Dear Referee:

Thank you very much for the review, which is very constructive and helps a lot to improve our manuscript. We have answered your comments, questions etc. point by point in below.

Best regards,

Clemens Wastl, and co-authors, 26.11.2018

Responses to RC1:

General Comments:

8. Does the title clearly reflect the contents of the paper?
In its present form the title does not really reflect that stochastic perturbations are independently applied to parametrization schemes and prognostic variables.

The title has been adapted accordingly. "Independent perturbations for physics parametrisation tendencies in a convection permitting ensemble (pSPPT)".

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
Figures 3-7 must be substantially improved. There are no ordinate labels and no references to the different panels in the figure caption.

Figures 3-7 (including the captions) have been adapted accordingly.

Specific Comments:

— Abstract: "Both schemes … lead to statistically significant improvements in the probabilistic performance compared to the original SPPT". The authors have shown statistical significance of the different schemes with respect to the reference experiment of no perturbations not with respect to SPPT (see also Page 7, lines 11-12: "The statistical significance of the score differences between the three experiments and the reference run is defined by using a bootstrapping confidence test.")

This sentence in the abstract has been modified: "… compared to a reference run without stochastic physics."

— Page 4, line 22: "The tapering function : : : - it is not necessary in some regional models (e.g. WRF, COSMO)." Are there any references? Do you know why?

We do not exactly know the reason for this. The implementations of SPPT in the WRF and COSMO model use especially adapted settings for the perturbations, as stated by Leutbecher et al. (2017): "The differences include the variance of the perturbations, the space and time auto-correlation of the random pattern, the shape of the distribution that is sampled…" We have tried several settings for C-LAEF as well, but without using a tapering function the SPPT implementation was not stable.

The reference "Leutbecher et al., 2017" has been added to this statement.

— Page 6, lines 8-10: “A potential drawback of the pSPPT approach is a possible duplication in attributing errors across schemes which can introduce inherent correlations between the perturbations applied to one physics scheme and the output of a later scheme (Christensen et al., 2017).” In contrast to SPPT the pSPPT approach enables switching of the tapering function. Christensen et al. 2017 state about the effects of tapering in SPPT “: : : this method cannot represent uncertainty in the vertical distribution of convective heating. SPPT does not perturb fluxes at the surface or top of atmosphere, introducing inconsistencies between the perturbed tendencies in a column and these fluxes.” When balancing the disadvantages from tapering inconsistencies or a possible duplication in attributing errors across schemes, it seems not unreasonable to try out the pSPPT approach. Moreover, it could be expected that errors naturally appear all along the production line of the parametrization schemes and will be processed through the different schemes anyway, presumably being inherently correlated at the output. The main fundamental reason for keeping the parametrization chain deterministic and balanced is that in deterministic parametrisations similar tendencies on input produce similar outputs and thus large scale correlations of the input tendencies on the grid scale will not be altered by the parametrization chain. Therefore, in SPPT only the final output of the chain is perturbed with large scale stochastic patterns which are tied to resolved physical processes on the grid scale. Christensen et al., 2017 state: “SPPT also imposes large spatio-temporal correlation scales when perturbing tendencies to represent the correlation of model uncertainties in space and time, but these correlation scales have not been measured and are not tied to physical processes.” Because the correlation length scales of the stochastic patterns are the same one can still expect that the correlations of the input tendencies survive not only the SPPT but also the intermediately perturbed parametrization chain of pSPPT.

Thank you very much for this statement. Yes you are right, we are only using one scale pattern for all the parametrisations and therefore the correlation between input and output tendencies of the complete parametrization chain is kept.

— Page 12, lines 13-15: “Perturbing the physical schemes separately and considering this perturbed fields in the subsequent parametrisation (pSPPT) results in a positive effect on the stability of the model. In this case the tapering function could be switched off for microphysics, radiation, and shallow convection without any problems.” Page 12, lines 23-24: “Hence, an interaction of of the uncertainty of one physical scheme in the subsequent one is considered in pSPPT and ipSPPT which seems to increase the consistency of the model,” What does consistency of the model means? Does it mean that it runs more stable? It is a little bit surprising that the tapering function could be switched of in the ipSPPT as well, because there is much more uncorrelated noise than in the two other schemes? Where does the stability comes from?

Yes you are right, we are talking about the stability of the model (not consistency). In case of pSPPT and ipSPPT not only the tendencies are passed from one physical parametrisation to the subsequent one, but also the uncertainties (perturbed tendencies). In SPPT the uncertainty is not considered in the physical parametrisations, it is applied at the end of the time step. Hence, in case of pSPPT and ipSPPT the model can react in each parametrisation on the uncertainties coming from the previous one, which is more consistent in our mind than just adding perturbations at the end of the time step.
As written at the end of section 2.2.1 we have used a two weeks test period in 2011 to try different settings of the stochastic schemes and to test the stability of the model when switching off the tapering function. In case of SPPT and no tapering function used we had about 10% model crashes, with ipSPPT about 2% and no crash with pSPPT. This means that ipSPPT is less stable than pSPPT (which is not surprising), but more stable than SPPT, which is of course a bit surprising. Unfortunately we did not look into the details about where the stability in ipSPPT comes from. In our later studies and investigations we focused

completely on the pSPPT approach and did no more consider ipSPPT because as stated in section 2.2.3, some physical relationships within a parametrisation scheme could be violated in this case. But as you say in your last comment, looking only on the scores of this study, it could be worth to enhance this approach.

We have changed this phrase to: "… seems to increase the stability of the model"

— Pages 12-13, lines 32-6: "The ipSPPT approach is a modification of pSPPT where the tendencies of the variables T, U, V, and Q receive separate perturbations. As shown in Sect. 3, this approach obtains the best probabilistic scores overall, even though the method is considered critical from a physical point of view. A major concern with the ipSPPT approach is that the balance between the quantities resulting from one parametrisation scheme can be disturbed (Palmer et al., 2009). For example, the microphysics scheme provides an increase of temperature at a certain point due to condensation processes which are also decreasing the water vapor content. This equilibrium is destroyed if temperature and water vapor content tendencies are perturbed with opposite signs. On the other hand, it cannot be assumed that T and Q have exactly the same error characteristics, as it is supposed in SPPT and pSPPT. Furthermore, in SPPT and pSPPT the wind direction is never altered stochastically, since the tendencies of the U and V components are always using the same stochastic pattern." Page 6, lines 23-25: "The first SPPT version in the IFS model (Buizza et al., 1999) has also used such separate patterns for the different parametrised tendencies. However, it has been removed in the revised SPPT scheme (Palmer et al., 2009) because some physical relationships within a parametrisation scheme could be violated in this way (see Sec. 5)." The ipSPPT is the winner while pSPPT performs very similar to SPPT, except for the surface in January 2017. This latter improvement might be effectively attributed to switching of the tapering function in pSPPT. I would follow the authors that the possibility of altering the wind direction in ipSPPT is a good candidate for explaining the superiority of the ipSPPT in generating reasonable spread, especially in complex terrain like the Alps in winter and for convection in summer. While U and V should be perturbed differently, the T/Q imbalance could be avoided, if these variables are perturbed with the same or correlated stochastic patterns, as it has been described by Christensen et al., 2017 for the handling of processes in their iSPPT scheme: "It is likely that the 'true' errors in the parametrization schemes are neither perfectly correlated as in SPPT nor perfectly uncorrelated as in iSPPT. A further interesting line of enquiry would be to introduce correlations between the noise patterns used for different parameters. Instead of using two independent patterns in iSPPT, perturbation patterns for the wet processes could be partially correlated with each other, while perturbations for the dry processes could also be partially correlated." Have you tried to modify the SPPT by simply perturbing U, V with different stochastic patterns, or do you know, if anyone else has done this?

Thank you very much for this interesting statement and the ideas you raised. As mentioned in the paper, from a physical point of view it is not reasonable to perturb T and Q separately because it disturbs the relationship between these two quantities. However, applying different perturbations to the U and V component of wind is very interesting. A colleague from Hungary (Mihaly Szucs) tried this approach for some test cases with the AROME model and the results were quite promising (increase of spread near the surface and no obvious model degradation. The results of this study can be found on
http://www.rclace.eu/File/Predictability/2015/LACE_stay_MihalySzucs2015.pdf
It would be very interesting to compare this approach with the ipSPPT experiment presented in this paper. However, as mentioned in the previous comment, the focus of our later studies was put on the pSPPT approach.