Response to R2

R2 general comments

The manuscript presents applications and evaluation of the Eulerian transport model FALL3D-v8, alongside a companion paper (Part 1), which presents the model physics and some limited verification. The applications are of SO2 and volcanic ash transport using a data insertion scheme, evaluated using satellite imagery; the transport and deposition of volcanic ash evaluated using tephra samples; and the transport and deposition of radionuclides evaluated using deposition measurements. This manuscript pertains to 'model evaluation', using the language of the journal. Quoting GMD's scope, they suggest that "where evaluation is very extensive, a separate paper focussed solely on this aspect may be submitted. . .typically, this comprises a comparison of the performance of different model configurations or parameterisations." In the case where the manuscript contains "substantial conclusions about geoscience rather than about models, and such papers are not suitable for submission to GMD." In its current form, in my opinion, the manuscript does not sufficiently evaluate the model to fit the scope of publication in GMD. Very little of the manuscript focuses on model evaluation – i.e. the model's ability to reproduce real world physics – with too much focus on the data used to evaluate. The paper has, however, passed the access review stage, suggesting that the topical editor has deemed the manuscript acceptable for GMD's scope. Therefore below I provide my suggestions for revisions to improve the manuscript, followed by technical comments.

We thank R2 for their review of our paper. Please find specific responses below to each point raised.

1. The manuscript does not sufficiently evaluate the model physics. There are many ways that this could be done in harmony to the companion paper. For example, one evaluation could be through simulation of an ash cloud using the emissions terms (1)-(4) in Sect. 3.2.3 detailed in the companion paper. An additional important evaluation case is the effect of including the fourth-order Runge-Kutta scheme in the solving scheme. The superiority of the new aggregation scheme in v8 over that used in v7 should also be demonstrated.

Most model physics parameterizations are inherited from previous versions and have been discussed and evaluated in several other papers (including sensitivity analyses). Moreover the fourth-order Runge-Kutta scheme was already evaluated in Part 1 (see Sect. 4 of the companion paper). In terms of model physics in the current manuscript, we evaluate the model's ability to simulate SO_2, long-range transported fine ash (using HAT and SUZUKI options), radionuclides and ash deposition. However, this is not the case of the new multi-class aggregation scheme, which we plan to consider in future studies.

2. The manuscript should include an example using emissions term (5) from Sect. 3.2.3 in the companion paper (i.e. resuspension). Desert dust would be a sensible choice if the authors wish to move the model away from being purely volcanological. This will better demonstrate dispersion from within the boundary layer.

Thank you for these suggestions. However, these emission schemes derived from mineral dust are already in v7.x and have been tested for volcanic ash resuspension events. See, for example:

- Mingari, L. A., Collini, E. A., Folch, A., Báez, W., Bustos, E., Osores, M. S., Reckziegel, F., Alexander, P., and Viramonte, J. G.: Numerical simulations of windblown dust over complex terrain: the Fiambalá Basin episode in June 2015, Atmospheric Chemistry and Physics, 17, 6759–6778, https://doi.org/10.5194/acp-17-6759-2017, 2017.
- Folch, A., Mingari, L., Osores, M. S., and Collini, E.: Modeling volcanic ash resuspension - application to the 14-18 October 2011 outbreak episode in Central Patagonia, Argentina, Nat. Hazards Earth Syst. Sci., 14, 119–133, https://doi.org/10.5194/nhess-14-119-2014, 2014.

3. A huge section of the paper is taken up by description of the satellite detection algorithms. This level of detail should not appear in the main text, which should focus on the model. No reasoning is given for using a bespoke satellite detection and retrieval algorithm here. A previously published algorithm should be used using an available data source to improve transparency. For example from SACS (https://sacs.aeronomie.be/) or some similar openly available source. This point is emphasised by the manuscript stating that (Line 321) 'it should be noted that the retrievals presented here are preliminary and require further cross-validation with other satellite retrievals'.

We agree with this reviewer comment. In response, we have removed almost entirely the satellite detection and retrieval sections from the main body of the manuscript. We now provide them as Appendices instead. We have added justification for using the previously published ash (Prata and Prata, 2012) and SO_2 algorithms (Prata et al., 2003) in new subsections in our revision. The relevant subsections are Sects. 3.1.1 and 3.2.1 of the revised manuscript. As is emphasised in our revision (and in our specific responses below), we note that SACS only provides polar-orbing SO_2 retrievals which do not fit our desired method of validation (high temporal resolution and large scale spatial coverage).

4. The choice of the 1986 Chernobyl accident seems an odd choice given the relative improvement in measurements during the Fukushima-Daiichi accident. This would also allow the authors to demonstrate the decay scheme for Strontium-90.

We agree with the reviewer that the Fukushima-Daiichi accident is a good choice for testing the model. We hope that the Fukushima case will be the subject of a future work. Here we decided to focus on the Chernobyl case since in that case the radionuclides were dispersed on a continental-scale area and a lot of measurements are available both at short and long distances.

5. It is unclear why there is so much emphasis on data insertion. The paper generally reads as justification for using data insertion, which has already been shown in Wilkins et al. (2016). Either the volcanic ash or SO2 example should be dropped as a single example shows that the model is capable of data insertion.

It was not our intention for this paper to be purely about data insertion. However, we concede that in its original form it could appear that way. To address this we have completely restructured the manuscript around the new FALL3D-8.0 test suite (as suggested by RC1). The new revised manuscript more concisely explains the reasons for selecting each validation case study (see Sections 3.1, 3.2, 3.3 and 3.4 of the revised manuscript) and how the various new features of the FALL3D-8.0 model have been tested. We also note that we have included several advancements based on the conclusions from the Wilkins et al. (2016) study, which include adding a source term to data insertion simulations (i.e. Puyehue-Cordon Caulle example), the first example of SO_2 data insertion (Raikoke example) and using CALIPSO lidar measurements to constrain the vertical distribution of the ash/SO_2 clouds.

R2 technical comments

Abstract: Acronyms (i.e. SAL/FMS) should not be defined in the abstract.
There is nothing wrong with defining an acronym in the abstract if it is referred to again within the abstract. Since we quote SAL and FMS scores in our revised abstract we have left the acronym definitions in.

Line 16: Change to ‘15+ year track record’
Done.

Section 2.1.: If keeping, a brief description of SEVIRI is needed.
Done. See Sect. 3.1.1 of the revised manuscript.

Section 2.1.1.: I would strongly urge the authors to use an ‘off-the-shelf’ product, but if keeping then it needs to be made clear that this is a bespoke algorithm relevant to this test case.
As far as we are aware, there are no “off-the-shelf” SO_2 retrieval products for geostationary satellites. Moreover, the method we are using has been published before and the previously published papers are already referred to in the text - so it is not a “bespoke algorithm” as the reviewer states. We have moved the specific assumptions used for our implementation of the 7.3 micron retrieval to the Appendix based on suggestions from RC1 and RC2.

Eq 1: The subscript ‘ash’ should not be italic and it would be clearer is Twc was replaced with -0.5K in the equation
We have changed subscript ‘ash’ so that it is not italic. However, we disagree with the reviewer’s suggestion to replace the threshold definitions with constant values. Defining the thresholds as variables makes it much simpler to refer back to them in the text and in figures. For example, we quote the threshold settings in Figs. A1 and B2 (of the revised manuscript). It also makes it more straightforward to refer back to them in future studies if different threshold settings are used. In the end this comes down to style/preference. We also note that RC1 had no comment on these definitions.

Eq 3: Place -2K in the equation
See above comment.

Section 2.1.2, Eq 5: What geometric thickness of the cloud is assumed here? Is this the same thickness that is used in the insertion scheme?
This is an interesting question. We actually do not specify the geometric thickness of the ash cloud in the radiative transfer calculations as we are retrieving the total column mass. Eq. 5 (of the original manuscript) shows the relationship between the geometric thickness and the optical depth; however, in the radiative transfer modelling, we set the (total column) optical depth directly (from 0 to 9.9 in 0.1 increments) rather than compute it from the geometric thickness.

For the data insertion, we use the geometric thickness based on the CALIOP observation shown in Figure 4b of the original manuscript.

Eq 8: Put -2.5 K in equation
See above comment.

Line 172: Specify what you mean by ‘meteorological clouds’, i.e. water and ice clouds.
Done.

Line 178: “As mentioned above”, specify the section.
We have re-worded this sentence (as the old subsections have been merged). See Appendix B of the revised manuscript:
“Considering that ΔT_{SO2} calculated via Eq. (B1) is a function of the total column density of SO_2, we can retrieve the total column amount by constructing this function from offline radiative transfer calculations.”

Line 186: Which gases?
H_2O, CO_2, O_3, N_2O, CO and CH_4. These have been listed in the revised manuscript (see Appendix B).

Line 187: What ‘amounts’ are you referring to?
The total SO_2 column densities (in Dobson Units). This has been clarified in the revised text (see Appendix B).

Line 203: Is this vertical distribution also the same slab as used in the satellite retrieval? The Puyehue-Cordón Caulle eruption was known to have complex multi-layered cloud structures. How has this been dealt with in the satellite retrieval and insertion?
As mentioned above, we do not specify the ash layer thickness in ash retrieval. For the data insertion, we have selected a time when the ash layer was evidently uniform (i.e. almost representing a ‘slab’ in the vertical) in the CALIOP backscatter data (see Fig. A4b of the revised manuscript). Note that the vertical distribution of the ash layer only needs to be specified at the data insertion time.

Lines 212 and 216: FMS and SAL need defining in their first appearance in the mainbody of text.
Done.

Section 3: These validation metrics are only valid for the data insertion scheme. Please provide metrics for the other test cases. How S, A and L are combined into a single metric needs to be detailed.

How S, A and L are combined was detailed in the original manuscript at line 226-227.

However, to make his absolutely clear we have reworded the sentence to "After identifying objects for both the observation (satellite retrievals) and model fields, we compute the SAL as the sum of the absolute values of S, A and L, which results in an index that varies from 0 (best agreement) to 6 (worst agreement)". In addition, we have added the RMSE (along with its definition) as a quantitative validation metric for the other two case studies (Etna-2013 and Chernobyl-1986). See Sect. 4 of the revised manuscript for the definitions of the validation metrics.

Line 255: Are the 'ash mass loading areas' the areas of the satellite pixels or the meteorological/output resolution of the model? How are the alternate resolutions compared?

The ash mass loading areas are the areas of the model grid boxes. The satellite observations are first regridded (using nearest neighbour resampling) to the model grid. Then the areas where ash is present (above the defined threshold) in both the model and observations can be compared for spatial overlap. This was stated in the original manuscript on lines 200-203: "To insert IR satellite retrievals of volcanic ash and SO_2 (described in Sects. 2.1.2 and 2.2.2) into FALL3D, the satellite retrievals were re-sampled (using nearest neighbour sampling) from their native projection into a regular 0.1 X 0.1° latitude-longitude grid, consistent with the FALL3D grid." And reiterated in lines 227-228: "All comparisons between observations and model simulations are made using a regular 0.1 X 0.1° latitude-longitude grid". In our revised manuscript these details are provided in Sects. 3.1.1 and 3.2.1.

Sections 4.1 and 4.2: These sections are evaluating the data insertion method, which has been evaluated in previous work, rather than the model itself.

These sections evaluate the model's data insertion scheme ***as well as*** the model's emission sources without data insertion (e.g. HAT and SUZUKI options). The SO_2 simulation option (Section 4.2 of the original manuscript) has never been evaluated before so we're not sure what the reviewer is basing this assertion on. We hope that our revised and restructured manuscript has made this point clearer.

Section 4.1.: What is the grain size distribution used in this case? Assuming it is the retrieved effective radius, how many bins are used etc?

The simulation uses 12 particles bins (from 4 mm to 1 micron) assuming a particle size distribution that depends on column height (in this case this peaks at 125 microns). However, for validation against the satellite we only consider the 3 PM10 bins (sizes of 1, 4 and 8 microns) in order to be consistent with the range of effective radii retrieved by the satellite (1-10 microns).

Lines 259-263 and 264-269 can be cut. This information is superfluous to the model
We agree with the reviewer that this information is not relevant to this section of the paper. To address this we have re-organised the paper so that the description of the Puyehue eruption comes earlier on in the paper (Section 3) in a new “Validation cases” section (see Section 3.1 of the revised manuscript).

Line 299: Change ‘meteorological clouds’ to relevant cloud type
Done.

Lines 206-310 can be cut.
As stated above, we have restructured the paper so that the relevant pieces of information come earlier on in the manuscript. This comment suggests the complete removal of the description of the SAL metrics, discussion of previous authors that have used SAL and the description of the thresholds used to compare the model with observational data. As these pieces of information are crucial for understanding the validation, we have decided to retain this information in our revised manuscript.

Line 370: It is not clear to me why ARW was run first. Why was this initial step needed?
ARW was run first to generate the high resolution wind fields required to drive the FALL3D simulations. A more detailed explanation for using ARW can be found in our responses to RC1’s technical corrections, response to Line 377 comment. Details of the ARW model configuration are also included in the revised manuscript Sect. 3.3.2.

Line 393: ‘factor 3 error band’ should just be ‘a factor of 3’.
Done.

Line 399: ‘nuclear accidents’
Done.

Line 414: In contrast to the lengthy explanation of the observations in the other test cases, nowhere in this section does it specify what was actually measured. This section also needs discussion on how general this set up is. For example, the FALL3Dv8 would be unable to be used to model the recent 2017 release of ruthenium-106 in Europe nor iodine-135/xenon-135 during Fukushima.
The referee is correct. As stated in Part 1 (Table 3), FALL3D-8.0 admits 5 different isotopes not including iodine-135/xenon-135. We plan to enlarge this list in future versions, including decay chains to other unstable elements. Regarding observations, high quality deposition measurements of particulates (i.e. radioactive isotopes) were measured and what is used in the present study for validation. We state this and also provide the validation dataset (as part of a GitLab repository) and a description of the model setup for the radionuclides simulations in the revised manuscript (see Sect. 3.4).

Line 416: The supplementary material contains nothing on how it accounts for diffusion, deposition nor decay.

We intended to refer to the radionuclide simulation animations that we provided as Supplementary Material. To avoid confusion we have removed the reference to Suppl. Mat. in the revised manuscript.

Section 5: Many of the conclusions are about the satellite detection scheme and applications of the model, rather than evaluation of the model itself.

Our revised conclusions now summarise the validation results for each case study (rather than specifics about the satellite retrievals). We also summarise next steps in terms of the test suite and future model inter-comparison studies. See response below.

Line 456-459: Model performance has not been discussed anywhere else in the manuscript and therefore does not serve as a conclusion/future work. This should be removed unless performance is explicitly detailed elsewhere in the manuscript.

We agree and have removed these statements. We have revised the concluding remarks so that they are based on the test suite, which is what this paper is about:

"Future developments of the test suite include adding more case studies, model inter-comparison studies that make use of the validation datasets provided here and validation of probabilistic forecasts. In terms of model utilities, we plan to introduce the option of ensemble forecasts and to incorporate data assimilation in future versions of FALL3D."