Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-235-RC3, 2018

Geoscientific Model Development Discussions

Interactive comment on "The aerosol-climate model ECHAM6.3-HAM2.3: Aerosol evaluation" *by* Ina Tegen et al.

Miller (Referee)

ron.l.miller@nasa.gov

Received and published: 30 November 2018

This is a well-written article that represents an impressive amount of work to collect a wide range of observations for evaluation of an aerosol model. I have some suggestions for improving clarity, but the article should be suitable for publication subject to minor revision. If the authors have any questions, they can contact me at ron.l.miller@nasa.gov.

1. The article is a detailed and extensive evaluation of the current model. The article could have even more lasting value if it anticipated and aided development of future model versions.

My main suggestion is to add more quantitative metrics to certain figures. These met-

rics can be useful when future generations of the aerosol model are developed and evaluated. For example, Figures 19 and 20 will be of limited use in future development, because the current model assessment is mainly visual and qualitative (i.e. are the points clustered around the 1:1 line?). These figures would benefit from the summary statistics that are listed in Figures 21 or 23 (e.g.), so that similar figures of future model versions can be compared using these statistics. This suggestion also applies to Figures 5, 6 along with 22a and b. Figures 4, 8, 18 and 24 could also be improved with summary metrics. More generally, the authors should think about how they might assess future model versions and augment the figures so that any eventual improvement can be quantified.

The article would also benefit from a discussion in the conclusions about what the authors consider to be key model errors that should guide future development.

Technical Comments (page, line number)

2, 18: add citation to Huneeus et al. ACP 2011?

2, 30: somewhere in this paragraph, it would be helpful to state explicitly that the fields calculated by the chemistry module MOZ are prescribed in this study, so that only the aerosol calculation is fully interactive. (I suggest this distinction because Section 2 describes the capabilities of both the HAM and MOZ modules: a joint description that I like for placing the HAM module in context.)

3, 1: add 'or bin' after 'sectional'?

Section 2.2 ('ECHAM'): it would be helpful if the horizontal and vertical resolution of the model version used in this study was listed here, rather than postponed to Section 3.

Table 1 caption: 'Mode boundaries for the ' missing word?

5, 20: 'parameterization for *ocean* temperature dependence'?

6, 8: 'individual sectors'? Define or give examples of 'sectors' here (instead of in the

following paragraph)?

6, 21: 'T63' This is one reason why the ECHAM resolution would be more helpfully defined in Section 2.2 than Section 3.

6, 27: 'Interannual variability of biomass burning is not considered.' But is the decadal-averaged burning updated each year via interpolation so that it doesn't change abruptly each decade?

7, 13: 'Dust particle emissions are driven by...' How is the ratio of emitted silt particle sizes to clay sizes determined? This is relevant to the coarse-mode burden. which later is found to be unrealistically small.

7, 19: 'Tegen et al. (2002), who identified potential source areas...' It's worth noting that the sources are also identified using a calculation of paleo-lake extent (that I consider innovative).

7, 30: 'These regional correction factors...' What observations and criteria were used to arrive at this correction?

8, 5: 'As a marine source...' What physical variables control emission of DMS? (wind speed? ocean temperature?)

Section 2.3.2 ('Aerosol Microphysics') The omission of nitrate aerosols should be noted, since this could contribute to an underestimate of PM or AOT.

9, 11: 'in-cloud scavenging scheme' Observed scavenging depends upon whether the aerosol is hydrophilic, and this will vary with aerosol species (with sulfates being very hydrophilic and some common dust minerals as hydrophobic). Does droplet activation account for varying aerosol composition within each size mode?

9, 20: '1.52 + 0.0011' It should be noted that this represents relatively little shortwave absorption by dust, in better agreement w/ AERONET retrievals (Sinyuk et al GRL 2003) than the original Patterson et al. (1977) laboratory measurements of far-traveled

Saharan dust samples.

9, 27: Does dust nucleate ice particles by enabling heterogeneous freezing? Does nucleation vary with the aerosol mixture in each size mode, given that some aerosols like dust are much more efficient?

10, 14: 'are relaxed' What is the time scale for relaxation? Does it vary with height?

10, 21: 'do not vary on daily or interannual time scales' But is emission updated each year to reflect slow decadal trends, or is emission held constant over the simulation period?

10, 22: After 'satellite retrievals' insert '(labeled "GFAS")'?

11, 9: 'respective AERONET stations' How many years of measurements are available for the stations? Do some stations have short records that may not be representative of the simulation decade?

11, 14: 'data-assimilation grade product based on Dark Target retrievals' What are the advantges of this product compared to the standard off-the-shelf version of MODIS AOT?

11, 20: 'and with compiled number size distributions...by Heintzenberg et al. (2000)' Maybe move this phrase down as few sentences to where this measurement set is discussed in detail?

12, 7: IMPROVE. it should be noted that some of these sites are at elevation in regions where the topography is not well-resolved by the model. Also, why did you not evaluate dust using these measurements? e.g. see:

VanCuren, R., and T. Cahill, Asian aerosols in North America: Frequency and concentration of fine dust, J. Geophys. Res., 107(D24), 4804, doi:10.1029/2002JD002204, 2002.

13, 5: Why restrict the comparison of MODIS AOT to 2007? Is this year representative

of the entire simulation?

13, 17: 'This may point to missing aerosol sources' Maybe say 'aerosol species' instead of aerosol 'sources'? Also, Bauer et al. (2015) argue that nitrate aerosols and ammonium sulfate represent a significant fraction (a little more than half) of the anthropogenic contribution to PM2.5 over central North America. The omission of these aerosol species from HAM could contribute to the underestimate of AOT.

Bauer, S. E., K. Tsigaridis, and R. Miller (2016), Significant atmospheric aerosol pollution caused by world food cultivation, Geophys. Res. Lett., 43, 5394–5400, doi:10.1002/2016GL068354.

Figure 4 caption: 'Error bars show the variabilities of the measurements.' How is the variability defined? Standard deviation of daily values?

14, 16: 'slightly'? 0.01 seems like a large improvement compared to -0.3.

Figure 8: Could labels be added to the top of each column explaining the quantity plotted (AOT, AE and SSA)?

15, 16: 'underestimating the particle size' Is it known why this discrepancy over the Sahara is largest in the NH Fall?

15, 22: 'the too low particle size ... could result in too high SSA' Why would coarse particles reduce the SSA? Is there a simple explanation or reference for this?

15, 34: 'too strong mixing' vertical or horizontal or both?

Figure 10 caption: replace 'for over' with 'over'?

16, 6: 'may not be representative' This is especially true for stations where topography is poorly resolved.

16, 26: Table 3 includes median values from AeroCom. It should be stated that this median is derived from models and is not necessarily in agreement with observations.

Also, this comparison between HAM2.3 and AeroCom would be more useful is some measure of model diversity was added to the AeroCom column of Table 3.

16, 30: 'However...emission fluxes also depend on the size range considered...' This is a very relevant caveat. To me, it suggests that the comparison of HAM and AeroCom emission be replaced by a comparison of aerosol burden.

16, 34: 'in agreement with the AeroCom average burden of 19.2 Tg' Again, the AeroCom burden is not necessarily indicative of the observed value. As an alternative, Kok et al. 2017 calculate a global burden of around 20 Tg that is better constrained by observations:

Kok, J.F., D.A. Ridley, Q. Zhou, R.L. Miller, C. Zhao, C.L. Heald, D.S. Ward, S. Albani, and K. Haustein, 2017: Smaller desert dust cooling effect estimated from analysis of dust size and abundance. Nature Geosci., 10, no. 4, 274-278, doi:10.1038/ngeo2912.

17, 14: 'R on log scale' Does this mean that R is calculated by correlating logarithms of concentration and not concentration itself? This is fine, but it should be explained that such a correlation emphasizes the ability of a model to correctly simulate variations in concentration over large distances from the source, rather than subtle differences over more limited regions.

Figure 12 caption: 'Simulated vertical profiles are shown as monthly and regional averages.' What is the duration of the flight data sets that are compared to the monthly averaged model concentration in this figure?

Table 3 caption: does 'sedimentation flux' refer to gravitational settling?

19, 8: 'while the CLIM and NUDGE results underestimate BC' The values from these simulations are indeed smaller than the GFAS values, but all seem to be with in the undertainty of the measurements.

20, 15: 'the correlation is better' ... 'R values of 0.78 for the correlation of annual AEs for both the CLIM and NUDGE simulation' If the correlations are the same, how can

one be better?

21, 6: 'my be' (may be?) Why 'may be'? Doesn't Soffiev attribute the temperature dependence to specific physical processes like surface tension and solubility (or is the temperature dependence entirely empirical)?

21, 18: 'spume drops' Could you define spume drops and their relation to the emitted size distribution for the benefit of readers like myself :) who are not specialists in sea-salt aerosols?

page 22: the years of comparison for sea salt are confusing. Line 1 cites the year 2010. But line 5 claims that 2006 is used in Figure 22, whose caption says 2007. In fact, Figure 23 does seem to be based on 2010 according to the caption, but this is not noted in the text (line 15 onward).

22, 22: 'For high latitude stations ...' This generalization doesn't seem to be true for the only NH high latitude station. Also, why are the NUDGE and Gong-T models with better physics behaving so poorly at Cape Grimm (41 s), where temperature effects might be expected to be important? The improved agreement at Marion Island, just 6 degrees poleward, suggests individual stations might include large regional effects that are unique to each island.

23, 21: 'Mineral dust and sea salt aerosol distributions must be well characterized in order to be able to make meaningful statements on the importance of anthropogenic aerosol effects.' I agree with this, but could you justify this point more fully? For climate change, we don't need to characterize natural sources if their radiative forcing is time-independent. To be sure, dust sources expand and contract with time (e.g. due to variations in hydroclimate). Nonetheless, it is hard to evaluate simulated anthropogenic species if the natural species are poorly simulated, because common observed variables like AOT or PM2.5 include the effect of all species together. (Thus, most AeroCom models get reasonable values of AOT despite widely varying fractions of natural and anthropogenic species.)

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-235, 2018.