

We thank our two reviewers for their helpful comments on our manuscript.

Response to Reviewer 1:

Main comments

The paper's message that sub-grid variability is important and should be accounted for in aerosol model development is well-taken, but also easier to say than do. With their experience of looking into those aspects, the authors must have interesting views on the following questions. Is high resolution required? Line 223 gives an interesting statement in that respect. Do the authors have references or experiences to back up that statement that 10 km is a length scale characteristic of aerosol and CCN distributions? If high resolution is required, does that mean that low resolution simulations should not be trusted? Can low resolution be made to behave like high resolution by imposing subgrid distributions or stochastic parameterisations?

Our results suggest that high resolution may be required in regimes with complex non-linear thermodynamics. The resolution required may be different in different regimes.

There have been a number of previous studies that have quantified the scales of aerosol variability. For example:

- Anderson et al. (2003) used autocorrelation analysis to show that most of the variation in the aerosol properties existed on scales of 40-160 km.
- Weigum et al. (2012) analysed aircraft measurement of black carbon over the remote Pacific ocean and found that BC variability occurs on scales smaller than 80 – 160 km.
- Targino et al. (2005) performed a similar analysis using autocorrelation functions of aircraft data from clean and polluted regions in the free troposphere. They found the spatial scales of variability to be on the order of 10 km.
- Shinozuka and Redemann (2011) compared the horizontal variability of aerosol optical depth during two contrasting phases of the Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS) campaign. In the first phase, which was dominated by local emission sources, AOD demonstrated considerable variability at scales of 20 km, whereas the second phase, which was dominated by long-range transport, showed very little variability at these scales.

While there is considerable variability in these results, one can see that aerosol variability typically exists on scales in the range of tens of kilometres. We chose to run WRF-Chem at 10 km as this should capture the significant scales of aerosol variability while also being able to use a convective parameterisation (in order to make a more realistic comparison to global models, which typically cannot resolve convection).

There have been previous studies that show that a number of techniques employed to account for aerosol sub-grid variability can lead to improved results (e.g. adaptive grid – Garcia-Menendez et al, 2010; plume-in-grid – Karamchandani et al., 2006; PDFs/stochastic grids – Cassiani et al., 2010). To address this point, I have added the following sentences to the final paragraph of the paper:

“One of the major challenges to future modelling is determining how to account for the sub-grid variability of these aerosol processes. Several methods such as adaptive grid techniques and stochastic parameterisations are being developed to target areas where sub-grid variability is significant. This paper increases our understanding of the underlying mechanisms most affected by sub-grid variability and will help guide future development of these methods in order to more accurately predict the aerosol effect on climate.”

Other comments

Lines 134-137 and Figure 2b: it is unfortunate to have chosen to illustrate the effect of inner domain averaging with a variable (ammonia emissions) which does not get averaged in the method. Ammonia surface concentrations would have been a better choice. Having said that, lines 115-120 in the previous section could be interpreted as saying that emissions have also been averaged – but we are now told it is not the case. I guess variables that get passed from module to module are averaged while variables that are only used within one particular module are not. It would be helpful to clearly list in a Table in section 2.1 which variables are averaged in the AA setup.

To alleviate confusion, I have changed the plot to show ammonia surface concentrations rather than emissions. Additionally, I edited line 115 to make it clearer that emissions are treated as a process rather than fields that could be averaged.

As there are a large number of gases included in the gas-phase chemistry model (RADM2), I have added the sentence below to the description of the model configuration, rather than listing the species in a table.

“Gas-phase atmospheric chemistry is based on the Regional Acid Deposition Model, version 2 (RADM2), which includes 21 inorganic and 42 organic chemical species with 158 reactions, of which 21 are photolytic (Stockwell et al, 1990). These species, along with the aerosols described above, are all averaged in the ‘aerosol averaged’ simulations.”

Edit to line 115: “These processes include emission, photolysis, dry deposition, vertical mixing and wet deposition by convective transport, gas-phase chemistry, and aerosol microphysical processes.”

Lines 156: How was the length of the spin-up period chosen? Typically spin-up should be long enough for tracer mass budgets to balance for given boundary conditions. 2 days is probably too short, and I am unclear whether both real and averaged aerosol mass budgets should balance, or only the real one.

The concentrations of the aerosol and gaseous species stabilised after two days in both the averaged and unaveraged simulations. We also tested the results using a 5-day spin up period, and there were no differences in the results.

Lines 177-181: According to previous sections, averaged variables also include “gases”. What are the gases represented in MADE/SORGAM?

See response to first comment.

Lines 207-208: So which aerosol types/modes are PM10 emissions emitted into?

PM10 emissions from the TNO inventory remain unspiciated and are emitted into a generic ‘PM10’ variable. They are coarse mode aerosols.

Line 228: Could we have more details on this coarse-graining procedure?

I added the following statement to Line 228:

We calculated the differences by first coarse-graining the results from the high resolution simulation to the grid of the low resolution run to which it is being compared **by taking the average of the high resolution output residing within each low resolution cell.**

Lines 236-241: For the sake of completeness, a Table summarising the correlations studied, and the resulting correlation coefficients, would be good.

This has been added as Table 3.

Section 3.1.1: The causes for the lack of water uptake in AA80 are well investigated, but there a disconnect between the paragraph discussing ammonia (lines 325-340) and the paragraph discussing vertical profiles (lines 341-354). Should the sentences beginning lines 351 and 359 say that the causal chain begins with underestimated ammonia in the LRHA regime? That conclusion seems partly confirmed by the discussion in section 3.1.2.

Indeed. The underestimation of nitrate in the BL is likely due to the underestimation in ammonia due to the reasons discussed in lines 325-340. While we cannot conclude this for certain based on the vertical profiles, I have added a statement to line 351 to remind readers of this link.

Edit to line 351: "It is therefore this underestimation in BL nitrate, which is likely due to the underestimation in ammonia, that leads to an underestimation in aerosol water content (Figure 6c), and, ultimately, extinction (Figure 6d)."

Table 3: It would be useful to state in the caption that the LRHA regime is not shown.

Added: "Results from LRHA are not shown." to caption of Table 3.

Figure 7: What is the unit of x axis?

The units of the x-axis are the same as the y-axis. I have added this to the caption to clarify

Line 459: From this point, the word "mixing" is used to mean averaging or dilution over a grid-box. I am not sure that it is the best use of the word, because of the risk of confusion with vertical mixing. I recommend using averaging instead.

Good point – I changed 'mixing' to 'averaging'.

Line 516: Again, a Table showing the list of variables tested and the resulting correlations would be useful.

This has been added as Table 5.

Figures 14 and 15: Figure 15 is only used to make a small point, so its contents could be merged into Figure 14. 3

Figure 15 shows results from a distinct run where only aerosols are averaged, instead of both aerosols and gases. As this comparison is different from the comparisons being made in Figure 14, I chose to keep them separate to alleviate confusion.

Technical comments

Line 207: Typo: components

Corrected.

Line 325: There is a full stop missing somewhere in this sentence.

Corrected.

Line 449: Extra word: in some

Corrected.

Caption of Figure 13: Something has gone wrong with square brackets.

Corrected.

Captions of Figures 14 and 15: Why the plural in "FRA10 simulations"? There is only one FRA10 simulation according to Table 1.

Corrected.

Line 559: Remove closing bracket.

Corrected.