We thank Dr. Folch 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).

General Comments

- **The FPLUME model actually outputs MER and vertical distribution of mass from plume height. However, it seems that, in the online ICON-FPLUME coupling, only MER is considered whereas a Suzuki parameterisation (with fixed parameter values of 4 and 5) is imposed for the vertical distribution of mass. Is there a particular reason for this? To derive both ESPs (i.e. MER and profile) from FPLUME would seem a much more consistent approach. Note that, in addition, FPLUME gives size-resolved vertical profiles, whereas the authors assume the same values for the Suzuki parameters (4 and 5 according to eq.1) for all particle/aerosol bins.**

  In our work, we only used MER from FPlume and vertically distributed the mass by a Suzuki distribution instead of using the vertical distribution of mass from FPlume. Although, we agree that using the FPlume profiles for mass would be more consistent, we decided on our approach due to two main reasons: 1) based on offline analysis we figured out that the mass profiles for the predefined bin sizes strongly depends on the assumption of initial total grain size distribution (TGSD). As information on the TGSD is often lacking, using FPlume mass profiles lead to a less generic approach and large uncertainties, 2) the definition of ash modes in ICON-ART is only realistic for ash dispersion in the atmosphere and differs from TGSD at the vent. Thus, we would have to convert the FPlume size bins into ICON-ART mode which makes it complicated to control the emission into ICON-ART.

  We would agree that the FPlume profiles might be the ideal solution to study size distributions close to the eruption in test cases. However, for operational purposes where quick responses are needed, considering the FPLUME profiles might not be optimal.

- **Vent conditions (Sec.2.3.1). As noted in the text, FPLUME is sensitive to vent conditions, particularly in the case of low plumes. Insufficient momentum (low exit velocity) and/or insufficiency energy at inlet can**

yield, together with other atmospheric factors, to lack of model convergence. This is not necessarily a model shortcoming, but actually reflects configurations in which a plume could not be sustained and therefore collapses. Please check if this could be the case for columns below 10 km, particularly if exit velocities drop below 80-90m/s (this seems to be the case according to values from Table 1). Rather than using Mastin-derived MERs (see next comment), I suggest re-running FPLUME for the Raikoke phases 01, 02, 04, and 10 with higher values of exit velocity (see if this converges FPLUME). On the other hand, it is also true that FPLUME is a model for sustained plumes. Application to short-lived burst-like transient puffs may certainly fail.

Thanks for this comment. We re-run our simulations with higher exit velocities and were able to simulate also the smaller eruption phases online. We note that these modifications affect only the small phases that lead to a negligible effect (<10%) on total mass.

- **I do not see the need for the Mastin-derived modelling option in this study. Which is the purpose of it? First of all, with respect to FPLUME-rad only affects the strength of the source (given that same Suzuki is used in both cases). It should, therefore, imply only a scaling of the Amplitude component of the SAL metric. Unfortunately, comparison (e.g. in Figure 5) is not given to check this. If this is modelling option not relevant to the paper it could be removed. Results and conclusions would be unaltered. The paper could then focus on: i) comparison with previous off-line (Muser 2020) and, ii) comparison rad/norad. This would be simpler and easier to follow.**

  We do not show the SAL values for exactly the reason that you mentioned here: The difference in the SAL analysis between FPlume-rad and Mastin-rad is mainly the different Amplitude value for ash. However, we want to show the Mastin-rad experiment here as well because this parametrization is commonly used in volcanic dispersion forecasts. We argue that online treatment of plume dynamics improves the mass of ash in the model. We highlighted our findings in L. 287-288:

  Thus, neglecting meteorological effects and other plume-related processes in the case of the Raikoke eruption (offline treatment), as it is often done in volcanic dispersion forecasts, results in higher MER especially in the long continuous phase of the eruption and subsequently increased ash emissions into ICON-ART (Fig. 2).

- **ICON model configuration (section 2.4). Is there any particular reason for running ICON globally? Can the model be run only over a limited area? If not, could you comment on the grid approach over the area of**

interest or, if none, is the resolution of the R3B07 configuration uniform across the globe

ICON model can be used in limited area mode (for regional modeling) or with nests (for zooming into particular areas and see scale interactions). We performed global simulation, because 1) it is the same ICON configuration used at German Weather Service for global weather forecast, thus leading to seamless and consistent initialization; 2) it was comparable with the configurations used by Muser et al (2020). Nevertheless, the icosahedral grid of ICON ensures a uniform resolution across the globe (exceptions are nest, but then a refinement of the grid at specific locations is wanted), because the globe is divided in triangles of equal size.
We added the following sentence in the manuscript (L. 184):

The global icosahedral grid of ICON ensures a uniform resolution across the globe.

- **Figure 5. Why it only shows results for FPLUME-Rad? Why the other FPLUME.norad and Mastin-Rad are never shown? On the other hand, can you comment on the poor Structure-component results for ash? Any particular reason for doing 6-hourly model output averages? Considering that Himawari-8 observations can be available 4-times hourly, other approaches having less impact on the results could be considered.**

In this figure we focus on validation of the plume dispersion. We do not show the Mastin-Rad case in the SAL plot, because this experiment only impacts the Amplitude as already answered above. The differences in the SAL values for FPlume-rad and FPlume-norad are only small and a SAL comparison between both experiments does not give additional insights. The impact of radiation on vertical distribution of the plume is discussed in section 3.3.

Our SAL analysis uses modeled and observed data at every full hour as we wrote the output for the 3D-fields only every hour. Therefore, we also used the Himawari data at every full hour. The reason for the 6h averages is mainly that we can get rid of the gaps in the gridded Himawari data. These gaps arise when mapping the data from the native format. Another reason is that we can better compare the results with the maps in Fig. A3.

- **Averaging over such long intervals smooths out (e.g. peaks) and has substantial impact on instantaneous model results. How is SAL affected by this? Can this mask disagreements?**

Averaging smooths out peaks in both observational data and model data. In our analysis, we do not compare long term averages with instantaneous model results, so the SAL analysis does not compare ‘smoothed’ observed data with ‘raw’ modeled data. We attached the results with averaging over 4 and 5 hours for ash here as a comparison (stars: 6h; dots: 4 / 5 h). In the manuscript we still show the SAL analysis with 6 hour averages, because it is easier to compare to Fig. 4 (plotted grid lines agree with dots in SAL plot).

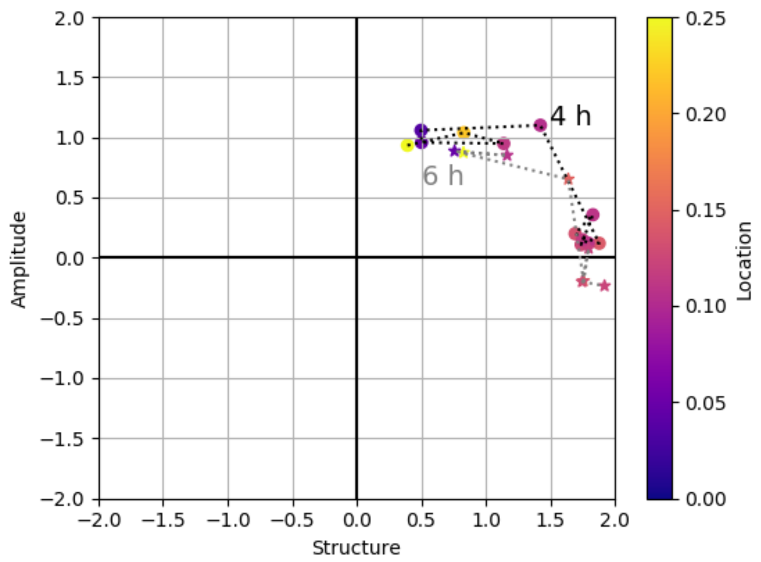

Fig. 1: Ash SAL values for 4-h-averages (black curve with colored dots) and 6-h-averages (grey curve with colored stars).

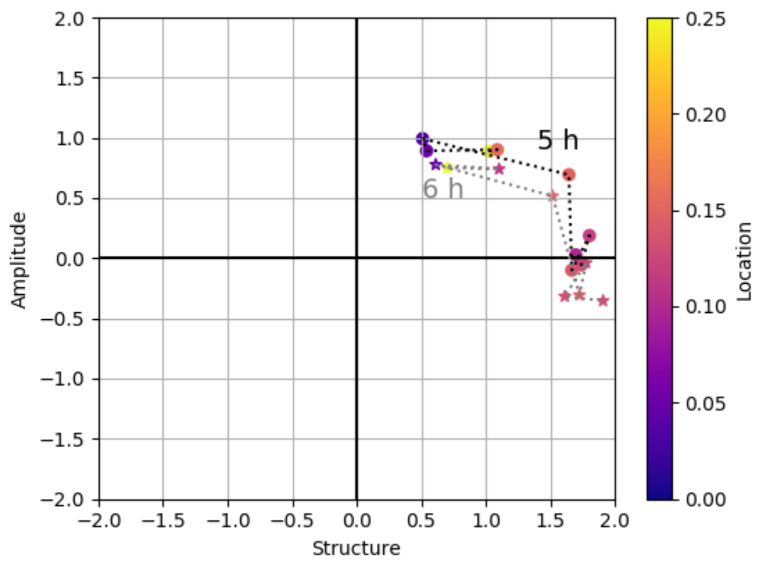

Fig. 2: Ash SAL values for 5-h-averages (black curve with colored dots) and 6-h-averages (grey curve with colored stars).

Specific Comments

- **L36 (and throughout the text). Should multiple reference citations be ordered chronologically?**

  We corrected the citations.

- **L103. Could you specify the size ranges for each aerosol mode?**

  We added to section 2.4. (L. 216-218)

  The three insoluble modes are emitted as lognormal distributions with median diameters of 0.8, 2.98, and 11.35 µm, respectively. The standard deviation is 1.4 for each mode.

- **L166. Sc layer?**

  Changed to ‘stratocumulus layer‘ (L. 177).

- **Table 1. Please specify that heights are a.s.l. (and not above vent as required by FPLUME) and that emission rate of SO2 is computed using (2). Also, the emission rate of ash is not reported.**

  We specified heights as above sea level and the emission rate of SO2 is computed using Eq. 2. However, this table only shows the input values that are fixed for the individual phases. The emission rate is not fixed within the phases and is therefore shown in Fig. 2.

- **Figure 2. Why is FPLUME (red dots) given in all phases?**

  The values refer to both the FPlume-rad and FPlume-norad scenario. We changed the label to FPlume-rad / FPlume-norad to clarify this.

- **Figure 3. Is the color scale adequate? Should the higher value be at around 2 gm-2 or similar?**

  We double-checked the plotting script, the color scale is correct. The highest values are selected based on the values appearing in the plots.