# Author response to anonymous referee #2

Thank you to referee #2 for reading the manuscript and for his/her valuable comments. In the following, the referee's comments are reproduced, and my responses are in blue. Please note that I am instructed by the journal to give responses before preparing a revised manuscript, but I highlight here any changes that I plan to make in the revision.

Review of the GMD submitted article referenced as: Title: The "ABC-DA system" (v1.4): a variational data assimilation system for convective scale assimilation research with a study of the impact of a balance constraint Author(s): Ross Noel Bannister MS No.: gmd-2019-318 MS Type: Model description paper Iteration: Minor Revision

The paper presents a very comprehensive technical description of a toy data assimilation system based on the previously published ABC-model formulation and codes. Its content is extremely detailed and can easily serve as useful introduction to any scientist, including young scientists and post-graduate students, interested in uploading and using the codes for research or education. The technical content is complemented by a showcase example of the scientific use of the system (the study of the impact of the ABC-associated geostrophic balance constraint in the B-matrix model).

The article is clearly written and the figures are of good quality. The presentation as a whole matches the expected goal which is to provide a scientific introduction to the ABC-DA system. As a reviewer of this article, I do not consider that my role is to evaluate or critizise whether the ABC-model and DA formulation will be a useful scientific tool per se. The present paper will however enable the scientific data assimilation community evaluate that usefulness in practice. I therefore recommend the paper for publication after very minor revisions.

Hereafter follow my minor comments:

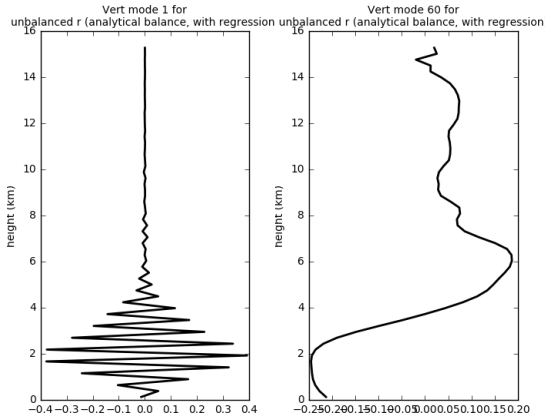
1. about §4.1: from what platform can a user upload UM data for initializing the very first steps? Or are such data available with the ABC system packages?
   Ensembles play an important role in many cases of the toy applications. Can you say a few words about how the size of the super-ensemble is set, and what would be a "reasonable" limit of size?

   (a) The size of the sample UM data is unfortunately too large for the GitHub repository. I have though added the sample UM data (and also some pre-prepared ABC ensemble initial conditions) to the web site of the Data Assimilation Research Centre at Reading and provided a link to there from the GitHub repository. Users downloading the ABC-DA system from GitHub now see a URL to the data.

   (b) The second point overlaps with a comment from another reviewer. A new appendix (appendix A) will be added which deals with the amount of information needed to determine aspects of the covariance model (CVT). Essentially there are about $10^5$ pieces of information of the covariance model that need to be determined during the calibration (things like vertical modes, and spectra), but $28 \times 10^6$ pieces of information is provided in the form of 260 super-ensemble members. This is more than adequate.

2. About Fig 6b: this particular plot is actually little discussed in the core text. My question is, noting that the vertical lengthscale increases with increasing vertical mode (i.e. the more nodes on the vertical, the deeper the penetration scale of the mode), is this behaviour due to the fact that the plot holds for the unbalanced part of the scaled density? i.e. one expects the opposite property for the balanced part of scaled density (low-order vertical modes of balanced scaled density would have the largest vertical lengthscales?). Is this correct? I suggest a short comment about Fig 6b could be added in §4.3 or §4.4.

   (a) I actually find that, for the vertical modes, the more nodes there are, the shorter the penetration scale. For info, a low mode of unbalanced scaled density is shown on the left below (short vertical scale) and a high mode is shown on the right (long vertical scale). They appear in this order because the eigensolver evidently outputs modes in ascending eigenvalue order. Such a correspondence between mode and vertical lengthscale is in fact found to be shared with vertical modes of streamfunction (which also share their vertical modes with the balanced component of scaled density). A comment will be added to the penultimate paragraph of Sect. 4.3.5 to state the above correspondence between the number of nodes and the vertical scale, and that the same (not the opposite) property exists for the balanced part of scaled density.




3. about §4.7: it is stated that the ABC-DA system is flexible enough to host a variety of DA methods, like 4D-VAR or Ensemble-Variational formulations. It seems indeed clear from the article that methods based on variational formulations, including iterative steps such as a minimisation and the computation of a gradient, are allowed. However, what about methods like Extended or Ensemble Kalman filters, or versions of Optimal Interpolation, i.e. methods where the Gain Matrix (G) would be somehow explicitly computed, and a direct inversion step involving G would be implied? Similarly, what about methods involving a number of computational steps in observation space as for LETKF (Local Ensemble Transform Kalman filters)? Can the author elaborate in only a few lines on these algorithms, in order to provide an insider view about how easy/how difficult/how different the implementation of such methods in ABC-DA would be?

   (a) As I see it, the extended or ensemble Kalman filters (of whatever flavour), or any of the methods that compute an explicit gain matrix would not follow from such a variational method presented. Some further comments about possible developments will be given at the end of the summary section concerning the possible extension of the system to ensemble-variational or hybrid systems, including the hybrid gain system of [1].

4. §6, line 776: typo "... that that ..." => "... that the ..."

   (a) This will be corrected.

5. line 774-780: in the discussion of the "control-ability" of the v-component of the wind fields. Is this weak control-ability due to the specific formulation of the ABC toy model? (my guess is "yes"). Can you comment this more in the discussion?

   (a) I think that I was wrong to suggest that $v$ is not controllable in the DA experiments. Instead I need to say that $v$ is updated in the wrong way. I will add more in the discussion (and remove the idea that it is due to low controlability). I suggest the the use of observations not at the analysis time (and hence the approximations due to 3DFGAT) might be a reason why the results for $v$ are not good, although this is not proven.

6. 6) §4.2-4.3-4.4 & §5: One general question I have is whether the ABC-DA system can allow the use of a full, total field, B-matrix (that is, one without any balance modeling)? If a total-field B-matrix would be feasible, then the corresponding ABC-DA system could be a valuable reference system for impact studies on specific B-modeling. Can you comment on this?

   (a) I will discuss in the new appendix (appendix A) that the number of pieces of information needed to determine a full B-matrix empirically is not really feasible.

# References

[1] Stephen G Penny. The hybrid local ensemble transform Kalman filter. *Monthly Weather Review*, 142(6):2139–2149, 2014.