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

Author response

Below we first detail our responses to the anonymous referees' comments. We then append a copy of the revised manuscript with tracked changes.

"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:

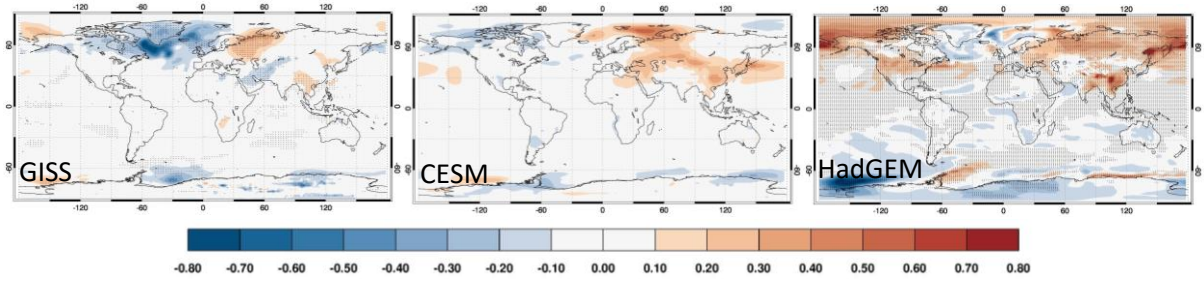
   "There are a number of possible avenues for future work to isolate the particular processes that lead to this model diversity in more detail; for instance studies imposing the aerosol field from one model into others would remove the diversity introduced by translating emissions into aerosol concentrations, 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.



Temperature change minus global mean (K)

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.

"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 #2

The authors are extremely grateful to the reviewer for their extremely helpful and positive comments. We very much appreciate the time taken to do provide these comments, which have helped highlight some areas of the paper where we were unclear or not precise enough.

Below we detail our responses to each minor comment in turn. We hope that these responses will satisfactorily address all the points raised. The referee's comments are included in italics, with our response to them and relevant changes to the manuscript in normal font.

Minor comment 1:

"*Only temperature (and no other climate) responses are addressed, and this should be reflected in also in the title.*"

Modified.

Changes made:

1) 'Climate' changed to 'temperature' in the title

Minor comment 2:

"*The descriptions of the three models in section 2.1 should be harmonized. It is especially important to provide the readers with a detailed enough summary of the aerosol and sulfur cycle treatments in each model – currently quite little is told about CESM1 and GISS-E2 aerosol/sulphur. The treatment of aerosol-cloud interactions within each model should also be briefly summarized.*"

We have attempted to harmonise the descriptions of the models through providing some additional details on CESM1 and GISS-E2, while slightly cutting down unnecessary text in the HadGEM3 description (we note that another reviewer actually thought our description of HadGEM3 was already too long, and so providing the right level of detail without hurting the flow and main message of the paper is a difficult balance). We have also, at the suggestion of the third reviewer, collated key details of the models into a table for easier reference.

Changes made:

1) In HadGEM3-GA4 description, removed:

   "…, dynamically resolving the stratosphere"

   "…, which includes 4 soil layers and 5 plant functions types. Although in principle this can be run in a fully interactive 'Earth System' mode with dynamic vegetation and a carbon cycle,…"

   "More detailed description and evaluation of the atmosphere and land surface schemes can be found in Walters et al. (2014)."

"Critical to our study is the representation of aerosols; we…"

"…, which is described and evaluated in…"

"The remaining aerosol species are emitted directly in the particulate phase, and…"

"…can then undergo advection, wet and dry deposition, and…"

2) In HadGEM3-GA4 description, inserted:

"(Walters et al., 2014)" in first and second sentences.

"HadGEM3-GA4 can be run with a choice of two aerosol schemes of differing complexity – CLASSIC (Bellouin et al., 2011), and GLOMAP (Mann et al., 2010). Here we use the simpler CLASSIC scheme, which is less computationally expensive, and is also the aerosol scheme that was used for CMIP5 simulations with the predecessor of this model (HadGEM2). CLASSIC is a mass-based scheme, meaning that only aerosol mass (and not particle number) is tracked, and therefore all aerosol species are assumed to be externally mixed."

"…mass…" in the sentence: "Cloud droplet number concentration and effective radius are determined from the mass concentration of these aerosols…"

plus minor connecting words so that sentences still read correctly after the phrases removed above.

3) In CESM1 description, removed:

"…modal aerosol scheme…"

"…from anthropogenic and natural…"

4) In CESM1 description, added:

"CAM5-Chem uses the MAM3 modal aerosol scheme (Liu et al., 2012), which is the same as used for the CESM1 submission to CMIP5. Both aerosol mass and particle number are prognostic, and the scheme simulates sulfate, black carbon, primary organic matter, secondary organic aerosol, dust, and sea salt aerosol species as an internal mixture in Aitken, accumulation, and coarse modes."

"The model includes emissions of natural and anthropogenic SO2 and natural DMS as sulfate precursors, and…"

"Aerosols-cloud interactions allow for the effect of aerosols on both cloud droplet number and mass concentrations (Tilmes et al, 2015)."

5) In GISS-E2 description, split second paragraph in to two and moved "nitrate, elemental and organic carbon along with secondary organic aerosols and natural sea-salt and mineral dust"

from the last paragraph to the new third paragraph.

6) In GISS-E2 description, replaced "SO_2 from anthropogenic and natural sources…" with "SO_2 from these sources…"

7) In GISS-E2 description, added:

   "GISS-E2 has a choice of three aerosol schemes of varying complexity – OMA (Koch et al., 2011; 2006), MATRIX (Bauer et al, 2008), and TOMAS (Lee and Adams, 2012). Following the GISS-E2 CMIP5 configuration, we use here simpler mass-based OMA scheme, which includes sulfate, …"

   "Aerosols are parameterised as an external mixture of dry and dissolved aerosol, with particle size parameterised as a function of relative humidity (Schmidt et al., 2006)."

   "includes natural emissions of DMS, and natural and anthropogenic emissions of SO2."

   "…, such that cloud droplet number concentration and autoconversion rate depend on the local concentration of aerosol."

8) Added a new table (Table 1; previous Table 1 is now Table 2) with key model details, as described in response to Referee #3 Minor Comment 3.

Minor comment 3:

"*P5L12: What does 'mass based' scheme mean in this context when modes and bins are also treated? P5: Is aerosol microphysics (condensation, coagulation, etc.) treated in CLASSIC?*"

We mean that only the mass concentration of each aerosol species (as opposed to number concentration) is tracked within each of the Aitken, accumulation, and dissolved modes. We have clarified this part of the description. Because only the mass of aerosol within each mode is tracked, microphysics is parameterised to allow transfer of mass between the different modes, based on the mass concentrations of each mode.

Changes made (also included in response to Minor comment 2 above):

1) Added "CLASSIC is a mass-based scheme, meaning that only aerosol mass (and not particle number) is tracked, and therefore all aerosol species are assumed to be externally mixed" in the description of CLASSIC

Minor comment 4:

"*P6L9-10: Does this mean that chemistry is solved online? The formulation here seems overly complicated.*"

Yes – this has been clarified.

Changes made:

1) In CESM1 description, added "…online…" in "…we use an online representation of tropospheric and stratospheric chemistry…"

Minor comment 5:

"*P7L1: 'aerosol-coating of dust': Dust is an aerosol particle itself; do you mean (secondary) coating of dust?*"

Yes – have amended to clarify.

Changes made:

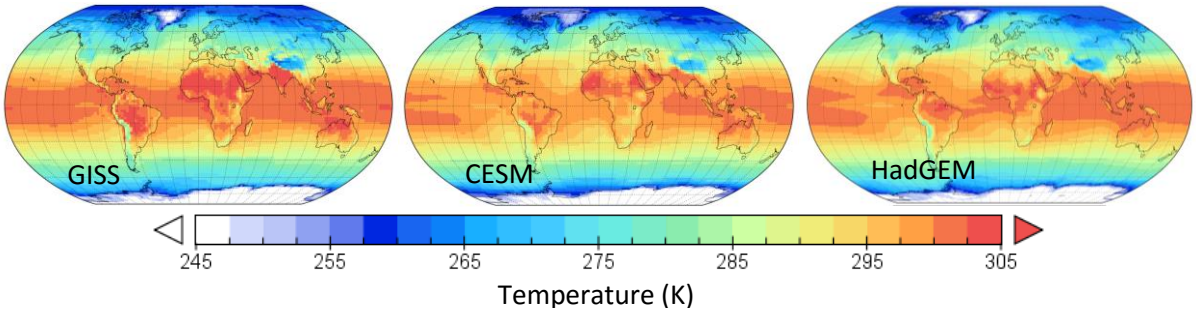
1) Changed "aerosol-coating of dust" to "secondary coating of dust"

Minor comment 6:

"*How different are the control climates between the different models? Would you expect this to impact your results?*"

The control climates are fairly similar between the models – an annual mean climatology is plotted below for comparison. If compared with observations, all three models have similar magnitude temperature biases. GISS is a bit too warm in the tropical oceans, CESM is a bit too warm over the northern mid-latitude land, all three - although especially HadGEM - are too warm in the Southern Ocean, and possibly too cold over the polar regions, by a few degrees in each case. On average, GISS is about ½ a degree warmer than HadGEM, which is about ½ a degree warmer than CESM.

In terms of whether this would impact our results – we do not think it could have a substantial difference to the models' responses to an aerosol emissions perturbation. Firstly, because the changes in SW flux themselves explain much of the diversity in the models' temperature responses. The effect of climate feedbacks and climate sensitivity may play a role in setting the exact magnitude of the final response, but these vary between models anyway, unrelated to the climate state, and so this is part of the structural uncertainty we wish to explore. To our knowledge, studies that have looked for example at the time-dependence of climate sensitivity and feedbacks in transient warming scenarios generally find that it varies slowly, and so inter-model variations in climate sensitivity are likely much more important than the base climate state, unless this were to be very different.



Minor comment 7:

"*P7L27: Are the runs restarted from an earlier simulation? 50-year spin-up by itself doesn't seem sufficient for a coupled model.*"

Yes, the runs were restarted from previous coupled simulations that had already been run for present-day conditions, though not necessarily with the exact model set-up that was used here. The 50 years is not intended to spin-up the control runs, but rather to allow the response to the perturbation to establish itself. We have expanded the experimental setup section to clarify this. Previous studies which apply an abrupt forcing (e.g. Andrews et al. (GRL, 2012, doi:10.1029/2012GL051942) have generally seen that most of the global surface temperature response is realised within this timeframe, and from inspection of the time series of global temperature changes, this seemed to be the case here as well.

Changes made:

1) In Section 2.2 (Experimental Setup), added ", initialised from a present-day state," to the description of the control simulations, and "from the same initial state," to the description of the perturbation simulations.

2) In first sentence of Section 3, replaced bracketed phrase "the first 50 years were discarded as spin-up" with "the first 50 years are discarded to allow the response to the perturbation to establish itself".

Minor comment 8:

"*P10: Both HadGEM and CESM1 simulate H2O2 and O3 oxidation pathways in the aqueous phase, so including both pathways cannot be an explanation to fast conversion to SO4 in HadGEM. This should be explicitly stated.*"

We agree that including both pathways cannot explain any differences in the SO_2 oxidation rates between HadGEM3 and CESM1, only for HadGEM3 and GISS-E2. We have added a sentence explicitly stating this.

Changes made:

1) Added additional sentence at the start of fourth paragraph of Section 4.1:

   "CESM1 includes the same oxidation pathways as HadGEM3-GA4, and in fact has a slightly shorter SO2 lifetime still, and so the differences between these two models have different origins."

Minor comment 9:

"*P16L6-8: Do you refer to sulphate aerosol above cloud top here? Simulated cloud distributions can have large impacts also in other ways, e.g. the background aerosol amount (clean/polluted) has large impacts on indirect effects, which start to saturate at high aerosol concentrations.*"

Yes – as suggested we've rephrased this sentence to add a mention here of other ways cloud distribution has a potential impact via the saturation of indirect effects (plus reference).

Changes made:

1) Added to first paragraph of Section 4.2 (section in square brackets was already there):

   "For instance, [the radiative effect of sulfate aerosol is modulated by the reflectivity of the underlying surface in the radiation scheme (Chýlek and Coakley, 1974; Chand et al., 2009), which may often be a cloud-top.] The low contrast with a highly reflective cloud surface means that sulfate aerosol above a cloud top will have a reduced direct radiative forcing. Blocking of radiation by clouds will also reduce the direct radiative effects of any aerosols within or below them (e.g. Keil and Haywood, 2003). Additionally, aerosol indirect effects can saturate in regions with a high level of background aerosol (e.g. Verheggen et al., 2007; Carslaw et al., 2013), meaning that the potential for indirect radiative forcing can also vary with the location of clouds. On top of diversity in indirect effects, and in the climatological distribution of clouds, different dynamical changes in cloud cover could also alter the all-sky flux."

2) Inserted additional references (Keil and Haywood, 2013, Verheggen et al., 2007, and Carslaw et al., 2013) in bibliography

Minor comment 10:

"*P16L20-21: Can you speculate which dynamical processes cause the increase in cloudiness when sulphate is removed? Based on 2nd indirect effect one would assume decreased cloudiness.*"

Indeed, we also expected decreased cloudiness from the 2nd indirect effect, and so the observed increases in GISS are presumably dynamical in origin. Dynamical feedbacks can be complex and chaotic, and cause and effect hard to untangle. Moreover, despite all having local warming at the surface in east China, all three models have quite different regional cloud changes, and so whatever dynamical processes are at play are not robust. Therefore we do not wish to speculate further here.

"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 #3

The authors wish to express their sincere gratitude to the anonymous referee for their invaluable comments and positive appraisal of our study. They have provided thorough and 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 original comment is included in italics, with our response and change to the manuscript in normal font.

Comment 1:

"*While the perturbation applied is well specified, the model output and diagnostics retrieved seems to vary a lot. I realize it's hard to do anything about this once the simulations are done, but for later studies I would encourage the authors to use a wider output protocol. E.g. clear-sky vs all-sky fluxes should be possible to diagnose for all these climate models, and for sulphate perturbations their difference can be very instructive due to differences in treatment of the indirect effect.*"

We absolutely agree. This shortcoming was mentioned also by the first reviewer, and the lessons from this study are indeed being learnt in the discussion of potential future collaborations, which have developed following presentations of the results in this study. As described in the responses to Referee #1, we have in fact also taken the step of extending the simulations with CESM1 for a short period to diagnose the previously missing clear-sky SW flux (which we expect has lower variability than temperature, and so probably doesn't need the same 150-year averaging period), and so the discussion in Sections 4.2 and 4.3 has been updated with this new data.

Comment 2:

"*Page 8, line 23++: For a sudy such as this one, a good diagnostic of TOA RF is very useful. It can be extracted from relatively short and inexpensive fSST runs, as was done here for HadGEM3. I would encourage the authors to add this also for the two other models, and to take the results into their intercomparison discussions.*"

We do agree with the reviewer that including additional simulations would be helpful to get a more precise measure of the radiative responses. However, we have opted already to use the available time to extend the coupled simulations with CESM in order to diagnose clear-sky fluxes as requested by the first reviewer, which we decided was a more critical deficiency. Although more thorough RF diagnostics would be nice for consistency, we do not anticipate they would qualitatively change any of our findings, and we strongly believe that our analysis with the presently available diagnostics already robustly supports the points we make in the conclusions. Given the number of single model studies that have appeared recently in the literature and that have not always considered structural uncertainties, we believe these conclusions are already of sufficient importance and urgency to merit publishing this paper now, rather than incur the further delay and additional costs of additional simulations.

Comment 3:

"*Page 11, line 30-31: The GISS-E2 model has had some problems with its nitrate implementation, and e.g. pulled these results from AeroCom Phase II. Is this issue resolved for the simulations presented here? (I assume so, but still ask since nitrate here seems to be one of the drivers of intermodel differences.)*"

The GISS-E2 configuration used here is the AR5 version, meaning that it does still suffer from the issue of too high a nitrate burden, and probably an overly strong nitrate response as a result. This was, in fact, one of the first things we considered as a possible cause of the discrepancy between GISS and the other models. However, as we discuss in the paper, although there is some partial compensation by increases in nitrate, it turns out to still be a fairly minor factor in the inter-model differences in this study.

Comment 4:

"*Page 13, line 1-10: This section is very interesting, but briefly presented. I would suggest expanding it somewhat, perhaps adding some comparison plots? This would make the study even more useful for future model work.*"

P13, L1-10 discusses the comparison against AERONET, for which there is already a comparison plot in the supplement and we are not sure that there is much scope to expand on it. However, we think the reviewer may have meant Page 14, where we discuss the fractional change in AOD, which turns out to be much larger in HadGEM3 than in GISS-E2 or CESM1. In this case then yes, we do agree that this was rather interesting and could merit some more detail. We have therefore expanded the discussion and added two extra Supplementary Figures here showing firstly how the sulfate fraction of total AOD varies considerably between HadGEM3-GA4 and GISS-E2, and then also comparing the non-sulfate AOD to show that this is in fact similar in these two models, and so the discrepancy in the fraction of total AOD removed is primarily due to disagreeing on the sulfate optical depth only.

Changes made:

1) At the end of the third paragraph of Section 4.1.1, added:

   "This is illustrated further for the two extreme cases, HadGEM3-GA4 and GISS-E2, in Supplementary Fig. S3, which shows that the fraction of climatological AOD made up by sulfate is consistently higher across the east Asian region in HadGEM3-GA4 than in GISS-E2. However, the total non-sulfate AOD is fairly similar across the region in these two models (Supplementary Fig. S4), indicating that the stark difference in the fractional contribution of sulfate comes primarily from HadGEM3-GA4 simulating much greater sulfate AOD alone. Given that regionally GISS-E2 appeared to underestimate total AOD, this would then suggest that either the higher sulfate AOD in HadGEM3-GA4 is more realistic, or else both models underestimate the non-sulfate AOD."

2) Added new Supplementary Figure (S3), showing fraction of total AOD made up by sulfate in GISS-E2 and HadGEM3-GA4.

3) Added new Supplementary Figure (S4) showing total non-sulfate AOD (i.e. total AOD minus sulfate AOD) in GISS-E2 and HadGEM3-GA4.

4) Renumbered other Supplementary Figures accordingly.

Comment 5:

"*Page 15, line 12-30: This section discusses wet deposition results vs observations, and link good performance to a realistic SO4 distribution. However, isn't this also very dependent on the representation of precipitation? The China/Asia region has a lot of variability both in actual and modeled precipitation, and until it's shown that these compare to a reasonable degree I would be cautious about the above interpretation of wet deposition.*"

A valid point. We have added a caveat to this part of our discussion by noting that precipitation will influence the amount of local wet deposition, and so it is difficult to draw definite conclusions from this comparison. (Although, because wet deposition is the primary sink of sulfate aerosol, to some extent regionally it must balance the source of aerosol regardless of the precipitation, and so a large underestimate in the amount of wet deposition could be indicative of too low production of sulfate aerosol). At any rate, we do not rely on this single measure to determine which model is more accurate, but note that it appears consistent with the other observations that we compare with in suggesting that GISS-E2 likely simulates too little sulfate in the region.

Changes made:

1) Added sentence to end of wet deposition paragraph:

" This overall picture seems consistent with that of the other observational measures looked at here, although it should be noted that wet deposition rates are dependent not just on the column sulfate burden but also on the amount and distribution of precipitation however, and so biases in wet deposition could also be due to incorrect precipitation distribution rather than sulfate."

Comment 6:

"*Page 17, line 17-19: It's hard to assess if e.g. "a 3-fold larger clear-sky SW change" is significant without some indication of the internal variability. Since the results in this paper are mostly from 150-year integrations, I would encourage the authors to add more information on the year-to-year variability (i.e. just the standard deviation of the result across the integration, or similar) throughout the manuscript.*"

We agree with the reviewer that some desirable detail on the significance of the results was either omitted or hard to find, which we have tried to rectify. In our SW and surface temperature plots for GISS-E2 and HadGEM3 we did already include a measure of significance by stippling the plots and stated in the text which temperature responses were significant, but we have now extended that by including ± 2σ uncertainty values in the Table of global and regional responses for all variables that there were sufficient data to calculate it for. This includes the clear-sky SW changes in HadGEM3 and GISS, for which the discrepancy is seen to be extremely significant (around 23 standard deviations). Extending this to all variables in the Table is complicated by the fact that the very long control

simulation used to assess variability in GISS-E2 only output basic climate diagnostics and not more detailed aerosol-related diagnostics, and we have no equivalent long or ensemble control simulation at all for CESM. The SW changes and final temperature response are ultimately what we are interested in most though, so we do not think that this is too restrictive (and one can generally use the value given for HadGEM3 to get at least an order-of-magnitude estimate of the likely uncertainty where a value isn't available for the other models). We deliberately avoided estimating the significance of other variables from the year-to-year variability in these simulations though, because we do not think this necessarily leads to an accurate measure of the long-term 150-year variability which is the relevant quantity here, and on which we base our uncertainty analysis.

Changes made:

1) Added ± 2σ uncertainty values to the Ch0-Con differences in Table 2 (formerly Table 1), for all variables for which long/multiple control runs data were available (all of HadGEM3 + temperature and radiative fluxes for GISS). Added statement to Table 2 caption:

   "For models and variables where data was available, error ranges are quoted for the Ch0-Con values and indicate ± 2 standard deviations, evaluated in HadGEM3-GA4 from an ensemble of six 150-year control runs with perturbed initial conditions, and in GISS-E2 from twelve 150-year segments of a long pre-industrial control run. Values quoted without error ranges indicate that uncertainty was not evaluated."

Comment 7:

"*Table 1: The numbers listed here seem to have an unrealistically high precision (e.g. -0.034810…) Please give a reasonable number of significant digits, and also include some indication of the internal variability in each model (see previous comment).*"

We agree that the precision that the numbers are quoted to is implausibly high – this is an oversight that appears to have crept in from an old version of the table, and the numbers should have been truncated to fewer significant figures in the submitted version. This has now been corrected. See our response to the previous comment for discussion of internal variability and estimating significance – we have added error values into Table 1 for the variables and models for which these we had these figures.

Changes made:

1) Values in Table 2 truncated so that Ch0-Con values are at most 3 significant figures. Values for individual simulations have been truncated to at most the same number of decimal places as the Ch0-Con anomalies for that variable.

2) Added significance estimates to Table 2 as detailed in response to Comment 6

Minor comment 1:

"*Abstract (p2): "...and reinforces that caution must be applied when interpreting the results of single-model studies." I believe the results of this paper show that we should be cautions also in interpreting multi-model studies. They are usually just ensembles of opportunity, with little or no observational*

constraint beyond what is already taken into the model parametrizations. Hence their average values are not necessarily closer to reality, but instead just indicative of the present model diversity."

We have modified both the abstract and conclusion so as to not limit our statement to single-model studies.

Changes made:

1) In the abstract, changed 'single-model studies' to 'modelling studies'

2) Changed the corresponding line in the second last paragraph in the conclusion ("…and imply that care must be taken not to over-interpret the results of studies performed with single models") to:

   "…and imply that care must be taken not to over-interpret studies of aerosol-climate interaction if the robustness of results across diverse models cannot be demonstrated"

Minor comment 2:

"*Page 3, line 31-32: The Phase II AeroCom study (Myhre et al. 2013, ACP) which you cite later probably belongs in this company.*"

We agree, and have added a reference to this paper in that section as well.

Changes made:

1) Added 'Myhre et al., 2013' to bracketed list of HTAP and AeroCom references.

Minor comment 3:

"*Section 2.1: The description of HadGEM3-GA4 is very long compared to the two other models. Could the descriptions be clarified and made more uniform? Perhaps through a table of the most relevant model parameters/physical processes included?*"

Agreed – this is something that has been mentioned by another reviewer as well, though the other reviewer favoured more detail for CESM1 and GISS-E2 rather than less for HadGEM3-GA4. We have slightly cut down superfluous details in the HadGEM3 description while adding several additional details to the other model descriptions and slightly re-ordering them to make the descriptions more uniformly structured. As recommended, we have also added a new table which includes key references and features of the three models for easy reference.

Changes made:

1) Numerous changes to model descriptions which are detailed in responses to Referee #1 Minor Comment 1 and Referee #2 Minor Comments 2, 3, 4, and 5, which harmonise the model descriptions.

2) In the first paragraph of Section 2, added:

   "The models are briefly described below, and the key references and features are also

summarised in Table 1."

3) Added a new table (Table 1; previous Table 1 is now Table 2) with key model details. Updated all previous instances of "Table 1" in the text to "Table 2", and updated the caption of the existing table to Table 2. Added caption to new table:

   "Table 1: Key references and features of the three models and their aerosol schemes used in this study"

Minor comment 4:

"*Page 18, line 1-2: The SO4 forcing is not very sensitive to the vertical distribution, compared e.g. to absorbing species. See e.g. Samset and Myhre, GRL 2011, doi:10.1029/2011GL049697.*"

This is very true. We have removed that speculation, and found a different (partial) explanation:

Changes made:

1) Removed "For instance, the forcing per unit AOD will be influenced by the vertical distribution of the aerosol (Myhre et al., 2013a), which could vary between models in different parts of the world."

2) Replaced with:

   "The sulfate efficiencies in Myhre et al. (2013) are calculated relative to all-sky direct radiative effect, and so local differences in vertical profiles and cloud screening may therefore change the relationship – however they also evaluated clear-sky forcing normalised by AOD for all aerosol species combined, and again found HadGEM2 to be higher than GISS ModelE."

3) Additionally, at end of this section, added text indicated in the mark-up below:

   "However, the study also found that, globally, the atmospheric component of HadGEM2 had a slightly larger~~very similar~~ forcing efficiency to CAM5.1 both for sulfate (all-sky) and all aerosols (clear-sky), unlike the somewhat smaller regional efficiency found here for HadGEM3-GA4 compared with CESM1. Given that our regional values from GISS-E2 and HadGEM3-GA4 also seem to conflict qualitatively with the global values from the AeroCom study, this would suggest that either the global comparison is not relevant on regional scales, or else the radiative efficiency is very sensitive to changes in model configuration and version.

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

Additional changes:

1) Added "cloud radiative interactions" to list of key discrepancies in the abstract

2) Added Boucher et al. (2013) reference to the overview of aerosol radiative effects in the Introduction

3) Added Meinshausen et al. (2011) reference for the CMIP5 greenhouse gas concentrations used in the model description (Section 2.1)

4) Added additional paragraph to the Experimental Setup (Section 2.2):

   "Additionally, shorter atmosphere-only simulations were performed with HadGEM3-GA4 (identical in setup except that sea-surface temperatures (SSTs) and sea-ice cover are prescribed to year-2000 values) to diagnose the effective radiative forcing, as well as the SO_2 oxidation rates and SO_4 wet deposition rates for this model, referred to in Section 3, Section 4.1, and Section 4.1.1. In CESM1, the SO_2 burden, surface SO_4 concentration, clear-sky radiative flux, and cloud cover referred to in Sections 4.1.1, 4.2, and 4.3, were all diagnosed from a 30-year extension of the control and perturbation coupled simulations, rather than from the original 200 years."

   Removed phrase "where sea-surface temperatures (SSTs) and sea-ice cover were prescribed to year-2000 values" where it originally occurred later on in 4th paragraph of Section 3.

5) HadGEM3-GA4 plots for SW flux change and surface temperature change have been replotted to fix an error in the location of a small number of the significance stipples. The discussion of the plots is unaffected.

6) HadGEM3-GA4 plots for surface air temperature changed to show 1.5m temperature anomaly rather than surface temperature, as this is probably more consistent with the other model's surface air temperature diagnostics. The (Ch0 – Con) change is almost identical though, and the discussion is not affected (except that global mean temperature changes from 0.114 K to 0.115 K))

7) Changed name of Section 4.1 from "Differences in simulated aerosol amounts" to "Differences in simulated aerosol amounts and aerosol optical depths"

8) Added clarification at start of 2nd paragraph of Section 4.1 of the source of chemistry diagnostics:

   "For GISS-E2 and HadGEM3-GA4, more detailed chemistry diagnostics were available from a 5-year period of a HadGEM3-GA4 atmosphere-only control simulation, and a 5-year period of the GISS-E2 coupled control simulation. For these two models,…"

9) Added new penultimate paragraph to Section 4.1 (and references therein):

"The AOD changes per unit burden change are summarised in Table 2, and it is clear that there is a large diversity between the models. The possible contributors to diversity in the AOD per unit burden are extensive, and a full analysis of them is beyond the scope of this paper. Host model effects, such as different cloud climatologies and radiative transfer schemes, are one likely contributor. Stier et al. (2013) suggests that one third of total diversity originates there. Relative humidity, which drives water uptake (hygroscopic growth), is also diverse among models. For example, Pan et al. (2015) find that over India, boundary-layer RH is the main source of diversity. At the more basic level, assumed composition and hygroscopic growth curves also often differ between models – in this case, the aerosol scheme used for HadGEM3-GA4 assumes that all sulfate is in the form of ammonium sulfate, whereas CESM1 and GISS-E2 both assume a mixture of ammonium sulfate and sulfuric acid, and additionally all three models use different sources for their hygroscopic growth parameterisations (Bellouin et al., 2011; Liu et al., 2012; Koch et al., 2011; and references therein)."

10) Added CESM1 to Zhang et al. surface SO_4 comparison figure and IMPROVE comparison figure (Supplementary figures S5 and S7), and added CESM1 station biases in each case to the text in Section 4.1.1.

11) Added CESM1 to climatological column SO_4 figure (Supplementary Figure S6)

12) OMI SO_2: Added an extra Supplementary Figure (S8), to additionally compare column SO_2 in GISS-E2, CESM1, and HadGEM3-GA4 with satellite observations from the Ozone Monitoring Instrument (OMI). Split the 2nd last paragraph in Section 4.1.1 (dealing with wet deposition observations) into two, in order to insert a short paragraph about OMI SO_2 as follows:

"…Returning to Asia, we therefore also tried evaluating the models against column sulphur dioxide observations. We use the gridded, monthly mean Level 3 observations from the Ozone Monitoring Instrument (OMI) (Krotkov et al, 2008) (available from http://disc.sci.gsfc.nasa.gov/Aura) which is flown on the Aura satellite, averaged over eight years from 2005 - 2012. Over the E. China region the mean OMI SO_2 is 0.153 Dobson Units (DU), and all three models appear to overestimate this substantially, with very similar regional mean SO_2 columns of 0.282 DU for HadGEM3-GA4, 0.272 DU for GISS-E2, and 0.259 DU for CESM1. Spatially, all three models have more diffuse SO_2 fields than the OMI observations, where the SO_2 burden seems much more localised around source regions (Supplementary Fig. S8). This may be partly due to the coarse resolution of the models compared with the 0.25°satellite product, but also suggests that the lifetimes for SO2 may be too long in both models, or transport processes too efficient. The surprisingly similar column SO_2 burdens in all three models suggests that, at least on regional scales, column SO_2 may not constrain SO_4 burden that well.

An alternative observational measure which to an extent reflects a column-integrated quantity is the deposition rate, and for the two extreme cases of HadGEM3-GA4 and GISS-E2 ~~W~~we therefore also try comparing against observations of sulfate wet deposition. We use the 3-year mean wet deposition

data from 2000-2002 described in Vet et al. (2014)…"

13) Clear-sky SW flux data for GISS-E2 replaced with data from a different clear-sky diagnostic which should be more comparable to the way this variable is calculated in the HadGEM3-GA4 and CESM1 diagnostics. HadGEM3-GA4 clear-sky data in Table 2 updated (was previously diagnosed at the surface, now diagnosed at the TOA to be consistent with all-sky diagnostics). Also updated GISS-E2 and HadGEM3-GA4 clear-sky and all-sky SW flux changes in Figure S10 to show TOA flux changes, using the updated GISS diagnostic.

14) Third and fourth paragraphs of Section 4.2 largely re-written to reflect new GISS-E2 clear-sky SW diagnostic, which points to a larger role for cloud interactions in reducing the sulfate radiative forcing in this model. They now read:

"For the extreme cases of HadGEM3-GA4 and GISS-E2, comparing the changes in clear-sky TOA SW flux with the all-sky TOA SW flux anomalies (Table 2 and Supplementary Fig. S10) reveals that for clear-sky conditions, there is in fact a much smaller regional discrepancy between these two models: Over the E. Asia region GISS-E2 has a 4.1 Wm-2 clear-sky SW flux change, whereas HadGEM3-GA4 has a 5.1 Wm-2 flux change. HadGEM3-GA4 still has the larger radiative change, but nowhere near the 6-fold difference that is seen in the all-sky flux (Section 3, and Table 2). This much reduced difference between GISS-E2 and HadGEM3-GA4 in the clear-sky compared with all-sky anomaly is hard to apportion quantitatively though, because 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 appears to almost completely offset the direct forcing of reduced SO4, and results in a far smaller all-sky flux change than clear-sky flux change over E. China (0.9 Wm-2 all-sky compared with 4.1 Wm-2 clear-sky). HadGEM3-GA4 by contrast has very little difference between all-sky and clear-sky flux changes (5.3 Wm-2 and 5.1 Wm-2 respectively (Table 2)). The changes in cloud amount over east China are somewhat more mixed (Supplementary Fig. S11c), though area-averaged, the overall cloud change is a small decrease, which should enhance the all-sky flux change. However, spatially as well as in magnitude the HadGEM3-GA4 all-sky flux change is exceptionally similar to its clear-sky flux change, and does not resemble the pattern of cloud changes (comparing Supplementary Figs. S10e,f, and Fig. S11c), which suggests that aerosol radiative effects are larger than the effect of the small cloud cover changes, and still dominate the all-sky flux changes. Therefore, the very similar regional all-sky and clear-sky SW flux changes in HadGEM3-GA4 implies that unlike in GISS-E2, aerosol indirect effects in HadGEM3-GA4 probably roughly compensate for the presence of clouds reducing the direct effect, so that the change in all-sky combined direct and indirect forcing is similar to the change in clear-sky direct forcing when sulfate is removed."

15) Added additional Supplementary Figure (S11) showing regional cloud cover changes in all three models (referred to in Section 4.2 where it previously said 'not shown')

16) Added additional caveat to end of Section 4.2:

"We note though that clear-sky diagnostics will be influenced by choices within the models of how aerosol water uptake is determined under the artificial assumption of clear-sky conditions. The all-sky SW flux change, which drives the final climate response, is regionally still the most directly comparable quantity, reflecting the total radiative effect of the aerosol change."

17) In second paragraph of Section 4.3, added/changed marked-up text in following sentence:

"This is not directly comparable with previous studies like Myhre et al. (2013a), as we use a regionally-averaged number instead of globally-averaged, and for the numerator we use the change in clear-sky TOA SW flux as the best available measure of aerosol direct radiative effect, rather than the ~~clear-sky~~ direct radiative forcing diagnosed either from double radiation calls or simulations with fixed meteorology.

18) In third paragraph of Section 4.3, changed regional radiative forcing efficiency values for HadGEM3-GA4 and GISS-E2 to reflect new diagnostics used in Table 2. This results in GISS-E2 have a much higher value than HadGEM3-GA4, rather than just a somewhat higher value. It also changes the flux change normalised by sulfate burden change so that GISS-E2 is now bigger than HadGEM3-GA4, rather than smaller. Relevant comparative statements in this paragraph were therefore changed as shown in mark-up:

"As noted in Sect. 4.1 and 4.2, over the eastern China region HadGEM3-GA4 has a 6-fold larger mean AOD reduction (-0.29) compared with GISS-E2 (-0.047), but only slightly~~a 3-fold~~ larger clear-sky SW change (5.1~~8~~ W m-2 compared with 4.1~~.8~~ W m-2). As a result the regional radiative efficiency for HadGEM3-GA4 is much smaller than~~only about half~~ that of GISS-E2: ~~(~~-17.6~~20.3~~ W m-2 compared with -~~39.1~~87.2 W m-2~~)~~ per unit AOD change (Table 2). If instead of AOD we normalise by the change in sulfate burden ~~instead of the AOD~~ integrated over the same region~~, however,~~ we find a similar~~the opposite~~ relationship: HadGEM3-GA4 has a smaller~~larger~~ regional mean change in clear-sky SW flux per Tg sulfate than GISS-E2: ~~(~~-145~~67.1~~ W m-2 Tg-1 compared with -256~~117.7~~ W m-2 Tg-1~~)~~. Proportionally though, the discrepancy is not as great when normalising by change in sulfate burden, due to the ~~The~~ much larger AOD per unit mass of sulfate simulated in HadGEM3-GA4 ~~therefore outweighs the smaller radiative response per unit AOD~~. Curiously Myhre et al. (2013a) reported results that were qualitatively the inverse of what we show here, finding that the atmospheric component of GISS ModelE~~2~~ has a smaller sulfate radiative forcing than that of HadGEM2 (HadGEM3's predecessor, with a very similar aerosol scheme) when normalised by AOD, although still~~but~~ larger when normalised by column-integrated sulfate burden.

19) In final paragraph of Seciton 4.3, inserted the word "all-sky" into sentence:

"In their case, they found CAM5.1 to have approximately 2.25 times higher all-sky direct

radiative forcing per unit AOD than GISS-E2."

20) In Section 4.4, replaced (Samset et al., in preparation) with (Samset et al., 2016) and updated reference in the reference list, since this paper is now published.

21) Last paragraph of Section 4.4, removed four instances of the word 'regional' when referring to the Shindell (2012) study which looked at forcings imposed in different latitude bands, to avoid confusing with the more localised usage of 'regional' throughout the rest of the paper to refer to the China /East Asia region. In the last sentence of this section, replaced 'regional forcings' with 'forcings at different latitudes'.

22) Other minor grammatical and readability changes – see tracked changes in full manuscript for details

Compete revised manuscript with tracked changes:

Regional and global ~~climate~~ temperature response to anthropogenic SO_2 emissions from China in three climate models

M. Kasoar[1], A. Voulgarakis[1], J.-F. Lamarque[2], D. T. Shindell[3], N. Bellouin[4], W. J. Collins[4], G. Faluvegi[5], and K. Tsigaridis[5]

[1]{Department of Physics, Imperial College London, London, UK}

[2]{NCAR Earth System Laboratory, National Center for Atmospheric Research, Boulder, CO, USA}

[3]{Nicholas School of the Environment, Duke University, Durham, NC, USA}

[4]{Department of Meteorology, University of Reading, Reading, UK}

[5]{Center for Climate Systems Research, Columbia University, and NASA Goddard Institute for Space Studies, New York, NY, USA}

Correspondence to: M. Kasoar (m.kasoar12@imperial.ac.uk)

Abstract

We use the HadGEM3-GA4, CESM1, and GISS ModelE2 climate models to investigate the global and regional aerosol burden, radiative flux, and surface temperature responses to removing anthropogenic sulfur dioxide (SO_2) emissions from China. We find that the models differ by up to a factor of six in the simulated change in aerosol optical depth (AOD) and shortwave radiative flux over China that results from reduced sulfate aerosol, leading to a large range of magnitudes in the regional and global temperature responses. Two of the three models simulate a near-ubiquitous hemispheric warming due to the regional SO_2 removal, with similarities in the local and remote pattern of response, but overall with a substantially different magnitude. The third model simulates almost no significant temperature response. We attribute the discrepancies in the response to a combination of substantial differences in the chemical conversion of SO_2 to sulfate, translation of sulfate mass into AOD, cloud radiative interactions, and differences in the radiative forcing efficiency of sulfate aerosol in the models. The model

with the strongest response (HadGEM3-GA4) compares best with observations of AOD regionally, however the other two models compare similarly (albeit poorly) and still disagree substantially in their simulated climate response, indicating that total AOD observations are far from sufficient to determine which model response is more plausible. Our results highlight that there remains a large uncertainty in the representation of both aerosol chemistry as well as direct and indirect aerosol radiative effects in current climate models, and reinforces that caution must be applied when interpreting the results of ~~single-~~modelling studies of aerosol influences on climate. Model studies that implicate aerosols in climate responses should ideally explore a range of radiative forcing strengths representative of this uncertainty, in addition to thoroughly evaluating the models used against observations.

1 Introduction

Short-lived atmospheric pollutants such as aerosols have very inhomogeneous spatial distributions. This means that, unlike long-lived greenhouse gases such as CO_2, the radiative forcing due to aerosols is highly variable, and the resulting climate response may be strongly influenced by the region of emission and the prevailing circulation patterns. There is increasing interest in trying to understand how aerosol forcing from different regions affects the climate, both locally and remotely. For example, Shindell and Faluvegi (2009) and Shindell et al. (2012) looked systematically at the response of temperature and precipitation to single-species forcings imposed in different latitude bands, and showed that the influence of remote forcings on certain regions can often outweigh and even have opposite sign to the influence of local forcings. Teng et al. (2012) investigated the global temperature response to drastically increasing carbonaceous aerosols only over Asia, finding a strong remote effect on US summertime temperatures.

One of the most important anthropogenically-sourced aerosol species is sulfate (SO_4) (e.g. Myhre et al., 2013b). Sulfate-containing aerosols are formed following chemical conversion of gaseous sulfur dioxide (SO_2) emissions from fossil-fuel combustion, as well as natural sources such as volcanic SO_2 and ocean dimethyl sulfide (DMS) emissions (e.g. Andres and Kasgnoc, 1998; Andreae and Crutzen, 1997). Sulfate particles strongly scatter incoming shortwave (SW) radiation, which helps to increase the planetary albedo and cool the surface. They also act as cloud condensation nuclei, leading to additional cloud droplets forming in supersaturated conditions, which increases cloud albedo and again cools the Earth system

(Boucher et al., 2013). Historically, cooling from sulfate aerosol, predominantly in the more industrialised northern hemisphere, has been implicated by a range of modelling studies in disrupting climate since the mid-20th century. For instance, Booth et al. (2012), Hwang et al. (2013), and Wilcox et al. (2013) discussed the importance of historical aerosol cooling in modulating large-scale temperature and precipitation patterns, while other studies such as Bollasina et al. (2011), Dong et al. (2014), and Polson et al. (2014) have looked at the impact of historical aerosols on regional climate features such as the monsoon systems or Sahelian rainfall.

The few studies that have investigated specific regional aerosol forcings (e.g. Shindell and Faluvegi, ~~(~~2009~~)~~; Shindell et al., ~~(~~2012~~)~~; Teng et al., ~~(~~2012~~)~~) typically used a single climate model at a time to investigate the climate response to idealised, historical, or projected forcings. However models vary considerably in their representation of aerosols and their radiative properties, resulting in a large uncertainty in aerosol radiative forcing (e.g. Myhre et al., 2013b; Shindell et al., 2013a). When investigating the climate response to regional aerosol emissions, such uncertainties are likely to be confounded even further by the variability between models in regional climate and circulation patterns, and variation in the global and regional climate sensitivity (the amount of simulated warming per unit radiative forcing). To best interpret the findings of single-model experiments with regional aerosol forcings, it is therefore critical to understand the range of uncertainty in the climate response that may arise as a result of structural and parametric differences between climate models.

We investigate here the range of variability that can arise in the translation of a regional emission perturbation to a climate (temperature) response, between three different state-of-the-art global climate models. We select as a case study the removal of SO_2 anthropogenic emissions from the region of China. Since China is currently the largest anthropogenic source region of sulfur dioxide (Smith et al., 2011) and hence anthropogenic aerosol, this regional perturbation represents a substantial modification to global aerosol levels, with the additional characteristic of being localised over a particular part of the world. This aspect of our experiment is distinct from many previous model intercomparison studies, which have typically compared the climate response in models forced by global historical trends in aerosols (for example, Shindell et al., 2015; Wilcox et al., 2013), or which have only considered the impact of regional emissions on long-range pollution transport and on radiative forcing (for example the HTAP and AeroCom experiments (HTAP, 2010; Yu et al., 2013; Kinne et al., 2006; Schulz

et al., 2006; Textor et al., 2006; Myhre et al., 2013)), but have not investigated the range of model climate responses to a regionally localised emission perturbation. The potential importance of remote climate effects due to the strong zonal asymmetry created by such regional emissions has therefore not yet been explored in multi-model studies. Single-model studies such as Teng et al. (2012) suggest though that regionally localised forcings can produce significant climate teleconnections in at least the longitudinal direction.

In the following sections we first describe the three models employed, and our experimental setup (Sect. 2). We then present the results of the radiative flux and surface temperature responses to the removal of Chinese SO_2 (Sect. 3), and analyse the possible reasons for differences between the model responses (Sect. 4). Finally, in Sect. 5 we present our conclusions.

2 Model descriptions and experimental set-up

The three models we employ are the Hadley Centre Global Environment Model 3 – Global Atmosphere 4.0 (HadGEM3-GA4), the Community Earth System Model 1 (CESM1), and the Goddard Institute for Space Studies ModelE2 (GISS-E2). To allow the climate system to freely respond, the models are all used in a fully coupled atmosphere-ocean configuration. These three models all feature explicit aerosol modelling, and include both direct and indirect radiative effects of aerosols. However the models all vary in the details of the parameterisations used, the dynamical cores, radiation and cloud schemes, model grids and horizontal and vertical resolutions, land surface and ocean components, etc. This lack of common structural features makes these three models well suited to contrast against one another and probe the range of potential model uncertainty, as we do here. The models are briefly described below, and the key references and features are summarised in Table 1.

2.1 Model descriptions

2.1.1 HadGEM3-GA4

For HadGEM3, we use the Global Atmosphere 4.0 version of the model (Walters et al., 2014) in a standard climate configuration with a horizontal resolution of 1.875° longitude x 1.25° latitude in the atmosphere, with 85 vertical levels and the model top at 85km~~, dynamically~~

~~resolving the stratosphere~~. The atmosphere is coupled to the JULES land surface model (Walters et al., 2014)~~, which includes 4 soil layers and 5 plant functional types. Although in principle this can be run in a fully interactive 'Earth-System' mode with dynamic vegetation and a carbon cycle,~~. ~~h~~Here we prescribe fixed vegetation and also ~~prescribe~~ globally-uniform observed mass-mixing ratios for CO_2, CH_4, and other long-lived greenhouse gases, taking their year-2000 values from the CMIP5 historical dataset (Meinshausen et al., 2011). A zonally-uniform present-day ozone climatology is also prescribed in the radiation scheme, derived from the SPARC dataset (Cionni et al., 2011).~~ More detailed description and evaluation of the atmosphere and land surface schemes can be found in Walters et al. (2014).~~ The atmospheric model is~~ also~~ coupled to the NEMO dynamical ocean model (Madec, 2008) and CICE sea-ice model (Hunke and Lipscombe, 2008), which are run with a 1° horizontal resolution, and 75 vertical depth levels for the ocean.

HadGEM3-GA4~~Critical to our study is the representation of aerosols; we~~ can be run with a choice of two aerosol schemes of differing complexity ~~—~~ CLASSIC (Bellouin et al., 2011), and GLOMAP (Mann et al., 2010). Here we use the simpler CLASSIC ~~aerosol~~ scheme~~, which is described and evaluated in Bellouin et al., (2011),~~ which is less computationally expensive, and is also the aerosol scheme that was used for CMIP5 simulations with the predecessor of this model (HadGEM2). CLASSIC is a mass-based scheme, meaning that only aerosol mass (and not particle number) is tracked~~ prognostically~~, and therefore all aerosol species are assumed to be externally mixed. ~~.~~ The~~CLASSIC~~ scheme ~~is a mass-based scheme which~~ includes an interactive representation of sulfate in three modes (Aitken, accumulation, and in-cloud), fossil-fuel black carbon, fossil-fuel organic carbon, and biomass-burning aerosol in three modes (fresh, aged, and in-cloud), dust in six size bins, and sea-salt in two modes (jet and film), as well as an offline biogenic aerosol climatology. The scheme can also include a representation of nitrate aerosol, but this option was not used here.~~ All species are considered to be externally mixed.~~

The sulfate component of the scheme (Jones et al., 2001) includes tracers for two gas-phase precursors: SO_2 from anthropogenic and natural sources, and DMS from natural sources. These are emitted into the atmosphere and can undergo advection, wet and dry deposition, or oxidation using prescribed 4D oxidant fields (Derwent et al., 2003). In CLASSIC, oxidation of SO_2 to SO_4 aerosol can proceed through three possible reaction pathways: in the gas phase by reaction with OH, or in the aqueous phase by reaction with either H_2O_2 or O_3.

The radiative transfer scheme of Edwards and Slingo (1996) is used with six spectral bands in the shortwave, and~~The remaining aerosol species are emitted directly in the particulate phase, and a~~ all aerosol species~~ can then undergo advection, wet and dry deposition, and~~ interact~~ion~~ with radiation. The hygroscopic aerosols (sulfate, organic carbon, biomass-burning aerosol, sea-salt) can also interact with clouds via their role as cloud condensation nuclei. Cloud droplet number concentration and effective radius are determined from the mass concentration of these aerosols, which affects the simulated cloud lifetime (2nd indirect effect) and cloud brightness (1st indirect effect) as described in Bellouin et al. (2011) and Jones et al. (2001).

2.1.2 CESM1

CESM1 is run in its standard CAM5-Chem configuration (Tilmes et al., 2015) with a horizontal resolution of 2.5° longitude x 1.875° latitude, and 30 vertical levels, with the model top at approximately 40 km. The atmosphere is coupled to the Community Land Atmosphere version 4 land surface model (Lawrence et al., 2011). In the present configuration, the vegetation distribution is fixed at its 2005 distribution and the CO_2 concentration is specified. The atmosphere model is coupled to the POP2 ocean and CICE4 sea-ice models, with an equivalent resolution of 1°.

In the present CAM5-Chem configuration (Tilmes et al., 2015)~~,~~ we use an online representation of tropospheric and stratospheric chemistry so that no chemical constituents are specified, other than specifying the long-lived greenhouse gases' concentrations in the surface layer. CAM5-Chem uses the MAM3 modal aerosol scheme (Liu et al., 2012), which is the same as used for the CESM1 submission to CMIP5. Both aerosol mass and particle number are prognostic, and the scheme ~~simulates~~ sulfate, black carbon, primary organic matter, secondary organic aerosol, dust, and sea salt aerosol species as an internal mixture in Aitken, accumulation, and coarse modes.

The model includes emissions of natural and anthropogenic SO_2 and natural DMS as sulfate precursors, and~~T~~ the gas-phase chemistry is coupled to the~~ modal aerosol scheme~~ MAM3 aerosol scheme~~(Liu et al., 2012),~~ so that the rate of formation of sulfate aerosols is dependent on the chemical state of the atmosphere. SO_2 ~~from anthropogenic and natural sources~~ can be converted to SO_4 through three oxidation pathways: by OH in the gas phase, or by either H_2O_2

or O_3 in the aqueous phase. In addition, the surface area of the prognostic tropospheric aerosols is used to compute heterogeneous reaction rates that affect gas-phase chemistry.

Shortwave radiative transfer is based on the RRTM_SW scheme (Clough et al., 2005) with 14 spectral bands, and ~~A~~aerosols interact with climate through both absorption and scattering of radiation. ~~and~~ Aerosol-cloud interactions allow for the effect of aerosols on both cloud droplet number and mass concentrations (Tilmes et al, 2015)~~cloud-aerosol interactions~~.

2.1.3 GISS-E2

GISS-E2 is run in the configuration used for CMIP5 with a horizontal resolution of 2.5° longitude x 2° latitude, and 40 vertical levels, with the model top at 0.1 hPa (80 km). The atmospheric model is coupled to the dynamic Russell ocean model with horizontal resolution of 1° latitude x 1.25° longitude, and 32 vertical levels as described in Schmidt et al. (2014) and Russell et al. (1995).

Well-mixed greenhouse gases are prescribed as described in Miller et al. (2014), but methane is only prescribed at the surface and is otherwise interactive with the chemistry. The ozone distribution is prognostic throughout the simulated atmosphere, and the chemical mechanism is described in Shindell et al. (2013b). In general, other atmospheric gas and aerosol constituents are also simulated online and interact with each other (via oxidants in both the gas and aqueous phases, heterogeneous chemistry, aerosol-influenced gas photolysis, and secondary ~~aerosol~~ coating of dust) and with climate (via radiative effects of ozone and methane, water vapour change due to chemistry, and aerosol direct and indirect effects) in a manner consistent with the physics of the rest of the GCM as described in Sect. 2~~3b~~ of Schmidt et al. (2014~~3~~).

GISS-E2 has a choice of three aerosol schemes of varying complexity – OMA (Koch et al., 2011; 2006), MATRIX (Bauer et al, 2008), and TOMAS (Lee and Adams, 2012). ~~The aerosol scheme (Koch et al. 2011; 2006)~~Following the GISS-E2 CMIP5 configuration, we use here simpler mass-based OMA scheme, which includes sulfate, nitrate, elemental and organic carbon,~~n (Koch et al. 2011; 2006)~~ along with secondary organic aerosols and natural sea-salt and mineral dust. Aerosols are parameterised as an external mixture of dry and dissolved aerosol, with particle size parameterised as a function of relative humidity (Schmidt et al., 2006). ~~T~~For the sulfur scheme includes natural emissions of DMS, and natural and anthropogenic emissions of SO_2. ~~particular,~~ SO_2 from ~~anthropogenic and natural~~ these sources

can be oxidised to SO_4 aerosol through two reaction pathways: by OH in the gas phase, or by H_2O_2 in the aqueous phase.

Aerosol direct effects are calculated following the Hansen et al. (1983) radiation model, with six spectral bands in the shortwave. ~~Other aerosols include nitrate, elemental and organic carbon (Koch et al. 2011; 2006) along with secondary organic aerosols and natural sea-salt and mineral dust.~~ Aerosol indirect effects are calculated as described in Menon et al. (2010), such that cloud droplet number concentration and autoconversion rate depend on the local concentration of aerosol.

2.2 Experimental setup

For this study we investigate the surface temperature response to an idealised regional emission perturbation, on a centennial timescale. Each model has a control simulation, initialised from a present-day state, which is forced with the same anthropogenic emissions of aerosols and their precursors following the year-2000 ACCMIP emission inventory (Lamarque et al., 2010). The control simulations are run for 200 years with continuous year-2000 conditions. For each model, we then also run a 200-year perturbation simulation from the same initial state, in which SO_2 emissions from energy production, industry, transport, domestic use, and waste, are set to zero over the region of China, defined here to be the rectangular domain 80°-120°E, 20°-50°N. These emission sectors contribute 98.7% of the anthropogenic SO_2 emitted from this region, so this corresponds to a near complete removal of SO_2 emissions from this highly polluting area of the globe. Quantitatively, this perturbation reduces global anthropogenic SO_2 emissions from around 104 Tg yr^{-1} to 86 Tg yr^{-1}, a reduction of around 17 Tg yr^{-1}, or 16.5%.

Additionally, shorter atmosphere-only simulations were performed with HadGEM3-GA4 (identical in setup except that sea-surface temperatures (SSTs) and sea-ice cover are prescribed to year-2000 values) in order to diagnose the effective radiative forcing, as well as the SO_2 oxidation rates and SO_4 wet deposition rates for this model, referred to in Section 3, Section 4.1, and Section 4.1.1. In CESM1, the SO_2 burden, surface SO_4 concentration, clear-sky radiative flux, and cloud cover referred to in Sections 4.1.1, 4.2, and 4.3, were all diagnosed from a 30-year extension of the control and perturbation coupled simulations, rather than from the original 200 years.

3 Radiative forcing and climate response

We investigate the change in the mean state of the models by taking averages over the last 150 years of the 200-year-long simulations (the first 50 years are~~were~~ discarded to allow the response to the perturbation to establish itself~~as spin-up~~), and taking the difference between the perturbation simulation and the control simulation. As well as plotting maps of 2D variables, we also calculate area-weighted means of temperature, short-wave radiative flux, and aerosol optical depth, both globally and for an east China region (E. China) defined as 100°-120°E, 20°-40°N. This region is found to contain the most intense changes in sulfate aerosol in all three models, and is used from here on to quantify the magnitude of local changes over China. The global- and regionally-averaged quantities, with associated uncertainties where available, are tabulated in Table 2~~1~~, along with the total sulfate burdens over the globe and E. China, and the ratios of AOD to sulfate burden and SW flux to AOD changes.

The anticipated immediate consequence of removing SO_2 emissions from China is that there will be a reduction in the amount of sulfate aerosol formed, leading to a positive shortwave (SW) radiative forcing. Figure 1 shows the changes in net downward top-of-atmosphere (TOA) SW radiative flux in each of the three models. For HadGEM3-GA4 and GISS-E2, the plot is stippled in locations where the change exceeds two standard deviations, estimated for HadGEM3-GA4 from the grid-point standard deviations from six ~~year-2000~~ 150-year-long year-2000 control ~~runs~~ simulations with perturbed atmospheric initial conditions, and for GISS-E2 from 12 non-overlapping 150-year sections of a 1900-year-long pre-industrial control simulation that had reached radiative equilibrium. Such uncertainty analysis has not been performed for CESM1 due to lack of the necessary unforced simulation output for the version of the model used here. 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.

Figure 1 reveals that there is a very substantial variation between the models in the intensity of the local radiative flux change over China. GISS-E2 shows a fairly weak increase in net downward SW flux over E. China, with a local increase (from Table ~~1~~2) of 0.91 W m^{-2} and an insignificant global mean change (-0.034 W m^{-2}), whereas HadGEM3-GA4 shows a very pronounced change of 5.3 W m^{-2} locally over E. China, and a global mean value of 0.28 W m^{-2}. CESM1 lies in the middle, with a moderate local SW flux change of 4.2 W m^{-2}, and 0.19 W m^{-2} in the global mean. Between GISS-E2 and HadGEM3-GA4, there is a 6-fold increase in the intensity of the local SW radiative flux change over E. China.

Because these are fully coupled simulations, the change in the TOA SW flux does not provide a measure of the shortwave radiative forcing, since the underlying climate has been allowed to adjust, potentially allowing feedbacks on clouds, and snow and ice cover. A complementary pair of atmosphere-only simulations~~, where sea-surface temperatures (SSTs) and sea-ice cover were prescribed to year-2000 values,~~ were ~~run~~ performed with HadGEM3-GA4 to diagnose the effective radiative forcing (ERF) – the change in TOA radiative flux when feedbacks due to the slow response of the ocean are prevented (Andrews et al., 2010). The global SW ERF due to removing SO2 from China in these fixed-SST simulations is 0.18 W m^{-2}, 35% smaller than the 0.28 W m^{-2} change in the fully coupled case. However, locally over the E. China region, the fixed-SST ~~change~~ SW ERF was found to be 4.2 W m^{-2}, which is only 21% lower than the 5.3 W m^{-2} value in the fully coupled experiment. Moreover, the spatial map of the SW flux anomaly over China is very similar between the two experiments (Supplementary fig. S1). At least in HadGEM3-GA4, over E. China the change in sulfate therefore appears to be the dominant driver of the change in TOA SW flux, and the local change in SW flux over this region is reasonably representative of the local radiative effect of the sulfate perturbation even in the fully-coupled simulations with this model. The same is less true of the global-mean values because of positive feedback from ice melt in the Arctic, and also some small but widespread changes in cloud cover, which globally add up to a sizeable effect in the coupled simulations (not shown).

Based on the fully coupled simulations, the substantial differences in the intensity of SW flux changes over China ultimately translate to very pronounced differences in the strength of the resulting climate response. Figure 2 shows the change in surface air temperatures between the perturbation and control ~~runs~~ simulations for each of the three models, clearly demonstrating that temperature effects extend far beyond the more localised radiative effects. Again stippling indicates that the response exceeds the 2σ level in HadGEM3-GA4 and GISS-E2. The difference between GISS-E2 and HadGEM3-GA4 is particularly striking. Apart from a small warming in parts of eastern China and north-east Europe by around 0.1-0.3 K, there is virtually no coherent temperature response across the rest of the globe in GISS-E2. The global mean temperature change (Table 2~~1~~) is -0.028 K and is not significant. In contrast HadGEM3-GA4 displays significant warming across almost all of the northern hemisphere, with much larger increases in temperature between 0.4-1 K in many regions, not only in China but also in much of the US, northern Eurasia, and the Arctic. The global mean temperature response is +0.12~~1~~ K. CESM1 sits again in the middle, with clear warming responses between 0.2-0.5 K over

much of eastern Europe, Asia, and the western Pacific. Overall the warming response is still less strong and less widespread than in HadGEM3-GA4, with a global mean warming of +0.054 K.

The spatial pattern of warming over Europe and Asia in CESM1 bears some qualitative similarity though to the pattern over the same region in HadGEM3-GA4, suggesting that there may be a similar mode of global response to heating over eastern China in these models, at least across the Eurasian continent. The dynamical mechanisms through which local aerosol emissions are translated to remote response are beyond the scope of this manuscript though. Whether GISS-E2 would have displayed the same pattern had the radiative forcing over China been stronger is impossible to tell from these results; given the small magnitude of the SW flux change it seems that most of the spatial pattern in the temperature response in GISS-E2 can be attributed to internal variability – the largest changes in temperature seen in this model are in fact a region of cooling over the north-west Atlantic, which is mostly not significant and appears instead to be the result of particularly large internal variability in this region.

4 Exploring drivers of diversity

We investigate the differences between these models that lead to such a large variation in the predicted temperature response. We explore below a number of possible sources of discrepancy.

4.1 Differences in simulated aerosol amounts and aerosol optical depths

We address first the possibility that differences in the aerosol schemes themselves, lead directly to very different aerosol loadings between the models, despite the identical change in SO_2 emissions applied. Figure 3 shows the change in column-integrated SO_4 in each model as a result of removing ~~Chinese~~ SO_2 emissions from China. The models vary in both the distribution and magnitude of SO_4 reductions. In particular, HadGEM3-GA4 has the reduction in SO_4 burden fairly concentrated over China. CESM1 and GISS-E2 simulate ~~more diffuse~~ changes in SO_4 which extend further downwind from the source region, giving a larger spatial footprint, although CESM1 still has large reductions over China as well.

For GISS-E2 and HadGEM3-GA4, more detailed chemistry diagnostics were available from a 5-year period of a HadGEM3-GA4 atmosphere-only control simulation, and a 5-year period of the GISS-E2 coupled control simulation. For these two models, ~~T~~the~~is~~ difference in spatial extent of the SO_4 field from Chinese SO_2 emissions seems to be due to ~~particularly~~ faster conversion of SO_2 to SO_4 in HadGEM3-GA4, resulting in much more concentrated changes in SO_4 close to the source. ~~For GISS-E2 and HadGEM3-GA4 where more detailed diagnostics were available, we find that t~~The SO_2 lifetime is around 1.8 times shorter in HadGEM3-GA4, associated with around 45% higher wet oxidation rates in this model. This difference is due in part to the inclusion of an additional wet oxidation pathway in HadGEM3-GA4: whereas GISS-E2 only includes wet oxidation of SO_2 by H_2O_2 (around 730 kg(S) s^{-1} globally integrated), HadGEM3-GA4 includes wet oxidation by both H_2O_2 and O_3, each of which contribute similarly in this model (around 540 kg(S) s^{-1} and 520 kg(S) s^{-1} respectively).

Globally integrated, HadGEM3-GA4 and GISS-E2 simulate fairly similar reductions in SO_4 burden, at -0.070 Tg and -0.077 Tg respectively (Table 2~~1~~). This, combined with the more spread-out SO_4 field in GISS-E2, means that locally over eastern China HadGEM3-GA4 has a much more intense reduction in SO_4 burden, with 50% of the global reduction occurring over E. China in HadGEM3-GA4 (-0.035 Tg), compared with only 21% (-0.016 Tg) in GISS-E2.

~~–~~CESM1 includes the same oxidation pathways as HadGEM3-GA4, and in fact has a slightly shorter SO_2 lifetime still, and so the differences between these two models have different origins. CESM1 in fact~~, by contrast,~~ simulates almost double the global change in SO_4 burden as the other two models, with -0.136 Tg. This means that although the SO_4 reduction spreads further from the source in CESM1 than in HadGEM3-GA4, CESM1 still has a similar reduction to HadGEM3-GA4 locally over E. China as well (-0.039 Tg), which is also evident in Fig. 3.

Given that HadGEM3-GA4 and GISS-E2 simulate a similar global reduction in SO_4, it is surprising that there is such a difference in the magnitude of their climate responses. Also, given that CESM1 simulates a much larger global reduction in SO_4 than the other two models, it is similarly surprising that this model does not have the largest response. A partial explanation may be found by inspecting the change in total aerosol optical depth (AOD), which is a more direct measure of the radiative properties of the aerosol column. Unfortunately, the AOD diagnosed by the models is not completely equivalent: HadGEM3-GA4 diagnosed clear-sky AOD, which is done in this model by calculating the relative humidity in the cloud-free portion of each grid-box, and using this adjusted humidity to calculate the size of the aerosol

droplets in the optical depth calculation (Bellouin et al., 2007). However CESM1 uses the unadjusted grid-box relative humidity to calculate the droplet sizes in its optical depth calculation, thereby providing an all-sky AOD calculation (Neale et al., 2012). GISS-E2 diagnosed both all-sky and clear-sky AOD, and unless otherwise stated we compare here its clear-sky AOD, as it is more directly comparable with satellite retrievals of AOD (Kahn et al., 2010; Levy et al., 2013). Figure 4 shows these changes in AOD at the 550nm wavelength for the three models.

As with the radiative flux change, there is a large range in the magnitude of local AOD reduction, with E. China AOD reductions ranging from 0.047 in GISS-E2 to 0.287 in HadGEM3-GA4, i.e. about six~~6~~ times higher (Table 2~~1~~). This is comparable to the approximately 6-fold range of SW flux change found over this region. Globally averaged, HadGEM3-GA4 also has a much larger AOD reduction than GISS-E2; 0.0042 compared with an almost negligible 0.0003 in GISS-E2, despite these two models having a similar change in global SO_4 burden. The much lower globally-averaged value in GISS is partly due to a very small but quite zonally-uniform compensating increase in nitrate aerosol~~,~~ (absent in HadGEM3-GA4), which occurs across the northern hemisphere (not shown). However, the global change in sulfate-only optical depth in GISS-E2 is still only half that in HadGEM3-GA4 (not shown), and locally around eastern China we find the increase in nitrate optical depth in GISS-E2 is at least an order of magnitude smaller than the decrease in sulfate optical depth, and so nitrate compensation does not substantially contribute to the discrepancy in local AOD changes. We therefore still find that HadGEM3-GA4 simulates a considerably larger change in sulfate optical depth per unit change in SO_4 burden at both global and local scales. Having the largest change in AOD per unit change in aerosol burden (Table 2~~1~~) appears to be key to this model simulating the largest climate response.

Comparing the clear-sky and all-sky AOD for GISS-E2 (for which we have both diagnostics), we find that the simulated reduction in E. China all-sky AOD (-0.183) is much larger than the reduction in clear-sky AOD (-0.047). We cannot be sure that the same would apply to CESM1, but it suggests that we might expect the all-sky values we have for CESM1 to be larger than the equivalent clear-sky values. Given this, it is all the more surprising to find reductions of all-sky AOD in CESM1 for the E. China region of -0.076 and for the global mean of -0.0013 (Table ~~1~~2), which lie in between the clear-sky values of GISS-E2 and HadGEM3-GA4 even though CESM1 had the largest change in SO_4 burden both locally and globally.

The AOD changes per unit burden change are summarised in Table 2, and it is clear that there is a large diversity between the models. The possible contributors to diversity in the AOD per unit burden are extensive, and a full analysis of them is beyond the scope of this paper. Host model effects, such as different cloud climatologies and radiative transfer schemes, are one likely contributor. Stier et al. (2013) suggests that one third of total diversity originates there. Relative humidity, which drives water uptake (hygroscopic growth), is also diverse among models. For example, Pan et al. (2015) find that over India, boundary-layer RH is the main source of diversity. At the more basic level, assumed composition and hygroscopic growth curves also often differ between models – in this case, the aerosol scheme used for HadGEM3-GA4 assumes that all sulfate is in the form of ammonium sulfate, whereas CESM1 and GISS-E2 both assume a mixture of ammonium sulfate and sulfuric acid, and additionally all three models use different sources for their hygroscopic growth parameterisations (Bellouin et al~~, due to different simplifications and data sources~~., 2011; Liu et al., 2012; Koch et al., 2011; and references therein).

The changes in SW radiative flux and the final climate response seem to correlate with the changes in AOD much better than with the changes in SO_4 burden for HadGEM3-GA4 and GISS-E2, where over China there is a 6-fold difference both in AOD and in SW flux change between these two models. For CESM1, the all-sky AOD change~~s~~ over E. China ~~are~~is about 1.6 times larger than the clear-sky change~~s~~ in GISS-E2 (Table 2~~1~~). If we used instead all-sky AOD from GISS-E2 (not shown in Table 2~~1~~), we find that the AOD change over E. China is more than 2 times smaller in CESM1 than in GISS-E2. However, the change in TOA SW over the same region is about 4.7 times larger in CESM1, and so it seems that unlike the discrepancies between HadGEM3-GA4 and GISS-E2, differences in the AOD response cannot explain the difference in the magnitudes of radiative flux change between CESM1 and GISS-E2 (see Sect. 4.2~~3~~).

4.1.1 Validation of aerosol fields

To get an indication of whether the model-simulated AODs are realistic in the region of interest, we compare the mean AOD from each model's control run with station observations in Asia from the AERONET radiometer network (Holben et al., 2001). Because of the limited number of stations in the region with long data records, we use the observed AOD at 500 nm from all

AERONET stations able to provide an annual mean estimate for at least one year, averaged over all years for which an annual mean was available, (generally ranging between 1998 and 2014 in different stations), and compare this with the annual mean AODs at 550 nm from the three models, masked to the locations of the AERONET stations (Supplementary fig. S2). Focusing on stations in E. China (eight in total), we find that HadGEM3-GA4 compares best with AERONET in this region with a mean station bias of -22%, whilst both GISS-E2 and CESM1 appear to be biased lower in this part of the world, with mean biases of -56% and -60% respectively.

We also calculate the area-weighted mean AOD as observed by the MODIS and MISR satellite instruments. The MODIS (Moderate Resolution Imaging Spectroradiometer) instrument is flown on both the Terra and Aqua satellites, whilst MISR (Multi-angle Imaging SpectroRadiometer) is flown on Terra. For MODIS we use the collection 6 combined Deep Blue + Dark Target monthly AOD product at 550 nm (Levy et al., 2013) (available from https://ladsweb.nascom.nasa.gov/), averaged from both Terra and Aqua satellites, and take a 10-year average from 2003-2012 (2003 being the earliest year that data from both satellites is available). For MISR we use the best estimate monthly AOD product (Kahn et al., 2010) version 31 (available from https://eosweb.larc.nasa.gov/) at 550 nm over a 15-year averaging period, from 2000-2014 (2000 being the earliest year MISR data is available). For MODIS the area-weighted E. China mean AOD is 0.51, whilst for MISR it is 0.31, so regionally there is a considerable uncertainty in these observations. HadGEM3-GA4 overestimates the AOD compared with both instruments (though only slightly so when compared to MODIS), with a regional average AOD of 0.58, whilst GISS-E2 and CESM1 underestimate it with regionally-averaged AODs of 0.23 for both models. Globally the two instruments are in better agreement, with MODIS giving a global average AOD of 0.17 and MISR giving 0.15. Again HadGEM3-GA4 overestimates global AOD compared with both instruments (0.22) whilst GISS-E2 and CESM1 both underestimate it (0.13 and 0.12). Given that CESM1 diagnosed all-sky AOD, whereas satellite retrievals are only possible for clear-sky conditions, the underestimate for this model is likely greater than these numbers suggest.

There is considerable variation in the observations as well as the models. Globally-averaged, GISS-E2 seems to compare best against MODIS and MISR, though tentatively HadGEM3-GA4 seems to have the more accurate AOD over China, comparing best regionally with both AERONET and MODIS, though poorer against MISR. This suggests that the more concentred

sulfate aerosol burden and larger AOD reduction simulated by HadGEM3-GA4 over this region may be more realistic. WHowever we note though that since these observations only measure total AOD and cannot differentiate by species, the comparison cannot show for certain that the higher sulfate optical depth specifically is more realistic in HadGEM3-GA4. The AOD reduction over E. China due to removing Chinese SO_2 represents 50% of the climatological total AOD in HadGEM3-GA4 over the region, compared with 34% in CESM1 and only 20% in GISS-E2. Even if the total AOD in HadGEM3-GA4 is more realistic, there is still considerable variation between the models as to what fraction of that total AOD is due to Chinese SO_2 emissions. This is illustrated further for the two extreme cases, HadGEM3-GA4 and GISS-E2, in Supplementary Fig. S3, which shows that the fraction of climatological AOD made up by sulfate is consistently higher across the east Asian region in HadGEM3-GA4 than in GISS-E2. However, the total non-sulfate AOD is fairly similar across the region in these two models (Supplementary Fig. S4), indicating that the stark difference in the fractional contribution of sulfate comes primarily from HadGEM3-GA4 simulating much greater sulfate AOD alone. Given that regionally GISS-E2 appeared to underestimate total AOD, this would then suggest that either the higher sulfate AOD in HadGEM3-GA4 is more realistic, or else both models underestimate the non-sulfate AOD.

For HadGEM3-GA4 and GISS-E2, for which sulfate mixing ratio diagnostics were available for individual model levels, wTo try and better constrain whether the sulfate content (rather than total aerosol) is correct, we therefore also compared against the surface sulfate observations conducted in China reported by Zhang et al. (2012) for 2006-2007 (Supplementary fig. S53). However, all threeboth models performed extremely poorly, with HadGEM3-GA4 having a mean bias of -71% (-66% if urban stations are excluded), CESM1 a mean bias of -71% (unchanged if urban stations are excluded), and GISS-E2 having a mean bias of -87% (-86% when urban stations are excluded). Although HadGEM3-GA4 and CESM1 isare slightly closer to the observed values, the large underestimation despite the relatively good column AOD in HadGEM3-GA4 comparison suggests that at least thise model has difficulty representing the vertical profile of sulfate aerosol, and so this comparison with surface measurements may not be that particularly useful in constraining the sulfate optical depth or column-integrated burdens. Large underestimations of surface sulfate concentration over East Asia have been reported previously for two other models, MIROC and NICAM, by Goto et al. (2015), suggesting that this is a problem common to many current generation models.

It seems plausible that any differences in the processing of sulfate aerosol would apply to all polluted regions, and not just over China. Indeed, the spatial pattern of the climatological sulfate burden over other major emission regions ~~like~~ such as the United States shows a similar characteristic to that over China, with HadGEM3-GA4 and CESM1 having ~~a~~ higher burdens close to the emission source regions, whilst GISS-E2 has a more diffuse sulfate distribution (Supplementary fig. S6~~4~~). With this in mind we also validated the ~~se two~~ models against surface sulfate observations from the Interagency Monitoring of Protected Visual Environments (IMPROVE) network in the United States (Malm et al., 1994), a dataset with a far more extensive record than the Zhang et al. (2012) dataset for China. Taking 61 IMPROVE stations which have data for at least six~~6~~ years between 1995 and 2005, we find that over the United States ~~both~~ all three models are in fact biased ~~slightly~~ high, with GISS-E2 performing relatively better with a mean bias of +10.1%, but HadGEM3-GA4 somewhat worse with +44.5%, and CESM1 worse still with +86%. However, in the case of HadGEM3-GA4 we find that the larger mean bias ~~in HadGEM3-GA4~~ comes mainly from an incorrect spatial distribution (Supplementary fig. S7~~5~~), with a high bias on the West Coast but a pronounced low bias in surface SO_4 on the East Coast. Consequently, this comparison would suggest that HadGEM3-GA4 in fact has too little sulfate around the principal US emission regions on the East Coast, even though over that area HadGEM3-GA4 actually has a larger column-integrated sulfate burden (Supplementary fig. S6~~4~~) and a larger AOD (not shown) than GISS-E2, as was the case for China. This suggests that HadGEM3-GA4 again fails to capture the vertical profile of sulfate, underestimating surface concentrations over this region despite having a high column-integrated burden.

Validation with surface observations therefore seems insufficient to constrain which model performs better with regard to the more climate-relevant column-integrated quantities of sulfate burden and AOD. Returning to Asia, we therefore also tried evaluating the models~~validating HadGEM3-GA4 and GISS-E2 usingagainst~~ column sulphur dioxide observations~~sulfate wet deposition observations, which should be less sensitive to the precise vertical profile of sulfate in the models.~~ . We use the gridded, monthly mean Level 3 observations from the Ozone Monitoring Instrument (OMI) (Krotkov et al, 2008) (available from http://disc.sci.gsfc.nasa.gov/Aura) which is flown on the Aura satellite, averaged over eight years from 2005 - 2012. Over the E. China region the mean OMI SO_2 is 0.153 Dobson Units (DU), and all three models appear to overestimate this substantially, with very similar regional mean SO_2 columns of 0.282 DU for HadGEM3-GA4, 0.272 DU for GISS-E2, and 0.259 DU

for CESM1. Spatially, all three models have more diffuse SO_2 fields than the OMI observations, in which by contrast the SO_2 burden seems much more localised around sources (Supplementary Fig. S8). This may be partly due to the coarse resolution of the models compared with the 0.25° satellite product, but also suggests that the lifetimes for SO_2 may be too long in all three models, or transport processes too efficient. The surprisingly similar column SO_2 burdens in all three models suggests that, at least on regional scales, column SO_2 may not constrain SO_4 burden that well.

An alternative observational measure which to an extent reflects a column-integrated quantity is the deposition rate, and for the two extreme cases of HadGEM3-GA4 and GISS-E2 ~~W~~we therefore also try comparing against observations of sulfate wet deposition. We use the 3-year mean wet deposition data from 2000-2002 described in Vet et al. (2014) and provided by the World Data Centre for Precipitation Chemistry (http://wdcpc.org, 2014), taking the 6 stations located in China. We exclude the station in Guizhou province in southern China where HadGEM3-GA4 has a bias of +590% and GISS-E2 a bias of +253%. This station only provided data for one year and was flagged as having a high uncertainty in the Vet et al. (2014) dataset; it is also located in a mountainous region and so it could equally be that the models cannot resolve the specific local conditions. Removing this station from the analysis we find for the remaining 5 stations in China that HadGEM3-GA4 performs well with a mean bias of -3.9%, compared with -64.8% for GISS-E2. This gives an indication that HadGEM3-GA4 has more realistic sulfate deposition directly over China~~,~~ (though the sample size is very small), and supports the earlier findings from the comparison against AERONET and MODIS. If we broaden the analysis to include all stations described as being broadly in Asia – an additional 32 stations – then the mean bias for HadGEM3-GA4 is worsened (-41.8%), whilst the bias in GISS-E2 is slightly improved (-54.1%). HadGEM3-GA4 still performs better over the Asian region as a whole, though less dramatically so (Supplementary fig. S9~~6~~). This overall picture seems consistent with that of the other observational measures looked at here, although it should be noted that wet deposition rates are dependent not just on the column sulfate burden but also on the amount and distribution of precipitation, and so biases in wet deposition could also be due to incorrect precipitation distribution rather than sulfate.

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).

4.2 Differences in cloud effects

Sulfate aerosol exerts indirect radiative effects by modifying cloud properties. The strength of these indirect effects is highly uncertain (e.g. Boucher et al., 2013) and differs substantially between the models, having been shown to contribute substantially to inter-model variation in historical aerosol forcing (Wilcox et al., 2015). Differences in the underlying climatologies of the models, particularly with regard to cloud distributions, could also be important. ~~since~~ For instance, the radiative effect of sulfate aerosol is modulated by the reflectivity of the underlying surface in the radiation scheme (Chýlek and Coakley, 1974; Chand et al., 2009), which may often be a cloud-top. The low contrast with a highly reflective cloud surface means that sulfate aerosol above a cloud top will have a reduced direct radiative forcing. Blocking of radiation by clouds will also reduce the direct radiative effects of any aerosols within or below them (e.g. Keil and Haywood, 2003). Additionally, aerosol indirect effects can saturate in regions with a high level of background aerosol (e.g. Verheggen et al., 2007; Carslaw et al., 2013), meaning that the potential for indirect radiative forcing can also vary with the location of clouds. On top of diversity in indirect effects, and in the climatological distribution of clouds, different dynamical changes in cloud cover could also alter the all-sky flux.

In our case, the good correspondence between higher (clear-sky) AOD change in HadGEM3-GA4 and higher (all-sky) SW flux change in this model ~~would~~ might suggest that the cloud effects are not the root cause of the larger radiative response in this model. However, the origin of this good correspondence in fact appears to be somewhat dependent on how clouds modify the radiative effects of sulfate aerosol:

~~Additionally clear-sky SW flux diagnostics were available for HadGEM3-GA4 and GISS-E2 (Supplementary fig. S7), and comparing~~For the extreme cases of HadGEM3-GA4 and GISS-E2, comparing the changes in clear-sky TOA SW flux~~them~~ with the all-sky TOA SW flux

anomalies (Table 2 and Supplementary Fig. S10) reveals that for clear-sky conditions, there is in fact a much smaller regional~~we still find a large - albeit smaller (3-fold rather than 6-fold) -~~ discrepancy between these two models: Over the E. Asia region GISS-E2 has a 4.1 Wm^{-2} clear-sky SW flux change, whereas HadGEM3-GA4 has a 5.1 Wm^{-2} flux change. HadGEM3-GA4 still has the larger radiative change, but nowhere near the 6-fold difference that is seen in the all-sky flux (Section 3, and Table 2). This much reduced difference between GISS-E2 and HadGEM3-GA4 in the clear-sky compared with the all-sky anomaly is hard to apportion quantitatively though, because compared with the clear-sky change, the all-sky response incorporates all the contributing factors described above:~~both aerosol indirect effects and also dynamical feedbacks on clouds. 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~~ 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~~not shown~~)~~., which~~ Combined with the screening effect of clouds, this appears to almost completely~~partially~~ offset~~s~~ the direct forcing of reduced SO_4, and results in a far smaller all-sky flux change than clear-sky flux change over E. China (0.9~~1~~ Wm^{-2} all-sky compared with 4.1~~1.8~~ Wm^{-2} clear-sky). HadGEM3-GA4 ~~has mixed changes in cloud amount over East Asia (not shown) and~~ by contrast has very little ~~a smaller~~ difference between all-sky and clear-sky flux changes (5.3 Wm^{-2} and 5.1~~8~~ Wm^{-2} respectively (Table 2)). The changes in cloud amount over east China are somewhat more mixed (Supplementary Fig. S11c), although area-averaged, the overall cloud change is a small decrease, which should enhance the all-sky flux change. However, spatially as well as in magnitude the HadGEM3-GA4 all-sky flux change is exceptionally similar to ~~its~~the clear-sky flux change~~s~~, and does not resemble the pattern of cloud changes (comparing Supplementary Figs. S10e,f, and Fig. S11c), which suggests that- changes in aerosol radiative effects are larger than the effect of the small cloud cover changes, and still dominate the all-sky flux changes. Therefore, the very similar regional all-sky and clear-sky SW flux changes in HadGEM3-GA4 imply~~ies~~ that unlike in GISS-E2, aerosol indirect effects in HadGEM3-GA4 probably roughly compensated for the presence of clouds reducing ~~the~~ the

direct effect, so that the change in all-sky combined direct and indirect forcing is similar to the change in clear-sky direct forcing when sulfate is removed.~~., explaining why there is a bigger discrepancy between these two models in the all-sky forcing. Nonetheless, the fact that there is still a 3-fold difference in clear-sky flux indicates that even in a cloud-free world, there would be large disagreement in the models' SW forcing over China, and so cloud responses are not the primary driver of the discrepancies, although cloud feedbacks are clearly important in modulating the final magnitude of the discrepancy.~~

~~Diagnostics for clear-sky radiative fluxes and cloud amount were not available for CESM1, so we are unable to make a similar comparison for this model.~~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 in 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}, ~~which is considerably smaller than~~compared with the 4.1 Wm^{-2} clear-sky change in GISS-E2 (Table 2). However, whereas ~~However, whereas~~in GISS-E2 the all-sky SW flux change (0.9 Wm^{-2}) was ~~seen to~~ then ~~be~~ more than four times smaller than this clear-sky flux change~~in GISS-E2 (0.9 Wm^{-2})~~, ~~conversely~~ in CESM1 the all-sky SW flux change is instead almost two times larger~~wice as big as~~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 SO_2 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. We note though that clear-sky diagnostics will be influenced by choices within the models of how aerosol water uptake is determined under the artificial assumption of clear-sky conditions. The all-sky SW flux change, which drives the final climate response, is regionally still the most directly comparable quantity, reflecting the total radiative effect of the aerosol change in the different models.

4.3 Differences in aerosol forcing efficiency

An additional source of discrepancy between the models lies in differences in the aerosol radiative forcing efficiency – the direct forcing that results from a given aerosol optical depth or burden (e.g. Samset et al, 2013). A previous model intercomparison looking at radiative forcing as part of the AeroCom Phase II study found that, on a global scale, there was a large variation in the radiative forcing due to aerosol-radiation interactions per unit AOD between different participating models (Myhre et al., 2013a) ~~on a global scale~~. As a result, whether a model simulates AOD changes correctly, for instance, may not particularly constrain the resultant direct forcing even, let alone the indirect forcing or eventual climate response.

Globally-averaged, the changes in radiative flux and AOD are too small in our experiments to calculate an accurate ratio, but instead we calculate here a regional radiative efficiency ~~for HadGEM3-GA4 and GISS-E2~~ by taking the change in clear-sky SW flux over the 100-120E, 20-40N E. China region ~~(Sect. 4.2)~~, and dividing by the ~~clear-sky~~ AOD change over the same region (Table 2~~1~~). This is not directly comparable with previous studies like Myhre et al. (2013a), as we use a regionally-averaged number instead of globally-averaged, and for the numerator we use the change in clear-sky TOA SW flux as the best available measure of aerosol direct radiative effect, rather than the ~~clear-sky~~ direct radiative forcing diagnosed either from

double radiation calls or simulations with fixed meteorology. Consequently, we use this metric here mainly to qualitatively highlight differences between the models.

As noted in Sect. 4.1 and 4.2, over the eastern China region HadGEM3-GA4 has a 6-fold larger mean AOD reduction (-0.29) compared with GISS-E2 (-0.047), but only slightly~~a 3-fold~~ larger clear-sky SW change (5.1~~8~~ W m^{-2} compared with 4.1~~.8~~ W m^{-2}). As a result, the regional radiative efficiency for HadGEM3-GA4 is much smaller than~~only about half~~ that of GISS-E2: ~~(~~-17.6~~20.3~~ W m^{-2} compared with -~~39.1~~87.2 W m^{-2}~~)~~ per unit AOD change (Table 2). If instead of AOD we normalise by the change in sulfate burden ~~instead of the AOD~~ integrated over the same region,~~, however,~~ we find a similar~~the opposite~~ relationship: HadGEM3-GA4 has a smaller~~larger~~ regional mean change in clear-sky SW flux per Tg sulfate than GISS-E2: ~~(~~-145~~67.1~~ W m^{-2} Tg^{-1} compared with -256~~117.7~~ W m^{-2} Tg^{-1}~~)~~. Proportionally though, the discrepancy is not as great when normalising by change in sulfate burden, due to the~~ The~~ much larger AOD per unit mass of sulfate simulated in HadGEM3-GA4~~ therefore outweighs the smaller radiative response per unit AOD~~. Curiously Myhre et al. (2013a) reported results that were qualitatively the inverse of what we show here, finding that the atmospheric component of GISS ModelE~~2~~ has a smaller sulfate radiative forcing than that of HadGEM2 (HadGEM3's predecessor, with a very similar aerosol scheme) when normalised by AOD, although still~~but~~ larger when normalised by column-integrated sulfate burden. The reason for the discrepancy is not clear, though the aforementioned fact that we calculate our numbers for a specific region means that there may be important local factors. The sulfate-specific forcing efficiencies in Myhre et al. (2013) are calculated relative to all-sky direct radiative effect, and so local differences in vertical profiles and cloud screening may therefore change the relationship – however they also evaluated clear-sky forcing normalised by AOD for all aerosol species combined, and again found HadGEM2 to be higher than GISS ModelE.~~ For instance, the forcing per unit AOD will be influenced by the vertical distribution of the aerosol (Myhre et al., 2013a), which could vary between models in different parts of the world.~~

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 values given for HadGEM3-GA4 and GISS-E2 ~~values~~ (-17.6 W m^{-2} and -87.2 W m^{-2} respectively) were calculated using clear-sky AOD. ~~Normalising the clear-sky flux change by all-sky AOD in~~ GISS-E2 ~~(which~~ provided both clear-sky and all-sky AOD diagnostics, and using instead the all-sky AOD change~~)~~ from GISS-E2 gives~~ instead~~ a

smaller value of -22.4 W m^{-2} per unit AOD,~~,~~ which suggest~~ings~~ that when compared like-for-like, CESM1 (with -28.4 W m^{-2}) may in fact have the~~a~~ greater regional radiative efficiency. More directly comparable between all three models is the regional clear-sky 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 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.

~~Making an equivalent comparison for CESM1 is hindered by the lack of clear-sky diagnostics available from this model for these simulations. What we can note is that if we instead use the all-sky change in SW flux over East China, normalising by AOD we find a much larger SW change per unit AOD in CESM1 than in HadGEM3-GA4 or GISS-E2 (-55.0 W m^{-2} compared with -18.6 W m^{-2} and -19.6 W m^{-2}) (Table 1). Normalising by all-sky AOD in GISS-E2 (which provides both clear-sky and all-sky diagnostics) however gives a comparatively even smaller value (-4.95 W m^{-2}). Normalised by the change in regional sulfate burden instead, CESM1 sits in the middle with -107.7 W m^{-2} Tg^{-1}, compared with HadGEM3-GA4's -153.5 W m^{-2} Tg^{-1} (quite close to its clear-sky normalised value), and GISS-E2's -56.6 W m^{-2} Tg^{-1} (much smaller than its clear-sky normalised value). These results suggest that either CESM1 has a large radiative efficiency per unit AOD which compensates for its much smaller AOD per mass of sulfate, or else there are large cloud responses – either due to a particularly strong aerosol indirect effect, or a dynamical reduction in local cloudiness – which considerably amplify the radiative effect of a relatively small AOD reduction in this model.~~

The Myhre et al. (2013a) AeroCom intercomparison found that globally, the atmospheric component of CESM1 (CAM5.1) ~~did indeed~~ had~~ve~~ a much higher sulfate radiative efficiency than the atmosphere-only version of GISS-E2. In their case, they found CAM5.1 to have

approximately 2.25 times higher all-sky direct radiative forcing per unit AOD than GISS-E2. However, the study also found that, globally, the atmospheric component of HadGEM2 had a slightly largervery similar forcing efficiency thano CAM5.1 both for sulfate (all-sky) and all aerosols (clear-sky), unlike the somewhat smaller regional efficiencies found here for HadGEM3-GA4 compared with CESM1. Given that our regional values from GISS-E2 and HadGEM3-GA4 also seem to conflict qualitatively with the global values from the AeroCom study, this would suggest that either the global comparison is not relevant on regional scales, or else the radiative efficiency is very sensitive to changes in model configuration and version.though, this probably does not provide a strong indication of which factor is more likely the dominant driver of the relatively large response in CESM1 despite its modest AOD change.

4.4 Differences in climate sensitivity

So far we have discussed mainly factors which influence the translation of a change in aerosol precursor emissions to a radiative heating, and these varied strongly between the models. There is a final step in arriving at the climate response, which is the translation of a given radiative heating into a surface temperature change. The climate sensitivity – the amount of warming simulated per unit radiative forcing – is also well known to vary considerably between models, globally (Flato et al., 2013) and regionally (Voulgarakis and Shindell, 2010), and this will additionally impact the strength of the final response. Climate sensitivity is typically estimated from a 2x or 4x global CO_2 simulation, giving a large response and a large forcing from which to calculate the ratio. For GISS-E2, a climate sensitivity value of 0.6 K $(W\ m^{-2})^{-1}$ was found in the IPCC AR5 report from a 4x CO_2 simulation (Flato et al., 2013) using the regression method of Gregory et al. (2004) to estimate radiative forcing. For CESM1, a value of 1.1 K $(W\ m^{-2})^{-1}$ is obtained from values from a 2x CO_2 simulation (Meehl et al., 2013), noting that in this case the radiative forcing was calculated using the stratospheric adjustment method (Hansen et al., 2005). For HadGEM3-GA4, we use a 100-year 2x CO_2 simulation that was performed separately as part of the Precipitation Driver Response Model Intercomparison Project (Samset et al., 2016in preparation), which gives a value of 1.1 K $(W\ m^{-2})^{-1}$ based on the Gregory method.

While CESM1 and HadGEM3-GA4 both have very similar climate sensitivities, we see that GISS-E2 has a particularly small sensitivity – in fact, the smallest value of all the CMIP5 models reported in the AR5 report (Flato et al., 2013). This presumably compounds the fact

that GISS-E2 simulates the smallest SW flux change of the three models, ensuring that the resulting surface temperature response is comparatively smaller still. Differences in climate sensitivity do not seem to explain any of the variation in the magnitude of the response between CESM1 and HadGEM3-GA4, at least based on these values. However, it is worth noting that the climate sensitivity values that we report are derived from global CO_2 forcings, whereas in our case we are looking at the translation of a very regional forcing into a global response. It is not trivial that the global-mean temperature response to a regionally localised forcing is a function only of the resulting globally-averaged forcing, and in particular it may be that different models are more or less sensitive to forcings in specific regions. Unfortunately we know of no study that has calculated climate sensitivity to regional forcings in single or multi-model frameworks. Shindell (2012) calculated ~~regional~~ climate sensitivities to forcings imposed in different latitudinal bands for the GISS-E2 model, finding that there is considerable ~~regional~~ variation relative to the global climate sensitivity. In that study, estimates of the response to ~~regional~~ forcings at different latitudes in ~~3~~three other global climate models, based on the GISS-E2 ~~regional~~ sensitivities, are found to largely agree to within +/- 20% with the full simulations however, suggesting that regional sensitivities (relative to a model's global sensitivity) may not vary that much between models.

5 Conclusions

By applying an identical regional perturbation to anthropogenic SO_2 emissions in three different climate models, we observe three markedly different resulting climate responses, ranging from virtually no coherent surface air temperature response in one model (GISS-E2), to pronounced surface warming all across most of the northern hemisphere in another (HadGEM3-GA4). The third model (CESM1) sits in the middle in terms of both magnitude and spatial extent of the temperature response. This huge variation in climate response corresponds to a similarly large variation in the SW radiative flux change following the reduction in sulfate aerosol. All three models show a fairly localised increase in net downwards SW radiation over China as a result of reduced SO_2 emissions from this region, however the magnitude of this radiative heating is substantially greater in HadGEM3-GA4 than in CESM1, which is substantially greater still than in GISS-E2. The response in GISS-E2 is so weak that temperature changes are largely not detectable above the internal variability of the model. The stronger heating in CESM1 and HadGEM3-GA4 produces much more pronounced temperature changes, and even though the

radiative heating is localised over China, the temperature responses in these two models are much more spread out, particularly in the zonal direction. This is consistent with the findings of Shindell et al. (2010), who found that the temperature response to inhomogeneous aerosol forcings is more uniform and extends much further from the forcing location in the zonal direction than in the meridional direction.

Comparing the models, we find ~~very~~ different ~~changes in the~~ SO_4 mass changes due to removing SO_2 emissions from China, very different ratios of AOD change per mass of sulfate, and very different radiative flux changes per unit AOD change. These differences are compounded further by very large variations in cloud ~~responses~~interactions, as well as variations in climate sensitivity, and ~~the~~ feedbacks on other aerosol species such as nitrate, which diversify the response further. ~~In addition to differences in the total changes in sulfate and AOD, we find there are also substantial differences in the spatial distribution of the changes, attributed to differences in the rate of chemical conversion of SO_2 to SO_4 which influences how concentrated the aerosol changes are around the emission region. This implies that even if both the AOD per sulfate burden and the forcing per unit AOD were identical among the three models, they would still have different distributions of radiative forcing.~~

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 t~~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.

In addition to differences in the total changes in sulfate and AOD, we find there are also substantial differences in the spatial ~~e~~distribution of the changes, attributed to differences in the rate of chemical conversion of SO_2 to SO_4 which influences how concentrated the aerosol changes are around the emission region. This implies that even if both the AOD per sulfate burden and the forcing per unit AOD were identical among the three models, they would still have different distributions of radiative forcing.

There are no direct observations of sulfate radiative forcing, nor of sulfate optical depth or vertically-integrated burden, and so we have tried validating the aerosol component of the models with a range of surface and satellite-based measurements of total aerosol optical depth, surface sulfate concentration, column SO_2, and sulfate wet deposition. All the models have biases, and no model performs best against all the observational datasets used. Tentatively HadGEM3-GA4 seems to perform best over China against observations of both total AOD and sulfate wet deposition, though over some other parts of the world this model performed slightly poorer, e.g. for global AOD and US surface sulfate concentrations. However, the main conclusion is that comparison against all existing observational measures is unable to satisfactorily constrain which model response is more realistic, given that the ratios of both AOD change per sulfate burden change and SW flux change per AOD (Table 2) are found to vary so substantially between the models. The model with the largest sulfate mass change (CESM1) did not have the largest radiative or climate response, and two models with a similar AOD change (CESM1 and GISS-E2) had markedly different radiative and climate responses. Given the range of discrepancies that we find in all steps along the conversion of SO_2 change to SO_4 change to AOD change to radiative forcing to temperature response, it seems that knowing how accurate a model is with respect to either sulfate concentrations or total AOD is far from sufficient to determine whether the climate response to a regional aerosol perturbation is similarly accurate.

There are several 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 not only the column-integrated sulfate burden, but also 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.

We have only looked here at surface temperature, which is a particularly direct measure of the climate response. The response of other, less well-constrained, climate variables such as precipitation might be expected to show even greater variation. Our results show that there remains a very large uncertainty in current climate models in the translation of aerosol precursor emissions into a climate response, and imply that care must be taken not to over-interpret ~~the results of~~ studies of aerosol-climate interaction ~~performed with single models~~if the robustness of results across diverse models cannot be demonstrated.

On a more optimistic note, we remark that in the two models which showed the more substantial change in SW radiative flux (CESM1 and HadGEM3-GA4), both also show a remarkably strong remote temperature response to a relatively localised northern-midlatitude heat source, with qualitatively similar temperature change patterns that extend across much of the hemisphere, indicating that there may be some agreement on the response to a given regional forcing, if not on the forcing itself.

Data availability

Model output data from all simulations described here is available upon request from the corresponding author.

Acknowledgements

MK and AV are supported by the Natural Environment Research Council under grant number NE/K500872/1. Also, we wish to thank the European Commission's Marie Curie Actions

International Research Staff Exchange Scheme (IRSES) for funding MK's placement at NASA GISS and Columbia University and facilitating interactions on this work with the US colleagues, as part of the Regional Climate-Air Quality Interactions (REQUA) project. Simulations with GISS-E2 used resources provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. Simulations with HadGEM3-GA4 were performed using the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council. We specifically thank Dr. Fiona O'Connor, Dr. Jeremy Walton, and Mr. Mohit Dalvi from the Met Office for their support with using the HadGEM3-GA4 model.

References

Andreae, M. O. and Crutzen, P. J.: Atmospheric Aerosols: Biogeochemical Sources and Role in Atmospheric Chemistry, Science, 276, 1052-1058, 1997.

Andres, R. J. and Kasgnoc, A. D.: A time-averaged inventory of subaerial volcanic sulfur emissions, J. Geophys. Res., 103, D19, 25251-25261, 1998.

Andrews, T., Forster, P. M., Boucher, O., Bellouin, N., and Jones, A.: Precipitation, radiative forcing and global temperature change, Geophys. Res. Lett., 37, L14701, doi:10.1029/2010GL043991, 2010.

Bauer, S. E., Bausch, A., Nazarenko, L., Tsigaridis, K., Xu, B., Edwards, R., Bisiaux, M., and McConnell, J.: Historical and future black carbon deposition on the three ice caps: Ice core measurements and model simulations from 1850 to 2100, J. Geophys. Res. Atmos., 118, 7948–7961, doi:10.1002/jgrd.50612, 2013.

Bellouin, N., Boucher, O., Haywood, J., Johnson, C., Jones, A., Rae, J., and Woodward, S.: Improved representation of aerosols for HadGEM2, Technical Note 73, Hadley Centre, Met Office, Exeter, UK, 2007.

Bellouin, N., Rae, J., Jones, A., Johnson, C., Haywood, J., and Boucher, O.: Aerosol forcing in the Climate Model Intercomparison Project (CMIP5) simulations by HadGEM2-ES and the role of ammonium nitrate, J. Geophys. Res., 116, D20206, doi:10.1029/2011JD016074, 2011.

Bollasina, A. M., Ming, Y., and Ramaswamy, V.: Anthropogenic Aerosols and the Weakening of the South Asian Summer Monsoon, Science, 334, 502-505, 2011.

Booth, B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N.: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, Nature, 484, 228-232, doi:10.1038/nature10946, 2012.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S. K., Sherwood, S., Stevens, B., and Zhang, X. Y.: Clouds and Aerosols. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K.,

Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503, 67-71, 10.1038/nature12674, 2013.

Chand, D., Wood, R., Anderson, T. L., Satheesh, S. K., and Charlson, R. J.: Satellite derived direct radiative effect of aerosols dependent on cloud cover, Nature Geosci., 2, 181–184, 2009.

Chýlek, P. and Coakley, J. A. Jr.: Aerosol and climate, Science, 183, 75-77, 1974.

Cionni, I., Eyring, V., Lamarque, J. F., Randel, W. J., Stevenson, D. S., Wu, F., Bodeker, G. E., Shepherd, T. G., Shindell, D. T., and Waugh, D. W.: Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing, Atmos. Chem. Phys., 11, 11267-11292, doi:10.5194/acp-11-11267-2011, 2011.

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., and Brown, P. D.: Atmospheric radiative transfer modeling: a summary of the AER codes, J. Quant. Spectrosc. Radiat. Transfer, 91, 233–244, 2005.

Derwent, R. G., Collins, W. J., Jenkin, M. E., Johnson, C. E., and Stevenson, D. S.: The global distribution of secondary particulate matter in a 3D Lagrangian chemistry transport model, J. Atmos. Chem., 44, 57–95, 2003.

Dong, B., Sutton, R. T., Highwood, E., and Wilcox, L.: The Impacts of European and Asian Anthropogenic Sulfur Dioxide Emissions on Sahel Rainfall, J. Climate, 27, 7000–7017, doi:10.1175/JCLI-D-13-00769.1, 2014.

Edwards, J. M. and Slingo, A.: Studies with a flexible new radiation code, I: Choosing a configuration for a large-scale model, Q. J. Roy. Meteorol. Soc., 122, 689–719, 1996.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., Qin, D., Plattner, G.-K., Tignor,

M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Goto, D., Nakajima, T., Dai, T., Takemura, T., Kajino, M., Matsui, H., Takami, A., Hatakeyama, S., Sugimoto, N., Shimizu, A., and Ohara, T.: An evaluation of simulated particulate sulfate over East Asia through global model intercomparison, J. Geophys. Res. Atmos., 120, doi:10.1002/2014JD021693, 2015.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, L03205, doi:10.1029/2003GL018747, 2004.

Hansen, J. E., Russell, G. L., Rind, D., Stone, P., Lacis, A., Ruedy, R., and Travis, L.: Efficient three-dimensional models for climatic studies. Mon. Wea. Rev., 111, 609–662, 1983.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, Ja., Perlwitz, Ju., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., and Zhang, S.: Efficacy of climate forcings, J. Geophys. Res., 110, D18104, doi:10.1029/2005JD005776, 2005.

Hemispheric Transport of Air Pollution (HTAP): Hemispheric Transport of Air Pollution 2010. Part A: Ozone and Particulate Matter, Air Pollution Studies No. 17, [Dentener, F., Keating, T., and Akimoto, H. (eds.)], United Nations, New York, 2010.

Holben, B. N., Tanré, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J. S., Chatenet, B., Lavenu, F., Kaufman, Y. J., Vande Castle, J., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N. T., Pietras, C., Pinker, R. T., Voss, K., and Zibordi, G.: An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, J. Geophys. Res., 106(D11), 12067–12097, doi:10.1029/2001JD900014, 2001.

Hunke, E. C. and Lipscombe, W. H.: CICE: the Los Alamos sea ice model documentation and software user's manual, Version 4.0, LA-CC-06-012, Los Alamos National Laboratory, New Mexico, 2008.

Hwang, Y.-T., Frierson, D. M. W., and Kang, S. M.: Anthropogenic sulfate aerosol and the southward shift of tropical precipitation in the late 20th century, Geophys. Res. Lett., 40, doi:10.1002/grl.50502, 2013.

Jones, A., Roberts, D. L., Woodage, M. J., and Johnson, C. E.: Indirect sulphate aerosol forcing in a climate model with an interactive sulphur cycle, J. Geophys. Res., 106, 20293–20310, 2001.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A., and Holben, B. N.: Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res., 115, D23209, doi:10.1029/2010JD014601, 2010.

Keil, A., and Haywood, J. M.: Solar radiative forcing by biomass burning aerosol particles during SAFARI 2000: A case study based on measured aerosol and cloud properties, J. Geophys. Res., 108(D13), 8467, doi:10.1029/2002JD002315, 2003.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., Berntsen, T., Berglen, T. F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevåg, A., Kloster, S., Koch, D., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Lesins, G., Liu, X., Lohmann, U., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, O., Stier, P., Takemura, T., and Tie, X.: An AeroCom initial assessment – optical properties in aerosol component modules of global models, Atmos. Chem. Phys., 6, 1815-1834, doi:10.5194/acp-6-1815-2006, 2006.

Koch, D., Schmidt, G. A., and Field, C. V.: Sulfur, sea salt and radionuclide aerosols in GISS ModelE, J. Geophys. Res., 111, doi:10.1029/2004JD005550, 2006.

Koch, D., Bauer, S., Del Genio, A., Faluvegi, G., McConnell, J. R., Menon, S., Miller, R. L., Rind, D., Ruedy, R., Schmidt, G. A., and Shindell, D.: Coupled aerosol-chemistry-climate twentieth century transient model investigation: Trends in short-lived species and climate responses, J. Climate, 24, 2693–2714, doi:10.1175/2011JCLI3582.1, 2011.

Krotkov, N. A., McClure, B., Dickerson, R. R., Carn, S. A., Li, C., Bhartia, P. K., Yang, K., Krueger, A. J., Li, Z., Levelt, P. F., Chen, H., Wang, P., and Lu, D.: Validation of SO2 retrievals from the Ozone Monitoring Instrument over NE China, J. Geophys. Res., 113, D16S40, doi:10.1029/2007JD008818, 2008.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmos. Chem. Phys., 10, 7017-7039, doi:10.5194/acp-10-7017-2010, 2010.

Lawrence, D.M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, J. Adv. Model. Earth Sys., 3, DOI: 10.1029/2011MS000045, 2011.

Lee, Y.-H., and Adams, P. J.: A fast and efficient version of the TwO-Moment Aerosol Sectional (TOMAS) global aerosol microphysics model, Aerosol. Sci. Technol., 46, 678–689, doi:10.1080/02786826.2011.643259, 2012.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989-3034, doi:10.5194/amt-6-2989-2013, 2013.

Liu, X., Easter, R. C., Ghan, S., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N, Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D., Toward a Minimal Representation of Aerosol Direct and Indirect Effects: Model Description and Evaluation, GeoSci. Mod. Dev., 5, 709-739, doi:10.5194/gmd-5-709-2012, 2012.

Madec, G.: NEMO ocean engine, Institut Piere-Simon Laplace (IPSL), France, No. 27, ISSN No. 1288–1619, 2008.

Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A., and Cahill, T. A.: Spatial and seasonal trends in particle concentration and optical extinction in the United States, J. Geophys. Res., 99, 1347–1370, 1994.

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., and Johnson, C. E.: Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model, Geosci. Model Dev., 3, 519-551, doi:10.5194/gmd-3-519-2010, 2010.

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., Gettleman, A., Lawrence, D. M., Sanderson, B. M., and Strand, W. G.: Climate Change Projections in CESM1(CAM5) Compared to CCMS4, Journal of Climate, 26, 6287-6308, 2013.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213-241, 10.1007/s10584-011-0156-z, 2011.

Menon, S., Koch, D., Beig, G., Sahu, S., Fasullo, J., and Orlikowski, D.: Black carbon aerosols and the third polar ice cap, Atmos. Chem. Phys., 10, 4559–4571, 2010.

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., Del Genio, A. D., Kelley, M., Lo, K. K., Ruedy, R., Shindell, D. T., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Chen, Y.-H., Cheng, Y., Clune, T. L., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Menon, S., Oinas, V., Pérez García-Pando, C., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Sun, S., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: CMIP5 historical simulations (1850-2012) with GISS ModelE2, J. Adv. Model. Earth Syst., 6, no. 2, 441-477, doi:10.1002/2013MS000266, 2014.

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmos. Chem. Phys., 13, 1853-1877, doi:10.5194/acp-13-1853-2013, 2013a.

Myhre, G., Shindell, D., Breon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T. F., Qin, D., Plattner, G.-K., Tignor,

M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013b.

Neale, R. B., Chen, C.-C., Gettleman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, D., Lamarque, J.-F., Marsh, D., Mills, M., Smith, A. K., Tilmes, S., Vitt, F., Morrison, H., Cameron-Smith, P., Collins, W. D., Iacono, M. J., Easter, R. C., Ghan, S. J., Liu, X., Rasch, P. J., and Taylor, M. A.: Description of the NCAR Community Atmosphere Model (CAM 5.0), NCAR Technical Note TN-486+STR, National Center for Atmospheric Research, Boulder, Colorado, USA, 2012.

Pan, X., Chin, M., Gautam, R., Bian, H., Kim, D., Colarco, P. R., Diehl, T. L., Takemura, T., Pozzoli, L., Tsigaridis, K., Bauer, S., and Bellouin, N.: A multi-model evaluation of aerosols over South Asia: common problems and possible causes, Atmos. Chem. Phys., 15, 5903-5928, doi:10.5194/acp-15-5903-2015, 2015.

Polson, D., Bollasina, M., Hegerl, G. C., and Wilcox, L. J.: Decreased monsoon precipitation in the Northern Hemisphere due to anthropogenic aerosols, Geophys. Res. Lett., 41, doi:10.1002/2014GL060811, 2014.

Reddy, S., Seland, O., Stier, P., Takemura, T., and Tie, X.: An AeroCom initial assessment – optical properties in aerosol component modules of global models, Atmos. Chem. Phys., 6, 1815-1834, doi:10.5194/acp-6-1815-2006, 2006.

Russell, G. L., Miller, J. R., and Rind, D.: A coupled atmosphere-ocean model for transient climate change, Atmosphere-Ocean, 33(4), 683–730, 1995.

Samset, B. H., Myhre, G., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Diehl, T., Easter, R. C., Ghan, S. J., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Penner, J. E., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., and Zhang, K.: Black carbon vertical profiles strongly affect its radiative forcing uncertainty, Atmos. Chem. Phys., 13, 2423-2434, doi:10.5194/acp-13-2423-2013, 2013.

~~Samset et al., in preparation.~~ Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Andrews, T., Faluvegi, G., Fläschner, D., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J.-F., Olivié, D., Richardson, T., Shindell, D., Shine, K. P., Takemura, T., and Voulgarakis, A: Fast and slow

precipitation responses to individual climate forcers: A PDRMIP multimodel study, Geophys. Res. Lett., 43, doi:10.1002/2016GL068064, 2016.

Schmidt, G.A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y.-H., Cheng, Y., Clune, T. L., Del Genio, A., de Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, J. Adv. Model. Earth Syst., 6, no. 1, 141-184, doi:10.1002/2013MS000265, 2014.

Schulz, M., Textor, C., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Dentener, F., Guibert, S., Isaksen, I. S. A., Iversen, T., Koch, D., Kirkevåg, A., Liu, X., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, S., Seland, Ø., Stier, P., and Takemura, T.: Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations, Atmos. Chem. Phys., 6, 5225-5246, doi:10.5194/acp-6-5225-2006, 2006.

Shindell, D. T.: Evaluation of the absolute regional temperature potential, Atmospheric Chemistry and Physics, 12, 7955-7960, doi:10.5194/acp-12-7955-2012, 2012.

Shindell, D. and Faluvegi, G.: Climate response to regional radiative forcing during the 20th century, Nat. Geosci., 2, 294-300, 2009.

Shindell, D., Schulz, M., Ming, Y., Takemura, T., Faluvegi, G., and Ramaswamy, V.: Spatial scales of climate response to inhomogeneous radiative forcing, J. Geophys. Res., 115, D19110, doi:10.1029/2010JD014108, 2010.

Shindell, D. T., Voulgarakis, A., Faluvegi, G., and Milly, G.: Precipitation response to regional radiative forcing, Atmos. Chem. Phys., 12, 6969–6982, 2012.

Shindell, D. T., Lamarque, J.-F., Schulz, M., Flanner, M., Jiao, C., Chin, M., Young, P. J., Lee, Y. H., Rotstayn, L., Mahowald, N., Milly, G., Faluvegi, G., Balkanski, Y., Collins, W. J., Conley, A. J., Dalsoren, S., Easter, R., Ghan, S., Horowitz, L., Liu, X., Myhre, G., Nagashima, T., Naik, V., Rumbold, S. T., Skeie, R., Sudo, K., Szopa, S., Takemura, T.,

Voulgarakis, A., Yoon, J.-H., and Lo, F.: Radiative forcing in the ACCMIP historical and future climate simulations, Atmos. Chem. Phys., 13, 2939-2974, doi:10.5194/acp-13-2939-2013, 2013a.

Shindell, D. T., Pechony, O., Voulgarakis, A., Faluvegi, G., Nazarenko, L., Lamarque, J.-F., Bowman, K., Milly, G., Kovari, B., Ruedy, R., and Schmidt, G.: Interactive ozone and methane chemistry in GISS-E2 historical and future climate simulations, Atmos. Chem. Phys., 13, 2653-2689, doi:10.5194/acp-13-2653-2013, 2013b.

Shindell, D. T., Faluvegi, G., Rotstayn, L., and Milly, G.: Spatial patterns of radiative forcing and surface temperature response, J. Geophys. Res. Atmos., 120, doi:10.1002/2014JD022752, 2015

Smith, S. J., van Aardenne, J., Klimont, Z., Andres, R. J., Volke, A., and Delgado Arias, S.: Anthropogenic sulfur dioxide emissions: 1850–2005, Atmos. Chem. Phys., 11, 1101-1116, doi:10.5194/acp-11-1101-2011, 2011.

Stier, P., Schutgens, N. A. J., Bellouin, N., Bian, H., Boucher, O., Chin, M., Ghan, S., Huneeus, N., Kinne, S., Lin, G., Ma, X., Myhre, G., Penner, J. E., Randles, C. A., Samset, B., Schulz, M., Takemura, T., Yu, F., Yu, H., and Zhou, C.: Host model uncertainties in aerosol radiative forcing estimates: results from the AeroCom Prescribed intercomparison study, Atmos. Chem. Phys., 13, 3245-3270, doi:10.5194/acp-13-3245-2013, 2013.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bull. Amer. Meteor. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Teng, H., Washington, W. M., Branstator, G., Meehl, G. A., and Lamarque, J.-F.: Potential impacts of Asian carbon aerosols on future US warming, Geophys. Res. Lett., 39, L11703, doi:10.1029/2012GL051723, 2012.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I., Iversen, I., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: Analysis and quantification of the diversities

of aerosol life cycles within AeroCom, Atmos. Chem. Phys., 6, 1777-1813, doi:10.5194/acp-6-1777-2006, 2006.

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Ma, P.-L., Liu, X., Ghan, S., Bardeen, C., Arnold, S., Deeter, M., Vitt, F., Ryerson, T., Elkins, J. W., Moore, F., and Spackman, R.: Description and evaluation of tropospheric chemistry and aerosols in the Community Earth System Model (CESM1.2), Geosci. Model Dev., 8, 1395-1426, doi:10.5194/gmd-8-1395-2015, 2015.

Verheggen, B., Cozic, J., Weingartner, E., Bower, K., Mertes, S., Connolly, P., Gallagher, M., Flynn, M., Choularton, T., and Baltensperger, U.: Aerosol partitioning between the interstitial and the condensed phase in mixed-phase clouds, J. Geophys. Res., 112, D23202, doi:10.1029/2007JD008714, 2007.

Vet, R., Artz, R. S., Carou, S., Shaw, M., Ro, C.-U., Aas, W., Baker, A., Bowersox, V. C., Dentener, F., Galy-Lacaux, C., Hou, A., Pienaar, J. J., Gillett, R., Forti, M. C., Gromov, S., Hara, H., Khodzher, T., Mahowald, N. M., Nickovic, S., Rao, P. S. P., and Reid, N. W.: A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus, Atmospheric Environment, 93, 3-100, doi:10.1016/j.atmosenv.2013.10.060, 2014.

Voulgarakis, A., and Shindell, D. T., Constraining the sensitivity of regional climate with the use of historical observations. J. Climate, 23, 6068-6073, doi:10.1175/2010JCLI3623.1, 2010.

Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., Lock, A. P., Morcrette, C. J., Stratton, R. A., Wilkinson, J. M., Willett, M. R., Bellouin, N., Bodas-Salcedo, A., Brooks, M. E., Copsey, D., Earnshaw, P. D., Hardiman, S. C., Harris, C. M., Levine, R. C., MacLachlan, C., Manners, J. C., Martin, G. M., Milton, S. F., Palmer, M. D., Roberts, M. J., Rodríguez, J. M., Tennant, W. J., and Vidale, P. L.: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations, Geosci. Model Dev., 7, 361-386, doi:10.5194/gmd-7-361-2014, 2014.

Wilcox, L. J., Highwood, E. J., and Dunstone, N. J.: The influence of anthropogenic aerosol on multi-decadal variations of historical global climate Environ. Res. Lett., 8, 024033, 2013.

Wilcox, L. J., Highwood, E. J., Booth, B. B. B., and Carslaw, K. S.: Quantifying sources of inter-model diversity in the cloud albedo effect, Geophys. Res. Lett., 42, 1568–1575, doi:10.1002/2015GL063301, 2015.

Yu, H., Chin, M., West, J. J., Atherton, C. S., Bellouin, N., Bergmann, D., Bey, I., Bian, H., Diehl, T., Forberth, G., Hess, P., Schulz, M., Shindell, D., Takemura, T., and Tan, Q.: A multimodel assessment of the influence of regional anthropogenic emission reductions on aerosol direct radiative forcing and the role of intercontinental transport, J. Geophys. Res., 118, 700-720, doi:10.1029/2012JD018148, 2013.

Zhang, X. Y., Wang, Y. Q., Niu, T., Zhang, X. C., Gong, S. L., Zhang, Y. M., and Sun, J. Y.: Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols, Atmos. Chem. Phys., 12, 779-799, doi:10.5194/acp-12-779-2012, 2012.

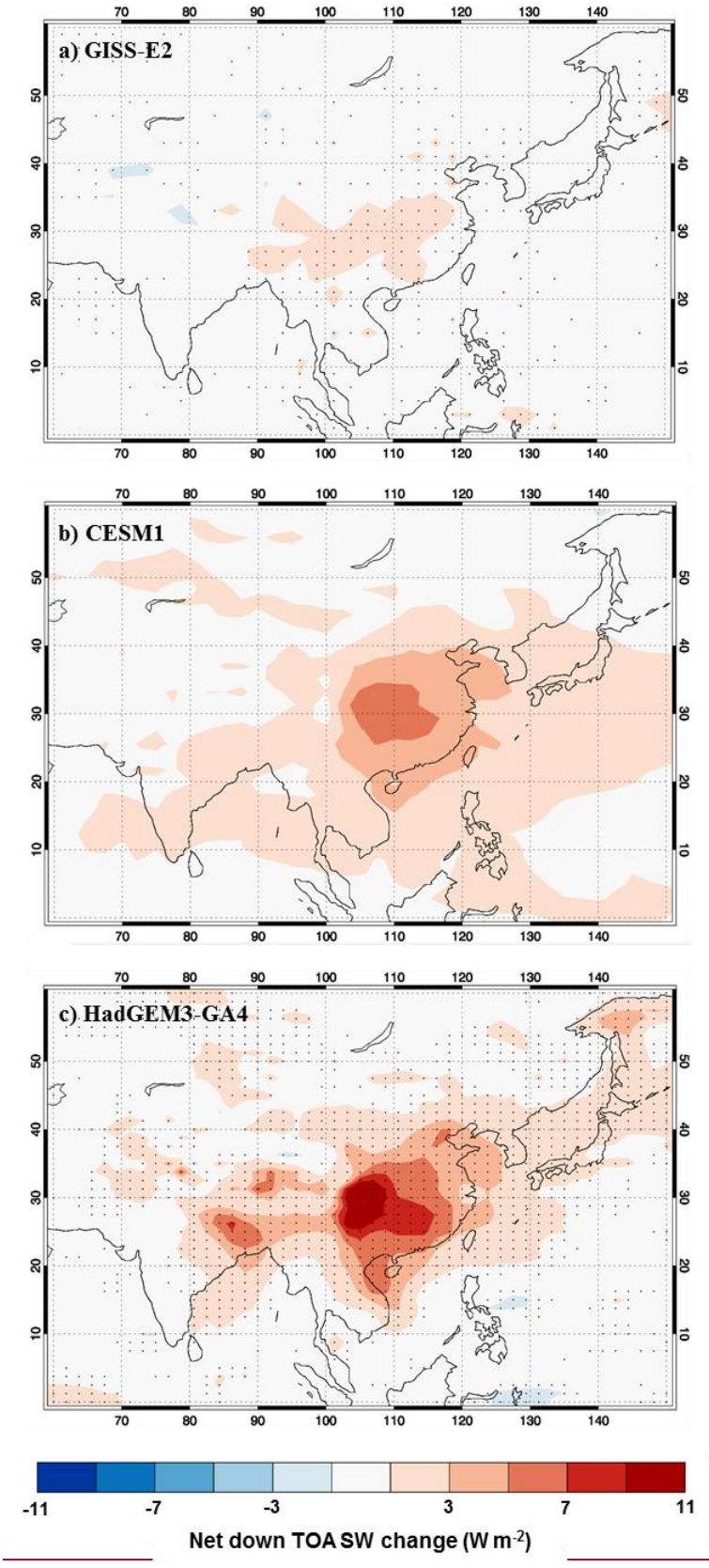
	HadGEM3-GA4	CESM1	GISS-E2
Primary model reference	Walters et al. (2014)	Tilmes et al. (2015)	Schmidt et al. (2014)
Aerosol scheme references	Bellouin et al. (2011) Jones et al. (2001)	Liu et al. (2012)	Koch et al. (2011) Koch et al. (2006)
Resolution (longitude x latitude)	1.875° x 1.25° 85 vertical levels, model top at 85 km	2.5° x 1.875° 30 vertical levels, model top at 40 km	2.5° x 2° 40 vertical levels, model top at 80km
Aerosol tracers	Sulfate, fossil-fuel black carbon, fossil-fuel organic carbon, biomass-burning, dust, sea salt	Sulfate, black carbon, primary organic matter, secondary organic aerosol, dust, sea salt	Sulfate, nitrate, black carbon, organic carbon, secondary organic aerosol, dust, sea salt
Indirect effects included	Yes (1st and 2nd)	Yes (1st and 2nd)	Yes (1st and 2nd)
SO_2 oxidation reactions included	OH (gas phase) H_2O_2, O_3 (aqueous phase)	OH (gas phase) H_2O_2, O_3 (aqueous phase)	OH (gas phase) H_2O_2 (aqueous phase)
Chemistry	Offline (prescribed 4D oxidant fields)	Online	Online
Shortwave radiation	Edwards and Slingo (1996) 6 spectral bands	Clough et al. (2005) 14 spectral bands	Hansen et al. (1983) 6 spectral bands

Table 1: Key references and features of the three models and their aerosol schemes used in this study

		HadGEM3-GA4			GISS-E2			CESM1		
		Con	Ch0	Ch0-Con	Con	Ch0	Ch0-Con	Con	Ch0	Ch0-Con
Global	Total SO_2 (Tg)	0.637	0.592	**-0.045 ± 0.001**	1.151	1.075	**-0.076**	0.553	0.503	**-0.050**
	Total SO_4 (Tg)	1.569	1.499	**-0.070 ± 0.004**	1.091	1.014	**-0.077**	1.459	1.323	**-0.136**
	Mean AOD	0.217	0.213	**-0.0042 ± 0.0004**	0.131	0.131	**-0.0003**	0.123	0.122	**-0.0013**
	Clear-sky TOA SW flux (W m^{-2})	286.0	286.2	**0.184 ± 0.06**	289.0	289.1	**0.052**	288.7	288.8	**0.076**
	All-sky TOA SW flux (W m^{-2})	242.3	242.6	**0.279 ± 0.10**	241.0	241.0	**-0.034 ± 0.06**	236.7	236.9	**0.186**
	Mean temperature (K)	288.6	288.7	**0.115 ± 0.05**	289.0	289.0	**-0.028 ± 0.04**	288.0	288.1	**0.054**
	Δ AOD/Δ SO_4 (Tg^{-1})			**0.0603**			**0.0042**			**0.0094**
	Δ Clear-sky SW/Δ AOD (W m^{-2})			**-43.8**			**-173**			**-58.5**
	Δ All-sky SW/Δ AOD (W m^{-2})			**-66.4**			**106**			**-145**
E. China (100°E-120°E, 20°N-40°N)	Total SO_2 (Tg)	0.035	0.006	**-0.029 ± 0.0002**	0.033	0.005	**-0.028**	0.030	0.001	**-0.028**
	Total SO_4 (Tg)	0.050	0.015	**-0.035 ± 0.0003**	0.043	0.027	**-0.016**	0.054	0.015	**-0.039**
	Mean AOD	0.576	0.289	**-0.287 ± 0.002**	0.232	0.185	**-0.047**	0.227	0.151	**-0.076**
	Clear-sky TOA SW flux (W m^{-2})	296.3	301.4	**5.06 ± 0.08**	294.3	298.4	**4.10**	305.35	307.51	**2.16**
	All-sky TOA SW flux (W m^{-2})	228.8	234.2	**5.34 ± 0.3**	233.32	234.22	**0.90 ± 0.3**	224.16	228.36	**4.20**
	Mean temperature (K)	287.6	287.9	**0.382 ± 0.07**	288.965	289.014	**0.049 ± 0.07**	289.110	289.404	**0.294**
	Δ AOD/ Δ SO_4 (Tg^{-1})			**8.23**			**2.94**			**1.96**
	Δ Clear-sky SW/Δ AOD (W m^{-2})			**-17.6**			**-87.2**			**-28.4**
	Δ All-sky SW/Δ AOD (W m^{-2})			**-18.6**			**-19.3**			**-55.0**

		HadGEM3-GA4			GISS-E2			CESM1		
		Con	**Ch0**	**Ch0-Con**	**Con**	**Ch0**	**Ch0-Con**	**Con**	**Ch0**	**Ch0-Con**
Global	**Total SO2 (Tg)**	0.6370	0.5917	**-0.0453**	1.1511	1.0753	**-0.0758**	-	-	**-**
	Total SO4 (Tg)	1.5689	1.4993	**-0.0696**	1.0907	1.0142	**-0.0765**	1.4594	1.3234	**-0.1360**
	Mean AOD	0.21692	0.21272	**-0.00420**	0.13122	0.13090	**-0.00032**	0.12336	0.12208	**-0.00128**
	Mean TOA SW (Wm^{-2})	242.274	242.553	**0.279**	241.030	240.996	**-0.034**	236.678	236.864	**0.186**
	Mean temp (K)	289.677	289.791	**0.114**	288.987	288.959	**-0.028**	288.047	288.102	**0.054**
	Δ AOD/Δ SO4 (Tg^{-1})			**0.0603**			**0.00418**			**0.00941**
	Δ TOA SW/Δ AOD ($W\ m^{-2}$)			**-66.443**			**(+)105.723**			**-144.961**
100-120E, 20-40N	**Total SO4 (Tg)**	0.050229	0.015419	**-0.034810**	0.042600	0.026605	**-0.015995**	0.054137	0.015172	**-0.038965**
	Mean AOD	0.57565	0.28904	**-0.28661**	0.23156	0.18459	**-0.04697**	0.22705	0.15071	**-0.07634**
	Mean TOA SW ($W\ m^{-2}$)	228.828	234.171	**5.343**	233.319	234.224	**0.905**	224.160	228.355	**4.195**
	Mean temp (K)	288.687	289.095	**0.407**	288.965	289.014	**0.049**	289.110	289.404	**0.294**
	Δ AOD/ Δ SO4 (Tg^{-1})			**8.23**			**2.94**			**1.96**
	Δ TOA SW/Δ AOD ($W\ m^{-2}$)			**-18.642**			**-19.268**			**-54.952**

Table ~~1~~2: Area-integrated SO_2 and SO_4 burdens, area-weighted annual means of AOD, net down clear-sky and all-sky TOA SW flux, and surface temperature, and ratios of the ~~mean~~ changes in ~~TOA SW~~ AOD to change in SO_4 burden, and SW flux to change in AOD, for the globe and the E. China region 100°E - 120°E, 20°N - 40°N. Values are shown for each model for the control simulation (Con), the simulation with no SO_2 emissions from China (Ch0), and the difference (Ch0 – Con). AOD is diagnosed for clear-sky conditions in HadGEM3-GA4 and GISS-E2, and for all-sky conditions in CESM1. ~~Global SO_2 burden was calculated only for HadGEM3-GA4 and GISS-E2.~~ For models and variables where data was available, error ranges are quoted for the Ch0-Con values and indicate ± 2 standard deviations, evaluated in HadGEM3-GA4 from an ensemble of six 150-year control runs with perturbed initial conditions, and in GISS-E2 from twelve 150-year segments of a long pre-industrial control run. Values quoted without error ranges indicate that uncertainty was not evaluated.


a) GISS-E2
b) CESM1
c) HadGEM3-GA4
70 80 90 100 110 120 130 140
10 20 30 40 50
-11 -7 -3 3 7 11
Net down TOA SW change (W m-2)

1

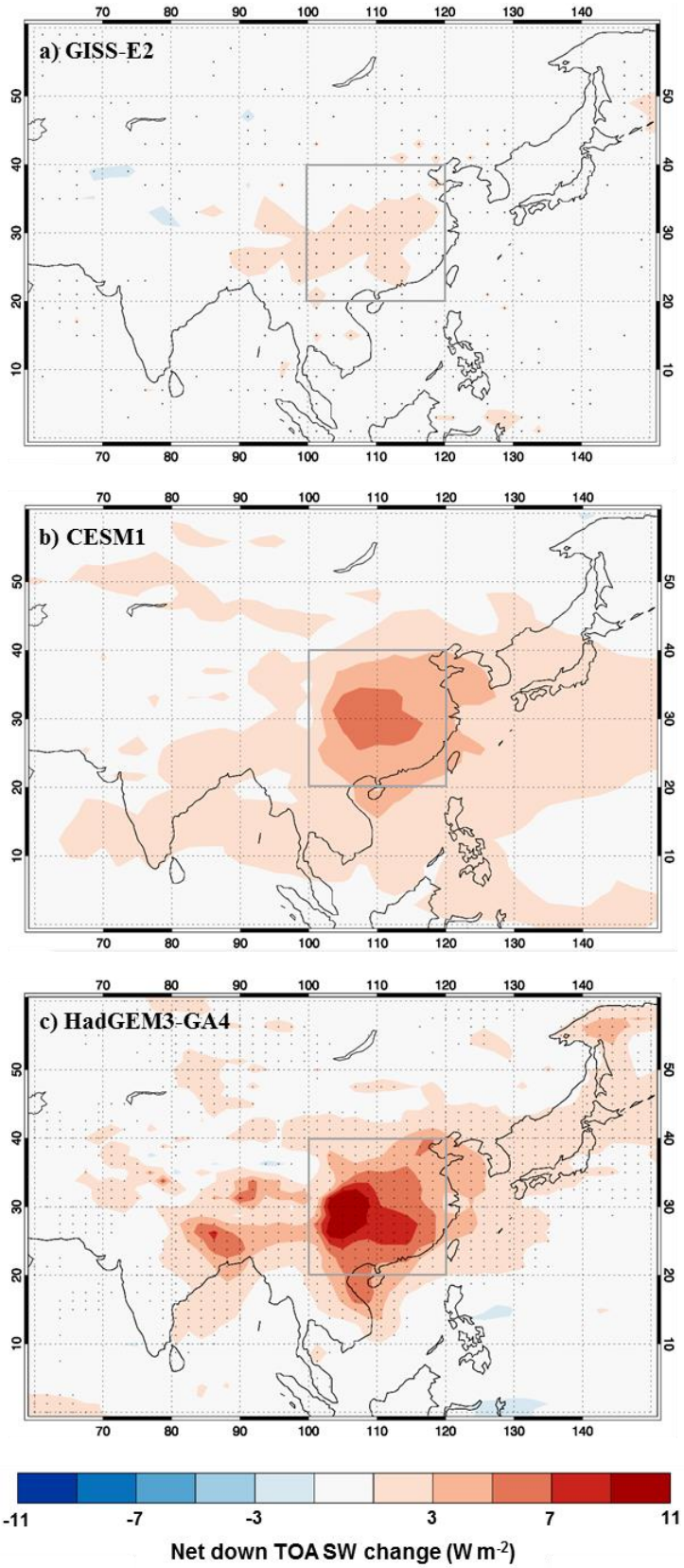
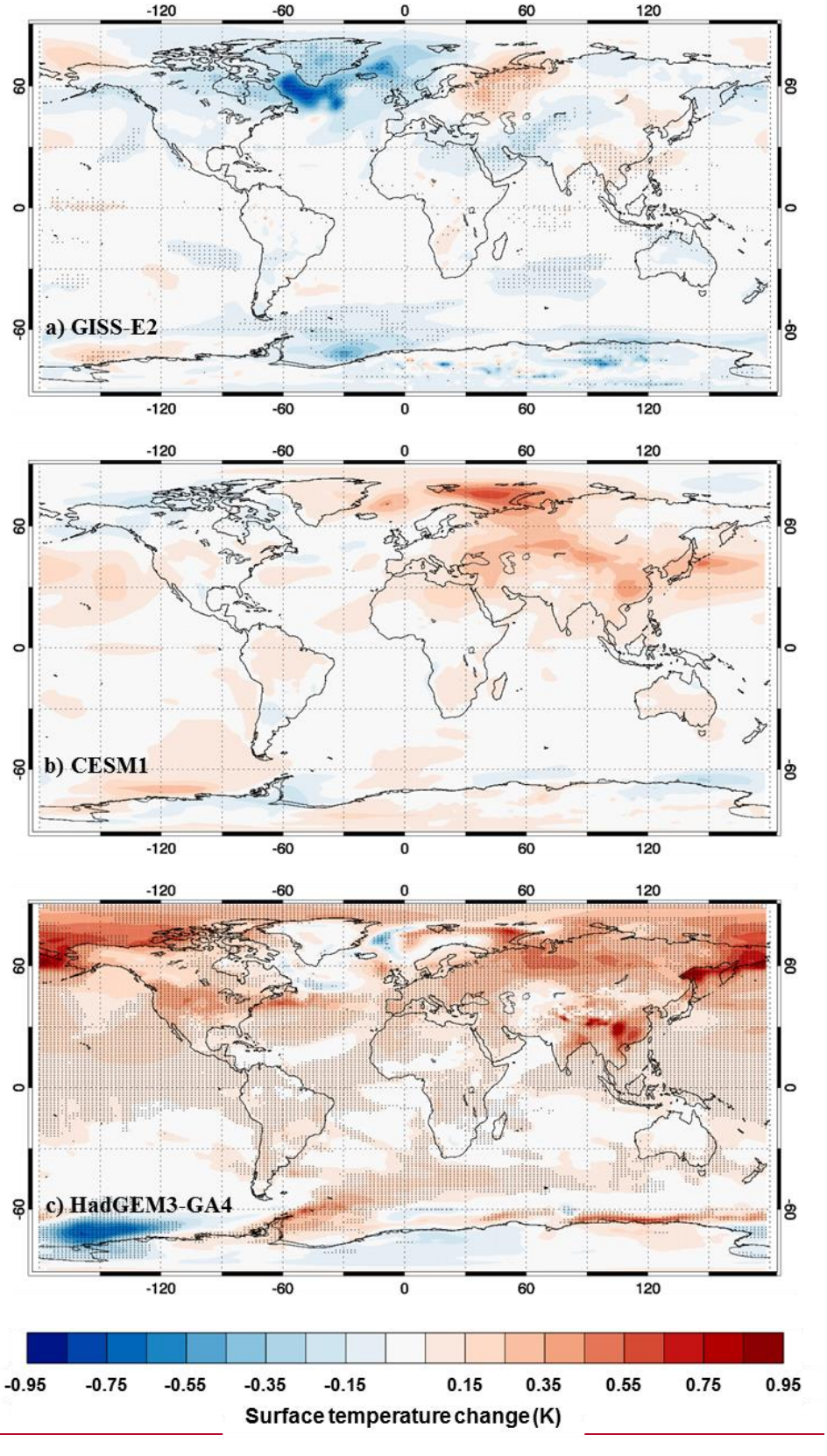


Figure 1: Change in net downward TOA SW flux due to removal of anthropogenic SO_2 emissions over China for a) GISS-E2, b) CESM1, and c) HadGEM3-GA4. Differences are

calculated as the 150-year annual mean of the perturbation simulation minus the 150-year annual mean of the control simulation. Plots focuses on the Asian region as changes outside this domain were found to be minimal. Stippling for GISS-E2 and HadGEM3-GA4 indicates that the change in that grid-box exceeded two standard deviations. Significance was not evaluated for CESM1 as multiple 150-year control runs were not available to assess internal variability for this model. The grey box denotes the E. China (100°E - 120°E, 20°N - 40°N) region which is used in Table 2 and throughout the discussion.


-120
-60
0
60
120
60
0
-60
a) GISS-E2
b) CESM1
c) HadGEM3-GA4
-0.95
-0.75
-0.55
-0.35
-0.15
0.15
0.35
0.55
0.75
0.95
Surface temperature change (K)

1

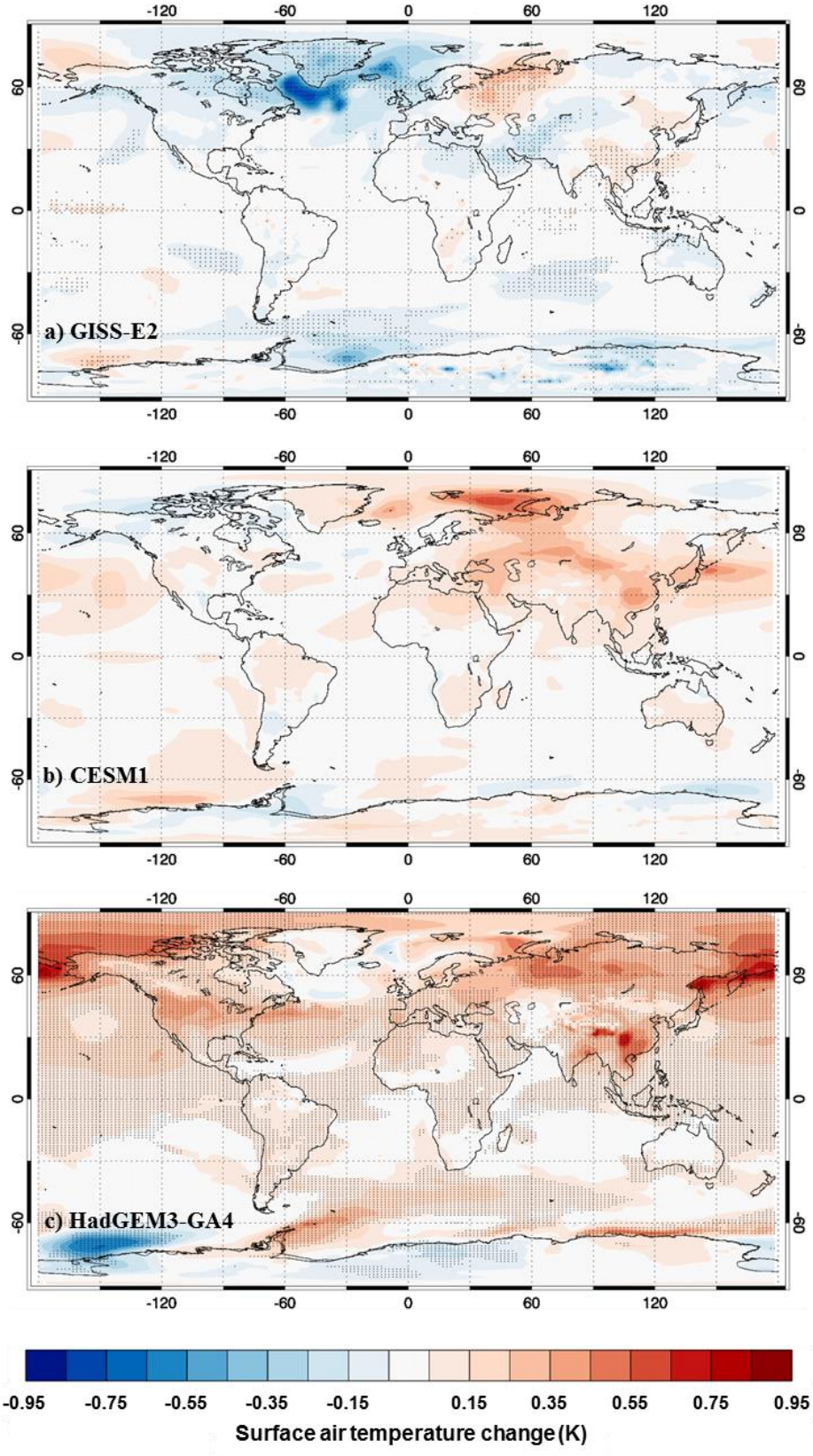


Figure 2: Global changes in surface air temperature due to removing anthropogenic SO_2 emissions from China for a) GISS-E2, b) CESM1, and c) HadGEM3. Differences are for 150-

year annual means of perturbation simulation minus control simulation. Stippling for GISS-E2 and HadGEM3-GA4 indicates changes exceeded two standard deviations for that grid box.

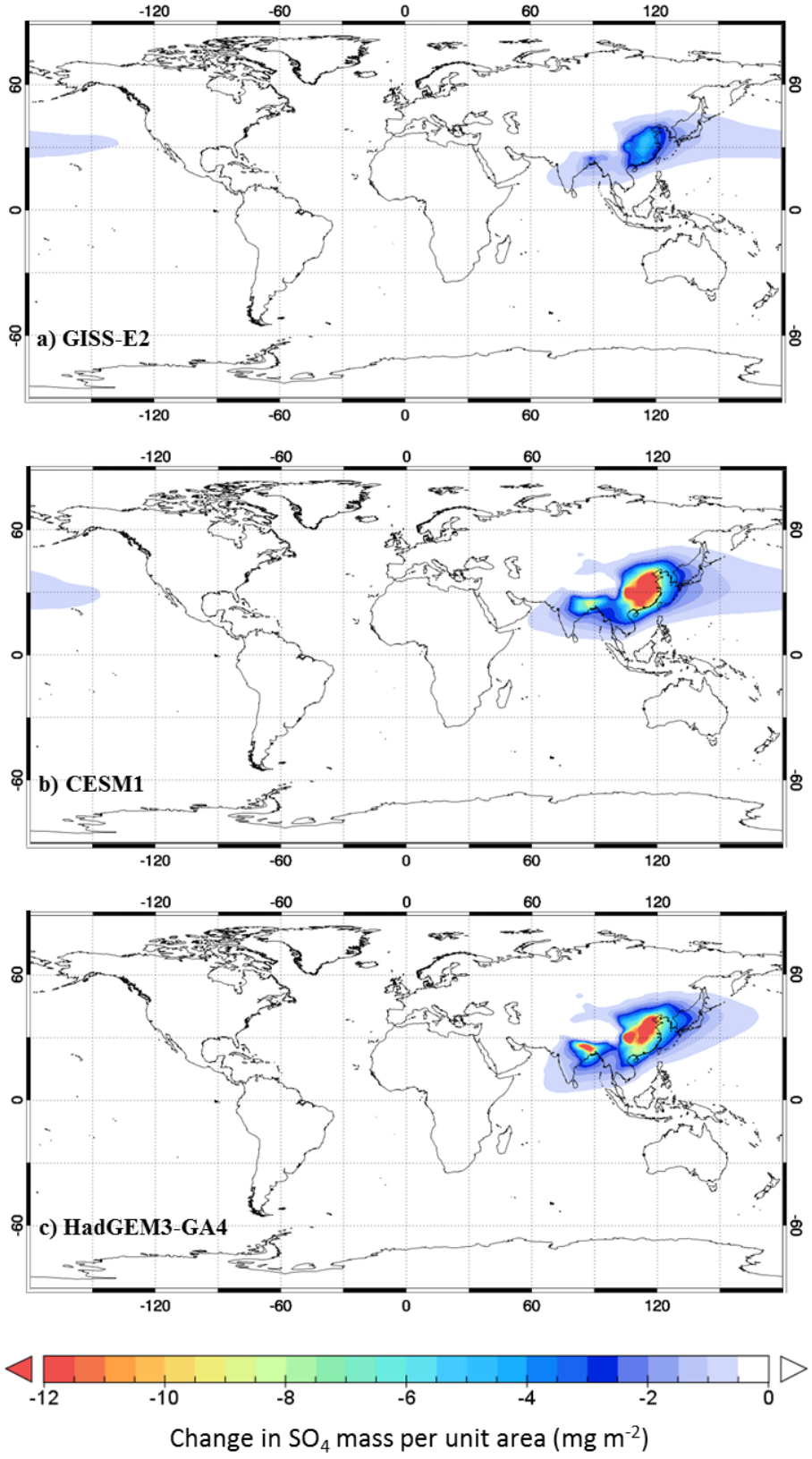


Figure 3: Global changes in column-integrated SO_4 burden due to removing anthropogenic SO_2 emissions from China, for a) GISS-E2, b) CESM1, and c) HadGEM3-GA4. Differences are calculated as perturbation simulation~~run~~ minus control simulation~~run~~, averaged over 150 years.

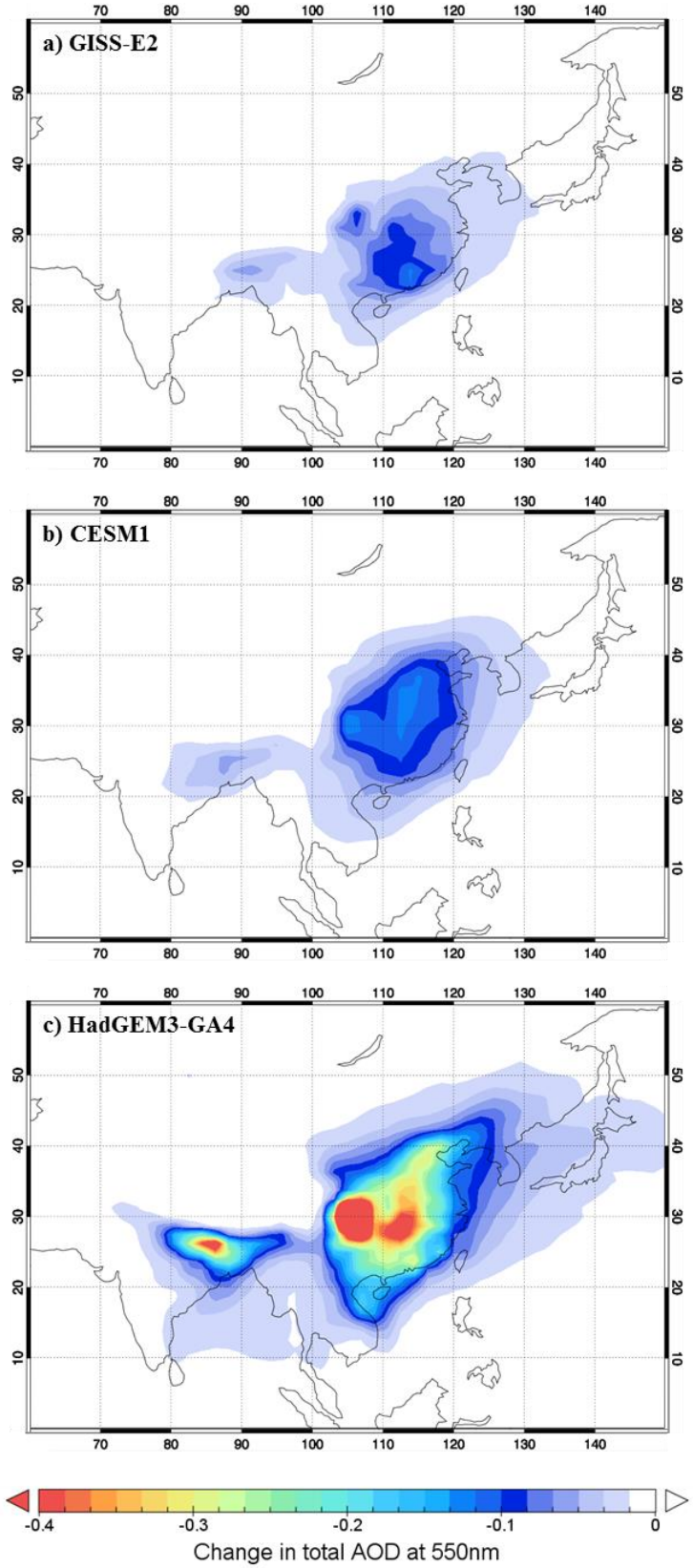


Figure 4: Change in AOD at 550nm due to removing SO_2 emissions from China for a) GISS-E2, b) CESM1, and c) HadGEM3-GA4. For HadGEM3-GA4 and GISS-E2, AOD is calculated for clear-sky conditions, whereas for CESM1 AOD is calculated for all-sky conditions, which

will generally result in higher values within each simulation. Differences are calculated as perturbation run minus control run, averaged over 150 years. The plot region focuses on Asia as changes outside of this domain were minimal.