Response to the Interactive Comments During the Discussion Phase

Manuscript Ref: acp-2016-790

Comparing multiple model-derived aerosol optical properties to spatially collocated ground-based and satellite measurements

I. B. Ocko and P. A. Ginoux

We sincerely appreciate the thoughtful reviews of our manuscript, and thank the referees and Editor for their time. The suggestions have undoubtedly and considerably enhanced the manuscript.

Specifically, we have improved the analysis by (i) developing a new section with regional evaluation of model performance with two new instruments and three new figures, (ii) providing quantification of results throughout the text, and (iii) converting AERONET AOD to 550 nm from 440 nm. We have also refined and reformatted the conclusions, and added 12 new references.

Below, we have responded point-by-point to comments and provided information on the modifications in the text.

Responses to Interactive Referee Comment #2 (Anonymous Referee #1):

Comment 1: The manuscript by Ocko and Ginoux presents a comparative study of two versions of the GFDL model, an older (CM2.1) and a newer one (CM3), against optical properties data from AERONET and CALIOP. The manuscript is clearly written and of interest to the science community, especially those using any version of the GFDL model. The analysis focuses on 4 urban locations and 3 sites influenced by significant biomass burning. Those sites, although spread around the globe, are not representative of the global atmosphere, since they represent a very small fraction of the surface of the Earth with exceptionally high pollution levels, at least seasonally. In addition, the coarse model resolution is not capable of resolving the very localized heavy pollution of the urban centers studied, which can lead to spurious conclusions. Although I understand that there is value in comparing a global model with urban data and the authors made a considerable effort to justify that, I firmly believe that the absence of comparisons against places where the model has a chance to give good results is critical in assessing model performance. The apparent incapability of the model to resolve urban pollution also greatly degrades model skill, ending up with a not so flattering model performance, even the newer version of it, despite the great amount of work invested over the years, which resulted in large improvements in the parameterizations since the older version. I do not recommend publication in the present form, at least not until some analysis is included from locations where there is either regional pollution or cleaner conditions.

> **Response:** We understand the referee’s concerns, and have added three new figures (Figs. 3, 4, and 5) and a new section of the paper (Section 4.2.1) to analyze model performance on a regional scale. To add to our existing evaluation of model performance with AERONET and CALIOP data, we use MODIS and MISR data to evaluate model AOD in all regions of the world, and calculate correlation coefficients to provide quantitative assessment. We show that in cleaner regions, both models successfully reproduce AOD magnitudes. In many polluted regions, there is an improvement in AOD from CM2.1 to CM3, but the seasonality performance declines. In addition to regional analysis of overall AOD from each model, we parse out the AOD by aerosols species, in order to better understand model biases. This added analysis provides context and motivation for the rest of our study, where we pinpoint and more deeply evaluate key regions where the models do not perform well. Through analysis of multiple aerosol parameters and spatially collocated instruments, we are able to better characterize model successes and failures. This will provide important information for future model improvements. We thank the referee for the suggestion to include an analysis of regional and cleaner conditions, and in doing so we have considerably enhanced the paper while providing the foundational context for the rest of our analysis.

Comment 2: Section 3.1 (about the older model description) has some very strong assumptions about aerosol modeling. These include the absence of nitrate (6.1), the concentrations (not fluxes) of sea salt that scales with wind speed over the ocean (6.22) (what happens over land?), the zero sea salt over 850hPa (6.23), the offline aerosols coming from different (thus inconsistent) sources (6.31-7.1), the fixed 80% RH for optical calculations which is not even used for BC and OC (7.4-5). I understand that this is an older generation model that is probably not used any more, but in any case with such assumptions the correlations with measurements is expected to be poor. The fact that the new model performance is not greatly better is very surprising. I believe that the authors made the choice of using and presenting that old model to contrast the improvements in the newer model, something very useful for both the users of the GFDL model and its output (so they will look at both model versions) but also for the people that only care about the current model skill (that will look only the newer version comparisons). However, especially for the audience that belongs to the first group, the model performance probably degrades, as presented here (e.g. Figures 4-5, 15.8-9, and 19.16). This comparison though is biased towards the urban stations where the models are not expected to perform well, which is something that even the authors acknowledge (11.29-30). A fair comparison really needs background (not necessarily clean) stations. A great example for this is Oklahoma (10.22-11.6 and figure 4), which is the only urban station captured. This is not a surprise, since the station is not in a city, but downwind of one, and represents regional pollution.

> **Response:** The referee is correct in the assumption of why we compare both CM2.1 and CM3 with observations. To provide the larger context for the basis of our study, and to offer a fair comparison, we have included a new section (Section 4.2.1), new figures (Figs. 3, 4, and 5), and new instruments that look at background regions. While the older generation model (CM2.1) is not used much anymore, and CM3 does indeed improve AOD magnitudes in almost all regions of the globe, there is a decline in seasonal performance from CM2.1 to CM3. By further investigating key regions that are problematic in models, we are able to pinpoint model successes and failures such that future model generations can improve aerosol distributions and optics. Further, we have added text to emphasize the significance of the Oklahoma analysis (in providing a more representative characterization of model performance) as compared with other highly polluted locations (lines 13.5-13.6): *"The site in Oklahoma is in a rural environment compared to the other urban sites we have chosen for model evaluation, and therefore represents areas with background pollution."*

Comment 3: Another argument against comparing with background and even remote stations can be found when comparing the results of Naik et al. (2013), presented in 8.30-32: The global AOD biases are within 5% or 2%, while the differences presented here are significantly larger, and frequently exceed a factor of 2 (section 4.2.1). I understand the motive to accurately capture the very high pollution regions where aerosol-climate interactions maximize, but these are not representative of the global atmosphere and should not be used as a metric of model skill, as is done here.

> **Response:** We agree with the referee that it is important to provide a spatially broader analysis, especially as to not bias the impression of overall model skill based on a selective analysis. To represent the global atmosphere at large, and as discussed in our response to Comment 1, we have added a new section (Section

4.2.1), new figures (Figures 3, 4, and 5), and new instruments to our existing study. We show that overall CM3 improves aerosol AOD magnitudes, but seasonality deteriorates. In unpolluted regions, both models perform well.

Further, we clarify in the text that the purpose of this study is to not adjudicate overall model performance, but rather to use a specific set of tools (multiple aerosol parameters and collocated instruments) to characterize model strengths and weaknesses to aid in future improvements. The modified text reads (lines 2.27-2.31): *"Here we show that comparing multiple model-simulated aerosol properties – from two prominent, related climate models with vastly different aerosol treatments – to available datasets from spatially collocated ground-based and satellite instruments is important for determining model biases. By characterizing model strengths and weaknesses, we are able to provide feedback to improve emission scenarios and aerosol properties for future model generations."* And (lines 3.4-3.7): *"Because the aerosol treatments in the two models are starkly different, as we present in Section 3, comparing multiple optical properties with spatially collocated instruments is especially useful in identifying possible sources of error which are otherwise challenging to determine."*

Comment 4: The discussion is overly qualitative at times, in too many places to be able to enumerate. There are several examples, most of which include wording like "slight", "reasonably", "somewhat", "a better/worse/nice job", "better magnitudes", "fairly well", "correlates well", etc. More quantitative statements need to be used throughout.

Response: We have considerably increased the quantification of our analysis. We have omitted several qualitative statements, supplemented the discussion with correlation coefficients, and also provided correlation coefficients for all model and AERONET comparisons in Figures 7 and 8, as well as for model and MISR/MODIS comparisons in the new Figures 3, 4, and 5.

Comment 5: Specific comments
1.14: please put the names of the models in the abstract.
1.24-27: Longwave aerosol absorption is also an important climate driver.
3.5: : : :treatments IN THE TWO MODELS are: : :
3.11: Delete first instance of word "instruments".
3.11: Describe a bit more the cities, e.g. population, including any other information that might be useful for the reader. Throughout the manuscript there are scattered information, e.g. types of fuels burned in the area, meteorological conditions, etc. This is a good place to have them all together.
4.6-8: BC has an Ångström exponent of 1 across the visible spectrum when externally mixed (see paragraph 112 in Bond et al., 2013), while a spectral dependence is measured for coated BC aerosols. Since BC is homogeneously mixed and not coated in this study, this statement is probably misleading.
4.13: To my knowledge, hardly any model uses interpolations when doing comparisons, primarily because the model uncertainties are probably larger than the concentration

gradient in a grid box. Unless the authors believe the opposite, which would then require to justify why this approach was not followed, I recommend dropping the sentence.
4.25-26: How do you use temporal colocation with CALIOP, which only has day/night profiles at specific times a day? Simply take the level 3 product and compare with the modeled monthly mean? If yes, this is not what colocation means.
5.29: delete extra dot.
7.23: : : : Second, SOME (please say which) aerosol: : :
7.25: Aerosol indirect effects are not considered in this study (5.16-17), so either drop this sentence or remind the reader.
7.27: : : :to be HOMOGENEOUSLY internally mixed: : :
7.23-27: Is there nitrate aerosol in this version of CM3? I know there is from recent publications of the same group, but is it present in this current study?
8.12: "Transportation" –> "Transport".
8.15: "property" –> "properties".
Figure 3: How do you break down the per-component AOD when internal mixing is assumed? This is important information to be in the text, e.g. in 9.21.
9.28: Why the Jaegle et al. (2011) paper is cited? Is this parameterization used in CM3? Please say so, if yes.
Section 4.2.1 is too long. I propose splitting it in two (or three, given my request for background stations), with the second part starting 13.16.
11.7-8: Delete "Upon further investigation".
12.9: Fix typo in punctuation.
12.18: model shows –> models show.
12.30: delete both commas.
13.27: scale –> magnitude.
13.28: capture –> include.
14.3-17: Alta Floresta experienced severe deforestation at the beginning of the dataset used in the manuscript, which later declined significantly. This is probably why the error bars are too large during the dry season: not because of the strong interannual variability, but due to the steep decline of biomass burning in the area over the years. You might want to consider using a shorter period of time from the available long time series, one that is more representative of the simulated period.
14.24: shown –> present.
17.3: I might have missed it, but what is the assumption for the vertical distribution of biomass burning emissions in CM3?
18.26: properly –> accurately.

> **Response:** We thank the referee for the careful and thoughtful review of our manuscript. We have made the requested modifications and clarifications, which have substantially improved the manuscript.