Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-34-RC1, 2020

Geoscientific Model Development Discussions

Interactive comment on "Global aerosol simulations using NICAM.16 on a 14-km grid spacing for a climate study: Improved and remaining issues relative to a lower-resolution model" *by* Daisuke Goto et al.

Anonymous Referee #1

Received and published: 1 May 2020

This manuscript presents one of the first multi-year simulations with an aerosol–climate model at a resolution approaching the convective "grey zone", in this case using the NICAM atmospheric model with the SPRINTARS aerosol scheme at a grid spacing of 14km. The study compares the meteorological fields, aerosol mass and optical thickness, and aerosol radiative forcing, between this configuration and a lower-resolution version of the same model. They additionally consider the output from the high-resolution model coarsened to the lower resolution to distinguish the effects of simulation vs data resolution. In many respects, this is a therefore a significant and illu-

minating study and very much appropriate for publication in GMD. However, there is one major point as well as a number of minor points which should be addressed before acceptance:

Major points

The main area of concern with this study relates to the ability of a 3-year simulation to adequately capture interannual variability. The very large (and zero-containing) confidence intervals quoted up front in the abstract, e.g. -91% to $+18\%$ and -49% to $+223\%$ suggest that this is simply not a long enough period to adequately constrain what are presented as headline results. If this is to be presented as a major advance in its own right compared to single-year high-resolution simulations, rather than merely a technical step towards a long enough simulation to produce robust interannual statistics, this needs to be clearly quantified and justified in the study. Alternatively, it might be more appropriate to submit results when a longer and more statistically-significant dataset is available, if the conclusions drawn from 3 years do not add appreciable confidence to those from single-year studies.

There are also a very large number of figures in the paper, and I would strongly recommend condensing to a smaller number that illustrate the important results, and moving additional plots to the supplement if they do not contribute to the main conclusions.

Minor points

p.2, line 7: while 56km is a fine resolution compared to most global aerosol–*climate* models, it is coarser than some of those used for operational global aerosol *forecasting*, as collected in the ICAP initiative, for example.

p.2, lines 12–13: what does "column burden of the aerosol wet deposition" mean?

p.2, line 13: should this be "cloud to precipitation" rather than the other way around?

p.2, lines 17–18: it is not clear why "differences. . . between the different horizontal grid spacings are not explained simply by the grid size" follows from the results quoted above as suggested. Is this related to the coarsened-HRM data, which hasn't been mentioned yet at all in the abstract?

p.2, line 25–p.3, line 1: what are the confidence intervals on these ARF values, and therefore are the differences statistically significant?

p.2, lines 4–6: it is commonly found that tuning parameters require quite different values at different model resolutions; therefore any suggestion that tuning using the LRM can be applicable to the HRM or vice versa requires proper justification.

p.8, line 2: does this really mean monodisperse particles of a single fixed size as suggested, or the more usual unimodal size *distribution* with a fixedmode and width?

p.10, line 10: what is meant by "flux of the ARI"?

p.9, lines 6–20: given that a pre-industrial reference is mentioned elsewhere for the radiative-forcing calculations, please state what emissions are used for the pre-industrial case as well as for present-day/2010.

p.9, line 8: are biomass-burning emissions used for specific years in the simulation, or is this 2005–2014 period used to construct a climatology instead? If the latter, please consider the impact on the results (especially comparison to specific observations) of not capturing the real-year interannual variability in fire emissions.

p.9, lines 11–13: what does a conversion from "POM" to "particulate organic aerosols" mean? Why is such a factor necessary, and different for anthropogenic vs

biomass-burning? Or is that actually supposed to be referring to conversion between organic *carbon* (OC) and POM?

p.10, line 4: is it correct that only Terra is used for MODIS AOD, and not Aqua? If so, why? Both are mentioned below for the CERES_EBAF data. Also, are Dark Target or Deep Blue products used, or both?

p.12, lines 11–23: it's quite hard to identify the differences in these plots. Please include some kind of difference or statistical comparison plot (as is done in Fig. 6 for the aerosol fields).

p.15, line 16: please quote the actual MODIS AOT value here.

p.16, line 23: are't these are ground-based, not satellite, observations?

p.16, line 25–p.17, line 3: please consider the behaviour of the "coarsened" HRM data here, which should indicate whether the differences relate to the model resolution itself, or the fact that the observation sites are simply less representative of the coarser grid boxes.

p.22, line 7: why does underestimation of the emission inventory indicate the importance of using finer horizontal resolution? One does not obviously follow from the other.

p.22, line 22: please define what you mean by "secondary" and "tertiary" products.

p.23, line 17: "incredibly" is too strong a word here.

p.23, line 24–p.24, line 1: the description of LRM-macro is rather confusing here. Please introduce both LRM-macro and VLRM-macro properly in the main model-description section instead of suddenly throwing them in during the discussion section.

p.26, line 1: CALIPO ⟶ CALIOP.

p.26, lines 3–23: Figure 17 suggests that the LRM consistently matches the SP2-observed profile shape better than the HRM, especially at higher altitudes (ARCTAS-B being the exception in absolute terms due to a bias although the *shape* is still a better match with the LRM). However, the discussion seems to gloss over this result which goes against the general theme of "HRM performs better". This needs to be identified in the text, and consideration given to how it fits with the rest of the results and conclusions.

p.28, line 20–p.30, line 16: the terminology (ARF, IARF etc.) used throughout this section is non-standard and confusing. Please consider using terminology like RFari, ERFaci etc. consistent with IPCC usage for clarity. Most of the values quoted in this section also need uncertainty estimates or some other method to enable an assessment of whether their differences are significant or not, otherwise they are not that meaningful.

p.29, lines 13–14: the second half of the sentence, "probably because the light-absorption. . . " doesn't make much sense. Please rephrase more clearly.

p.30, lines 22–23: this appear to quote variability of "meteorological fields" in $\mathrm{Tg\,yr^{-1}}$, which doesn't make sense for any usual meteorological quantity. Please check what is actually meant here.

p.31, lines 1–7 etc.: please explain how the "difference caused by meteorological variabilities" is being quantified here as this is not at all clear. I think this is based on some measure of the variation over the three years of simulation, but please clarify explicitly *what* measure of this is being used, and also its associated uncertainty or confidence range given the small number of years.

p.32, line 3: 3 years is a long simulation at such resolution, however it's still very short for any study involving interannual climate variability, as discussed above. Al-

though this is one of the novel aspects of the work, describing it as "very pioneering" in this context is unjustified.

p.33, line 6: does "carbon" here refer to BC, OC, or both together?

p.33, lines 18–19: why should the lack of agreement with observations over China be attributed to an urban-density issue at this resolution that doesn't apply elsewhere, rather than to a possible error in the emissions inventory, for example?

p.34, lines 2–6: this terse list of "variabilities at relevant sites" isn't really appropriate for the presentation of conclusions. Please clarify what the actual conclusion is here instead.

p.34, lines 13–24: this needs to take account of the apparent better perfomance of the LRM with respect to SP2 profiles mentioned above.

p.35, lines 6–7: "negatively large[r]" is rather confusing. Does this mean "strongly/more negative"?

p.35, lines 15–18: again, this conclusion is rather unclear. What distinguishes the "relevant" fields from the "others"? Please rephrase and clarify.

p.35, lines 23–25: it's not clear what this means. What tuning exactly has been carried out here using the LRM for the HRM? Does this tuning have a bearing on the results presented in the manuscript?

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-34, 2020.