Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-292-RC2, 2020

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Ensemble prediction using a new dataset of ECMWF initial states – OpenEnsemble 1.0" *by* Pirkka Ollinaho et al.

Hannah Christensen (Referee)

hannah.christensen@physics.ox.ac.uk

Received and published: 18 December 2020

Ensemble prediction using a new dataset of ECMWF initial states - OpenEnsemble 1.0

Pirkka Ollinaho , Glenn D. Carver , Simon T. K. Lang , Lauri Tuppi , Madeleine Ekblom, and Heikki Järvinen

This paper describes a new dataset of perturbed initial conditions which enables researchers to initialise ensemble forecasts using the OpenIFS. The initial condition perturbation methodology is that used operationally (a couple of cycles ago) at ECMWF. The authors also present and provide a workflow manager to assist in producing ensemble forecasts using these initial condition files. Results are presented which demonstrate the impact of different initial condition perturbations available in the

Printer-friendly version

Discussion paper

dataset on forecast skill.

I was happy to review this manuscript, as I believe the dataset and workflow manager described within will be a real boon for the research community. I recommend the manuscript be published subject to the correction suggested below, which I hope will improve the clarity of the manuscript.

» General comments

1. Several of your figures show subtle differences between the forecast skill with EDA, SV or both perturbations. It would be helpful to assess the statistical significance of the differences between different initialisation methods. This will demonstrate that the database contains sufficient independent start dates for researchers' needs. For the significance you will need to compare forecasts pairwise for the same dates, as some dates will be more predictable than others. You could include the 95% significance levels on your scores in the figures (e.g. as in Christensen et al, 2017, DOI:10.1002/qj.3075), though I appreciate this may be difficult for the figures showing results from many experiments.

2. Do you plan to make more start dates available, and with what kind of frequency?

» Specific comments

L21-24 While it is harder for the academic community to contribute to ensemble forecasting research at the moment, it is not impossible. But it does have to be carried out in close collaboration with an operational centre, e.g. through an ECMWF special project. You should soften this statement to acknowledge this.

L45-48 "Although the . . . quality early on". This statement doesn't seem to fit here. You could move it up to the end of paragraph L32-36

L73 "physical grid" – you should expand this for those unfamiliar with the IFS, to indicate you mean the grid on which the model physical parametrisations are run

L92 & L106 On the first line, 50 SVs are mentioned, while on the second, 25 SVs are generated. This needs to be clarified.

L124 clarify, e.g. to: “the highest model resolution available in OpenEnsemble is TL639 (instead of TCO639 as used at ECMWF)”

L125 This seems to be a good place to highlight the three resolutions available in OpenEnsemble, which I don’t think you mention explicitly in this section at the moment

L137-152 including Table 1. This was an extremely confusing section. - what are “pan”, “psu”, “pua” and “pert” on L137 - L139-140 highlight briefly why you need files on both GG and SH grids (different variables are available on different grids, as expanded in the appendix) - the file names are long and unintuitive – what do all the bits of the filename mean? ICM? UA vs T? There doesn’t seem to be an indicator of whether the files are model levels or surface levels in the name. - Table 1 is very perplexing. What does “use as” mean? - descriptions are missing for rows 5, 6, 7 in table 1 - the contents of the CL file are not included in a table in the Appendix. This would be helpful - while running the model, is the assumption that all the surface fields are held constant, or do they evolve? I suppose the land model is included, so those variables would evolve, but what about SST? - L147-152 I was confused why some of the files contained only EDA perturbations, while others contained EDA plus SV. Again the filenames “pan”, “psu”, “pua” and “pert” are rather unintuitive so don’t help the reader understand what’s going on - what has become apparent by this point is that perhaps the EDA and SV perturbations are not applied to all the variables in the different input files needed by the IFS. It would be helpful to explicitly say which variables are perturbed by each method in section 3.3

L154-160 and Table 2. I found it helpful for you to include the cdo commands, and this does make sure that the user is absolutely clear as to how to use the data. However I found Table 2 confusing. In particular the third column “manipulation” was a mystery.

L244-245 Not only do they not grow, they shrink over the first 48 hours. Why is this?

Can you comment here? In addition, the initial spread seems to be too large when using the EDA perturbations – again, why is this?

L264-265 “the normal CRPS . . .(Fig 3)” – this sentence needs to be rephrased – perhaps too many “the”s

L268 “generates less spread than using only EDA perturbations” this figure is not showing spread and error, but is showing CRPS, so this statement should be linked through to a resultant lower CRPS.

L327 Is the backscatter scheme not also available, e.g. as legacy code?

Warmest wishes, Hannah Christensen

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-292, 2020.