Response to reviewer comments for "Towards a process-oriented understanding of the impact of stochastic perturbations on the model climate"

Initial submission on Aug 24, 2023, EGUSPHERE-2023-1938

Deinhard, M. and Grams, C.M.

Re-submission of revised manuscript on May 12, 2024

Dear Mr Pedram Hassanzadeh, dear reviewers,

thank you for the constructive feedback on our paper. Attached you will find our responses to the reviewer's comments in blue.

Reviewer 2

My thanks to the authors for their efforts to address the reviewers' concerns – for both your responses and your edits to the manuscript.

Thank you again for your comments that helped to improve the manuscript!

In particular, thank you for addressing the question of significance of some of the results. I have a few remaining comments on this aspect – one request for the testing CI (Figures 3 & 9) and a query about a potential typo. Other comments are kindly-intended observations:

- Fig 3/9: the thick lines for each experiment line have been used to demonstrate significant difference from 1 (ie different from the IC only experiment). Visually, this is not very easy to distinguish (thick versus thin – are there any thin lines in Fig 9?) and the confidence interval for testing should be stated. As you recognise, this test does not demonstrate how different the MU experiments (Fig 3) are to each other. And I take your argument about interpreting the significance of small differences from 1 being significant as an indication of the likely significance of differences between experiments. Thank you for not over-elaborating on the differences in the text. As a reader, I would remain cautious of some of the smaller differences from what I can see.
  Thank you for this comment. It is true that the thick and the thin lines in Figure 3 and 9 are hard to distinguish. Nevertheless, we kept the Figures as they are, because we think that also other ways of showing the significance (such as marked lines) reduce the readability of the Figure. Regarding the absence of thin lines in Figure 9: Apparently, all values shown in the Figure are significant at the indicated level of confidence. The latter is now specified explicitly in the caption of Figures 3 and 9.
- Figure 9 – you state '1000-sample bootstrapping'. Is this correct? The other tests all claim 10,000-sampling.

Thank you for spotting this, but this is actually not a typo. We chose a lower number of repetitions for the bootstrapping on purpose, as the dataset underlying Figure 9 is by far larger than the other ones. In order to perform the bootstrapping in reasonable amount of time, we therefore decided to go with 1000 iterations only.

- Fig 6/7/8: thank you for adding the confidence intervals. It is striking that here you choose the 10-90% confidence range (as opposed to 95% range for Figure 2). This perhaps reflects that the experiments (or differences between CF and PF in Figure 8) overlap when using 95%. Again, I don't request further changes or response – but this relatively low confidence range (80%, after all) does suggest that one should not read too much into some of the differences: they rather "indicate" something interesting.

  Thank you for this comment, which we agree with. We believe that it would be possible to obtain a more robust signal at higher confidence level with a larger data base, but this was not feasible in this project.

Thanks once again for a very interesting contribution to the literature.