**Responses to Reviewer 2 comments on "Intercomparison and evaluation of aerosol microphysical properties among AeroCom global models of a range of complexity" by G.W. Mann et al.,**

Reviewer 2

*This paper discusses the differences in particle size distribution simulated by twelve global aerosol microphysical models when model results are compared with observations.*

We thank the reviewer for their constructive comments and provide responses to each below, with reference to changes in the manuscript that have been made. Reviewer comments are shown in italics and numbered for reference. Excerpts of the manuscript shown in bold-italics.

*1) The paper is quite comprehensive and covers a lot of ground. However, while it shows comparisons with data, it does not quantify well the error of the models versus the data, and errors are obscured by logarithmic plots. As such, I would recommend a better quantification of errors for each plot provided (e.g., provide a statistic).*

We have added, to all Figures that compares model to observations, values of normalized mean bias and Pearson correlation coefficient to give a quantitative measure of the skill of the multi-model mean.

*2) In addition, the discussion of potential causes of model errors is missing. It is not so useful just to show that the models result in errors versus data, it is necessary to examine why this might occur. Along those lines, it is important to clarify the weaknesses and identify what is missing in the models examined for understanding why the models might result in errors.*

We have made clear at the outset of the paper that the purpose of this paper is to assess how well the models as a whole rather than individual models. We therefore assert that attempting to attribute the actual cases of each model bias is outside the scope of this paper. Nevertheless, we have in the paper sought to identify cases where there is a process missing from all or most of the models (e.g. where 7 of the 12 models include only binary homogeneous nucleation) or where there are sources of mass (e.g. secondary organic aerosol) that are likely to be underpredicted in the majority of the models.

For example we refer the referee to the following 3 excerpts from the conclusions which summarize the common sources of bias which we have identified:

**"…there are some important biases common among the models at many of the EUSAAR/GUAN sites. For example there is a strong underprediction of accumulation mode particle concentrations during winter, which is likely due to inadequately constrained particle number sources (both primary and secondary) or underprediction of growth due to a general underprediction of wintertime sources of mass (for example from secondary organic aerosol), or both. The results also show that model Aitken mode concentrations are too high during winter and too low during summer, which may reflect an underprediction of particle growth (to larger sizes) in winter and an underprediction of nucleation events in the summer."**

*And*

**"…The Southern Ocean low bias in total and Aitken particle number concentrations may be due to the models not adequately capturing the observed emission of sea-spray at sub-100nm sizes (e.g. O'Dowd and Smith, 1993; Clarke et~al., 2006; Pierce and Adams, 2006)"**

*And*

**"..Further work to compare the models against size distribution observations at higher temporal resolution is required to better characterise primary and secondary particle**

**sources. Greater understanding of the role of secondary organic aerosol and other components (e.g. nitrate) in affecting nucleation and particle growth in the boundary layer is also required.."**

*3) Finally, the literature survey on aerosol modeling could use improvement.*

We have addressed the specific suggestions for improvement made below.

*Below are specific comments.*

*4) ": : :observed number size distributions in a range of environments showing two distinct peaks, generally found in the 10 to 100nm and 100 to 1000nm dry diameter range (Raes et al., 2000)." There is another peak below 10 nm near roadways due to combustion emissions. Please clarify and cite a reference.*

We have revised the sentence:

***Near to particle sources, the Aitken peak often indicates freshly emitted primary particles.***

To instead say

***Although combustion sources generate particles as small as 10nm dry diameter, these particles rapidly evolve to larger sizes due to coagulation (e.g. Jacobson and Seinfeld, 2004) and global models directly emit the particles in the mid-Aitken size range (e.g. Dentener et al., 2006).***

*5) The authors discuss global models treating sectional schemes. They are missing Jacobson (2001), which was the first model to treat multiple interacting aerosol size distributions each with a sectional representation on the global scale.*

We have changed the existing sentence:

***…and have more recently become established in several global models (Adams and Seinfeld, 2002; Spracklen et~al., 2005a, 2011; Yu and Luo, 2009; Lee and Adams, 2010; Bergman et~al., 2012).***

To instead say

***…and in the 2000s became estab*lished in several global models *(Jacobson, 2001; Adams and Seinfeld, 2002; Spracklen e*t~al., 2005a, 2011; Yu and Luo, *2009; Lee and Adams, 2010; Bergman et~al., 2012).***

*6) The authors state, "Since AR4, many climate modelling centres have incorporated new aerosol modules that include size-resolved aerosol microphysics. This represents a major shift in model sophistication (Ghan and Schwarz, 2007) and improves upon previou "first generation" aerosol schemes in which aerosol optical properties and cloud droplet concentrations were based on the simulated mass of several externally mixed aerosol types, each assigned a prescribed size distribution. The microphysical aerosol schemes calculate and transport the number concentration and component mass in several size classes of particles and can also represent both external and internal mixtures. Separate transport of size-resolved number and mass allows growth processes such as condensation and aqueous sulphate production to realistically conserve particle number while adding mass, and enables new particle formation and coagulation to provide explicit sources and sinks for particle number, which has been shown to be important in capturing changes in aerosol in response to changing emissions (Bellouin et al., 2013)."*

*The authors are stating this as if these aerosol modules or treatments are new. However, this representation has been included in other global models for over a decade (e.g., Jacobson, 2001; Adams and Seinfeld, 2002). Please clarify.*

We have changed the existing text:

***Since AR4, many climate modelling centres have incorporated new aerosol modules that include size-resolved aerosol microphysics. This represents a major shift in model sophistication (Ghan and Schwarz, 2007) and improves upon previous ``first generation'' aerosol schemes in which aerosol optical properties and cloud droplet concentrations were based on the simulated mass of several externally mixed aerosol types, each assigned a prescribed size distribution.***

To instead say

***In recent years, many more modelling centres have incorporated aerosol modules with size-resolved aerosol microphysics into climate models. This represents a major shift in model sophistication (Ghan and Schwarz, 2007), improving upon previous ``first generation'' aerosol schemes in which aerosol optical properties and cloud droplet concentrations tended to be based on the simulated mass of several externally mixed aerosol types, each assigned a prescribed size distribution.***

*7) The authors then state, "The microphysics models explicitly simulate the evolution of the particle size distribution, and use this to determine aerosol optical properties and cloud condensation nuclei concentrations. In so-doing, they represent aerosol interactions with clouds and radiation consistently with the underlying physics of the fundamental aerosol processes."*

*True, however, the authors should clarify that the fundamentals of cloud physics is not treated in these models, as their cloud treatments are largely empirical (e.g., cloud water is often an empirical function of CCN concentration and/or some other parameters), and only one updraft is often assumed to occur in each grid column. The major weakness of these models is the treatment of clouds. As, such, the authors should focus on specifying what these weaknesses are so that readers can understand better the limitations of the models and why results may not always be so accurate.*

*One reason the treatment of clouds in these models is so critical for this paper is because the paper wants to propose multi-model mean size distributions for use in further analysis. However, aerosol size distributions in the presence of clouds are significantly affected by the size distributions of cloud and precipitation particles and the physical interactions of aerosol particles of different size with such cloud and precipitation particles of different size (Jacobson, 2003). As such, it is not realistic to expect a model that treats aerosol sizes explicitly but parameterizes cloud interactions with aerosols to obtain the aerosol size distribution within or below clouds before or after precipitation occurs. Parameterization of cloud treatment in these models may partly explain the model errors versus data seen in Figures 18 and 19 of the present paper, for example.*

We have added the following text to the end of that paragraph (page 30848, line 25)

***We note however that climate model representations of cloud processes tend to be highly parameterized, and characterizing aerosol-cloud interactions in these models continues to be a major challenge.***

*8) Table 1. Please clarify in the text or table whether each model treated one aerosol size distribution only. In other words, was it possible to have two particles of the same size but different composition in any model? Please state so explicitly.*

We have added an extra column to Table 1 to indicate whether or not each scheme has multiple distributions whereby the scheme can resolve two different types of particle (with different composition) at the same size. We also added the following text to the caption:

***The ``Multi-dist'' colu*mn indicates whether the scheme *includes multiple distributions, i.e. whether it is possible to have two particles of the same size but different composition.***

*9) Table 1. Please clarify whether the tracers are all components within aerosol particles of each size, or are some of the tracers present in only some sizes, or are some of the tracers gases?*

The column labeled tracers indicates the total number of transported aerosol tracers (i.e. the sum of the number concentrations and component mass concentrations over all size classes). To make this clearer we have added the following text to the caption for Table 1.

***The "Tracers" colu*mn indicates the total number of transported aerosol tracers for each scheme (the sum of the number concentrations and component masses over all size classes).**

*10) Table 2. Why do none of the models treat the combustion mode of emissions (< 10 nm). Please discuss.*

Combustion sources of particles smaller than 10nm tend to be from point or line sources. The timescales for coagulation of these smallest particles are too short (order of minutes, e.g. Jacobson and Seinfeld, 2004, [JS04]) to be representative of an emissions source on the ~100km gridscale of the global models participating in the intercomparison. We have already added a sentence to the text to clarify that point (citing the JS04 paper) – see response to comment 4) above.

*11) Table 2. What about organic carbon emissions, especially in the nanoparticle range? Figures 1ff. Please define "diversity."*

*Organic carbon emissions from biofuel, fossil-fuel and biomass burning sources are included in the models. Most of the models (but not all) co-emit the organic carbon together with black carbon from these same sources, resolving the emission as a number flux of carbonaceous particles consisting of an internal mixture of BC and OC. The "Primary size" column in Table 2 refers to the assumed size for these primary emitted carbonaceous particles. To clarify this, we have extended the existing text in the caption of Table 2*

***The primary size assumptions correspond to the geometric mean diameter values (nm) for primary carbonaceous particle emissions.***

*To instead say*

***The "Primary size" column refers to the geometric mean diameter values (nm) assumed for primary carbonaceous emissions, which most (but not all) models treat as a source of particles consisting of an internal mixture of BC and OC.***

*12) Figure 3. It would be useful to see how the mean vertical profiles from the models compare with HIPPO data.*

Schwarz et al. (2010) already compared black carbon profiles from the AeroCom phase 1 models against HIPPO-I measurements. We agree that it would be useful to compare the AeroCom phase 2 models against all 3 phases of the HIPPO campaign (for example following a similar analysis to that applied in the Kipling et al., (2013) paper already referred to in the paper. However, we feel that such an analysis is outside the scope of this particular study.

*13) Figures 5. Are the same 8 models used for all observation locations, or are the best 8 models chosen separately for each location?*

*The plot should show the same 8 models used at all locations; otherwise, the results are meaningless since you would not know if even a single model could predict the data well consistently.*

No, as described in section 2.4, the assessment of which models are ``central'' is done locally, so can be a different 8 models at the different observation locations.

We do not agree with the reviewer that this approach means the results are meaningless.

As explained at the outset of the paper, we do not seek to assess the skill of the individual models. Rather we assess the 12 models as a whole, using the central two-thirds mean and diversity to indicate their combined skill against the measurements.

*14) Figure 5 and other figures. What does Dp > 3, 10, 14 nm mean? Does it mean that the data at some stations are > 3 nm and at others, > 14 nm? If so, please say so.*

Yes, that is correct. The different size thresholds (3,10 and 14nm) correspond to the cut-off diameter for the type of CPC used at the different measurement sites. To clarify that we have added the following sentence to the caption for Table 3:

***When comparing to the CPC m*easurements, we derive from the models particle *concentrations larger than 3, 10 and 14 nm*. These *size thresholds correspond to the cut*-off diameters for the different *type of particle counter used in the measurements at each site.***

*15) Figures 6, 7, 8, 9. Are the same 2/3rds of models used each month, or are the best 2/3rds of all models selected a month at a time? I hope the same 2/3rds of models are used each month.*

Again, as we explain in section 2.4, the assessment of which models are central is done locally – in time (for each month) as well as in space (at different sites). So the central two-thirds of models in wintertime may not be the same 8 models as the central models in summertime (for example). We agree that this needs to be borne in mind when assessing the results. However, we do not consider it to be a significant drawback since the majority of the models are included in this central two-thirds.

*16) Figures 10, 11, 12. Are the same 8 models used at all sites?*

No, the same 8 models are not necessarily the same at all sites. Please see the response to points 13 and 15.

*17) Figure 18, 19. The model mean size distributions do not compare well with the data. Please discuss the potential reasons in terms of missing or simplified model processes, that might explain the errors. Right now, in Section 3.2.5, the text only describes some of the errors but no reasons for the errors.*

We do discuss the reasons for the biases in the last paragraph of section 3.2.5 when considering Figure 20 (which summarizes the size distribution comparisons shown in Figures 18 and 19). To make this clearer to the reader, we have added the following sentence near the start of that section (line 14, page 30878 in the ACPD version).

**To summarize these *comparisons, in Fig.~\ref{fig-24} w*e compare the models' simulated *number and size in the Aitken and* accumulation modes to observed *values shown in Heintzenberg et~al.~*(2000), which were derived via *log-normal fits to the size distribution measurements.***

*18) For Figures 18 and 19, please quantify the magnitude of the error (e.g., provide an RMS error over all sizes). Finally, please state, based on the results, whether the models can, therefore, be used reliably to predict aerosol size distribution.*

As described above, we summarize the size distribution comparisons in Figure 20, comparing the Aitken and accumulation mode number and sizes from the models. We have now added to the plots the values for the normalized mean bias and Pearson correlation coefficient for the central two-thirds model mean against the observations. We consider this to be a more

appropriate measure to score the models than the RMS error averaged over all sizes, since it assesses more meaningfully the ability of the models to capture the meridional variation of the size distributions seen in the observations.

*19) Similarly, for other comparison plots, please provide the magnitude of error. Are we talking about factors of < 2 or factors > 10 or something in between. A better quantification of error is needed throughout the paper.*

We have improved the quantification of model error by adding, to all Figures that compare the models to observations, values of the normalized mean bias and Pearson correlation coefficient.

*20) Also, there is little discussion of the causes of error. A section of the paper should be devoted to this. In this discussion, the limitations of the models should be discussed in detail. In the end, readers should be able to tell whether these models are reliable or not and the magnitude of their reliability. Also, what percent of the models are reliable?*

As already explained, and clearly state in the paper, we do not seek to score the individual models here. We therefore consider an assessment of which percentage of models are "reliable" (by some measure) to be outside the scope of the paper. We do however, throughout the manuscript, attempt to identify potential causes of the model-observation discrepancy, in terms of processes or sources missing or biased in the majority of models.

*It would be useful to have a table or figure identifying the location of all the surface sites.*

We have added a new Figure (between the current Figures 4 and 5) showing the locations of each of the comparisons to observations applied in the paper with a different symbol used for the different type of comparison applied. In the revised manuscript we refer to this in the section 3.2 revising the current text:

**The datasets used are listed in Table 3 and are briefly described here**

*To instead be*

**The datasets used are listed in Table 3 and are briefly described below. Their locations are shown on a global map in Figure 5.**

*References:*

*Adams, P. J., and Seinfeld, J. H., Predicting global aerosol size distributions in general circulation models, J. Geophys. Res., 107, 4370, doi:10.1029/2001JD001010, 2002.*
*Jacobson, M. Z., Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols, Nature, 409, 695-697, 2001*
*Jacobson, M. Z., Development of mixed-phase clouds from multiple aerosol size distributions and the effect of the clouds on aerosol removal, J. Geophys. Res., 108 (D8), 4245, doi:10.1029/2002JD002691, 2003*