We thank Dr. Mingari for the constructive comments and suggestions, which helped us a lot to improve the manuscript.

We decided to re-run our simulations due to different comments in all three reviews. This leads to slightly different results, but does not change the main arguments of the paper. Please find our answers to your comments below (Reviewer's comments in **bold**, our replies in standard font, and modifications of the text in blue).

- **l. 51, p. 2: "They described the gravitational spreading of the umbrella cloud by the model of Costa et al. (2013). Collini et al. (2013) highlighted a good agreement in ash transport simulations with satellite observations for the Cordon Caulle eruption 2011 by a combined WRF/ARW-FALL3D forecast system." These introductory sentences seems out of context. Please better contextualize this discussion.**

  We rephrased this paragraph and removed details that are not relevant for our paper (L. 49-53):

  Marti et al. (2017) overcame this issue by coupling the MMB-MONARCH-ASH transport model (Nonhydrostatic Multiscale Model on the B-grid – Multiscale Online Nonhydrostatic AtmospheRe CHemistry model – ASH) with the 1D plume model FPlume, which calculates the MER and the mass distribution in the column online. Another example is the study by Collini et al. (2013), who combined the WRF/ARW forecast system with FALL-3D and highlighted a good agreement in ash transport simulations with satellite observations for the Cordon Caulle eruption 2011.

- **Eq. (1): The expressions for the vertical profile seem to be wrong in Marti et al. (2017) and also here. The Suzuki distribution should be normalized by the integral of S, instead of the maximum S. For discrete point sources, you should normalize using the sum of S requiring: E = \sum S***

  We realized that the description here was misleading. We clarified it in the manuscript and added the following lines (L. 130-133):

  Eq. 1 explains the shape of the emission profile used here, which is also plotted in Fig. A1 in comparison with other profiles.To ensure the correct total ash mass emission and units when the particles are released into ICON-ART at discrete point sources in each model layer between the bottom and top height of the plume, we further normalized Eq. 1 by the integral of S*(z) (Rieger et al., 2015).

- **It seems that you distinguish between the terms "eruption phase" and "puff" in some parts of the manuscript, while sometimes are used as synonyms. In order to avoid confusion, it would probably be more convenient to unify the terminology and use only "eruption phase". In my opinion, "puff" is a bit ambiguous for a complex multi-phase eruption like Raikoke.**

  We replaced all ‚puffs' by ‚eruption phase' and additionally included the following explanation in L. 78-79:

  Throughout the paper, we define 'eruption phase' as one distinct time period in which the volcano was erupting.

- **What do you mean by "insensitive" here? Variations of 10% in column heights didn't affect MER estimations? Please clarify.**

  We added 'vent conditions' here (L. 160):

  The resulting MERs are insensitive to the input vent conditions (temperature, velocity, volatile fraction) in the range of 10%.

- **Figure 2 (title): "MER of very fine ash (<30 um)" → It should be "<32 um", right?**

  We corrected the title.

- **l. 217, p. 9: "The method assesses predefined objects based on a threshold value" You don't say how these objects were defined. What threshold value have you used?**

  We used a threshold of 0.2 g/m² and 2.5 g/m² for ash and SO_2, respectively. We added the following paragraph (L. 262-265:):

  To define objects in the SAL analysis, we used a threshold of 0.2 g m-2 for modeled and observed ash, because this is the detection threshold for the Himawari-8 ash retrievals. For SO_2, a threshold of 2.5 g m-2 for model and observations is used to remove background SO_2 concentrations in Himawari-8 data. This was necessary, because we did not initialize the model with realistic background conditions and, therefore, can only compare the observed and modeled SO_2 plume from the eruption.

- **l. 230, p. 9: Where do these gaps come from? The mean averaging you applied to fill gaps conserves the total mass? Or are you adding new mass with this approach?**

The raw Himawari-8 data column loadings are dense and without gaps in the native format. The mass evolution in Fig. 4 is calculated based directly on this native format (no mass changes). However, for the SAL analysis we needed to map the data on a regular lat-lon grid. The re-projected pixels are adjacent near the centre of the scan, pixels get larger towards the edges (latitudes>30°N and similarly for longitudes ±30° from the subsatellite point) and the separation between adjacent pixels also increases. In the region of the Raikoke volcano the scan angles and adjacent pixels are ~2 km apart (the nominal H8 resolution for IR). This leads to gaps on the lat-lon grid. There may be some mass changes but actually we only retrieve mass loading (not mass) so there are some implicit assumptions about the area of pixels etc which are more important. We added to the manuscript (L. 258-260):

These gaps in the satellite data arise during mapping from the native format onto a regular lat-lon grid as needed for the SAL analysis and are due to the increasing pixel sizes towards the edges of the retrieval domain.

- **l. 232, p. 9: Regridding mass loading to a coarser grid by a linear interpolation is not the best approach, as mass conservation is not ensured. Do you have an idea how much is the total mass difference induced by the interpolation method?**

We agree that linear interpolation is not the best approach. However, it does not makes much difference in the mass loading though as the field is smooth and the error is therefore negligible. An example is shown in the following figure for ash column loading in g/m² for two time averages (top: no interpolation; bottom: with interpolation).

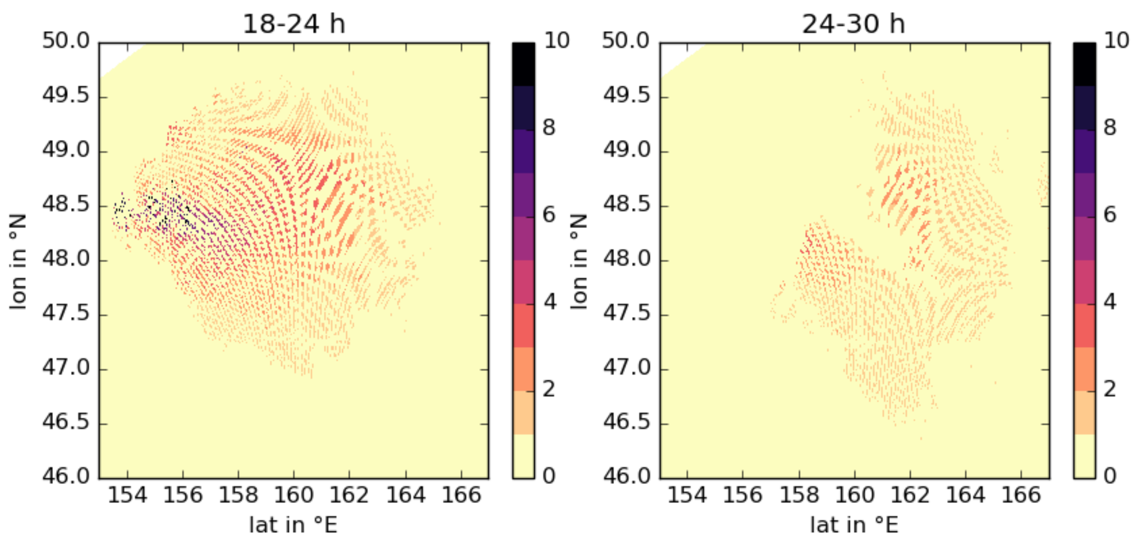

Fig. 1: Himawari ash 6-h- averages for 18-24 h (left) and 24-30 h (right) after June 21, 12 UTC without filling the gaps in the plume that arise due to mapping on a regular latitude-longitude grid.

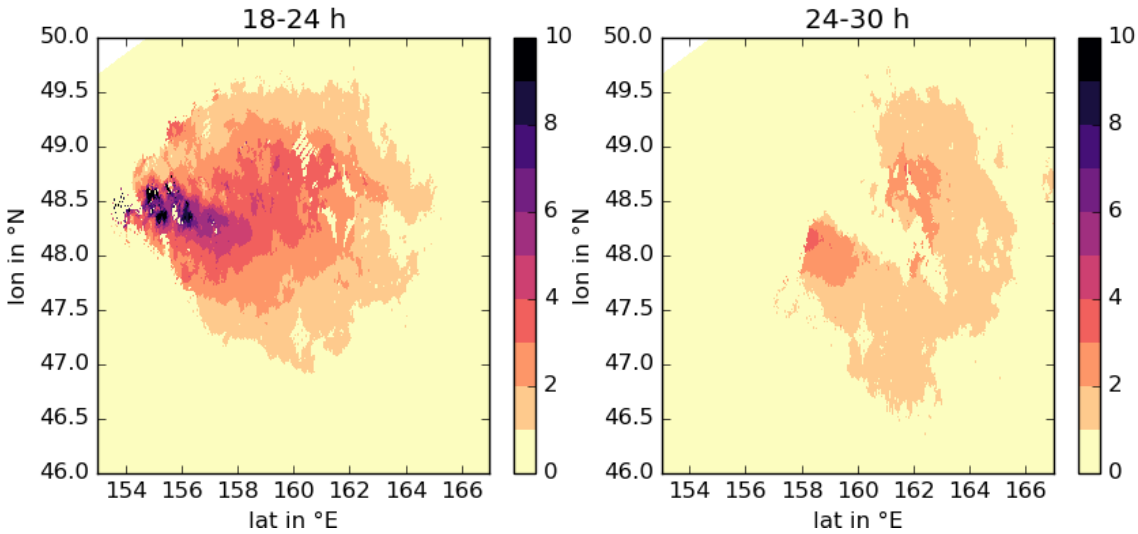

Fig. 2: Himawari ash 6-h- averages for 18-24 h (left) and 24-30 h (right) after June 21, 12 UTC with filling the gaps in the plume that arise due to mapping on a regular latitude-longitude grid.

- **l. 260, p. 10: "Thus, we conclude that the online treatment of plume development improves the ash loading prediction during the first hours and days after the eruption."This statement is not correct. It cannot be concluded from the results presented in Section 3.1 that the online treatment improves simulations. In fact, FPlume-norad outperforms Mastin-rad, and you cannot say that FPlume-rad is better than FPlume-norad. I think the only valid conclusion here is: "the FPlume experiments (ie, FPlume-rad and FPlume-norad) showed better agreement with observations"**

  With online treatment, we refer to the experiments where MER are calculated within the simulation depending on the atmospheric state. Thus, the conclusion “that the online treatment of plume development improves the ash loading prediction during the first hours and days after the eruption” is valid. We clarified in L. 194-196:

  The experiments FPlume-rad and FPlume-norad calculate the ESPs online within the simulation, whereas in Mastin-rad the ESPs are derived offline independent of the atmosphere and vent conditions.

- **l. 264, p. 11: SAL requires defining model and observation objects. For instance, you can identify clusters of grid cells with mass loading exceeding a given threshold. However, you haven't mentioned what threshold was used. This threshold can be defined based on the detection limit of the satellite retrievals. For example, Prata et al. (2021) assumed 0.2 g/m2 for volcanic ash. Otherwise, model-observation**

comparisons wouldn't be fair. Please clarify what thresholds you have assumed for ash and SO2.

Done (see answer to comment above).

- **l. 282, p. 12: Results in Fig. A2 are really good. I think it would be worth including Fig. A2 in the main body of the paper, probably replacing Fig. 3.**

  We produce these two figures (3 and A2) for different reasons: Fig. 3 should give the reader a short overview on the location and the dispersion of the cloud without showing too many details. Figure A2 should give supporting information to understand the values of the SAL analysis and is much too large for the main text. We therefore leave the order of the figures as it is.

- **l. 295, p. 12: Why are you defining two threshold for volcanic ash instead of using a single threshold for total ash? Why didn't you define a threshold for the giant mode? What threshold have you used for ash in Fig. 6?**

  We included a subsection (Sect. 2.2.1) in the Methods part describing the choice and need of the threshold values.

- **l. 296, p. 12: I see no reason to remove those 'steps'. Why top height is greater than zero before the eruption starting time in Fig. 6? This has to do with the smoothing? Since data was smoothed, you should at least indicate the vertical resolution of the model. Is it comparable to the differences found in top height between rad and no-rad experiments?**

  We attached the figure without the smoothing and with the 'steps' here, but we decided to keep the figure with the smoothed lines in the manuscript. The differences in the plume top height between the smoothed data and not-smoothed data are only small. We agree that the top height >0 km before the eruption start, which results from the smoothing, is misleading. We therefore adjusted the strength of smoothing and added the following sentence in L. 335-336:

  The difference in height between FPlume-rad and FPlume-norad remains, regardless of the use of this filter. However, the increasing plume height already starting before the beginning of the eruption is a result of the filtering.

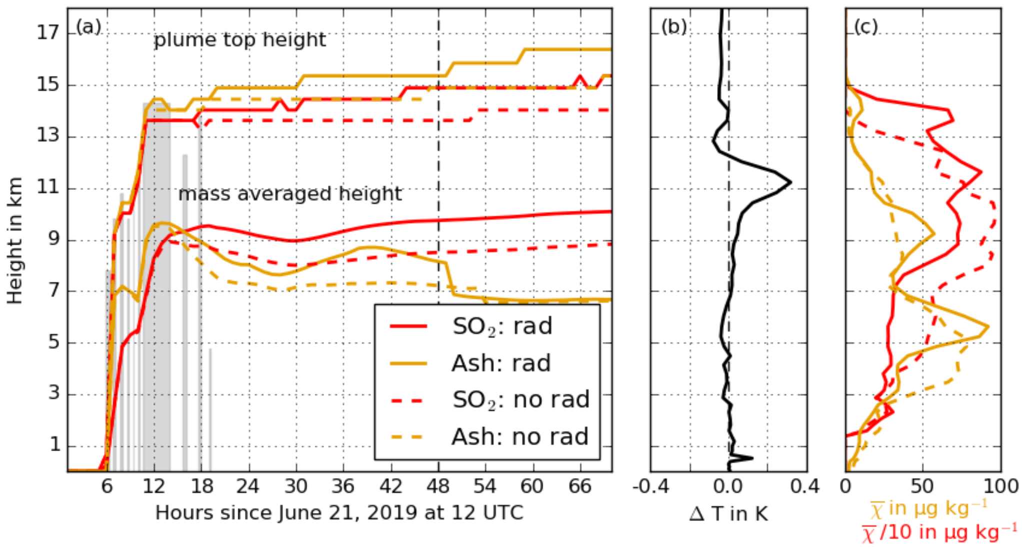

Fig. 3: a) Temporal evolution of the SO2 (red) and ash (yellow) plume top height and mass-averaged height for the FPlume-rad (solid) and FPlume-norad (dashed) experiment. The grey bars indicate the duration and height of the 10 individual eruption phases; b): vertical profile of the temperature difference between FPlume-rad and FPlume-norad in the ash plume 48 h after the start of the simulation. c): vertical SO2 (red) and ash (yellow) profile averaged over the plume after 48 h for the FPlume-rad (solid) and FPlume-norad (dashed) experiment.

- **Figure 6: How is the mass averaged height computed? Is a vertical average? In this case, the average is weighted by mass concentration or by mixing ratios? Or is a horizontal average? In this case, the average is weighted by mass loading (in g/m2). Are vertical profiles also averages? Or are they computed at specific locations? How this average is performed? Do exclude grid cells without ash/SO2 from the average?**

  We included the following information (L. 329-332):

  The plume top height in a) is defined as the maximum height of all grid cells in the plume that were separated from background mixing ratios as explained in Sect. 2.2.1. The average plume height in a) is the mean height weighted by the mass of all grid cells considered as inside the plume. The values in b) and c) were horizontally averaged over the whole detected plume, again excluding grid cells outside the plume.

- **l. 306, p. 13: "in the FPlume-norad case still shows the overlap of the different phase dependent profiles" Where is shown this overlap? Please clarify.**

We removed this sentence, as it does not make sense here anymore.

- **l. 318, p. 14: "the vertical distribution of the total ash mass"**
  What do you mean by "total mass"? Is the total mass within a model layer?

  We used a wrong title in the plot and also here in the text (see next comment). We corrected this to mass concentration in kg/m3.

- **Figure 7(c): "SO2 mass loading in kg" -> mass loading should be in units of g/m2 as in Fig. 3. Why are you showing "mass" in (a) and "mass loading" in (c)? What is the difference?**

  Here, the title was wrong. In the first manuscript, both plots showed mass loading in g/m2 for the individual grid cells. However, we changed the quantities to concentrations in kg/m3 as this is independent of the height of the grid cell and easier to understand. Nevertheless, the argumentation in the text is not affected by this.

- **Equation (3): Obtaining the median radius for a multimodal log-normal distribution is not a trivial problem. Are you sure the median radius is given by such a simple formula? Or this expression only defines a "characteristic radius"?**

  We agree with the reviewer and used the word ‚characteristic radius‘ instead throughout the text.

- **Figure A1(left) is not relevant.**

  We provided this figure not only to explain our profile, but also for others as a reference and to reproduce our settings without further complications. Thus, we leave this figure as it is.

Technical corrections and minor comments:

- **l. 126, p. 5: Remove parentheses from Marti et al. (2017)**

  Done.

- **l. 130, p. 5: Correct citing format in Rose and Durant (2009)**

  Done.

- **l. 143, p. 5: It is necessary a reference for the Mastin empirical relationship**

Done.

- **l. 155, p. 6: Specify: "ellipsoid" -> "Earth ellipsoid"**

  Done.

- **l. 166, p. 6: What do you mean by "marine Sc layer"?**

  Done.

- **l. 174, p. 7: "The Raikoke eruption 2019" -> "The 2019 Raikoke eruption"**

  Done.

- **l. 204, p. 8: "Himawari-8 Ash and SO2" -> "Himawari-8 Ash and SO2 retrievals"**

  Done.

- **l. 212, p. 8: "The ash retrievals were corrected (...) completely cloud covered". Please clarify this sentence.**

  We extended the sentence like this (L. 137):

  The ash retrievals were corrected by a mask that accounts pixels that contain meteorological clouds but which were classified as completely cloud covered.

- **l. 330, p. 15: "Fig. 7 also shows" -> "Fig. 7d also shows"**

  Done.

- **l. 331, p. 15: "the radius is higher on average because" ->"the radius is higher on average according to the FPlume-rad experiment because"**

  Done.

- **Figure 7 (caption): "and SO2 mass (b)" -> "and SO2 mass (c)".**

  Done.