Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2021-638-RC1, 2021


# Comment on acp-2021-638

Anonymous Referee #1

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-RC1, 2021

**General comments:**

In this contribution, Leadbetter et al. examine whether a dispersion model ensemble constructed using input from a meteorological ensemble model can help to understand and quantify the impact of meteorological uncertainty on dispersion model predictions. Using the Met Office ensemble prediction system (MOGREPS-G) in combination with the NAME dispersion model, two hypothetical emission scenarios (radiological and volcanic ash scenario) are investigated.

The paper is well written and the presented work advances our understanding on how to perform a quantitative evaluation of the meteorological uncertainty for dispersion model simulations by considering ensemble meteorology. This work demonstrates that dispersion model predictions based on ensemble meteorology generally performed better than the deterministic dispersion simulation. This information is particularly important for dispersion modellers and VAACs, which in the future may be required to provide more detailed and quantitative uncertainty analyses for their dispersion model predictions. I recommend publication after addressing some minor comments discussed below.

**Specific comments:**

My main comment is related to the period over which the simulations are performed. As explained in L. 153, all the simulation are initialised during a single winter season (2018-2019), but I miss an explanation to why these months were selected for this study and how this choice could potentially influence the results. For example, if persistent weather conditions were present during these four months, this could strongly bias the resulting skill of the ensemble.

Also, weather conditions during summer can be very different, therefore I wonder if the statistics would change as well if one were to redo the same experiments for a different period of the year or equally for another winter season. The authors do indicate this potential limitation (L.426-433) and I appreciate that it would require a large amount of extra work to include more months in the study, but I think the paper would benefit from a longer discussion of the potential impact of the selected study period on the robustness of the conclusions.

In this study two very different scenarios (near-surface radiological release versus tropospheric/stratospheric volcanic ash eruptions) are explored to examine if the ensemble meteorology produce more skilful dispersion model predictions. Reading through the manuscript, I did miss some discussion that compared the two scenarios in terms of their relative improvements when using the ensemble forecast (i.e. is it more important to consider the ensemble forecasts for boundary layer releases or volcanic eruptions or do they both benefit equally from the ensemble forecast?). In the final paragraph of the paper (L.438-440), the authors mention that on average the ensemble meteorology outperforms the deterministic model. I think the paper would benefit from a short discussion on whether the impact of using the ensemble meteorology is more significant for one of the two scenarios.

Finally, I also noticed is that throughout the manuscript (text, legends and figure captions), many of the units are not written in the format as outlined in the ACP submission guidelines (e.g. mg/$m^3$ instead of mg $m^{-3}$). I have annotated most of them in the technical comments below, but please check carefully throughout the manuscript.

**Technical corrections/suggestions:**

L12: '… at those later time steps for deposition than for air concentration.' Based on figure 6 and 7, I am not sure that the differences in BSS between deposition and air concentration at the later timesteps are significant enough to support this statement. Would it be possible to give an estimate of the uncertainty in the average BSS reported in these figures (e.g. standard deviation)?

L47: comminicate -> communicate

L65: need a space between 45 %.

L78: The acronym SNAP needs to be spelled out.

L84: In the following paragraphs several of these case studies are discussed, but no citation/discussion is presented for the ensemble studies of Eyjafjallajökull and Grimsvötn. I think this should be included.

L105: used the wrong quotation mark ‘dimension’

L109: ‘…use analysis data…’ From the text it is not directly clear what analysis model data / meteorology is. I think a short description would be useful, as it is a key part of the analysis presented in this paper. Or alternatively, the authors could refer to L.186.

L119: ‘…that NWP ensembles including ensembles may…’. I do not understand this sentence, do you mean: ‘…that NWP ensembles may…’?

L130: space needed between 50 m

L131: 1 μm should be in roman font.

Figure 1: What does the green box represent in panel a? Also, a colour scale for both figures (I assume the colours represent topography) is missing.

L135: ‘…full 48 hours were output.’ Is there a reason to select the 48 hours? In most cases, contaminants will be airborne for longer than the 48 hours. Do we get a different result if we take the output for longer/shorter integration times? I feel some additional argumentation for selecting the 48-hour accumulation period is needed here.

L136: “10km” space between value and unit

L138: “12km” and “25km” need space between value and unit

L142: “g/hr” should be replaced by g $hr^{-1}$

L143: $kg/m^3$ should be replaced by kg $m^{-3}$

L144: “ash was modelled for 24 hours.” Related to the comment for L135, is there a reason why the simulations length is chosen to be 24 hours for the volcanic simulations?

L.151: “20km” space between value and unit

L.153: The period selected for the simulations is one winter season (NDJFM 2018-2019). The aim of the study is to capture a range of meteorological conditions and I am worried that the current selection only captures a limited number of the possible conditions. If a certain weather pattern was dominant during these months (e.g. for the 2018-2019 winter season the NAO was mostly negative), this might have a large impact on the skill of the dispersion model. Furthermore, mass is removed from the model atmosphere by wet and dry deposition, which I assume can be very different in summer especially for the boundary layer releases. All of this will influence the Brier skill score (BSS). I think some discussion regarding this point should be included.

L155: remove “the”

L161: repetition of definition ‘Volcanic Ash Advisory Centre’ (L139), so can be omitted here.

L176: “20km” space between value and unit

L177: “30km” space between value and unit

L185: “focus is on the two forecasts at 00 and 12 UTC.” I am slightly confused by this sentence, as in the next sentence also the 06 and 18 UTC forecasts are used. Do the authors mean that both for the ensemble and deterministic configurations only the 00 and 12 UTC forecasts are used, but for the analysis meteorology all forecasts are included?

L206: “…thresholds for the radiological scenario were chosen to reflect typical distances…” There are no references to any literature to indicate that the found values in figure 2 are indeed typical distances for 48 hrs after the occurrence of a radiological dispersion events. I think some information supporting the typical values should be included.

L212: exceed -> exceeded

L213: “100km” space between value and unit

Figure 2: exceeeded -> exceeded. Also need to change the units in the legend to Bsq m$^{-3}$

L214: “650km and 1000km” space between value and unit

L215: “700km and 1300km” space between value and unit

L222-223: mg/m$^{3}$ -> mg m$^{-3}$

L225: mg m$^{-3}$

L226: “1 km” space between value and unit and remove italics

Figure 3: exceeeded -> exceeded. Also need to change the units in the legend to g m$^{-3}$

L231: “250km” space between value and unit

L232: “1600km to 2400km” space between value and unit

L238: What could be the impact of the coarser resolution of the ensemble meteorology on the actual dispersion simulated? Would it be possible to downscale the deterministic and analysis meteorology to the same resolution of the ensemble meteorology and run the dispersion simulations with the reduced resolution for one of the cases to test the impact on the results?

Figure 4: units in the legend should be Bqs m$^{-3}$ instead of Bqs/m$^{3}$

Figure 5: 50*kBqs/m*$^{3}$ -> 50 kBqs m$^{-3}$ and 5*kBq/m*$^{2}$ -> 5 kBq m$^{-2}$. Also, many of the crosses in are overlapping making it hard to understand the density of the crosses. I think a box plot would show the same information more clearly.

L270: “10 and 30 percent” Are these values for the data shown in figure 5 or is this related to all the different thresholds? Based on figure 4, I would think that the fraction of negative Brier skill scores can be much higher when considering the highest thresholds?

Figure 6: Should this be “Averaged Brier skill score…”? The second comment is related to a possible bias in calculating the average Brier skill score (BSS). As mentioned by the authors, the BSS can range from -infinity to 1 (L266), which is why the adjusted BSS was introduced in equation (3). However, in figure 7, several panels show at least one simulation member for the adjusted BSS to be approximately -1. This indicates that the actual BSS was <<-1. If very low BSS occurred for several individual simulations, you could end up with a negative average BSS, even though the rest of the simulations could be near perfect with a score of approximately 1. Therefore, I think it would be useful to report the range of the actual BSS values (or just the minimum) to understand if this could have impacted the average BSS score presented in the paper.

L281: ‘Brier skill scores increase with forecast time,…’ How much does the increased number of grid points where we have a plume influence the sensitivity of the Brier skill score? During the initial stages of the simulations, only a small number of points have concentrations above any threshold in all three (ensemble, deterministic and analysis meteorology) simulations, while after e.g. 24 hours the plume has spread over a much larger region (as shown in figure 10). If you misrepresent one grid point of the analysis in the earlier stages for the deterministic and/or the ensemble simulation, will this not have a larger impact on the Brier skill score calculated by equation (2) than the same single grid point error after 24 hours? Is it possible that part of the increase in the Brier skill score with time we see in figure 6 and the reduced range in values in figure 7 for later timesteps is caused by the increased plume size?

L284: “12km” space between value and unit.

Figure 7: Similar to figure 5, I think a box plot would show the same information more clearly. There also seems to be a single simulation in the dataset that seems to perform much better than all the other simulations and shows up on all the panels in this figure. Is this the same simulation for all panels and is there something special about this simulation?

L303: I think this should be ‘This resulted in an average Brier skill score of 0.76.’

Figure 8: Please use a different colour for the green or red contour, as it is hard to see the difference. Also replace ‘*kBqs/m*$^{3}$’ -> ‘kBqs m$^{-3}$’.

L305: ‘200*kBqs/m*$^{3}$’ -> ‘200 kBqs m$^{-3}$’

Figure 9: What does the blue triangle in panel a represent? What is the altitude for these wind speed and directions? Are these altitudes at which the contaminants are released in the model?

L315: '2*kBq/m*$^{2}$' -> '2 kBq m$^{-2}$'

L334: Why are you using the Brier skill score for the volcanic simulations and not the adjusted Brier skill score like it was done for the radiological scenarios?

L339: 2mg/m$^{3}$ -> 2 mg m$^{-3}$

Figure 10: Please use a different colour for the green or red contour, as it is hard to see the difference. Also replace 2mg/m$^{3}$ -> 2 mg m$^{-3}$.

Figure 12: Please use a different colour for the green or red contour. Also please change 5mg/m$^{3}$ -> 5 mg m$^{-3}$.

L365: increase -> increases

L433: small -> smaller