Manuscript ID: hess-2021-601

## Response to Reviews

We appreciate and would like to thank two reviewers for taking the time and effort to review our revised manuscript. We have improved our manuscript based on reviewer's helpful comments. Our point-to-point response are below (comments of the referee are in black, our responses are in purple, texts quoted from the revised manuscript are in italic).

## Point-to-point response

## Response to referee comment Referee #1

**R1-1:** The new version of the manuscript by Xia et al. has addressed all comments I raised. The authors included several new paragraphs based on my questions for clarification as well as those from the other reviewers. I see this made their paper way more accessible to a wider audience. I have no concerns in recommending the publication of the paper in its present form, except for a few language corrections which should be accomplished. I list some of them below but I recommend the authors to double check the manuscript and pay some more attention to the new paragraphs as they include several typos.

Response: Thanks for the reviewer's comments. We have revised the English language based on reviewer's detailed comments below. We also did a careful proofreading for the revised manuscript.

**R1-2:** L343-344: "Because the change of objective function value for all evaluations in $\boldsymbol{\psi}$," a verb is missing here

Response: We have revised the sentence to be: "*Because the objective function value for all evaluations in* $\boldsymbol{\psi}$ *changed after each iteration*, …" (In L343-344 of the revised manuscript)

**R1-3:** L351: missing "is" in "since it more likely"?

Response: We have added the missing "is". (In L351 of the revised manuscript)

**R1-4:** L353: missing "were" in "that not explored"?

Response: We have added the missing "were" (In L353 of the revised manuscript)

**R1-5:** L362: missing space in "thefitting"

Response: We have added the missing space between "the fitting" (In L362 of the revised manuscript)

**R1-6:** L447-452: please revise the English of the entire comment to Figure 5. What do the authors mean with "is worst"? Maybe they meant "worse"? Still it's not clear what they mean with worse/worst. Please comment the figure in a more clear way. In line 450 "Cali-Tem in worst" maybe that "in" was a "is"?

Response: We have added the explanation of what "the worst" mean. We "the worst" is defined in terms of how far the parameter value of the optimization solution is from the true solution. We have revised the entire comment to Figure 5 to be: "*In general, the solution in the Cali-Both scenario is closer to the True solution than solutions in the other two scenarios in terms of parameter values. Calibrated values proposed by the Cali-Both scenario are closest (relative to other scenarios) to the True values for four parameters (i.e.,* $\boldsymbol{D_H^{back}}$, $\boldsymbol{D_V^{back}}$, $\boldsymbol{H_{Secchi}}$, $\boldsymbol{c_H}$*). Moreover, besides the Manning coefficients, the calibrated values proposed by the Cali-Both scenario are not the worst (relative to the other two scenarios) for any other parameter. In contrast, calibrated values proposed by the Cali-Tem scenario are worst (i.e., the parameter values are farthest from the true solution, relative to other scenarios) for five parameters (i.e.,* $\boldsymbol{v_H^{back}}$, $\boldsymbol{D_H^{back}}$ $\boldsymbol{v_V^{back}}$, $\boldsymbol{D_V^{back}}$, $\boldsymbol{c_e}$ *) and calibrated parameter values for the Cali-Vel scenarios are worst for* $\boldsymbol{L_{oz}}$, $\boldsymbol{H_{Secchi}}$, $\boldsymbol{c_H}$. *This indicates that calibrating to both temperature and velocity can help to prevent the value of the 9 calibration parameters from being very far from the corresponding value for the True solution.*" (In L445-L454 of the revised manuscript)

**R1-7:** Figure 5: Please add the legend for the lines

Response: Thanks for reviewer's comments. We have added the legend for the lines in Figure 5.

**R1-8:** L551: getS

Response: We have added the missing "s" after "get" (In L553 of the revised manuscript)

**R1-9:** Figure 7: missing "is" in the description of the final solution hexagon

Response: We have added the missing "is" in the caption of Figure 7. (In L581 of the revised manuscript)

**R1-10:** L727: it worthS

Response: We have changed it to be "it is worth mentioning that …"(In L729 of the revised manuscript)

**Response to referee comment Referee #2**

**R2-1:** I want to thank the authors for putting great effort in the revision of the manuscript. Although I think that the authors addressed most of the technical remarks well, the major issue that is in my opinion present in this study requires further discussion. I am afraid that the authors reply does not fully dispel my doubts in the study design. I will again address this issue in the next section. In the following section I address minor line-by-line comments.

Response: Thanks for reviewer's time and careful review. We have revised the manuscript based on reviewer's comments on the study design. We have responded in the next comment. We have also revised manuscript based on reviewer's minor line-by-line comments below.

**R2-2:** I appreciate that the authors stress the fact of using only synthetic data in their case study in the revised version of the manuscript. Yet, I would like pick up the discussion on the synthetic case study and come back to the arguments that were in my opinion not fully addressed.

I raised two major concerns in my previous review: i) synthetic observation data that were generated with the same set of equations that are then fitted to these date would favour a combined calibration (velocity and temperature), as both data would constrain the function space stronger than each of the variables individually; ii) real world settings are usually trade-off problems in the calibration of multiple variables.

While I can accept the reply to argument ii), I think the reply to argument i) misses the main issue. The authors argue that the model is physically based and should therefore be capable of reproducing observations of physical variables. Further, as the model is physically based it is capable of generating realistic data that can be used as observation data in a test case. The authors also refer to Baracchini et al. (2020) as this study supports the use of velocity data and temperature data for a 3D lake hydrodynamic model calibration.

The theoretical argument may be reasonable, but it misses the fact the measurement uncertainty in real observation data can be substantial. The study of Baracchini et al. (2020) used measured observation data for velocity and temperature. Baracchini et al. (2020) mention that measured flow velocities are often close to their measurement accuracy in one of the case studies while the difference between the variance in temperature measurements and the measurement accuracy is substantially lower. A synthetic example that uses simulated data as observation data cannot account for this property of the observation data. To make the synthetic study more comparable to a real case setting overlaying the synthetic data with an error model may be a probable solution. Although I fully understand that this is likely infeasible to be still considered in this study this aspect should be addressed. I think the distributions of the observation data and their uncertainties can strongly influence the performance of the simultaneous calibration.

Response: We sincerely thank the reviewer for elaborating their argument pertaining to the design and use of synthetic data in this study. We do agree with reviewer that our synthetic example did not account for uncertainty of observation data, and we agree that this point warrants a discussion in the manuscript.

The reviewer also mentioned the discussion of Baracchini et al. (2020) in relation to accounting for the measurement uncertainty of temperature and velocity. We believe that their discussion can be elaborated further, and in-fact, supports the use of DYNO. In Baracchini et al. (2020)'s study, the cost function is the square of temperature (or velocity) difference between computed and measured value divided by the observational uncertainty. The observational uncertainty in Baracchini et al. (2020)'s study is defined as the maximum value of two elements: 1) the instrument precision, and 2) the temporal dynamic variability (i.e., standard deviation of observations) at the measurement location over a short-term period (e.g., ±6h to ±12h). The first element is a fixed value in their case (0.002℃ for temperature and 0.8cm $s^{-1}$ for velocity). . Baracchini et al. (2020) mentioned that the computed and measured velocity value is close to velocity measurement accuracy of 0.8cm $s^{-1}$ (Teledyne RDI Workhorse Sentinel) while

temperature computed and measured value is orders of magnitude larger relative to temperature measurement accuracy in their study. Hence compared with temperature observations, velocity has much less impact on the cost function than temperature and therefore it might have an issue to calibrating both temperature and velocity simultaneously. In our opinion, Barachhini et al. (2020)'s discussion and findings add more value to our study. The issue of velocity having much smaller impact on the cost function than temperature is true for traditional approach of using a fixed weight in adding the cost function of temperature and velocity together as the objective function. However, the proposed new objective function in our study handles the error of each variable separately (the error functions may also account for measurement error) and dynamically normalized their error so to balance their impact on the objective function (and hence their impact on calibration). The issues mentioned by Barachhini et al. (2020) related to the difference in terms of the magnitude of temperature and velocity's cost function value is exactly what our new proposed objective function (DYNO) attempts to rectify. We thank the reviewer for bringing up this issue since it reminded us to point out in the manuscript the advantage of the new objective function DYNO to incorporate the error in each variable.

However, as mentioned earlier, we do acknowledge that measurement errors are important and should be incorporated in future studies to understand effectiveness of DYNO in multi-constituent calibration settings.

In Section 3.6 of our revised manuscript we have added following discussion: "*Moreover, the synthetic observation data used in our analysis did not account for the measurement uncertainty of observation data. Further investigations related to the impact of measurement errors on the calibration setup proposed here will also be beneficial. It is important to note that the measurement uncertainty and distribution of different variables could be different (and thus, our new objective function formulation DYNO could be very useful in balancing the calibration process in such a scenario). For example, Baracchini et al. (2020) reported that the measured and computed velocity value (in the magnitude of 1 cm* $s^{-1}$ *for velocity in hypolimnion layer) is close to velocity measurement uncertainty 0.8cm* $s^{-1}$ *(the velocity measurement instrument precision) while the computed and measured temperature value is an order of magnitudes larger relative to temperature measurement uncertainty in their study. The difference in terms of measurement accuracy and measurement value could lead to a different magnitude of error function value for each variable (temperature or velocity). (In their study, the error function is the square of temperature (or velocity) difference between computed and measured value divided by the observational uncertainty). Baracchini et al. (2022) pointed out that such discrepancy hinders the use of different kinds of data (e.g., temperature and velocity) simultaneously because the impact of velocity on the cost function is almost negligible compared with temperature observations. Hence, they carried out a separate discussion for both types of observation data. Their argument is true if the calibration objective function is a sum of temperature or velocity's error function with a fixed weight. In this case, the difference of the error function value's magnitude might lead to a biased calibration to the variable that has a larger impact on the error function.*

*However, our proposed new objective function DYNO dynamically normalizes the error function value of each variable using the maximum and minimum value of each variable's error*

*function value obtained during the calibration and hence balances the impact of each variable on the objection function. Hence DYNO is designed to work well in scenarios where the error function values of each variable are significantly different due to differences in measurement uncertainty and the distribution of each variable's observations."*

Baracchini, T., Hummel, S., Verlaan, M., Cimatoribus, A., Wüest, A., and Bouffard, D.: An automated calibration framework and open source tools for 3D lake hydrodynamic models, Environmental Modelling & Software, 134, 104787, 2020.

Line-by-line comments

p.4 L141 - L143 Please rephrase the sentence. The formulation (e.g. don't) is highly informal.

Response: We changed the sentence to *"Real observations for velocity are not available in our case as well."* (In L141 of the revised manuscript)

p.8 L256 Remove assume.

Response: We have removed "assume" (In L256 of the revised manuscript)

p.9 L266 This section gives...

Response: We have added the missing "s" after "give" (In L266 of the revised manuscript)

p.13 L382 Remove about. It is exactly 8 times 24.

Response: We have removed "about" (In L382 of the revised manuscript)

p.14 L427 Please revise or remove this sentence. It sounds unnecessary and very vague to me.

Response: We have removed the sentence in Line 426-427 of the manuscript in last submission.

p.14 L443 Please rephrase 'is set be the middle'.

Response: We have rephrased the sentence to be: "The parameter value of the uncalibrated solution (in Table S1) uses the mean of the calibration range in Table 2" (In L442 of the revised manuscript)

p.15 L447 ...in the other two scenarios...

Response: We have added the missing "the" in "… in the other two scenarios …". (In L446 of the revised manuscript)

p.15 L450 - L452 Please revise the two sentences. The wording sounds odd. Avoid using vague wording such as 'good' and 'bad'.

Response: Thanks for reviewers' suggestion. We have revised this two sentences to be: "*In contrast, calibrated values proposed by the Cali-Tem scenario are the worst (i.e., the parameter values are farthest from the true solution, relative to other scenarios) for five parameters (i.e.,* $\boldsymbol{v_H^{back}}$, $\boldsymbol{D_H^{back}}$ $\boldsymbol{v_V^{back}}$, $\boldsymbol{D_V^{back}}$, $\boldsymbol{c_e}$*) and calibrated parameter values for the Cali-Vel scenarios are worst for* $\boldsymbol{L_{oz}}$, $\boldsymbol{H_{Secchi}}$, $\boldsymbol{c_H}$*. This indicates that calibrating to both temperature and velocity can*

*help to prevent the value of the 9 calibration parameters from being very far from the corresponding value for the True solution.*" (In L450-L454 of the revised manuscript)

p.23 L671 - L675 Please revise this section. The formulation (e.g. don't) is highly informal. Further, either it is a physical law or some empirical observation by a human. I would avoid this formulation.

Response: We have revised this section to be: "*Our analysis is based on synthetic observation data from the physical model since we do not have real velocity measurements. These physical models are based on physics laws. The analysis from modelling can provide some implications for the real-world situation. Hence, it is worthwhile to repeat the analysis based on real data if there are real velocity measurements available in the future.*"