Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-260-AC2, 2018

Geoscientific Model Development Discussions

Interactive comment on "Methods of investigating forecast error sensitivity to ensemble size in a limited-area convection-permitting ensemble" *by* Ross Noel Bannister et al.

Ross Noel Bannister et al.

r.n.bannister@reading.ac.uk

Received and published: 5 February 2018

Responses to reviewers for, "Methods of investigating forecast error sensitivity to ensemble size in a limited-area convection-permitting ensemble" (gmd-2017-260)

February 2018

This document is our general (undetailed) response to the main criticisms of reviewer 2. A second (more detailed) part will follow should the editor allow the paper to be revised with a reasonable chance of acceptance in GMD.

Response to reviewer 2

We would like to thank reviewer 2 for his/her comments and criticism of our manuscript sent to GMD.

The reviewer's report contains many positive comments about the paper, and in fact does show that there are many original and useful aspects to the work presented, despite the apparently negative overall score given.

Some of the points of reviewer 2 overlap with those of reviewer 1, namely the single case study (para. 2 of reviewer 2's report), and the apparent lack of originality (para.

3). We defend these issues in the response to reviewer 1 (reproduced below). We respond to other of the main issues raised by reviewer 2 as follows.

1. Content and focus. We accept that the manuscript could be shortened (see para. 4 of reviewer's report) and that the focus could be improved (paras. 3 and 5). The focus could be improved by covering how each issue is important to ensemble forecasting or DA. That said though, the two disciplines (ensemble forecasting and DA) are inherently linked, and could/should become aspects of the same problem in time, and so it is arguable that papers that are relevant to both should be welcomed. The extra references could be added (para. 6).

2. Scientific interpretation. We could reduce the interpretation in the light of the single case limitation (para. 7), although we think that there is still scope for interpretation as the results are still valid, although not necessarily applicable to all weather situations.

3. More points on novelty. In addition to our defense of originality covered in the response to reviewer 1, reviewer 2 does actually acknowledge that many of the results are indeed new and interesting (paras. 8, 9, 12), although he/she has chosen to make emphasis on other parts of the paper when making his/her overall assessment (as stated in our response to reviewer 1, our paper is a mixture of new and standard diagnostics to help paint an overall picture).

   (a) With respect to the comments about Sect. 7 (para. 12 of reviewer's report) about the assumed overlap with Menetrier et al. (2014): please note that there are some profound differences between our work and theirs, namely the (robust) finding of exponentially-shaped correlation functions with our work.

   (b) With respect to the comments about the kinetic spectrum (para. 11): we acknowledge that the mere computation of a kinetic spectrum is not new, but

its study in an ensemble context, we believe, is new. The reviewer questions why doing an average over ensemble members is necessary. In the same way that performing an ensemble average of a field can and does affect the size of features (as we reminded readers in Sect. 4.1 in the context of rainfall), this can also be true when looking at the effective resolution of an ensemble vs an individual member. (We could make more of this in a revision.)

4. Simplicity. The reviewer commended the work of Sect. 3.3 (linear independence tests), but commented that the method is complicated (para. 9 of reviewer's report). The method used is actually extremely simple (much simpler than, e.g. rank histogram computations in our opinion). The editor or reviewer is invited to see the code, which is made available with the paper.

The reviewer's report contains some minor and detailed aspects which we can address in a way to improve the paper, should the manuscript be taken further. We also hope that the editor will see that the overall assessment of this manuscript is not justified (e.g. that the presentation is actually much better than the "fair" assessment given).

Relevant extract from report for reviewer 1

1. Single case study. One of the comments concerns the presence of a only single case study, and this limitation is indeed highlighted by the authors (para. 2 of reviewer's report). It does mean that the specific results do not necessarily represent firm conclusions, but it certainly does not mean that the results are not useful, e.g. they are probably representative of the particular weather regime studied (especially as the case comprised several days' data). Menetrier et al. (2014), which the reviewer cites (para. 3), is also based on a single case, demonstrating that this need not be a show-stopper. The 'single case' limitation (which

was outside of our control) is the reason why we have emphasised the methodology, rather than the specific results, and to seek publication in this particular journal. Indeed, parts of the methodology could be adopted by others studying other systems.

2. Originality. We believe that the reviewer's opinion on originality is not a fair judgment. The new aspects of the paper are not highlighted in the reviewer's report, even though they are present in the paper. A version of this manuscript was originally sent to another leading journal, and the only reason why it was not accepted there was because of the single case study limitation, which should not be a problem given the scope of GMD. Importantly in that submission, both reviewers' comments were otherwise very positive about the work (it was described as "state of the art", and a "significant contribution to the field"). Of course, that was the outcome of a separate editorial process, but it does serve to highlight that it is possible, as in the present case, to get a distorted view of a piece of work from a small number of reviewers. It is not clear to us why the present reviewer should hold their opinion, but it might be the case that he/she was expecting that every figure should represent a brand new diagnostic. It is usual for a study to use a range of new *and* standard diagnostics as part of an overall picture. Many of the figures in the paper are in fact, we believe, based on new developments in the field of ensemble forecasting and/or data assimilation, which the reviewer has not discussed in his/her report, and this may wrongly lead the editor to believe that the paper contains no or little original content. To emphasise our point, these are areas of the manuscript that we think especially have not been explored before in the context of ensemble forecasting/DA:

   (a) The way that a large ensemble can be generated from an existing smaller ensemble (Fig. 2, Sect. 3.2). This technique will almost certainly be of interest to other research groups who would like to extend their ensemble systems.

   (b) The linear independence tests to show that the members do develop linear independence (Fig. 3, Sect. 3.3). This is a simple but essential diagnostic to confirm the usefulness of the above method.

   (c) Study of the kinetic spectrum in an ensemble context (Figs. 9 and 10, Sect. 6). This suggests how errors in kinetic energy of a finite ensemble change as a function of scale – very important information to have when designing and interpreting ensemble data assimilation systems.

   (d) Study of the form of the correlation functions of variance errors, in particular finding an excellent fit to an exponential form (Figs. 13 and 14, Sect. 7.3), and how this could be used to generate variance fields that have a prescribed form of sampling error characteristic of a finite ensemble (Eq. 7). The exponential fit makes the length-scale analysis different from that covered previously, e.g. Menetrier et al. (2014), Pannekoucke et al. (2008), and Raynaud and Pannekoucke (2012/3), which looked at parabolic or Gaussian forms (7th minor comment or reviewer's report).

   (e) A potential new test of whether an ensemble is large enough to meaningfully neglect sampling error (Sect. 7.2, and Fig. 13).

   (f) Analysis of the errors in the sub-sampling for many diagnostics (e.g. the fit to the exponential, Fig. 14).

   (g) The application of the above and standard diagnostics to the high-resolution Met Office Unified Model.