

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2  
<https://doi.org/10.5194/hess-2021-601-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on hess-2021-601

Anonymous Referee #2

---

Referee comment on "A novel objective function DYNO for automatic multivariable calibration of 3D lake models" by Wei Xia et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-601-RC2>, 2022

---

### Overview:

The manuscript is well-written and clear in its intent and results. It is close to publishable if the authors place their approach to calibration in the appropriate context. I am classifying this as "major revisions" because I think the issues are important, but they should not necessarily take a lot of work to implement.

### Specific Comments

1. I would have liked to see the introduction have a little that explains the context for the authors choice to test their calibration against a calibrated model rather than against observations. This choice gives them more data to compare against, but with the drawback that the "true" data are biased in exactly the same way as their calibrated results. This is an acceptable, but limited approach -- acceptable because it is useful in understanding and illustrating the new calibration method -- but limited in that it cannot be used to say anything about what might occur when compared to real world data. This idea is emphasized in comment 2 below.

2. I strongly disagree with the penultimate sentence of the abstract, that the study "suggests that both temperature and velocity measures should be used for hydrodynamic model calibration in real practice." Similar language is found elsewhere in the paper. The model "suggests" nothing and we must remain skeptical about the value of calibration in a real world context without direct illustration of its importance. Unfortunately, the authors' methodology does not support this suggestion or any suggestion about real-world practice. The authors are testing their calibration against results of a calibrated model -- not the real world. The ability to more precisely capture the calibrated model (by a variant of the same model) cannot be used to imply that real world will be also represented more accurately. This is a fundamental confusion of "precision" -- how close are my answers to grouped together, with "accuracy" -- how close do my answers reflect the real world. The authors have not included any comparisons of their model to real-world data hence they cannot make any statements or suggestions about likely accuracy in representing real-

world phenomena.

2. If the authors want a stronger paper, they can either compare to real-world observations in a different time period as a classic validation test, or they can examine some important phenomena that are not directly included in the observational data set. For example, the timing of global overturns of the lake are arguably an important phenomena that can be computed from real-world observed data -- the model results for the different calibrations in predicting timing of overturns could be analyzed. I may be wrong, but I suspect that the differences between the various models may not be as significant when compared in their prediction of a large-scale phenomena. The paper would still be publishable, but future researchers would be able to see that control of model biases are likely more important than calibration in capturing real world behaviors.

3. The authors should include a comparison to an uncalibrated run (using conventional default values).

4. The authors should discuss the choice of calibration parameters -- how were they chosen, why were they chosen, and what does it mean to hold these as constant parameters. Arguably, a sensitivity study should have been done prior to choosing the calibration parameters. I am somewhat concerned about whether it is physically meaningful to calibrate as fixed parameters values that arguably depend on time-varying physics (e.g., Secchi depth, Ozmidov length scale, Dalton number, Stanton number).

5. The authors should provide some discussion about the physical meaning of the errors in the results. For example, the Ozmidov length scale in the "true solution" is 0.015, but the "best" calibration has more than double this value; at the same time, the background vertical diffusivity in the calibration is about 3/5 of the true value and the Secchi depth is overestimated by 13%. These all exert controls on vertical mixing, and the wide disparities for different calibration methods makes me question as to whether calibration can really be effective to capture the complex, time-varying behaviors of vertical mixing in a stratified system with the given turbulence model. I do not expect the authors to solve such a problem, but I think they should discuss the implications. Replacing Table 5 with a well-constructed bar graph (with appropriate normalization) would help the readers see which parameters are doing all the work.

Path to acceptance:

The authors need to carefully limit their language so that the manuscript reflects what can truly be understood from the model and refrain from speculation about real-world behaviors unless they specifically bring new real-world comparisons into the manuscript.