Dear Editor,

Thank you very much for your email of June 15, 2018, informing us of valuable suggestions to improve our manuscript 'Improvement of model evaluation by incorporating prediction and measurement uncertainty' (hess-2017-66). We wish to express our gratitude to the Hydrology and Earth System Sciences for encouraging us for the revised version of this paper, as well as to anonymous reviewers who provide valuable suggestions and professional revision of our manuscript. We have carefully considered your comments and revised the manuscript. The detailed responses are as follows:

===

Response to the editor

1) Comments: I would suggest to highlight methodological contribution to the literature in the manuscript.

Response: Thank you for the valuable suggestion. As suggested, we revised the manuscript accordingly. Please find the attached manuscript. In fact, the objective of this study is indeed to develop a new framework for model evaluation by incorporating prediction and measurement uncertainty. This methodological contribution is because in traditional indicators (such as Nash–Sutcliffe model efficiency), the deviation between the measured and predicted data is expressed by the absolute distance $(O_i - P_i)$ between the paired data points. This method is questionable because it fails to incorporate prediction and measurement uncertainty. Thus, the idea behind the CDFA was to replace the point-to-point comparison with the deviation between uncertain measured data and predicted data expressed as cumulative distribution functions. In fact, this is a modification of traditional good-of-fit indicators by replacing the calculations of their $O_i - P_i$ term by using stochastic distances between the paired probability density functions (PDFs). Thus, this CDFA could be used during the calibration and validation process if PDFs could be obtained for both prediction data and measurement data. Based on the results obtained from this study, we found that the model performance worsened when a larger error range existed, and the choice of PDF affected the model performance, especially for non-point source (NPS) pollution predictions. These proposed methods could be extended to other goodness-of-fit indictors and other watershed models to provide a substitution for traditional model evaluations within an uncertainty framework. Thus, the authors do believe our method could be a substitute of traditional goodness-of-fit indictors and they could be used for the calibration and validation process in the future.

===

Response to the reviewer 2

1) Comments: The proposed approach is more about a post calibration process for model evaluation rather than a new technique that can be implemented into the model during the calibration to gain knowledge from prediction and measurement uncertainty. This should be clarified in the manuscript. I would also suggest to comment the technical complexity in implementing this approach within the model (may be in the conclusion).

Response: As suggested, we revised the conclusion section in order to make clear description of this proposed approach. The revised conclusion is as follows:

"In this study, two new methods were proposed and employed to evaluate model performance within an uncertainty framework: the CDFA and the MCA. Using the CDFA and the MCA, both prediction and measurement uncertainty could be considered for model evaluation in a post calibration process, and the possible impacts of error range and the choice of PDFs could be quantified for a real application. Based on the results, the model performance worsened when a larger error range existed, and the choice of PDF affected the model performance, especially for NPS pollution predictions. These proposed methods could be extended to other goodness-of-fit indictors and other watershed models to provide a substitution for traditional model evaluations within an uncertainty framework. Thus, the new approaches could be a substitute of traditional goodness-of-fit indictors and they could be used for the model evaluation process.

However, it should be noted the proposed CDFA and the MCA would serve for model evaluation in a post calibration process rather than a new calibration technique due to the technical complexity in implementing this approach within the model calibration. With the results presented, fixed PDFs or error range for prediction data could not be founded due to insufficient knowledge and natural randomness. Thus modellers should better assess the error range of measured data for their use in watershed simulations, and more data should be gathered to obtain a real measurement error range and a proper PDF for the predicted data. Further explanations are also suggested for the inherent uncertainty of hydrological and pollutant transportation processes. More case studies should be conducted to test the IDA, CDFA and MCA in future practical analyses of other watershed models."

Thank you very much for your wonderful job. Hope that our responses are satisfactory. Best regards.

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