Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC2
https://doi.org/10.5194/acp-2021-638-RC2, 2021


# Comment on acp-2021-638

Anonymous Referee #2

---

Referee comment on "Assessing the value meteorological ensembles add to dispersion modelling using hypothetical releases" by Susan J. Leadbetter et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-638-RC2, 2021

---

Can we assess the value meteorological ensembles add to dispersion modelling using hypothetical releases?

ACP-2021-683

This paper aims to assess the value of using meteorological ensembles with dispersion models for a range of meteorological conditions. They perform 2 hypothetical releases and compare the skill of a dispersion simulation using deterministic forecast to the skill of dispersion simulations using ensemble forecasts. They then assess the value of the respective forecasts using the Brier Skill Score. Overall, the paper is well structured, and the description of the simulations is clear. There are however several places where a more detailed discussion of how to interpret the BS and BSS is needed (see general comments below).

General comments

- Line 120, 178 and 420-422. The discussion of configuring ensembles to perform better for certain variables and certain parts of the atmosphere is interesting and would benefit from a lengthier description. In this study, simulations are performed using the MOGREPS-G meteorological ensemble. Has this ensemble been optimised to produce a maximum growth rate of the ensemble spread at a certain forecast lead time? Would differently configured ensembles be more suitable for dispersion applications?
- Line 331. The authors correctly state that the BSS provides a comparison of the performance of the ensemble relative to the deterministic forecast and does not provide information about the individual performance of the ensemble. Therefore, if the deterministic forecast is accurate the BSS can be negative even if the ensemble forecasts are also representative of the analysis. I would like to see this argument in

the introduction section if possible as it's an important point for interpreting these relative skill scores. This is particularly exemplified in figures 12 and 13. By eye the ensemble forecast appears to perform in a very similar manner to the deterministic forecast, but the BSS shows that relatively, this ensemble is worse.

- Line 204, 282, 291 and elsewhere. The Brier Score is calculated for a single output grid square. Does the size of the grid matter? For example, the authors state that the ensemble runs perform better than the deterministic runs at later time steps and hypothesise that this is due to increased ensemble spread at later times. Another reason could be that the plume has spread out more at later times reducing the potential for a double penalty issue. This issue also highlighted in figures 5 and 6, do the negative BSS occur when the plume is narrow, i.e. at the start of the simulations? When calculating BSS at the grid scale small displacements in the plume location can result in large differences compared to the analysis. This occurs particularly when the size of the eddies causing dispersion are large compared to the width of the plume. Would it be possible to show the BSS vs area covered by plume, in an analogous way to fig 7. Finally on line 282; do the authors know why there is a difference in the rate at which the Brier skill score increases with forecast time for different flight levels? One explanation could be that the plume spreads more rapidly at the lower levels due to increased turbulence and remains tightly constrained at upper levels?
- Although not the aim of this paper, it would be of value in the conclusions to discuss how forecasters/decision makers might make use of ensemble dispersion forecast output.

Specific comments

- Why is the title posed as a question? The answer is clearly yes, but ensembles add value is what is being addressed here.
- Line 55. The authors refer to the computational expense of running a statistical emulator. In my experience statistical emulators are built precisely because they can mimic the response of a dynamical model but much faster because they only rely on statistical relationships. Perhaps I have misunderstood the meaning of this sentence?
- Line 75. Not all ensemble systems perturb both the initial model state and the model physics. Therefore 'and' should be 'and/or' in this sentence.
- Line 96. In this section there is reference to the Brier skill score and use of lagged ensembles. These terms should be explained. For example, does the 'most recent ensemble' refer to the lagged ensemble with the shortest lead time?
- Line 99. The term 'dispersion ensembles' is somewhat ambiguous. All Lagrangian model dispersion simulations are ensembles in the sense that they release an ensemble of particles and track their motion. I guess the authors are referring to dispersion simulations run using ensemble of meteorological fields. This is a bit wordy but should be explained in full the first time to avoid ambiguity.
- Line 151. Here the authors use a 20kmx20km horizontal grid spacing, but earlier (line 136) they use a 10km x 10km grid spacing. Why as a different grid spacing used for the two scenarios?
- Line 213. Do the authors have a reason or hypothesis for why the highest threshold is exceeded around 100km from the release location for all the release locations?
- Line 239. The analysis and deterministic met have the same grid spacing while the

ensemble met has coarser grid spacing. Does this impact the results? If so, why not coarse grain the analysis and deterministic met to the same grid spacing as the ensemble met?

- Lines 358-400. These two paragraphs are a repetition of the methodology and are not conclusions. Therefore, it is not appropriate for them to be in the conclusions section.
- Table 1. What time period are the accumulations over?
- Figure 2, 3 and 4. Are the averages over all releases?
- Figure 3. Why is there a larger spread in the average maximum distances for the different concentration thresholds in the lowest layer (FL000-200) compared to the higher layers? Is this due to deposition of particles to the surface? If the particles did not deposit to the surface does this difference in spread decrease?

Typographical errors

- Line 146. There is an extra space before 800m.
- Line 308. 'at and' should be 'and'