Response to Reviewers

We thank the two reviewers for their comments on the manuscript. We have responded to every reviewer comment (highlighted in italics).

General comment

The introduction does not adequately pose the scientific question nor state the importance of this research. (Reviewer #1)

We have revised the introduction to improve the description of the paper, including in particular some of the summary comments from the last section to emphasise the purpose of the age of air section of the paper. The introduction now addresses the next point made by the reviewer.

Why is it important to calculate the actual age of the emissions as opposed to the photochemical age? (Reviewer #1)

The main objective of the age of air calculation is to quantify the effects of photochemical retardation due to aerosols within a biomass burning plume. Quantifying the retardation is achieved by computing the physical and photochemical age of the plume. We have clarified this point in the introduction.

Referee #1

You should discuss the results of Parrington et al (2012, 2013) including limitations of the methods employed …and the benefits of the new method

Limitations of the photochemical age method are briefly discussed in regards to OH assumptions in line 7 page 8735. We have extended this section to include other limitations of this method and therefore the results in Parrington et al 2013. We have also included the advantages of our new method.

Wouldn't it make more sense to use a finer spatial resolution in the model?

To address this reviewer response we have run the model at other spatial resolutions, including nested model simulation over the BORTAS study region at a resolution of 0.5° latitude and 0.667° longitude during July 2011. As discussed below, this provides a more difficult test for the emission inventories.

Can you speculate on the impact to your conclusions?

We have now run the model at higher spatial resolutions and driven it using different emission inventories for biomass burning to assess uncertainties introduced by emissions and atmospheric transport. In summary using the nested resolution does not change the outcome the results of the paper although they do highlight the incorrect spatial distribution of CO emissions from biomass burning that was less obvious using the coarse resolution. We have changed the manuscript to include these points as described in more detail below.

We have run GEOS-Chem at 2° latitude and 2.5° longitude using FLAMBE biomass burning emissions to test the transport in the model. Emissions of CO from FLAMBE appear to be broadly qualitatively consistent with GFED3 emissions. The model shows similar skill in capturing the variability of the biomass burning plumes with both emission inventories but FLAMBE has a larger positive bias particularly within air masses identified as biomass burning plumes. This bias might also be due to errors in atmospheric transport at high altitudes. The GEOS-Chem model distributes the GFED3 biomass burning emissions evenly though the boundary layer, without consideration of injection heights above the boundary layer which would subsequently affect transport of emitted CO (see below).

We have also run the tagged CO model using the native-resolution nested grid over North America for July 2011 to cover the BORTAS period. The nested model has similar skill to the coarser

resolution in reproducing all the BORTAS aircraft data (Pearson r^2 = 0.28). We find that the model has better skill at reproducing results at lower levels below 4 km (r^2 = 0.36) compared to higher altitudes (r^2 = 0.02). In the revised paper we argue that the Spearman's rank correlation is the more appropriate metric, reflecting our ability to identify plumes but not necessary reproduce the elevated values. We find that Spearman's rank correlation is 0.65 below 4 km and 0.19 above 4 km.

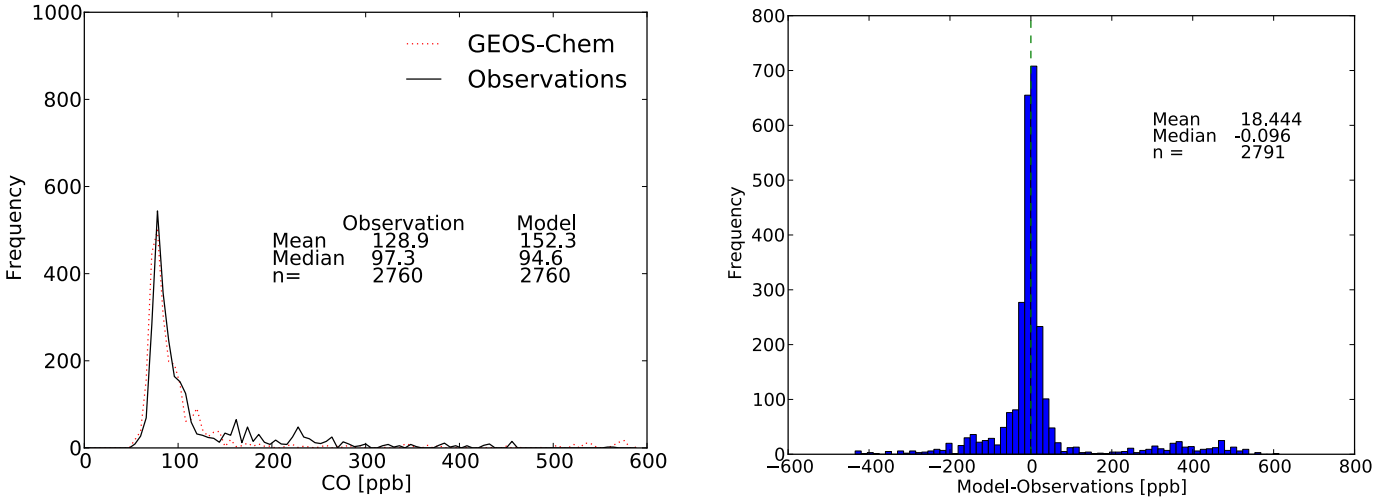

Figure 1 Statistical comparison of model and observed CO from BORTAS-B for the nested grid. The observations have been averaged over the 0.5 x 0.667 degree model grid. Left panel shows the frequency distributions and the right panel shows the frequency distribution of the model minus observed CO residuals. Mean and median values are shown inset of each panel.

We find that the nested model overestimates the CO in the plumes up to a factor of 5. This large discrepancy is due to overestimating CO emissions from the GFED3 emissions dataset. In the coarser model run, the high CO values on a fire hotspot are reduced, as expect, because of spatial averaging of small-scale features. The mean residual CO is 18.4 ppb with a standard deviation of 132.3 ppb and a median of -0.1 ppb.

Figure 2 shows the vertical resolution of this bias according to pressure bins. The largest discrepancy is between 750 and 850 hPa with a mean proportional difference of 2. There is a small positive model bias between the surface and 900 hPa and a small negative bias above 750 hPa. This strengthens the hypothesis that the model does not distribute CO emissions well.

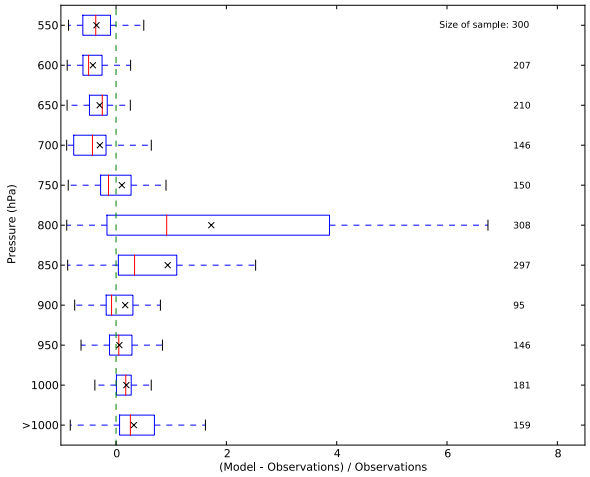

Figure 2 Relative model error in the GEOS-Chem simulation of CO during BORTAS-B as a function of altitude described by the box and whiskers approach. The red line and grey cross denotes the median and mean values, respectively.

I would like to see some actual data-model comparisons along the flight tracks

Figure 3 shows observed and model CO concentration along all the flight tracks combined for the nested resolution. The model reproduces background CO but do not reproduce the very highest values of CO associated with biomass burning. This bias is more prominent at the higher resolution, supporting our hypothesis that the distribution of CO emissions in the model is wrong. We have left this figure out of the manuscript because it does not add anything to the readability of the paper, especially given the additional text added as part of the revision (see above).

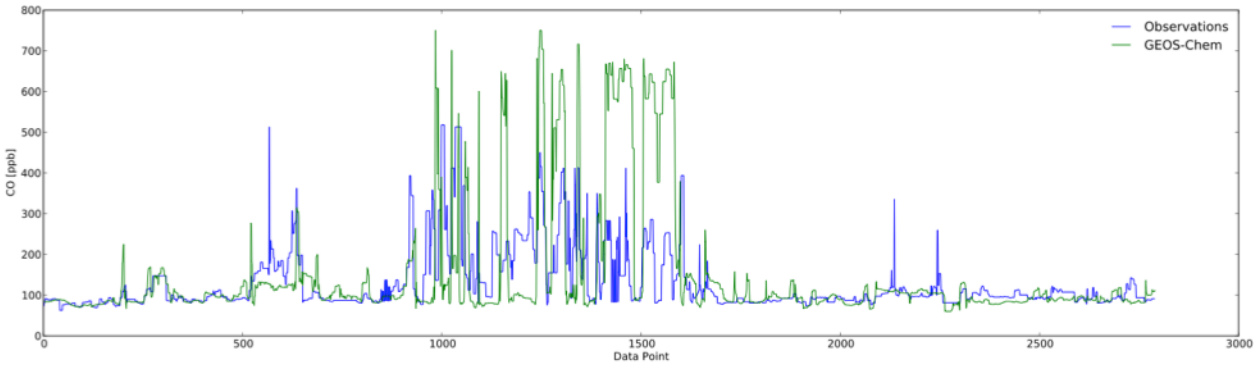

Figure 3 CO concentrations along all the flight paths during BORTAS-B from GEOS-Chem (green) and observations (blue).

It seems as if you are saying that even though your model has very little skill (r2 = 0.2) at reproducing individual observations, I should believe your conclusions because your median and mean are similar. I recommend you provide a more convincing justification.

This a fair point. See above response related to our revised calculations.

In short, using the higher resolution model for the age of air analysis strengthens our results. We now show the r-squared value of 'difference between photochemical and physical effective age' and 'effective age' to have risen from 0.60 to 0.80 inside the plume but only risen from 0.23 to 0.28 outside the plume. This strengthens our claim that the major reason for the difference in the two age measures is due air being within a BB plume or not (and therefore the levels of OH, most likely controlled by aerosol concentration). The age of air calculation will be affected by the incorrect spatial distribution of the CO, but not by the inconsistencies with the amount of emissions as the calculation does not take into account the size of the CO emissions. The updated plots will be included in the manuscript.

Referee #2

Consistency in referring to it as 'physical' or 'effective' would help

We thank the reviewer for raising this point. To clarify this point we have consistently renamed this metric the effective physical age (or EPA).

What effect would having a lower OH value have on your conclusions? Is it sensitive to OH?

All EPA values we report in our domain are well below the e-folding lifetime set by OH therefore diffusion is the dominant effect. We have run the analysis with a doubled and halved lifetime and the ages differ by less than one day. As stated in the paper (page 8735, line 13 -14), halving (doubling) the OH concentration would double (halve) the photochemical lifetime, but this was chosen by Parrington et al 2013 to reflect the photochemical environmental at that latitude and season.

If "correct" OH were used, would the effective age match the photochemical age? Is this a way of deducing mean OH?

We see your argument. The "correct" OH would be OH sampled along the flight track for the age of air calculation. The photochemical age calculation uses only one value of OH, representative of the background atmosphere, to underpin the whole calculation. The difference between the physical and photochemical ages would then provide information about the integrated OH concentration along the flight track, assuming that atmospheric diffusion played no role in mixing the air. We think the compounded errors associated with every stage of such a calculation would preclude a reliable estimate for mean OH.

*A brief description of how photochemical age was calculated would hel*p

In the interest of brevity we referred to the Parrington et al 2013 paper in our previous manuscript. However, we now nclude a brief description for readers as well as references with more a more detailed description of how photochemical age is calculated.

Contradictory *terms in abstract, p 8728 and 8733*

Thank you for pointing these out. This has been corrected.

Line 15 - 16: How can a median be a range?

This has now been changed to "median difference… 3 days" - This has also been updated to match the new higher resolution model run on the advice of reviewer #1

Page 8729 line 3: Most frequent is the mode.. why is it a range and not a single number?

This has been amended to the following statement: " The interquartile range for observed CO is 87--126 ppb with a mode of 90 ppb. This is consistent with background CO measurements during the NASA ARCTAS-B campaign…"

Page 8730 line 12: Should be "are due" not "is due"

This has been changed in the revised manuscript.

Page 8731 Line 11: add W after 50 degree

This has now been amended. We have also amended the definition of the latitude averaging area to the correct co-ordinates (45 - 60 degrees north) to match the figure.

Line 15 onwards: confusing section. … did not get the same numbers… some ambiguity in which numbers specifically you are referring to … suggest editing this section to make it clear and unambiguous

We have reworded the paragraph to make it clearer. The changes are stated below.

- (Line 15 - end of line 18) "At the western boundary, emission from 2008 have a median age of 7 days, 2009 has a median age of 10 days and 2010 and 2011 have a median age of 5 days. The age of air at the western boundary in 2009 has a larger age range than the other years with generally older air (from 2 to 20 days). We find the older age of emissions during 2009 is due to fewer fires along the western boundary so that sampled air originates from Alaska and further upwind."

- (Line 21 - line 22) "The eastern boundary for all years and at all altitudes have median values between 11 and 13 days and a range of 2 - 23 days."

- (Line 24 - line 28) “The age of emissions clearly gets older moving towards the Atlantic, consistent with westerly flow, with an increase in median age of approximately 3 days for all years except 2009, where the age difference between the west and east boundary is approximately one day in the lower troposphere and no differences for higher altitudes. We attribute the lack of ageing in 2009 to the less fresh emissions in Canada during July of this year. Boreal biomass burning in 2009 was seen earlier in the season and was more prominent in Alaska and Siberia.”

Page 8732 Line 10: give co-ordinates for Thunder Bay

We have added coordinates for Thunder Bay.

Page 8732 Line 11: *Does 16-17 July equate to 3-4 contour on the plot? It would be easier to follow if the text and the figure matched up.*

We agree with the reviewer on this comment. The text has been amended to match the colour bar for ease of comparison.

Page 8733 Line 1-2: explain “mean fractional difference”

What we mean by ‘mean fractional difference’ is the mean (or median in line 2) of the relative difference of age to the size of the effective physical age (age differences/physical ages) for both within and outwith the plume. We agree with the reviewer that this needs more explanation . We have been split the statement into two separate points to make our message clearer as well as updating the numbers to correspond with the new higher resolution nested results.

Page 8733 Line 18-19: First mention of sub-grid scale vertical mixing. Why do you come to this conclusion. Maybe better to explain earlier in the manuscript or and least explain it here.
The model bias can be partially explained by the sub-grid scale vertical mixing. The model injection height of the smoke plume, in the absence of other information, remains in the boundary layer while in reality the majority of pollutants could be injected into the free troposphere. This has implications for the subsequent atmospheric transport of emitted gases. We have now improved this explanation in the revised manuscript.

Page 8734 Line 12 - 15: have you described this HYSPLIT evidence earlier in the text? Explain this more clearly - We have not explained it earlier in the text. Here we simply use this as a comparison with a paper (Griffin et al, 2013) where HYSPLIT is used briefly to estimate the age of pollutants they observed. We have now explained this further in the manuscript in the concluding remarks.