Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2021-858-RC1, 2021

Comment on acp-2021-858

Anonymous Referee #1

Referee comment on "Refining an ensemble of volcanic ash forecasts using satellite retrievals: Raikoke 2019" by Antonio Capponi et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-858-RC1, 2021

Review of Refining an ensemble of volcanic ash forecasts using satellite retrievals: Raikoke 2019 by Antonio Capponi, Natalie J. Harvey, Helen F. Dacre, Keith Beven, Cameron Saint, Cathie A. Wells and Mike R. James.

This paper presents a filtering method for volcanic ash simulations and it is applied in a case of study for the Raikoke 2019 eruption. They use a a large ensemble of simulations and column integrated ash load retrievals from a geostationary satellite for filtering the ensemble. They show the methodology and their results include improvement in the forecasts of ash, also showing how the filtered forecast could affect the planning a set of flight routes over the Pacific Ocean, compared with the raw ensemble forecast.

The manuscript explain in details the context and the method (including most of their assumptions and limitations). It show the results in a clear and concise way and it presents the conclusions accordingly. The article is well structured, well written, and its length is according to the content. The figures are clear and provide interesting information. Therefore, I recommend for publication in ACP after minor corrections.

I would like to ask to the authors some details on the averaging and collocation procedures of the Himawari mean ash column loadings retrievals and their uncertainties ; and, if possible, clarification on the correlated LHS sampling algorithm, and a few clarifications on the general overview of the algorithm (see the minor comments below). I am also wondering which of the decisions were made for avoiding the filter divergence, giving the large number of perturbed parameters of the ensemble and their implications of this known issue in particle filters, which is particularly relevant for high-dimensional problems.

Giving the length of this review, I would be happy if the authors can address the general comments above and the related minor comments below, but I understand that addressing all of them could take more time than expected for a minor review.

Minor comments

L14: The filtering technique was developed in Beven and Binley (1992), not here. The rejection metrics, design of the ensemble, use of the retrievals, etc. are unique for this work (not the technique itself). I would propose to change the word "technique" by "system" or similar, to reflect the large amount of work done in this paper that it not related to the method itself.

L18:L20 I suggest (but not strictly needed) to rephrase this sentence: "The ensemble members are filtered, based on their level of agreement with the ash column loading and their uncertainty of the Himawari satellite retrievals, to produce a constrained posterior ensemble" .

L31: "at a given time" is not needed

L32: which dynamics? Atmospheric? Ash? Wouldn't be clearer : "...VATDMs solves numerical representations of equations related to ash processes in the atmosphere to evolve…" ?

L35: Distributions over which dimension? Are you meaning spatial distributions (as in L40)? Size distributions? Temporal distributions?

L39: I suggest to remove "forward". It could be the inverse model, depending on the retrieval system.

L40: I suggest to avoid the word "distribution" here. "volcanic ash column loading" (instead of volcanic ash distributions (known as column loadings)) is clear enough.

L41: Would you add retrievals errors too?

L42: wouldn't be clearer: "This error information … "?

L50: Indeed, this is the only the "Jo" or the observational part of the cost function (y-H(x)), where y are the satellite retrievals and H(x) the simulations of the retrievals by the VATDM. The VATDM errors shall be included in this observational error covariance matrix (usually denoted by R or Po). The other term in the simplest form of the variational formulation refers to the prior errors (x-xb)^T B^-1 (x-xb), which is the difference of the control with the prior.

L53: I see a small change of the colour of the text after "plume" (but it could be only an artefact of my pdf viewer).

L50, L57-L60 Please reconsider to use the word "sequential". Variational data assimilation can also provide estimates of the system states sequentially (for example the 3DVar or 3DVar-FGAT), and Kalman filtering ("sequential" in the manuscript) can also use a fixed time window (but usually it is shorter than the variational method), for example in the NOAA's GFS.

L64: Shouldn't be better to write "... this estimates a probability density …", or "empirical pdf" ? Can the the full (usually continuous) pdf be computed by perturbing the VATDM parameters and meteorology?

L65:66 Just a suggestion: it could be clearer that, instead of using the "filtering pdf" wording, the authors use "prior pdf" in L65 and "posterior pdf" in L66.

L71: Please add "unbiased" before Gaussian.

L72: It is not clear for me why they are more sensitive to the tails of the prior distribution.

L74: I suggest some rephrasing like :"Bayesian inference is used in particle filtering to constrain simulation parameters with observations".

L75: I suggest to replace "derived" by "computed".

L87:103: Wouldn't be better to move this to Sect. 4?

L124: Advanced Himawari Imager (AHI).

L125. Please indicate if this is the retrieval used in this work, or at least link the Met Office retrieval with Francis et al. (2012) in the text.

L147: This implies that it is needed that all the 10-minute (within 1 hour) retrievals are also flagged as "clear", to have a "clear" regrided pixel, right? Isn't this too restrictive?

L149: What about the regridding of the uncertainties?

Table 1: Any particular reason for not using the control MOGREPS-G for your control run?

L188:189: There is no information on the tropical cyclone on Figure 1.

L189: Please choose Himawari-8 or Himawari in the manuscript.

Figure 1: Please add latitude and longitude to the Figure. The space between the panes could be decreased and you could save space by including only one colorbar for all the panels.

L203: Following L167, the NAME model outputs 6h averaged values, and you are comparing with 1-hour averaged Himawari retrievals. This difference in temporal collocation can be important. Why this mismatch? Can you compare them with a consistent time collocation, by setting the model outputs to 1h averages, for example? Would you expect changes in your results?

L231: I am not sure if this sensitivity is used later in this manuscript. If it is not used, this sentence can be removed.

L233: It was later in the text that I understand that the resampling mentioned in L226 was, in fact, to create a full new 1000 member ensemble sampled (and not resampled) from the posterior pdf, and it was not meaning the usual resampling technique (that basically produces copies of existing members, weighted by the posterior empirical pdf). Is there any way to clarify this here, also indicating (and thus repeating L162) that the new ensemble is run from T0 up to Tn? Is this right or I misunderstood?

Figure 2: This figure could be clearer. Shouldn't be "posterior pdf" in box 5 instead of "parameters resampling"?. The resampling is done in the "LHS" box, or not?. Are the

“ensemble creation” and “NAME runs” the same step (i.e., do you consider than an ensemble is the set of perturbed variables/parameters, or an ensemble is the set of NAME outputs?)?

L243 : Please define MOGREPS the first time it appears in the text. Also, is MOGREPS the same as MOGREPS-G? (please check the text for consistency).

L296: Wouldn’t have more sense to, instead of assuming a constant release duration and constant value of the parameters, to assume a parametrised temporal variation of them, following the qualitative information wrote in this paragraph?. How much this assumption could affects the filtering of the ensemble? Can you envisage how to improve this issue taking advantage of the high temporal resolution of Himawari retirievals? (see my very last comment)

L336: Is the filtering of the “clear” pixels (ie., a subset of the no-matching pixels) biasing your HR metric? With this matching pixels procedure, you are removing from your dataset those pixels where the model simulates ash but the satellite indicates “clear”. Could this play a role in the possible overestimate of the ash horizontal extension shown in Figure 6?

L353: Please see comment on L149. I guess that you are comparing with the regridded uncertainties from the Himawari retrieval. The information on the regridding of these uncertainties is missing in the paper.

L360: As the thresholds are defined later in the text, I suggest to add “Sect. 4.2.4” just before “Fig. 3d”

L368: Same as L360

L366 : Do you mean the absolute value of the difference between simulated and observed values, or you are allowing negatives values of PDs?

L371:372 Both HR and MPD are normalised by the number of matching pixels (Eq. 2 for HR and the averaging of PD for MDP), and I do not see any obvious argument for this statement as the main reason. Aren’t these dynamically adjusted thresholds implemented in an attempt to avoid filter divergence rather than justified by the number of retrievals?

L397: I understand that you cannot show all the ENS, but why did you skip ENS03, while it is the ENS with most retrievals in Table2? (and you also provide ENS03 in the supplementary dataset).

Figure 4 : This figure is very interesting. “Each parameter in the box plots is normalized by dividing each individual value from the ensemble members by the mean of that entire parameter range from the selected ensemble”: Is this meaning that the normalisation is different for the blue and orange boxplots? How can they be compared side by side? Wouldn’t be better to normalise by the fixed sampling range of Table 1, such that a zero value means the lower bound of sampling range, and an unity value the upper bound of the sampling range?

Figure 4: As panel (d) shows relatively small differences of both boxplot colours in comparison with those of (a) , I am wondering how Figure 4 looks like in the following iterations of the filter. Are they similar to panel d? Can you identify signs of filter divergence? It is worth to add this in the Appendix?

L425: The LHS sampling with correlated variables is not trivial, and the Cholesky reference do not add useful information, unless all the process is described. Since this step is fundamental in the method, could you add a reference or explain how the correlated LHS is done?

L428: I do not understand why do you use the posterior pdf for some parameters and keep using the uniform pdfs for the internal parameters. Could you provide a justification for this?

L432: This is a qualitative statement and it should be presented as such. Unless you perform the formal proof of it, I would recommend to rephrase this sentence. In addition, could you state explicitly which EPS shows distributions similar to a normal distribution (particle density, duration)?

Figure 5: Very nice figure. Similar to a comment above, since the ENS06 shows small peaks in the H, DFAF and MER panels, I am wondering what would be the equivalent of these figures for ENS07 to ENS11.

Figure 5: Just a comment: it is interesting to see that in the H panel for ENS01 the distribution is very skewed but while the algorithm iterates, the right tail of the distribution start to be heavier and the peak smaller. Is this continuing in later iterations? Do this have implications?

Figure 5 If the ENS01 perturbations have very well defined range limits and the filtering is done by rejection criteria, how it is possible to draw samples from values outside these range limits in the next iterations?. For example, the particle density, duration, MERF and DAF panels have tails of the distribution that are away from the ENS01 shading. Are you smoothing the pdfs in the ABC procedure?

Figure 6: I suggest to plot the probability of 0-30 percent of ENS03 in orange colour, to avoid confusion with ENS01. Also, please add x and y axis labels and values.

Figure 6: Why do you compare mean values of satellite burden with these probabilities, and not with the mean value (or median, etc) of your ensemble? Could you add a panel with this information? I think that the probability maps are good, but the mean/median value information could be missing.

L435: Why do you have such large mismatch in the area covered by ash loading? It is because limitation in the Himawari retrievals (and clouds)? Or it is because you assumed a constant flux of ash emission during the period? Other reasons?

L470: Why ENS08 at T10 and not ENS10? It is because you are interested in 12-hour forecasts?

Figure 7: I cannot easily see the colour shading of ENS08 for ash > 2 mg/m2 in this figure. I would suggest to only keep the 0.2 threshold (cyan) for ENS08.

L495: Can you be more specific in the NCAR dataset used here? Why using NCAR reanalysis data if you can use the same meteorological (control) simulation of your ensemble?

L479: Is there any independent set of observations to validate your results?

Figure 8: Please add latitude and longitude axis and labels.

L537: As in the abstract, I would replace “methodology” by “system”.

L543:544: Please note that this is not true for two of the six panels of Figure 5.

L546: This is a design choice, not a result of the filtering procedure. Please see my comment above on the justification of leaving these internal parameters unconstrained versus the others.

L557: After this work, is there any plan to revise the set of parameters to perturb? For example, to add more flexibility on the emission timing and duration pulses for other eruptions; to add some degree of freedom to the temporal variability of the parameters and variables; or improved assumptions on the vertical distribution of the emitted ash?.