In the following, we repeat the reviewer's comments in normal font and give responses in *blue italic*.

Response to Anonymous Referee #2

In this study, the authors apply the radioactive tracer-based constraint obtained by Kristiansen et al. (2012) to assess the aerosol lifetimes simulated by 19 global models. They find that modelled lifetimes vary over a wide range but are generally too short. To interpret their constraint, the authors propose to represent aerosol lifetime as a succession of increasingly long timescales related to the strengths of aerosol sinks, themselves a function of altitude.
The paper is interesting and well-written. It falls short of pinpointing the exact causes of model diversity in aerosol lifetimes, but demonstrates the usefulness of their constraint, which should help modellers improve their aerosol schemes in the future. Figures and Tables are informative and well-chosen.
In my comments below, I recommend clarifying the discussion in places to avoid seemingly contradicting aspects of the analysis and to support the reader in understanding the implications of the results. Addressing those comments should not amount to more than minor revisions.
We thank anonymous reviewer#2 for the positive feedback on our paper.

1 Main comments

• The study is not interested in emission injection heights. In section 3, this is justified by citing Croft et al. (2014), but that study in fact finds otherwise: to quote their abstract, "Global mean lifetimes are shown to strongly depend on the altitude of injection." So the statement on line 26 page 24521 is incorrect and the lack of interest in injection height (and the lack of a corresponding discussion in section 6.2) is unjustified. This silence is difficult to reconcile with the analysis in sections 6.1 and 6.3: if removal timescales, and specifically the first, τ1, are so crucially dependent on the vertical location of the aerosols (page 24536, lines 17–20) then surely the injection height must matter?
When it comes to the altitude-dependency it is important to consider the differences between global mean lifetime and e-folding lifetime, as discussed in depth by Croft et al (2014). The reviewer is correct that they found that the global mean lifetimes are indeed dependent on the injection altitude. However, they also state that "... the simulated e-folding times were relatively insensitive to emission parameters (altitude, location and time)." Thus, the two different definitions of lifetime were not equally dependant on the injection height. In our paper (Line 26, P24521 in the ACPD version) we clearly specify that the e-folding lifetime was unaffected by emission altitude, which is correct and justified by the findings of Croft et al 2014.

We further clarify the differences between the two lifetimes in our paper in section 4 (Methods): "They [Croft et al 2014] also show that the global mean aerosol lifetime strongly depends on the altitude where emissions are assumed to occur, while the e-folding lifetime is relatively insensitive to emissions parameters such as altitude, location and time, suggesting that e-folding times allow a robust comparison between modelled and measurement-based lifetimes".

In our study we focus on the e-folding lifetime over the time-scale τ2 and our results are derived from this lifetime definition, which is shown to be relatively insensitive to the emission altitude. However, for τ1, this is indeed relevant, but this time-scale is not the focus of this study.

In section 6.1 we also discuss the modelled global mean lifetime as well as τ1. But since we do not have measurements that can be used to evaluate modelled global mean lifetimes, or τ1 we do not focus on these lifetimes. It is still of interest to us to discuss and show on a model-model comparison perspective the differences between e-folding and global mean lifetimes, like in Section 6.1.

Throughout the paper we consistently specify which definition of lifetime (e-folding or global mean lifetime) we are referring to, and hence we feel the differentiation between global mean lifetime and e-folding lifetime is made clear. We also specify that we are focusing on time-scale τ2 and not on τ1. Thus, it should be clear why altitude dependency is not of interest in this study.

• Figure 7 is also interesting to look at within the multiple timescale framework of section 6.1. However, there seems to be a mismatch between the timescales one could derive from Figure 7 and those quantified in Table 4. In Figure 7, it is remarkable that the models all show a large variability in lifetime before day 40, followed suddenly by a well-behaved increase in lifetime with time. That sudden change cannot be a signature of different horizontal patterns, so it must be due to the vertical distribution, but why such a good agreement on a 40-day timescale? And why doesn't that timescale appear more clearly in Table 4?
The occurrence of the less variable steady state is related to the emissions. Between day 10-40 there are low-level emissions of ^{137}Cs, about one order of magnitude less than during the highest emissions around day 5. Because the emission was a point source (and not global) and its geographical distribution is rather local, one continues to see the impact of single deposition events, making the instantaneous lifetime fluctuate considerably. On day 40 all emissions of ^{137}Cs end in the model simulations. The PBL is soon emptied from most ^{137}Cs, and the remaining ^{137}Cs (mainly in the free troposphere) is more widely spread and gives rise to a more stable instantaneous lifetime, and thus a quasi-steady state with less variability.

We included the sentences above to the second paragraph of Section 5.3 to make this clear.

The lifetimes in Table 4 represent average (mean or median) lifetimes over certain time periods and therefore are not picking up the sudden change around day 40 related to the burdens.

• The author reserve a special treatment to high-latitude stations, perhaps because of the importance of transport to the Arctic. They imply several times, most clearly on Page 24529 lines 14–16, that transport to those high-latitude stations somehow tell us more about wet deposition. But is that really the case? One reason could be that those stations are reached after a long transport, but arguably Ulan Bator is the most remote station when accounting for transport pathways (Figure 2). Another reason is that extra-tropical precipitation frequency and intensity are more difficult to model accurately. Perhaps, but 75% of total precipitation is in the Tropics. So do high-latitude stations really tell us something that other stations cannot?
We are taking a closer look at the high-latitude stations in relation to previous studies which have shown that models struggle to capture Arctic aerosol concentrations, and particularly their seasonality, and this at least partly seems to be related to model shortcomings in removal processes. We introduce this topic already in the Introduction. We do not mean or say that looking at high latitude stations tells us more about wet deposition in general. So we agree with the reviewer that, perhaps, the Arctic stations do not tell us something that the other stations can't. However, we still find this an interesting topic to discuss.

2 Other comments

• Page 24517, lines 16–17: It would be the right place to remind modellers of the difference between lifetime and residence time, the latter being the ratio of burden to deposition/emission/production. It is unfortunate that modellers never use the proper terminology.
We agree with the reviewer and have added two sentences:
"Several definitions of lifetime and residence time (the ratio of burden to deposition/emission/production) exist but the terminologies are often used inconsistently. We encourage future studies to give clear information about which lifetime definitions that are used."

• Page 24517, line 27: the “while” is not strong enough to convey the contradiction between the results of Samset et al. (2014), which hint at the need for shorter lifetimes in models, and comparisons against observations in the Arctic, which suggest the need for longer lifetimes. As stated later in the introduction and in section 6.6, the regional dependence of lifetime likely explains the contradiction, but it means that comparisons are to be used cautiously when making statements about the quality of simulated lifetimes in models.
We agree with the reviewer that the wording was not strong enough to highlight the contradiction. Therefore, we have rephrased this: “Thus, a short aerosol lifetime appears necessary to reproduce such observations (e.g., Samset et al., 2014). On the other hand, the models generally underestimate the aerosol concentrations closer to the surface, and this would get worse with a shorter model lifetime.”

• Page 24518, line 26: Is there a reference for the fact that Cesium uptake by aerosol is proportional to surface area? Aren’t soluble particles preferred in any way?
Papastefanou (2008) explains in detail the attachment of radioactive decay products to aerosols particles. The book is available here: http://file.zums.ac.ir/ebook/281-Radioactive%20Aerosols%20(Radioactivity%20in%20the%20Environment,%20Volume%2012)-Constantin%20Papastefanou-00.pdf, or here https://books.google.co.uk/books?id=kKtr7CH9kvMC&printsec=frontcover#v=onepage&q&f=false See in particular section 3.3. Attachment of radon decay products to aerosol particles (page 17-19). Similar processes to those of radon are likely to occur for cesium. The process is also detailed elsewhere in the book. We have added the abovementioned reference to the manuscript.

• Page 24521, line 20: The description of model experiments is not sufficiently detailed. What did you ask modellers to do exactly? What was the emission rate? For how long are the emissions imposed? Where are the emissions injected? How long are the simulations?
The emission are detailed in Stohl et al (2011), and available as a supplementary to that paper (http://www.atmos-chem-phys.net/12/2313/2012/acp-12-2313-2012.html). The modellers were asked to use exactly those emissions.
We have added a few sentences to the paper in the particular paragraph to better describe the emissions:

“The simulations extended from 11 March until at least 5 June 2011 for when the last measurement of the radionuclides were taken. A total of 36.6 PBq of ^{137}Cs and 15.3 EBq of ^{133}Xe were released by all models. The emission rates vary significantly over the emission period considered (11 March – 20 April 2011).” “The releases were divided into three vertical layers between 0–1000 m above ground level (Stohl et al., 2012).”

• Page 24523, lines 14–26: This paragraph also looks like a good opportunity to formalise the difference between lifetime and residence time.
We are not sure what the reviewer means here. Lifetime is the more general term and can include processes such as chemical transformations or radioactive decay. However, when referring to the atmosphere as a reservoir, lifetime and residence time are typically used as synonyms. For instance, a common definition is: “The average time a molecule spends in a reservoir (such as the atmosphere) is called the lifetime or residence time”. We define the lifetimes used clearly in our paper and do not rely on semantic differences.

• Page 24526, lines 24–26: I do not follow the reasoning here: since the radioactive tracers must end up deposited at the surface, using ground- based measurements should not be a problem when assessing lifetimes, irrespective of how high the tracer goes during its time in the atmosphere.
Yes as long as the measured and modelled data to derive the lifetime are the same – both ground-based as here- it is no problem to assess the lifetime from these data and the lifetime should be

comparable. The issue we are trying to explain here is not to defend using such lifetime estimates, but rather as we say earlier "Both the measurements and the median model give slightly longer lifetimes (by about 0.3–1.0 days) for the earlier phases than for later time periods". One could think of this as illogical and could expect longer lifetimes for later time periods due to transport in the upper troposphere. But since we are here using only ground-based data (both from model and measurements) we might not expect to see the same increase in lifetime with time.

• Page 24528, line 7: At least as long as models are driven by or nudged to analyses... Free-running models are not assessed in this study.
As Table 1 highlights, two models are free-running, given as "internal meteorology" (NorESM, ULAQ-CCM).

• Page 24529, line 4: How to interpret that temporal trend? Is it due to an accumulation of the underestimation, i.e. if an aerosol had to be deposited before time T, then obviously it should also have been deposited before time T+1?
We are not sure what the reviewer means here. Obviously, the statement about T and T+1 is correct, but we do not understand what the reviewer is asking us to clarify here.

• Page 24529, line 13: It is worth mentioning here that both horizontal and vertical spatial dimensions matter.
We included "temporal and spatial occurrence and dimensions".

• Page 24537, section 6.4: This section is speculative because the authors lack coordinated sensitivity experiments by all models, e.g. doubled or halved wet deposition rates, emitting in September instead of March, etc. This is unfortunate, but probably too late to fix. It is now up to individual modellers to run those sensitivity experiments in their model development cycle.
We agree with the reviewer that this section is slightly incomplete as it does not include experiments by all models, however results are shown from four models which span the range in modelled lifetimes . The section is meant to bring some ideas for further model experiments and we hope modellers will take on this task using their individual models.

3 Technical comments
• Page 24522, line 14: Missing hyphen "decay-corrected"
Corrected.
• Page 24525, line 24: Replace "of" with "is"
With "is" the sentence does not read correctly, so we have kept "of".
• Page 24526, line 1: "a too quick removal": too quick a removal
Corrected.

Figures 4 and 5: Rigorously speaking, those are Tables, not Figures.
We agree but we would like to keep the color-coding in the table which we think is a useful addition, and hence these needed to be converted to figures.

Figures 4 and 5: The choice of colour is not great: green/red are widely used to denote "ok/not ok", whereas here they both denote "not ok". Blue/red would be a more usual choice to indicate an under/over-estimate.
We have changed the color-coding to blue/red.