Response to comments from reviewer # 2:

"Evaluation of biomass burning aerosols in the HadGEM3 climate model with observations from the SAMBBA field campaign"

Reviewer comments in normal font

Responses in italics, ***change to manuscript text shown in bold***

The authors present an evaluation of the performance of two aerosol simulation modules used within HadGEM, against a range of observations. The manuscript is thorough and well written, and falls within the scope of the journal. Some of the arguments made require some further clarification or quantification. After this and some other minor issues are dealt with, I recommend publication in ACP.

Thanks for the positive feedback

Main comments

My main question is about the scaling factors of GFED emissions used. I understand that they were chosen so as to "match the magnitude of AOD observations" (p16,l7) / "give good agreement between modelled and observed mid-visible AOD" (p9,l27). However, no further description of how the exact scalings were chosen is given, and yet the scaling factors make it all the way into the abstract, where they read as a conclusion about emission underestimates by GFED. Even slightly different scaling factors might change the discussion of AOD comparisons in section 3.1, in particular regional features such as discussed e.g. on page 16, lines 18-20. From the global/regional AOD values given in Figures 2-4, which reflect the entire aerosol distribution in the model, it seems that there's some range of scaling factors that might fit (as the authors also discuss on page 31. I'd suggest adding some sensitivity studies here, quantifying (through model bias calculations or similar) that any AOD comparison improvement in GLOMAP over CLASSIC is robust, and maybe toning down the focus on the scaling factors to be more in line with the authors own discussions in the Conclusion section.

Further, it's difficult to properly see from Figures 2-4 what regions are most affected by the improved aerosol treatment. I suggest adding some difference or ratio plots (maybe just versus one of the MODIS collections) to highlight the changes.

AOD scaling issue

I can understand the lack of satisfaction with the global scaling factors. The need to scale the BB emissions is a problematic issue that many other modelling studies have encountered (references given in section 2.3.1). Prior to scaling the simulation of AOD in BB regions was lacking in both CLASSIC and GLOMAP-mode and so the purpose of applying the scaling factors was to highlight this issue, show the approximate magnitude of the discrepancy, and show that it varied between CLASSIC and GLOMAP-mode due to differences in the two aerosol schemes.

As the reviewer has surmised, the scaling factors were not calculated precisely but resulted from experimenting with different scaling factors and assessing the results across the AERONET and MODIS AOD comparisons we present. We also looked at the modelled BBAOD from CLASSIC and GLOMAP-mode to ensure these were consistent in magnitude across the main BB source regions after scaling.

1) *To make this clearly in the manuscript the following text has been added in section 2.3.1:*

Page 10, Line 8:

"These scaling factors were not calculated precisely but were found to give good overall correspondence between modelled and observed peak AODs over continental BB source regions in the tropics, and tuned to ensure a consistent AOD contribution from BB emissions in CLASSIC and GLOMAP-mode, over the BB source regions."

2) *The agreement between CLASSIC and GLOMAP-mode BBAOD is explained more clearly in section 3.1:*

Page 17, Line 6:

"The magnitude of BBAOD is also very similar in both models as BB aerosol emissions were scaled separately in each model to ensure the modelled AOD approximately matched MODIS and AERONET AODs observed over the main BB source regions during peak BB months where BB was the dominant contributor to modelled AOD."

3) *As recommended by the reviewer we have also added plots of unscaled AOD to Figures 2 – 4 to allow the reader to see the impact of scaling on total AOD. This is now discussed in section 3.1:*

Page 17, Line 11:

"The impact of the BB aerosol emission scaling factors is shown by comparing the total modelled AOD from scaled and unscaled simulations. The emission scaling factors have a relatively modest impact on the global distribution of AOD when assessed on an annual mean basis. This is due to

the highly seasonal nature of BB emissions. Nevertheless, even in annual means, it is clear that AODs over tropical South America and Africa are somewhat lower than observed (from both MODIS collection 5 and 6) in the un-scaled simulations. The scaling factors bring modelled AOD closer to the observations, although the benefit is clearer in later figures (3 – 5).”

Other minor edits have been made in section 2.4 (experimental design) to include the description of the unscaled simulations. Also some minor edits in section 3.1.1 and 3.1.2 (AOD assessment) to contrast their result with the scaled result.

In section 3.1.2 we add:

Page 18, Line 11:

“Over the BB regions the modelled AOD is generally underestimated in the un-scaled simulations compared to MODIS. In the scaled simulations total AODs agree very well with MODIS, especially over South America and Indonesia. However, some discrepancies between modelled and observed AODs remain...”.

Page 18, Line 23:

“This leads to an overestimate of modelled AOD over central Africa in the scaled simulations.”

4) *For the benefit of the reviewers and editor, below we also include scatter plots of model versus AERONET and MOIDIS AOD below (Figs R1 & R2).*

This demonstrates the difficulty we had in deriving an appropriate global scaling factor. Whilst it is obvious that AODs in BB-affected regions are biased low prior to scaling there is a great deal of spread. The magnitude of the discrepancy varies between different regions / AERONET sites and to some extent as a function of AOD. There may be various reasons for this spread, such as the lack of sub-sampling of the model to observations space-time points, and biases in simulated water uptake and in other aerosol sources. Even after selecting regions and months where BB emissions are the most dominant, other aerosol sources still contributed 20-50% to modelled AOD (in the scatter plots we filtered data to remove points where BBAOD contributed less than 50% to the scaled AOD). Observations may also be biased in some regions. For instance, we have low confidence in the high MODIS AODs (0.6 – 1.2) over the SE Atlantic Ocean due to low sampling statistics (data points circled in Fig R2 below). Intensive observations over the SE Atlantic via the ORACLES, CLARIFY and AEROCLO-SA will help elucidate this problem (Zuidema et al., 2016). We do have reasonable confidence in the high AODs observed over Southern Amazonia and Southern Africa during Aug – Oct where AERONET (Alta Floresta and Mongu) and MODIS are in reasonable agreement. For these reasons it did not seem possible (or appropriate) to devise a precise method for deriving the global scaling factors.In the case of GLOMAP-mode the scaling*

factor of 2.0 results in good agreement between modelled peak AODs over Amazonia and Southern Africa (Aug – Oct) and the AERONET observations and MODIS observations (Figs R1 & R2 below). In the case of CLASSIC this was scaled to the nearest 0.1 to give the best consistency between CLASSIC and GLOMAP-mode for the BBAOD and AOD in the main BB regions equalled that of GLOMAP-mode, as shown below in Fig R3, and in Figs 2 – 5 of the manuscript.

*Zuidema, P., J. Redemann, J.M. Haywood, R. Wood, S. Piketh, M. Hipondoka, P. Formenti, Smoke and Clouds above the Southeast Atlantic: Upcoming Field Campaigns Probe Absorbing Aerosol's Impact on Climate, Bulletin of the American Meteorological Society, DOI: http://dx.doi.org/10.1175/BAMS-D-15-00082.1, 2016.

In the abstract we do not claim that this scaling factor is due to bias in GFED3.1. Rather we note how the scaling factor depends on assumptions within the aerosol schemes:

Page 1, Line 19:

"Finally, good agreement between observed and modelled AOD was gained only after scaling up GFED3 emissions by a factor of 1.6 for CLASSIC and 2.0 for GLOMAP-mode. We attribute this difference in scaling factor mainly to different assumptions for water uptake and the growth of aerosol mass during ageing via oxidation and condensation of organics."

Similarly, in the conclusions we do not put forward these scaling factors as a conclusion about the magnitude of GFED3.1 biases or as recommended scaling factors for other models to adopt. Rather we discuss the various issues and areas of uncertainties that need resolving to make sense of this scaling problem.

5) *To stress this further the follow text has been added to paragraph 5 of the conclusions:*

Page 33, Line 5:

"Moreover, due to the difficulties of comparing large-scale models with limited observations, these scaling factors are not precise but rather indicate the approximate scale of the AOD biases."

Figure 7 is another example where I don't quite follow the authors argument that GLOMAP is a clear improvement (page 19-20). It seems that some observed species ratios are closer to CLASSIC, some to GLOMAP, and unifying MAC to 10 m2/g also seems to pull both ways. Please add some quantification of the improvement here.

Yes, the results in Figure 7 show that the composition from the two aerosol schemes are pretty similar with no clear winner. The only notable difference is a slightly higher BC mass fraction in GLOMAP-mode, which is generally closer to the observations.

For instance averaging over the four locations in Figure 7 the average BC fractions are:

	Obs as given	Obs with MAC 10 m2/g	CLASSIC	GLOMAP-mode
Avg BC frac %	9.5	9.2	6.2	7.9

After unifying the MAC assumption the comparison still comes out in favour of the higher BC fraction in GLOMAP-mode. I agree that we can not be very confident in this comparison as there are large uncertainties in absorption and BC measurements. We therefore do not mention this result in the abstract and now describe the apparent improvement more cautiously in section 3.2:

Page 21, Line 23:

"GLOMAP-mode gives slightly higher BC mass fractions than CLASSIC and in general GLOMAP-mode BC mass fractions are closer to observed values... ***Tentatively,*** *GLOMAP-mode therefore shows* ***some*** *improvement over CLASSIC, although it still appears to underestimate BC mass fraction..."*

And In the conclusions:

Page 32, Line 9:

*"...These variations are a challenge for the models to capture. Whilst the dry BC mass fraction and SSA in GLOMAP-mode (7-10 %; 0.85 – 0.87) are closer to the observed values than CLASSIC (5-9 %; 0.91)****,*** *the modelled variability between source regions is lower than observed.*

As the authors themselves point out (p29,l17), sensitivity tests of the injection height assumptions of BB aerosols would be very useful - both for the present analysis, and for the aerosol community. Adding 1-2 simulations here would further increase the relevance of the paper.

Yes, we did run some earlier experiments looking at the sensitivity to vertical injection height. At present forest emissions are elevated from 0 – 3km and savannah emissions at the surface. We tried (i) injecting all BB emissions at the surface, (ii) injecting all BB emissions uniformly from the surface to 3km). The results are included in the additional plot (Fig R4) below. Scenario (i) made a huge difference to the vertical profile of aerosol over Amazonia leading to very high concentrations in the lowest few 100m of the atmosphere. This was in part due to the fact that emissions in the model do not have a diurnal cycle and so concentrations build up in the night-time. In the comparisons for the DABEX and SAFARI-2000 regions the sensitivity to injection height was less, mainly as these are downwind

regions with elevated BB plumes and atmospheric convection has had time to mix the surface emissions upward. However, Injecting all BB emissions from 0 – 3km rather than at the surface still has some impact on these downwind regions. These sensitivity experiments were completed in an earlier configuration of HadGEM3 and have not been repeated for the GA7 configuration that we have used in this manuscript.

The main focus of the study has been on comparing the two aerosol schemes as they are rather than experimenting with improvements. As far as we can judge here, the models perform reasonably well with the current (rather crude) injection height assumptions. Therefore, we made a decision to leave the experimentation with injection heights as a topic to re-visit and explore in a separate study. To evaluate this properly would require more observations, especially for instance taking advantage of MISR plume heights and/or CALIPSO data. An AEROCOM activity is also focussing on this very issue and we have participated in this submitting HadGEM3 results. Hopefully this will shed some light on the injection height issue.

Minor/technical issues

- Many references are made to two as-yet unavailable manuscripts ("Darbyshire et al., in preparation"). As comparisons to SAMBBA is a main point of the paper, and topics such as data averaging are important for reproducibility, this was somewhat annoying when trying to understand the analysis presented here.

These papers are still in preparation at the present time. The first of these (2016a) relating to the vertical distribution of aerosol is less critical and the citation has been removed. Marenco et al. (2016) is already cited here (section 2.3.4, paragraph 1) and provides evidence of the vertical distributions of aerosol during SAMBBA from both the lidar and insitu measurements. The second (now the only) reference to ("Darbyshire et al., in preparation 2016") relates to a study summarizing the physical and optical properties of aerosols measured during SAMBBA. I understand it is not ideal for the reviewer and reader but it is planned that this paper will be submitted to ACP this autumn. This will contain detailed descriptions of the flight patterns, instrumentation and data processing methods. Only a summary can realistically be given in the present manuscript but this has been bolstered somewhat in section 2.5, paragraph 1 to explain the basic approach to data selection and averaging:

Page 13, Line 26:

"The regional averages of aerosol particle size distribution, composition and optical properties are based on data from straight level runs sampling the regional haze. Data sections corresponding to plume penetrations (identified from spikes in CO, CO2, BC and aerosol scattering) were filtered out prior to averaging."

As so often in dedicated observational campaigns, peer reviewed papers cannot be always be prepared / submitted in series as this would lead to undue delays in the duration for open submissions to the journal.

- P5l21: Please give a brief explanation of the MetUM, even if it's probably known to most readers.

A little more detail has been added at the start of section 2.1.

Page 5, Line 21:

"This work uses global simulations of the Met Office Unified Model (MetUM), a state-of-the-art atmospheric general circulation model with a range of configurations for numerical weather prediction and climate simulations. Here we use the MetUM within the framework of the third generation Hadley Centre Environment Model (HadGEM3) (Hewitt et al., 2011)."

- P6l20: So SO4 and BC emissions are climatologies, but BB emissions have annual variations? Or do they change through interpolations? Please clarify.

For both anthropogenic and biomass burning emissions we take the mean emission from the period 2002 – 2011 and run without any year to year change in emissions. The BB emissions do contain a seasonal cycle of course so we use monthly varying emissions. To make this clearer in section 2.1. We add, when referring to the GFED emission.:

Page 6, Line 17:

"We use monthly mean emissions averaged over the period 2002-2011."

And in reference to the anthropogenic emissions:

Page 6, Line 24:

"However, we keep annual emissions constant at the 2002 - 2011 mean rate."

- P7l15: Will the assumption of no sea salt transport over land in CLASSIC affect AOD comparisons in BB-emission coastal regions? Does this contribute to differences to GLOMAP?

Yes, but the contribution of sea salt to AOD in the major BB regions is very small. The main aerosol sources that affect the comparison are anthropogenic emissions, dust (mainly West Africa), and biogenic (especially Amazonia and the Congo basin). These are mentioned in the discussion of results.

- P12l1: Add "of" between "prevalence" and "moist".

Done

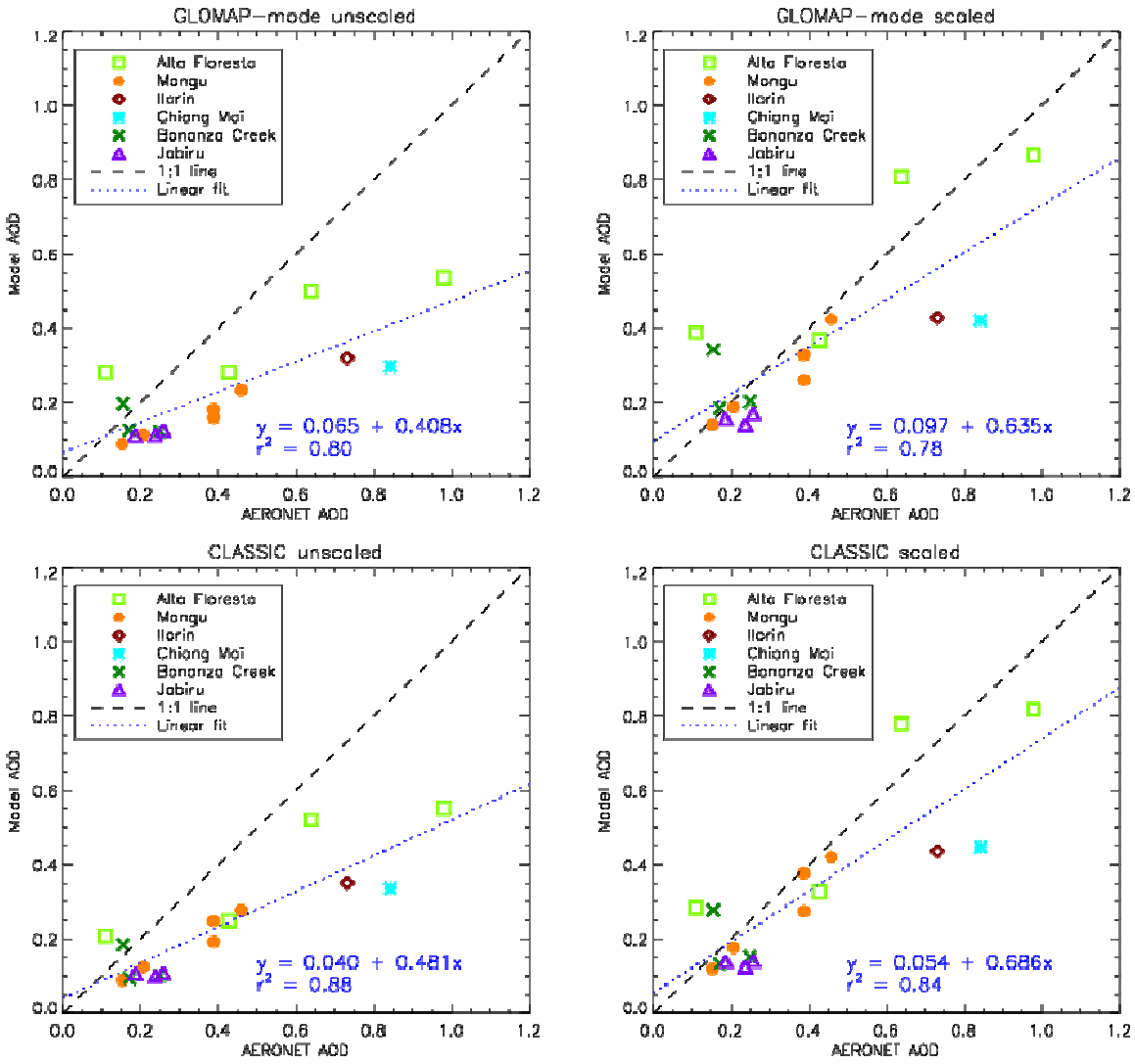

Fig R1. *Model versus AERONET AOD. Observations are the long-term monthly mean averages shown in Figure 5 of the manuscript, except here data points are selected only if CLASSIC or GLOMAP-mode showed that at least 50% of the AOD was associated with BB emissions (i.e. BBAOD / AOD > 0.5) in the scaled simulations. The linear fit is a straight-forward linear regression (blue dotted line). Model results are shown with and without the global scaling of BB emissions.*

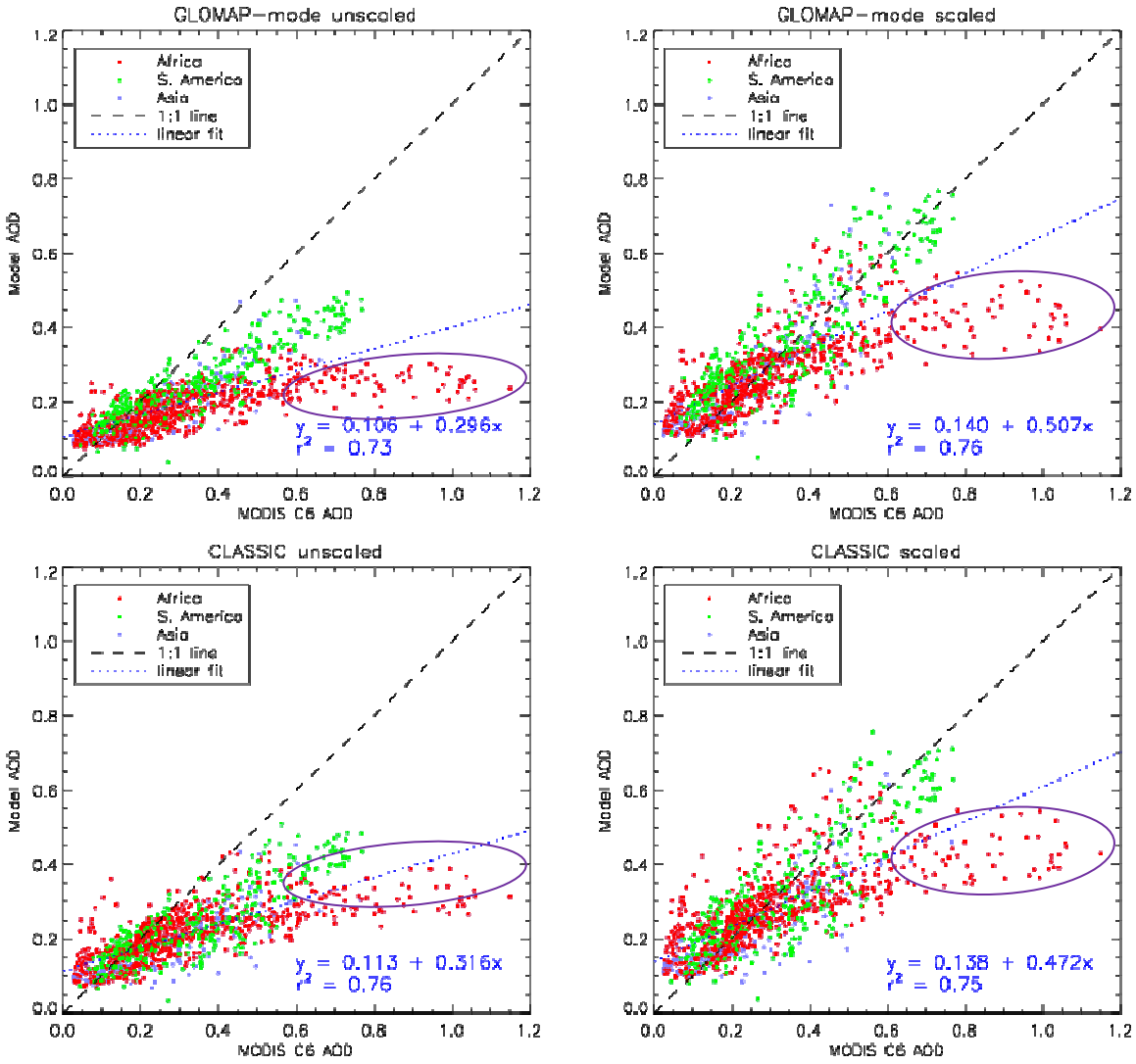

Fig R2. *Model versus MODIS C6 AOD. Observations and model data points are the long-term monthly mean averages for September as used in Figure 3 of the manuscript, except here data points are shown only if CLASSIC or GLOMAP-mode showed that at least 50% of the AOD was associated with BB emissions (i.e. BBAOD / AOD > 0.5) in the scaled simulations. The linear fit is a straight-forward linear regression (blue dotted line). Model results are shown with and without the global scaling of BB emissions. Circled are data points over SE Atlantic ocean, which may be overestimated by MODIS.*

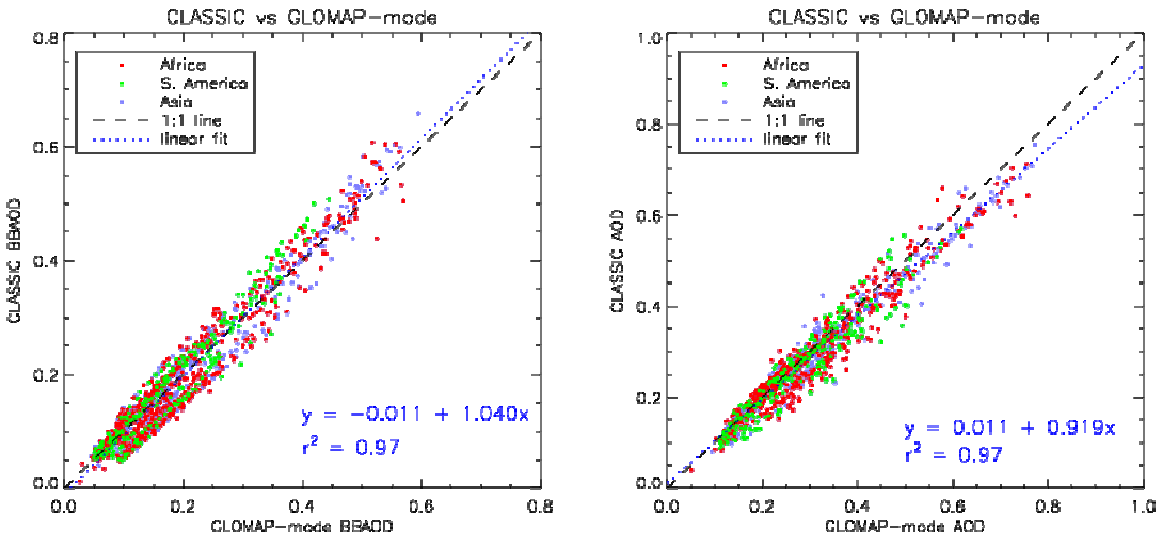

Fig R3. *CLASSIC versus GLOMAP-mode BBAOD (left) and AOD (right) for the scaled simulations. Model data points are the long-term monthly mean averages for September as used in Figure 3 of the manuscript, except here data points are shown only if CLASSIC or GLOMAP-mode showed that at least 50% of the AOD was associated with BB emissions (i.e. BBAOD / AOD > 0.5) in the scaled simulations. The linear fits are straight-forward linear regressions (blue dotted line). The plot demonstrates that the scaling factor of 1.25 applied to the BB emissions in the CLASSIC simulation leads to good agreement with the BBAOD and AOD from the scaled GLOMAP-mode simulation.*

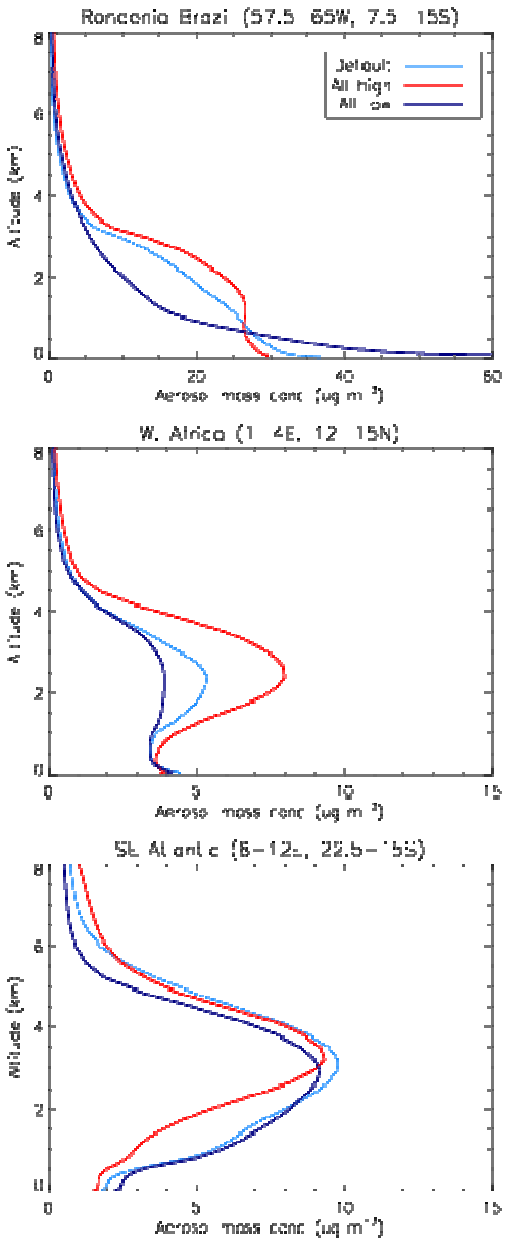

Fig R4. *Vertical profiles of fine-mode aerosol mass concentration for the SAMBBA, DABEX and SAFARI-2000 regions (as in Figure 14 of manuscript). All lines are from earlier simulations with GLOMAP-mode based on GA6 (the previous Global Atmosphere configuration compared to simulations used in the manuscript). The light blue indicates a simulation using the default vertical injection assumption, red indicates a simulation where all BB emissions are injected uniformly from 0 – 3km, and dark purple indicates a simulation with all BB emissions injected at the surface.*