**Response to Reviewer #2:**

This paper developed a new algorithm called BEAR for accurate quantification of input errors in water quality modeling. The precondition of the BEAR algorithm is that the input uncertainty should be dominant and that the prior information of the input error model can be estimated. Results of both synthetic data and observed data indicated the efficiency of the algorithm. Overall, the paper is well rewritten and the topic is suitable for the journal. However, the following issues should be further explained and clarified before its submission:

We thank the reviewer for the overall positive assessment of the manuscript and helpful comments, which have helped to improve our study. We have responded to each point in turn in the following sections. The comments from the reviewer are provided in blue text and our responses are organized point-by-point in black text. The manuscript text after changes is shown in "*black italics*" and the equation and section number are shown in yellow highlight.

It should be noted that the method name will change from the "Bayesian error analysis with reshuffling" into "Bayesian error analysis with reordering". This is based on suggestions by one of the reviewers, as the word "shuffling" implies randomness in the reordering, while the reordering in our method is determined by the model residual error. The term "reordering" better reflects the deterministic nature of error quantified via this new method. Besides, the abbreviations of methods (T, D, R) will be changed to the full names (Traditional, IBUNE, BEAR).

1) There have been many studies focusing on the uncertainty of input data errors for hydrologic modelling, and many methods including Bayesian algorithm can be used for handling the issue. However, the gap between previous studies and this study was not explained clearly in the Introduction. The motivation of this study should be clearly clarified.

Thanks for your suggestion. The research gap and motivation will be modified in the Introduction as follows:

"*Input uncertainty can lead to bias in parameter estimation in water quality modeling (Chaudhary and Hantush, 2017, Kleidorfer et al., 2009, Willems, 2008). Improved model calibration requires isolating the input uncertainty from the total uncertainty. However, the precise quantification of time-varying input errors is still challenging when other types of uncertainties are propagated* *through to the model results. In hydrological modeling, several approaches have been developed to characterize time-varying input errors, and these may hold promise for application in WQMs. The Bayesian total error analysis (BATEA) method provides a framework that has been widely used (Kavetski et al., 2006). Time-varying input errors are defined as multipliers on the input time series and inferred along with the model parameters in the Bayesian calibration scheme. It leads to a high-dimensionality formulation, which cannot be avoided (Renard et al., 2009) and restricts application to cases where event-based multipliers (the same multiplier applied to one storm event) need to be used. In the Integrated Bayesian Uncertainty Estimator (IBUNE) (Ajami et al., 2007) approach, multipliers are not jointly inferred with the model parameters, but sampled from the assumed distribution and then filtered by the constraints of simulation fitting. This approach reduces the dimensionality significantly and can be applied in the assumption of the data-based multiplier (one multiplier for one input data) (Ajami et al., 2007). However, this approach is less effective because the probability of co-occurrence of all optimal error values is very low, resulting in an underestimation of the multiplier variance and misidentification of the uncertainty sources (Renard et al., 2009). From the above, a new strategy should be developed to avoid high dimensional computation and meanwhile ensure the accuracy of error identification.*"

2) More detailed steps about how to use the BEAR algorithm should be explained. Besides, the advantages of the BEAR algorithm compared with conventional methods should be more clearly clarified for making clear understanding from readers.

Thanks for your suggestion. The detailed steps of the BEAR method will be added in Appendix A (see the following Appendix A), and an illustration example will be moved from the methodology part to Appendix A to make the explanation more clear. In addition, the comparison with conventional methods will be clarified as follows:

"*The application of the BATEA framework is limited by high dimension computation (Renard et al., 2009). In quantifying the data-varying errors (rather than the event-varying errors in the study of BATEA (Kavetski et al., 2006)), the computational dimension is easily excessive and the BATEA probably becomes impractical (Haario et al., 2005). Therefore, the BATEA method is not considered in the comparison. In this study, three methods are compared to evaluate the ability of the BEAR method in quantifying input errors. The first one is the "Traditional" method, regarding* *the observed input as error-free without identifying input errors (i.e. Eq. (2)), while the other two methods employ a latent variable to counteract the impact of input error and build the modified input (i.e. Eq.(4)). One of them is the "IBUNE" method, where potential input errors are randomly sampled from the assumed error distribution and filtered by the minimization of the objective function (Ajami et al., 2007). Although the comprehensive IBUNE framework additionally deals with the model structural uncertainty via the Bayesian Model Averaging (BMA) method, this study only compares the capacity of its input error identification part. The last one is the "BEAR" method developed in this study. This new method adds a reordering process into the "IBUNE" method to improve the accuracy of input error quantification.*"

3) Actually, the availability of prior knowledge of the input data error is important for modelling, but is also a difficult issue. It may be not enough only mentioning this issue in Conclusion. At least more discussions and the potential solutions should be provided.

The reviewer raised an important point. The discussion about this will be added in Section 4.2:

"*The availability of prior information of the input error relies on the studies about benchmarking the observational errors of the water quality data and hydrologic data. When the prior information is not available, the selection of the proper input error model is important. Comparing the error parameter estimations in Figure 3, the $\mu$ and $\sigma$ estimations are less biased from the reference values in add-inferred scenario than in mul-inferred scenario. It illustrates that the compensating effect between the input error and parameter error is weaker in the additive form of the input error. However, this is probably related to the specific model structure, as exponent b in BwMod has a stronger interaction with the multiplied errors than the additive errors. Thus, more comprehensive comparisons should be taken to explore the capacity of different input error models in different model applications.*"

4) The quality of some Figures in the manuscript should be improved to make all information clear.

Thanks for pointing this out. We will improve the quality of all the figures, including improving the resolutions and modifying the colors or placing of legends.

Appendix A: The illustration of the BEAR method

Table A 1 An example illustrating the BEAR method

| row | time step $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1st iteration (random sample) | | | | | | | | | | | | | | | | | | | | |
| 1 | sampled input error | 0.07 | -0.12 | 0.07 | 0.16 | 0.05 | .0.07 | 0.07 | -0.03 | 0.03 | -0.08 | 0.09 | -0.11 | -0.11 | -0.08 | -0.29 | 0.14 | 0.03 | -0.08 | 0.14 | -0.17 |
| 2 | input error rank $k$ | 13 | 3 | 14 | 20 | 12 | 17 | 15 | 9 | 10 | 7 | 16 | 4 | 5 | 6 | 1 | 19 | 11 | 8 | 18 | 2 |
| 3 | residual error $\varepsilon$ | -0.29 | 0.49 | -0.58 | -0.98 | -0.78 | 0.29 | -0.66 | 0.59 | -1.31 | -0.31 | -0.87 | 0.76 | 0.46 | 0.54 | 0.25 | -0.80 | -0.07 | 0.56 | -0.23 | 0.40 |
| | MSE | 0.40 | | | | | | | | | | | | | | | | | | | |
| | 2nd iteration (random sample) | | | | | | | | | | | | | | | | | | | | |
| 4 | sampled input error | -0.01 | -0.02 | 0.03 | 0.03 | -0.09 | 0.00 | -0.02 | 0.06 | 0.11 | 0.11 | -0.09 | 0.01 | -0.12 | -0.11 | 0.00 | 0.15 | -0.08 | 0.04 | -0.02 | 0.11 |
| 5 | input error rank $k$ | 9 | 6 | 14 | 13 | 3 | 10 | 8 | 16 | 17 | 18 | 4 | 12 | 1 | 2 | 11 | 20 | 5 | 15 | 7 | 19 |
| 6 | residual error $\varepsilon$ | -0.13 | 0.23 | -0.43 | -0.41 | -0.21 | 0.70 | -0.23 | 0.09 | -1.88 | -1.52 | 0.20 | 0.17 | 0.53 | 0.60 | -0.43 | -0.72 | 0.36 | 0.12 | 0.47 | -0.82 |
| | MSE | 0.47 | | | | | | | | | | | | | | | | | | | |
| | 3rd iteration (updating the error rank via the secant method) | | | | | | | | | | | | | | | | | | | | |
| 7 | calculated pre-rank $K$ | 5.8 | 8.7 | 14.0 | 8.0 | -0.3 | 22.0 | 4.3 | 17.3 | -6.1 | 4.2 | 6.2 | 14.3 | 31.3 | 42.0 | 4.7 | 29.0 | 10.0 | 16.9 | 14.4 | 7.6 |
| 8 | ranked rank $k$ | 6 | 10 | 12 | 9 | 2 | 17 | 4 | 16 | 1 | 3 | 7 | 14 | 19 | 20 | 5 | 18 | 11 | 15 | 13 | 8 |
| | 3rd iteration (reordering errors according to the updated error ranks) | | | | | | | | | | | | | | | | | | | | |
| 9 | reordered input error | -0.02 | 0.00 | 0.01 | -0.01 | -0.11 | 0.11 | -0.09 | 0.06 | -0.12 | -0.09 | -0.02 | 0.03 | 0.11 | 0.15 | -0.08 | 0.11 | 0.00 | 0.04 | 0.03 | -0.02 |
| 10 | residual error $\varepsilon$ | -0.23 | 0.20 | -0.34 | -0.24 | -0.12 | 0.19 | 0.14 | 0.08 | -0.40 | -0.31 | -0.22 | 0.03 | -0.17 | 0.26 | -0.09 | -0.55 | 0.11 | 0.14 | 0.27 | -0.23 |
| 11 | MSE | 0.06 | | | | | | | | | | | | | | | | | | | |

The implementation of the BEAR method contains two main parts: sampling the errors from an assumed error distribution and reordering them with the inferred ranks via the secant method. An example is illustrated in **Error! Reference source not found.** and the explanation about the specific steps is presented in the following contents.

(1) In the 1st iteration ($q$=1), the errors are randomly sampled from the assumed error distribution (row 1), and then they are sorted to get their ranks (row 2). This error series is employed to modify the input data, which corresponds to a new model simulation and model residual (row 3).

(2) Repeat the step (1) in the 2$^{nd}$ iteration ($q$=2) as two sets of samples are prerequisites for the updating via the secant method. The results are shown in row 4, 5 and 6. **Error! Reference source not found.** demonstrates that the ranges of the error distribution are the same between the true input errors (black line) and the sampled errors (blue and green lines) as they come from the same error distribution under the condition that prior knowledge of the input error distribution is correct. However, the value at each time step is not close.

(3) At the 1$^{st}$ time step ($i$=1) in the 3$^{rd}$ iteration ($q$=3), the pre-rank $K_{1,3}$ is calculated via the secant method (illustrated as the following equation). The details are demonstrated in red boxes.

$$K_{1,3}=k_{1,2}-\varepsilon^{p}_{1,2}\frac{k_{1,2}-k_{1,1}}{\varepsilon^{p}_{1,2}-\varepsilon^{p}_{1,1}}=9\text{-}(\text{-}0.13)\frac{9-13}{-0.13-(-0.29)}=5.8$$

(4) Repeat the step (3) for all the time steps. The calculated pre-ranks are shown in row 7.

(5) Sort all the pre-ranks to get the integrity error rank (row 8).

(6) According to the updated error ranks (row 8), the sampled errors in the 2nd iteration (row 4) are reordered. The example for the 1st time step is demonstrated in black boxes. The error rank at 1st time step is updated as 6, and the rank 6 corresponds to the error value -0.02 in 2nd iteration. Therefore, -0.02 is the input error at the 1st time step in the 3rd iteration. Following this example, the sampled errors at all the time steps are reordered. The results are shown in row 9. **Error! Reference source not found.** demonstrates that after reordering the errors with the inferred ranks, the estimated errors are much close to the true input error.

(7) The reordered input error will lead to a new input data, a new model simulation and a new model residual. The residual error is shown in row 10.

(8) If a defined target about the residual error is achieved, the input error estimation is accepted; Otherwise, $q$=$q$+1, repeat step (3)~(7) until $q$ is larger than the maximum numbers of iteration $Q$.

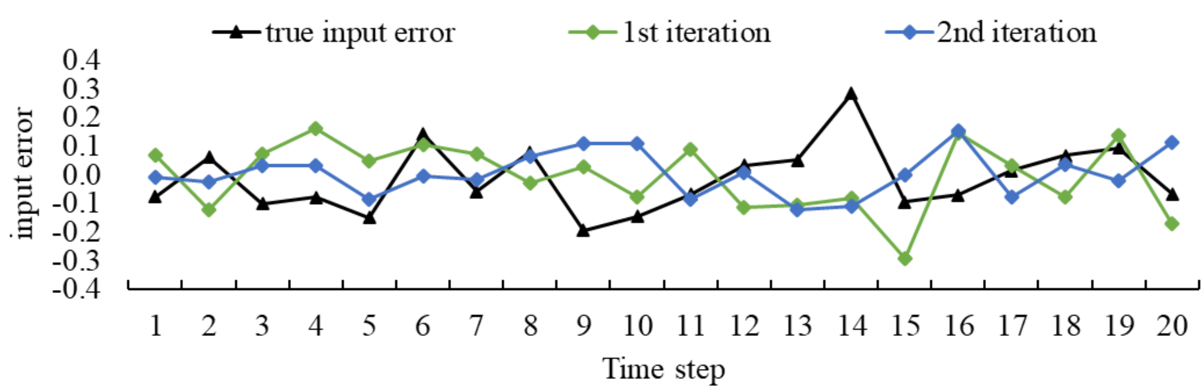


Figure A 1 Demonstration of the input error estimation in **Error! Reference source not found.** at the 1st and 2nd iteration where the input errors are randomly sampled

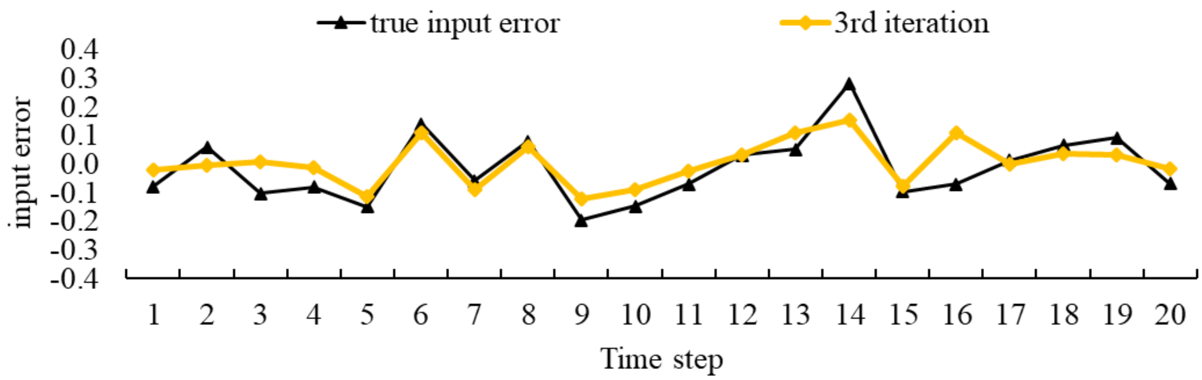


Figure A 2 Demonstration of the input error estimation in **Error! Reference source not found.** at the 3rd iteration where the input errors are reordered according to the updated error ranks