

This paper proposes a single objective function, which aims at producing realistic model simulation. The paper is well written and well organized, which makes it easy to read.

In my opinion, the methods used do not justify the main conclusion of the paper, which is that the proposed objective function is better than available alternatives.

Provided that the objective function introduced by the Authors is new, the way the goodness of this objective function is evaluated, and the way the exponent is chosen, is not convincing.

More specifically, the authors propose the objective function:

$$F_1 = \sum |Q_{obs} - Q_{mod}|^\alpha \quad (1.1)$$

Which generalizes the function

$$F_2 = \sum (Q_{obs} - Q_{mod})^2 \quad (1.2)$$

In order to select the best exponent α , the authors calibrate the model with different values of α , and pick the one that optimizes another objective function, CL.

1. CL is another single objective function (it is not a multi-objective function). Therefore, the first question that arises is why not using CL directly for model calibration?
2. In a second experiment, the Authors compare the simulations obtained by calibrating on the Nash Sutcliffe, to the ones obtained with calibration on F1. The comparison is done by evaluating model performance on the individual components of the function CL. However, because the exponent of the function F1 has been chosen as to optimize CL, one can say that CL has been used as an objective of model calibration. The comparison is unfair, as it is quite obvious that the model performance will be better with respect to an objective function to which the model has been optimized than with respect to an objective function to which the model has not been optimized.
3. In a third experiment, the Authors compare the simulation obtained with respect to F1 to the envelope of curves obtained by multi-objective calibration on the individual components of CL. They show that their optimal model lies within the envelope. Again, because the exponent of their function has been chosen to optimize CL, their simulation would lie very close to the Pareto front of the 4 objectives of CL. Again, this is not a strong test for assessing the quality of their objective function.
4. In a fourth experiment, the Authors perform time validation and show model performance on FDCs. This is, in my opinion, the only valid and independent test performed in this study, and it shows that the observed FDC can be very distant from the simulations. This test does not really support the conclusions of this study.
5. I think the Authors should provide other means to identify the parameter alfa in F1. For example, by choosing a value that minimizes the heteroscedasticity of the residuals.
6. The Authors should also look at the problem of estimating lambda in a Box Cox transformation, as this is very much related to their problem. Standard least squares leads to a maximum likelihood estimator that is essentially function F1 with exponent 2 (or equivalently, the Nash Sutcliffe efficiency). A box cox transformation results in a slightly

different likelihood, which depends on lambda, and there are different ways of estimating lambda, including that of maximising the likelihood.

7. The comparison of different objective functions should use widely used metrics, it is ok to use NS and CL, with inclusion of some other objective functions previously proposed, such as the Kling Gupta.
8. The evaluation should be done with metrics independent on the objective functions used, such as FDCs, QQ plots, indices that are deemed important by hydrologists such as the baseflow and flashiness index, and preferably in the calibration period.