Response to Nina Kristiansen on her review:

We colored our response to Nina Kristiansen in blue.

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.
We thank Dr. Kristiansen for her detailed and helpful review. We are confident that our modifications are in compliance with her expectations, which would certainly improve our manuscript.

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.
  We have rewritten this phrase: "Thus, the proposed system addresses the special challenge of analyzing the vertical profile of volcanic ash clouds given only column mass loading data as retrieved by geostationary satellite imagery."
- 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.
  We added this information to the abstract. The phrase concerning identical twin experiments now reads: "As initial validation of ESIAS-chem, the system is applied to simulated artificial observations of both ash-contaminated and ash-free atmospheric columns using identical twin experiments. Thus, in this initial performance test the underlying meteorological uncertainty is neglected."

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.
Thank you for this comment. Indeed, extending the assimilation window does not guarantee a better performance. Anyhow, even in the case of weak wind shear, the confidence of the analysis ensemble on the time and height of the volcanic ash emissions increases with increasing assimilation window length. This has been addressed in the new Fig. 11 where you see a histogram of the relative emission factor of the analysis for the weak wind shear test case. We have changed the sentence: „In case of increasing wind shear, the performance of the analysis may benefit from

extending the assimilation window, in which new observations potentially contribute valuable information to the analysis system. "

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.
  We agree. We have added "explosive" before volcanic eruptions in this phrase.
- Line 33: You might want to include that radar observations also are uncertain and have limitations.
  Thank you for this comment. We added this information to the sentence. It now reads: "Only limited observations of volcanic ash emission parameters are available (e.g. eruption plume heights retrieved from radar measurements, Arason et al., 2011), which are affected by their specific uncertainties and limitations, e. g. by orographic shielding."
- 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'.
  Thank you for this helpful comment. We have added this information. The point we wanted to highlight is the large variance of the eruption rate given a specific plume height (Fig. 1 in Mastin et al. 2009). We agree, that "heterogeneous" is insufficient to explain this relation. Thus, we have changed the sentence to: "Statistical models base on observational data from historic volcanic eruptions, which are sparse and show a large variance in eruption rate for given plume heights. For example, Mastin et al. (2009) calculated an uncertainty by a factor of four in estimating the emission rate for a plume height of 25 km using their statistical model."
- 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)
  As for the statistical models, we have added some information about the uncertainty of the plume-scale models. The sentence has been expanded: "Physical plume-scale models require orographic details of the volcano (e. g. crater size) but also meteorological fields and parameters (e. g. wind entrainment coefficients), which are often poorly known and render these models highly uncertain. Costa et al. (2016) identified the wind entrainment coefficient as main source of uncertainty leading to up to two orders of magnitude differences for the estimation of mass eruption rates for weak volcanic eruptions. In their analyses of the eruptions of the Eyjafjallajökull, Iceland, in 2010 and Grímsvötn, Iceland, in 2011, Woodhouse et al. (2015) found a comparable range of uncertainty depending on the choice of the wind entrainment coefficients."
- 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?
  We have rewritten the sentence: "Column mass loading observations as

available from, for example, the Spinning Enhanced Visual and InfraRed Imager (SEVIRI) on board Meteosat Second Generation (Schmetz et al., 2002) are beneficial for source inversions as they provide measurements of the horizontal extent of the volcanic ash cloud with a frequency as high as 15 minutes, which is used for analyzing the temporal evolution of the volcanic eruption column. "

- 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.

Thank you for providing the additional references. We address these now in the introduction. Further, we added some more information to the paragraphs describing the approaches in the literature for analyzing volcanic eruptions. The full section now reads: "First estimations of volcanic ash emissions from the 2010 Eyjafjallajökull eruption in a high temporal and vertical resolution were made by Stohl et al. (2011) and later by Kristiansen et al. (2012) and Kristiansen et al. (2015). Their algorithm bases on the inversion technique of Eckhardt et al. (2008), in which an optimal combination of distinct emission packages is estimated using a least squares method. The method showed to provide reliable a posteriori estimates of the time-varying emission profiles. Stohl et al. (2011) include errors from a priori estimates, retrieval errors and model errors and discussed results in terms of relative error reduction subject to assumptions made. Schmehl et al. (2012) initiate the volcanic ash analysis using an ensemble of simulations with random emission strengths and wind fields. Their best estimate of the volcanic ash concentration is found iteratively using a "genetic algorithm variational approach". Herein, rather strong assumptions on the emission profile are made: the emissions are kept fixed for the simulation duration; emissions are placed into a single model layer; wind fields are only adjusted in the model layer containing volcanic ash emissions. However, the method provides a quick and easy to implement first estimate of the volcanic ash concentrations in the atmosphere. Yet, the strong assumptions may render the approach unfeasible for longer lasting volcanic eruptions in which the

emissions vary more strongly. Another data assimilation method for estimating the volcanic ash emissions was proposed by Lu et al. (2016). They developed an adjoint-free, ensemble-based four-dimensional variational data assimilation (4D-var) method. The method showed reliable estimates of the true emission profile in their experiments using synthetic, vertically integrated satellite observations. However, they do not address the uncertainty estimate of their analysis.
Zidikheri et al. (2016) and later Zidikheri et al. (2017b) developed an assimilation system that aims to analyze the horizontal distribution of volcanic ash column mass loading rather than the emission strength. This study was extended by Zidikheri et al. (2017a) to additionally estimate the height and the particle size distribution of volcanic ash emissions using a parameter refinement method. Here, an ensemble of source parameter values has been applied. Using a proper metric (in their case the pattern correlation coefficient) the ensemble is evaluated against observations. The best fitted ensemble member is taken as analysis. The method is easy to implement for a fast analysis of a volcanic eruption as only the upper and lower bounds of the considered source parameters need to be defined. However, the number of model runs used to find the analysis increases exponentially with the number of parameters. Rough estimates of the parameters' uncertainty are provided by the spread of the top 2 % ensemble members with respect to the metric (Zidikheri et al., 2017b), which does not take uncertainties in the observed quantities into account. Wilkins et al. (2014) used the "data insertion" method, in which observed volcanic ash column mass loadings act as virtual sources for volcanic ash with a predefined vertical distribution. The algorithm was successfully applied to the eruptions of Eyjafjallajökull, Iceland, 2010 (Wilkins et al., 2016b) and Grímsvötn, Iceland, 2011 (Wilkins et al., 2016c). Given the lack of vertical information in column mass loading retrievals of volcanic ash, the data insertion method needs assumptions about the vertical distribution of the volcanic ash content in the atmosphere. Thus, this larger source of uncertainty for the volcanic ash analysis is ignored. The data insertion scheme has also been implemented as a first step towards an ensemble-based data assimilation scheme in the FALL3D-8.0 atmospheric transport model (Prata et al., 2021).
Fu et al. (2017) developed a mask-state algorithm for ensemble Kalman Filters to reduce the size of the state vector to be optimized. More recent applications of the ensemble Kalman Filter and its variants are provided by Pardini et al. (2020) and Osores et al. (2020). By estimating the source parameters of the volcanic eruption, the approaches using the ensemble Kalman Filter assume constant emission parameters between two assimilation steps. This is a rather strong assumption on the emissions especially if observational data is sparse or far away from the volcano. However, keeping this assumption in mind the ensemble Kalman Filter methodology provides an estimate on the analysis uncertainty."

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.
  Thank you for this suggestion to improve the general description of the method. We have updated the caption of Figure 1 as you suggested. It now reads: "Schematic of the ESIAS-chem analysis workflow. The analysis is initiated with

an ensemble of emission packages at time $t=t_0$ and restarted, when new observations become available (left side, cf. Sect. 2.1). Here, t_{i+1} corresponds to the observation time. Previously calculated simulations with emission packages within the time interval t_0-t_i may be restored (upper panel). Simulated volcanic ash is compared with perturbed observations for the whole simulation (i. e. from t_0 to t_{i+1} (upper center panel). The resulting volcanic ash concentrations are passed to the DENM minimization algorithm that produces an ensemble of emission profile analyses (right panel, cf. Sect. 2.2) by finding an optimal combination of the pairwise distinct emission packages. This ensemble of emission profile analyses is evaluated by the particle filter and resampling method to assign a weight to each emission profile according to the fit of the resulting volcanic ash with the observations. Emission profiles are replaced if their corresponding volcanic ash content does not fit well to the observations (lower panel, cf. Sect. 2.3)." Further, we agree that the naming convention of some variable is misleading. We have added a table describing all variables. Also, we have changed the subscripts for referring to ensemble members to superscripts. Thus, ensemble member i is now referred by $(\cdot)^{(i)}$. We are confident that these modifications meet your expectations.

- 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.

  Thank you for your comment on the uncertainty. We are aiming to make our point clearer. Our approach accounts for the analysis uncertainty in different ways. First of all, the minimization is performed for an ensemble of emission profiles, where each ensemble member uses perturbed observations and different a priori emission profiles. Further, the ensemble of emission profile analysis is valued by the particle filter algorithm, which assigns weights to each ensemble member and replaces statistically valueless emission profiles, i. e. emission profiles with too little weights. In this way, the ensemble members are comparable in explaining the observed volcanic ash content. The probability is the relative number of ensemble members that simulate volcanic ash concentrations (or column mass loadings) above a threshold. We have added this information to the particle filter section 2.3: "Qualitatively, the strategy of particle filtering applied here can be expressed as follows: By replacing the valueless ensemble members (i. e. those with too little weight) each ensemble member has comparable skill to match the observations. Hence, the probability of an event (e. g. volcanic ash concentrations above a certain threshold) can directly be extracted from the relative number of ensemble members that simulate this event."

  Further, Stohl et al. (2011) in their Fig. 3 give the relative reduction of the assumed a priori uncertainty of the ash emissions by the inversion algorithm, with assumptions made and the results further detailed in their section 3.2. We are sorry that we did not appreciate these findings in a pertinent way, what is now made up for. We wrote: "Stohl et al. (2011) include errors from a priori estimates, retrieval errors and model errors and discussed results in terms of relative error reduction subject to assumptions made." We also add some remarks on our error estimates with particle filter approach. We like to point out that our approach uses the uncertainty estimation to provide reasonable error simulations of the volcanic ash cloud, which has the potential to identify areas with high volcanic ash content without direct observations.

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.
  We have added some further information on the advantages and disadvantages of the Nelder-Mead algorithm to Sect. 2.2. The section now reads: "The minimization problem posed by (1) is solved using the Nelder-Mead algorithm (Nelder and Mead, 1965). The Nelder-Mead minimization algorithm is a combinatorial optimization method without constraints and without the need to compute the function derivatives. It has proven to be robust, especially in cases where the function to be minimized has discontinuities or the function values are noisy (see McKinnon, 1998). This is expected to be likely in highly variable volcanic eruptions especially given highly uncertainty, and thus noisy, observations. Additionally, the Nelder-Mead algorithm can easily account for bounded regions, in our case positive semi-definite ash loads, and needs relatively few function evaluations (mostly 1-2 function evaluations per iteration, Lagarias et al., 1998).
  The idea of the algorithm is to move a simplex on the surface of the cost function to find an improved model state in a N-dimensional space. The version of the Nelder-Mead method used in this study follows Gao and Han (2012) and utilizes adaptive parameters controlling the step size for each iteration of the minimization. The version has been implemented for parallel operation (Klein and Neira, 2014; Lee and Wiswall, 2007). In our application the Nelder-Mead algorithm is used to find the optimal combination of the pairwise distinct emission packages. Hence, a factor a_i, which needs to be scaled by the algorithm, is assigned to each emission package.
  Due to its simplicity the Nelder-Mead algorithm is easy to implement but it is likely to find a local rather than the global minimum of the cost function (which is also a problem for least-square minimization techniques with poor initial guesses, as for volcanic eruptions). Thus, we have added some adjustments to the algorithm. First, we perform the minimization only for integers (including 0). Thus, only integer values are accepted for the scaling factors a_i of the emission packages. By applying this constraint it is assumed that the introduced errors are of lower order than the error introduced by the temporal resolution of the emission packages. Further, the minimization is restarted with larger perturbations of the vertices once the optimization fails for many iterations. Finally, the minimization is started for an ensemble of Nelder-Mead analysis. As perturbed observations are used as input to the minimization procedure, the solutions (here emission profiles) produced by the analysis ensemble are assumed to map the uncertainty given by the observations onto the emission rates. Thus, the minimization

algorithm is called hereafter discrete-grid ensemble Nelder-Mead method (DENM)."

- "The minimization was performed in N_0^N, which has been found to be more effective than the minimization in R^N" –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?
  Thank you for stating this point to be not clear. We allow only integers as solutions for the minimization. Tests showed that the minimization is less trapped in local minima and the convergence to the solution is faster. We have made this point clearer as can be seen in the response to the previous point.
- How is the initial simplex determined for each ensemble member? You say later it is arbitrary but some more details here would be useful.
  Indeed, we start the minimization from an arbitrarily chosen initial simplex (emission profiles in our case). We have tested the algorithm starting from an emission profile with an umbrella shaped vertical mass distribution that varies temporally in strength and plume height. For the given "true" emission profile in our study, we found better performance using the arbitrary initial simplex. It is likely that this is only due to the chosen true emission profile in the nature run. In an application to real volcanic eruptions, we will again test the arbitrary initial emission profile against the time-varying umbrella shaped emission profile. In our algorithm, the initial simplex can be freely chosen, which allows to adapt the method to the characteristics of the current volcanic eruption and its assumed degree of uncertainty. We added this information to section 2.4 where we introduce the arbitrary initial emission profile (we have omitted to use the term simplex to avoid further confusion about the terminology): "The algorithm was tested using a time-varying initial emission profile with umbrella-shaped vertical mass distribution. Due to the chosen true emission profile in this idealized study (cf. Sect. 3) the minimization using the initial emission profile with umbrella-shaped vertical mass distribution shows larger errors. In the application of the algorithm to a real volcanic eruption the performance of the analysis using umbrella-shaped initial emission profile may exceed the performance using arbitrary emission profile. Hence, ESIAS-chem is designed to adjust the initial emission profile to the characteristics of the current volcanic eruption."
- It would also be nice to see some comment on computational effort (i.e. run time) for this system.
  As the core focus in this study is the reconstructability of the 3D ash field based on wind shear driven sequences of 2D column field imagery, numerical efficiency was not our primary concern. The run time of the ensemble system is an informative value about the applicability as early warning system. However, as for other methods in the literature, we have decided to not concentrate on the computational performance. Thus, we have adapted the simulations to the available compute resources (especially granted wall clock time). We run the ensemble of emission packages subdivided into chunks of 60 members. Further, we increased the number of iterations in the Nelder-Mead minimization to 15,000 (including restarts), which is not feasible in a realistic early warning scenario. However, we chose 15,000 iterations in order to track the performance of the minimization. It was found that the costs reached the minimum value after ~1,000 iterations. With this setup, the run time of the system is not competitive with other algorithms nor representative for a realistic application.

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….
  We apologize to not have been clear on this point. We mean that the particle filter formulations do not need Gaussian error statistics or unbiased model states as it is necessary in other data assimilation approaches (although generally the ensemble needs to have a large enough spread such that the solution is within the spread). We have added this information to the manuscript: “It is noted that in the particle filter method no assumptions of the statistical forecast error characteristics and the observation error were made (the errors do not need to be normally distributed and the model state does need to be unbiased as other data assimilation methods require).”

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.
We apologize for the unclear description of the methodology. We have carefully revisited section 2 and changed the wording such that the distinction between the ensemble of emission packages and the analysis ensemble is clearer. We have added some more description to the text: „The ensemble statistics can now be computed using the ensemble member weights. For example, the ensemble mean is

$$\overline{x}=\sum_{i=1}^{N_{ns}} w^{(i)} x^{(i)}$$”

Further, we have added: “In ESIAS-chem, the particle filtering and resampling steps are applied after the ensemble of optimal emission profiles has been found by the DENM algorithm. A weight $w^{(i)}$ is assigned to each optimal emission profile. Residual resampling (Liu and Chen 1998) is used to replace emission profiles leading to too small weights by emission profiles with high weights (this step includes perturbing duplicated emission profiles). After resampling, the weights are normalized again ($w^{(i)}=1/N_{ens}$). Thus, the statistical informative value of the analysis ensemble is preserved.
Qualitatively, the strategy of particle filtering applied here can be expressed as follows: By replacing the valueless ensemble members of the analysis (i. e. those with too little weight) each ensemble member has comparable skill to match the observations. Hence, the probability of an event (e. g. volcanic ash concentrations above a certain threshold) can directly be extracted from the relative number of ensemble members that simulate this event.”

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.
  Yes, it is a possible extension, which is tested in an upcoming real volcanic eruption investigation. We have changed the sequence as follows: “ESIAS-chem is constructed such that it is applicable to other accidentally released matter and constituents, given constraining observations are available. Further, it is capable to be coupled with ensembles of meteorological fields to account for additional uncertainties resulting from meteorological forecasts. However, this idealized investigation focuses on the ability of the system to reconstruct the emission

profile and its uncertainty under perfect meteorological conditions. Thus, no meteorological ensemble is used."

- 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)?
  Thank you for pointing out that we were not clear in this point. The ensemble of emission packages is input for the Nelder-Mead minimization, which constructs an ensemble of emission profile analyses. These analyses are weighted by the particle filter algorithm. We have added the information on how to calculate the ensemble mean and the probability from the ensemble in Section 2.3 (see also our answer above). Further, we have clarified the input and output of the Nelder-Mead algorithm. To distinguish between the ensemble of emission packages and the ensemble of emission profile analyses (meaning analysis ensemble) we explicitly use these terms instead of only "ensemble" in section 2. In section 3, only the analysis ensemble is referred to. We have carefully revised the manuscript for other occurrences of the word "ensemble", which may cause confusion and changed the terminology accordingly.

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.
  We have added another sentence illustrating the identical twin principle more clearly: "The term identical twin refers to the fact, that observations and a priori knowledge are constructed from the same model and input data, in which only the parameters to be optimized (emission profile in our case) differ."
- L190: Is EURAD-IM an online or offline model? Please clarify and which meteorological data are used as driver/lateral boundary conditions.
  Thank you for this note. Indeed, we have forgotten to give the information about the meteorological model. We have added the following information: "As we consider the differences of feedbacks of the ash clouds on the meteorological evolution as not critical on the forecast time scale in our idealized tests the EURAD-IM is offline coupled with the Weather Research and Forecasting (WRF) model version 3.7 (Skamarock et al., 2008). Meteorological boundary conditions are taken from the ECMWF (European Centre for Medium-Range Weather Forecasts) analysis."
- 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/B978-008-100405-0.00017-3. (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.

  Thank you for providing further literature. We have surveyed the literature and changed our statement: “The uncertainty of volcanic ash column mass loading observations is about 40% (Western et al., 2015; Clarisse and Prata, 2016) or even higher (Wen and Rose, 1994; Kylling et al., 2014).”

- 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..

  We have added this comment to Eq. 7: “For applications to real volcanic eruptions, the observation error provided by the satellite retrieval per pixel should be considered.”

- 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 shear.” 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.

  We agree. The wording is insufficient to illustrate our point. The second sentence now reads: „With increased residence time in the atmosphere the volcanic ash at different heights becomes more horizontally split by wind shear. This effect can be exploited by increasing the assimilation window.”

- 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.

  As was suggested by reviewer 3, we have shifted the definition of the analysis metrics to Section 2. Thus, there is no more jumping between pcc and RMAE and we hope this clarifies the matter.

- 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?

  We chose a low threshold to investigate the error of the full volcanic ash cloud, not only the highest ash concentrations. However, the results are the same when choosing, for example, 200 µg/m3 as threshold, although the number of grid cell exceeding this threshold is low, especially in the later hours of the investigation. We have added this information to the text: “The relatively low threshold to calculate the RMAE was chosen in order to increase the number of grid cells to be analyzed and to investigate the full volcanic ash cloud rather than only the area of high concentrations.”

- 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.
  Thank you for the suggestion. We have added this result to the abstract: „For our test cases using an artificial volcanic eruption, we found an assimilation window length of 18 hours, i. e. 10 hours after the eruption terminated, to be sufficient for analyzing the extent and location of the artificial ash cloud.“
- 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.

  Thank you for pointing us to the literature. In our study, we state that we need observations at least 10 hours (or later) after the eruption has terminated in order to sufficiently estimate the volcanic ash cloud. The literature provided by you moreover states, that once an improved model state of the volcanic ash is found, the improvement rests for 6 to 24 hours. With our experiment, even though it is an idealized study, we show that the improved model state prevails until the end of the forecast time. However, we have not focused on the duration of the improvement in the model state due to the assimilation. All this makes it difficult to compare our results with the findings by Fu et al. Thus, we suggest to rather not compare with these publications.

- 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.
  You are right, the perfect model and perfect meteorology assumption contributed to the good performance of ESIAS-chem for the comparison of the vertical distribution of the volcanic ash. This is what we will investigate more clearly when applying ESIAS-chem to real volcanic eruptions, where meteorological forecast uncertainties impose a limiting factor to further improvements. However, the good estimate of the emission profile for the case study with strong wind shear hints that at least for strong wind conditions the vertical distribution of volcanic ash can

be sufficiently analyzed under realistic conditions. We have added the following comment to our statement: “However, this good performance in analyzing the vertical structure of the volcanic ash cloud is partly due to the perfect model / perfect meteorology assumption made in this study. The reliable estimate of the emission profile for the test case with strong wind shear suggest that the vertical structure of the volcanic ash is also sufficiently estimated under real conditions, where meteorological forecast uncertainties impose a limiting factor to further improvements. This needs to be proved in the application to real volcanic eruptions.”

- 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.
  We have chosen 24 hours instead of 18 hours because it is a realistic choice for operational applications. Indeed, we could have shown the 18 hours assimilation window case with equal results.
- 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.
  We understand your preference. It is certainly an option to concentrate first on the estimated emission profile for one assimilation window and see how it performs compared to other assimilation windows. Our intention was to first compare the results for the different assimilation windows before concentrating on one realistic choice (as mentioned before, we could have concentrated on the 18 hour assimilation window case, too). As a compromise, we have added the emission profile for the analysis ensemble mean for all assimilation window lengths in Appendix A. Also, we have added an investigation on the error distribution for estimating the emission rates (measured by a relative emission factor between the model and the nature run emissions) for all assimilation window lengths. Both is provided as supplement to this author response.
- 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.
  As mentioned above, we have added this information in Appendix A.
- 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.
  Thank you, we have updated the numbers in the abstract.
- 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.
  Clearly, this is one explanation for the results. However, we believe that our conclusion is correct. As you can see in our Fig. 12, right hand side, the upper and lower part of the volcanic ash cloud is horizontally displaced, which suggests that column mass loading observations should be useful to estimate the emission

profile. The problem here is the long residence time of the volcanic ash cloud above Iceland, which enables the mixing of volcanic ash emitted at different times at the lower part of the emission column. This prohibits a better performance of the system when increasing the assimilation window length as the volcanic ash cannot be attributed to the correct emission package once the volcanic ash has mixed.

- 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 mean.”. 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.
  The weak wind shear case, in which the increasing thickness of the nascent ash cannot help to analyze the height-time resolved emission profile, is a typical case of an ill-conditioned inversion problem. Any built-up sequence (upper level first-lower last, or vice versa) in a stagnant column can comply with later observations off the volcano, after inception of wind. The vertical profile of the eruption sequence remains beyond analysability with any inversion method, given only column thickness data. Additional observations, e. g. the Keflavik radar or similar height resolving observation systems, are required to further constraint the volcanic ash emissions. Yet, this is not part of this study.
- 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.
  You are right, the fact that the ensemble predicts an event by 90 % probability does not show its accuracy. As we need a large number of analyses to give reasonable estimates of the accuracy of the method and given the lack of computational time for such a large investigation, we decided to remove this sentence from the manuscript.

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.
  We have revised the conclusions. Now, Section 4 is “discussion and conclusion”. As you suggested, we have added some bullet points to summarize our main findings.

Technical corrections

- Line 33: “form radar” – change to “from radar”
  Done
- Line 285: “extend” – change to “extent”
  Done