Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-563-RC1, 2020


Hydrology and Earth System Sciences
Discussions

# *Interactive comment on* "Quantifying input uncertainty in the calibration of water quality models: reshuffling errors via the secant method" *by* Xia Wu et al.

**Anonymous Referee #1**

Received and published: 9 December 2020

Review of "Quantifying input uncertainty in the calibration of water quality models: reshuffling errors via the secant method" by Xia Wu, Lucy Marshall and Ashish Sharma

**Summary** 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:

- 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.
- 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.
- 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 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?
- 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)
- 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.

**Specific comments**

- L. 37-38: “. . .estimate the residuals between the measurements and proxy values. . .” -> yet, measurement error is not addressed
- L. 68: “variable” -> “scalar“– both, vectors and scalars represent variables
- Eq. 3: unnecessary, since given by equation (1)
- L. 84-91: repetitive, add details to the corresponding paragraph in the introduction
- L. 92: “innovation” -> rather “introduction” or simple “The secant method” as chapter header – the innovation was made before
- L. 98-99: Rank definition and concept -> requires further explanation



- 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.
- 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.
- L. 146: “. . .when 1000 proposed parameter sets. . .” -> is this suggested as general approach or an arbitrary choice for this study. Please explain.
- L.171ff: Please replace abbreviations T, D and R by their names. With all abbreviations that follow it is hard to keep track.
- LL. 190+196: “Sikorska et al, 2015” $\rightarrow$ missing in references
- Eq. 9: define parameter “b”
- L. 215ff: incomplete sentence
- 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
- 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
- L. 282-283: “This illustrates that the impacts of other. . .” -> unclear phrase, please clarify and re-formulate
- L. 291: “. . . could be regarded as the reference value.” -> Why? Please explain.

- 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?
- L. 297ff. “Compared with the IBUNE framework. . .” -> there is no real comparison made, please see major comments
- 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.
- 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?
- L. 358: “developed” -> “proposed” – the methods are already known but used in a way to address input error here.
- L 362: “. . . addresses the high dimensionality problem. . .” -> not shown

**Figures**

- General: Legends in figures should be improved, e.g. in terms of colors or placing
- General: Provide higher resolution and unify the legend (see especially Fig. 4 and 6)
- Figure 3: please use colors that are better distinguishable (see cases “T” and “R”)
- Figure 3(4): NSE = 1 is unrealistic. Please see major comments.
- Figure 4 (c3,c4): model predictions are clearly shifted. Please elaborate on this offset.



- 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.

**Tables**

- Tables 1-1 and 1-2: The tables could be presented as additional files but are not helpful in the main article
- Table 3: the “fixed” scenarios in the real test case are not fixed but provide small hyperparameter ranges

**Technical corrections**

- L. 128: double “,”
- L. 142: “sth” -> make “s” italic
- Eq. 8: unspecified symbol
- L. 314: “q increasing until the objective. . .” -> incomplete sentence

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-563, 2020.