**Response to Anonymous Referee # 4**

(Note: Reviewer comments are listed in grey, and responses to reviewer comments are in black. Pasted text from the new version of the paper is in italics.)

This work conducts numerous sensitivity simulations using a global chemical transport model (MOZART-4) by changing the aging timescale of the tagged BC from various geographical source regions. The authors optimize the aging timescale of BC emitted from each source region by minimizing errors in vertical profiles of BC mass mixing ratios between the simulations and HIPPO aircraft measurements. They investigate the contributions of BC from each source region to BC loading over each receptor region, and examine relationship between lifetime of BC and its aging timescale. This study is interesting and scientifically important. The subject is of great interest to ACP. For the most part, the manuscript is written clearly. However, the description and validity of the optimization approach in section 2.5 are not sufficient, which are cause for concern (see Major comments). Once these points are addressed satisfactory, the paper should in my opinion be suitable for publication in ACP.

Major comments:

We greatly appreciate the reviewer for detailed, valuable and constructive comments, especially those regarding the uncertainties of our study. The suggestions are extremely helpful to improving our manuscript.

My main concern is the validity of the optimization approach shown in section 2.5. In definition of the NMAE in equation (3), the authors normalize the absolute errors by the minimum of observed and modeled BC so that NMAE weights both high bias and low bias equally. However, I am not convinced that this approach can estimate the optimized aging timescale of BC. For example, we consider one data point for simplicity in the following two cases: Case A: BCo=4, BCm=2.5; NMAE = 1.5/2.5 = 0.6 (greater NMAE). Case B: BCo=4, BCm = 6; NMAE = 2/4 = 0.5 (smaller NMAE). According to this approach, case B shows better model results because of the smaller NMAE, although I think that case A shows better agreement between the modeled and observed BC values.

I guess that modeled BC (BCm) in equation (3) is sum of the 13 tagged BC from each source region, and contributions of the tagged BC from each region to the total BC over the HIPPO regions are largely different. I think that variations of the BC mass mixing ratio due to changes in the aging timescale are greater for the tagged BC from the large-contributed source regions (e.g., East Asia) and smaller from the small contributed regions. In this case, the approach may have limitations to estimate the optimized aging timescale for BC from the small-contributed regions, because the total BC mixing ratio is dominated by BC from the large-contributed regions. According to Table 1, the optimized aging timescales are 4 hours for BC from the large-contributed source regions (page 16959, lines 4-5) and are greater values (120-200 hours) from the small-contributed regions.

If I am misunderstanding the optimization approach, the authors should clarify their approach and discuss the validity of their approach.

The reviewer’s understanding towards the optimization approach is correct. To clarify the procedure, we expanded the description of the optimization approach in Section 2.5:

*“We perform 13 simulations, each with different constant aging timescales (i.e. 4, 8, 12, 18, 24, 27.6, 38.4, 48, 60, 90, 120, 160 or 200 hours). Every simulation tags BC from each of 13 regions (i.e., North America, East Asia, Canada, ...); as mentioned in Section 2.3,* $BC_m(j,k)=\sum_r BC_m(j,k,r)$*, where r denotes each region. We construct* $BC_m(j\ ,k)$ *using all possible combinations of* $BC_m(j,k,r)$ *from the 13 simulations. Then we check which combination of* $BC_m(j,k,r)$ *best matches BC observations. Note that we constrain the aging rates of BC emitted from Africa, South America, and Australia to be the same since these three regions are all biomass burning dominated sources in the Southern Hemisphere, which effectively reduces the total number of tagged tracers from 13 to 11. Thus, we determine the best-fit BC aging timescale for each source region (out of* $13^{11}$ *combinations in total) that minimizes MNAE.”*

We agree with the reviewer that there are multiple limitations in our original optimization scheme. As the reviewer mentioned, our modeled BC concentration is the sum of 13 regions which is typically dominated by only one or a few source regions for each place. Therefore, our optimization on BC aging is more accurate to these tracers emitted nearby than those experienced longer-distance transport. To differentiate the uncertainties of the optimized aging timescales for different regions, we add an additional table in the supplementary materials to represent the range of each optimized aging timescale (see Table S1 in the Supplementary material or below). For example, during HIPPO1, the optimized aging timescale for Middle Asian BC is 120 hours, but the uncertainty range is 4-200 hours, much broader than the aging timescale of EA tracer (4-4).

Table S1. Range of plausible aging timescales (units = hours) for 13 regions (For each HIPPO campaign, we assign a range of BC aging timescales for a region if the corresponding MNAE value is no larger than that of the optimized MNAE plus a small perturbation ΔE=0.01).

| | | CA | SU | EU | MA | EA | ME | NA | SE | IN | AF | SA | AU | RR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **HIPPO1** | Jan | 12-200 | 4-200 | 90-120 | 4-200 | 4-8 | 4-60 | 160-200 | 4-4 | 4-8 | 4-4 | 4-160 | 4-60 | 90-120 |
| **HIPPO2** | Nov | 90-200 | 4-200 | 120-160 | 4-200 | 4-4 | 4-60 | 4-24 | 4-8 | 4-8 | 4-90 | 90-200 | 90-160 | 160-200 |
| **HIPPO3** | Apr | 4-200 | 60-200 | 200-200 | 4-200 | 38.4-48 | 160-200 | 4-60 | 27.6-48 | 4-48 | 18-90 | 18-60 | 4-27.6 | 200-200 |
| **HIPPO4** | Jun | 38.4-90 | 4-18 | 120-200 | 4-200 | 4-8 | 4-200 | 4-18 | 4-8 | 4-48 | 4-38.4 | 4-120 | 4-8 | 200-200 |
| **HIPPO5** | Aug | 90-120 | 4-4 | 4-38.4 | 4-120 | 4-4 | 4-60 | 4-4 | 4-4 | 4-12 | 60-90 | 60-60 | 48-160 | 4-4 |

We agree with the reviewer that our optimized aging timescales are more robust for specific regional BC tracers than others, therefore we added following discussions on the uncertainties in optimized aging timescales in Section 6:

*“...Fourthly, the computed optimized aging rate is more accurate for tracers (i.e. source regions) with larger emissions and in closer proximity to the Pacific Ocean*

*(e.g. East Asia). This is because modeled BC concentrations over the Pacific (i.e. the location of HIPPO observations) for each latitude and longitude bin are typically dominated by only a few source regions, and the sensitivity of MNAE on each regional BC tracer is different (see Figure S5 in the supplementary material). For some source regions, observations in other remote regions would provide a better constraint for optimizing aging timescale in the model. More specifically, aircraft observations over the Atlantic Ocean could better constrain aging timescales for BC emitted from Africa and South America. As new observations become available, this study could be repeated to more accurately optimize the aging timescale for source regions with lower relative contributions to BC over the Pacific (e.g. Middle Asia)…"*

If possible, in addition to the present approach, it would be better to optimize the BC aging rate by other conventional approaches (e.g., mean normalized gross error, normalized mean error for each altitude range, taylor diagram, etc.) and evaluate the statistical values in a comprehensive manner, which could give the confidence in the validity of the estimates. If these evaluations are difficult, the authors should at least address limitations of their approach and discuss possible errors included in their results.

Excellent suggestion! In the previous example given by the reviewer:
"Case A: BCo=4, BCm=2.5; NMAE = 1.5/2.5 = 0.6 (greater NMAE). Case B: BCo=4, BCm = 6; NMAE = 2/4 = 0.5 (smaller NMAE). According to this approach, case B shows better model results because of the smaller NMAE, although I think that case A shows better agreement between the modeled and observed BC values."
We find that for this example, statistical values like mean absolute error normalized by observed value (MNAEo) and mean absolute error normalized by the average of observed and modeled values (NNAEa) will show a better agreement for case A than case B, the same as the reviewer expected. Following the reviewer's suggestion, we conduct additional tests with MNAEo and MNAEa, re-generated optimized BC aging timescales, and added the following tables to the supplementary material:

Table S2. Optimized aging timescales (units = hours) for 13 regions using $\mathrm{MNAE_o} = \frac{1}{\mathrm{N}}\sum_{\mathrm{nlat}}\sum_{\mathrm{nalt}}\frac{\mathrm{Abs}(\mathrm{BC_m}(j,k)-\mathrm{BC_o}(j,k))}{\mathrm{BC_o}(j,k)}$

| | | CA | SU | EU | MA | EA | ME | NA | SE | IN | AF | SA | AU | RR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **HIPPO1** | Jan | 60 | 60 | 120 | 60 | 4 | 12 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| **HIPPO2** | Nov | 200 | 90 | 60 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 8 |
| **HIPPO3** | Apr | 120 | 200 | 200 | 200 | 38 | 4 | 4 | 18 | 4 | 12 | 12 | 12 | 27 |
| **HIPPO4** | Jun | 18 | 4 | 200 | 4 | 4 | 200 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| **HIPPO5** | Aug | 8 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Table S3. Optimized aging timescales (units = hours) for 13 regions using $MNAE_a = \frac{1}{N}\sum_{nlat}\sum_{nalt}\frac{Abs(BC_m(j,k)-BC_o(j,k))}{(BC_m(j,k)+BC_o(j,k))/2}$

| | | CA | SU | EU | MA | EA | ME | NA | SE | IN | AF | SA | AU | RR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **HIPPO1** | Jan | 200 | 200 | 90 | 200 | 4 | 200 | 120 | 4 | 4 | 18 | 18 | 18 | 4 |
| **HIPPO2** | Nov | 200 | 200 | 48 | 200 | 4 | 4 | 4 | 4 | 4 | 90 | 90 | 90 | 160 |
| **HIPPO3** | Apr | 200 | 200 | 200 | 200 | 38 | 200 | 4 | 24 | 60 | 12 | 12 | 12 | 200 |
| **HIPPO4** | Jun | 18 | 8 | 200 | 200 | 4 | 160 | 4 | 4 | 4 | 4 | 4 | 4 | 160 |
| **HIPPO5** | Aug | 120 | 4 | 38 | 4 | 4 | 4 | 4 | 4 | 4 | 90 | 90 | 90 | 4 |

Comparing Table S2, S3 to Table 1, the optimized results are similar to those using the originally defined $MNAE_m$ which is normalized by the minimum of BCo and BCm, as shown below. Generally, they all show a longer aging timescale in high-latitude regions and a shorter aging timescale for other regions, although moderate differences do exist.

| | | CA | SU | EU | MA | EA | ME | NA | SE | IN | AF | SA | AU | RR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **HIPPO1** | Jan | 200 | 90 | 120 | 120 | 4 | 12 | 160 | 4 | 4 | 4 | 4 | 4 | 90 |
| **HIPPO2** | Nov | 200 | 160 | 160 | 90 | 4 | 4 | 4 | 4 | 4 | 90 | 90 | 90 | 200 |
| **HIPPO3** | Apr | 200 | 200 | 200 | 200 | 38 | 200 | 4 | 38 | 27 | 24 | 24 | 24 | 200 |
| **HIPPO4** | Jun | 60 | 4 | 160 | 12 | 4 | 160 | 4 | 4 | 4 | 4 | 4 | 4 | 200 |
| **HIPPO5** | Aug | 120 | 4 | 18 | 4 | 4 | 4 | 4 | 4 | 4 | 60 | 60 | 60 | 4 |

In general, we hope we can find a statistic which can judge model performance in both polluted and clean conditions since the latter usually cover a much broader spatial extent. Orders of magnitude overestimation or underestimation in these clean regions (though the absolute values are small) may also indicate large model biases which need further investigation. Thus, we normalize the absolute errors by the minimum of observed and modeled BC so that $MNAE_m$ weights both high bias and low bias equally. However, we agree with the reviewer that which error matrix we choose influences the optimization results and acknowledged this issue in Section 6:

*"…Thirdly, the optimized aging results may somewhat depend on the error matrix chosen. We conduct additional simulations with different error matrices (see Table S2, S3 in supplementary material). The results are overall similar, but in some cases moderate differences are found… "*

The authors chose the aging timescale between 4 hours and 200 hours in section 2.5. On the other hand, Table 1 shows that the best-fit aging timescales are 4 hours and 200 hours in many cases (minimum or maximum). I believe that conclusions of this paper (e.g., faster aging over the anthropogenic source regions and slower aging over high latitude regions) will not be changed even if the authors expand the range of the timescale, however the authors need to discuss the validity of their estimation.

We also agree with the reviewer that our approach do have limitations and the optimized results may partially depend on the performance of modeling processes beyond aging. In addition, this study only used 13 choices of aging timescales for regions, which cannot cover all the possibilities happening in the real atmosphere. Thus, every given number may actually represent a range. For example, '4h' and '200h' in Table 1 can mean ' less than 8h'(or fast aging) and 'greater than 160 hours' (or slow aging), respectively. Nevertheless, the goal of our aging optimization is not to provide a set of accurate aging timescales that can be directly used in models, but to utilize every HIPPO observation at most to inform the possible spatiotemporal pattern of aging timescales globally. Our results may indicate that the aging rate of BC may change by region and season, as opposed to a fixed aging timescale currently widely used in many chemical transport models. In addition, we extensively expanded the caveats on Section 6 to address the uncertainties of our approach:

We followed the reviewer's suggestions and added more discussion in Section 6:
*"We note that there are multiple limitations to our approach. Firstly, we assume that model parameterizations of wet and dry deposition, precipitation, transport, and emissions are realistic, even though these processes also affect BC distributions and have uncertainties (Vignati et al., 2010;Fan et al., 2012). Consequently, the optimized aging timescales may partially counter biases in these processes (i.e. other than aging), and may vary according to the model used. For example, as model resolution increases, aerosol-cloud interactions in climate models can be better resolved, which can improve the simulation of BC transport (Ma et al., 2013;Ma et al., 2014). Therefore, the optimized aging timescales might change if models with different cloud schemes or spatial resolutions are used. Secondly, due to limitations in computing resources, we carry out simulations assuming 13 discrete values for aging timescale Optimized aging timescale could have been more precisely determined with more simulations… The goal of the optimization presented here is not to provide precise aging timescales that can be directly used in models, since models differ significantly in their parameterizations of physical and chemical processes, particularly the wet scavenging. Also, BC aging includes complicated chemistry and physics, but is simplified in our modeling as a first-order conversion from hydrophobic to hydrophilic BC. Nevertheless, this study proposes a useful method to utilize all HIPPO observations and explore the spatiotemporal pattern of BC aging timescales globally."*

The authors should improve the description of the optimization approach in section 2.5. Please clarify how the 13 tagged BC are used in equation (3).

We expanded the description of optimization approach in section 2.5:
*"We perform 13 simulations, each with different constant aging timescales (i.e. 4, 8,*

*12, 18, 24, 27.6, 38.4, 48, 60, 90, 120, 160 or 200 hours). Every simulation tags BC from each of 13 regions (i.e., North America, East Asia, Canada, ...); as mentioned in Section 2.3,* $BC_m(j,k)=\sum_r BC_m(j,k,r)$, *where r denotes each region. We construct* $BC_m(j,k)$ *using all possible combinations of* $BC_m(j,k,r)$ *from the 13 simulations. Then we check which combination of* $BC_m(j,k,r)$ *best matches BC observations. Note that we constrain the aging rates of BC emitted from Africa, South America, and Australia to be the same since these three regions are all biomass burning dominated sources in the Southern Hemisphere, which effectively reduces the total number of tagged tracers from 13 to 11. Thus, we determine the best-fit BC aging timescale for each source region (out of* $13^{11}$ *combinations in total) that minimizes MNAE."*

Please describe the time information for the modeled and observed BC. Are these compared the same time? Please indicate time resolution of the observation data and model output (hourly, daily, monthly?) used in equation (3).

We compared the daily model output to the HIPPO observational data. We have clarified this issue in the revised manuscript. Please see section 2.5 or the revised text below:
*"The model output daily averaged BC mixing ratios. For every record in HIPPO data (averaged in every 10s), we find modeled BC mixing ratio at the same longitude, latitude, altitude, and on the same day correspondingly. In this way, modeled and observed BC mixing ratios are paired, and then are averaged respectively over latitude and altitude bins. "*

Additional comments:
Page 16951, line 14, "with 28 vertical levels": Please indicate the top boundary of the model..

We added the corresponding information in the revised manuscript (i.e., Section 2.1):
*" ...The model is run at a horizontal resolution of approximately 1.9 °×1.9 °(latitude × longitude) with 28 vertical levels from surface to approximately 2 hPa... "*

Page 16951, lines 15-16: Is MACCity inventory (Lamarque et al., 2010) used for the CMIP5 project?

According to the website http://accent.aero.jussieu.fr/MACC_metadata.php, "As part of two projects funded by the European Commission, MACC and CityZen, the ACCMIP and the RCPs emissions dataset have been adapted and extended on a yearly basis for the period 1990-2010. For anthropogenic emissions, emission data were interpolated on a yearly basis between the base years 1990, 2000, 2005 and 2010. For the years 2005 and 2010, the RCP 8.5 emissions scenario was chosen (4 emissions scenarios were developed in support of the IPCC-AR5: RCP 2.6, RCP 4.5, RCP 6, RCP 8.5). This 'extension' of the ACCMIP and RCPs emission dataset for the

MACC and CityZEN projects is referred to as MACCity (MACC/CityZen) emission dataset.” We modified the sentence (in Section 2.1) to make it more precise:

*“...Anthropogenic emissions are based on the MACCity emission inventory (http://www.pole- ether.fr/eccad), which is extended from the database used for IPCC Coupled Model Intercomparison Project (Lamarque et al., 2010). ”*

Page 16951, lines 9-10: Does "C" include ice content?

Yes, C is the sum of ice and liquid water content. We revised this sentence in Section 2.2 to make it clearer:

*“…C is the sum of cloud ice and liquid water content (kg $m^{-3}$)…”*

Page 16954, lines 12-13: Please show a few references at the end of the sentence.

Thanks, we added some references at this place, namely:

*Y., Shiraiwa, M., Gong, Y. G., Shao, M., Hu, M., Zhu, T., Zhang, Y. H., Carmichael, G. R., Wiedensohler, A., Andreae, M. O., and Poschl, U.: Size-resolved measurement of the mixing state of soot in the megacity Beijing, China: diurnal cycle, aging and parameterization, Atmospheric Chemistry and Physics, 12, 4477-4491, DOI 10.5194/acp-12-4477-2012, 2012.*

*Moteki, N., Kondo, Y., Miyazaki, Y., Takegawa, N., Komazaki, Y., Kurata, G., Shirai, T., Blake, D. R., Miyakawa, T., and Koike, M.: Evolution of mixing state of black carbon particles: Aircraft measurements over the western Pacific in March 2004, Geophysical Research Letters, 34, Artn L11803*

Page 16954, line 28: Schwarz et al. (2008"b") is not shown in the reference list.

Thanks for catching this mistake. We changed the citation here to *(Schwarz et al., 2008),* which is now in the reference list.

Page 16958, lines 6, Figure S1: Typo? Figure S3 in the present manuscript?

Thanks for catching this typo. It should be Figure S3 in the previous manuscript. We have fixed it.

Page 16958, lines 14-21: In the Southern Hemisphere during HIPPO5 (Figure 4), the improved BC (green) is smaller than the observed BC (black), however the original BC (red) is greater than the observed BC. If the model does not capture abrupt biomass burning emission events, the original BC would also be smaller than the observed BC. I am not satisfied with the author’s explanation, because differences in modeled BC values (green and red) are caused mainly by the wet deposition of BC.

Excellent point! We agree with the reviewer that factors such as wet removal also significantly influence the BC distribution. Since this study assumes BC aging timescales in all the southern hemispheric countries are the same, it may not be able to differentiate variability of BC aging across regions in the southern hemisphere. In addition, the optimized aging in these regions exhibit large uncertainties and may require additional aircraft campaigns over the Atlantic Ocean as well as the Indian Ocean to better calibrate the BC timescales in the southern hemisphere. In the revised manuscript, we followed the reviewer's suggestion and revised the whole paragraph. Now we have the following discussions in Section 3:

*"In a few cases, relatively large differences between the improved model and observations remain. These differences could be attributed to any number of factors (e.g., emissions, transport, cloud/precipitation, aging process, wet removal efficiency, etc.). For example, models could misrepresent BC wet deposition, originating from biases in precipitation. As shown in Figures S4 in the supplementary materials, though MOZART-4 generally captures well the spatial extent of precipitation during all HIPPO campaigns, biases occasionally appear when comparing to the NCEP reanalysis over the western Pacific. As another example, the model uses a monthly biomass burning emission inventory. This means that modeled emissions lack daily variation in biomass burning activities that could be important where biomass burning emissions dominate BC loading. Underestimates in BC mixing ratio may be partially due to abrupt emissions events that are not captured by the model. Lastly, since this study assumes that BC aging timescale in all the southern hemispheric continents is the same, we do not account for variability in BC aging rates from these regions that may exist in reality. "*

Equation (3) and Table 1, terminology: "Normalized mean absolute error (NMAE)" should be "mean normalized absolute error"? "Normalized mean bias (NMB)" should be "mean normalized bias" or "mean fractional bias"?

Thanks for correcting this. We have corrected all the uses of these terms in the paper. Now we are using mean normalized absolute error (MNAE) and mean normalized bias (MNB).