Response to referee 1. comments.

We would like to thank the referee for carefully reading our paper and for the helpful comments and suggestions. We have modified the manuscript according to these suggestions, and detailed answers to each comment are listed below. The reviewer comments are in italic and our answers are in normal font. In the modified manuscript the changes are shown in red font. Revised manuscript can be found from supplements of this post.

Specific comments:

Comment 1.

The discrepancies in projected climate features among current climate models have recently been related to the differences in representing the processes of aerosol and aerosol-cloud interaction in these models. This study addresses this issue by investigating whether arbitrarily eliminating the differences in models' aerosol forcing strength and distribution could limit the above-mentioned discrepancies. For such a purpose, the authors have designed two sets of equilibrium-climate simulations: firstly to use two climate models (NorESM and ECHAM6) driven by their own slab-ocean modules while masked with the same prescribed direct and first-indirect radiative effects of aerosols (MACv2-SP), and secondly to force one of these two models, NorESM to adopt derived aerosol forcing field from the other model, ECHAM6. Certainly, the results are not entirely a surprise that the two models even with largely the same aerosol foring distribution and strength would still produce different climate responses particularly over regional scale, for example as reflected from modeled precipitation. The comparison involving the two sets of runs lead the authors to a conclusion that the discrepancy between the two models appears to be largely resulted from the differences in model components beyond that of aerosol. The model simulations were designed straightforwardly and supportive to addressing the science issue of the study. The paper is well organized, and the result is clearly presented. The content of the paper is perfect for the readers of ACP. Nevertheless, there are a few issues the authors should adequately address before the paper could be accepted for publication.

The authors have drawn one of their major conclusions that "further improvements in the model aerosol descriptions can be expected to have limited value in improving our understanding. . .". Such a statement does not make any logical sense based on the results of the paper. Firstly, simply making any two models to have a nearly exact radiative effects of aerosols does not necessarily mean that they both had already been equipped with an improved representation of aerosol and aerosol-cloud interaction. Furthermore, we perhaps all agree that such representations in our current climate models are far from being ideal and in fact, even unable to correctly simulate some of the key physical processes. Therefore, no one could predict the outcome in terms of modeled climate features should an ideal aerosol module be eventually produced and included. Secondly, per the current modeling efforts in this study, the applied constraint of aerosol forcing does not even include that on cloud

response to aerosol perturbation through precipitation and other critical cloud features – as indicated by the authors, not mentioning that on aerosol resuspension through activation-dissolution-evaporation. Even putting aside these comments relating to rather specific processes, giving the well-known status of our current climate models, logically and realistically, the same conclusion made by the authors to the improvement of aerosols could be applied to any other major model components or aspects. Therefore, the above-commented statement, especially presented as a major one for the paper, is not logically meaningful and adequate, in addition, it does not accurately reflect the nature and science meaning of this very study.

Author response:
We agree with the reviewer that our conclusion was perhaps overstated, and we fully agree with the reviewer that aerosol descriptions in current models are far from being ideal. The root of our original argument stemmed from the fact that models with same simplified aerosol (or forcing) descriptions (ECHAM6.1 and NorESM1) show no less regional variability in their climate responses that models with more complex (albeit far from complete) intrinsic aerosols descriptions (Samset models). However, it is true that it does not make sense for us to argue that aerosol descriptions would not matter. Nevertheless, even if we would have perfect aerosol descriptions inside the global climate models, uncertainty arising from the differences in circulation responses between the models would likely still result in a significant uncertainty in regional climate responses. For this reason, we have changed our conclusion in the abstract accordingly.

Deleted:
"Hence, further improvements in the model aerosol descriptions can be expected to have a limited value in improving our understanding of regional aerosol climate impacts, unless the dynamical cores of the climate models are improved as well."

Added:
"Hence, even if we would have perfect aerosol descriptions inside current global climate models, uncertainty arising from the differences in circulation responses between the models would likely still result in a significant uncertainty in regional climate responses"

Comment 2.
In order to make a statement as strong as "differences in aerosol descriptions among different models are not the main cause of variation in the regional distributions of climate response among different models", one needs to compare the results produced by the versions of the two adopted models with their own intrinsic aerosol module without arbitrary constraints on forcing. Such a comparison would serve as a good reference to evaluate the real effect by eliminating aerosol forcing discrepancies.

Author response:
We agree that this was a too strong statement. In this study, we have not explored regional differences among the same model with different aerosol descriptions, as pointed out by the referee.

Change in the manuscript:

Deleted:
"This implies that differences in aerosol descriptions among different models are not the main cause of variation in the regional distributions of climate responses among different models. Rather, differences in model intrinsic dynamic responses appear to dominate the differences in regional climate responses."

Added:
"The lack of improvement in the correlation coefficients suggests that differences in aerosol descriptions are not the only cause of regional differences in climate signals between the models. Rather, the differences in model circulation responses appear to dominate the differences in regional climate responses."

Comment 3.
The use of the term "aerosol-cloud interaction" seems quite casual in certain places. Giving the nature of this study that dealing primarily with direct radiative effect alongside the so-called first indirect effect of aerosol, when discussing the context of the study itself, the authors should stay closely within the proper scope of their topic.

Author Response:
We agree that the use of the term "aerosol-cloud interaction" was too vague particularly in the abstract. First, we have modified the abstract to be more accurate and state specifically only the first indirect cloud effect. Later in the text this term is only used to summarize previous research.

Change in the manuscript:

Deleted:
"We carry out experiments of equilibrium climate response to modern day anthropogenic aerosols using an identical representation of anthropogenic aerosol optical properties and aerosol-cloud interactions, MACv2-SP, in two independent climate models (NorESM and ECHAM6)"

Added:
"We carry out experiments of equilibrium climate response to modern day anthropogenic aerosols using an identical representation of anthropogenic aerosol optical properties and the first indirect effect of aerosols, MACv2-SP, in two independent climate models (NorESM and ECHAM6)."

Comment 4.
The authors borrowed the results presented in Samset et al. (2018) in their discussions. It does tell us that Samset et al. indeed derived a much larger discrepancy among models with intrinsic aerosol scheme. On the other hand, one needs to realize

that Samset et al. did not include the same models that are adopted in this study, and the simulation design in that work (with fully-coupled ocean models and most importantly, based on preindustrial era only with perturbations adopted from current climate) are quite different. The performance of climate models with fully coupled ocean component would be different than that of the models with slab ocean module, e.g., likely occurring over high latitudes as discussed in many previous works. The authors should discuss the limitations of such a comparison.

Author response:
We agree that our comparison with Samset (2017) dataset has its limitations. However, we would like to point out the Samset et al. study for aerosol reduction is not based on pre-industrial era, but carried out at climate that has warmed by 1.5 K from pre-industrial due to elevated CO_2. Samset et.al have also included the same NorESM1 model although with a different ocean description. They have used fully-coupled ocean models whereas we have slab ocean, and it is known that ocean can play a key role in model discrepancies. It is noteworthy, however, that the spatial correlation coefficients do not differ much between any set of models (Samset et al. models and our models). The role of oceans is now discussed in more detail in the revised MS:

In page 12 we now write:
"The fully coupled ocean models in the Samset et al. (2018) dataset also feature long-term internal variability in the ocean states that adds to the level of natural variation with respect to our models with simpler slab ocean representations used in this paper. Therefore, we would expect the Samset et al. data to include more noise than our results with slab ocean configurations. Furthermore, it is important to note that differences in the ocean descriptions are known to have a large impact in the regional climate responses between different models (Deser et al.; Kay et al. (2016)). Overall, we would expect that due to these differences the climate signals obtained from fully coupled models would intrinsically correlate less well with each other than those from models with slab ocean configurations. Somewhat surprisingly, this turns out not to be the case."

Also we note that we discussed the role of ocean in page 13:
"However, it should be noted that the ITCZ shift is also sensitive to the type of ocean model used, and slab ocean models tend to exaggerate the change in ITCZ (Kay et al., 2016). "

Comment 5.
Pg. 4, Ln 2: remove "of" before "are based on".

Author response:
Done

Comment 6.
Pg. 4, Ln 5" "between aerosol optical depth and CDNC: : :" this seems implying that the modeled AOD rather than aerosol concentration is the primary input for applying MAC-SP constraint? Or, in fact CDNC itself has been prescribed based on MODIS

AOD independent to the model predicted aerosol properties?

Author response:
It is true that this sentence is ambiguous. To be more precise, we have modified the text accordingly:

Change in manuscript:
Deleted:
"The relation between aerosol optical depth and CDNC is derived from MODIS data."

Added:
"The cloud albedo effect in MACv2-SP is parameterized by modifying the model-intrinsic natural cloud droplet number concentration (CDNC) via a relation based on the total change in AOD. This parametrization is derived using MODIS data."

Comment 7.
Section 2.3: The description of NorESM-EF is not very clear. When masking the aerosol forcings of NorESM using ECHAM6 derived values, how did the cloud fields produced in NorESM be considered, for instance, what to do with non-zero first-indirect effect from ECHAM6 in a no-cloud grid in NorESM, or, how to mask direct forcing into cloud fields in NorESM? Could these details be the reason behind the discussed difference between NorESM-EF and ECHAM6-MACSP?

Author response:
We have now added a section to the appendix of this study to explain more detailed how NorESM-EF run was made, and refer to this section in the main text.

Change in manuscript:

Added into appendix :
The NorESM1-EF run employed radiative forcing extracted from the ECHAM6-MACSP run. First, multi-year monthly means of MACv2-SP aerosol radiative forcing (for TOA and surface radiative fluxes and atmospheric heating rates) were computed for ECHAM6-MACSP. Second, these values were interpolated to the NorESM horizontal and vertical grid and normalized by the monthly-mean incoming solar radiation at model top. Third, during the NorESM1-EF run, these normalized forcings were multiplied by the TOA incoming solar radiation at each radiation time step, and they were added to the radiative fluxes and heating rates computed without MACv2-SP aerosols.

This treatment ensures that the diurnal cycle of the aerosol forcing is approximately correct; in particular there is no aerosol forcing during the night. However, the computed forcing is independent of the clouds simulated by NorESM1. Thus, while the aerosol radiative forcing is computed correctly in a monthly-mean sense, its sub-monthly correlation with clouds is ignored. In principle, this could impact the differences between NorESM1-EF and ECHAM6-MACSP. The impact is, however, most likely small. If neglecting the sub-monthly correlation between clouds and aerosol forcing were to have a substantial impact on the

climate response to MACv2-SP aerosols, this should also show up in the differences between NorESM1-EF and NorESM1-MACSP. Yet the differences between NorESM1-EF and NorESM1-MACSP are very small (Tables 2 and A1), in fact much smaller than the corresponding differences between ECHAM6-MACSP and either NorESM1-EF or NorESM1-MACSP. This strongly suggests that the differences between NorESM1-EF and ECHAM6-MACSP are primarily caused by the use of a different climate model rather than by the subtle differences in radiative forcing.

Comment 8.

Pg. 12, Line 28: "identical anthropogenic aerosol representations in the models" is inaccurate

Author response:
This is true as the applied aerosol description results in a different total radiative forcing. We have modified the sentence mentioned in this comment so that it is unambiguous.

The sentence is now changed in the manuscript:
"We have here provided the first results on the equilibrium climate response of modern day anthropogenic aerosols using two different climate models, ECHAM6 and NorESM1, with the MACv2-SP (Stevens et al., 2017) anthropogenic aerosol representations."

Comment 9
Pg. 14, Ln 1: please correct "essentially equally".

Author response:
The corrected text now says "nearly as".

Change in manuscript:
However, the correlation coefficients for regional distributions of climate responses, averaged over equal run length, were nearly as good among our experiments with prescribed aerosols and slab ocean representation (0.78 for temperature and 0.41 for precipitation) and among Samset et al. experiments with model-intrinsic aerosols and the fully coupled ocean representation (0.79 for temperature and 0.34 for precipitation).

Comment 10.
Figure 1, 2, &4: the results of NorESM-EF run should be presented.

Author response:
This is done as suggested. We have included NorESM-EF results for radiative forcing, temperature and precipitation responses, and comparisons with NorESM and ECHAM6 responses.

Change in manuscript:
ECHAM-MACSP and NorESM1-EF difference is added to figure 1.

Figures 2 and 4 NorESM1-MACSP and NorESM1-EF difference are added to show the temperature and precipitation responses for NorESM-EF compared to NorESM1-MACSP.

Response to referee 2. comments.

We would like to thank the referee for a detailed analysis of our paper. Here we answer to all comments made by referee 2. Importantly, we have changed the term "dynamical response" and fixed typos pointed out by reviewer. Below is a list of our detailed answers to all comments as well as descriptions of the modifications made to the manuscript. In the modified manuscript we have marked all changes with red color.

Comment 1.
This paper considers the response of two different climate models to the addition of anthropogenic aerosols. The aerosols are specified in exactly the same form in the two models. The paper argues that whilst the global average temperature and precipitation responses are quite consistent between the two models, there are major differences between the regional responses. The conclusion is that it is the intrinsic differences in the dynamics of the circulation between the two models that determines the differences in the regional responses. This conclusion is supported by additional evidence. The first evidence is from the results of adding the aerosol radiative forcing evaluated in one model into the second – which gives a response more similar to that to the aerosol added to the second model than to that to the aerosol added to the first. The second evidence is from the response to aerosol in previously reported model experiments which show good spatial correlation, at least in temperature, to the responses in the new experiments reported here.

I see this as an interesting study, which usefully adds to the body of recent work emphasising the importance of the circulation response in determining regional climate change (and also in determining the geographical variation of internal variability that may dominate any clear climate change signal in the short term).

The first referee has already made some comments, which seem reasonable to me, about the general conclusions of this paper – e.g. whether a message of 'the usefulness of research on aerosol representation in models is fundamentally limited until we are more certain about circulation response' is a bit too sweeping.

My own comments are as follows:
General comment: Can you provide any information on the typical geographical distribution of differences in response between the two models you consider for other forcings? Perhaps other results from other experiments with these two models are available (either published or unpublished). Also recent papers on circulation response such as Shepherd (2014, Figure 4) have tended to show differences in winds rather than temperatures. To put your results in context it would therefore be interesting to see differences in, say, 850hPa winds. Also there tends to have been an emphasis on differences in the North Atlantic region. You are showing significant temperature differences in the North Pacific – are you aware of other work that has showed up differences in circulation response between models in that region?

Author response:
The reviewer asked for information about the typical geographical distribution of differences in response to different climate forcers between the two models considered in this study. Unfortunately we were not able to find such data. It would have been interesting to compare, for example, responses to heterogeneous aerosol forcing and homogenous greenhouse gas forcing, as done by Shindell et al. (2015). However, as already mentioned, the Shindell et al. paper does not include NorESM1 or ECHAM6 models. However, we note that in future we plan to quantify circulation response differences for a set of different climate forcers within a larger group of models, using PDRMIP data.

As suggested by the reviewer, we now also show also 850hPa level wind responses that were discussed by Shepherd (2014, Figure 4) and also recently by Li et al. (2018) (the references are added to the revised manuscript). A figure showing the 850hPa winter wind responses in the two models is now included in the appendix (Shephard (2014) and Li et al. (2018) also discussed wintertime wind responses). We added a short text about the wind response into the manuscript, noting that our results resemble those from the literature.

Change in manuscript:
Added to page 14:
"The lack of improvement in the correlation coefficients suggests that differences in aerosol descriptions are not the only cause of regional differences in climate signals between the models. Rather, the differences in model circulation responses appear to dominate the differences in regional climate responses. Figure C5 shows the average 850 hPa wind responses for ECHAM6-MACSP and NorESM1-MACSP experiments during for Northern hemisphere winter. The responses in the circulation fields vary significantly between the two models, with an annual average correlation coefficient of only 0.18 (DJF:-0.03; MAM:0.07; JJA:0.15; SON:0.19). The lack of robustness in atmospheric circulation responses between different climate models has been previously discussed by Shepherd (2014) for CMIP5 RCP8.5 scenarios and by and by Li et al. (2018) for HAPPI 1.5 K and 2.0 K warming scenarios. Shepherd (2014) argued that the differences in circulation responses cause variation in the regional temperature and precipitation responses in future climate scenarios. Li et al. (2018) showed that model consensus for circulation response is low even for atmosphere-only models forced with same time-varying SST and sea ice, anthropogenic greenhouse gases, ozone, land use, land cover, and aerosols. Both in Shepherd (2014) and Li et al. (2018) data the NH wintertime circulation response over the North Atlantic disagrees significantly between models. Also for ECHAM6-MACSP and NorESM1-MACSP the circulation response over the North Atlantic show differences in magnitude and pattern. Differences are also seen over the North Pacific region. Combined with the difference in the sea ice and surface albedo change in the North Pacific, these circulation changes can drive the temperature response differences in the region.

Comment 2.
p1 l19: 'unless the dynamical core of the climate models are improved as well'. 'Dynamical core' is often used to mean the part dealing with the dry thermodynamics and dynamics. If that is what you intend then I think that this may be too narrow a view – I don't see why the

moist processes, including coupling between circulation and clouds, shouldn't play a role as well.

Author response:
We thank the reviewer for this insightful comment. We agree that the term dynamical response is too narrow. Therefore, we have change the term, dynamical response, to circulation response as also used in paper by Shepherd, 2014.

Change in manuscript:
Hence, even if we would have perfect aerosol descriptions inside the global climate models, uncertainty arising from the differences in circulation responses between the models would likely still result in a significant uncertainty in regional climate responses.

Comment 3.
p3 l16: 'The original …

Author response:
Fixed as suggested

Change in manuscript:
The Original ECHAM model branched from an early version of the European Centre for Medium-Range Weather Forecasts (ECMWF) model for climate studies.

Comment 4.
p3 l21: 'were run

Author response:
Fixed as suggested.

Change in manuscript:
Here, both models were run with identical fixed modern-day greenhouse gas concentrations

Comment 5.
p3 l22: 'intrinsic slab ocean configurations of the models' – again in principle this might be something that controls the different responses of the model and doesn't fit naturally under the heading of 'dynamical core'.

Author response:

This relates also to the comment 2. We have now changed the term dynamical core to circulation response. Also, we have included the oceanic heat exchange as a source for model difference (p3 l22). The role of ocean models is also discussed more in page 12.

Change in manuscript page 3 line 22:

Oceans were simulated with the intrinsic slab ocean configurations of the models. This idealization removes the effect of natural and aerosol induced variations in ocean circulation and restricts our study to the response in atmospheric circulation, oceanic heat exchange, and sea ice dynamics only.

Modified manuscript page 12:
"The fully coupled ocean models in the Samset et al. (2018) dataset also feature long-term internal variability in the ocean states that adds to the level of natural variation with respect to our models with simpler slab ocean representations used in this paper. Therefore, we would expect the Samset et al. data to include more noise than our results with slab ocean configurations. Also, it is important to note that differences in the ocean descriptions are known to have a large impact in the regional climate responses between different models (Deser et al (2016).; Kay et al. (2016)). Overall, we would expect that due to these differences the climate signal obtained from fully coupled models would intrinsically correlate less well with each other than those from models with slab ocean configurations. Somewhat surprisingly, this turns out not to be the case"

Comment 6.
p4 l2: 'properties of are' > 'properties are'?

Author response:
This typo is fixed as suggested by the referee.

Change in manuscript:
The aerosol properties are based on aerosol climatology by (Kinne et al., 2013).

Comment 7.
p4 l10: 'constructed from two identical runs' – do you mean that for each model the control run was constructed from two runs, or do you mean 'two identical runs, one for each model'?

Author response:
Our control run is constructed from two almost identical runs via small perturbations on the initial states. The purpose of this approach is to remove some of the climate natural variability by averaging these two runs.

Change in manuscript:
The sentence "The control run (CTRL) included only natural aerosols, and was constructed from two runs for each model with small initial condition perturbations." is added to page 4

Comment 8.
p4 l14: 'experiments' seems an odd word to use about taking differences between simulations

Author response:
Word "experiments" was chosen to distinguish individual runs from differences between a pair of runs.

Comment 9.
p5 l3: 'from the MACv2-SP'

Author response:
"the" is added here as suggested.

Change in manuscript:
The total radiative forcing from the MACv2-SP anthropogenic aerosol description was found to be very similar for the two models (see Fig. 1).

Comment 10.
p5 l7: The panels in Fig A1 are very small.

Author response:
We have enlarged this figure and the panels are more visible in the revised MS.

Comment 11.
p5 l18: 'nearly all the variance' – do you mean 'variance' or 'difference'?

Author response:
Here we mean variance. This is related to our sensitivity analysis where we used FAST method to decompose the variance in our modelled model difference.

Change in manuscript:
Our analysis showed that differences in cloud cover and surface albedo can explain nearly all of the variance in the difference in total instantaneous shortwave radiative forcing between ECHAM6 and NorESM1.

Comment 12.
p5 l22: 'Previous research shows that the aerosol radiative forcing can also depend on the meteorology (surface winds and precipitation) produced by the models, partly driven by the natural variability of the climate system (Baker et al., 2015; Bony et al., 2015; Shepherd, 2014).' – my reading of these papers was that they were saying that it was the response to e.g. aerosol radiative forcing, that depends on the meteorology and that the relevant aspects of the meteorology were those that were also responsible for natural variability.

Author response:
The reviewer rightly points out that these papers do not discuss the effects of meteorology to the aerosol forcing. We have change these references to Fiedler et.al (2019) where they explicitly discuss the effects of model representation of weather to aerosol forcing.

Change in the manuscript:
Previous research shows that the aerosol radiative forcing can also depend on the meteorology (surface winds and precipitation) produced by the models, partly driven by the natural variability of the climate system (Fiedler et.al, 2019).

Comment 13.
p7 l2: 'disagree the most in high latitude regions' – part of this disagreement is well to the south of 60N

Author response:
This is true, clearly the models disagree also well to the south of 60N.

Change in manuscript:
The modeled regional temperature responses between ECHAM6 and NorESM1 simulations disagree the most in mid- and high latitude regions as seen in Figure 2c.

Comment 14.
p7 l3: Are you implying that the changes in surface albedo feedback cause the differences in temperatures? What reasoning are you using for that?

Author response: We think that this may indeed be the case, since changes in surface albedo are known to amplify changes in Arctic temperatures (albedo feedback).

Change in manuscript:
The modeled regional temperature responses between ECHAM6 and NorESM1 simulations disagree the most in mid- and high latitude regions as seen in Figure 2c. In high latitude regions temperature differences are associated with surface albedo responses (snow/sea ice) between the models (see Figure A2). Changes in surface albedo are known to amplify changes in Arctic temperatures (albedo feedback). Hence, differences in snow and sea ice responses may partly explain the difference in temperature responses in the high latitudes.

Comment 15.
p7 l6: The disagreement would be 'curious' if the zonal-mean temperature response at high latitudes was locally forced. Are you confident that it is?

After consideration we decided to remove the entire sentence to which this comment refers to. Particularly, it is not clear what role the changes in cloud cover have on the responses. Also, the point that high latitude responses may not be locally forced (at least fully) is a valid one.

Comment 16.
p8 Figure 3 etc: You are choosing to quantify the change in precipitation by the percentage change. This means, for example, that in Figure 4 there is a conspicuous difference in

precipitation response over much of Australia (where the actual precipitation is very small). I see that in the Samset et al (2017) paper they choose to show change in precipitation normalised by change in temperature. Have you considered carefully whether your choice is the most effective way to show change in precipitation.

Author response:
We prefer this style of representation and are inclined to keep it as it is. The choice in Samset et al. is, we believe, based on relative large differences in global temperature responses between the models (that do not exist between the two models here).In Samset et al., scaling with temperature was carried out to make the precipitation responses comparable to each other. However, the text in Samset et al. refers to absolute percentage changes.

Comment 17.
p8 l13: 'Africa'

Author response:
fixed as suggested by the referee

Change in manuscript page 8 line 15:
Both models consistently show an overall drying of the Northern Hemisphere, with some statistically significant regional increase in precipitation over the Northwest Africa.

Comment 18.
p10 l6: I'm not convinced that the change in vertical velocity can be regarded as a cause of the change in precipitation – don't the two go together as part of an overall coupled change in circulation and precipitation

Author response:
It is true that precipitation and vertical velocity go hand in hand. Therefore, the sentence "it cannot be concluded that change in precipitation is caused by the change in vertical velocity" is added to clarify this. Here we want to say that model disagreement in precipitation response is overall related to the difference in circulation response.

Change in manuscript:
it cannot be concluded that change in precipitation is caused by the change in vertical velocity. Probably, both the changes in vertical velocity and precipitation are related to changes in circulation.

Comment 19.
p10 l19: For your comparison between NorESM1-MACSP and NorESM1-EF you shown only the zonal-mean and give spatial correlation information. To me the argument that you are trying to make, that the these two simulations closely resemble each other, would be more convincing if you also showed a limit amount of latitude-longitude information – e.g. adding to the information in Figs 2 and 4.

Author response:
Referee 1 also pointed out about the missing information on the NorESM-EF run. We have included the following figures:

Change in manuscript:
Difference between ECHAM-MACSP and NorESM1-EF added to figure 1.

Figures 2 and 4 NorESM1-MACSP and NorESM1-EF difference are added to show the temperature and precipitation responses for NorESM-EF compared to NorESM1-MACSP..

Comment 20.
p10 l24: 'dynamical responses' – again I wonder if something like 'circulation responses' might be better (implying something more complex than simply dry dynamics).

Author response:
Word "dynamical" is change to "circulation" as suggested.

Role of climate model dynamics in estimated climate responses to anthropogenic aerosols

Kalle Nordling[1], Hannele Korhonen[1], Petri Räisänen[1], Muzaffer Ege Alper[1], Petteri Uotila[2], Declan O'Donnell[1], and Joonas Merikanto[1]

[1]Finnish Meteorological Institute, Helsinki, Finland
[2]INAR/Physics, University of Helsinki, Helsinki, Finland

Correspondence: Kalle Nordling (kalle.nordling@fmi.fi)

Abstract. Significant discrepancies remain in estimates of climate impacts of anthropogenic aerosols between different general circulation models (GCMs). Here, we demonstrate that eliminating differences in model aerosol or radiative forcing fields results in close agreement in simulated globally averaged temperature and precipitation responses in the studied GCMs. However, it does not erase the differences in regional responses. ~~We carry out experiments of equilibrium climate response to modern~~
5   ~~day anthropogenic aerosols using an identical representation of anthropogenic aerosol optical properties and aerosol-cloud interactions, MACv2-SP, in two independent climate models (NorESM and ECHAM6).~~ We carry out experiments of equilibrium climate response to modern day anthropogenic aerosols using an identical representation of anthropogenic aerosol optical properties and the first indirect effect of aerosols, MACv2-SP, in two independent climate models (NorESM and ECHAM6). We find consistent global average temperature responses of 0.48(±0.02) K and 0.50(±0.03) K and precipitation responses of 1.69
10   (±0.04)% and 1.79(±0.05)% in NorESM1 and ECHAM6, respectively, compared to modern-day equilibrium climate without anthropogenic aerosols. However, significant differences remain between the two GCMs regional temperature responses around the Arctic circle and the equator and precipitation responses in the tropics. The scatter in the simulated globally averaged responses is small in magnitude when compared against literature data from modern GCMs using model intrinsic aerosols but same aerosol emissions −(0.5–1.1) K and (1.5–3.1)% for temperature and precipitation, respectively). The Pearson cor-
15   relation of regional temperature (precipitation) response in these literature model experiments with intrinsic aerosols is 0.79 (0.34). The corresponding correlation coefficients for NorESM1 and ECHAM6 runs with identical aerosols are 0.78 (0.41). The lack of improvement in correlation coefficients between models with identical aerosols and models with intrinsic aerosols implies that the spatial distribution of regional climate responses is not improved via homogenizing the aerosol descriptions in the models. Rather, differences in the atmospheric dynamic ~~and high-latitude cloud~~ and snow/sea ice cover responses dominate
20   the differences in regional climate responses.

~~Hence, further improvements in the model aerosol descriptions can be expected to have a limited value in improving our understanding of regional aerosol climate impacts, unless the dynamical cores of the climate models are improved as well.~~ Hence, even if we would have perfect aerosol descriptions inside the global climate models, uncertainty arising from the differences in circulation responses between the models would likely still result in a significant uncertainty in regional climate
25   responses

1 Introduction

Making reliable predictions on future changes in regional climates is crucial for estimating how climate change will impact people and societies (Hawkins et al., 2016), but there are still large uncertainties related to climate change predictions on
5   regional scales (Giorgi and Francisco, 2000; Feser et al., 2011). Anthropogenic aerosol particles can be an important driver for regional climate change due to the near-instantaneous response of local aerosol concentrations to changes in emissions, their direct radiative properties, and their ability to modify cloud microphysical processes. However, reliable implementation of aerosol effects into global climate models has been challenging. Several aerosol processes are still not well-understood (Boucher et al., 2013), and there exists an enormous scale difference between the microphysical processes and the resolution
10   of global scale models (Carslaw et al., 2013).

Varying descriptions of aerosols and aerosol-cloud interactions cause a wide spread in aerosol radiative forcing and climate impacts between different GCMs (Wilcox et al., 2015). Shindell et al. (2015) compared historical CMIP5 runs with and without anthropogenic forcing from aerosols, ozone, and land use. The forcing showed a very large spatial variation with globally averaged values that ranged between $0.15\ \mathrm{Wm^{-2}}$ and $-1.44\ \mathrm{Wm^{-2}}$ (the aerosol contribution being between $-0.29\ \mathrm{Wm^{-2}}$
15   and $-1.44\ \mathrm{Wm^{-2}}$). The combined changes in aerosol, ozone and land use produced globally averaged transient temperature responses between $0.00\ \mathrm{K}$ and $-1.33\ \mathrm{K}$ over the twentieth century, with the spatial pattern of the temperature response varying significantly between the models. Overall, the inclusion of aerosols in CMIP5 models nevertheless improved the historical temperature trends compared to observations. This applied particularly to models including sophisticated parameterizations for aerosol cloud droplet activation (Ekman, 2014).

20   Besides reducing the global temperature, anthropogenic aerosols are also known to reduce global precipitation (Ramanathan, 2005) and to significantly modify the Asian monsoon (Bollasina et al., 2011; Salzmann et al., 2014) . Wang (2015) demonstrated that among CMIP5 models the changes in anthropogenic aerosols dominated the total precipitation changes from the pre-industrial era to the present day. Most of this change was caused by the remote impact of aerosols rather than by direct effects on local cloud processes, and cloud optical depth in all but heavily aerosol-loaded regions, such as in the Indian monsoon
25   region. Also for precipitation changes, an improved representation of aerosol-cloud interactions was found to be the key factor in reproducing consistent distributions of past precipitation change.

Improvements in model aerosol descriptions have not succeeded to remove the large uncertainty in aerosol climate effects. After CMIP5, the most representative multi-model results on aerosol climate impacts have been provided by Samset et al. (2018). They compared the equilibrium climate responses for complete removals of model intrinsic anthropogenic aerosols
30   among four state-of-art fully coupled climate models, with aerosol emissions from CMIP5 (Lamarque et al., 2010). In their study, removing the aerosols produced global-mean temperature increases between 0.5 and $1.1\ \mathrm{K}$ and precipitation increases between 1.5% and 3.1%. In another recent study, Kasoar et al. (2016) reduced anthropogenic SO_2 emissions from China in three independent climate models. There, identical emission reductions lead to simulated changes in aerosol optical depth and

shortwave radiative flux over China that varied by up to a factor of 6 between the models. The three models also exhibited large differences in their global and regional temperature responses. However, it is unclear to which degree the existing spread in aerosol climate impacts among current climate models results from differences in modeled aerosols or from differences in model dynamical responses to aerosols. Only standardized aerosol perturbations across different models can entangle these sources of uncertainties in aerosol climate effects (Stier et al., 2013).

Here, we explore how robust the aerosol climate response would be in modern GCMs if the anthropogenic aerosols and their cloud interactions could be modeled exactly. To assess this question we carry out long equilibrium climate simulations with fixed greenhouse gas concentrations and prescribed aerosol fields using the MACv2-SP aerosol description (Stevens et al., 2017) in two modern GCMs, NorESM1 and ECHAM6. The MACv2-SP is partly based on observational data and provides a simple representation of global aerosol optical properties. It also includes a simple empirical fit for aerosol-cloud-albedo effects. These experiments allow us to single out the contribution of climate model dynamics to the intermodel differences in the response to anthropogenic aerosols. We will compare our results against the dataset by Samset et al. (2018) to investigate the robustness of global and regional climate responses in modern climate models using interactive or prescribed aerosols.

2 Methods

2.1 Applied climate models and set-up

We carry out modern day equilibrium climate simulations with two independent climate models, ECHAM6.1 and NorESM1. ECHAM6.1 (Stevens et al., 2013) is the sixth generation of ECHAM general circulation model developed in Max Planck Institute with 47 sigma hybrid vertical levels, with the model top at 0.01 hPa and a horizontal resolution of $1.9° \times 1.9°$. The original ECHAM model branched from an early version of the European Centre for Medium-Range Weather Forecasts (ECMWF) model for climate studies. NorESM1 is the Norwegian Earth system model with 26 sigma hybrid vertical levels (the highest model level at 2.9 hPa) and $1.9° \times 2.5°$ horizontal resolution (Bentsen et al., 2013; Iversen et al., 2013; Kirkevåg et al., 2013). NorESM1 is based on the CCSM4 model operated at NCAR. Thus, the two models applied in our study do not share a common development history. Here, both models were ~~ran~~ run with identical fixed modern-day greenhouse gas concentrations. ~~Oceans were simulated with the intrinsic slab ocean configurations of the models. This idealization removes the effect of natural and aerosol induced variations in ocean dynamics and restricts our study to the response in atmosphere/sea ice dynamics only.~~ Oceans were simulated with the intrinsic slab ocean configurations of the models. This idealization removes the effect of natural and aerosol induced variations in ocean circulation and restricts our study to the response in atmospheric circulation, oceanic heat exchange, and sea ice dynamics only.

2.2 Standardized aerosol representation

MACv2-SP is a standardized representation of anthropogenic aerosol radiative effects, accounting for the direct radiative as well as the cloud albedo effect of anthropogenic aerosol (Stevens et al., 2017). However, the cloud lifetime effect is not taken

3

into account. Anthropogenic aerosols are represented by nine 3D time-varying Gaussian plumes defining the aerosol optical depth, single scattering albedo and asymmetry parameter. Four of these plumes represent aerosol emissions from biomass burning and the other five are associated with industrial emissions. The industrial plumes originate from Europe, North America, East Asia, South Asia and Australia and the biomass plumes from North Africa, South America, South central Africa and

5 Maritime Continent (Fig. 1 and Table 1 in Stevens et al. (2017)). The plumes differ in their annual cycle and optical properties, and have a realistic horizontal and vertical structure that represents the transports of aerosols with prevailing winds. The aerosol properties ~~of~~ are based on aerosol climatology by (Kinne et al., 2013), derived from ground-based sun-photometer networks (AERONET) merged onto background maps from global models participating in the Aerosol Model Intercomparison Project (AeroCom). ~~The cloud albedo effect is parameterized by modifying the model-intrinsic natural cloud droplet number~~

10 ~~concentration (CDNC). The relation between aerosol optical depth and CDNC is using from MODIS data~~ The cloud albedo effect in MACv2-SP is parameterized by modifying the model-intrinsic natural cloud droplet number concentration (CDNC) via a relation based on the total change in AOD. This parametrization is derived from MODIS data. MACv2-SP allows for a simple and observation-based representation of the changes in aerosol optical properties and cloud droplet number concentrations due to anthropogenic aerosols.

15 ## 2.3   Model experiments and analysis

Sets of 100-year equilibrium climate runs for the year 2005 were conducted with both models, with the last 60 years used for the analysis: (1) ~~The control run (CTRL) included only natural aerosols, and was constructed from two identical runs with small initial condition perturbations.~~ The control run (CTRL) included only natural aerosols, and was constructed from two runs for each model with small initial condition perturbations.; (2) The MACSP run included both natural and anthropogenic

20 aerosols for the year 2005. In addition, for NorESM1, a third run EF was carried out. This run employed the time-varying 3D aerosol radiative forcing field computed from the ECHAM6's MACSP run. A more detailed description of the implementation is given in the appendix A. A summary of the runs is given in Table 1.

Table 1. Summary of the performed model runs

Runs	Forcing	Models
CTRL	natural aerosols	ECHAM6, NorESM1
MACSP	MACv2-SP + natural aerosols	ECHAM6, NorESM1
EF	Forcing field from ECHAM6	NorESM1

Based on these runs, the following three experiments were defined to estimate the effect of anthropogenic aerosols: ECHAM6-MACSP (the difference between the MACSP and CTRL runs for ECHAM6), NorESM1-MACSP (MACSP minus CTRL for

25 NorESM1), and NorESM1-EF (EF minus CTRL for NorESM1). The analysis of the results was based on monthly-mean values of data, and focused on the effects of MACv2-SP aerosols on near-surface temperature, precipitation, surface albedo and total cloud cover. The statistical significance of the responses was evaluated using Student's t-test with an auto-correlation

4

correction according to Zwiers and von Storch (1995). The response uncertainties in global mean values were estimated by the standard error of means taking into account lag-1 auto-correlation according to Zwiers and von Storch (1995). The instantaneous radiative forcing was calculated using double radiation calls with and without MACv2-SP aerosols during the slab ocean runs.

3 Results

3.1 Aerosol radiative forcing

The total radiative forcing from the MACv2-SP anthropogenic aerosol description was found to be very similar for the two models (see Fig. 1). For ECHAM6, the MACv2-SP aerosol scheme produces a $-0.64\,\mathrm{Wm^{-2}}$ global average total shortwave radiative forcing at the top of the atmosphere (TOA) for the year 2005, with $-0.35\,\mathrm{Wm^{-2}}$ arising from direct and $-0.29\,\mathrm{Wm^{-2}}$ from indirect radiative forcing. For NorESM1, the same aerosol scheme produces a slightly higher global radiative forcing of $-0.69\,\mathrm{Wm^{-2}}$ at TOA, with $-0.36\,\mathrm{Wm^{-2}}$ direct and $-0.33\,\mathrm{Wm^{-2}}$ indirect radiative forcing. Figure C1 shows the maps of aerosol direct and indirect radiative forcings in the two models as calculated here. The largest difference in the total forcing was found over South East Asia up to $3.20\,\mathrm{Wm^{-2}}$, where also the largest absolute forcing was found in both models. Fiedler et al. (2019) have calculated both the MACv2-SP effective radiative forcing as well as the instantaneous radiative forcing using double radiation calls with fixed sea surface temperature for the two climate models used here. They showed that with fixed sea surface temperature the MACv2-SP aerosols produce an instantaneous radiative forcing of $-0.60\,\mathrm{Wm^{-2}}$ and $-0.68\,\mathrm{Wm^{-2}}$ in ECHAM6 and NorESM1, respectively. The correlation coefficient for the regional total forcing in the two models due to MACv2-SP is 0.97, and 0.90 for direct and 0.89 indirect forcings only. Thus, the regional differences in direct and indirect forcing somewhat compensate for each other.

We used a Gaussian process emulation technique (O'Hagan, 2006) to assess the causes for the regional differences in aerosol radiative forcing (see Appendix B for details). Our analysis showed that differences in cloud cover and surface albedo can explain nearly all of the variance in the difference in total instantaneous shortwave radiative forcing between ECHAM6 and NorESM1. Our sensitivity analysis reveals that in the regions with the largest radiative forcing (close to the center of the MACv2-SP plumes) the difference in model cloud cover dominates the difference in model shortwave forcing. Vice versa, in regions with low aerosol radiative forcing the differences in surface albedo dominates the differences in forcing. We note that these results apply only to fixed aerosol fields produced by the MACv2-SP representation. Previous research shows that the aerosol radiative forcing can also depend on the meteorology (surface winds and precipitation) produced by the models, partly driven by the natural variability of the climate system (Fiedler et al., 2019).
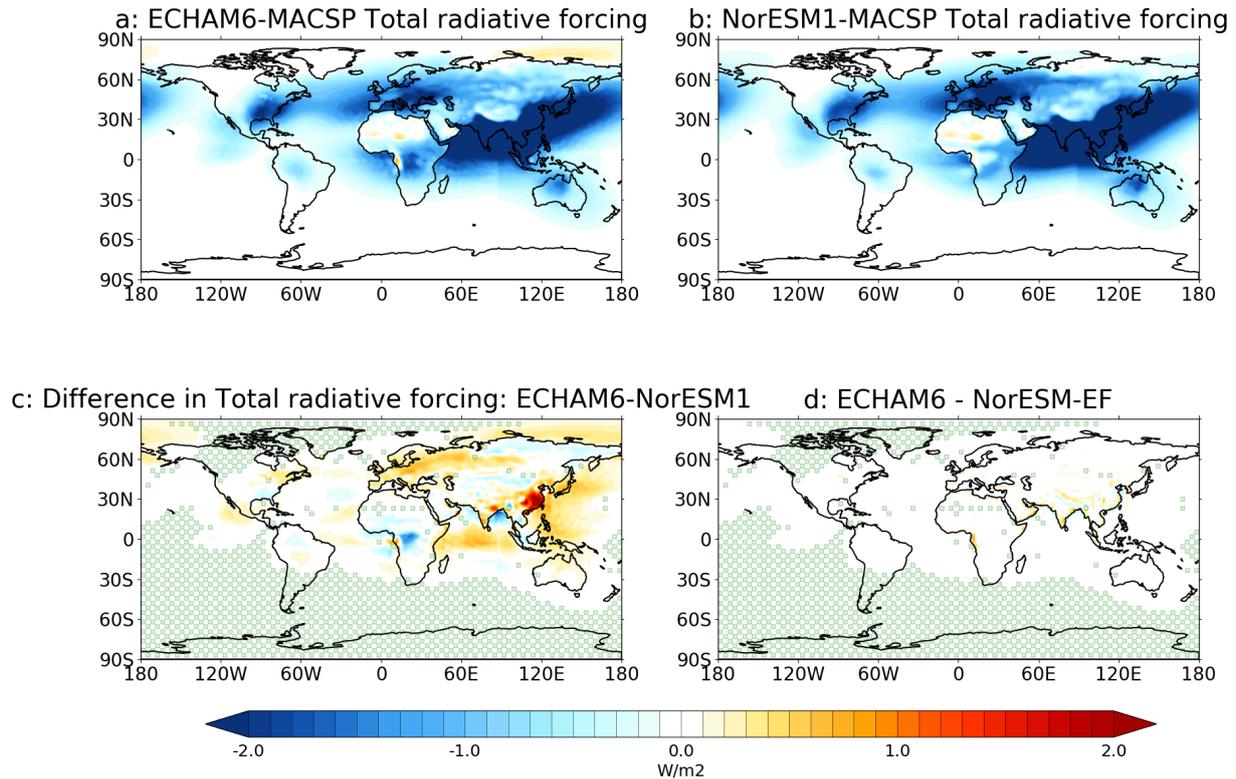
Figure 1. The total radiative forcing at top of the atmosphere produced by MACv2-SP aerosols. The top left figure shows the forcing in ECHAM6-MACSP experiment and the top right figure in NorESM1-MACSP experiment. Below is shown the difference in the forcing between the two models and difference between ECHAM-MACSP and NorESM1-EF runs. Small green circles mask the areas where results are not statistically significant at the p level < 0.05.

3.2 Climate response to the addition of anthropogenic aerosols

3.2.1 Temperature

We obtain a robust global temperature response of -0.5 K due to the inclusion of MACv2-SP anthropogenic aerosols in both models. For ECHAM6-MACSP experiment the global mean near-surface temperature response is $-0.50(\pm0.03)$ K, with regional values ranging from $+0.30$ K to -2.10 K. For NorESM1-MACSP experiment the global mean value is $-0.48(\pm0.02)$ K and the regional values range between $+0.39$ K and -2.28 K.

Figure 2 shows the regional temperature response to the inclusion of anthropogenic MACv2-SP aerosols. The spatial correlation between ECHAM6-MACSP and NorESM1-MACSP experiments is 0.81 for full experiments with 60+120 years of MACSP and CTRL runs in both models. Largest cooling in ECHAM6 is located in Southeast Asia whereas in NorESM1 the largest cooling is found near the Russian Far East and north of Japan, with a second minimum over the Greenland sea.
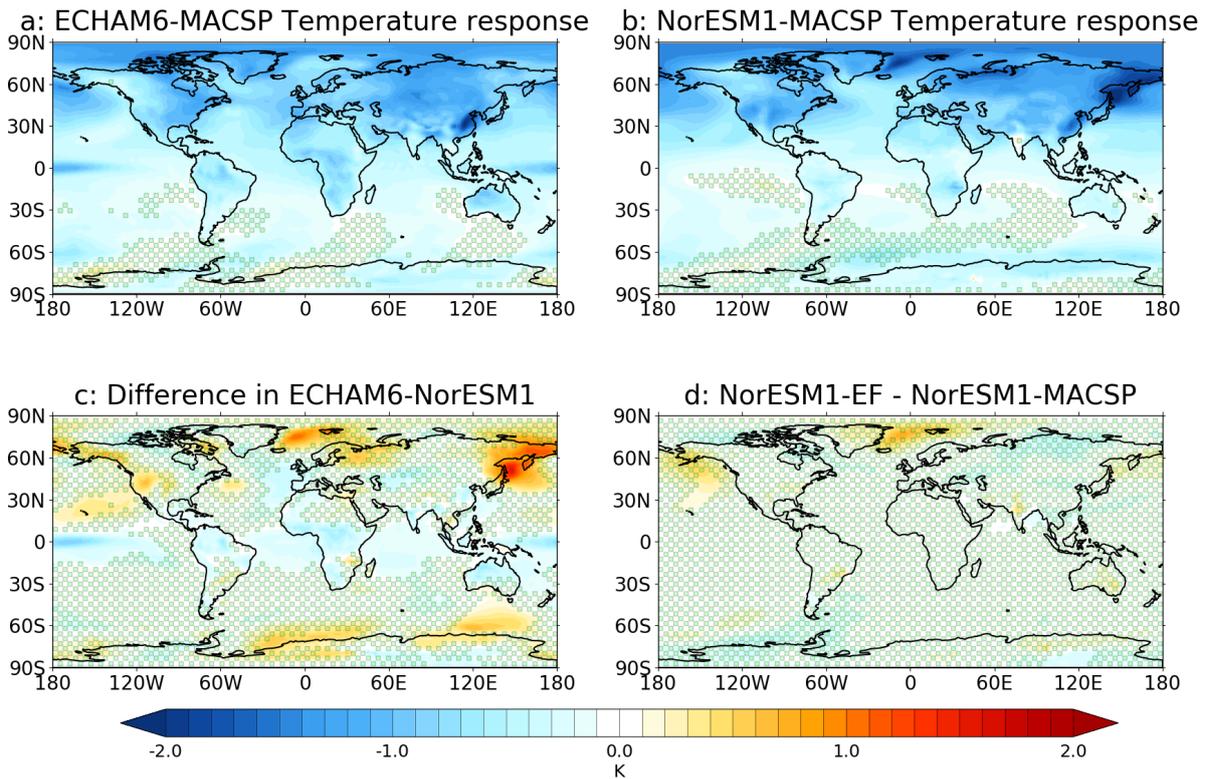
Figure 2. Near surface temperature response to the addition of anthropogenic (MACv2-SP) aerosols. The top left figure shows the response for ECHAM6-MACSP experiment and the top right figure for NorESM1-MACSP experiment. Below is shown the difference in the responses between the two models and difference between NorESM1-MACSP and NorESM1-EF. Small green circles mask the areas where results are not statistically significant at the p level < 0.05.

Small positive temperature responses are found close to the Antarctic coast in both models, but these temperature responses are not statistically significant and are related to natural variations in sea ice. We found some significant correlation between the regional aerosol forcing and regional temperature response in both models: 0.39 in ECHAM6 and 0.29 in NorESM1, respectively. Among the CMIP5 model considered in Shindell et al. (2015), the multimodel mean regional correlation between

5 the combined effective aerosol and ozone forcing and temperature response was slightly negative (-0.1), varying between negative values in some models and positive values among others.

Figure 3a shows the zonal-mean temperature responses obtained from ECHAM6-MACSP and NorESM1-MACSP experiments. These experiments show a moderate cooling due to anthropogenic aerosols across the Southern hemisphere latitudes, whereas in the Northern hemisphere the cooling response clearly strengthens towards the high-latitudes. ~~The modeled regional~~

10 ~~temperature responses between ECHAM6 and NorESM1 simulations disagree the most in high-latitude regions as seen in~~ ~~Figure 2c, also associated with largest differences in surface albedo feedback (snow/sea ice) between the models (see Figure~~

7
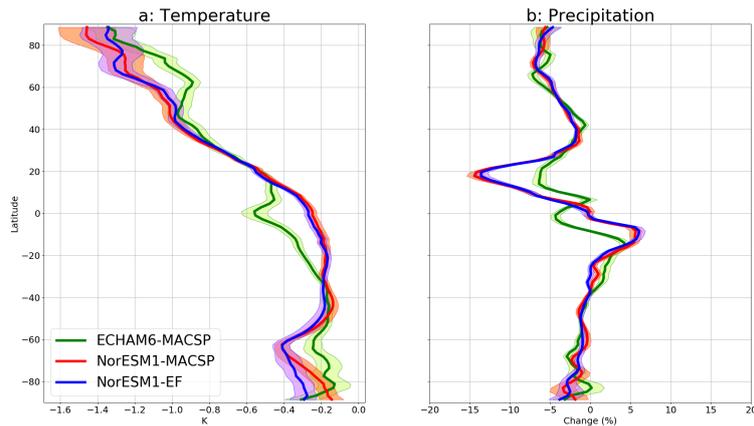
Figure 3. Impact of MACSP anthropogenic aerosols on zonal-mean temperature (K) and precipitation (%) in ECHAM6-MACSP, NorESM1-MACSP and NorESM1-EF experiments. The shaded area shows the standard error of mean as a function of latitude.

~~C2)~~. The modeled regional temperature responses between ECHAM6 and NorESM1 simulations disagree the most in mid- and high-latitude regions as seen in Figure 2c. In high-latitude regions temperature differences are associated with surface albedo responses (snow/sea ice) between the models (see Figure A2). Changes in surface albedo are known to amplify changes in Arctic temperatures (albedo feedback). Hence, differences in snow and sea ice responses may partly explain the difference in temperature responses in the high-latitudes. This feedback, together with ocean circulation feedback, also dominates at high-latitudes the regional differences in temperature responses to homogeneous greenhouse gas forcing among different climate models (Shindell et al., 2015). ~~The similarity in zonal-mean temperature response at high Northern latitudes in ECHAM and NorESM1 is curious, as NorESM1 shows a more positive surface albedo response (Fig. C2) and a more negative cloud cover response (Fig. C3), both of which should favor stronger cooling at high-latitudes.~~

3.2.2 Precipitation

The inclusion of anthropogenic aerosols results in a similar global reduction of precipitation in all experiments, with ECHAM6-MACSP showing a change of -1.79 ± 0.05 % and NorESM1-MACSP change of -1.69 ± 0.04 % in annual precipitation (Table 2). The regional changes of the precipitation patterns are shown in Figure 4. The spatial correlation between the precipitation responses in full ECHAM6-MACSP and NorESM1-MACSP experiments is 0.47, which is much lower than the corresponding correlation for temperature. In addition, while the temperature responses are negative almost globally, both positive and negative responses occur for precipitation, with relatively sharp edges between regions with different signs of changes. While similar large-scale features of precipitation changes can be seen in both models, their dislocation leads to a weaker regional correlation than for the temperature response. In both models, the relative changes in the convective precipitation are larger than the relative changes in large-scale precipitation. Also consistently across the two models, the seasonal response in the
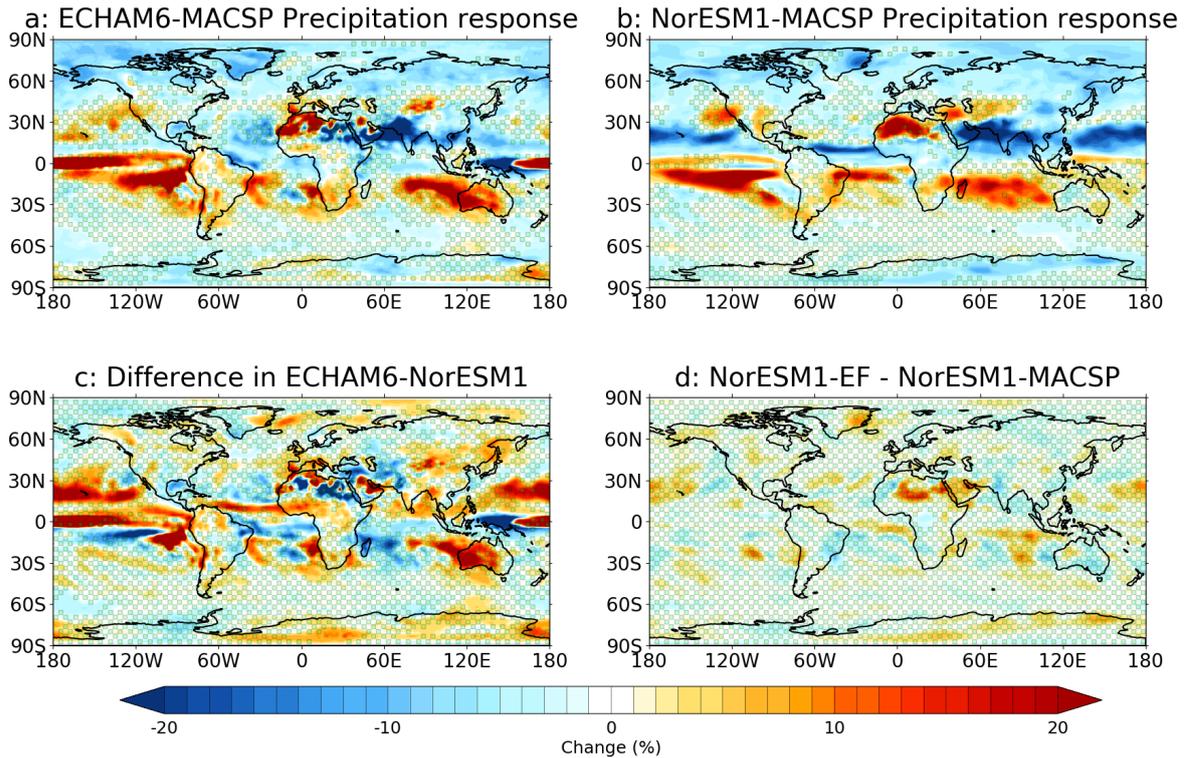
Figure 4. The upper left shows ECHAM6-MACSP experiment precipitation response to adding MACv2-SP aerosols and the upper right figure shows the same for NorESM1-MACSP experiment. Below is the intermodel difference in precipitation response and difference between NorESM1-MACSP and NorESM1-EF. The green dots mark the regions where the MACv2-SP aerosols do not have a statistically significant impact at the level p < 0.05.

Table 2. Summary of global mean change of temperature and precipitation due to modern day. Standard error of means are shown in brackets

	Near surface temperature	Precipitation (%)
ECHAM6-MACSP	-0.50 (± 0.03)	-1.79 (± 0.05)
NorESM1-MACSP	-0.48 (± 0.02)	-1.69 (± 0.04)
NoreESM-EF	-0.49 (± 0.01)	-1.82 (± 0.04)

total precipitation is similar, with the largest changes in June-July-August (see Table A1). Both models consistently show an overall drying of the Northern Hemisphere, with some statistically significant regional increase in precipitation over ~~Arfica~~ the Northwest Africa.

Both models show a maximum reduction in total precipitation around $15° - 20°$ N and a maximum increase around $10° - 15°$ S, associated with an asymmetric response in Hadley circulation across the equator (see Figs. 3b and 4). Changes in

precipitation in the tropics are also related to changes in vertical motion in the same region (see Fig. C4). This is suggestive of a southward shift of the Intertropical convergence zone (ITCZ) associated with a change in hemispheric temperature gradient (Broccoli et al., 2006). The inclusion of anthropogenic aerosols results in decreased precipitation in the South Asian monsoon region (defined here as the land region over 5°–25°N, 65°–110°E) (Fig. 3). In June-August, the monsoon precipitation is decreased by 12.8% in ECHAM6-MACSP and 15.3% in NorESM1-MACSP experiments. Reduction of monsoon precipitation due to the anthropogenic aerosols has also been reported in several previous studies (Ganguly et al., 2012; Li et al., 2018b; Polson et al., 2014; Bollasina et al., 2011). ~~Opposite to~~ In contrast with the seasonal cycle in temperature response, the largest precipitation response occurs in Northern hemispheresummer during the Asian monsoon season. The two models show a different response over the West African monsoon region (5° S – 25° N, 20° W – 20° E), with NorESM1-MACSP experiment showing a statistically significant reduction in precipitation of −5.3 % while ECHAM6-MACSP experiment does not show a significant change (−1.8 %). In the vicinity of the Australian continent, ECHAM6-MACSP experiment shows an area of increased precipitation extending from the Indian Ocean to the Western Australia, while in NorESM1-MACSP experiment, the increase is located entirely over the Indian ocean.

There appear to be several causes for the differences in the precipitation response between the two models. For instance, there is a relationship between the difference of the regional precipitation response and the difference in vertical velocity response (correlation coefficient 0.44 between Figure 4c and Figure C4c). However, it cannot be concluded that change in precipitation is caused by the change in vertical velocity. Probably, both the changes in vertical velocity and precipitation are related to changes in circulation. Also the difference in the initial equilibrium state of precipitation patterns correlates weakly with the difference in the precipitation response (correlation coefficient 0.23). Furthermore, differences in the cloud cover responses (see Figure C3) are also related to differences in precipitation responses (correlation coefficient 0.32). The vertical velocity correlates also with the total cloud cover response (correlation coefficient 0.41)

3.2.3   Comparison between the NorESM1-MACSP and NorESM1-EF experiments

We now briefly discuss the differences between the NorESM1-MACSP and NorESM1-EF experiments. As noted in Sect. 2.3, the difference between these experiments is that in NorESM1-MACSP, the radiative forcing due to the MACv2-SP aerosols is computed using NorESM1's own meteorology and own radiation scheme, while in NorESM1-EF, forcings from ECHAM6's MACSP run are applied. The forcing results are shown in Figure 1d. The minor differences seen in Fig. 1d are related to interpolating the radiative forcing between ECHAM6 and NorESM1 horizontal grids. The general finding here is that the results for these two experiments are very similar. The global-mean temperature response is −0.48(±0.02) K for NorESM1-MACSP and −0.49(±0.01) K for NorESM1-EF, while the global-mean precipitation responses are −1.69(±0.04)% and −1.82(±0.04)%. Also, the zonal-mean and regional temperature and precipitation responses in NorESM1-MACSP and NorESM1-EF are very similar (Figs. 2d, 3, and 4d). The spatial correlation in response between full NorESM1-MACSP and NorESM1-EF experiments is as high as 0.97 for temperature and 0.95 for precipitation, which are much higher than the correlations between NorESM1-MACSP and ECHAM6-MACSP responses (0.81 and 0.47). Indeed, with the exception of the global-mean precipitation response, for which the ECHAM6-MACSP value (-1.79 ±0.05%) falls between NorESM1-MACSP and NorESM1-EF,

Table 3. Intermodel correlations of regional temperature response for the Samset et al. (2018) models and our models. The average correlation coefficient between the Samset et al. (2018) models is 0.79 with a standard deviation of 0.05; the average correlation coefficient between the models used in this study and the Samset et al. (2018) models is 0.76. The correlations are calculated for 50 years with and 50 years without anthropogenic aerosols. Correlation for our whole dataset (60+120 years) is shown in brackets. The range is the standard deviation between results obtained for two different CTRL runs.

	CESM1	GISS	HadGEM2	NorESM1	ECHAM6-MACSP	NorESM1-MACSP
GISS	0.74					
HadGEMS2	0.83	0.79				
NorESM1	0.82	0.71	0.87			
This study						
ECHAM6-MACSP	0.75±0.01	0.72±0.02	0.75±0.01	0.74±0.02		
NorESM1-MACSP	0.80±0.01	0.68±0.01	0.79±0.0	0.85±0.01	0.78±0.02 (0.81)	
NorESM1 - EF	0.81±0.00	0.7±0.01	0.77±0.0	0.80±0.02	0.78±0.02 (0.82)	0.96±0.0 (0.97)

the responses in the two NorESM1 experiments are closer to each other than the ECHAM6-MACSP response. Therefore, it can be concluded that the differences in the effects of MACv2-SP aerosols between ECHAM6 and NorESM1 are mainly related to differences in the model ~~dynamical~~ circulation responses, not to the differences in the aerosol forcing fields.

3.3 Comparison to models with interactive aerosols

5   Finally, we compare the obtained equilibrium temperature and precipitation responses with prescribed MACv2-SP aerosols in ECHAM6 and NorESM1 against those equilibrium climate responses from four fully coupled climate models (CESM1, GISS, HadGEMS2, and NorESM1) with intrinsic aerosol schemes but the same aerosol emissions, reported by Samset et al. (2018). In the four models considered by Samset et al. (2018), the global average temperature responses were $-(0.5$ K, 0.5 K, 1.1 K and 0.6) K, and precipitation responses $-(1.5\%, 1.8\%, 2.6\%$ and 3.1%), respectively. We obtain similar temperature responses

10   of $-(0.48$–0.50) K and precipitation responses of $-(1.69$–1.82) % using the prescribed MACv2-SP aerosol description.

Tables 3 and 4 show the correlation coefficients for regional climate responses between all experiments in our and Samset et al. (2018) data sets. The correlations are calculated for equilibrium climate runs with equal time averaging over 50 years with and without anthropogenic aerosols both for our and Samset et al. (2018) datasets. Note that these coefficients do not depend on the magnitude of the average responses in the models, but only on the relative regional distributions of the responses. Perhaps

15   surprisingly, the average correlation coefficient for regional temperature response between interactive aerosol models (i.e., the Samset et al. (2018) models), 0.79, is almost identical to the correlation between our prescribed aerosol models (0.78). Also, the average correlation coefficient between experiments using interactive aerosols and a fully coupled ocean model (Samset et al. models) and experiments using prescribed aerosols and a slab ocean model (our models) is 0.76, nearly the same as for the fully coupled interactive aerosol models only. The similar regional correlation between different experiments is remarkable

Table 4. Intermodel correlations of regional precipitation response for the Samset et al. (2018) models and our models. The average correlation coefficient between the models is 0.34 with a standard deviation of 0.10; the average correlation coefficient between the models used in this study and the Samset et al. (2018) models is 0.38. The correlations are calculated for 50 years with and, 50 years without anthropogenic aerosols. Range of the correlation coefficient shows the standard deviations between results obtained for two different CTRL runs. The correlation for our whole dataset (60+120 years) is shown in brackets.

	CESM1	GISS	HadGEM2	NorESM1	ECHAM6-MACSP	NorESM1-MACSP
GISS	0.38					
HadGEMS2	0.42	0.43				
NorESM1	0.39	0.12	0.31			
This study						
ECHAM6-MACSP	0.42±0.03	0.28±0.03	0.36±0.03	0.12±0.07		
NorESM1-MACSP	0.5±0.05	0.34±0.03	0.49±0.0	0.38±0.03	0.41±0.02 (0.47)	
NorESM1 - EF	0.54±0.00	0.41±0.0	0.48±0.0	0.33±0.0	0.41±0.02 (0.47)	0.85±0.08 (0.95)

considering large differences in the aerosol descriptions between the different models. It appears that the differences in aerosol descriptions do not dominate the differences in regional temperature response.

The average correlation coefficient for regional precipitation changes within Samset et al. (2018) models with intrinsic aerosol descriptions is 0.34, while it is 0.41 within our models with prescribed aerosols. The average correlation coefficient for regional precipitation changes between the Samset et al. (2018) models with fully coupled ocean and our models with a slab ocean is 0.39, which is similar to the mean correlation within the Samset et al. models. The correlation coefficient between NorESM1 experiments using different aerosol descriptions and ocean models is now only 0.33/0.38. Thus, differences in aerosol descriptions, ocean models and atmospheric responses all contribute to differences in regional precipitation responses. The correlation coefficients for precipitation responses are, however, more uncertain than those for temperature responses, due to a stronger impact of natural variability.

Even long equilibrium climate runs cannot fully eliminate the natural climate variability on a regional level. With our full dataset (60 of MACSP runs+120 years of CTRL run) we obtain a spatial correlation of 0.47 between NorESM1-MACSP and ECHAM6-MACSP precipitation responses, a slight improvement over the correlation coefficient of 0.41(±0.02) for 50+50 year datasets. The spatial correlation for temperature improves from 0.78(±0.02) to 0.81. ~~The fully coupled ocean models in the Samset et al. (2018) dataset also feature long-term internal variability in the ocean states that adds to the level of natural variation with respect to our models with simpler slab ocean representations used in this paper.~~ The fully coupled ocean models in the Samset et al. (2018) dataset also feature long-term internal variability in the ocean states that adds to the level of natural variation with respect to our models with simpler slab ocean representations used in this paper. Therefore, we would expect the Samset et al. data to include more noise than our results with slab ocean configurations. Furthermore, it is important to note that differences in the ocean descriptions are known to have a large impact in the regional climate responses between

different models ((Deser et al., 2016; Kay et al., 2016)). Overall, we would expect that due to these differences the climate signals obtained from fully coupled models would intrinsically correlate less well with each other than those from models with slab ocean configurations. Somewhat surprisingly, this turns out not to be the case.

The dependence of the calculation of time-averaged correlation coefficients on the simulation length for our data is shown in Fig. 5. There, the blue and red shaded regions represent the level of expected variation in the regional correlation coefficients between two climate models obtained from equilibrium model experiments with and without anthropogenic aerosols. We obtained a correlation coefficient of 0.78 with a standard deviation of ± 0.02 for temperature response and $0.41(\pm 0.02)$ for precipitation after 50 years of simulation, these periods being representative for the Samset experiments but neglecting the impact of long-term ocean variations. The corresponding correlation coefficients for full model runs (60+120 years of simulation) are 0.47 for precipitation and 0.81 for temperature.
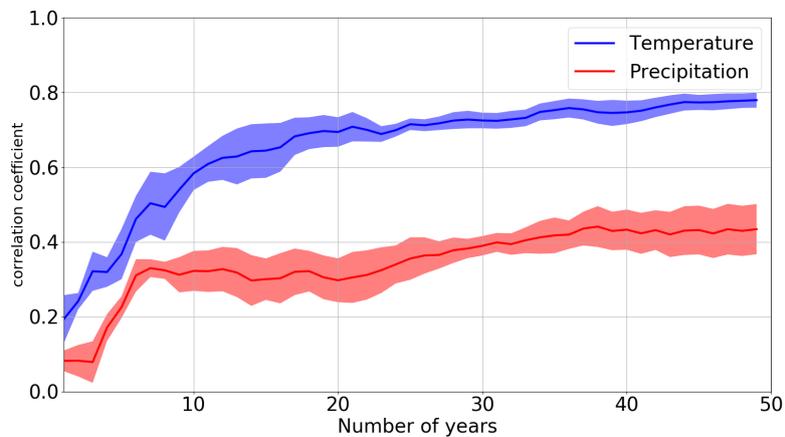
Figure 5. Correlation coefficient of temperature (precipitation) response as a function of the number of averaged years. Blue (red) is the correlation between the temperature responses to MACv2-SP aerosols in the two models. The shaded area shows the variation between different control runs. The same number of years is used for the CTRL run and MACSP run.

4 Conclusions

We have here provided ~~the first~~ results on the equilibrium climate response of modern day anthropogenic aerosols using two different climate models, ECHAM6 and NorESM1, with ~~identical anthropogenic aerosol representations in the models~~ the MACv2-SP (Stevens et al., 2017) anthropogenic aerosol representations The results were obtained both using the same representations of aerosol optical properties and cloud-albedo effect and for identical instantaneous aerosol radiative forcing fields in the models.

The MACv2-SP aerosols produced a very similar total instantaneous anthropogenic aerosol radiative forcing in the two models ($-0.64\,\mathrm{Wm}^{-2}$ in ECHAM6-MACSP and $-0.69\,\mathrm{Wm}^{-2}$ in NorESM1-MACSP experiments). We found that there are

differences up to $3.2\ \mathrm{Wm}^{-2}$ in the instantaneous regional aerosol forcing between the models when using the same aerosol representation.These differences can mostly be explained via differences in cloud fields and surface albedo in the models.

The addition of MACv2-SP anthropogenic aerosols produced very similar global average responses on temperature, $-0.48(\pm0.02)$ K and $-0.50(\pm0.03)$ K, and precipitation, $-1.69(\pm0.04)\%$ and $-1.79(\pm0.05)\%$ in NorESM1-MACSP and ECHAM6-MACSP experiments, respectively. The largest disagreements in regional temperature response were found at high-latitude regions associated with largest differences in surface albedo feedback (snow/sea ice), while the largest differences in regional precipitation response were located mainly in the tropics, ~~in part due to changes in the ITCZ~~. These key regional differences remained even when using exactly the same aerosol radiative forcing fields in both models. Several previous studies have discussed that the main driver for ITCZ shift is the northern hemisphere cooling due to anthropogenic aerosols (Broccoli et al., 2006; Hwang et al., 2013; Wang, 2015). Chiang and Bitz (2005) showed with Community Climate Model version 3 a connection between ITCZ shift and added Arctic ice cover. Based on these previous studies, it seems plausible that different responses in Arctic sea ice and snow cover in ECHAM6-MACSP and in the two NorESM1 experiments result in different high-latitude temperature responses, which in turn are reflected as differences in the ITCZ shift that drives the precipitation change at low latitudes. However, it should be noted that the ITCZ shift is also sensitive to the type of ocean model used, and slab ocean models tend to exaggerate the change in ITCZ (Kay et al., 2016).

We compared our results using uniform aerosol representations to a set of four current climate models using their intrinsic aerosol representations but the same aerosol emissions, reported by Samset et al. (2018). Among Samset et al. (2018) models the global responses to additions of anthropogenic aerosol varied between -0.5 K and -1.1 K for temperature and between -1.5% and -3.1% for precipitation. However, the correlation coefficients for regional distributions of climate responses, averaged over equal run length, were ~~essentiallye qually~~ nearly as good among our experiments with prescribed aerosols and slab ocean representation (0.78 for temperature and 0.41 for precipitation) and among the Samset et al. experiments with model-intrinsic aerosols and the fully coupled ocean representation (0.79 for temperature and 0.34 for precipitation).

~~This implies that differences in aerosol descriptions among different models are not the main cause of variation in the regional distributions of climate responses among different models. Rather, differences in model intrinsic dynamic responses appear to dominate the differences in regional climate responses~~

The lack of improvement in the correlation coefficients suggests that differences in aerosol descriptions are not the only cause of regional differences in climate signals between the models. Rather, the differences in model circulation responses appear to dominate the differences in regional climate responses. Figure C5 shows the average 850 hPa wind responses for ECHAM6-MACSP and NorESM1-MACSP experiments for Northern hemisphere winter. The responses in the circulation fields vary significantly between the two models, with an annual average correlation coefficient of only 0.18 (DJF:-0.03; MAM:0.07; JJA:0.15; SON:0.19). The lack of robustness in atmospheric circulation responses between different climate models has been previously discussed by Shepherd (2014) for CMIP5 RCP8.5 scenarios and by Li et al. (2018a) for HAPPI 1.5 K and 2.0 K warming scenarios. Shepherd (2014) argued that the differences in circulation responses cause variation in the regional temperature and precipitation responses in future climate scenarios. Li et al. (2018a) showed that model consensus for circulation response is low even for atmosphere-only models forced with same time-varying SST and sea ice, anthropogenic

greenhouse gases, ozone, land use, land cover, and aerosols. Both in Shepherd (2014) and Li et al. (2018a) data, the NH wintertime circulation response over the North Atlantic disagrees significantly between models. Also for ECHAM6-MACSP and NorESM1-MACSP the circulation response over the North Atlantic show differences in magnitude and pattern. Differences are also seen over the North Pacific region. Combined with the difference in the sea ice and surface albedo change in the North

5   Pacific, these circulation changes can drive the temperature response differences in the region.

Our results imply that in current global climate models the regional aerosol climate impacts cannot be better constrained by further improving aerosol descriptions alone. More extensive model comparisons are needed to explain the model discrepancies in response to aerosol forcing. Improvements on the dynamical cores, physical parameterizations and ocean models are needed to narrow down model uncertainties in the regional aerosol climate responses.

10   *Data availability.* Data and scripts used for data analysis can be obtained by contacting the corresponding author

Appendix A: NorESM-EF technical description

The NorESM1-EF run employed radiative forcing extracted from the ECHAM6-MACSP run. First, multi-year monthly means of MACv2-SP aerosol radiative forcing (for TOA and surface radiative fluxes and atmospheric heating rates) were computed for ECHAM6-MACSP. Second, these values were interpolated to the NorESM1 horizontal and vertical grid and normalized

15   by the monthly-mean incoming solar radiation at model top. Third, during the NorESM1-EF run, these normalized forcings were multiplied by the TOA incoming solar radiation at each radiation time step, and they were added to the radiative fluxes and heating rates computed without MACv2-SP aerosols.

This treatment ensures that the diurnal cycle of the aerosol forcing is approximately correct; in particular there is no aerosol forcing during the night. However, the computed forcing is independent of the clouds simulated by NorESM1. Thus, while the

20   aerosol radiative forcing is computed correctly in a monthly-mean sense, its sub-monthly correlation with clouds is ignored. In principle, this could impact the differences between NorESM1-EF and ECHAM6-MACSP. The impact is, however, most likely small. If neglecting the sub-monthly correlation between clouds and aerosol forcing were to have a substantial impact on the climate response to MACv2-SP aerosols, this should also show up in the differences between NorESM1-EF and NorESM1-MACSP. Yet the differences between NorESM1-EF and NorESM1-MACSP are very small (Tables 2 and A1), in fact much

25   smaller than the corresponding differences between ECHAM6-MACSP and either NorESM1-EF or NorESM1-MACSP. This strongly suggests that the differences between NorESM1-EF and ECHAM6-MACSP are primarily caused by the use of a different climate model rather than by the subtle differences in radiative forcing.

Appendix B: Sensitivity analysis of model aerosol forcing

We used a Gaussian process emulation technique (O'Hagan (2006)) to evaluate the regional differences in aerosol radiative forcing. First, we simply assume that the forcing difference depends only on the differences in model output values, and not on the actual values themselves. Second, we selected the differences in modeled output (total cloud cover, surface albedo, precipitation, surface temperature, surface wind u-component) as trial sets for these values. These can be described via a relation $Y = \eta(\boldsymbol{X})$, where $\boldsymbol{X} = [\Delta\alpha, \Delta\beta, ..., \xi]$, where α and β are total cloud cover and surface albedo, ξ is pure noise (Gaussian) variable. Next the function $Y = \eta(\boldsymbol{X})$ is inferred using a Gaussian Process prior emulator for a part of the yearly averaged radiative forcing data (in our case, 40 years). Each variable is assigned with a sensitivity index, which describes the relative sensitivity of Y to that variable. The sensitivity analysis of the estimated Y function was done by using Extended Fourier Amplitude Sensitivity Test (FAST) (Saltelli et al. (1999)). As an end result, FAST assesses the contributions of each emulator input variable (components of $X = (X_i)$) to the variance in emulator output variable (Y), where it's assumed the input variables X_i have an independent and identical distribution uniform prior. The inferred function Y is finally validated by comparing the emulated forcing field against validation data separate from the training data (here, 20 yearly averaged forcing fields from the model experiments).

Appendix C: Appendix figures C1-C5 and Table A1
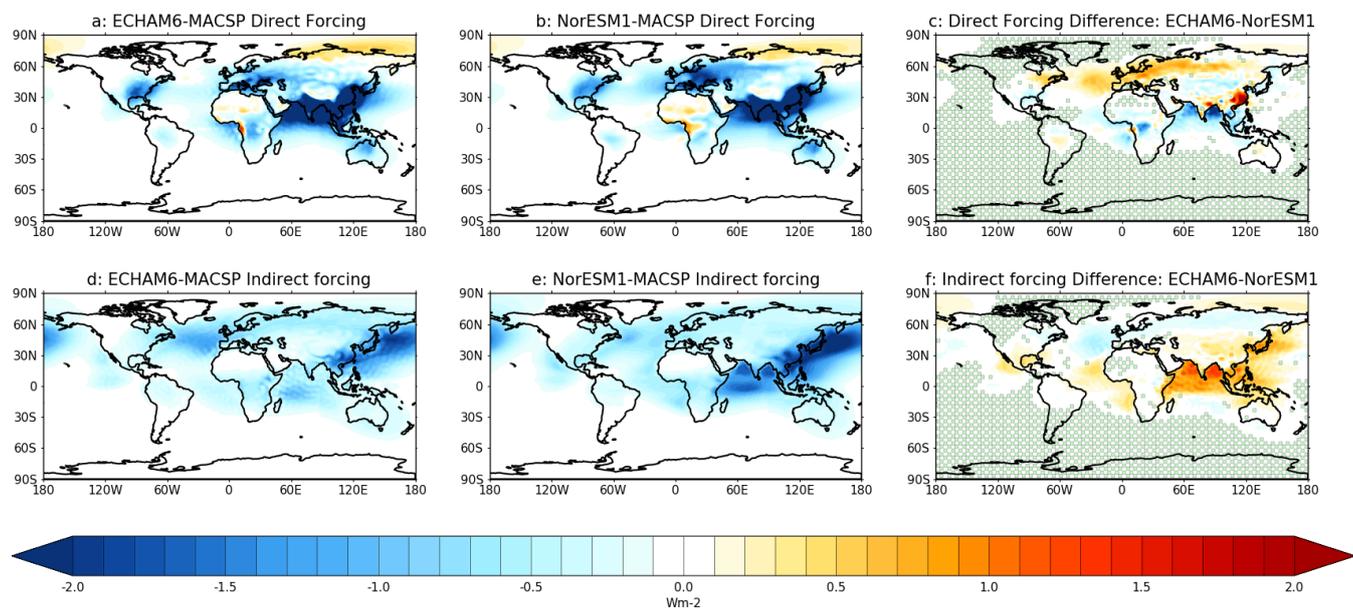
Figure C1. Instantaneous radiative forcing by anhtropogenic (MACv2-SP) aerosols. First row shows the direct radiative forcing and second row shows the indirect radiative forcing produced by MACv2-SP. Green masking in (c) and (f) indicates areas where the difference between the models in the instantaneous radiative forcing is not statistically significant (p > 0.05).
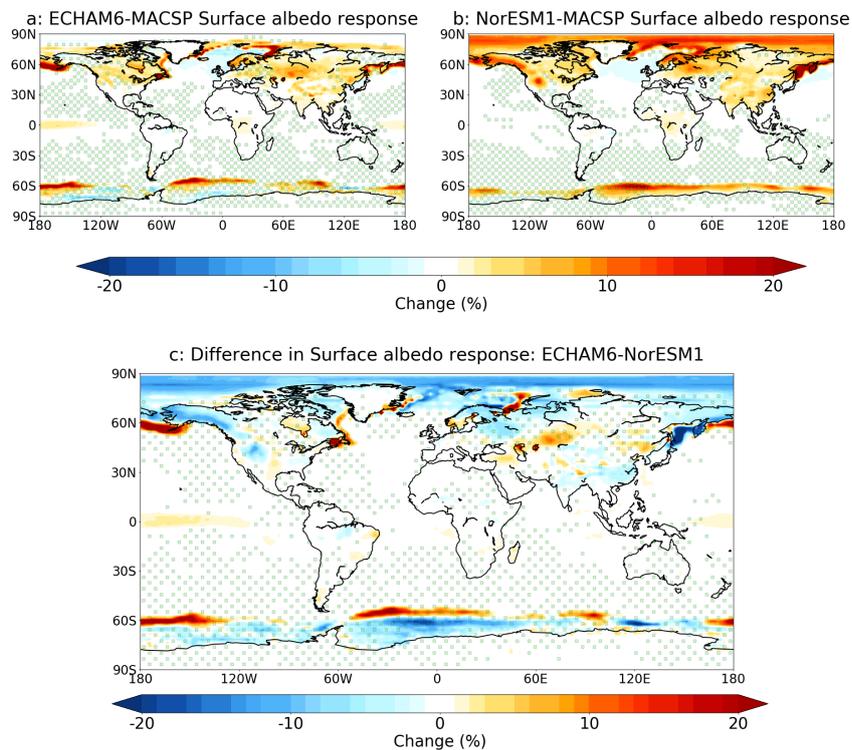
Figure C2. Surface albedo response to the addition of anthropogenic aerosols. a: response in ECHAM6-MACSP experiment; b: response in NorESM1-MACSP experiment; c: Difference in surface albedo response: ECHAM6-MACSP experiment minus NorESM1-MACSP experiment. The green dots present the the area where anthropogenic aerosols do not have a statistically significant impact at the level p < 0.05 (in panel c), or where the difference between the models is not statistically significant (in panels (a) and (b)).
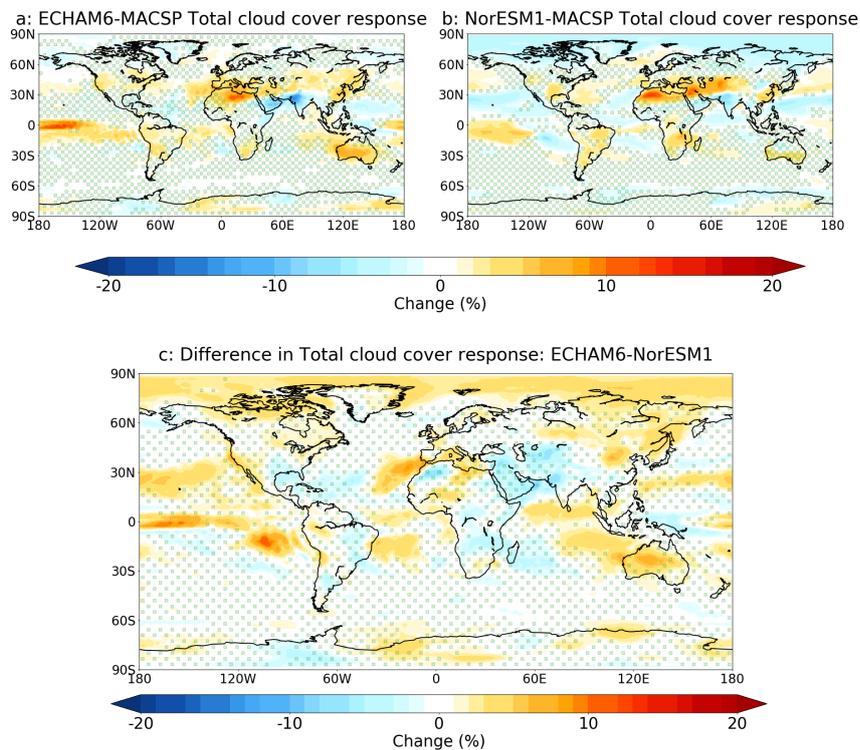
Figure C3. Total cloud cover response to the addition of anthropogenic aerosols. a: response in ECHAM6-MACSP experiment; b: response in NorESM1-MACSP experiment; c: the difference in responses between the experiments. The green dots presents the area where anthropogenic aerosols do not have a statistically significant impact at the level p < 0.05 (in panel c), or where the difference between the models is not statistically significant (in panels (a) and (b)).
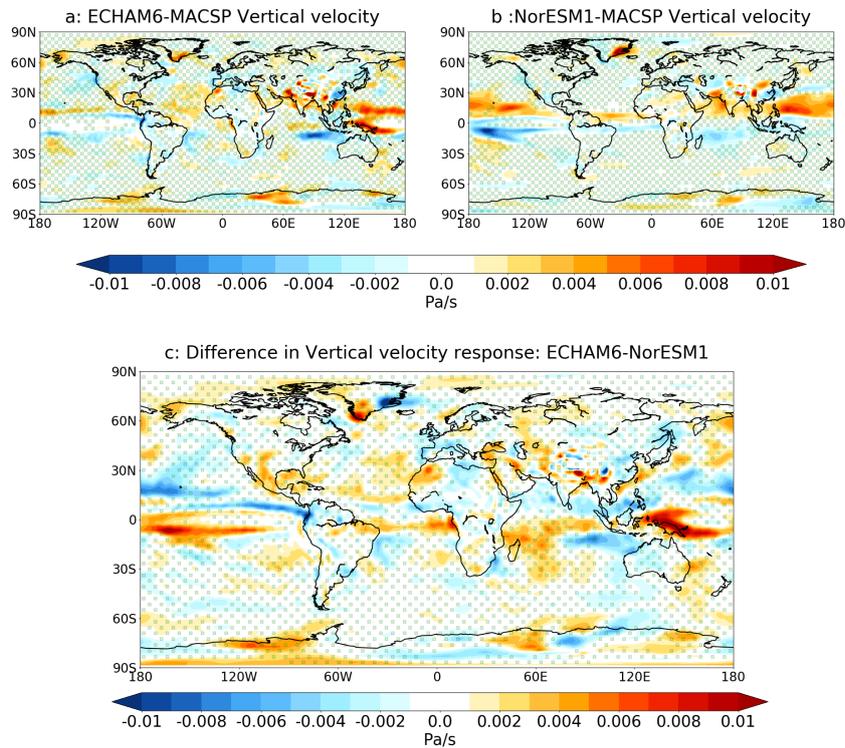
Figure C4. Vertical motion response at the 600hPa level to the addition of anthropogenic aerosols. a: response in ECHAM6-MACSP experiment; b: response in NorESM1-MACSP experiment; c: the difference in responses between the experiments. The green dots presents the area where anthropogenic aerosols do not have a statistically significant impact at the level p < 0.05 (in panel c), or where the difference between the models is not statistically significant (in panels (a) and (b))
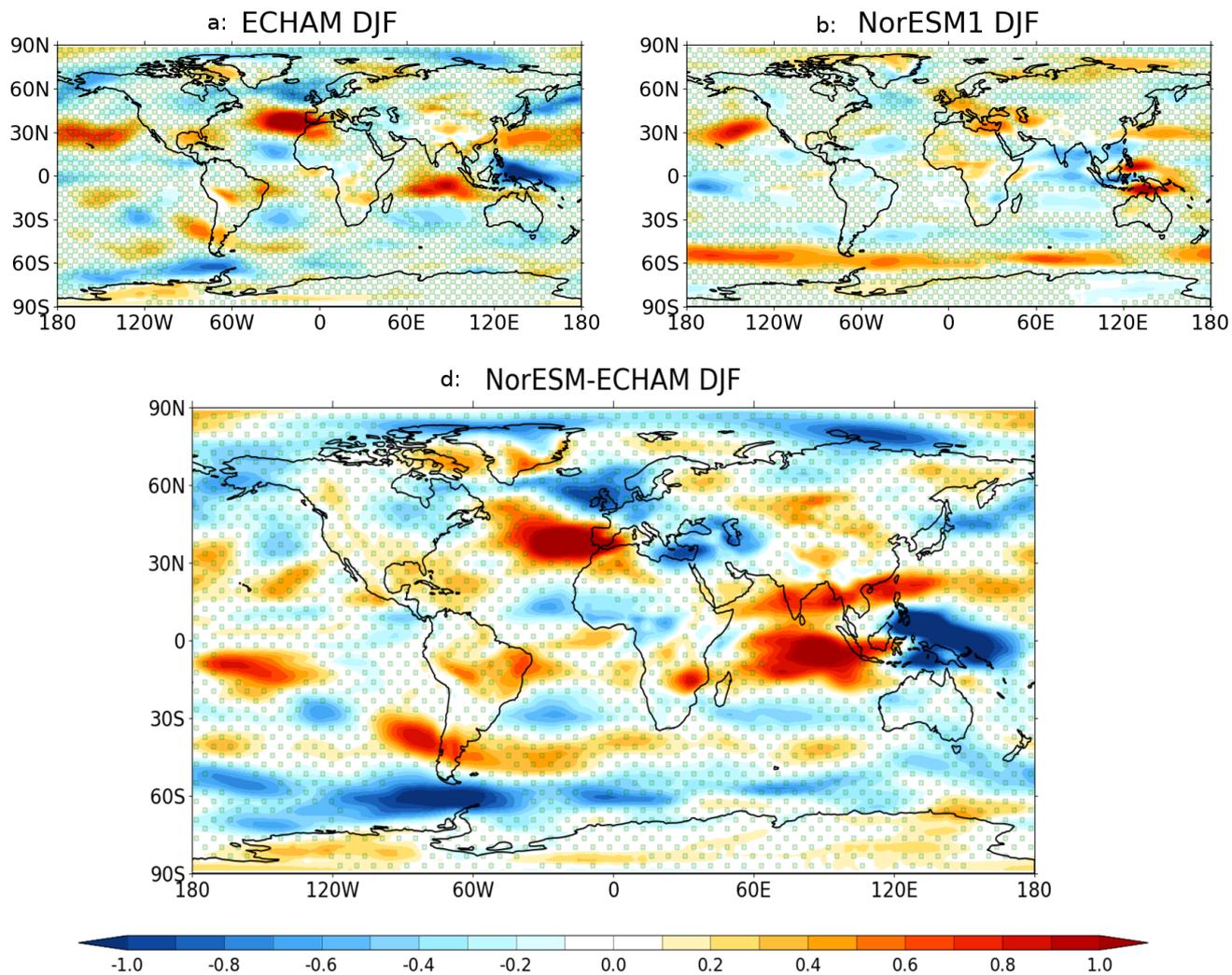
Figure C5. Lower tropospheric (850 hPa) zonal wind response to adding MACv2-SP anthropogenic aerosols for Northern hemisphere winter. The green dots presents the area where anthropogenic aerosols do not have a statistically significant impact at the level p < 0.05 (in panel c), or where the difference between the models is not statistically significant (in panels (a) and (b)). The units are in m/s.

Table A1. Summary of global mean change of temperature and precipitation due to modern day anthropogenic aerosols. Errorbars are standard error of means

	DJF	MAM	JJA	SON	Annual
Near surface temperature					
ECHAM6-MACSP	-0.54(± 0.03)	-0.50(± 0.03)	-0.44(± 0.02)	-0.51(± 0.02)	-0.50(± 0.03)
NorESM1-MACSP	-0.49(± 0.02)	-0.46(± 0.02)	-0.45(± 0.01)	-0.51(± 0.02)	-0.48(± 0.02)
NorESM1-EF	-0.51(± 0.02)	-0.47(± 0.01)	-0.46(± 0.01)	-0.50(± 0.01)	-0.49(± 0.01)
Total precipitation (%)					
ECHAM6-MACSP	-1.45(± 0.07)	-1.82(± 0.07)	-2.11(± 0.08)	-1.79(± 0.07)	-1.79(± 0.05)
NorESM1-MACSP	-1.62(± 0.07)	-1.53(± 0.07)	-2.08(± 0.07)	-1.52(± 0.06)	-1.69(± 0.04)
NorESM1-EF	-1.7(± 0.05)	-1.68(± 0.05)	-2.17(± 0.07)	-1.71(± 0.04)	-1.82(± 0.04)
Large scale precipitation (%)					
ECHAM6-MACSP	-1.62(± 0.22)	-1.65(± 0.12)	-1.22(± 0.2)	-0.77(± 0.16)	-1.31(± 0.1)
NorESM1-MACSP	-0.58(± 0.21)	-0.83(± 0.18)	-2.74(± 0.23)	-1.03(± 0.16)	-1.28(± 0.09)
NorESM1-EF	-0.74(± 0.18)	-0.98(± 0.15)	-2.77(± 0.22)	-1.03(± 0.09)	-1.37(± 0.08)
Convective precipitation (%)					
ECHAM6-MACSP	-1.36(± 0.12)	-1.91(± 0.11)	-2.56(± 0.1)	-2.34(± 0.1)	-2.05(± 0.06)
NorESM1-MACSP	-2.27(± 0.14)	-1.93(± 0.13)	-1.71(± 0.11)	-1.82(± 0.09)	-1.93(± 0.08)
NorESM1-EF	-2.28(± 0.11)	-2.08(± 0.09)	-1.83(± 0.08)	-2.12(± 0.09)	-2.08(± 0.06)

References

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, Geosci. Model Dev, 6, 687–720, https://doi.org/10.5194/gmd-6-687-2013, www.geosci-model-dev.net/6/687/2013/, 2013.

5   Bollasina, M. A., Ming, Y., and Ramaswamy, V.: Anthropogenic Aerosols and the Weakening of the South Asian Summer Monsoon, Science, 334, 502, http://science.sciencemag.org/content/334/6055/502.abstract, 2011.

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

10   pdf/assessment-report/ar5/wg1/WG1AR5_Chapter07_FINAL.pdf, 2013.

Broccoli, A. J., Dahl, K. A., and Stouffer, R. J.: Response of the ITCZ to Northern Hemisphere cooling, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL024546, http://doi.wiley.com/10.1029/2005GL024546, 2006.

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,

15   https://doi.org/10.1038/nature12674, http://www.nature.com/articles/nature12674, 2013.

Chiang, J. C. H. and Bitz, C. M.: Influence of high latitude ice cover on the marine Intertropical Convergence Zone, Climate Dynamics, 25, 477–496, https://doi.org/10.1007/s00382-005-0040-5, http://link.springer.com/10.1007/s00382-005-0040-5, 2005.

Deser, C., Sun, L., Tomas, R. A., and Screen, J.: Does ocean coupling matter for the northern extratropical response to projected Arctic sea ice loss?, Geophysical Research Letters, 43, 2149–2157, https://doi.org/10.1002/2016GL067792, https://agupubs.onlinelibrary.wiley.

20   com/doi/abs/10.1002/2016GL067792, 2016.

Ekman, A. M. L.: Do sophisticated parameterizations of aerosol-cloud interactions in CMIP5 models improve the representation of recent observed temperature trends?, Journal of Geophysical Research: Atmospheres, 119, 817–832, https://doi.org/10.1002/2013JD020511, http://doi.wiley.com/10.1002/2013JD020511, 2014.

Feser, F., Rockel, B., von Storch, H., Winterfeldt, J., Zahn, M., Feser, F., Rockel, B., Storch, H. v., Winterfeldt, J., and Zahn, M.: Regional

25   Climate Models Add Value to Global Model Data: A Review and Selected Examples, Bulletin of the American Meteorological Society, 92, 1181–1192, https://doi.org/10.1175/2011BAMS3061.1, http://journals.ametsoc.org/doi/abs/10.1175/2011BAMS3061.1, 2011.

Fiedler, S., Kinne, S., Huang, W. T. K., Räisänen, P., O'Donnell, D., Bellouin, N., Stier, P., Merikanto, J., van Noije, T., Makkonen, R., and Lohmann, U.: Anthropogenic aerosol forcing – insights from multiple estimates from aerosol-climate models with reduced complexity, Atmospheric Chemistry and Physics, 19, 6821–6841, https://doi.org/10.5194/acp-19-6821-2019, https://www.atmos-chem-phys.net/19/

30   6821/2019/, 2019.

Ganguly, D., Rasch, P. J., Wang, H., and Yoon, J.-H.: Climate response of the South Asian monsoon system to anthropogenic aerosols, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017508, http://doi.wiley.com/10.1029/2012JD017508, 2012.

Giorgi, F. and Francisco, R.: Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM, Climate Dynamics, 16, 169–182, https://doi.org/10.1007/PL00013733, http://link.springer.com/10.1007/

35   PL00013733, 2000.

Hawkins, E., Smith, R. S., Gregory, J. M., and Stainforth, D. A.: Irreducible uncertainty in near-term climate projections, Clim Dyn, 46, 3807–3819, https://doi.org/10.1007/s00382-015-2806-8, https://link.springer.com/content/pdf/10.1007%2Fs00382-015-2806-8.pdf, 2016.

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, Geophysical Research Letters, 40, 2845–2850, https://doi.org/10.1002/grl.50502, http://doi.wiley.com/10.1002/grl.50502, 2013.

Iversen, T., Bentsen, M., Bethke, I., Debernard, J. B., Kirkevåg, A., Seland, ø., Drange, H., Kristjansson, J. E., Medhaug, I., Sand, M., and Seierstad, I. A.: The Norwegian Earth System Model, NorESM1-M – Part 2: Climate response and scenario projections, Geosci. Model Dev, 6, 389–415, https://doi.org/10.5194/gmd-6-389-2013, www.geosci-model-dev.net/6/389/2013/, 2013.

Kasoar, M., Voulgarakis, A., Lamarque, J.-F., Shindell, D. T., Bellouin, N., Collins, W. J., Faluvegi, G., and Tsigaridis, K.: Regional and global temperature response to anthropogenic SO2 emissions from China in three climate models, Atmospheric Chemistry and Physics, 16, 9785–9804, https://doi.org/10.5194/acp-16-9785-2016, https://www.atmos-chem-phys.net/16/9785/2016/, 2016.

Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., Bitz, C., Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., and Bitz, C.: Global Climate Impacts of Fixing the Southern Ocean Shortwave Radiation Bias in the Community Earth System Model (CESM), Journal of Climate, 29, 4617–4636, https://doi.org/10.1175/JCLI-D-15-0358.1, http://journals.ametsoc.org/doi/10.1175/JCLI-D-15-0358.1, 2016.

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Rast, S., Giorgetta, M., Eck, T. F., and Stevens, B.: MAC-v1: A new global aerosol climatology for climate studies, Journal of Advances in Modeling Earth Systems, 5, 704–740, https://doi.org/10.1002/jame.20035, http://doi.wiley.com/10.1002/jame.20035, 2013.

Kirkevåg, A., Iversen, T., Seland, ø., Hoose, C., Kristjánsson, J. E., Struthers, H., Ekman, A. M. L., Ghan, S., Griesfeller, J., Nilsson, E. D., and Schulz, M.: Aerosol–climate interactions in the Norwegian Earth System Model – NorESM1-M, Geoscientific Model Development, 6, 207–244, https://doi.org/10.5194/gmd-6-207-2013, https://www.geosci-model-dev.net/6/207/2013/, 2013.

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, Atmospheric Chemistry and Physics, 10, 7017–7039, https://doi.org/10.5194/acp-10-7017-2010, http://www.atmos-chem-phys.net/10/7017/2010/, 2010.

Li, C., Michel, C., Seland Graff, L., Bethke, I., Zappa, G., Bracegirdle, T. J., Fischer, E., Harvey, B. J., Iversen, T., King, M. P., Krishnan, H., Lierhammer, L., Mitchell, D., Scinocca, J., Shiogama, H., Stone, D. A., and Wettstein, J. J.: Midlatitude atmospheric circulation responses under 1.5 and 2.0 °C warming and implications for regional impacts, Earth System Dynamics, 9, 359–382, https://doi.org/10.5194/esd-9-359-2018, https://www.earth-syst-dynam.net/9/359/2018/, 2018a.

Li, X., Ting, M., and Lee, D. E.: Fast Adjustments of the Asian Summer Monsoon to Anthropogenic Aerosols, Geophysical Research Letters, 45, 1001–1010, https://doi.org/10.1002/2017GL076667, http://doi.wiley.com/10.1002/2017GL076667, 2018b.

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliability Engineering & System Safety, 91, 1290–1300, https://doi.org/10.1016/J.RESS.2005.11.025, https://www.sciencedirect.com/science/article/pii/S0951832005002383, 2006.

Polson, D., Bollasina, M., Hegerl, G. C., and Wilcox, L. J.: Decreased monsoon precipitation in the Northern Hemisphere due to anthropogenic aerosols, Geophysical Research Letters, 41, 6023–6029, https://doi.org/10.1002/2014GL060811, http://doi.wiley.com/10.1002/2014GL060811, 2014.

Ramanathan, P. J. Crutzen J. T. Kiehl, D. R.: Aerosols, Climate, and the Hydrological Cycle, Science, 294, 2119–2124, 2005.

Saltelli, A., Tarantola, S., and Chan, K. P.-S.: A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output, Technometrics, 41, 39, https://doi.org/10.2307/1270993, http://www.jstor.org/stable/1270993?origin=crossref, 1999.

Salzmann, M., Weser, H., and Cherian, R.: Robust response of Asian summer monsoon to anthropogenic aerosols in CMIP5 models, Journal of Geophysical Research: Atmospheres, 119, 321–11, https://doi.org/10.1002/2014JD021783, http://doi.wiley.com/10.1002/2014JD021783, 2014.

Samset, B. H., Sand, M., Smith, C. J., Bauer, S. E., Forster, P. M., Fuglestvedt, J. S., Osprey, S., and Schleussner, C.-F.: Climate Impacts From a Removal of Anthropogenic Aerosol Emissions, Geophysical Research Letters, 45, 1020–1029, https://doi.org/10.1002/2017GL076079, http://doi.wiley.com/10.1002/2017GL076079, 2018.

Shindell, D. T., Faluvegi, G., Rotstayn, L., and Milly, G.: Spatial patterns of radiative forcing and surface temperature response, Journal of Geophysical Research: Atmospheres, 120, 5385–5403, https://doi.org/10.1002/2014JD022752, http://doi.wiley.com/10.1002/2014JD022752, 2015.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, Journal of Advances in Modeling Earth Systems, 5, 146–172, https://doi.org/10.1002/jame.20015, http://doi.wiley.com/10.1002/jame.20015, 2013.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., and Mauritsen, T.: MACv2-SP: a parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6, Geosci. Model Dev, 10, 433–452, https://doi.org/10.5194/gmd-10-433-2017, www.geosci-model-dev.net/10/433/2017/, 2017.

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, https://doi.org/10.5194/acp-13-3245-2013, www.atmos-chem-phys.net/13/3245/2013/, 2013.

Wang, C.: Anthropogenic aerosols and the distribution of past large-scale precipitation change., Geophysical research letters, 42, 10 876–10 884, https://doi.org/10.1002/2015GL066416, http://www.ncbi.nlm.nih.gov/pubmed/27134319http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4845713, 2015.

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, Geophysical Research Letters, 42, 1568–1575, https://doi.org/10.1002/2015GL063301, http://doi.wiley.com/10.1002/2015GL063301, 2015.

Zwiers and von Storch: Taking Serial Correlation into Account in Tests of the Mean, Journal of Climate, 8, 336–351, http://www.hvonstorch.de/klima/pdf/zwiers_storch_1995.pdf, 1995.