

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-205-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-205

Anonymous Referee #2

Referee comment on "Emulation of high-resolution land surface models using sparse Gaussian processes with application to JULES" by Evan Baker et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-205-RC2>, 2021

Major comments

In the manuscript titled "Emulation of high-resolution land surface models with sparse Gaussian processes with application to JULES," Baker et al. described a novel statistical emulation framework developed for land surface models that acknowledges the forcing data fed into the model, thereby providing predictions with high resolution. Land surface models generally have a resolution that is determined by the resolution of the forcing data. In this paper, the need of an emulator has been highlighted, as it is difficult for a land surface model to perform at 1 km resolution for the entire UK with high resolution. As I could understand, the output of the emulator is an estimation of the outputs of a numerical prediction model, when the numerical prediction model is computationally expensive. However, it is difficult to understand the overall development and calculation flow of the emulator, a limitation that could be overcome with a flowchart that would illustrate how the emulator calculates the covariance and ensemble average using the sample training data and how the Gaussian perturbation is added to it. In addition, it is difficult to understand the impact of tuning and calibration from Figure 3, as the lines overlap and are difficult to visualize. Finally, I suggest you indicate the several applications of the emulator by adding a summary and conclusion after Chapter 4. Overall, the manuscript needs revision before publication. Although there are some issues, I recommend the publication of this paper after the suggestions are addressed.

My major concerns are as follows:

(1) We get the full picture of the procedures involved only after reading the text. Since they are also presented in an extended manner, it is necessary to frequently go back to recall some of their components or figures. The inclusion of a flowchart describing the development of the emulator would be very helpful not only to give the whole picture to the readers since the beginning of the paper but also to show the location of each procedure in the whole sequence.

(2) I understand that the emulator can be used to extrapolate the model without extensive analyses, but I think it is necessary to discuss whether the emulator can be applied to other years instead of training data years. If an emulator is available, it would be interesting to discuss whether it is possible to replace the predictive simulations.

(3) It is expected that calibration and tuning of the model will bring it closer to the observed data. However, I believe that it would be valuable for the paper to add a discussion on the advantages of the emulator over data assimilation methods.

(4) I believe that the reader can better understand the usefulness of the emulator if Chapter 5 summarizes how this emulator can be applied.

My minor concerns are as follows:

Line 177–178: Please add the year to Running and Zhao's citation.

In Fig. 3, the lines overlap and are not visible, and hence, the figure is illegible unless you devise a way to show the ensemble spread in a semi-transparent way by representing the ensemble averages of initial, windows1, and windows2 as solid lines.

Line 311: I do not understand the intent in the following sentence: "Better observational data would improve the tuning; but the procedure itself does appear capable." Please explain this in more detail.