Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/hess-2021-601-RC3, 2022

Comment on hess-2021-601

Christoph Schürz (Referee)

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-RC3, 2022

Short summary of the manuscript

The authors present a new metric, the Dynamically Normalized Objective function (DYNO), the can be implemented in multi-variable calibration problems. Target variables in model calibration usually differ in their value ranges and error structures which can impede a balanced calibration for all variables. In an optimization procedure DYNO dynamically normalizes the error metric of each calibrated variable by the range its error metric that was identified until the respective iteration step. This range is dynamically updated in each iteration step. The normalized values are summed into a single metric that can be used in an optimisation search routine.
The proposed metric is tested in a synthetic hydrodynamic lake model study, where 9 model parameters were optimized by minimizing the the errors between synthetic observation velocity and temperature data and the simulated values at different locations and depths.

General evaluation and major comments

Overall, I think the manuscript is well structured and the methodology is well prepared. The presented tables and figures support the findings that were presented in the text and provide good insights in the functionality of the presented algorithm. While I think the overall quality of the manuscript in its present form is already good, I see some crucial points that require clarification. Other less significant points should also be improved to improve the quality and readability of the manuscript. I will outline my major concerns in the following and will address smaller issues in a line-by-line notation in the following section.

- *Synthetic study design:*

The observation data were generated with the same model structure that was later on investigated in the case study. Which means that "observed" variables (in this case velocity and temperature) were calculated with the exact same set of equations that in the following calculated the simulated variables. I think this property of the observation data could potentially impair the entire analysis and favour the simultaneous calibration of velocity and temperature.
As the authors outline in section 3.5 of the manuscript, multi-variable calibration problems are often affected by trade-offs between the variables for which performance metrics should be optimized (minimized). One reason for that is that models are simplification of the represented reality and parameters can affect multiple processes, sometimes in the opposite directions. Thus a change of a parameter value into one direction could improve the performance of one metric, while deteriorating the performance of another metric at the same time. This is different in the synthetic example where in fact the reality and the model are the same thing. Thus the "observed" time series of velocity and temperature perfectly agree with the model simplifications and assumptions.
Thus, in such a case providing both, temperature and velocity, to the search algorithm should better constrain the parameter response surface than only providing one of the two variables. If this hypothesis is sound, then the presented results would by affected by this effect. In this case only a real case scenario could provide an honest comparison of the three cases.

- *Number of iterations of the search algorithm*
  This comment is somehow related to the previous one. In theory it should be possible for the search algorithm to identify the "true" parameter set that was used to generate the synthetic observation data. This global minimum is present on the parameter response surfaces of all three calibration scenarios and has a value of 0 independent of the used metric (DYNO or the single performance metrics). Thus, when not ending up in a local minimum all three cases should converge towards this global minimum. As outlined above, I simply think that the Cal-Both scenario does this quicker due to the given reasons.
  In the presented results all "best" solutions did not find the global minimum. As far as I got it right from the text, each experiment involved 8 iterations with 24 parallel evaluations in each iteration step. Given a computation time of 5 hours per simulation run (according to the text and thus 5 hours per parallel iteration) one experiment takes 1.6 days. I am wondering if experiments with larger numbers of iterations were performed (that are maybe just not shown). I would be interested how the convergence of the three calibration experiments develops with larger number of iterations.
- *Consistency in the nomenclature*
  Overall I think the outline of DYNO and the explanation of variables is well done. Yet, I found some inconsistencies in the nomenclature and some mathematical definitions that must be improved. I will address these (at least the ones that I found) in the line-by-line comments.
- *English language*
  Overall I think the manuscript is well written. In some sections of the manuscript I found that the same errors repeat in every second line (e.g. singular/plural, missing articles). Although I am not a native speaker myself I had the impression that the manuscript could require some proof reading. Some sentences I had to read over and over again, but they still do not make sense to me. I addressed them as well in the line-by-line comments.
- *Non-color blind color theme in figures*
  Although this does not affect me, I would advise to change the color theme in the figures and refrain to use red and green at the same time (8% of males are affected by red-green color blindness).

Line-by-line comments

p.1 L13, p.4 L123 and further: Please use a consistent naming for Dynamically Normalized Objective function. Either all first letters caps or none. It differs throughout the text. Also only use the acronym after it was first introduced in the text.

p.1 L21 - 22: Please rephrase this sentence. Further, I think the statement is not universally true how it is formulated, that calibrating only one variable does not improve the calibration of another variable.

Eq. 2, Table 1: Consistent notation of *Sim* and *Obs* necessary. Either capital first letter or lower case.

Table 1: Please rephrase the sentence: "The set of variables the observation data of which is used in calibration"

Table 1: I would suggest to add the vector $\{x_1, ..., x_d\}$ to the definition of $\boldsymbol{X}$ as e.g. Table 2 refers to them.

Table1 and throughout the text: I would suggest to use the nomenclature $Sim^k_{j,t}$ (same for *Obs*) in the definition and throughout the text to indicate the time dependency. This nomenclature is in fact introduced later in the manuscript. Also the definition refers to the time *t* which is not indicated in the variable.

p.5 L153: ...multiple variables *with* a single objective function.

p.5 L159: Please remove \textit{value} as the distribution of NRMSE is not a single value.

p.5 L160: Same as above.

p.5 L160: Hence **,** ...

p.5 L163: Again consistent naming of DYNO or removing the full name as the acronym

was already introduced.

p.5 L166: ...all evaluations in $\boldsymbol{\psi}$ found so far?

p.5 L166: The variable $\boldsymbol{\psi}$ was not introduced at this point and is I think not defined in the manuscript.

p.5 L167: *The* mathematical formulation...

p.6 L172: ...all evaluation*s* ...

p.6 L172 - 174: Please rephrase this sentence. It is in my opinion not clear what is meant.

p.6 L180 - 185: The procedure as described is confusing and does not really make sense to me. A calibrated model exists from a previous study. This calibrated model was taken, but different parameters based on "expert guessing" were then used to run the model. So it is not the calibrated model anymore?

p.6 L186: Replace *a* by *the* as you did not use a parameter set but the parameter set $\mathbf{X}^{\mathbf{R}}$.

p.6 L193: ...the true value*s* of the parameter vector $\mathbf{X}^{\mathbf{R}}$ *are*...

p.6 L194: I would replace *value* with *set*.

p.6 L202 - 203: This sentence is not clear to me.

p.7 L226: coefficient*s*

p.8 L229: *the* Manning formulation

p.8 L229: ...which is a parameter tha should also be calibrated.

p.8 L232: ...*are* parameterized...

p.8 L237: *M*anning coefficient

Section 2.5: This section is actually a repetition of L146 - 150

p.9 L280: Remove *them*

p.9 L287: I would suggest to replace *a roughly* by *an approximately*

Caption Table 3: formulation*s*

p.10 L309: Use the acronym DYNO instead

p.10 L317: Replace *is* with *are*

Figure 2: The positions of the labels *Yes* and *No* in the stopping criteria are not clear to me.

p.12 L359 - 366: The argumentation in this paragraph could be affected by the affect of synthetic observation data as outlined in the first major remark.

p.13 L395: Missing end '.' of sentence?

p.13 L395 - 399: This sentence is confusing and might require rephrasing.

p.14 L406 - 416: I had the impression that many articles were missing here. This paragraph is hard to read in general and could potentially be revised.

p.15 L431: *the* calibration...

Figure caption Fig. 5 L451: in term*s* of...

Figure 6: Values in the darkest hex tiles are almost not readable.

p.19 L536 - 539: This sentence was not clear to me and might require revision.

p.20 L553 - 566: Again, the argumentation in this paragraph could be affected by the affect of synthetic observation data as outlined in the first major remark.

p.20 L573: Remove *these*

p.20 L574: suggest *to* have

p.21 L613: We conclude that the Dynamically Normalized Objective Function *that* we propose