Response to comments from reviewer #1:

"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***

This well-written, scientifically sound paper uses observations from the South American Biomass Burning Analysis (SAMBBA) campaign to evaluate biomass burnin aerosol properties in then HadGEM2 climate model. Two aerosol modules are considered: a simpler mass-based model (CLASSIC) and a more complex microphysical aerosol module (GLOMAP). The authors show that both schemes reasonably represent biomass burning aerosol properties once emissions are scaled up, and that scaling depends on the scheme considered. Overall, the microphysical representation of the aerosol produces a slightly superior biomass burning aerosol representation, particularly in the representation of optical property variability and size distribution. I find this study acceptable for publication in ACP after the authors address a few minor points, which I outline below.

Thank you for the positive feedback

1. Page 2, Line 19: Space between needed: "3km" -> "3 km"

2. Page 3, Line 3 (and thereafter): "Biomass Burning" -> "biomass burning"

3. Page 3, Line 5: Remove quotations around aerosol-radiation interactions.

Done *and made consistent through manuscript*

4. Page 3, Line 7: I do not agree with the statement that aerosol absorption suppress the hydrological society. What about the Elevated Heat Pump effect of Lau et al., Randles et al. etc. and similar studies?In some cases absorption suppresses precipitation, in others it can enhance precipitation. Please reword to state it can have either impact, depending on environmental conditions.

Yes, this is true. Strong absorption can sometimes increase precipitation regionally. The sentence has been expanded to make this point and I add the two citations the reviewer points to.

Page 3, Line 7:

"These effects may suppress the hydrological cycle by stabilizing the lower troposphere, although strong absorption can in some cases enhance precipitation regionally by increasing low-level convergence (Wu et al., 2013; Ramanathan et al., 2001; Lau et al., (2006); Randles et al., 2008)."

5. Page 6, Line 14: This is unclear. Are then same exact oxidant fields

(climatology) used for both the CLASSIC and GLOMAP runs? If not, why? Wouldn't

this make a difference?

The oxidant fields differ as the CLASSIC climatology had been generated separately from an earlier simulation. This is now made explicit in the text. Whilst it is true that these are not consistent the oxidants have no direct impact on the simulation of biomass burning aerosol in either CLASSIC or GLOMAP-mode. Impacts are limited to the relatively minor secondary impacts of oxidation of sulphate, which mix internally with BB aerosol. The comprehensive nature of implementing an entirely different aerosol and chemistry scheme (CLASSIC vs GLOMAP-mode) means that different oxidant fields have been adopted. It is difficult to unify these inputs without significant development and re-evaluation for a scheme (CLASSIC) that is no longer being actively developed. We now make this difference explicit in the text:

Page 6, Line 14:

"CLASSIC used a climatology of oxidants generated separately from an earlier simulation."

6. Page 7, Line 15: This seems odd. No sea salt transport over

land? You just wash it out once it crosses the coastline? Is this realistic?

The sea salt mass is not an advected tracer so only occurs over ocean. This point wasn't clear so I explain a little better:

Page 7, Line 14:

"CLASSIC uses a diagnostic scheme for wind-driven sea salt, i.e. sea salt aerosol is not transported but instead is diagnosed locally over ocean points as a function of wind speed and with a prescribed scale-height in the vertical (see Bellouin et al., 2011, Jones et al., 2001)."

7. Page 7, Line 7: I'm a bit confused about how emissions are prescribed in CLASSIC. Doesn't the OC:BC ratio on emission come from the emissions inventory (GFED)? Please clarify.

This point is now explained in more detail.:

Page 7, Line 27:

"The total aerosol mass emitted into the fresh mode is taken as the sum of BC and OC from GFED but the model makes its own assumptions regarding the proportion of BC and OC in each BBA mode. Each BBA mode is assumed to be an internal mixture of Black Carbon (BC) and Organic Carbon (OC) with an organic carbon mass fraction of 91.5 % for the fresh mode and 94.6 % for the aged mode".

8. Page 7, Line 25 (and thereafter): Put a space between number and unit (micrometer).

9. Page 7, Line 31: "Refractive Index" -> "refractive index"

10: Page 9, Line 14: "hygroscopic" -> "hydrophilic"

Done

11. Page 9, Line 16: Do you get optical properties on the fly or from a pre-computed look-up table?

Yes, we do use pre-computed look-up-tables. This is now made explicit (near end of final paragraph in section 2.2.2).

Page 9, Line 25:

"Aerosol optical properties are derived for each mode as function on aerosol mode diameter and RI using look-up-tables with pre-computed results from Mie theory. For these the RI is computed by volume-weighted averages depending on the mixture of components within any given mode."

12. Page 9, Line 19: You do not consider UV absorption (brown carbon)? Why not? Can you indicate if this should be the case here? Does SAMBBA lend support for a representation of brown carbon?

Neither CLASSIC or GLOMAP-mode currently have a representation of brown carbon but this as an important issue that we are considering for future developments to GLOMAP-mode. This is a challenging issue mainly the huge diversity of organics and uncertainty in measurements of brown carbon and also in how to represent this globally with only one or two tracers for organic carbon. SAMBBA did not lend any specific support for a representation of brown carbon, mainly as measurements of spectral absorption were not available from the instruments deployed in the field. The issue of brown carbon is mentioned at the end of the conclusions section.

13. Page 9, Line 29: You imply that QFED and FEER use global scaling. They do not. QFED,for example, uses biome-specific scaling. Please reword so that is clear while these inventories do scale their emissions, the scaling is more complex then multiplying by a single number globally as you do here. Considerable effort was made by these inventories in their scaling efforts.

Yes, this is an important distinction. I have altered this sentence to include this point and also shifted it further on in the discussion so that the difference in approach compared to the globally uniform scaling factors applied in atmospheric models is more obvious:

Page 10, Line 16:

"Note, observed AODs are also used to derive biome-specific or spatially varying scaling factors in some top-down emission estimation methods such as the Quick Fire Emission Dataset QFED (Darmenov and da Silva 2013) and the Fire Energetics and Emissions Research (FEER) (Ichoku and Elison 2014)...".

14. Page 10, Line 1: Why did you choose to use GFED as opposed to QFED or FEER? Curious.

For global climate modelling studies we typically use GFED as the bottom-up approach is well suited to the philosophy of Earth-System models that are built up from a detailed representation of physical processes. A top-down approach such as QFED or FEER may provide more accurate emissions for recent historical periods aiding evaluations between models and observations. However, most of our climate model experiments tend to align broadly with specifications from international MIP activities (CMIP5/6, AEROCOM, AerChemMIP) that use the GFED series. We therefore choose to evaluate the model using GFED.

15. Page 10, Line 30: When calculating optical properties, do you use the POM or OC mass?

Yes, as explained in section 2.3.2 the mass of organic aerosol mass is scaled up to POM for GLOMAP-mode but not for CLASSIC. The calculation of optical properties for the organic aerosol component in GLOMAP-mode is given in section 2.2.2.

16. Page 14, Line 7: Why do you just consider Aqua?

This was a pragmatic choice because we are aware that there was some degradation / drift in the calibration of MODIS Terra towards the later years of our study period (Polashenski et al., 2015). Also, the MIDIS collection 6, level 3 merged deep-blue and dark target product was only available from Aqua.

We add to section 2.6, page 15, Line 7:

"Terra products were not included as drift in the calibration of MODIS Terra in the later years of our observation window may have affected the retrieved AODs (Polashenski et al., 2015)."

17. Page 14, Line 16: Are you using the Level 3 product? Is this the *best to be using? Do you do any sub-sampling of your model to only consider satellite observation times? Would sub-sampling impact your comparison? You do say this on page 16 but I would comment here as well.*

Yes, we are using level 3 MODIS data (opening sentence of section 2.6). As we did not have the facility to sub-sample the model on observation space-time points the level 3 product is the most appropriate for our comparisons. In any case we further average the data over space and time to make our comparisons with the long-term (10-year) monthly means from the model resolution (~150km). To make this point more clearly we add the following text to the end of the paragraph, section 2.7.

Page 16, Line 9:

"For MODIS the level 3 data has been further averaged to generate 10-year (long-term) monthly mean AODs at the native resolution of the atmospheric model. These are compared against the long-term mean model values without any sub-sampling of the model data on observation space-time points. Sampling biases that may arise due to the lack of sub-sampling are discussed in section 3.1."

18. Page 16, Line 28: You say the model AOD is calculated on clear-sky RH. What does this mean? Does your model track 2 RH (clear- and cloudy- sky)? Or do you represent RH as a sub-grid PDF? Please explain.

The Unified Model large-scale cloud scheme does include a representation for the sub-grid PDF of total water. The scheme provides the mean relative humidity

calculated for the cloud-free portion of the gridbox and we use this for the aerosol hydration. The methods used to construct the PDF and calculate the mean clear-sky RH are quite complicated and so we choose not to go into this within the manuscript.

19. Page 17, Line 28: Ilorin will also be impacted by dust aerosol; is it therefore correct to conclude that your under-representation is due to biomass burning aerosol alone? Or could your dust simulations also be off?

Yes, dust undoubtedly affects the comparison too. We add this caveat to the text.

Page 19, Line 11:

"This again suggests an under-representation of BB emissions across West Africa during Northern hemisphere winter, although the low-bias could be partly caused by a low-bias in mineral dust aerosol from the Sahara."

20. Page 17, Line 32: How are you attributing this to BB aerosol? Couldn't it also be anthropogenic?

The peak in total AOD during June-Sept coincides with a peak in BBAOD and shows there is a BB contribution that comes in on top of a background of other aerosols (anthropogenic, dust, etc).

21: Page 18: Line 15: OA =POM?

OA is used for "Organic Aerosol". The acronym is given at the start of section 3.2 and used through the manuscript. We don't use the term POM because secondary organic aerosol also contributes to the organic aerosol mass in both models.

22. Page 19, Line 2: I still don't understand why both CLASSIC and GLOMAP do not have the BC:OA ratio coming from the emissions inventory (since that is built into the emissions).

*The issue with CLASSIC is addressed above in point 7. As explained in the text (section 3.2, paragraph 2) GLOMAP-mode does take the OC:BC ratio from the emission inventory when adding primary aerosol emissions to the model. However, the BC:OA ratio in aerosol mass concentrations can vary because (a) there are other primary and secondary sources of carbonaceous aerosol in the model, (b) OA added from primary emissions is assumed to be 1.4 * OC emission.*

23: Page 20, Line 9: Remove space before %

There is a single space before each "%" in this manuscript, as recommended in the manuscript preparation guidance.

24: Page 22, Line 24: "Sampling issues (e.g. altitude) may be a large source of representativeness error in the ... "

Suggested edit adopted (page 24, Line 19)

25: Page 23, Line 25: Do you not also sub-sample the model at AERONET observation times? Why not? Would't that matter?

We did not have the facility within our model configuration to sub-sample on AERONET observation times. As explained in section 2.7 we simply average over time to generate long-term monthly means. This is not ideal and we are working towards building more observation simulators into our modelling systems but this is not a trivial task.

26: Page 24, Line 27: Comment on why it's expected that a microphysical model can achieve more realistic variability in optical properties than a mass-based model. This is an important point.

This is a complex point. Microphysical models have the potential to achieve more realistic variability in optical properties than mass-based schemes but this isn't necessarily a given. In CLASSIC the microphysical properties were guided by in-situ measurements and this empirical approach can prove relatively accurate in representing the mean optical properties (Haywood et al., 2003; Johnson et al., 2008a).Variability in total optical properties in this approach is then dependent on external mixing between aerosol species and on the parameterized variability with relative humidity, rather than on a representation of microphysical processes. Note, the CLASSIC BBA species is itself represented optically as an internally mixed particle with specified mass fractions of BC and OC. Natural variability in optical properties can potentially be simulated much better with the more complex microphysical models. These can also represent the process of internal mixing between modes. However, these include many uncertain processes that mass-based models do not need. Therefore, we prefer not to set the expectation that one scheme is better than the other but rather assess the two approaches as objectively as possible. Overall GLOMAP-mode compares better than CLASSIC to the observed variability in optical properties as mentioned in the second sentence of the conclusions.

27. Page 25,

Line 30: "Scattering Growth Factor" -> "scattering growth factor"

Done

28. Page 30, Line 8:

This is a very important point. Can you comment on why the growth factor is smaller in the observations? Do any of the chemical data suggest reasons for this? Do you think this is characteristic of all BB or just over south America?

Section 3.5 contains a detailed discussion of the observations and modelled growth factors. There are huge discrepancies between the three observation sets (MH03, KH98, SW08) and the reasons for this are not fully understood. The models actually agree with the observations from MH03 (S. Africa) but the observed growth curves from KR98 (Brazil) are much lower. As we state, "This is difficult to reconcile, especially as both were derived from [the same] airborne humidified nephelometer system." We consider that: "Possibly the regional aerosol mixture (categorised as "regional air" in MH03) contained a substantial proportion of highly hygroscopic sulphate from industrial sources in Southern Africa and is therefore not representative of purely carbonaceous aerosol." The H-TDMA measurements reviewed more recently in SW08 suggest an intermediate level of water uptake. These include ground-based data from several Amazonian campaigns and will not suffer some of the problems associated with aerosol sampling on aircraft. In the conclusions we therefore suggest (final paragraph of section 3.5 & paragraph 3 of section 4) that the models overestimate the water uptake, and emphasize the range from the H-TDMA observations in SW08.

29: Figure 3: Could remove repeated colorbars (since all are the same) and just have one large one.

Done for all AOD contours plots (Figures 2 – 4).

What is the white in the squares in (e)? Missing data? Maybe grey this out.

Yes, missing data values from MODIS are plotted as white. I add this detail to Figure caption 2. Using white causes the areas to stand out as missing, whereas grey is harder to distinguish from the pale blue colours and could be mistaken for values around 0.05-0.1.

30: Figure 5: For AERONET, put the standard deviation on the month to indicate sub-monthly variability Could also do this for the model as shading.

The uncertainty in the monthly mean AODs have now been indicated in Figure 5 by plotting vertical lines showing +/- 1 standard error. Shading had been tried for this plot but made it too difficult distinguish between the 5 plotted lines.

In section 3.1, page 19, Line 22:

"The approach we have taken is to average over 10-years of data to gain more confidence in the long-term monthly means. The standard error on the monthly means AODs are generally much smaller than the differences between observed and modelled values, indicating that our results are not strongly biased by interannual variability of either the simulated or observed AOD. There main exceptions are for August at Bonanza Creek and Aug – Sept at Alta Floresta where the larger standard error in AERONET AOD indicates that interannual variability has a strong impact on the comparison."

31: Figure 8: The part labeled "dust-dominated" -> is there another measurement that confirms this? You see the same thing (almost) in SAFARI-2000 and there was not as much dust in southern Africa correct?

The evidence for this comes from Johnson et al. (2008a) where by sampling in and out of plumes the signature of the smoke aerosol size distribution is clearly distinct from the dusty background. By comparing samples in this way we found that the smoke contributed very little to particle numbers for d > 0.8um. In SAFARI-2000 you see the same shape to the PCASP size distribution but with the number concentrations for d > 0.8um an order of magnitude less. This could still be dust as there are some dust sources in Southern Africa (e.g. dry river beds along the Namibian coastline).

32. Figure 11: Again, one colorer for each 2 panels across.

Given that the contour levels do vary between panels it is maybe clearer for each panel to include it's own colour bar.

33. Figure 12: Rather than all the grey spaghetti lines, why not just shade max-min? Would be easier to see

Possibly, but because there are few data points in some months and sometimes outliers, shading the min-max range gives a somewhat misleading impression. The spaghetti lines were used so that you can see the min-max range and also a feel for the number of data points each month and how they are spread.