

RESPONSE TO REVIEW #1 OF GMD MANUSCRIPT:

“Introducing the Probabilistic Earth-System Model: Examining The Impact of Stochasticity in EC-Earth v3.2”

We thank the reviewer (hereby referred to as Reviewer #1) for their comments, which have helped us improve the clarity and quality of our manuscript.

MAJOR POINTS

- 1) **Reviewer #1:** *The order of the figures needs to be revised. (E.g. line 17, page 7, Figure 3 is mentioned before Figure 2).*

Our response: Yes, this was a clear oversight on our part. The revised manuscript now has figures numbered in the order they appear.

- 2) **Reviewer #1:** *“Some choices in the implementation of the stochastic perturbation needs to be motivated. For example what is the motivation behind the 3 temporal and spatial scales associated to STTP and ISTTP (Line 8, page 4). Also why the amplitude of the multiplicative perturbation factor is tapered in the boundary layer (given that the PBL scheme is a source of the kind of model errors that the stochastic perturbations are trying to represent?). Another choice that is not motivated is the use of parameter perturbation instead of stochastic tendency perturbation in the LAND experiment (Section 2.3).”*

Our response: We initially omitted some of these details for the sake of brevity, as most of the answers are covered in the cited papers. In response to this comment, we have now addressed these questions in the revised manuscript. The answers are as follows.

The three temporal and spatial scales associated to the SPPT/ISPPT perturbations come from the assumption that model error can come both from processes that happen at very fast timescales (and, therefore, small spatial scales), and also, independently, for processes happening on very slow timescales (and, therefore, large spatial scales). The three patterns are a compromise aimed at capturing a plausible range of relevant atmospheric processes.

The tapering in the boundary layer was implemented due to model stability issues. Numerical instabilities begin to accumulate when perturbations are added in the lowest model levels due to the delicate balance between model dynamics and vertical transport in this region.

Finally, for the stochastic land scheme, in fact tendency perturbations had been tested initially, but were found to have little to no impact on soil moisture in the model.

Parameter perturbations were therefore done instead, as these produce changes in soil moisture that more plausibly capture the model uncertainties.

- 3) Reviewer #1:** *“What is the motivation behind using the same perturbation for convection and large scale condensation in the ISPPT approach? The tendencies produced by these parametrizations are sometimes anti-correlated since when convection fails to remove instability from an atmospheric column large scale condensation tries to do that.”*

Our response: In the IFS, which forms the atmospheric component of EC-Earth, moist processes are split up into convection and large-scale condensation. The convection scheme transports moisture aloft, where it is detrained. This forms an input to the large-scale condensation scheme, which determines whether supersaturation has been reached and therefore how much cloud should be present. If needed, it then removes water vapour from the atmosphere with a corresponding increase in liquid water content. The intimate relationship between convection and large-scale condensation, whereby the outputs from one form the inputs to the other, motivated us to perturb these schemes together. Tests in an NWP setting indicate that grouping moist processes in this way gave skilful weather forecasts (Christensen et al, 2017, QJRMetS).

- 4) Reviewer #1:** *“It would be good to provide more discussion about the pathways in which the stochastic perturbations can change the mean. I agree in that the impact of SPPT and ISPPT suggests that the convective scheme is activated more frequently, however the discussion on how the stochastic perturbations can lead to this is not clear (e.g. line 14, page 10). In the discussion section it is stated that some perturbations can trigger convection in areas in which the unperturbed state has conditions close to those required to activate the convective scheme. However the opposite is also possible, some columns in which the unperturbed state is sufficient for the initiation of deep moist convection can be perturbed leading to a state in which these conditions are not met any more.”*

Our response: We would like to clarify that we do not assert that SPPT/ISPPT necessarily would be expected to trigger convection more or less frequently. Rather, our main assertion is that the inherent asymmetry in condensation triggers means that even a totally symmetric perturbation could lead to an increase in the mean cloud water content. This is because, while a perturbation away from e.g. deep moist convection would not be expected to change the cloud water content in the column (for that timestep), a perturbation towards it may trigger water to condense, which increased cloud water content. In other words, the impact of symmetric perturbations would be expected, if anything, to increase cloud water.

This increase in cloud water is suggested as a leading order mechanism (or pathway) towards larger scale changes with SPPT/ISPPT. This hypothesis is explained in the original manuscript on page 14, lines 9-16 (see the whole paragraph for a full discussion).

The comment the reviewer refers to (line 14, page 10) is specifically about precipitation *extremes*, which had been observed to increase with SPPT in previous studies. This was suggested as a potentially important pathway towards the observed changes in soil moisture.

The revised manuscript now contains some clarifying remarks to this effect in both sections.

5) Reviewer #1: *“As part of a first evaluation of the impact of SPPT, ISPPT and LAND upon the EC-Earth model it would be good to present some scores related to atmospheric circulation. Like for example MSE and biases for wind at different atmospheric levels and also for temperature at these same levels. The goal of the paper is focused on surface fluxes, but atmospheric circulation is also examined by studying for example the impact upon the Hadley cell. Although the impact upon the Hadley cell is relevant (particularly because SPPT and ISPPT seems to produce a large impact upon tropical convection), it would be good to provide these other scores for comparison with other systems.”*

Our response: We thank the reviewer for the suggestion, which we agree would be a helpful way to allow for easy comparison with other studies. We have extended Table 2 to add the MSE of temperature and zonal wind fields at various levels. This also helps illuminate the reviewer’s next question, and our answer to it.

6) Reviewer #1: *“It is not clear for me what is the motivation to study the QBO in the context of this paper. I understand that the impact upon different aspects of the atmospheric dynamics should be investigated but the inclusion of this particular aspect in a first evaluation has to be better motivated.”*

Our response: The reviewer makes a good point: this was not adequately motivated. Our primary motivation here comes from the paper Leutbecher et al. (2017), *Stochastic representations of model uncertainties at ECMWF: State of the art and future vision*, where the impact of SPPT on short to medium range forecasts is considered. On page 11 of this paper, it is documented that the biggest degradation of SPPT on the version of the IFS considered is on the upper level winds, where the QBO dominates. The MSE computations we added in Table 2 (c.f. our response to the previous comment) show similar behaviour. This is raised in *ibid* as being a point of concern due to the growing body of literature suggesting that the QBO is an important driver of European climate at seasonal timescales. We therefore wished to identify if a similar degradation occurs in the EC-Earth model for the schemes considered.

We have now added this motivation to the introduction of section 7.

MINOR POINTS

- 1) **Reviewer #1:** *“Line 10, page 5. What does exactly mean that parameters are correlated? Estimated parameters based on observation studies show that the value of these parameters in different soil types and conditions are correlated or that the joint sensitivity of these two parameters shows a certain degree of compensation between the impact of these two parameters (i.e. the effect of the increase in one of the parameters can be compensated by changes in the other parameter).”*

Our response: We meant the former. Estimations of parameters based on observation studies indicate that the two parameters show correlation across soil types. This correlation is neither zero nor one, indicating that the parameters are related whilst showing some independence. We therefore chose not to perturb the two parameters with the same pattern, nor with two entirely independent patterns. Rather, we introduce some dependence between the perturbations through definition of a third pattern, as described in the text.

The rationale for this is that the observed correlation suggests that some parts of parameter space may be unphysical (for instance extremely high values of gamma and low values of alpha). An independent perturbation would be likely to access these regions of parameter space (for instance simultaneous values of $1+r=1.8$ for gamma and $1+r=0.1$ for alpha). By tethering both parameters to a third base pattern, this possibility is reduced and the parameters are perturbed in a more similar way, whilst retaining some independence in the perturbation.

We have slightly rephrased the description in this section to make it clear that the correlation is based on observational studies.

- 2) **Reviewer #1:** *“Why performing 5 periods of 20 years each instead of a longer simulation. Using 5 different periods as ensemble members can artificially increase the ensemble spread and reduce the significance of the results. Also spin-up issues may be more important when several shorter periods are considered, particularly in the soil variables.”*

Our response: When constructing the experimental protocol, the authors recognized that there were several aspects it would be good to test, but that there were insufficient computer units to run all the configurations desired. We decided that the most important goal was to produce runs where the significance of any impacts could best be detected: for this reason, it was decided to produce an ensemble of simulations for each scheme as opposed to a single long run. The authors believe that comparison of distinct ensemble members allows for the most transparent and accurate assessment of the uncertainty in the computed metrics. While such an assessment from a single longer run would be possible by using techniques such as subsetting the data and/or bootstrap resampling, these techniques are essentially just trying to artificially create distinct ensemble members within the longer timeseries anyway. Therefore, simply producing an actual ensemble from definitely distinct initial conditions gives the cleanest methodology, in our opinion. We are not aware of any

reason why this may increase the spread in a way which is excessive compared to that diagnosed with other experimental protocols.

The initialisation of each ensemble member to a slightly different time period also allows us to cleanly assess any dependence of the rapid response on the initial ocean and land state. Spin up issues may have been expected to be problematic for the land state, but in practice we found that the main changes were the same across all the time periods, implying that the impact of the schemes are quite rapid and do not require a longer period to assess accurately.

We have included some more discussion on why our experimental protocol was chosen, based on the above discussion, in the revised version of the paper and hope that this will satisfy the reviewer.

3) Reviewer #1: *“Line 29, page 4. Remove parenthesis and “for details”.”*

Our response: We made the suggested change.

4) Reviewer #1: *“Line 29, page 6. Missing space before Spatial. Also indicate instead of indiate.”*

Our response: We made the correction.

5) Reviewer #1: *“Since convective precipitation is part of the products generated by the convective scheme and is linked to the other tendencies, is precipitation rate perturbed in the same way as the other tendencies produced by the parametrization? Same question but for the large scale condensation scheme.”*

Our response: Thank you for your insightful question. You are right, for complete consistency the fluxes in precipitation and evaporation should be perturbed using the same pattern as for SPPT. However, this is not currently implemented in SPPT. This could be part of the reason why the scheme does not conserve water, which is why the ‘humidity fix’ has been implemented to correct this. Having said that, testing is underway to develop a more consistent approach whereby precipitation and evaporation are perturbed, to bypass the need for the ‘humidity fix’.

6) Reviewer #1: *“Line 6 page 14. More frequent convective scheme activation can also explain why the PBL is drier.”*

Our response: We thank the reviewer for the insightful comment: we have added a comment on this in this section of the revised manuscript.

- 7) **Reviewer #1:** *“Figure 4 a, shows the biases in the precipitation for the control run. This bias pattern is strong and shows a clear maximum in the tropics. The authors indicate that the control configuration has been extensively tuned, however has the tuning been performed with this same model resolution?”*

Our response: The model was indeed tuned at the T255 spectral resolution which we used in this study. The tuning procedure for EC-Earth is carried out in order to obtain a realistic energy budget, with a particular focus on the net surface energy flux. In particular, precipitation biases are not directly tuned. Therefore, while the deterministic model has little bias in the energy fluxes at the surface, it does still have relatively notable biases in key variables like precipitation.

- 8) **Reviewer #1:** *“Line 16, page 15. Changes in the Hadley cell are caused by changes in evaporation? Or these two changes are driven by changes in tropical convection?”*

Our response: The reviewer makes a good point; the manuscript as it stood focused on changes to evaporation (latent heat flux) because of the focus on energy budget changes. However, it is of course possible that the first order impact is on tropical convection, and evaporation changes are a response to this. We have included some discussion of this in section 8.1 (Discussion) and 8.2 (Conclusions).

- 9) **Reviewer #1:** *“Figure 6: The changes in T2m over the sea ice in the ISPPT and LAND are very strong. It is surprising to see these changes in both experiments since none of these experiments seems to directly affect the sea-ice parametrization in any way (SPPT for example do not show a strong change in bias in this region). I suggest to check the sea-ice distribution and temperature in these experiments.”*

Our response: Being AMIP style experiments, the sea-ice distribution is a fixed field along with the sea-surface temperatures. Therefore, for both ISPPT and LAND, the cooling seen in the sea-ice regions are necessarily induced via atmospheric circulation changes, which can cause surface temperature (which is dynamic) to change. Discussion of potential mechanisms behind atmospheric circulation changes are included in the manuscript, and have been extended upon based on other comments by both reviewers. We therefore do not expand upon this further.

10) Reviewer #1: *“Figure 11. I suggest to use the same names as in the rest of the manuscript.”*

Our response: We made the suggested change.

11) Reviewer #1: *“Figure 8. Please correct the caption since the colors do not correspond to the ones on the legend (I assumed that the legend is correct).”*

Our response: We thank the reviewer for pointing out this silly mistake on our part. The caption has been edited to match the legend in the figure.

12) Reviewer #1: *“Line 10, page 8. This sentence is not clear, I can not see “each model simulation” but something that seems to be the mean of all simulations”*

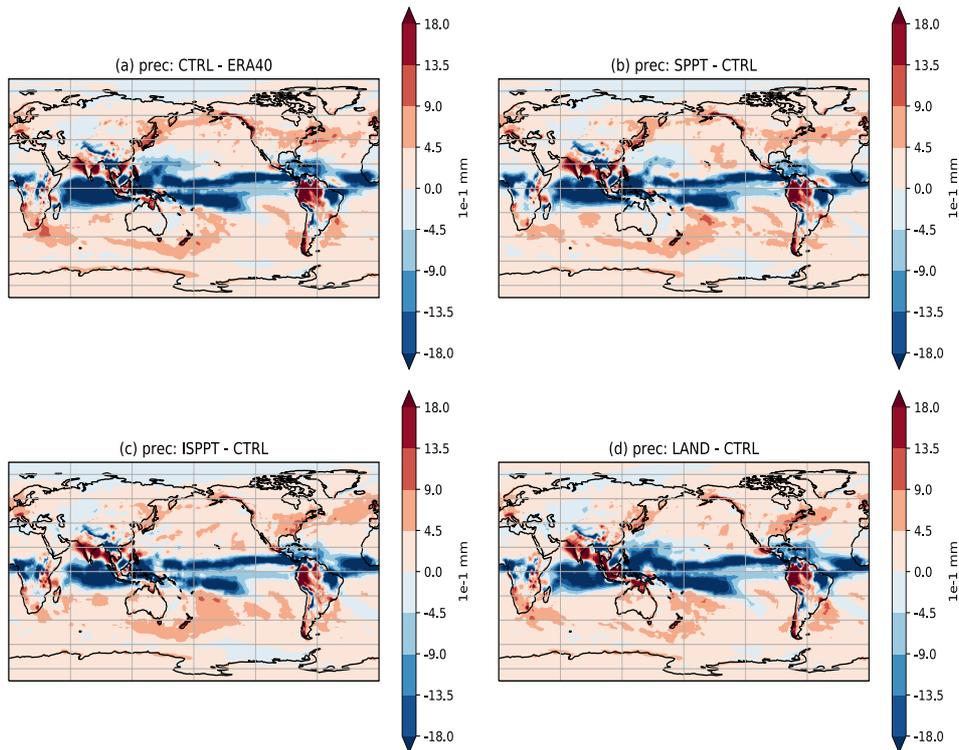
Our response: Our phrasing was somewhat unclear. We meant ‘the mean across all the individual differences’. We have rephrased this as follows: ‘Figure 8(a) shows the mean difference between the five SPPT-simulations and the corresponding CTRL-simulations (i.e. the average across the 5 differences), ...’

We hope this rephrasing makes the meaning clearer.

13) Reviewer #1: *“It would be better to use the same color scale for all panels in figures 3, 4, 5, 6 and 7. In most cases the range is similar. Another possibility is to show in all cases the bias with respect to ERA (again since the magnitudes are similar this should clearly show the improvement produced by the stochastic schemes and would be more easy to analyze). Also in this figures indicate what “M=” stands for. I assumed that this is the mean bias over the global domain.”*

Our response: We have added the explanation for “M=” in the captions now, a clear oversight: the reviewer was indeed correct in their assumption of its meaning.

We would like to respectfully disagree with the reviewer on the suggestion to change the spatial maps to show the bias with respect to ERA in each panel, as we believe this would make it harder, not easier, to analyse. As an example, consider the precipitation changes in figure 4. Note that the CTRL bias is nearly 3 times as large in magnitude as the impact of the stochastic schemes (with the former having a peak of ~1.8mm and the latter a peak of ~0.6mm). As a result, showing ERA biases side by side is not very illuminating, as the following figure demonstrates:



Showing the bias relative to CTRL immediately illuminates how the scheme is changing the mean state for all the variables. Colourscales are tailored specifically to each individual variable, and chosen so that the scale encapsulates 2 standard deviations around the mean bias. For a more easily comparable quantitative metric, we have used the mean bias (M=...) and the MSE table. We therefore suggest to leave the figures as they are.

14) Reviewer #1: *“Since the main goal is to perform analysis towards the development of a coupled stochastic modeling system, why a SPPT+LAND or ISPPT+LAND experiments where not performed?”*

Our response: In fact, such experiments were carried out. These were not ultimately included in the analysis proper, because the impact of SPPT/ISPPT is typically much larger in magnitude than the impact of LAND, which makes it difficult to assess the individual impact of either in a joint experiment. This piece of information should clearly have been included in the paper, and we thank the reviewer for making this point. We have now added a paragraph on this in the Experimental Setup section (section 3.1).