Thank you to both reviewers for their helpful comments, which we have addressed below.

Response to Reviewer 1

Major comments:

1) Normalized analysis has been shown to provide much more insight into the comparability of forcing-response relationships (Shindell and Faluvegi being a clear example). While it is clear that it is useful to provide the un-normalized numbers, the paper would greatly benefit from adding a description of the normalized results. For example, if ones takes the numbers from HadGEM/ECHAM/NorESM for BC, we see that the response in delta(T) is almost the same as the scale OC response. As it should. A table documenting the radiative forcing associated with each perturbation run should be included.

This work focusses on the climate impacts of perturbing emissions. This is a different focus from e.g. Shindell and Faluvegi who aim to assess the climate responses to regional forcing from aerosols/GHGs rather than the actual emissions.
In order to calculate climate responses for a given RF/ERF, i.e. the normalised responses, it would be necessary to run a different set of experiments (atmosphere-only rather than coupled) in order to find the RF or ERF from the emissions changes, since these cannot be derived from the coupled simulations. We therefore do not have the RF/ERF values that would be necessary to calculate normalised responses.

2) The figures only show stippling where models agree on the sign. That is a pretty low bar to pass (and I guess they still don't pass it). I would however provide estimates of the statistical significance based on the interannual variability. Similarly, zonal mean figures (5-7) are shown even for areas where models do not agree. What is the meaning of those figures in that case!?

We have changed the stippling in Figs 5-8 to denote statistical significance at the 95% confidence level. The distribution is generally very similar to that using previous method.
The zonal mean figures show the zonal means for the individual models as well as the multi-model means. Showing the individual models is useful to see differences (and similarities) between models. The multi-model mean zonal means are useful to compare with the map figures (which show only the multi-model means), and are useful to see when the sign of the response in the different models is in agreement.

3) The control experiment is much too short for the analysis that is being performed here, where the goal is to identify the response to a forcing much smaller than 2xCO2. As one can see for ECHAM, the global surface temperature is still trending at the end of the fifty years. Knowing that, it is necessary to show and discuss the trends in the climate state for the control experiment continued over the 50 years for which the perturbation is calculated. It would not be surprising if part of the "signal" was actually present in the control experiment as well. An approach might be to take into account the model drift over the 50 years.

We agree that it is not ideal to perform these experiments against a control state that still has a slight temperature trend. However, since the perturbations are applied to start of the control simulations, and we analyse the results over the same time period in each case, this underlying temperature trend is still present in the perturbed simulations as well as the control. In addition, we only consider differences between the perturbed and control simulations, rather than absolute values.
There are other studies in the published literature which have used similar integration-lengths and found these to be sufficient, e.g. Pausata et al (2014), Kristjansson et al (2005). Based on this we believe that 50 years is sufficient.

Added sentence to pg 3834 line 11:
These drifts are also present in the perturbation experiments, since these start from the control simulation at the beginning of the 50 yr period analysed.

4) There is no description of how the control experiments are performed. What is the level of GHGs? Where does the ocean initial state come from?

The description of the control experimental setup and GHGs is in the paragraph starting on pg 3832 line 18. The ocean initial state is spun-up from present-day (in the case of HadGEM, NorESM and CAM4, which use present-day GHGs) and pre-industrial (in the case of ECHAM which uses pre-industrial GHGs) states, based on CMIP5 initial conditions. The GHGs are fixed throughout; any differences occurring due to the different GHGs or initial ocean state will be much smaller than the differences due to different model dynamics/chemistry/aerosol/parameterisation schemes.

Amended text to describe the ocean state:
pg 3832, line 1:
The control simulations were first run for several decades using an initial ocean state based on present-day CMIP5 conditions for all models except for ECHAM-HAM, which used a pre-industrial control state (see below). The control and perturbed simulations were then run from this spun-up state for 50 yr, in order to separate a robust signal from the interannual variability.

Minor comments:

1) Author list: It is CICERO, not CISERO

Changed

2) Page 3825, line 5: There are much better and recent references to the impact of ozone on health and agriculture than HTAP. For example Tai et al., Nature Climate Change, 2014.

Added two references: Tai et al 2014 and Amann et al 2013

3) Page 3825, line 19-21: it might be useful to use the AR5 nomenclature (ACI-ARI)

Changed sentence to read:
“The **aerosol-radiation interactions and aerosol-cloud interactions** bring futher inhomogeneities,...”

4) Page 3826, line 23: there is a wealth of recent papers highlighting those connections. Please provide a better list of references.

Added references: Boucher et al 2013, Osborne and Lambert, 2014
More references are included in the rest of the paragraph about precipitation effects.

5) Page 3831: it is not really clear what the added value of including CAM4 is. This version does not have indirect effects. Why bother? Why only BC?

We agree that our reasoning was not clearly explained. The purpose of including the CAM4 simulations and the extra NorESM simulation was to explore further the BC results, to understand them further. This work was part of a project of which BC was one of the key focusses, and hence the CAM4 experiments were available to add to this study. The fact that CAM4 does not include the indirect effects is not a problem since the BC indirect effects are small relative to the direct/semi-direct effects

Added sentence to text on pg 3829 line 5:
The extra BC simulations were included in order to explore the BC results further, as this work was part of a larger project of which BC was a key focus.

6) Page 3833, line 5: ozone would be affected though, because of the methane impact.

Oxidant fields for the sulphate aerosol production calculation were a 2003-2010 average from the MACC reanalysis (Inness et al. 2013)
Added the above sentence to the text on pg 3038, end of line 11.

7) Page 3833: this whole section need to include a documentation of the aerosol budget. Also, it needs to include a discussion of the differences in precipitation between the models. Finally, the aerosols should be compared to existing papers such as Samset et al., Observational evidence for overestimation of modeled black carbon radiative forcing. *Atmos. Chem. Phys.*, 14, 12465-12477, doi:10.5194/acp-14-12465-2014, 2014.

We have added the lifetimes of BC,OC and SO4 to Table 1 (in addition to the burdens which were already shown). We have added more discussion to the paragraph starting on pg 3833 line 13, including discussion of existing papers:
“Despite all the models having the same emissions input, there is a large discrepancy between models in the vertical distribution of aerosols in the atmosphere, and in the total aerosol burden, which is typical for current global aerosol models (Textor et al 2007). HadGEM and ECHAM-HAM have relatively low total burdens of BC, **and short atmospheric lifetimes**, compared with NorESM and CESM-CAM4 (Table 1). Figure 2 shows vertical sections of the annual average, zonal mean BC mass mixing ratio in the control simulation for each of the models considered. HadGEM and ECHAM-HAM (Fig. 2a and b) have low concentrations of BC at high altitude, which means there is less BC above clouds. In contrast, NorESM and CESM-CAM4 show high BC concentrations extending to above 200 hPa throughout most of the northern hemisphere and southern hemisphere tropics (Fig. 2c and d). This has implications for the impact that removing anthropogenic BC emissions may have. BC at high altitude can have very strong direct effects if it is located above high-albedo cloud surfaces. In the models with higher concentrations of BC at high levels in the control simulations, more of this high-level BC can be removed in the BC perturbation

experiment, leading to a larger change in BC direct forcing. **The larger amount of high-level BC in NorESM and CESM-CAM4 (which uses aerosol input from OsloCTM2) is consistent with the AeroCom models discussed by Schwarz et al. 2013 and Samset et al. 2014 who found that these models have too much BC at high altitudes when compared with observations over the Pacific in the HIPPO campaign (Wofsy, 2011), and overestimate the BC lifetime. In HadGEM, the lower concentrations of BC at high altitudes and shorter BC lifetimes are likely due to recent modifications to the convective scavenging scheme, which were implemented in order to improve the correspondence with these observations. However, the BC lifetime of 3.4 days is rather short, so in this case the model is underestimating the high-level BC concentration. The true BC distribution is therefore somewhere in between that of HadGEM and NorESM/CESM-CAM4.** The OC burden in NorESM is considerably higher than in the other three models, **and its lifetime is correspondingly longer**. **The range of OC burdens between models is expected due to differences in OA burdens and OA/OC ratios between models (Tsigaridis et al. 2014).** NorESM and CESM-CAM4 have relatively low burdens of SO_4, **and short lifetimes,** compared with HadGEM and ECHAM-HAM. There are also differences in the vertical distribution of OC and SO_4 between models (not shown) but as these are scattering, rather than absorbing, aerosols the impact of the vertical distribution of the aerosol will have less of an impact on the results. **A more detailed evaluation of the models used here against observations is given in Stohl et al. 2015 and references therein.**"

We have added a supplementary figure showing observed (GPCP) precipitation for 2000-2010, the annual average precipitation in the control simulation for the multi-model mean and each individual model. We have added a paragraph to the end of Section 2 describing these:
"There are some differences between models in the precipitation patterns, particularly in the tropics (Fig. S1). All models suffer from the "double ITCZ" problem (i.e. there is an overly strong band of precipitation to south of the equator) which is a known problem in CMIP5 AOGCMs (Li and Xie, 2014). This is most pronounced in ECHAM-HAM (Fig. S1d). ECHAM-HAM and HadGEM also have region of very low precipitation around the Equator in the Pacific (Figs. S1c and d). There is some variation in the north-eastward extent of the North Atlantic storm track: in NorESM it extends too far north-east, while in CAM4 it does not extend far enough (Figs. S1e and f); in HadGEM and ECHAM-HAM it matches the observations well. All models have too much precipitation over the Himalayas and the Andes, which is probably due to inaccuracies in their representation of precipitation over high orography."

8) Page 3834, line 22: why are HadGEM and ECHAM similar when their SO4 burdens are so different?

The change in SO2 emissions is the same in each case, so the change in SO4 will be roughly the same, regardless of the different absolute burdens. Since we are looking at a difference in temperature between the control and perturbed runs, the baseline SO4 concentration should not impact the results too much.

9) Page 3836, line 2: include references discussing the shift in ITCZ

Added references:
Broccoli et al., 2006; Kang et al., 2008; Ceppi et al., 2013

10) Page 3836, line 23-25: how do we know that this is the causal link?

We have changed the text as follows. This is only hypothesis.
"There are broad regions over Russia and North America with increased precipitation. These are

collocated with regions of increased surface temperature, which would provide more available moisture through evaporation. The increased precipitation could also be due to the reduced aerosol concentrations in these regions."

11) Page 3836, line 26: the discussion of run-off might be much more useful if it is recast in terms of river flows, maybe for the largest rivers.

We agree that the discussion of the run-off changes is not very useful. We have therefore decided to remove plots and discussion of run-off since differences are very small and mostly not significant.

Figures of run-off and corresponding discussion have been removed from the paper.

12) Page 3838, lines 19-21: how do the authors know that they are not simply looking at noise?

We have included stippling in Figures S1-S3 to show significance. This shows that the temperature responses and sea-ice changes in individual models are significant.

We have also rewritten the corresponding paragraphs in the text (pg 3838 line 19 to pg 3839 line 20) to aid clarity of this discussion.

13) Page 3842, lines 5-9: it is not that clear that the ITCZ shift is related to climate sensitivity. Please substantiate!

This sentence does not refer to the ITCZ shift. Reworded to
"NorESM gives a weaker temperature response than the other two models."

14) Page 3842, lines 21-25: as the papers by Sand et al have indicated (among others) the location of the BC forcing is quite important.

Added sentence:
"...BC deposition on snow. As shown by Sand et al (2013b), this has a relatively large impact on surface temperature in the Arctic."

15) Page 3843, lines 25-26: I would expect that concentrations are available from those simulations. Therefore a statement other than "probably have" should be made.

The difficulty is actually in determining what the true BC distribution should be, since there are no observations that give a global, 3D picture of the present-day BC distribution.

Rewritten the paragraph on pg 3843:
"...partly to the different atmospheric BC distributions in the models**, as shown in Fig. 1**. Accurately representing the correct BC distribution in GCMs is very difficult **(Samset et al. 2015). Recent studies (e.g. Schwarz et al. 2013, Hodnebrog et al. 2014, Samset et al. 2015) have compared BC distributions in GCMs with data from observational studies such as the HIPPO campaign, which provided observations from a large spatial area over the Pacific (Wofsy, 2011). They found that the models had too much BC at high altitudes in these regions, and that the BC lifetime was generally too long. Recent modifications to the convective scavenging scheme in HadGEM (which are included in the model set-up used here) were designed to reduce the amount of high-level BC to give better agreement with the HIPPO observations. The result of these changes is that HadGEM has less BC at high levels globally than the other**

models (except ECHAM-HAM), and a much shorter BC and OC lifetime (Table 1). ECHAM-HAM also has less BC at high levels, and a short BC lifetime. In contrast, NorESM and CESM-CAM4 have much more high-level BC and longer BC lifetimes, which may overestimate the direct forcing from anthropogenic BC (consistent with Samset et al. 2015) and may therefore exaggerate the impact of removing anthropogenic BC emissions.

Response to Reviewer 2

- Page 3824, Line 17: There are four models in the study, but here the authors refer to "all three models".

Actually for the SO2 and OC experiments there were only 3 models.

Removed the word "three" to avoid confusion.

- Page 3824, Lines 17-18: Perhaps rephrase to "northern hemisphere mid and (especially) high latitudes".

Done

- Page 3825, Line 7: Typo: SCLPs -> SLCPs

Done

- Page 3825, Lines 9-10: According to the UNEP definition, methane is also included in SLCPs, so the authors could include all species up to methane here in their definition, but mention that they restrict their focus to the constituents with lifetimes of days to months, which therefore have a particularly inhomogeneous distribution.

Reworded as follows:
SLCPs have relatively short atmospheric lifetimes compared with well-mixed greenhouse gases (WMGHGs) such as CO2, **with most** remaining in the atmosphere for only days to months. **The exception is methane, which has a lifetime of around a decade, but here we focus on the shorter-lived species.** The impacts of SLCP emissions on climate therefore occur on relatively short timescales of less than 30 yr (Collins et al., 2013). The short atmospheric lifetime of **non-methane** SLCPs ...

- Page 3826, Lines 2-3: Need to also mention explicitly the cloud lifetime effect here.
-

Modified this sentence:
Hydrophilic aerosols also provide cloud condensation nuclei (CCN), allowing more smaller cloud droplets to form, which increases the cloud albedo and the cloud amount, and prolongs the cloud lifetime by inhibiting precipitation. This further contributes to the negative forcing (Boucher et al., 2013).

- Page 3826, Lines 6-7: Please mention why BC warms the surface when near it (reemission in thermal wavelengths).

Added:
low-level BC can warm the surface **by re-emitting radiation in the thermal wavelengths**, whereas …

- Page 3826, Line 25: Please add "global" before "temperature".

Done

- Page 3828, Line 12: Circulation changes are not really assessed in the paper, so I would remove this word from here.

Done

- Page 3828, Lines 23-25: Please clarify whether photolysis is affected by the aerosol tracers in the models.

No it is not.

Added sentence to pg 3828 line 25:
Photolysis is not affected by the aerosols in these models.

- Page 3829, Lines 11-13: Please mention whether stratospheric chemistry is simulated too.

No it is not

Added text to pg 3829 line 11:
“The UKCA TropIsop scheme is used to model gas-phase chemistry **in the troposphere**.”

- Page 3829, Lines 21-26: I would suggest mentioning how aerosol effects on clouds are simulated.

These are modelled using an aerosol activation parameterisation.

Added sentence on pg 3829 line 25:
“The effects of aerosols on clouds are modelled using an aerosol activation parameterization (Abdul-Razzak and Ghan, 2002).”

- Page 3830, Line 8: It is mentioned earlier that the gas-phase chemistry is not modelled online in ECHAM6-HAM2. Where do the oxidants fields used for aerosol production come from? Worth mentioning.
-

Oxidant fields for the sulphate aerosol production calculation were a 2003-2010 average from the MACC reanalysis (Inness et al. 2013)

Added the above sentence to the text on pg 3038, end of line 11.

- Page 3831, Lines 5-6: Is the BC/dust deposition effect on surface albedo accounted for in the two models mentioned earlier in the text? Please clarify.

No, this is not represented by the other models.

Added to sentence:
In the fully coupled NorESM1-M, albedo-effects of BC and mineral dust aerosols deposited on snow and sea-ice are also taken into account**; this process is not represented in the other three models**.

- Page 3831, Line 7: Please clarify again here that the NCAR CESM model is only used for the BC analysis.

Added to start of line 7:
“An additional model, NCAR CESM 1.0.4/CAM4, was used for the BC analysis only.”

- Page 3831, Line 27: Please add “globally” after “removed”.

Done

- Page 3832, Line 2: Was the spin-up performed for the control and the perturbation runs of equal length?

The spin-up was performed for the control only. The emissions perturbations were applied after the spin-up period.

Expanded and reworded this sentence (pg 3832, line 1):
The control simulations were first run for several decades using an initial ocean state based on present-day CMIP5 conditions for all models except for ECHAM-HAM, which used a pre-industrial control state (see below). The control and perturbed simulations were then run from this spun-up state for 50 yr, in order to separate a robust signal from the interannual variability.

- Page 3832, Line 26: Please mention explicitly what you mean by “other natural emissions”, as this currently sounds a bit vague.

Added text to this sentence:
“Other natural emissions**, including DMS and volcanic emissions,** are included,...”

- Page 3833, Lines 1-2: What about methane in CESM-CAM4 – how is it treated?

It is prescribed at present-day levels as in HadGEM and NorESM.

Added:
...present-day levels in HadGEM, **NorESM and CESM-CAM4** and at ...

- Page 3833, Line 12: It is a bit counter-intuitive that Africa has such strong anthropogenic emissions. Does this include agricultural biomass burning? If so, the distinction should be made clearer (i.e. whether there is absolutely no BB component in the anthropogenic emissions removed or whether there are exceptions).

All biomass burning, including agricultural BB, is not removed with the anthropogenic emissions.

Added to pg 3832 line 21, to clarify that this means all BB emissions:
Biomass burning emissions, **including those from agricultural biomass burning**, are from the …

- Table 1: Sulphate is not mentioned in the caption. Also, the caption says “three models”, whereas burdens for four models are shown.

Corrected

Caption changed to:
Summary of BC, OC and SO4 burdens in the control simulation for the four models.

- Page 3833, Lines 20-25: It would be useful to briefly mention here which model may be closer to reality when it comes to vertical BC distribution. Any ideas?

Added further discussion to this paragraph. In particular:
The larger amount of high-level BC in NorESM and CESM-CAM4 is consistent with the AeroCom models discussed by Schwarz et al (2013) and Samset et al (2014), who found that these models have too much BC at high altitudes when compared with observations over the Pacific in the HIPPO campaign (Wofsy et al., 2011), and overestimate the BC lifetime. In HadGEM, the lower concentrations of BC at high altitudes and shorter BC lifetimes are likely due to recent modifications to the convective scavenging scheme, which were implemented in order to improve the correspondence with these observations. However, the BC lifetime of 3.4 days is rather short, so in this case the model is underestimating the high-level BC concentration. The true BC distribution is therefore somewhere in between that of HadGEM and NorESM/CESM-CAM4.

- Figure 3: Please mention in the caption that these means are for the surface.

Added the word “surface” before “temperature”

- Page 3834, Lines 8-11: Are similar drifts also present in the perturbation simulations (so that they cancel out and do not affect the differences between perturbation and control runs)?

The perturbation experiments are started as perturbations from the control runs (at the beginning of the 50-year period we analyse). The drifts, are therefore still present in the perturbation experiments.

Added sentence to pg 3834 line 11:
These drifts are also present in the perturbation experiments, since these start from the control simulation at the beginning of the 50 yr period analysed.

- Page 3834, Line 15: Please add “in” after “interested”.

Done

- Figures 4 & 6: Please mention in the caption whether the SW TOA fluxes are calculated for clear-sky cases only.

These figures show the all-sky SW TOA fluxes.

Added “all-sky” before “TOA SW flux” in both figure captions.

- Figures 5-8: Making the fonts of some of the labels (e.g. on the colour bar, or above the panels) somewhat larger would help with the readability of the figures.

We will consider this point at a later stage as the figures may be larger in the final (ACP) format than in the current (ACPD) format.

- Page 3835, Line 25: Please amend typo ("smilar").

Done

- Page 3835, Lines 25-28: Please clarify why it is more likely to be the aerosol indirect effect rather than direct (from pollution outflow and associated radiative effects).

Good point.

Added text and amended text on lines 22-28:
" **This is consistent with the decreased aerosol concentrations in this region due to the reduced emissions in China. As well as the direct radiative effects, the reduced aerosols would also cause changes in cloud cover. It was shown by Wang et al. (2014)** that Chinese aerosols increased cloud cover over the North Pacific, so removing these aerosols would reduce cloud cover. A similar region of positive TOA SW flux change also occurs over the North Atlantic, **which is similarly due to aerosol-radiation and aerosol-cloud effects over this region resulting from the aerosol emissions reductions over the eastern USA.**"

- Page 3836, Lines 1-2: Please support this with an example reference (there are plenty).

Added:
(Rotstayn et al., 2000, Broccoli et al., 2006)

- Page 3836, Lines 16-19: The reduction in precipitation is seen south of the equator, whereas the reduction in SW TOA flux is maximum just north of the Equator. How do the authors explain this inconsistency?

This is consistent with the ITCZ moving northwards. As well as the precipitation shift there will also be a northward shift in cloud related to the ITCZ, which leads to the decrease in TOA SW to the north (where the cloud increases). This was already discussed in pg 3835 line 28- pg 3836 line 2.

Reworded this sentence to clarify this:
In the tropics there are regions of decreased TOA SW flux just north of the equator and increased TOA SW flux just to the south. These relate to a northward shift in the ITCZ, which increases the cloud cover north of the equator and decreases it to the south.

- Page 3836, Lines 20-23: Worth highlighting the Sahel wettening as well. And mentioning a few key references, as done for the South Asian case.
-

Added sentence to line 23:
There is a large increase in precipitation over the Sahel. This is consistent with the results of Rotstayn and Lohmann (2002) who found that present-day anthropogenic sulphate aerosol had contributed to reduced precipitation in the Sahel.

- Page 3836, Lines 23-25: Any ideas why much of Europe and parts of the US become drier? Possibly circulation adjustments? Or a northern expansion of the NH subtropical regions?

The drying in these regions is not statistically significant so we don't want to put too much emphasis on it. We hypothesize that it could be linked to the northward shift of the ITCZ and a corresponding adjustment of the Hadley Circulation.

Added to text on page 3836 line 25:
“Over much of Europe and the USA there is a decrease in precipitation. While this is not statistically significant, we hypothesize that this is linked to the northward ITCZ shift and corresponding changes in circulation.”

- Page 3837, Lines 9-10: Not in temperature, it seems, as in Fig. 4a HadGEM seems a bit higher than ECHAM.

Corrected:
“...despite having the largest increase in precipitation (Fig. 4 c and d).”

- Page 3837, Line 13: I would say, “qualitatively agree”, as the agreement on the magnitude is not apparent, with one model showing half the response. You could then add that two of the models show very good quantitative agreement too.

Added “qualitatively” before “agree” in this sentence. Added a final sentence:
“HadGEM and ECHAM-HAM show very good quantitative agreement in the response.”

- Page 3837, Lines 24-25: It would be appropriate to refer to Table 2 here. And generally to mention Table 2 a bit more often in the text, as it can help the reader make linkages.

Added reference to Table 2 in the following places:
Pg 3834 line 24 (after “Fig 4a”)
Pg 3837 line 22 (after “Fig 4”)
Pg 3840 line 22 (after “Fig 4a”)

- Page 3838, Lines 4-5: However, it is worth mentioning here the findings of the (very)recent paper by Myhre and Samset (2015), which claims that current models tend to underestimate BC forcing.
-

Agreed.

Added: However, we note the results of Myhre and Samset (2015) which indicate that climate models may underestimate the forcing from BC by around 10%.

- Page 3838, Line 18: I think the authors meant to write “high-latitudes” instead of “highaltitudes”.

Yes – corrected.

- Page 3838, Lines 15-18: Was the methodology for generating the different ensemble members in CESM-CAM4 different to that in NorESM?Were the initial conditions in any way more drastically perturbed in the extra member of CESM-CAM4? Or do the authors believe that this disagreement/agreement between members is a totally random non-linear feature?

In both cases there were two different control runs each with a corresponding perturbation run, and the method of generating the two different control runs was essentially the same (in both cases a different model start dump was used, but everything else was the same). The surface temperature in the control simulations (Fig 3) shows that the CAM4 simulations seem to be generally more noisy/have more variability between years than the NorESM simulations, which could explain why the CAM4 members diverge more than the NorESM members. So yes, we believe it is a random non-linear feature.

Added to pg 3844 line 16:
"It is also interesting to note the very similar behaviour of the two NorESM members compared to the quite different responses between the two CAM4 members. In both cases the two members were generated by initializing with a different atmospheric state but keeping everything else the same. This further emphasizes the importance of using more than model, since some models are more sensitive to small perturbations in the initial conditions than others."

- Page 3839, Lines 5-7: Yes, but what about the widespread warming over Eurasia?

This does not seem related to sea ice.

The warming over Europe in HadGEM is discussed separately. However, we appreciate that this is a bit confusing and unclear.

We have rewritten this whole paragraph (pg 3838 line 19 to pg 3839 line 20) to aid clarity, and have take this comment into account.

- Page 3839, Lines 7-10: It is not as clear as what the authors claim. Sea-ice decreases

in some parts of the Arctic, but the changes are fairly localised, and there are even areas of increased sea ice.

We agree that this section was not very clearly written.

We have rewritten this whole paragraph (pg 3838 line 19 to pg 3839 line 20) to aid clarity, and have take this comment into account.

- Page 3840, Line 2: Preferably rephrase "model-mean" to "multi-model mean".
-

Done

- Page 3840, Lines 17-19: It is a bit counterintuitive that precipitation decreases over

land, where most of the BC exists, and where most of the de-stabilisation of the atmosphere is expected due to the BC removal. It would be useful here to briefly discuss possible explanations.

Good point. The precipitation changes are driven by circulation changes, rather than local effects of BC.

Added discussion to pg 3840 line 3:
"It is interesting to note that over India, where the anthropogenic BC emissions are large, the removal of the BC emissions results in a decrease, rather than an increase, in precipitation. These precipitation changes are driven by circulation changes (e.g. the southward shift in the ITCZ) which dominate over the local effects on precipitation due to BC removal causing destabilization of the atmosphere."

- Page 3841, Line 2: I would add “qualitatively” before “agree”, given the very much smaller positive changes in ECHAM, as seen in Fig. 5f.

Done

- Page 3841, Line 15: Please mention that comparisons of the short-term instantaneous or effective forcing are “(not shown)”, as otherwise the reader may be misled to think that you are referring to the SW flux comparisons pursued in the manuscript, which I presume is not the case (as the latter are the effect and not the cause of what is discussed here).

Agreed this is not entirely clear.

Amended sentence at end of line 13 to read:
“The **TOA SW flux change** from the OC emissions perturbation ...”

- Page 3842, Line 4: Not all is caused by the ITCZ shift. The higher latitude changes probably have to do with the thermodynamic effect of temperature increases.

Yes.

Added to end of line 4:
“Further precipitation increases are seen in the northern hemisphere due to the increased temperature, and in the Indian monsoon region, linked to the reduced aerosol emissions.”

- Page 3843, Lines 14-16: I presume the authors mean that there are fluctuations that happen at frequencies lower than 50 years (otherwise the average effect of natural variability would be negligible). This needs to be made clearer here.

Amended lines 14-16:
“...may therefore be driven by **changes in circulation leading to, for example, the change in cloud and snow cover over Europe,** which overwhelm the relatively weak forcing from the BC emissions perturbation.”

- Page 3843, Line 17: I would suggest changing “some” to “large”.
-

Done

- Page 3843, Line 28: It would perhaps make sense to add a sentence at the end of this paragraph speculating that possibly the expected effect is somewhere in the middle.

Added as suggested:
“The true BC distribution is most likely somewhere in between these model estimates.”

- Page 3844, Lines 14-15: I suggest rephrasing to: “where natural variability is a relatively large contributor” (as internal variability is what it is and does not have special features in this study).

Done

- Page 3844, Line 20: The scenarios are idealised, but are they really “extreme”, given the drastic decreases expected for the aerosols examined here in the future? It might be worth comparing these reductions with changes between present-day and e.g. 2100 in a typical future emissions scenario, to put things in perspective. Also, it is worth discussing here or in the Discussion section the possible future role of nitrate aerosols, and whether their inclusion in the models could have led to different conclusions or not (both regarding the future role of aerosols in general and regarding the effects of the aerosol types examined here, e.g. sulphate).

Since we are simulating 100% emissions reductions, and these could never realistically be achieved, we believe that these are extreme scenarios, since future emissions could never be less than this (i.e. negative).

Regarding the nitrate discussion, we have added the following to pg 3845 line 2:
“We note that the models used in this study do not respresent nitrate chemistry. This means that they may be overestimating the climate responses to removal of SO_2 emissions, since reducing SO_4 would increase the potential amount of ammonium nitrate aerosol formation, counteracting some of the effects of the reduced SO_4 aerosol.”

- Page 3844, Line 22: Perhaps add “mainly” before “using”.

Done

- Page 3844, Lines 23-24: Suggested rephrasing: “: : : to capture the fast and slow responses due to these emissions perturbations, as well as the uncertainties in these responses.”

Done

- Page 3845, Line 10: Suggested rephrasing: “: : : AOGCMs due to responses in ocean temperature and circulation, sea-ice, and atmospheric circulation and cloud responses that are realised on long timescales.”

Done