Response to Anonymous Referee # 2
(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 paper presents a global modeling analysis to constrain the first-order aging timescale of black carbon based on observations from the HIPPO campaign. The analysis involves performing several sensitivity studies where the aging time scale was varied, with the BC tracer tagged according to different geographic source regions. Optimal aging time scales for each source region are then found by minimizing the error between simulated BC mixing ratios and HIPPO observations. The tagging of BC also allows quantifying the contribution of BC from different source regions to various receptor regions, including the Pacific Ocean, which is an area of interest as BC over this area is suspected to have significant climate impacts. This is an interesting paper, which makes innovative use of HIPPO data. It supports previous studies that found that the first-order aging time scale that is used in many global models should not be a fixed value, but should depend on local conditions. The paper fits well into the scope of ACP and is for the most part well-written. The discussion in Section 5 is very instructive. My main comments concern the optimization procedure in Section 2.5. I hope the reviewers can resolve these in a revised version.

We greatly appreciate the reviewer's thorough and constructive review. We believe revising the paper according to the reviewer's comments has considerably improved the paper. We have merged all of the suggestions into the revised manuscript. Please see our response to each comment below:

Major Comments:
The description of the optimization procedure in Section 2.5 needs to be improved (line 22 – 25). It sounds like the authors performed 13 simulations, with constant aging time scale for each of these simulations. It is not clear how the area-specific aging time scales are obtained from these 13 simulations. I believe that the constraint is used that BC(i; j; t) = Pnsource k=1 BC(i; j; t; k), and then BC(i; j; t) is reconstructed using all possible recombinations of BC(i; j; t; k) from the 13 sensitivity runs. Finally, it is checked which BC(i; j; t) best matches the observations. Please clarify this procedure.

Thanks for this great suggestion. The reviewer's description is correct. We have clarified the description of the optimization procedure following the reviewer's suggestion. Please see our revisions in Section 2.5 in the revised manuscript or below:

"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."

If this is what happens, my main concern is how stable the procedure is, i. e., given the large number of permutations to calculate candidate BC(i; j; t) values, it could happen that many different combinations of BC(i; j; t; k) lead to a similarly small error, and that the authors are fitting noise. One way to check this would be to use a testing and a training set, which might not be possible given the limited amount of observations. Another way to check this would be to visually inspect plots where the error is graphed as a function of parameter that is varied, keeping all other parameters constant. It should be very obvious if these curves look sufficiently smooth so that a robust minimum can be identified. The values listed in Table 2 do look questionable: two thirds of these values are either 4 or 200, which are the minimum and maximum values in the set of aging time scales used for the sensitivity runs. This could mean that the range of aging time scales chosen was not large enough. The authors discuss the physical interpretation of the optimized aging time scales on page 16957, but there are several examples that are hard to interpret. For example, for SU the time scale is 200 h for June and 4 h for August. Several other examples along this line can be found.

Excellent point. Following the reviewer's suggestion, we checked the sensitivity of the summed error to aging timescale changes. Specifically, we varied aging timescale in each region while keeping the aging timescale of other regions as previously optimized. We plotted the error versus the varied aging timescales for different regions, and added Figure S5 in the supplementary material:

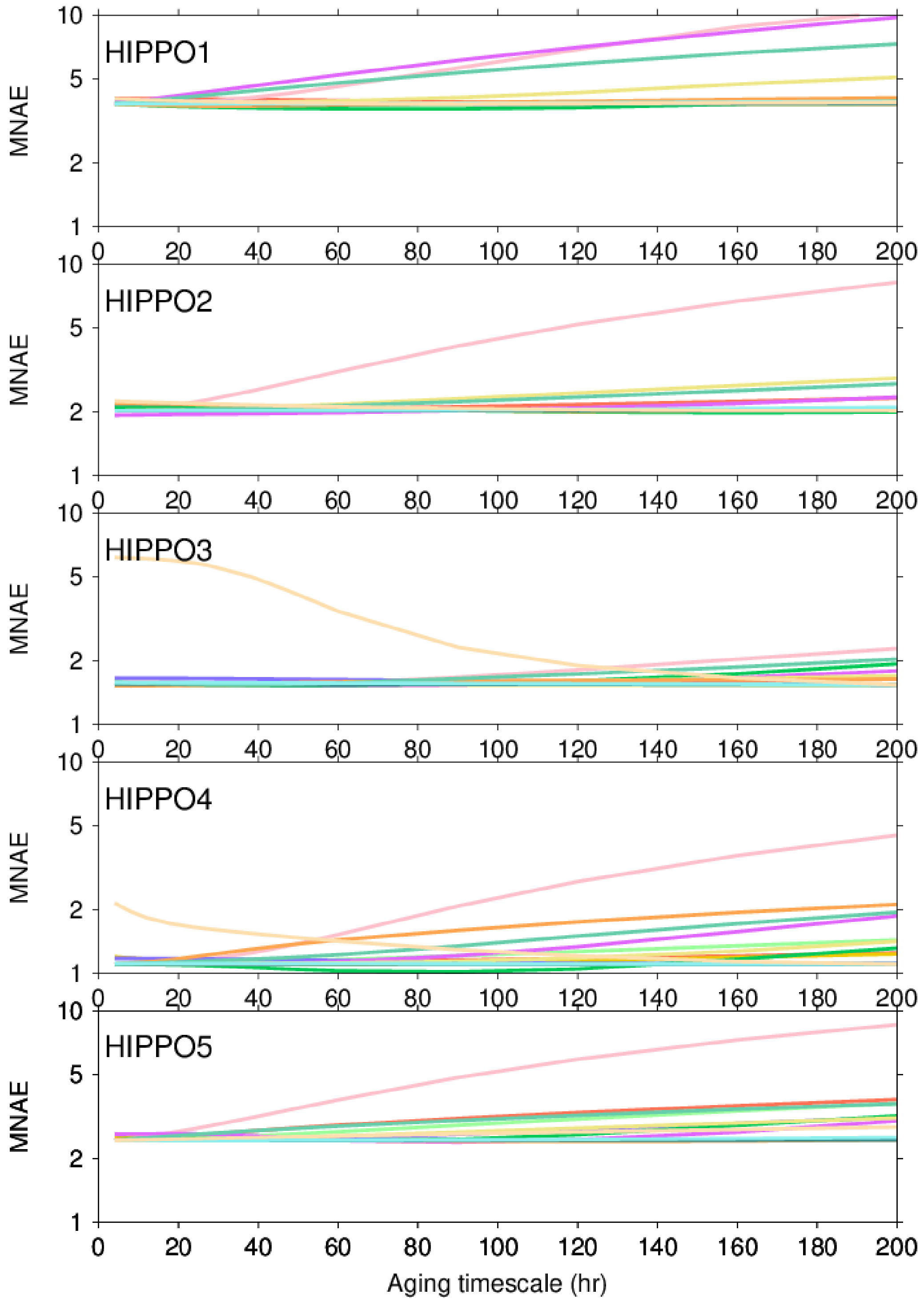

Figure S5. Mean normalized absolute error (MNAE) as a function of varying aging timescale for each region while keeping the aging timescale of other regions as optimized. Different colors denote 13 regions.

As shown in Figure S5, all curves are smooth. For some regions, it is easy to identify a robust minimum, while for some other regions the sensitivity of MNAE to aging timescale change is small. Therefore, our optimization on BC aging may be more robust to regions with significant contributions to BC burdens than insignificant contributors. The difference in sensitivity of MNAE on aging timescales in different regions inspires us to estimate the potential range of our results based on the information in Figure S5. 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. A narrow range may indicate that this optimized aging timescale is more robust than that with a broader range. Table S1 in the supplementary material (or below) illustrates the ranges for each optimized aging timescale when 0.01 is used for ΔE. We found that our optimized aging timescales are more robust for specific regional BC tracers (e.g. EA, IN, etc.) than others (e.g. MA, SA, etc.).

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

In the revised manuscript, we also include the following discussion 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)…"

We also agree with the reviewer that our approach do have limitations and the optimized results may partially depend on 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 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. 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. 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). 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."

Minor comments:

1. Rearrange the order of the columns in table 1 and 2, so that they show the source regions in the same order, to facilitate the comparison of the two tables for the reader.

Thanks for pointing out this mistake. We have changed the order of the columns in Table 2 to make it consistent with other tables.

2. Section 2.2: How do you know that using the updated dry and wet deposition schemes results in an improvement of the model performance?

Good question. First, the original model treats these processes in simplified, and to some extent arbitrary ways. The first order wet scavenging rate for BC is set to 20% of that for nitric acid, and dry deposition velocity of BC is fixed at 0.1 cm s^{-1} everywhere (Emmons et al., 2010). On the contrary, our employed deposition schemes are more physically or experimentally based, and have been shown to work well in other similar models by previous studies (Liu et al., 2011;Wang et al., 2011;Hodnebrog et al., 2014). Second, after implementing these schemes in MOZART-4, we also evaluated the model and found that MNAE is at least reduced by a factor of 2, if compared to the original model.

3. p. 16953, line 27: "approximately equal": Please quantify this statement.

We quantified the relative difference between the sum of tagged BC and untagged BC at the surface and at 500hPa. The relative difference is in most cases less than 1% with the largest bias less than 4%.

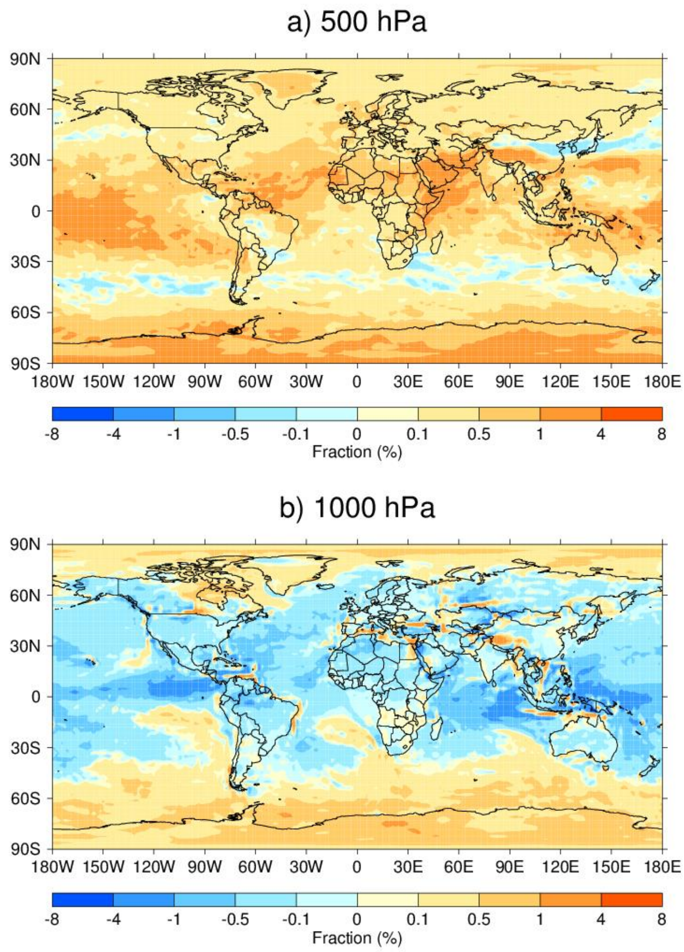

Figure R1. The relative difference between the sum of tagged BC and untagged BC a) at 500hPa and b) at 1000 hPa.

We follow the reviewer's suggestion and quantitatively described the difference in the revised manuscript (see below):

"The relative difference between the sum of the 13 regional BC tracers and the untagged BC is small (i.e., in most cases less than 1% with the largest biases less than 4%). Therefore, the sum of the 13 regional BC tracers is approximately equal to the untagged BC."

4. "Normalized mean absolute error" should be "Mean normalized absolute error"
Thanks for correcting this. We have corrected all the uses of this term in the paper. Now we use Mean normalized absolute error (MNAE).

5. In equation (3), are the simulated and observed BC values taken at the same time?
We compared the daily model output to the HIPPO observational data. We have clarified this issue in the revised manuscript in Section 2.5:

“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.”