

gmd-2020-292: Ensemble prediction using a new dataset of ECMWF initial states - OpenEnsemble 1.0

by Pirkka Ollinaho, Glenn Carver , Simon Lang, Lauri Tuppi, Madeleine Ekblom, and Heikki Järvinen

1 Overview

This paper introduces a data set of ECMWF ensemble initial conditions of three different resolutions for a period of one year (Dec 2016 to Nov 2017), which has been made freely available to the scientific community. Three different types of ensemble are made available – (1) a control member plus singular vector (SV) perturbations, (2) a control member plus ensemble of data assimilation (EDA) perturbations, and (3) a control member plus SV and EDA perturbations combined. Fifty ensemble members are available (plus the control members). The paper gives some explanation of how the data are accessed and manipulated using basic tools, and shows some examples of ensemble scores and a tropical cyclone example.

2 Recommendation

I would recommend that the paper is published in GMD as it highlights a very useful resource for atmospheric scientists. There are a number of changes that I would suggest to improve the clarity of the paper, as outlined below.

In my report the text of the paper is referenced by Lx (line x, as labeled in the manuscript), or by section/figure/equation/table number, and I often quote from the paper to help refer to the part that the comment refers to. Text that I suggest to be added are underlined, and items that I suggest to be removed are scored out.

3 Scientific and major points

1. L16-17 “Due to limitations in observations and in the data assimilation system, a measure of uncertainty remains in this state estimate.”: Surely there is no direct measure of uncertainty from a single estimate.
2. L101-102: Could you please clarify which boundary conditions are perturbed in the EDA (e.g. sea surface temperatures, etc.)?
3. L108 “For each singular vector perturbation an individual linear combination from all singular vectors is constructed.” I cannot make sense of this sentence. Should the first mention of “singular vector” in this sentence really be “ensemble of data assimilation perturbation”? That would then make sense – that each EDA perturbation is then complemented by a random combination of singular vectors.
4. L178 (step 3): Is this step specifically for the OpenIFS?
5. L237: “We use here operational ECMWF analyses from the forecast period as the truth.” I understand that the states that are available in this OpenEnsemble 1.0 are also ECMWF *analyses*. Could the authors remind the reader of the difference between the mentioned ECMWF analyses and the states that are down-loadable as part of OpenEnsemble 1.0?
6. L238: “The model output is truncated to a $1^\circ / 1^\circ$ regular grid ...” Does this apply to the forecasts and the nominal ‘truth’ analysis? What are the three available model resolutions (and the operational analysis) in terms of degrees?

7. L258: “The recommended ensemble size was set to be four to eight members for scientific testing.” All ensemble sizes are in this range or larger, so why aren’t all the CRPS values in Fig. 3a equal to their ‘fair’ values?
8. L269-270: “Noticeably, T_L159 resolution with EDA perturbations scores better in the Tropics than T_L639 with only SV perturbations active.” This seems to be true only for $t < \sim 168$ h.
9. L270-271: “The SV perturbations in the Tropics consist only of perturbations around active Tropical Cyclones, thus the relatively high fair-CRPS in the Tropics is expected.”
 - (a) This means that the authors have used the five SVs that are produced associated with tropical cyclones (as mentioned on L92), instead of the 50 NH/SH SVs produced. How can one use these five tropical cyclone SVs from the downloaded data (instead of the 50 NH/SH ones)?
 - (b) “... relatively high fair-CRPS ...” Relatively high compared to what (especially as the fair CRPS values for the tropics are smaller than those for the NH)?
 - (c) L276-277: “In TR the increase of skill becomes noticeable beyond forecast lead time of 96 h (120 h) for EDA (SV) perturbations.” This is quite a subjective measure of the time that the factor-1 and factor-1.2 lines split. For example you could argue that for the SV perturbations the factor-1 and factor-1.2 lines split between 24 and 48 hours into the forecasts. It really depends on how much one zooms into the plots.
10. Are there any plans to include other years’ data.

4 Presentational points

1. L102: Please define SPPT.
2. L135: Are these URLs stored in a permanent repository? Is there a parent web page that has a click-able list all of the files?
3. L136-137: What do pan, psu, pua and pert mean? See also point 5 below.
4. L143-144: “On top of these four files, OpenIFS requires static climatological files for radiation calculations and various namelists describing the hybrid sigma coordinates etc.” Are these files and namelists also available from the repository? If not, does this mean that OpenIFS cannot be run?
5. Table 1 is a little mysterious to me. I don’t know what the “Use as” column means – it appears to be just a list of files and empty elements. Also some of the files do not have a description. The caption of this table would benefit from some more explanation, e.g. enabling the reader to understand the filenames. See also point 3 above.
6. Similarly for Table 2, the “Manipulation” column is a little mysterious to me. I would have thought that this column would contain references to some procedures that need to be performed to manipulate the files. Instead it just contains a list of files in addition to those listed in the “Files needed” column.
7. All figures: Would it be possible to include the key to the lines in both panels? When I read the paper I had to zoom-into the plots a lot and it would be useful to have the keys right next to the lines themselves in all the panels.
8. L256: Would the authors consider defining the CRPS mathematically?
9. L264: This sounds confusing. Should it read, “As per construction, the ~~normal~~ CRPS scores smaller the normal CRPS score, the ~~better~~ the more ensemble members the ensemble contains (Fig 3).”?

10. L295-296: “Lang et al. (2012) show how especially the SV perturbations can rapidly alter the TC location and intensity.” This reads that Lang et al. (2012)’s study is of the same tropical cyclone. This cannot obviously be the case as typhoon Damrey happened after this publication. Would suggest that the sentence reads, ““Lang et al. (2012) show how especially the SV perturbations can rapidly alter the a TC’s location and intensity.”
11. L319-320: Suggested change, “Increasing the amplitudes of the initial state perturbations result in an increase of the forecast skill of the system, demonstrating that inflation tuning of the initial conditions can improve forecast skill.”
12. Appendix B:
 - (a) Are these commands performed in the shell or inside another piece of software?
 - (b) Most of the commands have an obvious syntax, but some generic descriptions would help (e.g. for one of the more obvious commands: `cdo -sp2gpl <input file> <output file>`).
 - (c) I would recommend repeating the definition of “CDO” and the reference to CDO documentation in this appendix.
 - (d) Where can CDO be obtained from?
 - (e) What is the difference between `-genbil` and `-remap`?
13. Appendix C: This figure would be more logically placed in the main text instead of in a separate appendix. Perhaps there is a reason for putting it separately?

5 Minor points (typos, etc)

1. L129: “... process and ~~amount~~ the number of observations ...”
2. L157: “... the +/- symmetry is ~~build~~ built into the files ...”
3. L181: “... tasks ~~for~~ e.g. ...”
4. L213: Also, the different initial perturbation types (SV and EDA) were tested separately and together in order to illustrate ...
5. L216: “Running large numbers of ensemble forecasts requires a substantial ~~amount~~ of computational resources.”
6. L243: “Both kinds of perturbations also improve ...”
7. L256: “Leutbecher et al. (2017) illustrated ...”
8. L262: “from 2019 onwards in the ECMWF operational ensemble ...”
9. L268: “... this is ~~inline~~ in line with ...”
10. L286: “... west of the Philippines ...”
11. L304: “... momentum exchange between the ocean surface and the atmosphere?”
12. L306: “ ... states covering a one year period ...”
13. L324: “In operational ensemble configurations, having ...”
14. L328: “... scheme (Buizza et al., 2008a) ~~but~~ and also an early version ...”
15. Table A2: I presume that z is the geopotential, which is missing from the description.
16. L 355: “... field subtraction can be done in spherical harmonics representation as well as in grid-point space.”