Response to Reviewer #1:

The study proposes and demonstrates an algorithm for quantifying input uncertainty called BEAR (Bayesian error analysis with reshuffling). It is claimed that the method is suitable to overcome restrictions of current state-of-the-art approaches like high dimensional computational problems or underestimation and misidentification of error sources. For this purpose, the algorithm employs the secant method to estimate a certain rank of error associated to input data from an underlying rank distribution of errors. After introducing the method, it is demonstrated on the task of total suspended solids modelling in, first, a synthetic case study and, second, a real test case. Thereby, both, the effectiveness and the limitations are shown and discussed. Finally, transferability of the method within the field of water quality modelling and potential routes of improvement are presented.

General comments:

The issue of uncertainty quantification in modelling is for sure one of high importance. By focusing on input uncertainty this study addresses a branch that is particularly challenging in this field. Contributions in this direction deserve attention and the topic of this manuscript is suitable for the journal. However, certain issues regarding content and presentation of the material require to be addressed:

We thank the reviewer for the overall positive assessment of the manuscript and helpful comments. 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 the proposed changes is shown in "*black italics"* and the equation and section number are shown in yellow highlight.

It should be noted that we are proposing 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.

1) Maybe it is just the presentation, but it was not straightforward to see how the method exactly works. Aside of a more detailed explanation, providing more illustrations to support explanations about how the method exactly works might help, e.g. displaying the secant method itself, error distribution in rank space, etc.

Thanks for your suggestion. We propose to address this by modifying the methodology in the following points to make the algorithm clearer:

(1) Summarize the main steps in the BEAR method upfront:

"*The BEAR method consists of the following key steps: 1) Sample the errors from the assumed error distribution to maintain the overall statistical characteristics of input errors; 2) Update the input error ranks via the secant method; 3) Reorder these sampled errors according to the updated error ranks, leaving the error magnitudes unchanged; 4) Repeat (2) and (3) for a few iterations until a defined target is achieved*."

(2) Integrate an example and its illustration in Appendix to explain the specific steps involved. More explanation for the rank estimation via the secant method and the reordering steps will be added (see an illustration of this in the following Appendix A).

(3) Separate the description of the BEAR method from the ABC-SMC calibration scheme. following suggestions from other reviewers, we propose this because the ABC-SMC calibration algorithm is not necessary in the BEAR method, and the core idea of the BEAR method (the reordering strategy in the rank estimation) can be easily applied in any other calibration algorithm, for example, MCMC and SMC algorithms.

2) By design, the BEAR method seems to shuffle and pick errors (by their ranks) such that maximum fit to the data is achieved. Is this a proper addressment of the input errors in terms of quantification of input uncertainty? For instance, in L.232 it is discussed that "method R always has much higher correlations with the true error series" and in L.243 it outperforms the other methods with highest NSE values. Both seem to be effects from the BEAR method searching for optimally fitting errors until exactly the error is found that minimizes the gap between model predictions and observations.

The reviewer raised an important question. The BEAR method works well under the circumstance where the input error is dominant in the total uncertainty, where minimizing the residual error has a similar effect as minimizing the input error. From this point of view, the sampling and reordering strategy in the BEAR method provides an effective way to identify the input error according to the residual error. This is what the Reviewer refers to as "searching for optimally fitting errors until exactly the error is found that minimizes the gap between model predictions and observations". Like current methods that BEAR seeks to demonstrate an improvement over, the input error compensates for other errors, a step that is constrained by accurate prior information of the input error distribution being available. However, the compensating effect in the BEAR method is more apparent because it is much more effective than other current methods in minimizing the gap. Thus, the accuracy of the input error model is particularly important in the BEAR method. The analysis and discussion in Section 4.2 will be modified to convey this as follows:

"The IBUNE method takes advantage of stochastic error samples to modify the input observations (Ajami et al., 2007). In Fig. 4 and Fig. 6, the uncertainty bands of modified inputs (blue parts) encompass the original input data, illustrating that the intrinsic quality of the input data plays an important role in the algorithm performance. Fig. 6 demonstrates that if the input error is insignificant in the residual, like in the O-fixed and O-inferred scenarios for the real case, the resultant simulations will fit the observed output (green line) well. Otherwise, the simulations are far away from the observed outputs (black line) due to inaccurate input observations (in the S-fixed and S-inferred scenarios in the real case). As per the finding in the previous study of Renard et al. (2010), if the σ of input errors is inferred with the model parameters, the IBUNE method will underestimate σ (in Fig. 3(1) and Fig. 5(a2)). If σ is fixed as per the prior information, the input modification and model simulation cannot be improved in the scenarios with large intrinsic σ of the input errors, demonstrated by a wider band in Fig. 6(b3) than in Fig. 6(b4). From the above, the data quality is more important than the availability of prior information for the IBUNE method, especially when the intrinsic σ of the input error is large.

However, the findings in the BEAR method are quite different. Although the BEAR method infers the input error also by minimizing the model residual error, it is much more effective than the IBUNE method. For the synthetic case (Fig. 3(c)) and real case (Fig. 5(c)), the model simulations via the BEAR method (red parts) are very close to the output observations (green line). In other

words, the estimated input error mainly depends on the output observation. Therefore, in the real case with the same output observation (Fig. 6(c)), the modified inputs are consistent among the different scenarios. If the input uncertainty is dominant over the output observational uncertainty, the BEAR method effectively employs more accurate information (output observations) to modify the less precise information (input observations).

To constrain the impacts of the other sources of error, accurate prior information about the input error model is important in the BEAR method. The fixed scenarios are assumed to have more accurate prior information than inferred scenarios. In the synthetic case, fixed scenarios always produce a higher NSE of the modified input (Fig. 3(5)) and a larger correlation in the estimation error (Fig. 3(3)) than inferred scenarios. In the real case in Fig. 6, the modified inputs in fixed scenarios are closer to the streamflow observation from the rating curve than the modified inputs in inferred scenarios.

To sum up, the role of prior information regarding the input error model is more important in the BEAR method than in the IBUNE method. A more accurate input error model can bring a more precise estimation of input errors by constraining the adverse impacts that other sources of errors may have."

3) Expectations are raised that the method overcomes issues of state-of-the-art frameworks like BATEA and IBUNE. Yet, no direct comparison is shown which makes it hard to see the benefit of the method. Both these methods are frequently mentioned and a comparison is claimed. So far, there is a comparison of cases abbreviated by "T" (traditional), "D" (distribution) and "R" (BEAR method itself). "D" is referred to be "similar to the basic framework of the IBUNE method". However, this does not provide an actual comparison.

The reviewer is correct that we propose denoting the BATEA and IBUNE methods instead of explicitly naming them in our comparison. We will change the abbreviation to the full name of the methods as per the reviewer's suggestion, and add more explanations about this comparison, 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."

4) The method is supposed to reduce "the potential search space for input errors" (L.360). I wonder whether this is the objective quantification of input uncertainty? Isn't it rather a comprehensive assessment of the errors and noise associated to input error and not searching in a sub-space of already collected errors and then selecting the one that fits best during predictions?

We apologize for the lack of clarification. We will add more explanations in the revision. Here "reduce the potential search space for input errors" (L.360) is because "*In a continuous sequence of data, the potential error values have an infinite number of combinations, while the error rank has limited combinations, dependent on the data length. For example, in Table A.1, the estimated error at the 1st time step could be any value. Even under the constraint of an input error ranging from the minimized to the maximized sampled errors (i.e. [-0.29,0.16] in the 1st iteration), error magnitude estimation still has infinite possibilities due to the continuous probability distribution the error represents. In contrast, the rank is discrete, having only 20 possibilities (i.e. an integer from [1,20]). From this point of view, it is far more efficient to estimate the error rank than estimate the error value*."

To avoid the misunderstanding the current manuscript created, we will delete "reduce the potential search space for input errors" (L.360), and change this sentence as follows:

"*The estimation focuses on the error rank rather than the error magnitude, which significantly improves the effectiveness of input error quantification.*"

5) Generally, a thorough discussion on the used error distributions is missing, e.g. why is a bias of 0.2 in the error function assigned without further discussion (l.211)

Thanks for your comments. We will add a clarification as follows:

"*If the input errors are estimated based on a rating curve, like the procedure in the following real case, the error distribution should be assumed as a Gaussian distribution and the mean should be 0. However, in order to test the ability of the BEAR method in wider applications, the systematic error bias equal to 0.2 has been considered in the synthetic case. An additive formulation (denoted as 'add' in Table 3) is adopted to illustrate the error generation in measurements, while the multiplicative formulation (denoted as 'mul' in Table 3) is specifically applied for errors induced from a log-log regression procedure, which is common in the water quality proxy processes (Rode and Suhr, 2007).*"

6) There is at least one article cited in the manuscript, that does not appear in the list of references (please see specific comments, l.190). Please assure correct referencing.

Thanks for your comments. We will correct all the missing references.

Specific comments

1) L. 37-38: "…estimate the residuals between the measurements and proxy values…" -> yet, measurement error is not addressed

Thanks for your comments. we will clarify this as follows:

"*In this process, the measurement errors can be ignored given the errors introduced from the surrogate process are commonly much greater than the measurement errors (McMillan et al., 2012).*"

2) L. 68: “variable” -> “scalar“– both, vectors and scalars represent variables

“variable” will be changed into “scalar”.

3) Eq. 3: unnecessary, since given by equation (1)

Equation (3) will be deleted.

4) L. 84-91: repetitive, add details to the corresponding paragraph in the introduction

Thanks for your suggestion. We will move these details into the introduction, as follows:

“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, results 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.”

5) L. 92: “innovation” -> rather “introduction” or simple “The secant method” as chapter header – the innovation was made before

We agree with the suggestion. The section titled *“innovation”* will be changed to *“introduction”*.

6) L. 98-99: Rank definition and concept -> requires further explanation

We will add further explanations, as follows:

"Here, the rank is defined as the order of any individual value relative to the other sampled values and determines the relative magnitude of each error in all data errors. For example, in the 1st iteration in Figure A 1, the error at 15th time step, -0.29, is the smallest value among all the sampled errors, therefore, its rank is 1."

7) L. 128ff: it sound like in ABC the requirements on the likelihood function are looser and therefore the method is easier to apply. However, requirements are also strict but ABC allows for Bayesian inference if the likelihood function is intractable. -> Please reformulate and clarify.

Thanks for your comments. Given the BEAR algorithm could be implemented via SMC, GLUE or SCE-UA, or any common model calibration approach, and this description about ABC confuses the main contribution of the paper (i.e. the core idea of BEAR method, which is to optimize input error ranks rather than input error magnitudes) we propose recasting the implementation of our optimization algorithm via SMC.

8) L. 132ff: Notation "OF" not explained. Overall, the introduction of ABC and SMC is not clear. Further, the motivation why SMC is used here is not given.

Thanks for your comments. "OF" here means "objective function". But according to the above reply, we propose recasting the implementation via SMC, which target is the likelihood function rather than the objective function. "*The SMC sampler is more computationally efficient than previous algorithms that have applied rejection sampling and MCMC samplers (Sisson et al., 2007, Jeremiah et al., 2011).*"

9) L. 146: "…when 1000 proposed parameter sets…" -> is this suggested as a general approach or an arbitrary choice for this study. Please explain.

According to the reply for 7), we propose recasting the implementation via SMC. If the BEAR method is implemented using a likelihood-based calibration procedure, the proposed parameter is compared to the previous entry in the chain and there is no need to set this stop criterion, just follow traditional convergence rules.

10) L.171ff: Please replace abbreviations T, D and R by their names. With all abbreviations that follow it is hard to keep track.

Thanks for your suggestion. We will change the abbreviations into the full names in the below descriptions and related figures.

"*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 to the "IBUNE" method to improve the accuracy of input error quantification.."*

11) LL. 190+196: "Sikorska et al, 2015" ! missing in references

We will add all the missing references.

12) Eq. 9: define parameter "b"

The definition is shown in Table 2. A clarification will be added:

"*where the descriptions of a and b are shown in Table 2*"

13) L. 215ff: incomplete sentence

We will complete this as follows:

"*The true output* $\boldsymbol{Y}^*$ *is the simulated TSS concentration via BwMod corresponding to the true input* $\boldsymbol{X}^*$ *and model parameters set as the reference values in Table 2.*"

14) L. 229ff: "calibrated via method T,…" -> misleading explanation. Please provide a more specific explanation of the calibration process under error scenarios T, D and R

Please see the reply for 10)L. 171ff

15) L. 257: ":…the impacts of model structural error and output data error cannot be ignored." vs. L.264: ":…other sources of uncertainty can be ignored" -> sound like a contradiction, please elaborate

Thanks for your comments. We will modify the description as follows:

"*In real-life applications, the impacts of model structural error and output data error exist and may impair the implementation of the BEAR method.*"

"*As the BEAR method works well under the assumption that input uncertainty is significant, other sources of uncertainties can be ignored in comparison,"*

16) L. 282-283: "This illustrates that the impacts of other…" -> unclear phrase, please clarify and re-formulate

We will add the clarifications as follows:

"Compared with sound estimations in the synthetic case where the modeling only suffers from the input error and parameter error, this undesirable result illustrates that the impacts of other sources of errors impair the error quantification when the prior information of input error is not accurate, regardless of the methods."

17) L. 291: "…could be regarded as the reference value." -> Why? Please explain.

Thanks for pointing this out. The observed streamflow from the rating curve cannot be considered as the reference value, it is just closer to the reference value than the simulated streamflow via GR4J. The explanations will be corrected as follow:

"*According to the results of the traditional method in Fig. 6(a), the outputs in the "O" scenarios (in (a1) and (a2)) capture the dynamics of observed TSS concentration better than the outputs in the "S" scenarios (in (a3) and (a4)). Thus, compared with the simulated streamflow via GR4J ("S" streamflow), the observed streamflow from the rating curve ("O" streamflow) should be closer to the true input data*."

18) L. 295-296: "…have an infinite number of combinations, while the error rank has limited combinations, dependent on data length." -> What is exactly meant here?

We will add the explanation as follows:

"In a continuous sequence of data, the potential error values have an infinite number of combinations, while the error rank has limited combinations, dependent on the data length. For example, in Table A.1, the estimated error at the 1st time step could be any value. Even under the constraint of input error ranging from the minimized to the maximized sampled errors (i.e. [-0.29,0.16] in the 1st iteration), error magnitude estimation still has infinite possibilities due to the continuous probability distribution the error represents. In contrast, the rank is discrete, having only 20 possibilities (i.e. an integer from [1,20]). From this point of view, it is far more efficient to estimate the error rank than estimate the error value."

19) L. 297ff. "Compared with the IBUNE framework…" -> there is no real comparison made, please see major comments

Please see the reply to the major comment 3).

20) L. 340: "for method R, an accurate input error model can constrain the adverse impacts…" -> wasn't this the problem to begin with? Please clarify this sentence.

This point will be clarified as follows:

"To constrain the impacts of the other sources of error, accurate prior information about the input error model is important in the BEAR method. The fixed scenarios are assumed to have more accurate prior information than inferred scenarios. In the synthetic case, fixed scenarios always produce a higher NSE of the modified input (Fig. 3(5)) and a larger correlation in the estimation error (Fig. 3(3)) than inferred scenarios. In the real case in Fig. 6, the modified inputs in fixed scenarios are closer to the streamflow observation from the rating curve than the modified inputs in inferred scenarios."

Please also see the reply to the major comment 2).

21) L 354-355: "However, the ability of these approaches needs further discussion in systems with correlated responses." -> Please clarify – what is the exact problem and why do ARMA models fit here?

Thanks for your comments. We will add clarifications as follows:

"*The part of each residual error correlated with the previous residual errors can be represented by an autoregressive moving average (ARMA) model (Kuczera, 1983) or autoregressive (AR) model (Schaefli et al., 2007, Bates and Campbell, 2001). This correlated part is removed from the residual error and the remaining part is considered to be impacted by the input error only. Thus, the correspondence between the input error rank and the residual error part is ensured and the latter process will be the same as the application of the BEAR method in BwMod. However, the specific settings of such an approach need further discussion in systems with correlated responses, for example, in the calculation of coefficients of the ARMA or AR model since the residual error changes in each iteration of calibration*."

22) L. 358: "developed" -> "proposed" – the methods are already known but used in a way to address input error here.

We will change *"developed"* to *"proposed"*.

23) L 362: "… addresses the high dimensionality problem…" -> not shown

"Address" will be changed to "avoid" and more clarification will be added as follows:

"*The introduction of the secant method links the error rank for each input data to its corresponding residual, which avoids the high dimensionality problem resulting from calibrating all the errors as a whole*."

Figures

1) General: Legends in figures should be improved, e.g. in terms of colors or placing
2) General: Provide higher resolution and unify the legend (see especially Fig. 4 and 6)
3) Figure 3: please use colors that are better distinguishable (see cases "T" and "R")

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

4) Figure 3(4): NSE = 1 is unrealistic. Please see major comments.

Thanks for your pointing it out. NSE is close to 1, not equal to 1. We will modify the demonstration to avoid this misreading. This occurs as Figure 3 shows the results of the synthetic case where the modeling only suffers from the input error and parameter error. The BEAR method is effective in isolating the input error and parameter error, which has been proved by the fact that NSE is much closer to 1. However, when the BEAR method is applied in real applications where other sources of errors will interfere, as Figure 6 shows, the fit to the output TSS observations reduces.

5) Figure 4 (c3,c4): model predictions are clearly shifted. Please elaborate on this offset.

The model applied is BwMod. When the input (streamflow) is large, the output (TSS concentration) will be reduced due to the wash-off effect. It is opposite to the hydrological model, where the large input (precipitation) will lead to a large output (discharge).

6) Figures 4 and 6: Maybe it is better to show these figures in the appendix and only present the most important subfigures in the main text

OK, we will move these two figures into Appendix.

Tables

1) Tables 1-1 and 1-2: The tables could be presented as additional files but are not helpful in the main article

OK, we will move these two figures into Appendix, integrating the descriptions and other figures in Appendix A to provide a more clear explanation about the BEAR method.

2) Table 3: the "fixed" scenarios in the real test case are not fixed but provide small hyperparameter ranges

Thank you for pointing this out. The small ranges will be changed into the fixed value in Table 2, as follows:

Table 1 Summary of the calibration scenarios in case studies

Scenario in the synthetic case	Notation	Input error model in the synthetic data generation	Prior information of input error model in calibration
1	*add-fixed*	$\boldsymbol{X}^o = \boldsymbol{X}^* + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(0.2, 0.5^2)$	$\boldsymbol{X}^o = \boldsymbol{X}^* + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(0.2, 0.5^2)$
2	*add-inferred*		$\boldsymbol{X}^o = \boldsymbol{X}^* + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mu, \sigma^2), \mu \in (-0.5, 0.5), \sigma \in (0, 5)$
3	*mul-fixed*	$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(0.2, 0.5^2)$	$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(0.2, 0.5^2)$
4	*mul-inferred*		$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(\mu, \sigma^2), \mu \in (-0.5, 0.5), \sigma \in (0, 5$
Scenario in the real case	Notation	Input data source in the real case	Prior information of input error model in calibration
1	*O-fixed*	Observations from the rating curve (USGS database)	$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(0, \sigma^2), \sigma$=0.103
2	*O-inferred*		$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(0, \sigma^2), \sigma \in (0, 1)$
3	*S-fixed*	Simulations from a hydrological model	$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(0, \sigma^2), \sigma$=0.764
4	*S-inferred*		$\boldsymbol{X}^o = \boldsymbol{X}^* \exp(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon} \sim N(0, \sigma^2), \sigma \in (0, 1)$

Technical corrections

1) L. 128: double ","

Thanks, the redundant "," will be deleted.

2) L. 142: "sth" -> make "s" italic

This will be changed to be italic.

3) Eq. 8: unspecified symbol

We will remove this unspecified symbol.

4) L. 314: "q increasing until the objective: : :" -> incomplete sentence

This will be corrected as follows:

"*Considering these two points, the BEAR method set q iterations in the algorithm (Fig. 1), and q increases until a defined target is achieved* ."

Appendix A: The illustration of the BEAR method

Table A 1 An example illustrating the BEAR method

		1st iteration (random sample)																		
row	time step i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
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 ε	-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 ε	-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 ε	-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 Table A 1 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 2nd 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. Figure A 1 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 1st time step (i=1) in the 3rd 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.

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

(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. Figure A 2 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.

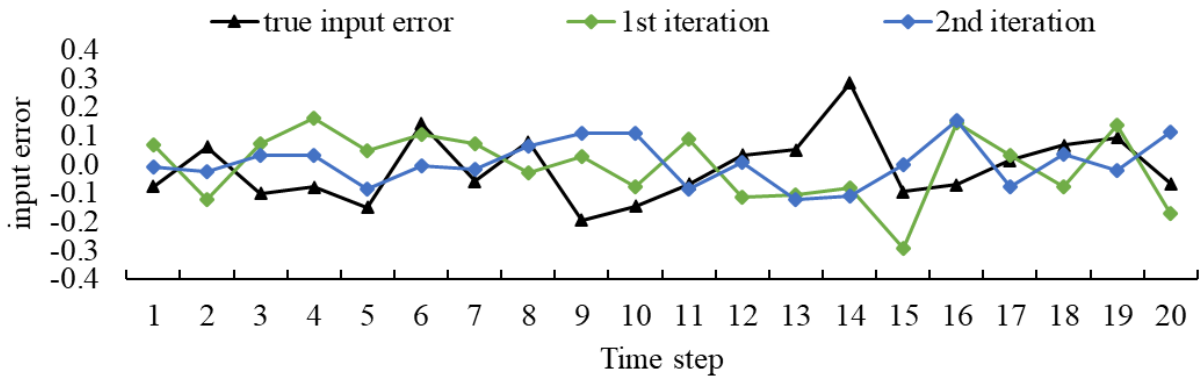

Figure A 1 Demonstration of the input error estimation in Table A 1 at the 1st and 2nd iteration where the input errors are randomly sampled

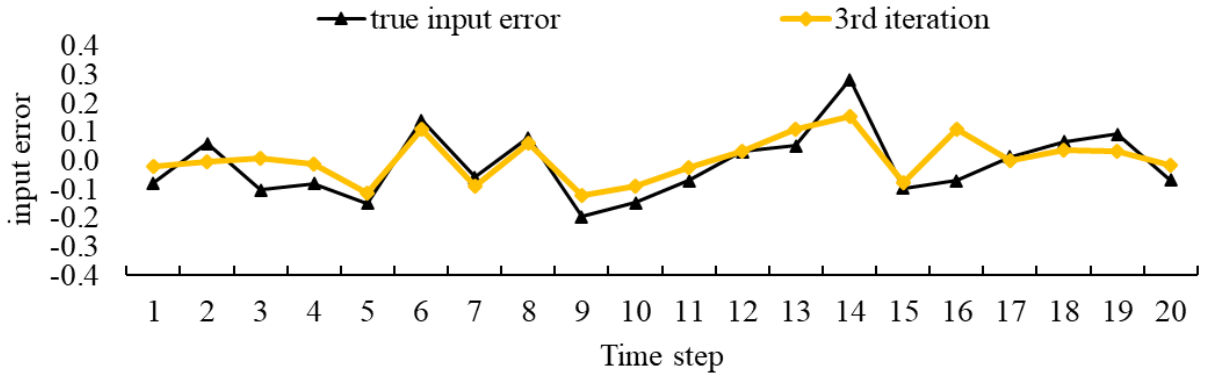

Figure A 2 Demonstration of the input error estimation in Table A 1 at the 3rd iteration where the input errors are reordered according to the updated error ranks