Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-685-RC3, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A robust objective function for calibration of groundwater models in light of deficiencies of model structure and observations" *by* Raphael Schneider et al.

Anonymous Referee #3

Received and published: 12 March 2020

Comments on the paper Hess-2019-685 entitled: "A robust objective function for calibration of groundwater models in light of deficiency of model structure and observations", by R. Schneider et al.

This work intends to show that the classical objective functions (OF) in the inversion of subsurface flow, such as the sum of squared errors (SSE) between simulated and observed heads, or the sum of absolute errors (SAE), are functions mainly dominated by a few large errors. If these errors are stemming from structural model discrepancies, then the inversion procedure would compensate on model parameters to lower the OF, but with sometimes the downside of rendering awkward or unphysical solutions.

Therefore, it is proposed to rely upon an OF based on the continuous ranked probability score (CRPS) reputed less sensitive to large residuals, as it measures the squared distance between the cumulated probability density (cumulated statistical distribution) of model outputs and its equivalent in terms of local observations (usually, a Heaviside function). A few examples of this reduced sensitivity to high residuals are provided on the basis of very simple examples such as a series of five values, or a continuous Gaussian distribution. Then, a comparison of CRPS, SSE, and SAE is carried out for two inversion problems dealing with actual watershed systems.

It seems interesting employing the CRPS, usually devoted to the analysis of multiple equiprobable realizations of a single variable, in the framework of a single realization of a single variable but distributed over time and space. However, in my opinion, the study partly misses its target because the applications are a priori free from model structural errors; at least, these errors are not explicitly considered in the analysis of the inverse sought solutions.

I have a few concerns of various importance with the present writing, and some specific points (in a non-exhaustive inventory) that let me think that the contribution is not mature enough for rapid publication in HESS. My suggestion is that the paper needs major revisions, including new numerical investigations, and a further complete round of review.

Regarding the main concerns:

1- The notion of structural error is not well defined. In a first approach, one could consider that structural errors are all the errors that do not directly target model parameter values. This could include: errors on the geometry of the modeled system, on initial and boundary conditions, and on source-sink terms. One could also consider that structural errors are those associated with features hardly inverted in view of their direct influence on the observed state variable. In that case, one could remove initial and boundary conditions, but also source-sink terms from the structural errors, as these characteristics of the model can be inverted in view, here, of hydraulic head measurements. Finally, in the specific case of the reported study, the model parameterization relies upon a parameterization of the zonation type, building a “block” system with uniform parameters within each block. A flawed delineation of these blocks could also be considered as a structural error. Even better, one could suggest that for the two actual tests cases reported by the authors, errors in the delineation of uniform blocks could be the main structural error generating high residuals on heads that will never be compensated by tuning the model parameters of each block. In the end, it seems important to better state what is meant by structural errors, then deliberately generate these errors in exploratory calculations before checking on the performance of a CRPS-based objective function.

2- The two actual test cases discussed in the paper are redundant, mainly because they deal with watershed systems of the same size, with the same density of streamflow routing in their surface compartment, and a very similar density of evenly spread locations monitoring the subsurface waters. Why to report on both? The authors would have been well advised to focus on a single system, and consider that an inverse solution becomes some kind of reference problem to which structural errors are added. Here, the first structural error I would give a try would be that of a flawed parameter zonation. Then, by providing us with a metric on model parameters distinguishing values inherited from the “reference” and the “flawed” problems, some proofs that CRPS outclasses SSE and SAE could be made available.

3- My understanding is that in many locations within the modeled system, the authors (for the principle of parsimony?) lump the measurements of heads at various times and in various layers of the subsurface to build an averaged information. I doubt that this information has the sensitivity of a single observation to both parameter and structural errors. Let us take for example the case of a point measurement of head located not so far from a boundary condition. This condition is flawed and prescribes a Neumann-type boundary with prescribed fluxes instead of a Dirichlet-type condition with prescribed

head. The Neumann flux is not sufficient in the wet periods to feed the system, but too high in the dry periods, thus rendering negative (positive) errors on the head at a short distance in the winter compensated by positive (negative) errors in the summer. As a result, the structural error is not seen by the data, as would render the true Dirichlet condition able to feed the system at will. This example is just for showing that averaging various measurements is probably not a good idea to reveal that structural errors exist. I must acknowledge that I have never seen in the literature inversions taking averaged errors over large periods at some locations as the basis for an OF. I guess that it is “dangerous” to proceed that way, but probably my knowledge of the literature is not sharp enough.

4- The authors employ the same cumulative distribution of residuals to build their OF, irrespective of the location where the distribution is used to measure the performance of the model. This implies that the distribution of residuals should be stationary over space (which differs from the assumption of ergodicity associated with the inference of a CRPS on the basis of a single realization, but could also go with. . .). I doubt that in the presence of structural errors, e.g., local errors on the system geometry, or its boundary conditions, the statistical distribution of residuals would be stationary. If the authors are right, the distribution should not be stationary in being skewed toward high residual values in regions under structural errors. By the way, the CRPS should give less weight to important residuals in regions where structural errors are plaguing the convergence of the inverse problem by only tuning the model parameters.

In addition to the above general comments, I have a few specific comments (a non-exhaustive list), mainly as the consequence of lack of clarity in the writing.

1- Line 111, Eq. 1. The CRPS seems to be not well defined if it is supposed to serve as an indicator concealed in an OF. With an integral from minus infinity to plus infinity and an expected value of zero (optimal residual) the CRPS will remain the same irrespective of the location where it is applied. My understanding is that for a variable X (here a residual) and an associated bound x, the CRPS should write as the integral between minus infinity and x of (Ps(x')-Po(x'))ˆ2dx', with Ps(x') the probability for the variable X of not exceeding the value x'. In this case, and for a residual value x at a given location, CRPS(x) measures the distance between x and zero.

2- Lines 135-143, Eqs 2 and 4... If the significance of the dPi is well exemplified in Fig. 1 (with differences between the left panel (CRPS) and the right panel (MSE)), the text does not mention this difference. In a CRPS dPi is the cumulated probability of not exceeding the value xi, when dPi in a MSE is the probability of x being within an interval bounded by xi-1 – xi, or something of the kind. I would change the notation to avoid misunderstandings and be clear on that in the main text.

3- Line 169. What means "a description of the unsaturated zone" in MIKE SHE, a simplification, a 3-D resolution of the Richards equations? A short explanation should be given as a reminder. Integrated hydrological models coupling surface and subsurface flow have many options to handle the subsurface including the vadose and the saturated zones, and very often these options condition how two different models respond differently to the same forward problem.

4- Section 3 "Model and data". As told earlier, I think that presenting a single model for a single study area would be enough. In general, the overall depiction of the models in terms of hydrological context is very poor. The reader ignores what are, for example, the mean discharges of the stream at the outlet of the system, their seasonal variability, the overall variability of heads within the subsurface, what is the hydro-meteorological forcing, what are the boundary conditions, etc. Even though the main question is not to go into the detailed features of the forward problem, a few words for fixing the context would be welcome. The hydrological context could condition the applicability of the CRPS as an OF; most of inverse problems are case-study dependents.

5- Line 175 -. It is stated that the hydrogeological model (the subsurface) encloses several "layers", which I think to be the representation of a geological stratification in the subsurface, with the consequence of generating vertical heterogeneity in the hydraulic

parameters. A few lines later, (230 and followings) it is stated that only "six different geological units' hydraulic conductivities are sought, which would mean that within a unit (a "block" sub-system), the conductivity is uniform over the various geological layers. Why to distinguish these layers in the model geometry if they are similar in terms of hydraulic parameters?

6- Line 235 and followings. The so-called "benchmark" appears here as drawn out of the blue. When the reader expects that it will be discussed on the application of the CRPS, SSE, and SAE, to the actual case-studies, a "synthetic" problem is presented based on the various responses of the OFs to continuous Gaussian distributions of residuals. In addition, the "benchmark" is not well presented at all, and the reader is required to conjecture on the calculations performed in the benchmark.

7- Section 4.2. Even though, associated tables and figures report on the fact that CRPS outperforms the other OF, all the material is in fact a blind test as we ignore what are the structural errors in the models rendering high residuals. As told earlier, I would focus on a single test-case, I would consider a given inverse solution as a reference problem, and then I would add deliberate structural errors, for example on the delineation of the unit blocks, by overestimating or under estimating the aquifer thickness is some areas, by modifying the boundary conditions, by artificially generating a few zones of preferential infiltration, etc... Then by inverting these various configurations, a comparison of the performances of the various OFs could be carried out. In the present form of the study, the CRPS appears better as a matter of fact completely dependent of the overall settings of the forward problem, but applicability to other contexts is compromised, and a better response to structural errors (even though these errors probably exist in the tested forward problems) is not proven.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-685, 2020.