Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-345-RC1, 2022

Comment on hess-2022-345

Keith Beven (Referee)

Referee comment on "Using simulation-based inference to determine the parameters of an integrated hydrologic model: a case study from the upper Colorado River basin" by Robert Hull et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-345-RC1, 2022

This is a highly sophisticated study, involving considerable work, that aims to introduce methods of simulation-based inference to hydrological models, methods that it seems have been used very successfully in other fields such as cosmology, particle physics and neuroscience. I have not taken the time to look at what has been done in those fields because it is clear from the current study that in hydrology no great advance has been made. The methodology effectively involves two important steps; 1. A dynamic emulator of a complex hydrological model (PAR-FLOW) (here a LSTM) and 2. A method of identifying a conditional joint parameter distribution (here a form of neural network). In both cases the aim is to greatly increase the efficiency of model calibration and in the latter case avoid the explicit specification of a likelihood function (albeit using a prior assumption of what that distribution should look like – here a simple bivariate Gaussian, which thereby implicitly implying a form of likelihood function or measure, though this is not discussed).

There are alternatives to both steps. Those potential alternatives are not compared in terms of efficiency but that is not the real problem with this study. The real problem is that it tells us nothing at all about simulating the Taylor River Basin in the upper Colorado Basin because the study uses only simulated data. The title of the paper is therefore already misleading. Indeed, all the problems of structural error or miss-specification of the model (both in terms of the process representations and their application at a 1km grid scale) and disinformation in the observations (especially since snowmelt is important in this region) are totally neglected. This is perhaps why there is no mention at all of any of the very many papers I have written about the limitations of physically-based models, the complex nature of responses surfaces (which however defined are NOT multi-Gaussian with real data), disinformation in observations used for model calibration and defining limits of acceptability.

While I certainly have no need of further citations, the more important point is that there is absolutely no point in publishing a paper that compares only model generated data with an emulator (particularly in just a 2 parameter space) without any resort to real observations. This is indicated by the threshold for the determinant of the posterior of 10^{-6}. Look at any response surface plots for any actual model applications, or dotty plots within GLUE applications, to see that this would be overwhelmed by model misspecification and observation errors.

This situation could, of course, be easily remedied by having a two part paper in which this first part is followed up by a second part that actually applies the method to the observational data. I suspect that this is already in preparation, but this part is not worth reviewing further without the 2nd part. I suggest this paper be rejected but that the authors be asked to resubmit in that 2-part form.

Keith Beven