“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 4^{th} 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 2^{nd} 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