# Authors’ response to referee comments on Kipling et al., “Constraints on aerosol processes in climate models from vertically-resolved aircraft observations of black carbon” (doi:10.5194/acpd-13-437-2013).

The authors would like to thank the two anonymous referees for their time and helpful comments on this paper, which we have taken on board and aim to address in a revised manuscript. The referees’ original comments are reproduced in blue italics below, with our responses indented in normal type underneath. Text added to the manuscript is shown in bold.

## In response to Anonymous Referee #1:

### *Major comments:*

*The authors have taken great care in conducting a point-by-point comparison of the model with the observations along the flight tracks for a large region of the Pacific. The authors state that this is a powerful tool for evaluating BC, whereas previous studies have compared monthly and grid mean model mixing ratios with aircraft profiles. A presentation of these comparisons under the previous methodology relative to the new methodology could help to document and support the argument that this new approach is worth the additional effort.*

> This is a very good point. The figure below shows (in black) BC MMR profile curves from HIPPO-1 based on the profiles identified for our burden analysis described in Section 4.1, averaged over the profiles in four of the latitude bands used in Schwarz et al. (2010). Note that these curves are not identical to those in that paper, because we use a different vertical-profile detection algorithm (and we have omitted the 67°S–60°S band because our algorithm identifies too few profiles there to calculate meaningful statistics).

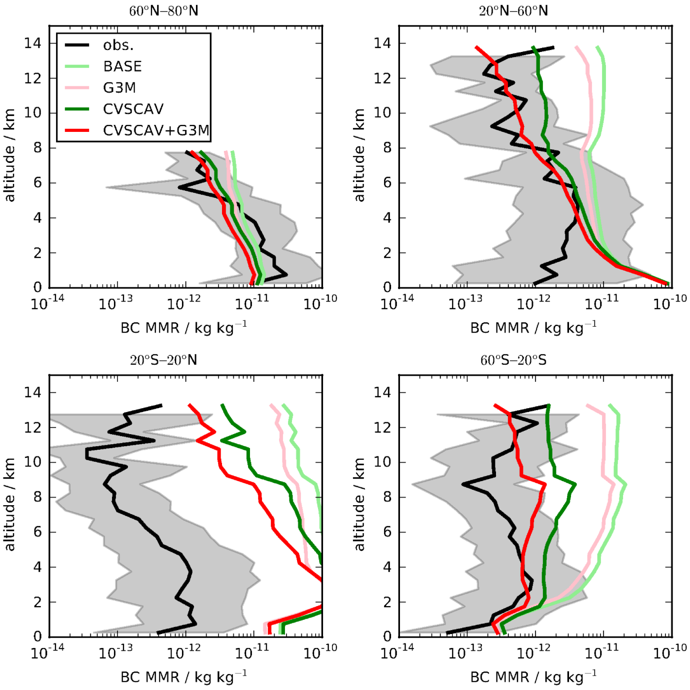

**Fig. 6.** Vertical profile curves of BC mass mixing ratio for HIPPO-1 and horizontally-matched locations in the January 2009 monthly-mean output from the HadGEM3–UKCA simulations. The shaded region shows the (geometric) standard deviation of the observed MMR values over the profiles in each latitude band, plus the ±30% measurement error.

The coloured curves show the profiles from the January 2009 monthly mean from each nudged HadGEM3–UKCA simulation, interpolated horizontally to the location of each profile and then averaged over the profiles in each latitude band.

As in our quantitative analysis in the manuscript, this is done in log-space: the curves represent the geometric mean over profiles, while the grey shaded region represents the geometric standard deviation over the observed profiles in each latitude band, plus the quoted 30% measurement uncertainty.

While it is possible to see improvements in certain regions from both CVSCAV and G3M, these appear much less clearly than in our point-by-point analysis, and in many cases are overshadowed either by the variability between profiles within a region, or by overall regional biases (e.g. in the tropics).

We have added this as Fig. 6, with the following explanation in a new section (lines 813–835 of revised manuscript):

> **6.4 Comparison with profile curves**
>
> **To compare the point-by-point analysis presented here with a more traditional approach, we have constructed profile curves from the HIPPO-1 observations for four latitude bands using the (geometric) mean and standard deviation over all the profiles identified for the burden analysis (as described in Section 4.1) in each latitude band. We have also constructed corresponding curves from the January 2009 monthly-mean output from the HadGEM3–UKCA simulations, by horizontally interpolating to the location of each profile identified in the observations.**
>
> **The results are shown in Figure 6. Although the construction is similar to that in Schwarz et al. (2010), the curves are not identical due to the different profile-detection algorithm used. While some improvement from both CVSCAV and G3M can be seen in these curves, this is rather less clear than in the point-by-point analysis. In many cases, the differences are overshadowed either by the variability between profiles within a region, or by overall regional biases.**
>
> **This demonstrates the usefulness of the point-by-point analysis presented here in allowing us to evaluate process-level changes to the model with rather more confidence than can be obtained from a more traditional approach.**

*In evaluating a global model, there is the possibility that compensating errors can occur, which can yield a closer agreement with observations for the wrong reasons. Thus, while these vertically-resolved aircraft observations provide a constraint on aerosol 'profiles', it is more difficult to argue that they provide a constraint on aerosol 'processes'. The authors should address this concern in the manuscript.*

– see also (19) and (22) below.

It is certainly true that compensating errors can be a problem when we look at potential improvements to processes in a global model. This is especially true when introducing a more physically-realistic parametrisation of one process into a model with significant parametric uncertainty: without adjusting other parameters to compensate, the model performance may actually worsen even though the new parametrisation is more realistic.

In this case, it would not have been surprising if the in-plume convective scavenging scheme introduced (or unmasked) new errors in HadGEM3–UKCA. That (without adjusting any other parameters) it shows such a strong improvement in this analysis while introducing very little in the way of new visible errors, coupled with the more physical formulation of the in-plume scheme, provides a strong indication (though not a formal attribution) that the lack of a realistic representation of this process is the primary cause of the disagreement between the BASE simulation and the observations. We have added the following text explaining this in the third paragraph of Section 6.1, where the improvement is described (lines 631–636 of revised manuscript):

> **This is supported by the fact that – without adjusting any other parameters in the model – the improvement is so strong, while introducing very little in the way of new visible errors which we might expect to see if the new scheme was compensating for errors in a different process (which would likely have a different structure).**

On the other hand, the smaller improvements seen in both models from changes to biomass-burning emissions are less definite: much of the difference is simply a shift in the global bias due to the lower total emissions in GFED3.1, and global bias is very likely to be the result of many partially-compensating errors in the model. We would suggest that a similar analysis in source regions might better distinguish emission effects from other sources of bias, but we would agree that the present study can provide only very weak evidence that GFED3.1 is a better representation of biomass-burning emissions than GFED2 – a point which we make clear in our conclusions, where we describe this as “smaller” [than the improvement from in-plume scavenging] and “not statistically significant”. We have clarified this in Sections 6.1 and 6.2 where we discuss this for HadGEM3–UKCA and ECHAM5–HAM respectively:

> **Unlike for convective scavenging, we cannot be confident that the small improvements seen here are genuinely attributable to better emissions, rather than compensating for biases elsewhere in the model.**

(fourth paragraph of Section 6.1; lines 656–659 of revised manuscript).

> ...; **however as in HadGEM3–UKCA, the improvement is not decisive enough to exclude the possibility that we are compensating for other biases in the model.**

(third paragraph of Section 6.2; lines 713–715 of revised manuscript).

We would therefore argue that, while compensating errors do have the potential to confound an analysis of this kind, the improvement seen from in-plume scavenging in HadGEM3–UKCA is sufficiently strong that we can be confident it is not an artefact of compensating errors, and therefore does represent a constraint on aerosol processes in climate models (specifically, that a representation of the scavenging of aerosol during convective transport is important for maintaining a realistic vertical profile).

We have added a new penultimate paragraph to the conclusions, acknowledging the potential difficulties due to compensating errors, and reiterating why we are nevertheless confident that the results presented for convective scavenging are real and reliable, while those for emissions are rather tentative (lines 890–904 of revised manuscript):

> **In an analysis of this kind, there is always the potential for compensating errors in different parts of the model to obscure which processes are poorly represented. This is particularly true in this case where we make a change in the biomass-burning emissions and compare to observations in remote regions; an analysis nearer the source regions might better distinguish emissions effects from other sources of bias. For convective scavenging in HadGEM3–UKCA, however, we see a very clear improvement from a more physically-realistic implementation without adjusting any other model parameters, with very little new error introduced; thus we conclude that we are not simply compensating for other errors, but that it is the convective scavenging itself which must be accurately represented in the model to obtain a realistic vertical profile.**

*Also, the manuscript does not include an examination of the potential role of the convective transport parameterization itself on the comparison with the observations. There is no indication of the entrainment and detrainment rates used in the models and how rates these might differ between the models. There is the possibility that errors in the entrainment/detrainment rates might be compensated by errors in a given parameterization for wet scavenging, or other aerosol processes. Some of these other processes are noted in the specific comments below. These issues should be addressed in the discussion.*

– see also (12) and (13) below

We agree that the convective transport parameterisation is likely to have a significant impact on scavenging by convective precipitation, and therefore on aerosol vertical profiles, and this is certainly an area worth further investigation. However, in the present study our intention was to focus on the impact of the coupling between the transport and scavenging schemes (simultaneous/in-plume or operator-split), rather than on the specific details of either (e.g. entrainment/detrainment rates, scavenging ratios or critical activation radii). Convective transport is one of many processes we hope to include in a follow-up study looking at the impact of a fuller range of processes on simulated aerosol vertical profiles.

It seems unlikely, however, that adjusting the entrainment or detrainment rates (thus altering the convective dynamics as well as tracer transport) would produce a compensating error with a similar structure to the difference between the operator-split and in-plume scavenging schemes.

The treatment of entrainment and detrainment in both convection schemes is quite complex, and the differences generally in the detail rather than being obvious in a simple overview. Despite significant similarities between the two schemes, there are a number of structural differences which mean that the parameters controlling entrainment and detrainment are not necessarily directly equivalent. We feel that a proper discussion of the effect of these on aerosol distributions, and the interaction with atmospheric dynamics, would require too much detail to include here. However this is certainly worth acknowledging, and we have added a new paragraph at the end of Section 5.2 briefly discussing the fact that the detail of the convective transport scheme is also likely to play an important role, citing Hoyle et al. (2011, ACP; doi:10.5194/acp-11-8103-2011) and Croft et al. (2012, ACP, doi:10.5194/acp-12-10725-2012):

> **Although in this study we focus on the impact of the coupling between convective transport and wet deposition, it is worth noting that the parameterisation of convective transport itself (in particular entrainment and detrainment) may have a significant impact on the vertical distribution of tracers, as demonstrated in Hoyle et al. (2011) and Croft et al. (2012).**

(lines 603–609 of revised manuscript)

***Specific comments:***

*1) Abstract: End of first paragraph: Please quantify what is meant by 'a rather different structure'.*

We have expanded the text to read: "**in HadGEM3–UKCA the discrepancy is dominated by excess aerosol in the tropical upper troposphere, while in ECHAM5–HAM2 areas of discrepancy are spread across many different latitudes and altitudes**" (lines 19–22 of revised manuscript).

*2) Abstract: second paragraph: Quantify what is meant by 'significantly improved with respect to observations'.*

We have expanded the text to read: "are **much improved with respect to the observations, with a substantial and statistically significant increase in correlation – this demonstrates** ..." (lines 28–29 of revised manuscript).

*3) Page 444, lines 3-4: Dust is omitted in HadGEM3. What influence does this have on the results? This excludes a considerable amount of aerosol surface area from the tropics. Could there be any feedbacks on the removal processes with this large coagulation sink excluded?*

Dust is not omitted in HadGEM3, but is represented by a separate scheme (as described in the text) and externally-mixed with respect to other aerosol represented by GLOMAP-mode. There may be feedbacks on removal processes in the tropical Atlantic from the Saharan outflow, if internal mixing of dust and other aerosol components is significant; however dust is unlikely to be a major aerosol component in the tropical Pacific region where we see the effect of convective scavenging in this study.

*4) Page 444, lines 18-21: The HadGEM3 model has a different threshold for aging from insoluble to soluble aerosol relative to ECHAM5-HAM2 (10-monolayer versus 1-monolayer). How does this influence the results? Does this delay the wet removal for HadGEM3 relative to ECHAM5-HAM2?*

Indications from a separate study in progress are that, although BC concentrations do decrease in many regions if this is changed from 10 monolayers to 1 in UKCA (which concurs with previous results), this effect is small compared to the differences between BASE and CVSCAV simulations seen here, or between any of the simulations and the HIPPO observations.

*5) Page 444, line 29: Please indicate that this is the large-scale cloud parameterization, if this is the case. How are the insoluble and Aitken modes treated? Is in-cloud impaction scavenging included in the parameterization? How does the fractional contribution of stratiform wet removal to total wet removal differ between HadGEM3 and ECHAM5-HAM2?*

We have clarified this in the text (sixth paragraph of Section 3.1; lines 271–277 of revised manuscript):

> In-cloud scavenging **by large-scale precipitation** assumes that 100% of the aerosol in the soluble accumulation and coarse modes is taken up by cloud water in the cloudy fraction of each 3D grid box, and is then removed at the same rate at which the large-scale cloud water is converted to rain. **(Nucleation, Aitken and insoluble modes are not scavenged.)**

There is no separate treatment of in-cloud impaction scavenging – as in most schemes of this type, the fixed fraction is considered to (coarsely) represent all in-cloud scavenging. Scavenging by convective precipitation is described separately a couple of sentences later.

In HadGEM3–UKCA, ~92% of in-cloud scavenging of BC is by large-scale/stratiform cloud in the BASE configuration; this is reduced to ~75% in CVSCAV, where convective scavenging is much more effective. We do not have these diagnostics from our ECHAM5–HAM2 runs; however Croft et al. (2009; ACP; doi:10.5194/acp-10-1511-2010) show this to be ~50% for an earlier version of the model.

*6) Page 445, line 4: How does the assumption of a fixed 30% cloud fraction influence the results? Is this also the precipitation fraction?*

We haven't looked at sensitivity to this particular parameter; the 30% comes from Giannakopoulos et al. (1999, JGR, doi:10.1029/1999JD900392), and ultimately from Walton et al. (1988, JGR, doi:10.1029/JD093iD07p08339). However, no physical justification for this choice is given there, and 30% seems quite high for a convective cloud/precipitation fraction (the two are assumed to be the same in boxes where convective precipitation occurs) over a ~1.5° grid box. For smaller fractions, wet scavenging by convective precipitation would be even less efficient, increasing further the discrepancy between the BASE simulation and observations seen here.

*7) Page 445, lines 8-10: Is the removal from the environmental layer at the level that the precipitation formed or otherwise?*

We have clarified this in the text (sixth paragraph of Section 3.1; lines 285–288 of the revised manuscript):

> The scavenged aerosol is removed from the grid-box mean tracers after the convection scheme has run – i.e. from the post-convection environmental air **at the level where the precipitation formed**, rather than the convective updraught itself.

*8) Page 447, lines 1-3: How vigorous is the below-cloud scavenging in ECHAM5-HAM2 relative to HadGEM3? How might differences between the two models in the parameterization for this process influence the comparison of the resultant mixing ratios between the two models?*

There may indeed be differences between the models due to their different below-cloud scavenging parametrisations, amongst many other processes. To address this, a follow-up study is under way, looking at the impact of a wider range of aerosol processes on simulated aerosol distributions. However, in both models the removal of BC is dominated by in-cloud scavenging. In HadGEM3–UKCA (BASE), below-cloud scavenging accounts for only ~1.1% of BC mass removal (and ~1.3% of that by wet deposition) globally. Our ECHAM5-HAM2 simulations did not record this information, but Croft et al. (2009, ACP; doi:10.5194/acp-9-4653-2009) tested a number of enhancements to the below-cloud scavenging scheme in ECHAM5-HAM, some of which have been adopted in ECHAM5–HAM2, and which may lead to stronger below-cloud scavenging (perhaps on the order of ~10%) but with in-cloud scavenging still representing the dominant sink.

*9) Page 448, lines 4-6: A reference to the figure number would be helpful here.*

We have added a reference **(Figure 1)** here (first paragraph of Section 4; line 420 of revised manuscript):.

*10) Page 449, lines 15-19: How well does nudging reproduce the observed synoptic conditions in terms of the cloud and precipitation fields? This approach is not exactly the same as using assimilated meteorological fields, such as in a chemical transport model. With nudging of the vorticity, divergence, temperature and pressure fields in a global climate model, how tightly are the cloud and precipitation fields controlled between simulations, how well do they match to the actual observed meteorology and how does this influence the comparison with the observed aerosol mixing ratios?*

The effects of nudging on large-scale cloud and precipitation were studied in Telford et al. (ACP, 2008), where it was shown to perform well, almost completely capturing the NH storm tracks for example. The effects on convection were studied in some detail in the SCOUT-O3 model intercomparison, where the nudged model performed at least as well as the CTMs in the same study (Russo et al., ACP, 2011; doi:10.5194/acp-11-2765-2011). Tracer transport was studied in a companion paper (Hoyle et al., ACP, 2011; doi:10.5194/acp-11-8103-2011), where again the nudged model performed well, although transport of short-lived species was still dominated by the convective parametrisation (the importance of which is rightly mentioned in another comment). While the meteorological fields in a nudged model will not perfectly track the reanalysis fields (or any given observations, or "reality") in a nudged model, studies such as those mentioned demonstrate that we can certainly expect a better match to (our best estimates of) historical meteorology than for a free-running model. We have added the Telford and Russo citations to address this point (first paragraph of Section 4.2; lines 475–482 of revised manuscript):

> **Although neither water vapour nor any cloud variables are nudged directly, Telford et al. (2008) show that large-scale cloud and precipitation patterns are reproduced well in a nudged model, while (Russo et al., 2011) show that for convection a nudged model performs as well as an offline chemical transport model (CTM) driven directly by meteorological fields from a reanalysis.**

Tracer transport (and Hoyle) is now discussed elsewhere in response to major comments.

*11) Page 451, lines 16-21: The biomass burning emissions are emitted uniformly between 50m-3km in HadGEM3, while the ECHAM5-HAM2 emissions are within the boundary layer. How does this difference in altitude of emissions contribute to differences between the models?*

Results from a separate study in progress suggest that (at least in HadGEM3–UKCA) changing this makes very little difference to BC or OC profiles – presumably because such differences in emission profile are quickly obscured by strong mixing processes in the lower troposphere (principally boundary-layer turbulent mixing and shallow convection). Even injecting all biomass emissions at the surface changes little; only if the emission height is extended well above 3km do we start to see any significant effect. We have added a mention of this in the third paragraph of Section 6.2, after we discuss the very limited impact of different emission profiles in ECHAM5-HAM2 (lines 723–727 of revised manuscript):

> **Similarly, using a boundary-layer-following emission profile in HadGEM3–UKCA (instead of the default fixed $\sim 50$ m to 3 km profile) makes little difference, indicating that the different emission profiles do not contribute significantly to the differences between the two models.**

*12) Section 5.2: In the description of the convective scavenging, please include some details about the entrainment and detrainment rates used in the two models and how these might differ for different types of convection, and at different latitudes and how this might influence the results.*

See third major comment re entrainment/detrainment.

*13) Page 453, line 15: Can the problem of too much aerosol aloft be related additionally to errors in the amount of aerosol that is entrained and detrained by the simulated convective plumes?*

See third major comment re entrainment/detrainment. However, it seems unlikely that convection is the primary cause of difference between HadGEM3–UKCA and HIPPO-3 observations in the northern hemisphere, for two reasons: (a) the difference is consistent between operator-split and in-plume scavenging – if entrainment was strong, and there was enough convective transport to produce too much aerosol aloft, we would expect this to be efficiently removed by the in-plume scheme; (b) as discussed under (16), convective activity in the northern mid-latitudes is relatively weak at this time of year.

*14) Page 453, line 18: Reference to the figure and row numbers might be added here.*

We have added a reference **(second row of Figure 3)** here (fourth paragraph of Section 6.1; line 648 of revised manuscript).

*15) Page 453, lines 26-29: The poor agreement is attributed to lack of realistic convective scavenging – but could other factors also play a role in this comparison? How well do the simulated precipitation fields agree with the observed precipitation? The lack of improvement for HIPPO3 illustrates the possibility for other factors/processes to confound the analysis. Can the possibility of compensating errors yielding a closer agreement for HIPPO1 and HIPPO2 be excluded? Could there be errors in the stratiform wet removal, which compensate errors in the convective wet removal?*

Certainly other factors may play a role, some of which are mentioned in this and other comments. However, the introduction of the in-plume scheme was motivated by physical realism; the fact that it does produce such a strong improvement for HIPPO-1 and -2 suggests a lack of confounding factors of similar magnitude. (If the errors had been masked e.g. by overly-strong large-scale scavenging, then we might not have seen such improvement without re-tuning the model.) HIPPO-3 does behave differently, which suggests that there are certain regimes in which the poor agreement with observations has a different cause, as is discussed in the text. (This may well relate to seasonal effects, as discussed in the next comment). Ultimately, such structural errors will only be detectable in perturbed-parameter ensemble simulations and then only if the bias exceeds the ensemble envelope; this is no trivial exercise, but will be attempted in the UK GASSP project.

*16) Each of the 3 HIPPO campaigns occurred in a different season. Could the authors comment on the seasonal cycle of convective precipitation and wet removal, particularly towards the mid-latitudes and how this might influence the results?*

Seasonal effects may indeed account for some of the difference between the three phases. Looking at convective precipitation from ERA-Interim, the convective precipitation in the northern tropical Pacific is fairy steady, while that in the southern tropical Pacific is strong only during southern-hemisphere summer (i.e. during HIPPO-1, with HIPPO-2 occurring as convection strengthens in this region and HIPPO-3 as it weakens. Mid-latitude convective precipitation over the Pacific appears to be stronger in autumn/winter in each hemisphere, which may explain why the lack of agreement between HadGEM3–UKCA and HIPPO-3 (March–April) in the northern mid-latitudes is relatively unaffected by changes to the convective scavenging. We have added a comment to this effect where we discuss the different behaviour of HIPPO-3 on in the penultimate paragraph of Section 6.1 (lines 667–670 of revised manuscript):

> **This is consistent with HIPPO-3 occurring in northern-hemisphere spring, when convective precipitation in the northern midlatitude Pacific is relatively weak.**

However, we don't really have enough data to quantify how much of the difference between phases of the HIPPO campaign is due to the seasonal cycle vs. other sources of variability, since we don't have multiple phases in the same season but different years.

*17) Page 454, line 10-11: The observed high burdens are attributed to localised biomass-burning plumes – are these plumes expected to be consistent between all seasons since this is seen for all campaigns? Could the models be transporting too much aerosol aloft, and as a result not allowing enough transport to the Arctic?*

In HIPPO-1, the high BC loading seen for the 60°N–85°N band in Schwarz et al. (2010). were due to two profiles very close together in space and time that showed very high loadings compared to the rest; thus this was viewed as a localised effect. However, for HIPPO-2 and HIPPO-3, there is a more consistent pattern of relatively high loadings in the Arctic. We agree that the model/observation discrepancy here is likely to be due to too little transport to the Arctic, either due to errors in the transport itself, or due to BC being removed too rapidly (probably by large-scale wet scavenging, since this affects both BASE and CVSCAV simulations in HadGEM3–UKCA). We have updated the text to reflect this (final paragraph of Section 6.1; lines 680–690 of revised manuscript):

> **The high burdens observed in the Arctic in HIPPO-1 were attributed to a localised biomass-burning plume (Schwarz et al., 2010) as they were dominated by two particular profiles which were close together. In HIPPO-2 and HIPPO-3, however, the high Arctic burdens are a more systematic feature of the profiles in this region. This suggests that the model is underestimating the transport of BC to the Arctic – either due to errors in the transport itself, or because it is removed too rapidly (probably by large-scale wet scavenging, since this affects both BASE and CVSCAV simulations).**

*18) Page 454, line 23: Please indicate where/when these strong biases are reduced and quantify the reduction.*

The reduction in overall bias is quantified in Fig. 5, where G3M and G3H are both closer to the zero-bias line than BASE; however we do not claim that this reduction is statistically significant. We have weakened the statement slightly, and added the following text explaining that (as can be seen from Fig. 4) the main visible reductions in bias are at lower levels around the equator (for all phases) and in at all levels in the southern mid-latitudes (for HIPPO-1):

> **Some** of the strongest biases are reduced in the G3M simulation**: in particular, at lower levels around the equator (for all three phases) and also in the southern midlatitudes (for HIPPO-1). This suggests that that part** of the tropospheric error ...

(third paragraph of Section 6.2; lines 707–711 of revised manuscript).

*19) Page 455, line 15-17: This sentence is confusing. The authors state that clearly the error might be in the parameterization of the scavenging process or in some other process. This indicates the problem. There is a difficulty to clearly attribute these discrepancies between the model and the observations to a single physical process. I am left unsure if we can actually use these observations to constrain aerosol processes in a global model – but they do provide a useful constraint on the aerosol profiles. Here and throughout, the authors should be careful to acknowledge these difficulties in their analysis.*

Notwithstanding the major comment about the extent to which we can constrain processes (which is addressed separately above), we have clarified this sentence (see below) – in essence we agree with the point made in the comment: we do not know from this analysis what process is responsible for the high burdens in ECHAM5–HAM2. Even though this model already has a in-plume convective scavenging scheme similar to that we have introduced in HadGEM3–UKCA, it remains possible that some other aspect of scavenging is responsible (and perhaps likely, given that it represents the main removal process), but that is only one of many possibilities. Certainly the vertical profile of the difference from observations is very different between the two models, which may be an indication that the process causing it is different.

> **We cannot determine from this analysis what process is responsible for the high burden in ECHAM5–HAM2 – a more detailed study of the role of the different processes in this model would be required. It may be still be some aspect of scavenging which is too weak, but equally the problem may elsewhere.**

(penultimate paragraph of Section 6.2; lines 746–751 of revised manuscript).

*20) Page 456, lines 25 onwards: Does a scavenging scheme that performs better in a nudged model environment capture reality more closely? Does nudging suppress any feedbacks in a global model, which might limit the validity of the conclusion that this is a more 'realistic simulation of the aerosol' during a flight. Nudging makes certain meteorological fields more consistent between simulations, but can you demonstrate that the cloud and precipitation fields are closer to reality and more consistent with each other between the simulations relative to the free-running simulations? Also, climate studies will generally use free-running simulations. Do the results indicate that the scavenging revisions should be adopted by the model, given the possibility for compensating errors in other processes and the lack of improvement for free-running simulations?*

See also response to (10). While the performance of the nudging schemes has not been specifically studied in these simulations, the studies described in (10) show that nudging can indeed reproduce cloud and precipitation fields well (notwithstanding the difference between reanalysis and "reality" in remote regions). Nudging is indeed likely to suppress any feedbacks that operate on a timescale longer than that of the relaxation in the nudging scheme, although in the configuration of HadGEM3–UKCA used here the feedback effects of aerosol on meteorology are not active in any case. The improved performance in the nudged model is a strong indication that the new scavenging scheme captures the physical process better than the existing scheme, and we would therefore recommend its adoption. Nevertheless, a full evaluation of its effects on the long-term climatology of a free-running model with all feedbacks active may bring to light interactions with other components of the model that (at least) require re-tuning of certain parameters.

*21) Page 458, line 9-11: Are the free-running simulations re-tuned between the implementation of the two scavenging schemes?*

No re-tuning has been done; we have clarified this where the sensitivity test configuration is described for both models (penultimate paragraphs of Sections 3.1 and 3.2; lines 317–318 and 406–407 of revised manuscript);

*22) Page 458, lines 18-20: While the observations are useful in evaluating aerosol distributions, I think that the insight into physical processes (e.g. convective wet scavenging) is more difficult to obtain within the framework of a single global model. This should be more carefully considered in the text and title of the manuscript.*

See major comments re constraining processes/profiles.

*23) Table 1: What is meant by aerosol feedbacks enabled and disabled (last line of table) and how does this influence the results?*

Aerosol feedback enabled/disabled refers to whether changes in aerosol are allowed to affect the model meteorology (via direct, semi-direct and indirect effects). We have clarified this in the table:

> Aerosol feedbacks **on meteorology (direct/semi-direct/indirect effects)**

Given that we primarily consider nudged runs, the resulting differences in meteorology are likely to be small.

*24) Fig 5: Were any free-running simulations conducted with the ECHAM5-HAM2? Is there the possibility for a significant difference between free-running simulations with different emissions in ECHAM5-HAM2?*

The comparison with free-running simulations was only done for HadGEM3–UKCA. It is possible that they may behave differently for ECHAM5–HAM2, although there seems little reason to expect this to be the case. As discussed in the text, the main context in which the different emissions would be likely to have more effect is if we compared to observations nearer to one or more major source regions.

**In response to Anonymous Referee #2:**

*Introduction, paragraph 1: aerosol radiative effects can also change the regional and local circulation, please mention this as a possibility here alongside the other effects already included.*

> We have added a mention of circulation effects, citing Roeckner et al. (2006; doi:10.1007/s00382-006-0147-3):
>
> > **Consequent changes to circulation patterns may lead to additional effects (e.g. Roeckner et al., 2006).**
>
> (lines 54–56 of revised manuscript).

*Page 439, Line 20, please include/consider references by Johnson et al (2004) and Johnson (2005) which are classic examples of the importance of vertical profile for direct and semi-direct effects.*

> We have added a citation of Johnson et al. (2004; doi:10.1256/qj.03.61) which seems to be of particular relevance (second paragraph of Section 1; line 66 of revised manuscript).

*Page 442, line 15 does "this paper" refer to Schwarz et al (2010)? – please make this clearer.*

> The words "this paper" do not occur here. However, we have replaced "in the above paper" with an abbreviated citation (**Schwarz et al.**), and "is not considered further in this analysis" with "**is not considered further in the present study**" to clarify this (penultimate paragraph of Section 2; lines 178 and 183–184 of revised manuscript).

*— Too many "however"s in the last sentence!*

> We have removed the first "however", without which the meaning is still clear.
>
> We have also updated this sentence following further analysis of the HIPPO SP2 measurements, which allows us to be a little more explicit about the regime in which this becomes a significant source of uncertainty (lines 179–184 of revised manuscript):
>
> > **In the cleanest regions (BC MMR less than ~0.5 ng kg$^{-1}$), where only a small number of particles were detected per minute, the sampling uncertainty of the observations is likely to contribute significantly to the scatter in the results; however this is not considered further in the present study.**

*Page 447, please add more details on how the nudging is done, as this is still a new technique in climate modelling.*

> The technique has been used in ECHAM at least since the 1990s – we have added a citation of Jeuken et al. (1996, JGR; doi:10.1029/96JD01218), which describes both the technique itself and its implementation in ECHAM. We have also added more description of the specific implementation in the versions of each model used here (choice of variables, relaxation time constant and levels). We would prefer not to get into the actual mathematics of the technique however, as this is well discussed in the references.

For HadGEM3–UKCA (Section 3.1, seventh paragraph; lines 298–309 of revised manuscript):

> **In order to capture the meteorological conditions at the time of the flight campaign, we use the technique of nudging (Jeuken et al., 1996). In the HadGEM implementation (Telford et al., 2008, 2012), potential temperature and horizontal wind are relaxed towards fields from the ERA-Interim reanalysis (Dee et al., 2011). The relaxation time constant is the "natural" one of 6 h (the time spacing of the reanalysis data); this choice is validated in Telford et al. (2008). The nudging is applied between levels 14 ($\sim$ 4 km) and 32 ($\sim$ 21 km) inclusive; levels 13 and 33 are nudged at half strength (i.e. with a 12 h time constant), and no nudging is performed on levels outside this range.**

For ECHAM5–HAM (Section 3.2, sixth paragraph; lines 393–400 of revised manuscript):

> **Once again, the large-scale dynamics are nudged towards ERA-Interim (Dee et al., 2011) reanalysis data, following Jeuken et al. (1996): temperature, vorticity and divergence are surface log-pressure are relaxed towards the reanalysis fields with time constants of 24 h, 6 h, 48 h and 24 h respectively on all model levels. The nudging is performed in spectral space, on all but the wavenumber-0 (global-mean) spectral component.**