

Geosci. Model Dev. Discuss., referee comment RC2  
<https://doi.org/10.5194/gmd-2021-378-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2021-378

Anonymous Referee #2

---

Referee comment on "Development and evaluation of the Aerosol Forecast Member in the National Center for Environment Prediction (NCEP)'s Global Ensemble Forecast System (GEFS-Aerosols v1)" by Li Zhang et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-378-RC2>, 2022

---

Review of "Development and Evaluation of the Aerosol Forecast Member in NCEP's Global Ensemble Forecast System (GEFS-Aerosols v1)" by Zhang et al. for publication in Geoscientific Model Development

The paper presents a description of the new GEFS-Aerosols modeling capability that is part of the FV3-based ensemble forecasts of the Global Forecast System (GFS). A number of experiments are performed with this system, and results are explicitly shown evaluating different biomass burning emissions assumptions and impacts of model horizontal resolution. Model results are compared to MODIS and VIIRS observations, AERONET and ATom data, and results from the GEOS-FP, MERRA-2, ICAP, and NGACv2 model-derived products. The model is shown to have considerably better performance relative to its predecessor NGACv2 system when compared to data sets and independent model products. Residual differences in the GEFS-Aerosols performance versus observations and models are speculated at.

The paper is overall well organized and the figures are for the most part clear (I detail some places below where I have suggestions to improve). I recognize here this is a significant update to the modeling capabilities for this major meteorological forecasting system, and I appreciate the progress the authors are making on this work. I nevertheless have a number of concerns about the paper as prepared here that I wish to see addressed before it can be published in a final form. I have many minor suggestions articulated below, but I here will lay out a few more major points.

First, the model description is lacking in some significant respects. In particular, there is no description of loss processes in the aerosol scheme and how they impact the simulation. This is unfortunate because in a number of places it is asserted that uncertainties in wet removal schemes explain differences between the model and observations. A general description of the approach would be helpful here, and it would be useful also to see differences in the large-scale and convective-scale precipitation between

the different resolution runs as a means to explore these differences. More generally, a budget analysis for a new modeling system is a useful add (see e.g., Textor et al. 2006, [www.atmos-chem-phys.net/6/1777/2006/](http://www.atmos-chem-phys.net/6/1777/2006/)) for some inspiration. It is helpful to see how the lifetime of your model is similar to and different than other systems.

Second, the comparisons between the GEFS-Aerosol simulation and the comparison datasets is in most cases only qualitative. There are any number of places where the performance is described as “very good” or “better” than this or that. For the most part these are not very helpful qualifiers, and in some cases I can’t reconcile the assertions with the graphics presented, or at least I don’t know what exactly is being highlighted. Better is something like the presentation in Figures 10 and Table 2, which are at least quantitative (well, semi-quantitative in Figure 10). These provide more objective measures of quality. Please address this in the revisions.

Third, and related, where discrepancies within the comparisons are noted there are appeals to wet removal schemes, plume rise model, dust emissions, and the like. Mostly these assertions are not grounded in anything presented in the paper. A compositional analysis that links underestimates in Europe to Saharan dust emissions (is that really the culprit?) would be helpful. Something similar (sensitivity tests?) to the points about wet removal too. I note a reference below that is relevant, but in particular it is pretty clear that this model suffers somewhat from a common problem in aerosol models with insufficient scavenging of especially black carbon in convective updrafts. Further expansion on this point should be included.

Finally, also noted below, the authors have chosen to evaluate the model performance with a focus on a perturbed period following the June 2019 Raikoke eruption. I note there is no indication of whether the model includes volcanic emissions at all, and Raikoke is evidently not in the simulation. If other pre-COVID periods were available for this evaluation I would prefer that, but at the least I think some acknowledgement of this state would be important to introduce as a caveat, probably most relevant to discussion of high northern latitude biomass burning.

Page 4, Line 24: EMC = Environmental Modeling Center  
([https://www.emc.ncep.noaa.gov/emc\\_new.php](https://www.emc.ncep.noaa.gov/emc_new.php))

Page 4, Line 26: I don’t see it explicitly, but I presume in the GFS-Aerosols member the aerosols are not in any way interactive with the radiation, clouds, etc. Please clarify that’s the case. Also, assuming so, how does GEFS-Aerosols differ from other GEFS members except for the prognostic aerosols? Is it meteorologically equivalent to another member of the ensemble?

Page 5, Line 11: Citations for FV3? I think it has quite a literature.

Page 5, Line 17: I have no context to understand what GFSv15 and GEFSv12 mean. Please clarify.

Page 5, Line 17: Here or somewhere nearby it would be relevant to state the model resolution of your simulations, including also the vertical coordinate. The horizontal resolution is referred to finally in the paper much later, but I don't see the vertical resolution discussed at all.

Page 5, Line 25: Please clarify if you are in fact getting DMS emissions from Lana et al. (2011). If so, that is a departure from GOCART, which uses the DMS seawater concentrations and then determines emissions dynamically based on surface wind speeds. If using DMS direct from Lana et al. (2011), for what year and seasonal variability are you assuming?

Page 5, Line 27: In the abstract you refer to a HRRR-based plume rise model, but here you say WRF-Chem. On page 11 you refer again to HRRR-based model heritage from the WRF-Chem. Please clarify this consistently throughout.

Page 6, Line 2: The GOCART model referred to here with the 5-bin sea salt is in Colarco et al. (2010), doi: 10.1029/2009jd012820

Page 6, Line 6: The "S" and "A" terms are not obviously defined in the text. I cannot find a reference for the FENGSHA scheme here, or at least the Tong et al. 2017 citation is missing in the references. Please state what "S" and "A" are (where they derive from) and add the citation.

Page 7, Line 1: Later in the text wet removal is appealed to in various places to explain the agreement (or lack thereof) with ATom data. I note there is no mention of loss processes and how parameterized in the model. Are the loss processes also in the same sequence as the emissions in GEFS-Aerosols? What is the process order?

Page 7, Lines 12 - 16: This text just reads out of place here as it is descriptive of the GEFS configuration and not the aerosols themselves. This belongs I think in Section 2.1.1.

Page 7, Lines 22-23: "aerosol optical properties from NASA" is not terribly descriptive. If from GEOS/GOCART please cite appropriate sources (e.g., Colarco et al. 2010, Colarco et al. 2014).

Page 8, Line 10: MERRA-2 do not provide forecasts, or anyway not in some form readily accessible. It is a reanalysis and I suspect you are looking at those products, which might just be described as state snapshots or averages.

Page 8, Line 32: The GEOS system I think referred and used here is the near-real time GEOS-Forward Processing (GEOS-FP) system. Suggest that terminology. And my understanding is the "branding" is no longer using "GEOS-5" but simply "GEOS".

Page 10, Line 6: What is the spatial resolution of the CEDS inventory used here? And are you in fact using the earlier CEDS inventory cited here and not more recently available versions that go through 2019?

Page 10, Line 19: What does "GOCART model background fields" refer to? I infer the oxidants. Please clarify. "Does "NASA GEOS/GMI" model refer to the MERRA-2 GMI version (<https://acd-ext.gsfc.nasa.gov/Projects/GEOSCCM/MERRA2GMI/>), or something else?

Page 10, Line 23: I find this description confusing and am not sure what is being described versus shown in Figure 4. GBBEPx is stated to blend emissions from several sources...is that really what it is doing, is blending QFED with other emission sources? QFED I think would not be referred to as "MODIS QFED" like here as it is not a MODIS product, but derived from MODIS observations. Second is a reference to 3BEM emissions which is merged with WF\_ABBA. But Figure 4 calls this "MODIS" which I don't understand. Finally, the plume rise model is mentioned to take input from FRP data. How does this relate to either of the emission products mentioned here?

Page 11, Line 13: It is really hard to read Figure 4, even blown up on a screen, in relation to the comments made about it. I can clearly see more fire spots across the northern latitudes in the GBBPEx emissions, but I cannot tell if the magnitudes are different or not in general because the points are too small to see. It is certainly not evident that emissions are greater in southern Africa (Line 15). My suggestion would be to show a temporal average (a month, a season) to make this point, and you can numerically refer to the relative number of fires observed if you need to.

Page 11, Line 21: What is different in Experiment 1 (prescribed parameters) versus Experiment 3 (real-time FRP data) regarding the plume rise? What are the prescribed parameters?

Page 11, Line 31: For here and elsewhere, when you are showing ICAP MME are you withholding NGACv2 from the ensemble mean, or including it? if the former, do you see a problem in how the clearly biased NGACv2 results shown later might confound the interpretation of comparisons?

Page 12, Line 1: I cannot tell what you mean by saying GEFS-Aerosols are under predicted in eastern Europe. Do you mean Russia at about 60 East?

Page 12, Line 2: It is really not clear how to say one of these models is better than the other. Some numerical statistics need to be presented in terms of biases, correlations. It is also not apparent from a single day comparison that would be the case.

Page 12, Line 23: MERRA-2 "reanalysis"

Page 13, Line 2: Maybe instead of "screening by" something like "...due to the presence of a stable stratiform cloud deck over the ocean that confounds the aerosol retrievals..."

I also want to point out here (and later in relation to Figure 7) that you have chosen an interesting period for analysis owing to the June 22, 2019, eruption of Raikoke in the northwest Pacific, which was a significant perturbation to the high latitude aerosol environment. There is no mention of volcanic emissions in GEFS-Aerosols until the conclusions where it seems like a future extension, so I presume Raikoke is omitted from the analysis. Likely the ICAP models (and for sure MERRA-2) do not explicitly account for the eruption, but they could catch some aspect of it through aerosol data assimilation. This needs to be noted somewhere in the discussion. Look especially at the high latitude MODIS points in Figure 7.

Page 13, Line 22: Please clarify here and elsewhere what we're looking at and how it's done. Figure 8 refers to Day 1 AOD forecast biases. I \*think\* what you are doing is running a 1 day forecast of the aerosol and then resetting the meteorology to the new analysis and making another 1 day forecast, and so on. So you are showing in Figure 8 the ~4 month mean of those 1 day forecast outcomes? How is that compared to the GEOS analyses mentioned here? Are you also looking at GEOS forecast outputs? Or the analysis itself? Are they compatible with what you are doing? Does it matter? Is this just a simple difference of the multi-month means?

Page 14, Line 1: How might you expect emissions to differ in the 2019 simulation years versus the 2014 valid year for the CEDS inventory used here?

Page 14, Line 28: Suggest adding some statistics of the comparisons tabularly in Table 1. It is hard to read the colors in Figure 10 quantitatively.

Page 15, Line 20: I don't see what you are referring to here, and if anything ICAP looks closer to the AERONET points in Figure 11b at the time indicated.

Page 15, Line 32: I'm not sure what is meant by saying GEFS is both comparable to but slightly lower than ICAP.

Page 16, Line 7: Only seven sites are shown.

Figure 14: What is the shading shown on the map in the top left corner?

Page 17, Line 25: There is nothing here that supports the assertion that a low bias over Europe is caused by underestimates in dust emissions. Please explain further.

Page 18, Line 6: There is nothing that supports the statement that under predictions are due to errors in emission or wet removal processes. Or, put differently, this statement doesn't really explain anything in the nature of the figure comparison shown.

Page 18, Line 24: I think I know what "log-Z AGL" means, but please explain.

Page 19, Line 28: I note that Figure 18 does not include any labeling Pacific versus Atlantic. Further, this is a very challenging figure to read without blowing up quite large on the monitor. I suggest you split into one figure for each species to allow more room. Finally, what you call in the text here "bias" is presented in the figure as a ratio. Please use consistent terminology.

Page 20, Line 3: Not sure about "all of the three experiments." I see two experiments.

Page 20, Line 29: How do you quantify "very good" performance?

Page 20, Line 32: I don't follow that the model is able to capture variations in the latitude-height profiles. Figure 18 shows that BC is overestimated in the models by a factor of 10 at higher altitudes. This is by the way a known problem in many models that they do not adequately scavenge BC (see e.g., Wang et al. 2014, doi: 10.1002/2013JD020824 and later).

Page 21, Line 4: Here injection height and scavenging are again appealed to for explanations for differences. What is the impact of the injection height parameterization, and how is that evaluated in this model?