

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-401-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-401

Anonymous Referee #2

Referee comment on "Using a surrogate-assisted Bayesian framework to calibrate the runoff-generation scheme in the Energy Exascale Earth System Model (E3SM) v1" by Donghui Xu et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-401-RC2>, 2022

The manuscript performed a per-grid calibration of the E3SM ELM model against a global monthly runoff dataset. The calibration was enabled by developing surrogate models for each grid of the ELM using Polynomial Chaos Expansion to mimic the response surface, which was chosen to be the root mean square error of monthly runoff for each grid. Subsequent analyses examined the spatial distribution of calibrated parameters with higher sensitivity and parametric uncertainty effects on simulated runoff. The paper is well organized, clearly written, and deals with an important topic of calibrating ELM and similar models. However, I have some concerns regarding the accuracy of surrogate and its effect on calibration, as detailed below.

- Line 15: "The main methodological advance in this work is the construction of surrogates for the error metric between the ELM and the benchmark data". But this is not entirely new as using surrogate in this manner has been done previously, e.g. Wang et al. (2014); Razavi et al. (2012) and references therein.
- Line 111: what's the difference between surface runoff and surface water runoff?

- Line 195 - to reduce the log likelihood to least-squares regression, further assumption is needed, which might include constant and known sigma. Please verify.

- Line 197 - I am not sure whether 1,000 samples are sufficient for burn-in, since MCMC often requires a large number (e.g., tens of thousands) of samples to converge. Including some convergence check statistics or plots in supplementary material would be helpful. Also, what is the MCMC algorithm being used here? Please include a reference for reproducibility.

- Fig. 1& 2 - there's some discrepancy between RMSE given by the surrogate and by ELM. Studies have shown that even small surrogate error can lead to large deviation of the inferred parameter posterior from the "true" posterior (Laloy and Jacques, 2019). I realize that it is not possible to calibrate ELM at global scale, but it seems possible to perform some quick test to validate the surrogate modeling approach. For example: for a few grids compare the posterior obtained using PCE and using ELM; In Section 3.5, step #4, compare the RMSE of ELM simulation with that of PCE.
- Line 355: If I understand correctly, 10,000 is the number of runs of the surrogate. It is not necessarily the case if ELM is run, because the convergence rate may be different given the surrogate error (Razavi et al., 2012).

- Fig. 11 - it seems that the same period of 1997-2010 is used to calibrate the model and validate the optimal parameters. Is data available after 2010 for validation, so that validation data is independent from calibration data?

- Some paragraphs are indented, some are not.

References

Laloy, E., & Jacques, D. (2019). Emulation of CPU-demanding reactive transport models: a comparison of Gaussian processes, polynomial chaos expansion, and deep neural networks. *Computational Geosciences*, 23(5), 1193-1215.

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., & Miao, C. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environmental Modelling & Software*, 60, 167-179.