"Regional and global climate response to anthropogenic SO2 emissions from China in three climate models" by M. Kasoar et al.

Author response to anonymous referee #1

The authors wish to express their sincere gratitude to the anonymous referee for their invaluable comments and appraisal of our study. They have provided plenty of thought-provoking points, and we very much appreciate the time taken to do so.

Below we detail our responses to each major and minor comment in turn. We hope that these responses will satisfactorily address all the points raised. The referee's comments are given in italics, below which we provide our responses and the details of any changes made in the manuscript in normal font.

Comment 1:

"*The advantage of a model intercomparison study is that it allows for a clean juxtaposition of models. Yet, fundamental model diagnostics differ between the models, and I find that this very much complicates the comparison and limits the ability to draw firm conclusions beyond the statement that the models differ. I find the lack of clear-sky shortwave fluxes for CESM most striking - clearly this is a standard diagnostic, and I know that CESM has this diagnostic implemented. So why is it not available for the runs provided here? Having the clear-sky shortwave diagnostic would greatly aid the discussion of cloud effects in Sect. 4.2. Similarly, why is AOD diagnosed differently across the models, which seems to inhibit firm conclusions about AOD differences and aerosol radiative efficiencies. And finally, why is there no measure of internal variability available for CESM? I understand that this has to do with the lack of ensemble control runs (available for HadGem) or one long control run (as for GISS), but why have such runs not been performed. Aren't the authors in charge of the simulations presented here? I think the paper could be much stronger if the above limitations were addressed and the model setup and experiments were designed such as to eliminate them.*"

We acknowledge that with respect to some variables an ideal comparison could not be made, and the conclusions we could draw are more limited as a result, because of inconsistencies in which standard diagnostics were saved from these simulations. With regard to the most notable deficiency identified here though, we have now performed extended simulations with CESM in order to output the clear-sky shortwave fluxes for a 30-year period, and have therefore been able to substantially expand on Section 4.2 as desired.

With regard to the discrepancies in the manner AOD is diagnosed across the models, this was not the authors' choice – unfortunately clear-sky AOD was not available from the present CESM configuration, and likewise all-sky AOD is not available from the present HadGEM configuration. We certainly agree that it would have been useful to have consistent diagnostics from CESM, but we include this model in the paper because the available diagnostics nonetheless provide an interesting additional angle, although we believe the results would have been valuable even based just on the two extreme cases of HadGEM and GISS.

Performing a very long, or an ensemble of control runs with CESM would require considerable additional time. We feel that the advantage of being able to include an additional state-of-the-art model outweighs the disadvantage of these lengthy additional simulations not yet being available. We have demonstrated a very large uncertainty in the climate model response to SO_2 emissions using three models. This is important to publish given the number of single model studies that have appeared recently in the literature and that have not always considered structural uncertainties in these papers. While performing additional simulations or implementing new diagnostics would certainly allow deeper investigation of the model differences, we maintain that our analysis in this paper robustly backs up the points we make in the conclusions, and that it is important to make this paper available to the community now rather than delay it.

Changes made:

1) Added CESM1 changes in clear-sky versus all-sky SW flux to Supplementary Figure S10

2) Removed sentence in Section 4.2 saying that similar comparison could not be made with CESM, and added three new paragraphs:

   "The picture is different again for CESM1. Comparing the clear-sky and all-sky TOA SW flux changes for this model (Supplementary Figs. S10c,d), we find that regionally, the clear-sky changes are much smaller than the all-sky flux changes – in fact, over China the clear-sky SW flux changes in CESM1 are considerably smaller in magnitude than the clear-sky flux changes of GISS-E2 (comparing Supplementary Figs. S10a,c). Averaged over the E. China region, the clear-sky flux change in CESM1 is only 2.2 Wm^{-2}, compared with the 4.1 Wm^{-2} clear-sky change in GISS-E2 (Table 2). However, whereas in GISS-E2 the all-sky SW flux change (0.9 Wm^{-2}) was then more than four times smaller than this clear-sky flux change, in CESM1 the all-sky SW flux change is instead almost two times larger than the clear-sky flux change: 4.2 Wm^{-2} regionally averaged.

   This is partly again due to cloud changes – in this case CESM1 has predominantly reductions in cloud amount over E. China (Supplementary Fig. S11b), which will have the effect of increasing the all-sky radiative flux change relative to the clear-sky changes. However, as with HadGEM3-GA4, these regional cloud reductions in CESM1 do not match up spatially with the maximum changes in all-sky SW flux seen in Fig. 1b and Supplementary Fig. S10d. Instead, the maximum changes in the all-sky SW flux change match closely the clear-sky SW flux changes (Supplementary Fig. S10c), which in turn correspond very well with the reduction in AOD (Fig. 4b). Both all-sky and clear-sky SW flux changes are maximum around where the AOD reduction is maximum, and in this location the all-sky flux change is still substantially greater than the clear-sky change. This implies that in CESM1 a large aerosol indirect effect, and/or effect of clouds increasing aerosol particle size through hygroscopic growth, overall amplifies the radiative effect of aerosols considerably in cloudy conditions, resulting in the much greater regional change in all-sky flux when aerosol is removed.

   Between these three models, then, the way that clouds modify the radiative balance is a major source of diversity over the E. China region in the response to removing SO2 emissions from China. In GISS-E2, the inclusion of clouds greatly reduces the radiative effect of a change in sulfate aerosol. In HadGEM3-GA4, the effect of including clouds is small, and does not change the clear-sky forcing substantially. Finally in CESM1, including clouds considerably amplifies an otherwise weak clear-sky radiative flux change."

3) Removed fourth paragraph of Section 4.3, comparing CESM radiative efficiency using the all-sky flux, and replaced with new paragraph using clear-sky flux, consistent with HadGEM3 and GISS:

"CESM1 seems to sit in the middle of the range, with a regional radiative efficiency of -28.4 W m^{-2} per unit AOD change (Table 2) – though again with the caveat that for CESM1, the AOD is an all-sky quantity, whereas the HadGEM3-GA4 and GISS-E2 values were calculated using clear-sky AOD. GISS-E2 provided both clear-sky and all-sky AOD diagnostics, and using instead the all-sky AOD change from GISS-E2 gives a smaller value of -22.4 W m^{-2} per unit AOD, which suggests that when compared like-for-like, CESM1 (with -28.4 W m^{-2}) may in fact have the greater regional radiative efficiency. More directly comparable between all three models is the regional flux change normalised by regional change in sulfate burden, which for CESM1 is -55.4 W m^{-2} Tg^{-1}. This is considerably lower than either HadGEM3-GA4 or GISS-E2, and indicates that the despite having at least average radiative efficiency per unit AOD, the very small translation of sulfate burden to AOD in CESM1 is a dominant factor which prevents this model from simulating a larger SW flux change and climate response than it already does. As noted in the previous Section though, this small clear-sky flux change per unit sulfate change is compensated by a large indirect effect as well as favourable regional cloud changes, meaning that the all-sky flux change per unit AOD is by far the largest is CESM1 (Table 2), and the all-sky flux change per sulfate burden change is then average in CESM1 (not shown, but readily calculated from Table 2). Similarly, the exceptional reduction in aerosol radiative effects due to clouds in GISS-E2 means that its all-sky flux change per unit AOD is almost exactly the same as that of HadGEM3-GA4 (Table 2), despite the clear-sky regional radiative efficiency being so much larger."

4) Added clear-sky flux changes for all three models to Table 2 (formerly Table 1)

Comment 2:

"*As a result of the above I am wondering what I am supposed to take away from the current paper, apart from the statement that there is large model uncertainty. The authors attempt to trace the uncertainty to different sources, including aerosol chemistry (Sect. 4.1), cloud-radiative effects and aerosol-cloud interactions (Sect. 4.2), aerosol-radiative interactions (Sect. 4.3) and climate sensitivity (Sect. 4.4). None of these seems to be the sole smoking gun, though. While I appreciate that there maybe is no single factor that explains most of the uncertainty, what kind of experiments would be needed to better understand the individual contributions of the above four factors? I think a discussion of this question is needed in the conclusion section.*"

Indeed, we believe we show that there is no single smoking gun, but several different factors which contribute to the uncertainty, which are all important. We reiterate again that this is the first time such a comparison has been made between three different models forced with the same regional emissions change, and so even the statement that the models differ considerably in their responses, and for a complicated mixture of reasons, is we believe an interesting finding from the available data. If the situation is that the response is very diverse because of several different reasons, this is important to document, even if it is not a simple conclusion. However, we have clarified the conclusions to better highlight what appear to be the largest sources of disparity. We agree also that some additional discussion in the conclusions of how further experiments could help elucidate this problem is worthwhile, and this has also been added.

Changes made:

1) Changes to third paragraph of the Conclusion as shown by markup below:

   "Specifically, we find that CESM1 simulates the largest reduction in sulfate burden both globally and locally. HadGEM3-GA4 has the smallest reduction in sulfate burden globally and the second largest reduction regionally, yet it produces by far the largest reduction in AOD both globally and regionally over E. China. This much larger change in AOD per change in sulfate burden in HadGEM3-GA4 results in the largest radiative changes and the largest temperature response in this model. Though both GISS-E2 and CESM1 both simulate much smaller changes in AOD than HadGEM3-GA4, still the SW flux changes and temperature responses produced are very different between these two models. An inferred larger aerosol-cloud interaction means that CESM1 simulates a particularly large change in all-sky SW flux relative to its fairly small AOD change, so although having a smaller response than HadGEM3-GA4, it is still much closer to it than GISS-E2. In GISS-E2 the clear-sky radiative forcing efficiency of sulfate is very large, but this is almost perfectly compensated for by large reductions in the direct radiative effect of sulfate when clouds are factored in.radiative effect of sulfate burden changes appears smallest, The absolute AOD change is also much smaller than HadGEM3-GA4 in this model, and this then combines with compensating increases in local cloud amount over China and nitrate aerosol globally to reduce the radiative response yet further, and finally a smaller global climate sensitivity than the other two models results in this being translated into a largely negligible temperature response."

2) Split second paragraph of Conclusions into two, and moved the second half ("In addition to differences in sulfate and AOD…") after the third paragraph.

3) In the paragraph after, in the sentence "However, the main conclusion is that comparison against all existing observational measures is unable to satisfactorily constrain which model response is more realistic", added:

   ", given that the ratios of both AOD change per sulfate burden change and SW flux change per AOD (Table 1) are found to vary so substantially between the models"

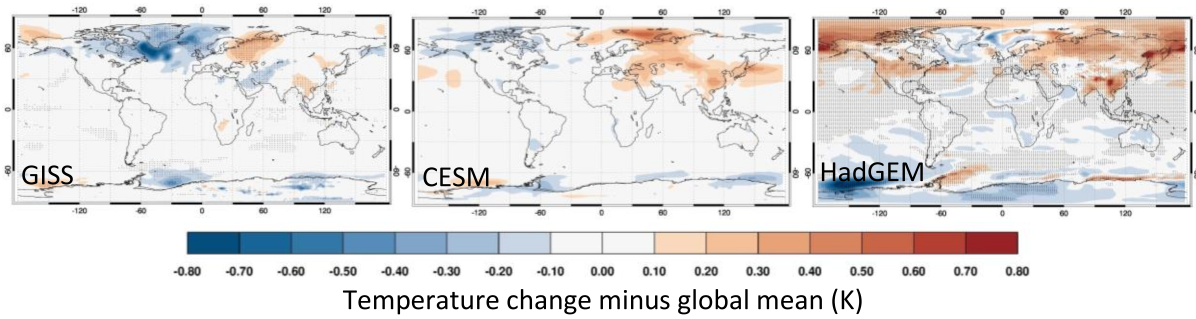
4) Added new paragraph to the Conclusions:

   "There are a number of possible avenues for future work to isolate the particular processes that lead to this model diversity in more detail; for instance studies imposing the aerosol field from one model into others would remove the diversity introduced by translating emissions into aerosol concentrations, while imposing surface temperatures and meteorology from one model into others could remove the diversity introduced by different background climatologies and climate sensitivities, although this may be difficult practically in complex climate models. A thorough assay of the range of parameter choices and formulae used in the aerosol schemes of various models could also help reveal where assumed aerosol properties diverge. However, without stronger observational constraints on aerosol radiative forcing, it is not clear that this alone could help make models more realistic. In particular, it seems that being able to better constrain the column-integrated sulfate burden, the AOD per sulfate burden, and the radiative forcing per AOD, would all also be needed. This represents a considerable observational challenge, and until it is possible, the considerable current diversity may be irreducible."

Comment 3:

"*Pattern of global temperature response: I am wondering to what extent the temperature patterns between the three models in Fig. 2 are more similar than acknowledged by the authors. What I mean is that GISS, while having no global-mean response, seems to show cooling in northern hemisphere regions in which CESM and Hadgem show relatively less warming (e.g., over the North Atlantic and Iran). Maybe the temperature patterns between the models look similar when the global-mean temperature change is removed? That would be interesting and point to robustness in the remote dynamical response.*"

This is an interesting suggestion, and we have now taken a look at this, but unfortunately it doesn't seem to show anything different – see plots below. Part of the problem we think is that what is seen in GISS is not really a response at all, but almost entirely noise.



Comment 4:

"*Reflecting on point 1, why is AOD diagnosed differently across the models? What is the motivation for this, and how to differences in the AOD diagnostics affect the results?*"

The first part of this is already addressed in the responses to Comment 1 – there was no deliberate motivation on the part of the authors, but unfortunately these are the diagnostics that were available from these model versions. And we still believe that the comparison is valuable. Regarding the second point here, it is consequently very difficult to know exactly how this will affect the results, however in Section 4.1 we do make comparison between the GISS-E2 all-sky AOD and CESM1 AOD, which should be more directly comparable, and we also note from the differences between the all-sky and clear-sky diagnostics in GISS-E2 that an all-sky diagnostic is likely to give larger values than the equivalent clear-sky diagnostic.

Comment 5:

"*At the end of section 4.1.1, I think a statement similar to the one on page 21, lines 23-25 would be helpful to wrap up this fairly complicated subsection, which simply seems to say that comparison to observations of current AOD doesn't help to constrain the model response.*"

This has been added.

Changes made:

1) Added new paragraph at end of Section 4.1.1:

   " Still, overall HadGEM3-GA4 seems to compare slightly better than GISS-E2 and CESM1 regionally over E. Asia against observations of total AOD, and better than GISS-E2 regionally against surface sulfate as well as wet deposition observations, although globally and over other regions this model is not necessarily found to compare better in general. This might hint that at least over China, HadGEM3-GA4 has more realistic sulfate optical depth, although none of these comparisons is very conclusive in that respect. Moreover, given that none of these observational measures directly constrains the sulfate radiative forcing, there is also no guarantee that performance with respect to these variables will necessarily translate to a more realistic climate response (see also Section 4.3)."

2) For greater clarification of the statement in the conclusion, also added an additional sentence at end of first paragraph of Section 4.3:

   "As a result, whether a model simulates AOD changes correctly, for instance, may not particularly constrain the resultant forcing and eventual climate response."

Comment 6:

"*Sect4.2, lines19, "what we would expect from a simple amplification of the radiative response due to indirect effects": Clear-sky shortwave changes will always be larger than all-sky shortwave changes because clouds mask some of the aerosol. So how can a comparison between clear-sky and all-sky changes inform about aerosol-cloud interactions (i.e., indirect effects)?*"

We agree that the highlighted sentence needed to be removed, as it is indeed mistaken. However we do still believe that the comparison made in the rest of this section, of the differences in the relative magnitudes of all-sky and clear-sky fluxes between the models, tells us something useful about the importance of cloud effects – although one cannot distinguish cleanly between microphysical and dynamical effects. (Indeed, the reviewer in their first comment also noted that: "Having the clear-sky shortwave diagnostic would greatly aid the discussion of cloud effects in Sect. 4.2", and so they presumably agree that something can be concluded from making such a comparison). In fact, the clear-sky flux changes need not necessarily be larger than the all-sky change if indirect effects are larger than direct effects, and this indeed seems to be the case for CESM, from the newly-added clear-sky diagnostics.

Changes made:

1) Removed:

   "In fact, in both models the clear-sky SW change turns out to be larger than the all-sky SW change, which is opposite to what we would expect from a simple amplification of the radiative response due to indirect effects. In particular GISS-E2 simulates an increase in cloudiness in East China when sulfate is removed, which…"

2) Replaced with:

> "…compared with the clear-sky change, the all-sky response incorporates all the contributing factors described above: the additional radiative forcing due to aerosol indirect effects, the screening of direct radiative effects due to clouds blocking radiation and providing a high albedo background, and also any dynamical changes in cloud cover.
>
> In this case, GISS-E2 is found to simulate a small increase in cloudiness in east China due to dynamical changes when sulfate is removed (Supplementary Fig. S11a). Combined with the screening effect of clouds, this…"

Comment 7:

"*Sect. 4.4: The idea to use global climate sensitivities derived for a uniform forcing to explain the local response to a highly localized forcings seems flawed to me to begin with, and indeed the authors find that global climate sensitivity does not help to understand the model differences. I suggest to condense this section into one or two sentences in the conclusion section.*"

The reviewer notes that we find the use of global climate sensitivities derived from a uniform forcing to be not particularly helpful in understanding the model differences – particularly between HadGEM3-GA4 and CESM1 (although GISS-E2 does have a known low climate sensitivity, which probably does contribute to this model having the lowest response along with the other factors discussed). However, we believe section is important partly to highlight this very fact. The comparison may be flawed, but yet global climate sensitivities are still typically used – very few studies have ever tried to calculate or use regional sensitivities. In meta-reviews like the IPCC AR5, it is typically implicitly assumed that the forcing due to inhomogeneous species like aerosols can be summed up with a global mean value for the forcing. As a result we believe this section still has value to draw attention to this. We already stress in this section that the comparison is flawed and that the global climate sensitivity to a uniform forcing should not be considered as equivalent to the climate sensitivity to a localised forcing, and highlight the lack of studies that have explored this issue.

Comment 8:

"*Instead, I would like to encourage the authors to expand their analysis of the changes in shortwave fluxes. The diagnostic approximate shortwave model of Donohoe and Battisti, J. Climate 2011 (Atmospheric and Surface Contributions to Planetary Albedo) would be a very valuable tool to understand the contribution of atmospheric and surface reflectivity to the changes in surface flux. One can further use the model for clear-sky and all-sky fluxes separately in order to distinguish aerosol effects (from the clear-sky use of the model) from cloud effects (when all-sky fluxes are used). I believe such an analysis has the potential to give much more insight and to grealy improve the paper.*"

We appreciate the reviewer's thoughts on potential further ways to expand on our analysis. We have considered the method suggested, but ultimately feel that our analysis in this paper already robustly backs up the points we make in the conclusions. Surface reflectivity changes appear to be unimportant to the responses over the East Asian region that we analyse (instance.g. we have verified, at least in HadGEM and GISS, that the local surface albedo is almost exactly the same in control and perturbation simulations), so in this case we do not feel that using the suggested additional model would change our analysis.

Minor comment 1:

"*Information about the shortwave radiative transfer schemes is missing in the model descriptions.*"

This information has been added.

Changes made:

1) Added to HadGEM3 model description:

   "The radiative transfer scheme of Edwards and Slingo (1996) is used with six spectral bands in the shortwave, and…"

2) Added to CESM1 model description:

   "Shortwave radiative transfer is based on the RRTM_SW scheme (Clough et al., 2005) with 14 spectral bands, and aerosols interact with climate through both absorption and scattering of radiation."

3) Added to GISS model description:

   "Aerosols direct effects are calculated following the Hansen et al. (1983) radiation model, with six spectral bands in the shortwave."

4) Added Edwards and Slingo (1996), Clough et al. (2005) and Hansen et al. (1983) to reference list.

Minor comment 2:

"*page 8, line 1: the East China box should be drawn in one of the figures for easier reference.*"

Done.

Changes made:

1) Box showing outline of E. China region added to all panels of Fig. 1.

2) Added to caption of Fig. 1:

   "The grey box denotes the East China (100°E - 120°E, 20°N - 40°N) region which is used in Table 1 and throughout the discussion."

3) Added sentence to end of second paragraph of Section 3 (where Fig. 1 is introduced):

   "For reference, Fig. 1 also shows the outline of the E. China region, which corresponds well to the region of maximum SW flux changes in all three models."

Minor comment 3:

"*caption figure 1: focuses –> focus*"

Corrected.

Changes made:

1) 'focuses' changed to 'focus' in Fig. 1 caption.