Thank you for you comments and suggestions. We believe the changes made have improved the article. What follows is a recap of the suggested changes, and how we have implemented them.

**Major Concerns:**

**- (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.**

We have added a flow chart broadly outlining the different steps required to fit an emulator in our framework.

**(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.**

- The goal of an emulator is indeed to replace new simulations. For JULES and other land surface models, our methodology would allow for predictions for other years, other than just the training years. Our framework deals with each grid cell and timestep separately, and so the emulator is valid for any potential future grid cell and timestep that is not outside the ranges of our training years. In other words, the emulator is valid for any future year, so long as that year has properties that are not far beyond the extremes of our training years. How a land surface emulator performs with years that are particularly extreme would be an interesting direction for future research, and would likely warrant training the emulator using entirely, or partially, artificial driving data that covers any credible future land environment.

We have added a paragraph to the discussion section outlining this potential.

**(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.**

Data assimilation is a related technique, but it does serve a slightly different goal. Data assimilation aims to improve one specific simulation, making it closer to observed data. Tuning also aims to improve simulations, but it does so obtaining better values for the static input parameters.   
One can try and use data assimilation techniques to do tuning as well, but it is not clear if this is an effective strategy(Rougier 2013). We have added some discussion of data assimilation to Section 4.

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

We’ve added a paragraph to section 4 explicitly making clear some of the applications of such an emulator. Emulators help with any goal where simulation runtime acts as a barrier to comprehensive analysis. Specifically, optimisation of land surface decisions is one interesting direction for future work.

**Minor Concerns**

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

done

**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.**

We have made the sampled ensemble members semi-transparent, which makes the individual lines more legible. We have also added text to the caption, making it explicitly clear that the red lines (if they were infinite) are a subset of the blue lines, which in turn (if they were infinite) are a subset of the grey lines). This information makes it more clear that the tuning procedure is shrinking the range of simulated GPP, and not simply moving it.

We have avoided adding ensemble averages or ensemble spread, as including these as well as the individual simulated ensemble members is not possible in a clear way, and we believe these summary statistics present an overly terse summary of the overall process to be presented alone.

**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.**

We have added to the text to clarify and expand.  
*(“Better observational data, with lower observational uncertainty and requiring fewer deletions / with less missing data, would likely further shrink the parameter space, improving the tuning results; but the overall tuning methodology itself appears capable.”)*

# Bibliography

Rougier, Jonathan. 2013. "‘Intractable and unsolved’: some thoughts on statistical data assimilation with uncertain static parameters." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.*