeau et al. (2004), who found that model outputs were highly sensitive to land use changes. This could be explained by the fact that most agricultural land was redistributed to forest and other land uses in the study of Payraudeau et al. (2004), which leads to significant changes in soil compaction and ground cover. However, these low values in our study could be due to minor land use changes during the period from the 1980s to 2007. As shown in **Table 2**, the fraction of forest area decreased gradually from 61.75% to 54.76%, whereas agricultural land increased from 25.68% to 33.47%. **Fig. 2f** indicates that the fertilizer input has only a

slight impact on in-stream TP loads. This was because P application was low in this watershed with the inorganic N being applied in greater amounts and more widely. Additionally, the major forms of P in mineral soils are plant-available soluble P, insoluble forms of mineral P and organic P. According to the mechanism of the SWAT model, P would be taken up firstly by plant uptake and then by erosion, and these processes would govern the turnover rates and transport of P (Arnold et al., 1998). Therefore, only a small proportion of P will finally flow into the water body as in-stream NPS-TP. In this sense, there might also be minor CV values if other representative attribute practices, e.g., tillage data, were selected. This indicates the degree of sensitivity due to single input data depends on two factors: the ratio of each individual input contribution to the total load (which is the case for management data) and the error in the individual input (which is more meaningful for land use maps).

4.2 Comprehensive evaluation of input data-induced uncertainty

As shown in **Fig. 3**, this demonstrated that input-induced uncertainty may be highly area-specific; i.e., dependent upon the scale of the drainage area and rainfall variability. For example, when multiple gauges (from 1 to 12) are used as model inputs, the simulated TP remained stable for the DX and no model uncertainty was observed. This could be due to the mechanism of SWAT, in which only the rainfall data from the closest gauge to the centroid were chosen and used as the sole model input for that specific sub-watershed. As shown in **Fig. 1**, there is only one sub-watershed in the DX region and the XN gauge is closest to its centroid; therefore, the rainfall data from the same gauge was used every time for this region. However, the CV values remained high for other outlets, ranging from 0.187 (CF)–0.448 (XX), suggesting that rain gauge density indicated different impacts under different spatial scales of drainage areas. In addition, using different DEM data, the CV values were relatively low for XX, DX, WX and CF, with an annual mean CV of 0.022–0.055, but the CV values were relatively high for HX and BY, with values of 0.152 and 0.136, respectively. This could be explained by the fact that there are more mountainous areas along XX, DX, WX and CF; therefore, the generated topography in these

398 regions, such as the watershed boundary, surface slope and other characteristic
399 parameters, could be extracted more easily by DEM data.
400 These results pose two significant scientific challenges for TMDLs. First, as model
401 uncertainty is difficult to quantify, the margin of safety (MOS) was often arbitrarily
402 assumed as 10% error. However, as shown in **Table 4**, this assumption is not highly
403 related to the reliability of the model system and supported the quantification of
404 TMDLs poorly. Specifically, ccompare to our previous studies (Shen et al., 2012b),
405 the uncertainties caused by input errors were greater than those resulting from model
406 parameters in 2001, 2005, and 2007, whereas uncertainties caused by inputs were
407 lower in the remaining years. Overall, the mean CV (0.168) for input-induced TP
408 uncertainty was slightly higher than that (0.156) for the parameter uncertainty, which
409 agrees with previous studies (Kuczera et al., 2006). Therefore, input data uncertainty
410 is critical in NPS modeling and efforts should be made to clarify this type of
411 uncertainty. Second, as illustrated in **Fig. 3**, the input data-induced uncertainty varies
412 considerably temporally and spatially due to the varying climate, underlying
413 topography, land use, soil type, and management (Shen and Zhao, 2010; Chen et al.,
414 2012). In this sense, a site-specific MOS, which might be more robust to any
415 particular sequence of input errors than current steady MOS, should be defined as a
416 priori.

417 ## 5. Conclusions

418 In this research, the impacts of four different input data types, including rainfall data,
419 DEMs, land use maps, and amount of fertilizer, on NPS modeling were quantified and
420 compared. Based on the results, input data-induced uncertainty is critical in NPS
421 modeling and efforts should be made to decrease this type of uncertainty. For the case
422 study, the mean CV value ranged from 0.101 to 0.271, which is slightly higher than
423 that for the parameter uncertainty. The study indicated that rainfall input resulted in
424 the highest uncertainty, followed by DEM, land use maps, and fertilizer amount.
425 Therefore, measures should be taken first to reduce this source of uncertainty by
426 adding rain gauges, modifying the selection mechanism of rain gauge in SWAT, and

using appropriate interpolation techniques. This paper also demonstrated the required input information would be reached if several key rain gauges and moderate-resolution DEMs are used. This paper provides valuable information for developing TMDLs in the Three Gorges Reservoir Area, and these results are also valuable to other model-based watershed studies for the control of model uncertainty. However, this conclusion might be only appropriate for NPS-TP and not for other pollutants, i.e., the generation and transportation of nitrogen differ substantially from those of NPS-P. Furthermore, the influence of soil type maps was not analyzed, because only one coarse soil map was available for the study region. More researches are needed if detailed input data sets are collected.

Author contribution

Z. Shen designed the experiments. L. Chen and Y. Gong developed the SWAT model and performed the simulations. L. Chen prepared the manuscript with contributions from all co-authors.

Data availability

The data could be obtained by emailing the first author.

Acknowledgements

This project was supported by the Fund for Innovative Research Group of the National Natural Science Foundation of China (Grant No. 51421065), the National Natural Science Foundation of China (No. 51409003 & 51579011), and Project funded by China Postdoctoral Science Foundation.

References

Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., and Srinivasan, R.: Modelling hydrology and water quality in the

451 pre-ailpine/alpine Thur watershed using SWAT, J. Hydro., 333, 413-430, 2007.
452 Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., and Lavabre, J.: Impact of
453 imperfect rainfall knowledge on the efficiency and the parameters of watershed
454 models, J. Hydro., 250, 206-223, 2001.
455 Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R.: Large area hydrologic
456 modeling and assessment - Part 1: Model development, J. Am. Water Resour. As.,
457 34, 73-89, 1998.
458 Balin, D., Lee, H., and Rode, M.: Is point uncertain rainfall likely to have a great
459 impact on distributed complex hydrological modeling?, Water Resour. Res., 46,
460 W11520, 2010.
461 Bárdossy, A., and Das, T.: Influence of rainfall observation network on model
462 calibration and application, Hydro. Earth Sys. Sci., 12, 77-89, 2008.
463 Beven K.: A manifesto for the equifinality thesis, J. Hydro., 320, 18-36, 2006.
464 Brazier, R.E., Heathwaite, A.L., and Liu, S.: Scaling issues relating to phosphorus
465 transfer from land to water in agricultural catchments, J. Hydro., 304, 330-342,
466 2005.
467 Chaplot, V., Saleh, A., and Jaynes, D.B.: Effect of the accuracy of spatial rainfall
468 information on the modeling of water, sediment, and NO3-N loads at the
469 watershed level, J. Hydro., 312, 223-234, 2005a.
470 Chaplot, V.: Impact of DEM mesh size and soil map scale on SWAT runoff, sediment,
471 and NO3-N loads predictions, J. Hydrol. 312, 207−222, 2005b.
472 Chaubey, I., Cotter, A., Costello, T., and Soerens, T.: Effect of DEM data resolution on
473 SWAT output uncertainty, Hydro. Proccess., 19, 621-628, 2005.
474 Chen, D., Dahlgren, R.A., Shen, Y., and Lu, J.: A Bayesian approach for calculating
475 variable total maximum daily loads and uncertainty assessment, Sci. Total
476 Environ., 430, 59-67, 2012.
477 Chen, L., Zhong, Y., Wei, G., Cai, Y., and Shen, Z.: Development of an integrated
478 modeling approach for identifying multilevel non-point-source priority
479 management areas at the watershed scale. Water Resour. Res., 50, 4095-4109,
480 2014.

481 Cotter, A., Chaubey, I., Costello, T., Soerens, T., and Nelson, M.: Water quality model
482 output uncertainty as affected by spatial resolution of input data, J. Am. Water
483 Resour. Ass., 39, 977-986, 2003.
484 Dixon, B., and Earls, J.: Resample or not?! Effects of resolution of DEMs in
485 watershed modeling, Hydro. Process., 23, 1714-1724, 2009.
486 Duncan, M., Austin Bfabry, F., and Austin, G.: The effect of gauge sampling density
487 on the accuracy of streamflow prediction for rural catchments, J. Hydro., 142,
488 445-476, 1993.
489 Gassman, P., Reyes, M., Green, C., Arnold, J.: The soil and water assessment tool:
490 Historical development, applications, and future research directions, Trans.
491 ASABE 50 (4), 1211-1250, 2007.
492 Han, J.C., Huan, G.H., Zhang, H., Li, Z., and Li, Y.P.: Bayesian uncertainty analysis in
493 hydrological modeling associated with watershed subdivision level: a case study
494 of SLURP model applied to the Xiangxi River watershed, China, Stoch. Environ.
495 Res. Risk. A., 28, 973-989, 2014.
496 Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error
497 analysis of conceptual rainfall-runoff models: Characterising model error using
498 storm-dependent parameters, J. Hydro., 331, 161-177, 2006.
499 Lin, S., Jing, C., Coles, N.A., Moore, N., and Wu, J.: Evaluating DEM source and
500 resolution uncertainties in the Soil and Water Assessment Tool, Stocha. Env. Res.
501 Risk Assess., 27, 209-221, 2013.
502 Liu, Y., and Gupta, H.: Uncertainty in hydrologic modeling: Toward an integrated data
503 assimilation framework, Water Resour. Res., 43, W07401, 2007.
504 Mango, L., Melesse, A., McClain, M., Gann, D., and Setegn, S.: Land use and climate
505 change impacts on the hydrology of the upper Mara River Basin, Kenya: Results
506 of a modeling study to support better resource management, Hydro. Earch Sys.
507 Sci., 15, 2245-2258, 2011.
508 McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall
509 uncertainty in hydrological modelling: An evaluation of multiplicative error
510 models, J. Hydro., 400, 83-94, 2011.

511 Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith,
512 T.L.: Model evaluation guidelines for systematic quantification of accuracy in
513 watershed simulations, Trans. ASABE, 50, 885-900, 2007.
514 Munoz-Carpena, R., Vellidis, G., Shirmohammadi, A., and Wallender, W.W.:
515 Evaluation of modeling tools for TMDL development and implementation, Trans.
516 ASABE, 49, 961-965, 2006.
517 Pai, N., and Saraswat, D.: Impact of land use and land cover categorical uncertainty
518 on SWAT hydrologic modeling, Trans. ASABE, 56, 1387-1397, 2013.
519 Payraudeau, S., Cernesson, F., Tournoud, M.G., and Beven, K.J.: Modelling nitrogen
520 loads at the catchment scale under the influence of land use, Phy. Chem. Earth,
521 29, 811-819, 2004.
522 Schuurmans, J.M., and Bierkens, M.F.: Effect of spatial distribution of daily rainfall
523 on interior catchment response of a distributed hydrological model, Hydro. Earth
524 Sys. Sci., 11, 677-693, 2007.
525 Seibert, J., and Beven, K.J.: Gauging the ungauged basin: how many discharge
526 measurements are needed? Hydro. Earth Sys. Sci., 13, 883-892, 2007.
527 Shen J., and Zhao Y.: Combined Bayesian statistics and load duration curve method
528 for bacteria nonpoint source loading estimation, Water Res., 44, 77-84, 2010.
529 Shen, Z.Y., Chen, L., Chen, T.: Analysis of parameter uncertainty in hydrological and
530 sediment modeling using GLUE method: a case study of SWAT model applied to
531 Three Gorges Reservoir Region, China. Hydro. Earch Sys. Sci., 16, 121-132,
532 2012b.
533 Shen, Z.Y., Chen, L., Hong, Q., Ding, X.W., and Liu, R.M.: Assessment of nitrogen
534 and phosphorus loads and causal factors from different land use and soil types in
535 the Three Gorges Reservoir Area, Sci. Total Environ., 454-455: 383-392, 2013b.
536 Shen, Z.Y., Chen, L., Liao, Q., Liu, R.M., and Hong, Q.: Impact of spatial rainfall
537 variability on hydrology and nonpoint source pollution modeling, J. Hydro., 472,
538 205-215, 2012a.
539 Shen, Z.Y., Chen, L., Liao, Q., Liu, R.M., and Huang, Q.: A comprehensive study of
540 the effect of GIS data on hydrology and non-point source pollution modeling,

Agr. Water Manage., 118, 93-102, 2013a.

Strauch, M., Bernhofer, C., Koide, S., Volk, M., Lorz, C., and Makeschin, F.: Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation, J. Hydro., 414-415, 413-424, 2012.

Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, A., and Srinivasan, R.: A global sensitivity analysis tool for the parameters of multi-variable catchment models, J. Hydro., 324, 10-23, 2006.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09, 2008.

Wellen, C., Arhonditsis, G.B., Labencki, T., and Boyd, D.: Application of the SPARROW model in watersheds with limited information: a Bayesian assessment of the model uncertainty and the value of additional monitoring, Hydro. Process., 28, 1260-1283, 2014.

Wellen, C., Kamran-Disfani, A., and Arhonditsis, G.B.: Evaluation of the current state of distributed watershed-water quality modeling, Environ. Sci. Technol. 49: 3278-3290, 2015.

Wu, S., Li, J., and Huang, G.H.: An evaluation of grid size uncertainty in empirical soil loss modeling with digital elevation models, Environ. Model Assess., 10, 33-42, 2005.

Wu, S., Li, J., and Huang, G.H.: Modeling the effects of elevation data resolution on the performance of topography-based watershed runoff simulation, Environ. Model Softw. 22, 1250-1260, 2007.

Xue, C., Chen, B., and Wu, H.: Parameter Uncertainty Analysis of Surface Flow and Sediment Yield in the Huolin Basin, China, J. Hydrol. Eng., 19, 1224-1236, 2014.

Zhang, P., Liu, R., Bao, Y., Yu, W., and Shen, Z.: Uncertainty of SWAT model at different DEM resolutions in a large mountainous watershed, Water Res., 53. 132-144, 2014.

571 Zheng, Y., Wang, W., Han, F., and Ping, J.: Uncertainty assessment for watershed
572 water quality modeling: A Probabilistic Collocation Method based approach, Adv.
573 Water Resour., 34, 887-898, 2011.
574

575 Table 1 The recorded annual mean rainfall data for each rain gauge (2000–2007)

Rain gauge	JL	GL	WG	TF	ZL	SY	CA	LM	XN	WX	WX2	XJB
MV/mm	1938	1648	1609	1416	1406	1358	1279	1255	1193	1079	1055	1005
SD/mm	445	334	309	260	357	235	222	243	264	235	269	180

576 MV indicates the mean value and SD represents the standard deviation

577

578 Table 2 The fraction of land use types within the watershed for different periods

Land use	1980s		1995		2000		2007	
	Area (km^2)	Percent (%)	Area (km^2)	Percent (%)	Area (km^2)	Percent (%)	Area (km^2)	Percent (%)
Farm land	622.5	25.68%	588.3	24.27%	613.3	25.30%	811.1	33.47%
Forest	1496.8	61.75%	1564.8	64.56%	1498.1	61.80%	1327.1	54.76%
Grass land	294.5	12.15%	261.5	10.79%	302.0	12.46%	267.2	11.02%
Water	8.9	0.37%	8.7	0.36%	8.9	0.37%	11.9	0.49%
Residential area	1.1	0.05%	0.6	0.02%	1.7	0.07%	6.3	0.26%

579

580

581 Table 3 The values of E_{NS} and R^2 of the SWAT model during the calibration and
582 validation period

Variable	Indicator	Calibration	Validation
Flow	E_{NS}	0.66	0.89
	R^2	0.79	0.95
Sediment	E_{NS}	0.73	0.67
	R^2	0.83	0.83
TP	E_{NS}	0.75	0.46
	R^2	0.86	0.79

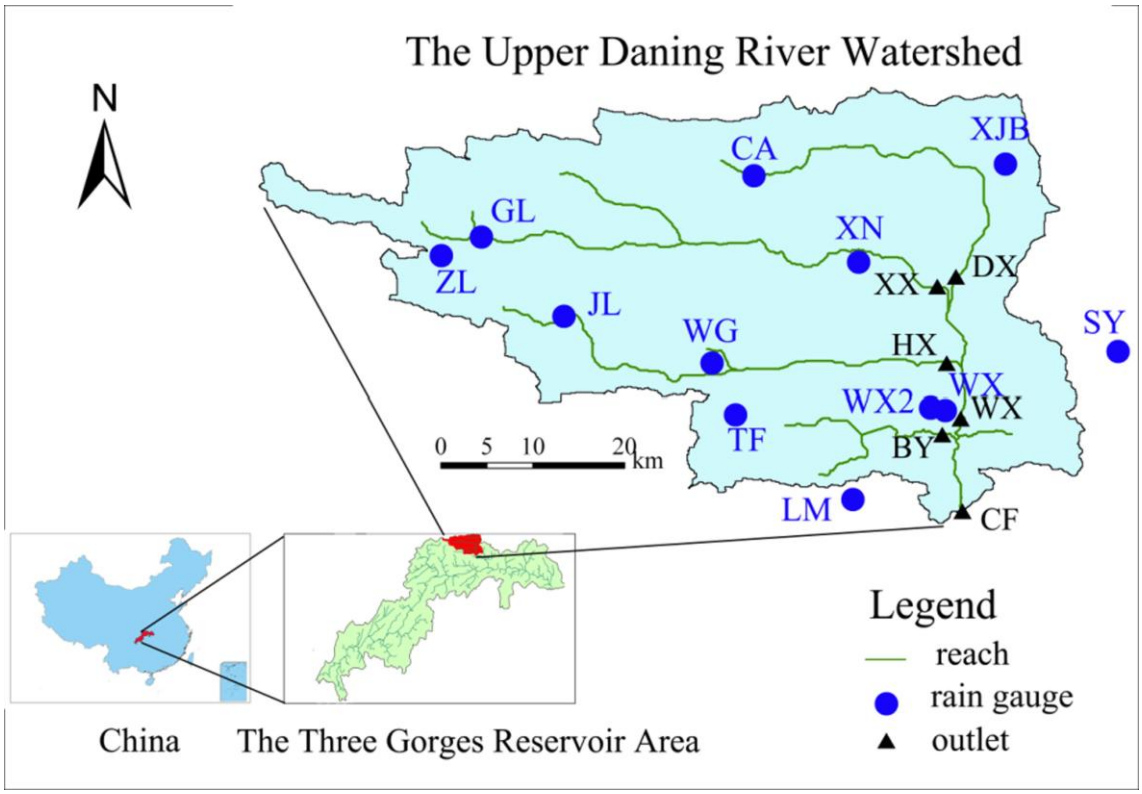
583

584 Table 4 The sensitivity of simulated TP (CV values) to different input dataset

Input data	2000	2001	2002	2003	2004	2005	2006	2007	Mean
Single gauge	0.419	0.421	0.332	0.587	0.319	0.417	0.284	0.410	0.388
Multi-gauges	0.433	0.362	0.240	0.287	0.141	0.256	0.098	0.241	0.249
NFGIS DEM	0.026	0.119	0.059	0.025	0.026	0.043	0.105	0.040	0.056
ASTER GDEM	0.189	0.276	0.225	0.105	0.198	0.255	0.383	0.274	0.197
Land use maps	0.022	0.013	0.018	0.018	0.024	0.036	0.009	0.024	0.027
Fertilizer amount	0.004	0.003	0.003	0.003	0.006	0.007	0.003	0.005	0.005
Input uncertainty	0.151	0.208	0.116	0.101	0.112	0.271	0.141	0.246	0.168
Parameter uncertainty	0.167	0.145	0.177	0.141	0.147	0.151	0.154	0.164	0.156

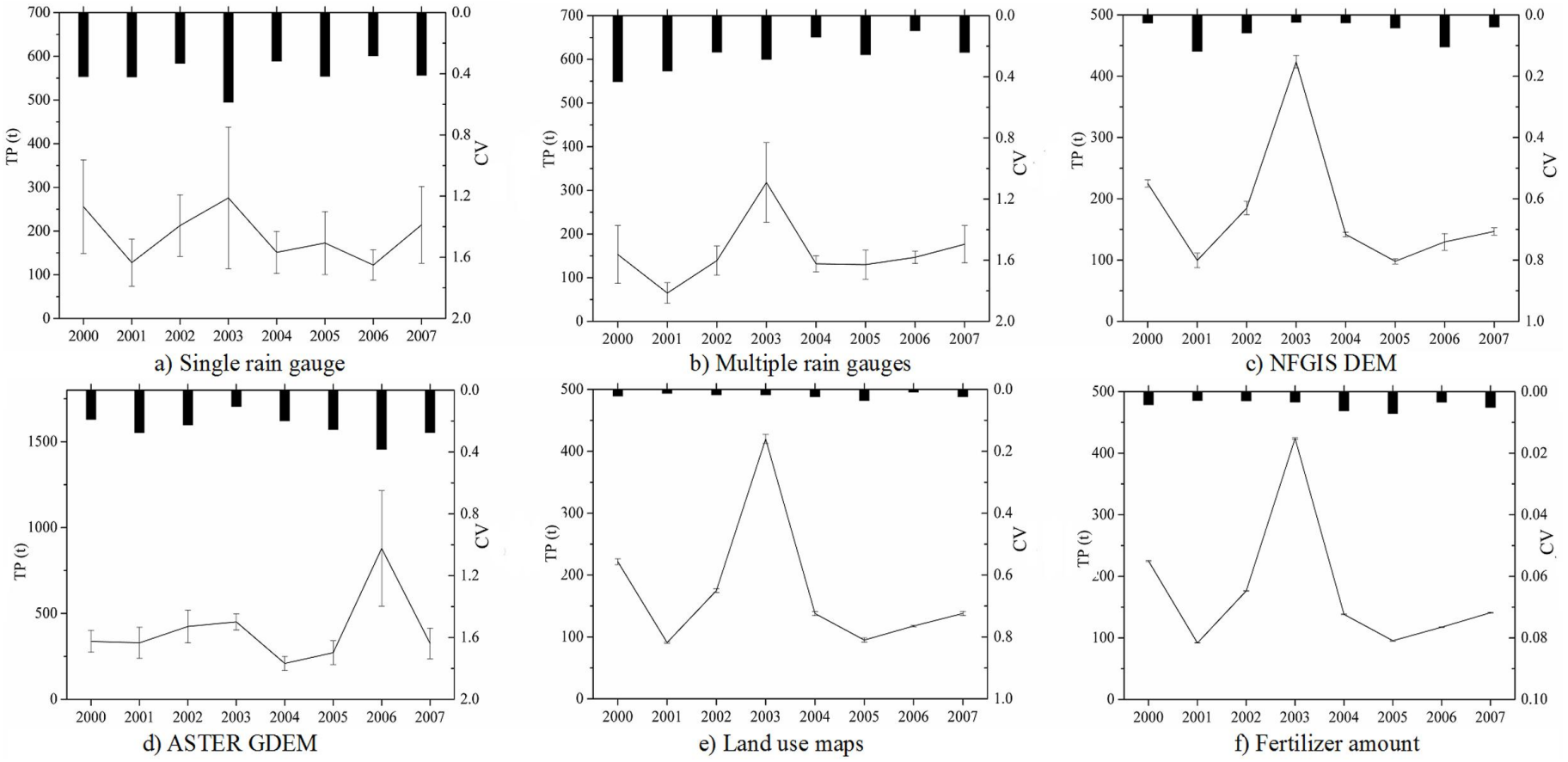
585

586



587

588 Fig. 1 Locations of and the rain gauges within the Upper Daning River Watershed



589

590 Fig. 2 Uncertainty of simulated TP induced by each input data, in which the line, error bar and inverted column indicate the mean value, SD and

591 CV values, respectively.

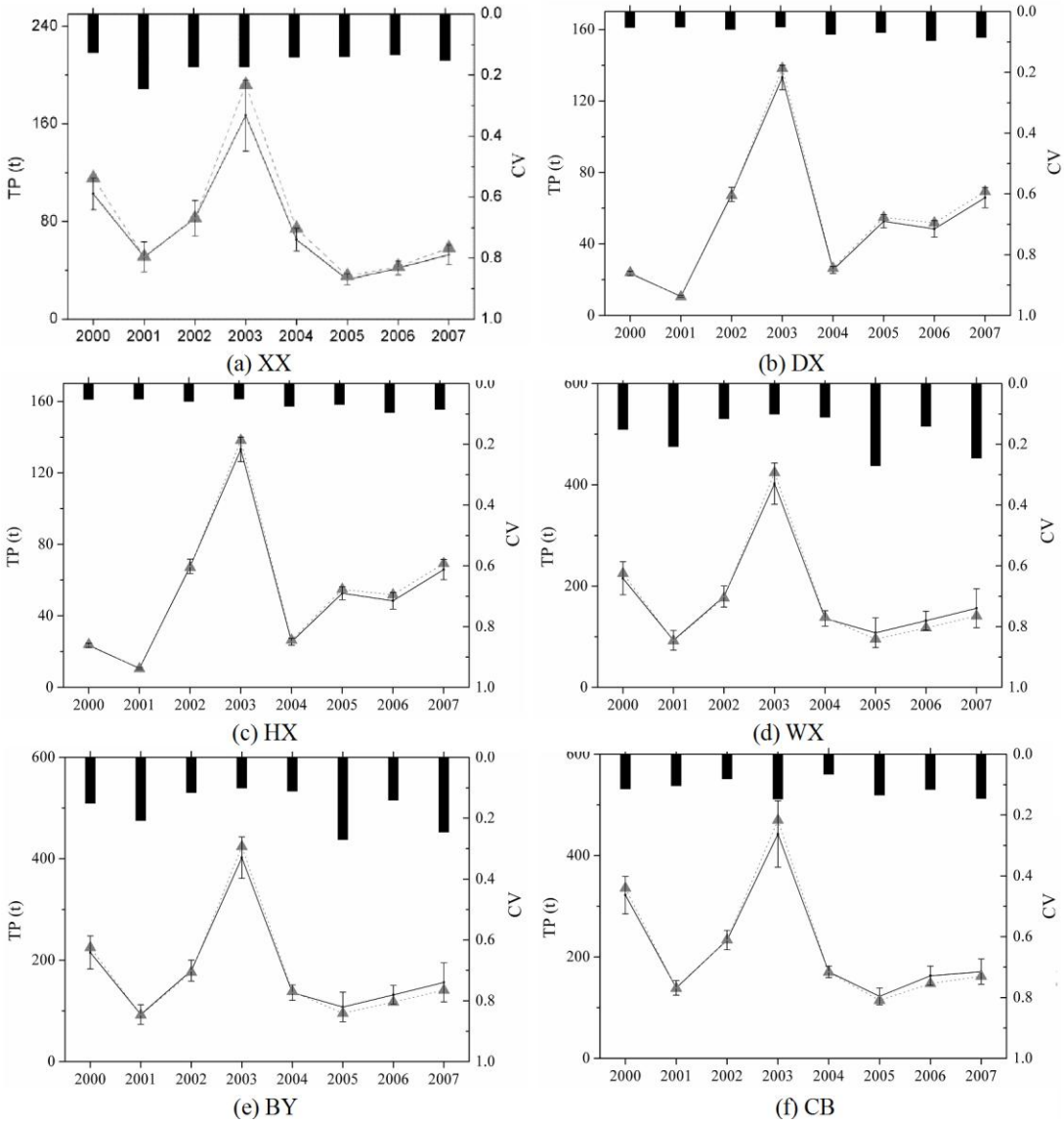


Fig. 3 Comprehensive uncertainty of input data-induced simulated TP, in which the line, error bar and inverted column indicate the mean value, SD and CV values, respectively.