Review of
"Particle filter based volcanic ash emission inversion applied to a hypothetical sub-Plinian Eyjafjallajökull eruption using the chemical component of the Ensemble for Stochastic Integration of Atmospheric Simulations (ESIAS-chem) version 1.0"
by Franke et al.

Review by Nina Kristiansen

General comments

The paper presents a methodology to estimate volcanic ash emissions as a function of time and altitude based on observations and modelled ensemble simulations. The main strength of the paper is that the methodology also gives an estimate of the uncertainty/errors in the estimated emissions, however this aspect could be more clearly demonstrated throughout the paper with a clearer description of how this uncertainty estimate is obtained. There is potential to make a few sections clearer and easier for the reader to follow, in particular the methodology section. The results shown are interesting and the figures clear. The paper is suitable for publication once the below comments have been addressed.

Specific comments

Abstract:

- "*The system validation addresses the special challenge of ash cloud height analyses in case of observations restricted to bulk column mass loading information, mimicking the typical case of geostationary satellite data*." – unclear what is meant here, please rewrite.
- The abstract should more clearly say that you are using an "idealized situation with artificial observations from a model run" and that you use many observations of both ash and ash-free areas. It should also say that meteorological uncertainty is not included.
- L20: "*This situation, however, can be remedied by extending the assimilation window*". I am not sure this is true, as in the results section you show Figure 10 for the weak wind shear situation that the emission profile is not well estimated and you don't show a better estimation when including more observations (i.e. increasing the assimilation window >24 hrs) for this case.

Section 1 Introduction:

- Line 32: "*Typically, volcanic eruptions occur as sequences of emissions with highly varying ejection mass and height*". This might be correct for explosive eruptions but not necessarily for effusive or passive degassing.
- Line 33: You might want to include that radar observations also are uncertain and have limitations.
- Line 36: "*Statistical models are based on observational data from only a few, highly heterogeneous volcanic eruptions*". I don't think the word heterogenous is right here. The issue with the statistical methods by Mastin/Sparks is mainly that it is biased to larger eruptions (very small number of smaller eruptions were included in the empirical estimates), but I wouldn't say that the eruptions considered were 'heterogeneous'.

- Line 38: "*physical plume-scale models require vent and magma details, which are poorly known, and thus making these models highly uncertain*." It might be more informative to include something on how sensitive the plume models are to the vent/magma details and the expected magnitude of errors associated with this. See the plume model intercomparison study by Costa et al 2016 (https://www.sciencedirect.com/science/article/abs/pii/S0377027316000366)
- Line 50: "*is the horizontally more complete picture of the volcanic ash extent*" unclear what is meant by more complete picture here – and more complete compared to what?
- Line 73: Would be good to include some further details of the advantages and limitations of each method which you mention. Also include more recent papers on data assimilation/insertion:

  Prata, A. T., Mingari, L., Folch, A., Macedonio, G., and Costa, A.: FALL3D-8.0: a computational model for atmospheric transport and deposition of particles, aerosols and radionuclides – Part 2: Model validation, Geosci. Model Dev., 14, 409–436, https://doi.org/10.5194/gmd-14-409-2021, 2021.

  Pardini, F., Corradini, S., Costa, A., Esposti Ongaro, T., Merucci, L., Neri, A., Stelitano, D., and de' Michieli Vitturi, M.: Ensemble-Based Data Assimilation of Volcanic Ash Clouds from Satellite Observations: Application to the 24 December 2018 Mt. Etna Explosive Eruption, Atmosphere, 11, 359, https://doi.org/10.3390/atmos11040359, 2020.

  Osores, S., Ruiz, J., Folch, A., and Collini, E.: Volcanic ash forecast using ensemble-based data assimilation: an ensemble transform Kalman filter coupled with the FALL3D-7.2 model (ETKF–FALL3D version 1.0), Geosci. Model Dev., 13, 1–22, https://doi.org/10.5194/gmd-13-1-2020, 2020.

  Fu, G.; Lin, H.X.; Heemink, A.; Lu, S.; Segers, A.; Velzen, N.V.; Lu, T.; Xu, S. Accelerating volcanic ash data assimilation using a mask-state algorithm based on an ensemble Kalman filter: A case study with the LOTOS-EUROS model (version 1.10), Geosci. Model Dev., 10, 1751–1766, 2017.

  Fu, G., Prata, F., Lin, H. X., Heemink, A., Segers, A., and Lu, S.: Data assimilation for volcanic ash plumes using a satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption, Atmos. Chem. Phys., 17, 1187–1205, https://doi.org/10.5194/acp-17-1187-2017, 2017.

Section 2:

- I generally find the methodology section difficult to follow. There are many technical terms and abbreviations to keep track of, and the descriptions are sometimes not clear. Perhaps an extension of Figure 1 (the flow diagram) to include further steps and references to methods/terminology/naming conventions would help. I also suggest expanding the figure caption of Figure 1 to explain what the figure shows which makes it easier for the reader to return to this figure later while reading subsequent sections.
- L 98: "*Stohl et al. (2011) and Kristiansen et al. (2015) aiming to estimate the optimal emission profile but not its uncertainty*". I think it is fair to say that this work did provide some uncertainty estimates (and included uncertainty in both the a priori, the observations and the model input – though this could of course be improved). In figure 3 of Stohl et al the uncertainty reduction from the a priori via the inversion to the a posteriori is shown. This shows how much influence the observations had, and which parts of the emissions were well constrained by the

observations and which were less constrained, and therefore is a form of uncertainty estimate. I do appreciate that what you are aiming to provide is different (but I still don't quite understand how the uncertainty is estimated!) but it would be good to include some more details here how your uncertainty estimate differ from this to make it clear.

Section 2.2:

- Some further comments on the advantages and disadvantages of the Nelder-Mead method would be good to include. For example, mentioning that the reason the method is suitable for "discontinuous"/"spiky"/"noisy" problems is because it does not use derivatives, but also that it doesn't use a convergence theory and doesn't necessarily find the minimum function value (but rather an 'improvement') – that is the key difference to the method used by Stohl etc.
- "*The minimization was performed in NN0, which has been found to be more effective than the minimization in RN*" –to help the reader please directly spell out what this means. I think you mean in model space rather than in observation space. And if this is the case is this more efficient because the number of ensemble members is smaller than the number of observations?
- How is the initial simplex determined for each ensemble member? You say later it is arbitrary but some more details here would be useful.
- It would also be nice to see some comment on computational effort (i.e. run time) for this system.

Section 2.3:

- L 140: "*It is noted that in the particle filter method no assumptions of the error statistics of the model state and the observations were made*." I don't understand how this relates to the uncertainty estimate you apply in the results section where you assume a 40% uncertainty on the observations….
- L 145 "*the ensemble members with high weights are duplicated and perturbed, replacing ensemble members with vanishing weights*." I don't quite understand how this works in practice with the unit ensemble members... please expand on this part.

Section 2.4

- L 150: "*In order to account for meteorological uncertainties, ESIAS-chem is capable to be coupled with ensembles of meteorological field*". But you have not used met ensembles here? Please clarify in text if this is only a possible extension.
- L 179: "*Finally, a particle filter step is applied. The weights, which result from the filtering step, are applied to the optimized emission profiles.*" Please provide some more details of how the "ensemble mean" is constructed (a term which you use in Section 3) and then also the uncertainty estimates based on the ensembles. I understand that each ensemble member simulates the emission released by a single emission package for an individual time step and height interval and then weighted by the likelihood in the particle filter step. How do you construct the ensemble mean from this? Are the ensemble members still associated with emissions from a single emission package but after the filtering with an amount of ash (rather than a unit release)?

Section 3:

- L187: "*Given the identical twin assumption the experiment is then to be made realistic in all other respects, as the two different weather conditions on our case*". This sentence is not entirely clear. I think you refer to the "twins" as the two different weather conditions. It might be better for the reader if you say this in the first sentence of this paragraph.
- L190: Is EURAD-IM an online or offline model? Please clarify and which meteorological data are used as driver/lateral boundary conditions.
- L203: "*The uncertainty of volcanic ash column mass loading observations is about 40 % (Kristiansen et al., 2015, and references therein*)": Some better references here might be:
    - L. Clarisse, F. Prata: Chapter 11 - Infrared Sounding of Volcanic Ash Editor(s): Shona Mackie, Katharine Cashman, Hugo Ricketts, Alison Rust, Matt Watson, Volcanic Ash, Elsevier, 2016, Pages 189-215, ISBN 9780081004050, https://doi.org/10.1016/B978-0-08-100405-0.00017-3. (https://www.sciencedirect.com/science/article/pii/B9780081004050000173)
    - Kylling, A.; Kahnert, M.; Lindqvist, H.; Nousiainen, T. Volcanic ash infrared signature: porous non-spherical ash particle shapes compared to homogeneous spherical ash particles. Atmos. Meas. Tech. 2014, 7, 919–929. 144.
    - Western, L.; Watson, I.; Francis, P. Uncertainty in two-channel infrared remote sensing retrievals of a well-characterised volcanic ash cloud. Bull. Volc. 2015, 77, 67.

- Equation 7: Here you use an observation error 40% of the observation value. It might be worth commenting here that when using real observations (instead of synthetic as in your case) then a better approach would be to use the retrieval uncertainty estimate for each single satellite pixel, and not a fixed uncertainty value..
- L219+223: "*The length of the assimilation window influences the performance of the data assimilation algorithm due to the influence of vertical and horizontal mixing and vertical wind shea*r." And "*Certainly, by increasing the assimilation window length the observations include more information, as the residence time of volcanic ash in the atmosphere is increased*". The wording here is a bit strange. The residence time of volcanic ash in the atmosphere doesn't increase with increased assimilation window. As the assimilation window increases the number of observations which are assimilated increases… Please rewrite.
- Page 11-12: There is a lot of jumping between the pcc and RMAE. Might be worth explaining first one and the results then the other one and the results.
- L265: 10 μgm^{-3} seems like a very low concentration threshold (considering the aviation thresholds starting at 200 ug/m3). What was the reasoning behind this threshold, and do the results change if the threshold is higher?
- L282: "*The analysis suggests that for the respective test cases an assimilation window of 18 hours, that is 10 hours after the artificial eruption terminated, is sufficient for ESIAS-chem to analyze the exact location of the volcanic ash cloud…*" This is a nice result (and backed up by the RMAE later) which I think you should put in the abstract.
  You could also here refer to work by Fu et al who also analysed this "effective duration" (i.e. the required temporal cycle to obtain improved forecasts). Fu et al. 2015, 2016, 2017 report between 6 and 24 hours.
    - Fu, G.; Lin, H.; Heemink, A.; Segers, A.; Lu, S.; Palsson, T. Assimilating aircraft-based measurements to improve forecast accuracy of volcanic ash transport, Atmos. Environ., 115, 170–184, 2015.

    - Fu, G.; Heemink, A.; Lu, S.; Segers, A.; Weber, K.; Lin, H.X. Model-based aviation advice on distal volcanic ash clouds by assimilating aircraft in situ measurements, Atmos. Chem. Phys., 16, 9189–9200, 2016.
    - Fu, G., Prata, F., Lin, H. X., Heemink, A., Segers, A., and Lu, S.: Data assimilation for volcanic ash plumes using a satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption, Atmos. Chem. Phys., 17, 1187–1205, https://doi.org/10.5194/acp-17-1187-2017, 2017.
    - Fu, G.; Lin, H.X.; Heemink, A.; Lu, S.; Segers, A.; Velzen, N.V.; Lu, T.; Xu, S. Accelerating volcanic ash data assimilation using a mask-state algorithm based on an ensemble Kalman filter: A case study with the LOTOS-EUROS model (version 1.10), Geosci. Model Dev., 10, 1751–1766, 2017b.

- L305: “*good performance of the ESIAS-chem analysis not only in terms of column mass loading but also in terms of the vertical distribution of the volcanic ash in the atmosphere*”. It might be worth re-iterating here that getting a good performance for the concentrations are possibly strongly affected by the use of “perfect meteorology” and that such good results are not expected using real observations.
- L306: I was surprised why the 18 hours assimilation windows wasn’t chosen here over the 24 hours as the 18 h seems to show equally good results up until now.
- You first show the validation using pcc and RMAE (Figs 6-8) for all assimilation time windows, and then the results for one of the assimilation windows (figure 9-10). I would prefer it the other way around so that I can see what the estimated emission profile looks like (for one of the assimilation windows) before it is validated and tested against the other assimilation windows. Also, because the main strength of your method is giving an uncertainty estimate for the emission profile this should be the focus of the results.
- Figure 9 and 10: It would be interesting to see “b” figure for all assimilation windows, to see how the emission profile changes as you assimilate more and more observations, and how the estimated uncertainty (c, d figures) also changes when including more observations.
- L322: In the abstract you say that for the strong wind shear condition the estimated emissions have “up to an error of only 10 %” but here you say relative errors are around 10-20 %. Also, in the abstract you say that in a situation with little wind shear the errors are “higher”, while here you say up to 60% and more. I would change the abstract to give the same numbers as here.
- L337: “*The results indicate that on 29 April 2010 the mixing of volcanic ash in the atmosphere is too effective*”. With much less wind shear on 29 April it seems that the problem isn’t too much mixing but that the emissions at different altitudes and times are transported in a similar way and thus cannot be easily separated by the assimilation.
- L340: “*However, the previous results show that even though the volcanic ash emission profile could not be properly estimated by the system on 29 April 2010, the vertical and horizontal distribution of volcanic ash in the atmosphere is fairly represented by the ensemble mea*n.”. This is a little worrying. The fact that the pcc and RMAE give such good scores even with such a “smooth” emission profile after the assimilation which deviates strongly from the nature run emissions (“truth”)... it does make me question whether the pcc and RMAE are appropriate statistical measures to be used here… But it might be more to do with the fact that with little wind shear many different emission profiles can equally well give a best fit with the observations and the fact that the Nelder-Mead method doesn’t necessarily find the minimum only an “improvement” as previously mentioned. I think the point that the emissions are not well estimated, but that the simulated concentrations and column loadings still fit well with the nature run would be a little bit more explored and discussed.

- L371: "*Thus, ESIAS-chem demonstrates to estimate the vertical distribution of volcanic ash in the atmosphere on both simulation days with a high accuracy*." A probability of 90% from the ensemble does not mean that the simulation is of high accuracy.

Section 4 Conclusions

- This reads a bit more like a discussion and future outlook. I would rename this to "discussion" and include another section for the Conclusions which summarizes the key results you have shown with a few bullet points.

Technical corrections

Line 33: "form radar" – change to "from radar"

Line 285: "extend" – change to "extent"