Atmos. Chem. Phys. Discuss., 10, C10813–C10828, 2010
www.atmos-chem-phys-discuss.net/10/C10813/2010/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Global dust model intercomparison in AeroCom phase I" *by* N. Huneeus et al.

Anonymous Referee #1

Received and published: 7 December 2010

General Comments

This paper presents a synthesis of several important global observational datasets and uses them as benchmarks against which to assess the skill of dust simulations from the 15 global aerosol models which submitted results to phase 1 experiments of the international AEROCOM model intercomparison initiative.

The benchmark observations datasets comprise a wide range of observation suitable for constraining the ability of the models to simulate different aspects of the global dust distribution and include

1) dust/iron deposition measurements from 5 existing global datasets (DIRTMAP sediment trap measurements from Tegen et al (2002), dust deposition measurements from

10 of the 16 sites in Ginoux et al (2001), ice core dust deposition measurements from Mahowald et al (1999), 28 measurements (from stations and cruises) from Mahowald et al (2009) and total/wet deposition of dust from the Florida Atmospheric Mecury Study (FAMS, Prospero et al, 2010).

2) long-term measurements of dust mass concentration from i) the University of Miami network (19 sites from the Pacific (SEAREX campaign, Prospero et al, 1989), and North Atlantic (AEROCE campaign, Arimoto et al, 1995)) with additional 2 long-term observations from sites from Zimbabwe and Australia, and ii) from short-term observations from cruises and long-term observations from additional sites described in Mahowald et al (2009).

3) Aerosol Optical Depth (AOD) and Angstrom Exponent (AE) from 25 dust-dominated stations in the AERONET sun photometer network based on a climatology over the period 1996-2006.

4) AOD and AE from 8 dust-dominated station with data specifically from the year 2000 (the year simulated by the models).

5) Coarse mode AOD from 22 of the 25 dust-dominated stations in 3) over the priod 1996-2006

6) Coarse mode AOD from 7 of the 8 dust-dominated stations in 4) for 2000.

Bringing together such an extremely wide range of observations into one synthesis for use by the AEROCOM community is a major undertaking and the authors should be congratulated for doing so. The description of the different measurements in section 2 reads well and the grouping carried out for the different datasets is very valuable work. Together, the datasets provide great opportunity (both within this paper and in future studies) to assess how well different models perform in simulating these five metrics which constrain different aspects of the model and are important for different interactions with the Earth System (e.g. the direct radiative effect and the supply of

ACPD
10, C10813–C10828, 2010

nutrients to the ocean). Providing the Tables S1, S2, S3 and S4 for the collected datasets is also of great benefit to the community.

The documentation of this kind of inter-model assessment and the performance of the models against the benchmark observational datasets is very valuable. The manuscript will be of interest to the global aerosol modeling community and the topic of the study is very suitable for publication in Atmospheric Chemistry and Physics.

The authors have clearly put a great deal of effort into processing the model output to facilitate the inter-comparison against these datasets. However, I feel that the manuscript as published in ACPD requires some considerable additional effort to improve the text and figures associated with section 3 of the manuscript which is currently not up to the standard of the observation-synthesis part of the paper. The abstract also requires substantial re-wording and as it stands currently doesn't do the article justice.

In the Main Comments and Specific Comments sections below I have tried to explain where I believe the manuscript specifically needs attention and provided suggestions for where and how the text needs to be improved before publication in ACP can be recommended.

Main Comments

It is a very challenging task to summarize the results from the 15 models against such an extensive compilation of measurements in a way that is accessible to the reader. The authors have attempted to include the comparisons against the observations of each of the models into the following Figures:

Figure 1 (scatter-plot of year-average total dust deposition against datasets in 1 above), Figure 3 (timeseries of monthly-mean total and wet deposition against FAMS datasets in 1 above) Figure 4 (scatter-plot of year-avg. surface dust mass concentration from 2ii above), Figure 5 (scatter-plot of year-avg. measured surface concentration from 2i above), Figure 6 (Hovmoller plot of monthly-means from model vs observations for

ACPD

10, C10813–C10828, 2010

dust mass observations at 21 sites as in 2i) Figure 7 (timeseries of monthly-mean surface dust mass at Barbados and Miami from 2i). Figure 10 (scatter-plot of year-avg AOD at 550nm at 25 sites of 1996-2006 dataset in 3 above) Figure 11 (Hovmoller plots of monthly-means from model vs observations for AOD_550 observations at 25 sites from 1996-2006 from 3 above) Figure 12 (scatter-plot of year-avg AOD at 550nm compared to 8 sites from 2000 in 4 above). Figure 13 (Hovmoller plots of monthly-means from model vs observations for AOD_550 observations at 8 sites from 4 above) Figure 14 (Hovmoller plots of monthly-means from model vs observations for coarse-mode AOD_550 observations at 22 of the 25 sites from 1996-2006 from 5 above) Figure 15 (Hovmoller plots of monthly-means from model vs observations for coarse-mode AOD_550 observations at 7 of the 8 sites from 2000 from 6 above) Figure 16 (scatter-plot of year-avg AE compared to 25 sites in 1996-2006 dataset in 3 above) Figure 17 (Hovmoller plots of monthly-means from model vs observations for AE observations at 25 sites from 1996-2006 from 3 above) Figure 18 (scatter-plot of year-avg AE compared to 8 sites from 2000 in 4 above). Figure 19 (Hovmoller plots of monthly-means from model vs observations for AE observations at 8 sites from 4 above)

In Figures 3 and 7, the data as presented is clear and the maps in Figure 1a and Figures 2, 8 and 9 are very useful to understand the location of the individual sites.

However, Figures 1, 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 contain too much information and I fear that the reader will not be able to take away the important aspects. In each of these Figures there are plots for the AEROCOM median model followed by a succession of individual plots for each of the 10-15 models that submitted data for that diagnostic. In the Hovmoller plots there are additionally an extra plot for the observations.

While this approach is admirable, and very complete – I think the manuscript would benefit from, moving the individual model plots to the Supplementary material and making each of these plots just present the results from the AEROCOM median model. Also, in the text for the results (sections 3.1, 3.2. 3.3, 3.4), the text should begin with

ACPD

10, C10813–C10828, 2010

an assessment of the skill of the AEROCOM median model and then, the general patterns for the individual model results that are highlighted in the current text are set into an overall context. With this approach, the reader has a chance of following the Figures whilst reading the text – at the moment I found it very difficult to keep track of the discussion in the text. Having only the AEROCOM median model in the Figures but having the individual plots available in the Supplementary material would allow detailed inspection of the performance of the individual models without detracting from the overall presentation of the results of the average behaviour of the models.

Moving the individual plots to the Supplementary Material would force the reader to concentrate on the performance of the AEROCOM-median-model against the observations. The authors have reached several general conclusions about specific deficiencies of some sub-sets of the models – and this is useful so should be kept in the text. But I believe it will greatly improve the manuscript if, in each case, the main analysis begins with the skill of the AEROCOM median model. This way, particular deficiencies that are common to all (or most) of the models can be drawn out first from the inter-comparison.

I recommend that, in the Hovmoller plots (Figures 6, 11, 13, 14, 15, 17, 19) in the main part of the paper, the authors, in addition to the observations and the AEROCOM median model, show the seasonal cycle for an AEROCOM multi-model normalised geometric standard deviation plot that illustrates the normalised variance among the models – this would summarise the variance of the models about the AEROCOM median and indicate where the model deficiencies in the median tended to be a symptom of all the models – or whether a substantial number of models did better/worse than the median.

Another particular issue that needs addressing in the manuscript before publication in ACP is related to the statistical measures that have been written out on each of the scatter-plots are RMS error, mean bias, correlation coefficient and standard deviation. The RMS error and mean bias are absolute measures of bias and are not very

ACPD

10, C10813–C10828, 2010

meaningful where the quantity being assessed varies over several orders of magnitude as for the dust deposition and mass concentration. For instance a poor performance at low and moderate burden sites could be masked by general good performance at high burden sites. To address this, I recommend that the authors add the normalized mean bias to the list of and also some measure of normalized mean error in addition to the RMS error. By doing so, one could then compare, for example, how well the AEROCOM median model performed at simulating the different metrics.

The discussion in section 4 reads much better than section 3 and the discussion of the role of different processes in the models is important. There should however, also be some discussion with reference to the role of model tuning here in increasing emissions to account for missing (possibly associated with sub-grid-scale processes) – could tuning the models to the Saharan dust plume be consistent with the general biases seen in the various AEROCOM models? Is there a role here for combining global models with regional models to better parameterize processes leading to emissions at the sub-grid-scale?

There are four main conclusions from the paper:

1) The models perform better at simulating the integrated properties AOD and AE than they do with total deposition and surface concentration.

2) Models generally simulate better surface concentration at stations downwind of the main sources than they do at remote ones. And do better at sites mainly affected by Saharan dust than those mainly affected by Asian dust.

3) Using the model-observation bias with respect to AOD and AE, the authors infer an emissions range that is narrower than shown by the models – i.e. an inference is drawnthat some models are under- or over-estimating emissions which is biasing their simulations.

4) From the current comparison one may conclude, with very large uncertainty, that the

ACPD

10, C10813–C10828, 2010

impact of dust on the ocean biogeochemical cycle is overestimated in most models.

The 1st conclusion is an interesting one and is founded on the finding that AOD and AE are generally within a factor of 2 of the observations whereas the surface concentration and total deposition are often biased by more than a factor 10. The authors should be careful here because the total deposition and surface concentration vary over several orders of magnitude whereas the AOD and AE varying only over 1 order of magnitude. So this finding is open to question here I feel. Perhaps the authors might consider changing this conclusion slightly so that it refers to the AEROCOM median model – it is clear from the Figures that the median model simulates quite well the surface mass concentration at U. Miami sites (Figure 5) and AOD (Figures 10 and 12) but systematically over-estimates the total dust deposition (Figure 1).

I consider this result to be very interesting and I wonder whether it could also be related to the 2nd conclusion. The global models run at coarse resolution and hence are missing emissions from processes at scales finer than the grid-scale. Also, we know that the dust emission parameters in the models are frequently tuned until they get the AOD in the main outflow from the Sahara approximately correct. Such tuning of the emissions upward will make the models compare better to the AOD and surface mass concentrations observations downwind of the key source regions but might be leading to the over-estimation of the dust deposition simulated by the models If the authors agree, then perhaps the revised manuscript could comment on this in the discussion and the conclusions be modified accordingly (and also the Abstract).

The 3rd conclusion is a novel outcome from identifying models with a bias in AOD and AE and inferring that those models must have a problem with emissions – and use this to provide a narrower range of emission simulated from the regions. This is an nice way of using the observations to better constrain the emissions fluxes – but I think it may in this case be compromised by the fact that the emissions fluxes quoted are actually over a different size range (some go up to 20 microns, others only to 8 microns – this is likely introducing a bias into the inter-comparison of the models – so I would

ACPD

10, C10813–C10828, 2010

be careful about making this kind of statement – the authors should at least provide a caveat about the different size-range for the emissions fluxes from the different models.

The 4th conclusion is very important – and yet it is not included in the Abstract – suggest that it should be added as a last point to the Abstract.

Specific Comments

Abstract:

i) Suggest to remove 1st sentence "Desert dust plays an important role. . .." from abstract as this is more Introduction and not needed in the Abstract which is rather long already at present.

ii) Suggest to move 2nd sentence "A large degree of diversity exists. . ." to before the sentence beginning "In general, models perform better. . ."

iii) Suggest to make the 1st sentence be the "We present the results of. . ." but change slightly at start to that it goes something like "This study presents the results of. . .."

iv) Need to re-word part of the sentence that says ". . .focusing on variables responsible for the uncertainties in estimating the direct radiative effect and the dust impact on the biogeochemical cycle". When you say that these variables are "responsible" for the uncertainties it sounds like you mean that they cause the uncertainty – rather I think you mean something else here – perhaps that they are variables which need to be well simulated for the two specific effects you refer to – need better wording here.

v) Suggest to extend the sentence ending "assessment of model performance" to also say something like ". . . and to identify common biases present in most of the models"

vi) Reword sentence that currently reads "These datasets form a benchmark dataset which is proposed for model inspection and future dust model developments" to something like "We use these observational datasets as a benchmark for the evaluation of the AEROCOM model simulated global dust distribution."

ACPD

10, C10813–C10828, 2010

Interactive Comment

vii) Suggest to reword sentence “Differences among models arise. . .” – here arise is the wrong word – perhaps you mean that all models are biased in this respect. Again I think the emphasis should be in identifiying common model deficiencies and that the skill of the AEROCOM median model could be used to identify systematic biases in the model – see my comments about the 1st and 2nd conclusions above in the Main Comments section.

viii) Suggest to remove or at least the sentences “This study also highlights important differences in models ability to . . .... deposition flux over Antarctica” and sentence after “The cause of this discrepancy could not be identified. . .” Again – I think you are pointing to a systematic bias among the models which is that they tend to over-estimate the deposition at remote regions.

ix) Add sentence something like “From the current comparison one may conclude, with very large uncertainty, that the impact of dust on the ocean biogeochemical cycle is overestimated in most models.”

2.1 Dust deposition – pg23787, line 13 – the Ginoux et al (2001) paper presents dust deposition observations from 16 sites but here only 10 are used (Table S1) – the observations at Shemya, Nauru, Samoa, Rarotonga, New Caledonia and Norfolk Island seem to have been omitted. Obsevations from several of these sites are included in the dust mass concentration observations (Table S2) but are not included in the total dust deposition. Was there a reason why these were left out?

2.4 Aerocom dust results – pg23795, lines 8-9 – the range of emissions given is 500-4400Tg/yr but the simulated size-range of dust particles over which this mass is distributed is different among the different models For instance from Table 1 some models simulate particles up to 25 microns whereas other consider only up to 8 microns. Consequently, the comparison of the emissions fluxes between the models is not a fair one because the models with a larger upper size limit will necessarily have a higher emission flux. This also affects the calculated lifetime and global total deposition fluxes.

ACPD
10, C10813–C10828, 2010

For instance, assuming a flat size-distribution of the emissions flux across the 5-25 micron size range, the dust mass emission for particles in the range 5-15 microns would be about a factor lower than that in the 15-25 micron size range.

The authors should add a caveat noting that some of the inter-model differences could be explained by different size range of the emitted dust. Note however that the deposition fluxes calculated at the site location should be consistent because the largest particles will fall out close to the source.

3.1 Dust deposition results – pg 23795, lines 25-26 and pg 23796 lines 1-2 – this is an example of where the presentation of the model results could be improved – the text says "All the models in this study mostly overestimate largely. . .." But it would be better here to illustrate this point by referring to the performance of the AEROCOM median model and how from Figure 1 it is clearly biased high – perhaps on average by a factor of 10 – this also illustrates how the statistics printed in the Figures are not very useful in interpreting the performance of the individual models in relation to the performance of the AEROCOM median model – see my specific point above about adding the normalized mean bias and normalized mean error (which would then give values that could be referred to when the skill of the AEROCOM median model in simulating the different quantities is assessed). I suggest also to delete the sentence "The level of overestimation varies largely from model to model" and instead give some solid numbers for range of the normalized mean bias in the models.

3.1 Dust deposition results – pg23796, lines 1-14 – there are many instances where statements like "Almost all models underestimate the, ". . .the degree of under-/overestimation varies largely from model to model". These findings need to be made more quantitative by adding reference to the normalized mean bias (nmb) values from the AEROCOM median model and the typical range of nmb values from the individual model plots.

3.2 Surface mass concentration results – pg23797, line 6 — I don't think it is appropri-

ACPD

10, C10813–C10828, 2010

ate to compare annual-average model values against cruise observations taken over a few days – suggest that Figure 4 only include the longer-term observations from the ground station data from Mahowald et al (2009) and omit the cruise observations. In fact one can see from the Figure that the long-term observations (diamonds) are not as biased against the model as the cruise observations (filled-in circles). I suspect some of the scatter may be due to the fact that the cruise observations are short-term observations and that dust deposition is so episodic. And remove the sentences that try to address this difference between the short-term measurements and the annual average of the model – I don't think an error bar fully addresses the issue.

3.2 Surface mass concentration results – pg23797, lines 8-11 – suggest to shorten this text as it is a bit convoluted – for instance suggest re-writing these 2 sentences as 1 as "The underestimation is mainly within a factor of 10 of the observations, except in remote regions of the Southern Atlantic Ocean," There are likely other parts of the Results part of the text that could be re-worded to be more succinct in a similar way. For instance also on lines 21-22 suggest to add a comma after (Figure 5) and delete the text "The cases where the underestimations exceed this limit correspond mostly to stations in Antarctica." and instead say "except for Antarctica (stations 1, 8 and 9)."

3.2 Surface mass concentration results – pg23798, lines 11-15 – suggest to shorten this text by deleting the text "The choice of colours is the same as the one applied in Figs. 2 and 5. In Fig 6 the groups have been labeled Low, Medium and High according to their regime of surface concentration." Instead insert "As is Figures 2 and 5," before "We continue to group" and replace "as done above" with "as Low, Medium and High according to their surface concentration regime".

3.2 Surface mass concentration results – pg23798, lines 19-20 – Again there is the qualitative statement "A large number of models (57%) mainly underestimate. . .." – suggest to make this more quantitative with reference to the normalized mean bias in the AEROCOM median model – and some general information about the range of nmb in the models.

ACPD

10, C10813–C10828, 2010

3.2 Surface mass concentration results – pg23798, line 26 – the text “not enough dust is transported to remote sites” – I don’t think this can be seen from Figure 6 which is currently being described – although it is probably clear from Figure 5 and S1 – be specific about where this is clear from the Figures.

3.2 Surface mass concentration results – pg23798, lines 22-26 – justify why you choose to discuss particularly the sites near Australia (rather than downwind of any of the other dust sources). Perhaps something like, “It is interesting to compare the seasonal cycle of dust between the two sites to the East of Australia which are geographically close but in quite different dust regimes.”

3.2 Surface mass concentration results – pg23799, lines 3-5 – Again the sentence “Most of the models fail to reproduce, to different degrees, the surface concentration in stations belonging to the 2nd group” needs to made more quantitative. Suggest to refer to the skill of the AEROCOM median model first (perhaps separate normalized mean bias scores could be calculated for the median model in the 3 groups?). Then, there should be a more quantitative statement about the skill of the models.

3.2 Surface mass concentration results – pg23800, lines 1-4 – Suggest to add a sentence at the start of this para something like “To test the simulated seasonal cycle in dust transport across the Atlantic, we compare the models to measurement. . .” (as you have it). Then suggest that the next 2 sentences can be deleted – i.e. those starting “These two stations. . .” And “Large model diversity. . .” Then sentences after that need to made more quantitative with reference to the normalized mean bias of the AEROCOM median model and how the other models perform in relation to that.

3.2 Surface mass concentration results – pg23800, lines 12-13 – re-word the sentence “The climatology in Fig. 7 illustrate that it is close to the seasonal cycle of the year 2000” It needs to be clearer what you mean here – I presume you mean that the seasonal cycle for the year 2000 is not unusual and follows the average from the 1996-2006 climatology. Since this is the case does one need to include Figures 6 and 7 at all? Do

ACPD
10, C10813–C10828, 2010

you they add anything new?

3.3.Aerosol Optical Depth results – pg 23801, lines 3-6 – again this sentence needs to be more quantitative – again, perhaps the general performance of the AEROCOM-median model should be presented first (it seems to do much better at simulating AOD than either the dust-dep or suface-mass) – can this be seen from the normalized mean bias (nmb) statistics? Then there should be more quantitative statements made about the individual models and the range of normalized mean-bias – maybe it is clear also that the model range of nmb against the AOD observations is narrower or closer to zero than for the dust-dep or surface mass? The authors should comment on this. Also re-word subsequent sentences particularly lines 9-13. Also re-word lines 20-22 to be more succinct – e.g. the observations show that for southern-African, northern-African and Middle-Eastern sites, the dustiest periods are December-April, July-September and May to September respectively. Or something like that. The rest of this section should also be re-worded with more quantitative statements about the mean bias of the AEROCOM-median model and more quantitative statements (involving nmb or normalized-mean-error or something like this) about deficiencies common to several models.

3.4 Coarse mode optical depth results – pg23803, lines 28-29 – is there also an alternative explanation for this – perhaps a different fine-mode aerosol type (e.g. biomass aerosol) might be contributing more AOD in the model than the observations? It doesn't necessarily follow that it is the size distribution of the dust that is at fault here.

3.4 Coarse mode optical depth results – pg23804, lines 1-6 – suggest to remove this Figure 15 (or move it to the Supplementary material) as it does not seem to offer any additional information here (the text could just state the result with "not shown".

3.5 Angstrom exponent results – pg23804, lines 12-15 – again the general statements should be made with reference to some statistical measure of the bias here (preferably nmb). Suggest again also to begin the description of the results with the AEROCOM

ACPD

10, C10813–C10828, 2010

median model. Parts of the rest of section 3.5 should be re-worded with reference to the nmb or normalized error where possible.

3.5 Angstrom exponent results – pg23806 – suggest to remove Figures 18 and 19 as the main reference to them seems to be that they behave similarly to the climatology from 1996-2006.

Figure 4 page 23839 – suggest to remove the cruise observations here as they cannot be described as annual-averaged observations. The short-term nature of these observations means that they are not comparable with the year-averaged model data. Or else change the caption to explain that they are not annual-averaged.

Figure S1 – legend has units ug/m3 but the caption says this is the relative difference – is the legend wrong or have you shown the wrong plot?

Typos & minor points

2.2 Surface concentration – pg23789, line 14 – replace “mean” with “means”,

3.2 Surface mass concentration results – pg23797, line 25 – replace “mayor” with “major”.

3.2 Surface mass concentration results – pg23798, line 20 – state what is meant by “first group”.

3.2 Surface mass concentration results – pg23798, lines 22,24,25,26 – add “New” before “Caledonia” and “Island” after “Norfolk”

3.2 Surface mass concentration results – pg23798, line 26 – add “from Australia” after “pathway”.

3.2 Surface mass concentration results – pg23799, line 17 – “as captured” rather than “with captured”?

3.2 Surface mass concentration results – pg23799, line 25 – what is meant by “year

ACPD

10, C10813–C10828, 2010

around phenomena" here? Do you just mean that there is some dust all year round? As you explain there is a pronounced high-dust season in Asia in the spring. Please re-word sentence to clarify what you mean here.

3.4 Coarse mode optical depth results – pg23803, line 18 – insert comma between "measurements" and "the coarse mode".

3.4 Coarse mode optical depth results – pg23803, line 21 – delete "with maxima in coarse AOD coninciding with the ones in total AOD" – that is obvious as you have already said it has the same seasonal cycle.

4.3 Emissions – pg23809, line 26 – replace "To illustrate this lets suppose" with "To illustrate this, let's suppose".

4.3 Emissions – pg23809, line 28 – insert comma between "underestimation" and "a larger fraction of fine".

4.3 Emissions – pg23810, line 5 – insert comma between "In the same way" and "the opposite case".

6 Conclusions – pg23815, lines 20-22 – insert comma between "conclude" and "with very large uncertainty".

6 Conclusions – pg23816, line 12 – replace "regional model performance" with "regional performance of the models" so as to be clear you are not referring to regional models.

Table 1 – pg 23831 – no mmr values are given for the 2 modes in the TM5 scheme.

Table 3 – pg 23833 – is there a typo here for the ECMWF emissions flux – it says 514 Tg/yr but the deposition fluxes suggest it should be about 6000 Tg/yr. Similarly the budgets don't add up for the CAM model – is this also a typo (or does the NaN sedimentation value account for this?

Table 4 – pg 23834 – the different models do not have equivalent dust size ranges –

ACPD

10, C10813–C10828, 2010

could the models be grouped into models with approximately the same size range so that one can at least compare the fluxes and lifetimes for the models that are similar.

Table S2 – caption – replace “Marin” with “Marine”.

Interactive comment on Atmos. Chem. Phys. Discuss., 10, 23781, 2010.

ACPD

10, C10813–C10828, 2010

Interactive Comment