Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-166-RC1, 2020

Geoscientific Model Development Discussions

Open Access

GMDD

Interactive comment

Interactive comment on "FALL3D-8.0: a computational model for atmospheric transport and deposition of particles, aerosols and radionuclides – Part 2: model applications" *by* Andrew T. Prata et al.

Anonymous Referee #1

Received and published: 1 July 2020

General Comments: The paper provides some examples of FALL3D-8.0 simulations in an attempt to validate the model and demonstrate its capabilities for volcanic eruptions and radionuclide accidents. Whilst these are interesting examples, unfortunately the paper falls far short of its part 1 sister paper's (Folch 2020) claim that it will be "a detailed FALL3D-8.0 model validation for several simulations that are part of the new benchmark suite of the code" and it does not stand up as a validation piece for the FALL3D model. I don't believe it can be accepted in its current form as a Part2 paper.

Particular concerns are that: - There is minimal reference to the FALL3D model code itself, which makes it questionable as to whether the paper is suitable for GMD. - It reads like 4 discrete pieces of work that have been thrown together without a coherent aim for what good validation looks like. - There is no description of the benchmark suite. In fact, the first mention of this is at the end of the paper. - There is no demonstration of the skill of FALL3D-8.0 compared to previous versions of the model - Whilst the title talks about applications, other papers have already been published using FALL3D for most of these applications so there is already far more detailed evidence in the literature for its use in these areas.

I believe the intentions of the authors are well-founded, but this current manuscript falls short. The examples neither prove the capability of the model to replicate ash or SO2 plumes (demonstrated by the poor agreement with time between the model and observations) nor demonstrate that v8 is an improvement over v7. It also doesn't provide any indication of how FALL3D-8.0 compares against other models for well-established datasets, which would be another route to provide validation of the model.

The authors potentially have two choices in my opinion: (1) to turn this into a paper that explores how new satellite retrieval techniques can be used to inform volcanic plume modelling for the Cordon-Caulle and Raikoke cases and submit elsewhere or (2) give this a major rewrite to focus on the role and construction of the benchmark suite, and how the model's performance has improved (assuming that this is shown in results) in v8 compared to v7 (or earlier versions).

Option 2 will require new simulations to be performed and the authors will need to put time aside to enable this.

I appreciate that lack of good data is a real challenge for these sorts of models, but a good model validation paper including a description of the approach to benchmarking would be valuable and so I encourage the authors to consider how this can be achieved.

I have provided some more lengthy feedback below.

Specific Comments:

It is unclear from the abstract what the purpose of the paper is – either a demonstration of the range of its applications, or the validation piece that accompanies the description of the new code in part 1. Based on lines 27-30 it seems to be validation, which is what I would agree this paper should be about, hence it would be better to structure the title and abstract around this. I would recommend removing the text “different application cases and” from Line 1 as a start.

The reference to the description of the eruption “in sec 4.1” on Line 55 is an indication that the structure of the paper could be improved. For the reader it would be preferable to describe this eruption (and the other case-studies) prior to this point if it is being used in section 2, as Section 4 is too late. I would suggest that the parts of Section 4 describing the four case-studies should come after the introduction to set the scene and to explain/justify why these examples have been chosen for the benchmarking and why they are the most pertinent to the aim of validating the model.

Given that the paper is submitted to GMD, it is surprising how many pages are devoted to explaining the derivation of satellite data (5 pages and 8 figures) compared to the actual analysis of model skill (5 pages). This doesn’t seem like a good balance. Particularly as the satellite data is not relevant for two of the case-studies. The detailed description of the retrievals in 2.1.1 and 2.1.2 is not a good fit for this paper, particularly as new unpublished techniques are introduced. It would be preferable for these techniques to be introduced and peer-reviewed in a remote sensing paper/journal and then just referred to in this model validation paper. An alternative would be to move these sections to supplementary material to keep the focus on the model, but this risks these techniques not being properly documented. I appreciate that there is a balance to be struck between needing to explain the data used and the actual validation, but currently the emphasis of the paper is not right. There is also too much focus on validating the data insertion scheme rather than the model. If this is meant to be an example of using a benchmark to test different approaches to the source description (i.e. no-insertion,

insertion, other options), then this should be made much clearer.

In Fig 9, a description is needed of what the different colour plumes are in the right-hand side panels. Using different colour contours for the same threshold in the left and central panels is really quite confusing. I can see what the authors have tried to achieve (i.e. to make them match the colours in the right-hand panel), but using different colours is misleading with respect to the contour scale. I would suggest removing the coloured line contours, as the plume extent is clear, or putting both in grey/black.

For a model validation paper, this is lacking a significant amount of detail about how the runs have been conducted. For the Cordon-Caulle and Raikoke simulations details are only given in the text about the horizontal grid and the data insertion time. Information is missing on the vertical grid, model timesteps and other model parameters, as well as the meteorology used. More critically, information is almost completely absent about the emissions (eruption source parameters) including: the depth of the plume used in the insertion case, the source term that has been used in the no-insertion case, how the continued eruption of the volcano is represented in the insertion case for Cordon-Caulle, species properties, etc etc. This is crucial information to understand how the model has been run and needs to be included, ideally in a new methods section. Reference should also be made to Table 1 in each section.

I'm not really sure what the Cordon Caulle example is trying to demonstrate. The first outputs in Fig 9a and 10a (2011-06-05 15:00 UTC) are a good example of how insertion at one timestep can correct for poor source information, but as the authors state the impact of this has disappeared by 48 hours. After this time neither set of simulations appears to validate well with the satellite data and are actually similar, because presumably they are using the same (although as highlighted above) unspecified source. This example doesn't really prove either (i) the value of insertion for a long eruption or (ii) the ability of the model to represent the plume. I suggest that the author's need to think more carefully about the purpose of this case-study and the reason for including it.

For the Raikoke case, the text should highlight that FALL3D does not have a chemistry scheme and no conversion is occurring. Are any loss processes being accounted for in the simulations? The text in lines 348-356 focuses on poorly referenced speculation as to why the observed SO2 could be increasing, but no mention is given as to whether the change in score is due to the modelled SO2 mass decreasing too rapidly. This needs to be considered in the text, even if it is just to rule it out. The Cordon Caulle and Raikoke case-studies are focussed on the data insertion scheme, which is a minor component of the model. To prove that v8 has introduced any enhancement to the model itself, these examples need to include a comparison for the no-insertion cases with the previous version of FALL3D.

I was pleased to see much more detail about the simulation set-up provided for the Etna case in the text, but this is needed for each case not just this one. The level of detail in lines 375 to 389 highlights just how much information is missing from the Cordon-Caulle description. But, as with other comments, I'd suggest these details would be better in a methods section.

Table 1: what is the difference between "fine ash" and "tephra" in the model? As a minimum this should refer back to Folch et al (2020) Table 3, but it would be much clearer to have this defined in the text.

The Poret 2018 paper is a very detailed study of this Etna eruption using FALL3D and so I am struggling to see what its inclusion in this paper brings. It has already been demonstrated that FALL3D can be applied to this sort of study. This could have been a very good opportunity to compare v8 of the model against a previous study, but this is not done. The main difference appears to be the use of much higher resolution meteorology, which is an input to the model, not the model itself. Is this the reason why the agreement between the model and obs appears (at least at face value from the log graphs in both papers) to be better in this paper? Much more detail on the reason for the differences is needed and to justify the inclusion of this case-study in this paper.

For the Chernobyl case, I am concerned that the authors have tuned the model's settling velocities to unphysical values in order to create a better match between the model and observations, but the text is vague enough to not make this clear. If this is the case, then are the authors suggesting that these are the values that have been implemented as the defaults within FALL3D? This seems unscientific and is just tuning the model to this specific case-study. More detail is needed here.

In Figure 17, an explanation is needed as to what the dashed lines show. In addition, to fit with Fig 15, the dashed lines should really be changed from a factor of 10 to x3. There would then be the basis for a much more meaningful discussion as to why the model performs so differently in these two cases, given that both are for ground deposits, and what the causes of these differences are. This would be much more powerful and useful for the scientific community.

Remove figure 18. Without any reference data to compare it to, it is just a pretty picture.

It is clear from the style that the text in the Chernobyl case study has been written by a different co-author to the rest of the text and some grammar improvements would be useful. As with the previous case-studies this example doesn't provide any real validation of the model, as would be required in a genuine part 2 paper.

The mention of the FALL3D Benchmark Suite in the first line of the Conclusions is a complete surprise. Surely this suite should be the focus of the entire paper in that case? Why is it not the common thread throughout? This would be far more appropriate for a paper for GMD. For example, describing: How does the suite work? How has it been coded up? Does it have Known Good Outputs that model tests are compared against? How many tests are run and does it test the code itself (e.g. some degree of unit tests or bit-comparability for different versions) or just the outputs? Are all the tests run for every commit? What metrics are used to show when the model is falling outside of acceptable performance? etc

As it stands, it's impossible to work out from this paper whether FALL3D-v8 is actually any good.

Technical Corrections:

Line 21: use of the word “aerosols” is incorrect here, as they all appear to be gases from the table in Folch et al (2020)

Line 24: Check/confirm the use of “chemical reactions” here, as Folch (2020) says that there is no chemistry as far as I can tell

Line 110: “new python implementation of the original FORTRAN”. This text seems superfluous here, unless you are also providing access to the new code. A separate paper introducing this retrieval code would streamline this paper and keep it focused on the model.

Line 126: “ERA5 reanalysis data” - Please provide more details on what this and a reference

Line 134: You use CALIPSO here but refer to CALIOP in the figure description. The two need to be consistent, and you need to add an explanation in the text as to what CALIOP is.

Line 190: You need to explain why this is relevant, i.e. that CALIPSO detects aerosol not gas, so the layer is only a proxy for SO2

Within section 2.2.2 and elsewhere you use both Himwari-8 and AHI. This is confusing, please choose one and stick to it throughout. I would suggest that more people are familiar with Himawari than AHI.

Line 213: add “of” to “validation air”

Line 271: It would be helpful to repeat the time used for the data insertion here to explain the 1:1 agreement in Fig 9a.

Lines 304-323: This is all introductory material, so would fit better earlier on and separated from the results

Lines 333-334: It would be helpful to give the scores for the insertion case here for clarity, i.e. I assume the FMS is 1. It is a little biased to compare at the exact time of insertion, a comparison even an hour later when the model has actually had some influence on the insertion case would seem fairer.

Lines 345-5: The simulations indicate that the model is able to track the Himawari observations when initiated with a Himawari source, but other satellites, e.g. TropOMI, show a much larger SO_2 plume on 23/24th, which in this case the data insertion approach does not capture as it is "tuned" to Himawari. This is an interesting question around whether the model should be validated against the same data source that is used for the insertion, it would be useful for the authors to comment on this in the paper.

Line 377: Can you provide some details as to how/if (for the horiz) and why (for vert and horiz) the resolutions used (i.e. resolution of 0.015deg and 60 vertical levels up to 11 km in) differ from the WRF input please.

Line 377: This is the first mention of "samples". What are these? These need to be explained earlier on.

Line 378: Please explain what Phi is, as this won't be known to readers of GMD.

Line 381: Explain what the two parameters at the start of this line are. Do they relate to coarse and fine?

Line 383-4: "and the resulting tephra ground load map is shown in Fig. 14." Resulting from what? The text needs to specify that this is 10hrs after the start time and corresponds to the end of the simulation. There is a lot of information missing here, which should be provided in a methods section, including: - Was the eruption source constant during this period? - What mass eruption rate was used? - Does this period correspond to when the measurements were taken? - Why is this different to the duration of the

Poret study simulation?

Line 393: “all points lie within a factor of 3 error band” - A factor of 2 would be the more conventional choice of statistic here. I suspect that the log scale is rather deceptive for the lack of agreement at the larger values! And it is the larger values that have the bigger impact and so are more important to get correct. It would be good to see this part of the graph on a linear scale if possible.

Line 394: with regard to the use of “10-3 km m-2” It would be helpful to understand the precision of the smallest values from both the obs and the model to understand the potential uncertainty at this end of the scale

Section 4.4: the grammar in this section needs some improvement

Line 399: “accident” should be “accidents”

Line 404: “estimations of such a source term is” – either needs to be “estimation” and “is” or “estimations” and “are”

Line 410-411 and Table 3: I am struggling to understand this. Firstly, the units need to be the same in the text and the table, either m/s or cm/s, for ease of interpretation. But secondly, the values in the table don’t agree with the ranges in the text: 0.0005 to 0.005 m/s for 137Cs - but the table has 0.04 m/s; 0.001 to 0.02 m/s for 131I - but the table has 0.06 m/s; Is there a unit issue here? Or am I missing something?; Some more explanation is needed to make this clear.

Line 416: use of “best case” – what were the other cases?!

Line 454-5: This is too old a paper to make this type of claim for the current breed of models.

Line 456-460: This text is not relevant as the paper is not about model development. Further developments to the benchmark suite would be more appropriate.

Overall, I would suggest that there are too many figures.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-166, 2020.