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. Modified manuscript can be find from supplements of this post.

Comment 1.
I see this as an interesting study, which usefully adds to the body of recent work em-
phasising 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 Perhaps other results from other experiments with these two models are available (either published or unpublished).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?

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?

Author response:

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.

Please also note the supplement to this comment:

https://www.atmos-chem-phys-discuss.net/acp-2018-1335/acp-2018-1335-AC2-supplement.pdf

__________________________________________