Reviews of HESS-2019-301

Title: Dynamics of hydrological model parameters: calibration and reliability

The objective of this paper is two-hold: 1) to develop and test the strategies of sub-period model calibration to generate temporarily varying optimized parameter sets and 2) develop the method to assess reliability of the optimized parameter set by evaluating parameter convergence behaviors. The paper presents the calibration results for one Chinese basin (two more in supplemental material) with focus on those two goals.

My first impression was the paper lacks focus due to two different objectives (distracted by each other), so I would lean to suggest putting more focus on the first goal and then reduce tone on the issue on parameter convergence evaluation. However, I found the results presented in this paper are interesting and reasonable overall. I have several specific comments below. In addition, the authors should work on textural improvements including fixing grammatical errors, punctuations (excessive use of parenthesis), vocabulary, and most importantly, more conciseness throughout the paper. I think the manuscript requires major revision before publication.

Specific comments:

- I am not sure about a list of the past studies on sub-period calibration (Page 2 Line 23-24). The most relevant paper to this study would be *Merz et al.*, [2011]. Please re-evaluate which references should be relevant to this sub-period calibration topic. Introduction should emphasize this topic than convergency behavior.

- Method for partitioning of the simulation period into sub-periods are not described in this paper but seems to be climatological based, i.e., dry, wet periods, and backbone for this calibration strategy. There is no information on this. Although a great deal of this topic is in another publication by the same author, I would like to see some summary of the paper, including what variables are used for clustering, and very brief clustering methods.

- For scheme 2. I understood that this is the same as scheme 1 except that one selected parameter is optimized per sub-period and the others are optimized for the entire simulation period. It is not clear to me what the motivation for this scheme is. And also wonder which parameter is exposed to sub-period calibration and how it is selected? Please clarify.

- Minor comments on the figures. Figure 2: I am not sure panel c is needed. It does not add anything meaningful to me. Table 1 would be enough. Figure 3. Panel c is specific to SCE and not general and I don't understand well about panel d. Figure 4. I don't think panel b is necessary. Also, RMSE for FDC is normalized by something?

- The methodology of parameter convergence assessment (3.2.2) is very specific to SCE, but not seems to be for the other algorithms. I think the concept works for the other global evolution algorithms, including DDS and even for multi-objective algorithms. My

recommendation is to generalize more technical descriptions on the procedures so that it is more applicable to such other algorithms.

- The most of hydrologic models struggle with dry basin calibration. For US basin, see *Newman et al.*, 2015, 2017. Interestingly Figure 8 shows dry period calibration also struggle converging the optimizing parameter values. I think this is something to discuss and would suggest showing (or mentioning) performance metrics for each 4 period for Scheme 3 and 4. My speculation is much better performance metrics for the wet periods than dry period, and reason why scheme 0 and 1 produce poor performance is due to poor performance during the dry period.

Merz, R., J. Parajka, and G. Blöschl (2011), Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, *47*(2), doi:10.1029/2010WR009505.

Newman, A. J. et al. (2015), Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrol. Earth Syst. Sci.*, *19*(1), 209–223, doi:10.5194/hess-19-209-2015.

Newman, A. J., N. Mizukami, M. P. Clark, A. W. Wood, B. Nijssen, and G. Nearing (2017), Benchmarking of a Physically Based Hydrologic Model, *J. Hydrometeorol.*, *18*, 2215–2225, doi:10.1175/JHM-D-16-0284.1.