Hydrol. Earth Syst. Sci. Discuss., 12, C6449–C6466, 2016
www.hydrol-earth-syst-sci-discuss.net/12/C6449/2016/

Hydrology and Earth System Sciences Discussions
Open Access

Interactive comment on "A comprehensive evaluation of input data-induced uncertainty in nonpoint source pollution modeling" *by* L. Chen et al.

L. Chen et al.

zyshen@bnu.edu.cn

Received and published: 31 January 2016

Dear Reviewer Thank you very much for your comment of January 8, 2016, informing us of valuable suggestions to improve our manuscript 'A comprehensive evaluation of input data-induced uncertainty in nonpoint source pollution modeling' (hess-2015-377). Our point-by-point responses are as follows. ==

Response to the Reviewer 1 (1) Your comment: The methodology in this paper is not very well described so I'm not entirely sure which analyses were carried out and what the motivations for these were exactly. Our respond: Thank you very much for your

valuable suggestions. I agree with the reviewer's idea that the methodology in previous version of our paper is not very well described. As suggested, we have checked the manuscript very carefully, while revised the methodology accordingly. Please find the attached manuscript.

(2) Your comment: It seems clear that the study doesn't go substantively beyond the authors' previous papers (Shen et al. 2012a, 2013). Our respond: I think the primary importance of this paper is its complete framework and detailed comparisons between different input datasets. Commonly, input data, which occurs because of variation in natural conditions, limitations in measurement, as well as a shortage of data, plays a role as a driving force of parameter calibration and may transform into a doubtful output. Therefore, imperfect knowledge of model input should, theoretically, be tackled comprehensively, and this has not been conducted yet by authors' previous papers (Shen et al. 2012a, 2013). The input uncertainty is very hard to quantify as various input datasets are required for the H/NPS models. Typically, the input datasets should include spatial data, which can be expressed as GIS maps and their linked attribute properties. As you mentioned, some important issues have been discussed in our previous papers (Shen et al. 2012a, 2013). However, due to the availability of data at that time, a lot of important issues have not been studied yet and we believe that further theoretical analysis and practical steps are needed when more data are available (these were mentioned in the conclusion section of Shen et al. (2012a, 2013)). Thus, we have been struggling to conduct a comprehensive assessment of input-induced uncertainty in TP modeling. After two-years further investigations, more datasets have been collected and four key types of input data, i.e., rainfall, topography, land use and fertilizer amount, are analysed, and their uncertainties are quantified. We believe this paper would provide 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. Specifically, the detailed differences between this paper and current previous papers (Shen et al. 2012a, 2013) are as follows: 1) In Shen et al. (2012a), we focused on the impact of rainfall spatial variability by means of

a variety of interpolation methods as follows: the Centroid method; the Thiessen Polygon method; the IDW method; the Disjunctive Kriging (Dis-Kriging) method, and the Co-Kriging method. I agree with the reviewer's idea that rainfall uncertainty might stem predominantly from the representativeness issues. Thus, we further focused on the representativeness of rainfall stations in this paper, while single-gauge and multiple-gauges simulations were conducted. 2) In Shen et al. (2013), we focused on the prediction uncertainty related to the combination of the available DEMs and land use maps by grouping the available GIS maps in all possible ways. However, due to the availability of data, only ASTER GDEM and land use data of 2007 were used at that time. In this study, two DEM sets, in terms of NFGIS DEM and ASTER GDEM were used and compared, while land use data were obtained from the 1980s (1980–1989), 1995, 2000, and 2007. Besides, only the impact of data resolution was studied in Shen et al. (2013). As GIS data may be available from alternative sources; therefore, another question is which specific data set should be used. Thus, we further studied the impacts of different data sources. 3) Besides, previous papers (Shen et al. 2012a, 2013) have focused on impacts of spatial data (GIS maps) but in this paper, we further focused on the impacts of attribute data.

(3) Your comment: The introduction and discussion lack references to fundamental papers on uncertainty in water quality modelling and input uncertainty propagation more generally. The authors focus largely on the SWAT literature, but should go beyond this to demonstrate sufficient grounding in the state-of-the-art of model uncertainty estimation. Our respond: Thank you for this valuable suggestion. I agree with the reviewer's idea that the introduction and discussion lack related references. In fact, we did not give more literatures about input uncertainty because they have already mentioned in our previous papers (Shen et al. 2012a, 2013). As suggested, we read more published references and picked up those ones related to rainfall, topography, land use and fertilizer amount. These references are added into their related sentences and we rewrote most sentences of the input uncertainty in a more concise way. Please find the attached manuscript. Specifically, the following references have been added in

the revised manuscript: Balin, D., Lee, H., and Rode, M.: Is point uncertain rainfall likely to have a great impact on distributed complex hydrological modeling?, Water Resour. Res., 46, W11520, 2010 Beven, K.: A manifesto for the equifinality thesis. J. Hydrol. 320 (1−2), 18−36 2006 Chaplot, V.: Impact of DEM mesh size and soil map scale on SWAT runoff, sediment, and NO3-N loads predictions, J. Hydrol. 312, 207−222, 2005 Chaplot, V., Saleh, A., Jaynes, D. B.: Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO3−N loads at the watershed level, J. Hydrol. 312, 223−234, 2005 Cibin, R., Sudheer, K. P., Chaubey, I.: Sensitivity and identifiability of stream flow generation parameters of the SWAT model. Hydrol. Process. 24, 1133−1148, 2010 Gassman, P., Reyes, M., Green, C., Arnold, J.: The soil and water assessment tool: Historical development, applications, and future research directions, Trans. ASABE 50 (4), 1211−1250, 2007 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., 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

(4) Your comment: The authors fail to distinguish between measurement uncertainty and calibration uncertainty and make confusing statements: "due to the availability of validation data" Our respond: Sorry for those confusing statements and these sentences have been revised to benefit our readers. In fact, we have distinguished measurement uncertainty and calibration uncertainty in one previous study (Chen et al., 2014). Based on our works, we found appreciable inherent errors exist in the measured data even when following strict quality assurance and quality control (QA/QC) guidelines. Harmel (2005) has mentioned measurement uncertainty may stem from errors in flow measurements, water quality sample collection, the processes of preservation, storage, transport and laboratory analysis. Given the river discharge data, errors from different sources such as river stage measurement or the interpolation of the rating curve, affect the measured data. In a thorough review (Harmel et al., 2006),

all possible errors in the H/WQ measured data were compiled. In comparison, model calibration is the process of estimating model parameters using a pair-wise comparison between the predicted and measured points, while the validation process involves running the well-calibrated model to check its performance. In traditional applications, model evaluation is usually conducted by a regression measure, most commonly the point-to-point pairs of predicted and measured data. The multi-objective functions are also used to calibrate models by the (predicted) point-to- (measured) point comparison to simulate the variety of hydrological events that may occur in a basin. Nevertheless, parameter identification is a complex non-linear problem, so the parameters could not be identified easily due to the numerous possible solutions obtained by optimization algorithms. Additionally, different parameter sets may result in similar predictions in a phenomenon known as equifinality. Thus, we have designed an interval-deviation approach (IDA) and incorporated it into likelihood functions with the support of interval theory (Chen et al., 2014). The proposed IDA was validated in a real application of the Soil and Water Assessment Tool (SWAT) and Generalized Likelihood Uncertainty Estimation (GLUE) in the Three Gorges Reservoir Area (TGRA), China. Compared with the traditional point estimates of observations and predictions, the IDA incorporated both prediction and measurement uncertainty into the process of model evaluation. This proposed IDA can be extrapolated into other forms of error indices and model to provide a substitute method of model evaluation within an uncertainty framework.

(5) Your comment: (P11423, L2), "simplification in natural randomness" Our respond: As suggested, P11423, L2 has been revised as "simplification in natural randomness".

(6) Your comment: (P11423, L4). The authors fail to distinguish between input uncertainty propagation and parameter uncertainty amplified through calibration with uncertain input data. The authors focus on the former which is well studied whereas the latter, despite some studies in the last decade, is rather new and much more interesting. Failing to discuss this weakens section 4.2 considerably. Our respond: Thank you for this valuable suggestions. In this study, input uncertainty propagation was used

to explain why the uncertainty of hydrology modeling would be transformed into even larger uncertainty of NPS modeling. Due to the mechanism of runoff–NPS formulation in the SWAT model, input uncertainty propagation could be due to the fact that sediment production was affected and dominated by hydrological processes as well as watershed response. For example, the impact of rainfall input data on NPS simulation was complex, as it depended not only on the degree of input error but also on watershed properties such as soils, geology, and river morphology. A similar result was obtained via TP simulation. In comparison, the cause of parameter uncertainty may be due to the value of the parameter being case-specific. The process of parameter calibration is a complex and subjective task. For these reasons, decisions regarding modeling should be based on available knowledge about the range and associated probability distribution function (PDF) for each parameter. In this sense, models are not re-calibrated to show the differences in model predictions because calibration masks the differences that may occur as a result of different input datasets. In addition, the un-recalibrated model results can show how good each dataset predicts stream flow and NPS before recalibration, which would indicate the effort required for calibration when using each dataset. We agree with the reviewer's idea that parameter uncertainty amplified through the calibration process is more interesting. In fact, we did recalibrate the SWAT when different datasets were used. Compared to the result of the un-recalibrated model, the relative error between predicted and observed data become smaller after recalibration, while ENS and R2 increased slightly. This can be due to the compensation mechanism of calibration process. Besides, we have been conducting researches on the interaction between soil input error propagation and parameter uncertainty amplified through calibration with uncertain input data.

(7) Your comment: Separate paragraphs should introduce land use and fertilizer data uncertainty as done for rainfall and topography. Our respond: As suggested, paragraphs about land use and fertilizer data have been added. About land use, we added: "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 pe-

riod 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." About fertilizer data, we added "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."

(8) Your comment: P11422, L2 and elsewhere: Do you mean "propagation" instead of "transitivity"? Our respond: As suggested, "transitivity" has been replaced by "propagation" in P11422, L2 and elsewhere.

(9) Your comment: P11422, L9-12: The meaning of the coefficient of variation is not clear in this context. Our respond: As suggested, P11422, L9-12 has been revised as "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."

(10) Your comment: P11422, L21: I wouldn't consider hydrological models "essential" for the stated purposes. Our respond: I agree that "essential" is a very strong word. As suggested, P11422, L21 has been revised as "important tools".

(11) Your comment: P11422, L22: The named models cannot explain anything about water quality deterioration as they don't have ecological impact components. Our respond: I agree the named models don't have ecological impact components but I think they can explain about water quality changes. One of the most outstanding achievements of last three decades is the development of hydrology and nonpoint source (H/NPS) models, which have enable hydrologists and environment expert to under-

stand human activities that affect basin systems. For example, the SWAT model is a physically-based, semi-distributed H/NPS model operated on a daily time step that simulates surface runoff, sediment, and nutrient and pesticide transport primarily from the watershed. The basic components of SWAT include hydrology, soil erosion, nutrient and pesticide leaching, crop growth, agricultural management and the generation of weather data. Specifically, a plant-growth component, which describes the sediment removal and water transpiration from the root zone, has been also incorporated based on heat units. The Modified Universal Soil Loss Equation (MUSLE) is used for estimating sediment yield. Those major forms of P, such as soluble P, insoluble mineral P and organic P associated with humus, are considered, while their movements and transformations are simulated. The soluble P loads would then be estimated based on the simulated runoff volume, the labile P concentration of the top 10mm soil, and a site-specific partitioning factor. The attached mineral and organic P is calculated based on the amount of eroded sediment. These soluble and attached P would be transported into the nearby reach segment and delivered within the river channel. All these components can help modeler to understand human activities that affect water quality.

(12) Your comment: P11423, L11: What is meant by "scenes"? Our respond: As suggested, "scenes" has been replaced by "images".

(13) Your comment: P11423, L12-14: This statements would need a reference. Our respond: As suggested, P11423, reference has been added and L12-14 has been revised as "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)." Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., and Lavabre, J.: Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models, J. Hydro., 250, 206-223, 2001. McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, J. Hydro., 400,

83-94, 2011.

(14) Your comment: P11423, L27-29: I think higher-resolution data should lead to better model performance; the question is whether this can be detected quantitatively! Our respond: In fact, the question about the quantitative impacts of data resolution on model predictions is also raised in this paper. To quantify the impacts quantitatively, the high-resolution GIS map is converted to coarser ones using the resample function of ArcMap. Then these resampled GIS data were used as model inputs, respectively, and their model performances are detected quantitatively based on the changes of goodness-of-fit indicators (Such as ENS). Moreover, we are trying to use more goodness-of-fit indicators, such as entropy, as quantitative approaches. Based on our results, higher-resolution data would not always lead to better model performance. One of the interesting results is that there exists a spatial resolution saturation level, beyond which further refinements to resolution do not improve model performance. In section 4.1, we has discussed about this issue.

(15) Your comment: P11424, L12-13: There is a sizable literature now on uncertainty in NPS modelling! Our respond: I agree that some works have been done on uncertainty in NPS modelling but I think most modelers might agree with the statement "there is relatively more uncertainty research about hydrological processes but less on NPS pollution". The problem is that this key statement is made with no attribution at all. To benefit our readers, we have added some key citations of hydrological (e.g., Beven, 2006; Balin et al., 2010) and NPS (e.g., Gassman et al., 2007; Wellen et al., 2015) literature to support our statement. Also, a pair of papers examining the effect of input uncertainty on NPS predictions have been cited in the introduction (Chaplot, 2005; Chaplot et al., 2005). Please find the attached manuscript.

(16) Your comment: The materials and methods fail to convey the details and the motivations for the uncertainty assumptions. Are these practically relevant? Our respond: As suggested, the details of the uncertainty assumptions have been added in the revised manuscript. In fact, per-input uncertainty assessment was done in our previous

papers (Shen et al., 2012a, 2013) and the database for planting, harvest, and tillage operations was based on our field investigations. Please find the attached manuscript.

(17) Your comment: What about issues of rainfall data quality (section 2.3.1) before interpolation is even considered (biases, gaps, etc)? Our respond: I agree the issues of rainfall data quality is very important. In fact, we investigated the effects of rainfall measurement errors in one previous study (Shen et al., 2015). A survey conducted by State Meteorological Administration (SMA) revealed that the error in rainfall measurements in China varied from 4.34% to 15.2%, with an averaged deviation equal to 6.52%. Based on our results, rainfall measurement error introduces considerable prediction uncertainty especially during high-flow periods, and this rainfall data quality issue can be reduced by a combination of dense rain gauge network, stochastic modeling and establishing a multi-event uncertainty analysis.

(18) Your comment: What about the issue of different maps and the issue of representativeness of spot land use information (section 2.2.3)? Our respond: More details about land use data have been provided in section 2.2.3. The following sentences have been added: "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 land use distribution characteristics during each period was analyzed according to use type of each land use 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."

(19) Your comment: Is the fertilizer data uncertainty distribution (section 2.3.4) applied to the spatial average or is it applied spatially differentiated? The former would be pointless in the light of the arguments put forward in this section. Is the normality

assumption realistic? Where does information on sigma come from? These choices will have influenced the results, i.e. that fertilizer data uncertainty was negligible here. It remains unclear how the per-input uncertainty assessment was done and whether the interaction of uncertainty sources was analysed at all (section 2.4). Our respond: Based on our limited local investigation, the initial annual applied urea and compound fertilizer was set as 450kg/ha and 300kg/ha for the potato-sweet potato rotation, while 150kg/ha and 225 kg/ha for the corn system, respectively. A survey conducted by local agricultural administration revealed the error or averaged deviation in the record fertilizer amount and this was used for the study area. Because there was not enough information available regarding the distribution of the fertilizer, normal distribution was used for fertilizer data in this study. In fact, we had compared those classic probability distribution functions (PDFs). Based on previous tests, without a realistic assessment of data distribution, the appropriate PDFs for each data may have little impact in the application of the hydrologic model. Therefore, we assumed that the fertilizer data were identically chosen from normal distribution spanning the feasible error range due to their simplicity.

(20) Your comment: P11425, L22: What is meant by "efficiently and effectively"? Our respond: I agree that these words are meaningless, so they have been removed from this paper.

(21) Your comment: P11426, L9: What is meant by "due to the lack of"? Our respond: In this study, 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; therefore, the use of this average information might result in fertilizer amount errors.

(22) Your comment: P11426, L16: I would say rainfall uncertainty stems predominantly (!) from representativeness issues. Our respond: I agree with the reviewer's idea so in this study, we focused on rainfall uncertainty comes from the lack of representative

rain gauges. In one previous study (Shen et al., 2012a), we has already focused on the impact of interpolation methods on the spatial rainfall heterogeneity but due to the availability of data, we did not focus on the representativeness of rainfall stations. In this study, rainfall data-induced prediction 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 (ENS) 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).

(23) Your comment: P11426, L20: Define E_NS. Our respond: As suggested, P11426, L20 has been revised as "the Nash–Sutcliffe efficiency coefficient (ENS) values".

(24) Your comment: P11426, L22: And then interpolated? Why was not the 'best' performing interpolation technique from Shen et al. (2012a) used? Our respond: In Shen et al. (2012a), the uncertainty introduced by spatial rainfall variability was determined using a number of different interpolation methods, such as the Centroid method; 2) the Thiessen Polygon method; 3) the Inverse Distance Weighted (IDW) method; 4) the Dis-Kriging method, and 5) the Co-Kriging method. Based on the results, it was also concluded that a global interpolation method such as IDW and Kriging would benefit both hydrological and NPS modeling in large watersheds. The reason why these techniques were not used was because these methods used the spatial correlation between each pair of stations to describe the variance over distance. Thus, minimum number of rainfall stations that are required by IDW and Kriging. However, most input datasets in this study, such as the single-station simulation, cannot meet the basic requirement of these methods. In fact, the expected rainfall spatial distributions were only generated by the centroid method was selected because it was the easiest to apply (Shen et al., 2012a). The centroid method, originally proposed by Wood et al (1990), has

already been incorporated into the SWAT model (Neitsch et al., 2002). The theory of this method is that the unknown points can be extracted by the values of nearby known points. In this study, the data from the rain gauge closest to the centroid of each sub-watershed was selected as the sole input for that particular sub-watershed. Then the areal rainfall was assumed to be homogeneous across sub-watersheds and inputted directly into SWAT utilizing the GIS interface.

(25) Your comment: P11428, L4: “Aggregate” and “average” should be swapped in this context. Our respond: As suggested, “aggregate” has been replaced by “average”.

(26) Your comment: P11429, L6: What is meant by “re-validate”? Our respond: Thank you for this remind because “re-validate” is wrong use. As suggested, P11429, L6 has been revised as “The ENS was used as the goodness-of-fit indicator to evaluate the model performance.” (27) Your comment: P11429, L12: That’s an unwarranted generalisation of a coarse grouping of different uncertainties. Our respond: I agree with this multi-input ensemble method might be a coarse grouping of different uncertainties. However, I think that the ensemble method can be used for a comprehensive evaluation of multi-input-induced model uncertainty. In this study, model structure was fixed and model uncertainty will stems predominantly from input errors. After the simulation, the ENS analysis was conducted with the previously model results, and this was done to further constrain the range of input dataset and to validate the new model taking conditional prediction uncertainty bounds into account. As suggested, we would try to find other methods to group different uncertainties.

(28) Your comment: P11429, L15-18: Reasoning unclear. Our respond: As suggested, P11429, L15-18 has been revised as: ”Based on the performance ratings by Moriasi et al. (2007), 0.5 was judged as a reasonable ENS value for TP simulation so a threshold of ENS$\geq$ 0.5 was defined to select acceptable SWAT runs (Liu and Gupta, 2007).” To provide a static state instead of subjective personal judgment, the performance ratings were adopted: very good (0.75-1), good (0.65-0.75), satisfactory (0.50-0.65), and unsatisfactory ($\leq$0.5).

(29) Your comment: P11429, L15: What’s meant by “interaction of input errors”? Our respond: Thank you for this remind. In fact, input errors cannot interact so we mean the interaction between input and model structure. As suggested, P11429, L15 has been revised as “In this study, model structure was fixed and model uncertainty will stems predominantly from input errors.”

(30) Your comment: P11429, L19: What’s meant by “grouped in all possible ways”? Our respond: In fact, we mean the stochastic combinations of various behavior input datasets. As suggested, P11429, L19 has been revised as “ In the next step, behavior input data (ENS$\geq$0.5), which refer to the phenomenon of equifinality and can be representative of a watershed system (ENS$\geq$0.5), were grouped to express the prediction uncertainty by using a multi-input ensemble method.”

(31) Your comment: P11430, L1, P11431, L8: What’s meant by “multi-input ensemble method”? Our respond: In this study, the ensemble of input-induced outputs was used to express the prediction uncertainty. Firstly, a threshold of ENS$\geq$ 0.5 was defined to select acceptable SWAT runs (Liu and Gupta, 2007). Then the input-induced model uncertainty was generated via sampling from the output distributions that are generated from these effective input datasets.

(32) Your comment: P11430, L6: Table 3 says 0.66 and 0.89. Our respond: Thank you for the wonderful job. As suggested, P11430, L6 has been revised as:” As shown in Table 3, for the flow simulation, the ENS were 0.66 and 0.89 in the calibration and validation periods, respectively.”

(33) Your comment: P11430, L8: Table 3 says 0.13 instead of 0.46. Our respond: In fact, 0.46 is right figure. Table 3 has been revised as: Table 3 The values of ENS and R2 of the SWAT model during the calibration and validation period Variable Indicator Calibration Validation Flow ENS 0.66 0.89 R2 0.79 0.95 Sediment ENS 0.73 0.67 R2 0.83 0.83 TP ENS 0.75 0.46 R2 0.86 0.79

(34) Your comment: P11430, L17: Which single rain gauge simulations? Unclear what

exactly was done here. Our respond: As suggested, P11430, L17 has been revised as "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."

(35) Your comment: P11430, L25: This plateau would be interesting to show. Our respond: Yes, I am also pleased to find a plateau exits when the number of rainfall station increases. This demonstrates that the information content in rainfall spatial variation is reached after a relatively small number of key gauges are used as model input (Seibert and Beven, 2009). It is encouraging that a small number of gauges distributed more optimally and perform well for logistical reasons (Bárdossy and Das, 2008; McMillan et al., 2011). In reality, there might not be many dense rain gauge networks similar to those used for this study; therefore, the fact that spatial rainfall variation is a function of key gauges rather than all gauges would indicate a wider range of applicability. For this study area (2,421 km2), the optimal number of gauges were identified as 6 beyond which further improvements to the SWAT predictions would not be found.

(36) Your comment: The results don't clarify what the benchmark is against which uncertainty scenarios are being compared. Our respond: In fact, the choice of benchmark depends on the targeted variable. When multiple-gauges are used, single-gauge simulation is used as benchmark, while when single-gauge stations are simulated, these rainfall stations were ranked from high to low based on the ENS values and the results of the highest-ENS simulation are used as the benchmark. To benefit our reader, we added the following sentence "These rainfall stations were ranked based on the ENS values, and combinations of m rain gauges (m ranged from 2 to 12) were used as SWAT inputs."

(37) Your comment: P11431, L19-21: Argument not convincing. Our respond: As suggested, P11431, L19-21 has been revised as "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)."

(38) Your comment: P11432, L6-10: Unclear. Our respond: As suggested, P11432, L6-10 has been revised as "It should be noted that comparison using un-recalibrated models is useful to evaluate the differences in model predictions because calibration masks the differences that may occur as a result of the input data sets. In addition, the un-recalibrated model results can show how good each dataset predicts stream flow before calibration, which would indicate the effort required for calibration when using each data set."

(39) Your comment: P11432, L8: What is meant by "due to the availability of validation data"? Our respond: As suggested, "due to the availability of validation data" has been removed from P11432, L8. In fact, "the availability of validation data" means "the availability of measured data".

(40) Your comment: P11433, L13: What's meant by "certainty-appropriate"? Our respond: As suggested, P11433, L13 has been revised as "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."

(41) Your comment: P11434, L7-11: Argument not convincing. Our respond: As suggested, the following sentences have added "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)."

(42) Your comment: P11435, L15: Is it realistic to suggest this uncertainty reduction? Our respond: In fact, this study focused on the error-transmission from input data to model predictions. Based on results, we can have a better understand of the impacts of input datasets on the functionality of watershed models. The study indicated that rainfall input resulted in the highest uncertainty, followed by DEM, land use maps, and

fertilizer amount. Therefore, measures should be taken first to reduce this source of uncertainty by adding rain gauges, modifying the selection mechanism of rain gauge in SWAT, and using appropriate interpolation techniques. As suggested, P11435, L15 has been revised as “Therefore, input data uncertainty is critical in NPS modeling and efforts should be made to clarify this type of uncertainty.”

(43) Your comment: P11435, L17: “As a complex function of” was never demonstrated in the paper. Our respond: As suggested, P11435, L17 has been revised as “Second, as illustrated in Fig. 3, the input data-induced uncertainty varies considerably temporally and spatially due to the varying climate, underlying topography, land use, soil type, and management (Shen and Zhao, 2010; Chen et al., 2012).”

(44) Your comment: P11435, L18-20: But how to define it a priori? Our respond: To define the MOS a priori, more site-specific studies should be conducted, especially in different hydrologic and environmental conditions. We are struggling on this issues in hopes that generalities about the relative appropriateness of site-specific MOS. As suggested, P11435, L18-20 has been revised as” In this sense, a site-specific MOS, which might be more robust to any particular sequence of input errors than current steady MOS, should be defined as a priori.”

Thank you very much for your wonderful job. Hope that our responses are satisfactory, and look forward to hearing from you.

Zhenyao Shen Professor School of Environment, Beijing Normal University Beijing, China, 100875 Tel/Fax: +86-10-5880 0398 E-mail: zyshen@bnu.edu.cn

References Chen L, Shen ZY*, Yang XH, Liao Q, Shaw L. Yu. An Interval-Deviation Approach for Hydrology/Water Quality Model Evaluation. Journal of Hydrology. 509C: 207-214, 2014. Harmel, R.D., Cooper, R.J., Slade, R.M., Haney, R.L., Arnold, J.G., 2006. cumulative uncertainty in measured streamflow and water quality data for small watersheds. T. ASABE 49, 689-701. Harmel, R.D., King, K.W., 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds.

T. ASABE 48, 1713-1721. Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Srinivasan, R., Williams, J.R., 2002. Soil and Water Assessment Tool User’s Manual:Version 2000. GSWRL Report 02-02, BRC Report 02-06, Publ.TexasWater Resources Institute, TR-192, College Station,Texas, USA. Shen ZY*, Chen L, Liao Q, Liu RM, Huang Q. Effect of rainfall measurement errors on nonpoint source pollution model uncertainty: a Monte-Carlo simulation of SWAT model output. Journal of Environmental Informatics.2015,16(1):14-26. Woods, R, Sivapalan, M., 1999. A synthesis of space-time variability in storm response: rainfall, runoff generation, and routing. Water Resour. Res., 35, 2469-2485.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/12/C6449/2016/hessd-12-C6449-2016-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 11421, 2015.