

Interactive comment on “Evaluation of the Plant–Craig stochastic convection scheme in an ensemble forecasting system” by R. J. Keane et al.

Anonymous Referee #2

Received and published: 22 December 2015

General comments:

The authors present results from a series of experiments comparing the performance of the Plant-Craig (PC) stochastic convection scheme to operational forecasts made using the MOGREPS ensemble forecasting system. Verification is performed using a number of techniques, considering both ‘deterministic’ and probabilistic verification techniques.

It is interesting to see the results of using the PC scheme in a near-operational ensemble prediction system. Separating the verification dates into strongly and weakly forced cases provided valuable information regarding strengths and weaknesses of the scheme – in particular, the PC scheme performs well compared to the reference forecast for the weakly forced cases. The discussion clearly discusses the merits and

C3451

shortcomings of the PC scheme as observed in the model results.

However, the paper could be improved to clarify some of the experimental details. Some of the figure labels are not very informative and would benefit from more information (see specific comments below). I also think the paper would benefit from additional analysis of the results. In particular, it would be interesting to consider the FSS for a wider range of averaging areas, which could reveal trends in the behaviour of the PC scheme. It would also be helpful to consider an additional, more-familiar verification technique, as the EAV is a very new approach and the FSS is not particularly widely used. This may make it easier for the reader to immediately understand the benefits of the new scheme. Finally, it could be interesting to consider forecast thresholds based on precipitation percentiles and not thresholds, which could be more indicative as to the performance of the two schemes at high rain rates. For these reasons I am recommending the paper for publication with major corrections.

Please see the ‘specific comments’ below for more details regarding these suggestions.

Specific comments:

P10200 L4 - ‘with a simple stochastic element only’ - somewhat confusing phrasing. Perhaps replace by ‘with a simple stochastic scheme only’

P10201 L10 perhaps also mention perturbed parameter approaches here since you mention the MO RP scheme later. They are also more commonly used than multi-parameterization approaches. e.g. Bowler et al, 2008; Christensen et al, 2015

P10201 L21 e.g: Teixeira and Reynolds, 2008

P10202 L15 does this mean that the PC scheme will have (next-to) no impact in seasonal and climate models, which have even coarser grid boxes?

P10204 L12 You reduce the mean mass flux per cloud and increase the cloud radius. You mention that Keane and Plant 2012 chose numbers to match those derived from

C3452

CRM studies of tropical oceanic convection. Is there a similar study that you used to estimate these parameters over the UK? Alternatively, could you motivate why the plume radius should be larger and have smaller mass flux per cloud from theory or from other independent studies?

P10206 L18 state how long the forecasts are

P10207 L7 You state the FSS is used for a 'deterministic' forecast such as a single member or the ensemble mean. You then state on P10211L5 that you evaluate the 'deterministic' forecasts, but you do not state whether you are referring to a single ensemble member or the ensemble mean. Please qualify in both paragraphs which of these you use in the paper.

If you are using the ensemble mean and not a single member, then that would explain the lack of skill for high thresholds - taking the ensemble average will smear out the precipitation fields.

P10207 L13 I find the FSS unintuitive, and would find a little more explanation helpful. For example, why is it normalised with respect to $\langle F^2 \rangle + \langle O^2 \rangle$? This means the expected value of the score can be written:

$$2 \langle FO \rangle / (\langle F^2 \rangle + \langle O^2 \rangle)$$

While this is indeed 1 if $F = O$ for every case considered and zero if F and O are uncorrelated, it has a form which is not otherwise motivated in the paper.

P10211 L5 clarify deterministic forecast

P10211 L8 at long lead times the GR scheme outperforms the PC scheme for grid point fields - the improvement in skill from the PC scheme drops off quickly with time. 54 hour forecasts are very short - did you consider any longer forecasts to see if this is a significant trend or if this is just noise? How long are operational MOGREPS forecasts run for? As you would want a new scheme to perform well over the whole range of the forecast.

C3453

P10211 L18 I agree that if fixed thresholds are used to compare observations and model data, the lack of skill at high thresholds is likely to be primarily due to the model bias in over-forecasting wet events, which gives a large mismatch in observation and forecast frequency at these thresholds. Have you considered comparing percentiles of the observed and forecast distribution instead of thresholds? This would remove the frequency bias, and would test the spatial and temporal distribution of rain instead.

P10212 L10 as you say in the text, the improvement is very scattered for the area averaging FSS, preventing a clear conclusion from being made. It is hard to tell if such noisy results are statistically significant. Can you either repeat the experiment for more start dates to improve the significance, or (if that is not possible, as I think you indicated in the text) could you repeat the analysis for a wider range of neighbourhood areas to see if general trends can be identified?

P10212 L1 clarify - is this statement (which is continued from the previous page) referring to both weak and strongly forced cases?

P10212 L11 if the aim is to average over a large number of calls to the scheme, why don't you evaluate this over the ensemble forecast instead of in space for a deterministic member? Please comment on this here.

P10212 L20 Remind the reader that the BSS is calculated with respect to the climatological forecast

P10213 L1 as for the FSS, computing the BS with respect to forecast percentiles may provide additional information about the performance of the schemes at high thresholds i.e. asking the question "which scheme tends to put its strongest precipitation in the right place at the right time?". When interpreted together with figure 10 this could provide more information.

P10213 L14 I am not familiar with the EAV score. Using a more well known verification technique (instead of, or as well as, the EAV) could be helpful for the reader here, such

C3454

as the decomposition of the Brier score, or considering RMS spread-error scatter plots (Leutbecher and Palmer, 2008), or reliability diagrams. This would also make it easier to compare the results from this paper with those in other papers, and in particular, with the new stochastic physics scheme recently adopted by the Met Office (Sanchez et al, 2015.)

P10214 L6-12 Most authors argue that the fundamental aim of a stochastic parameterization is to improve the reliability of the forecast by giving a flow-dependent indication of uncertainty in the forecast, which cannot be achieved using statistical calibration.

P10214 L26 Showing the spread alone as a function of time does not indicate the calibration of the forecasts. It would be interesting to see the ensemble spread vs error in the ensemble mean over the regions where you have verification.

P10215 L25-28 can you understand why the PC scheme degrades temperature and pressure?

Figure comments

For all figures, the axis labels and tick marks are quite small, so may need to be increased, depending on the final size of the figures in the paper. Figures 2, 3, 4 and 5 are particularly small.

For all figure captions, qualify statement "difference between the two schemes" by indicating whether this is PC-GR or the reverse.

For all captions qualify what kind of forecast are being used - e.g. Deterministic or ensemble, and what kind of deterministic scheme it is.

Figure 4 please add a zero line for clarity

Figure 6 remind the reader that the BSS is calculated with respect to the climatological forecast.

Refs:

C3455

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, 134(632):703–722.

Christensen, H. M., Moroz, I. M. and Palmer, T. N., 2015, Stochastic and perturbed parameter representations of model uncertainty in convection parametrisation. *J. Atmos. Sci.*, 72

Leutbecher, M. and T. N. Palmer, 2008. Ensemble forecasting. *J. Comput. Phys.*, 227(7):3515–3539

Sanchez, C., Williams, K. D. and Collins, M., 2015, Improved stochastic physics schemes for global weather and climate models, *Q. J. Roy. Meteor. Soc.*

Teixeira, J. and C. A. Reynolds, 2008, Stochastic Nature of Physical Parameterizations in Ensemble Prediction: A Stochastic Convection Approach *Mon Wea Rev*, 136(2)

Interactive comment on *Geosci. Model Dev. Discuss.*, 8, 10199, 2015.

C3456