We thank the Referees for taking the time to review our paper and for making helpful comments. We hope we have addressed all of the concerns below.

## Referee #1.

According to Appendix D, the impact of aerosols on autoconversion rates is isolated by ignoring modelled cloud droplet numbers and using fixed values instead. Doing so measures the impact of switching to a different set of cloud droplet numbers but does not isolate the impact of aerosols on autoconversion. Indeed, the reason why the prescribed numbers differ over land and ocean is to crudely represent the more polluted conditions over land. The easy option would be to take the prescribed values from the land/ocean averages of the PI simulation. A much more accurate configuration would be to use the distribution of cloud droplet number simulated in the PI simulation.

*In the simulations mentioned in Appendix D we set the droplet concentration (Nd) terms as used by auto-conversion equation to the constant values in both PI and PD simulations. Since both the PI and PD runs both use same Nd values there would be no impact of the aerosol changes between PI and PD on the autoconversion rates except via any changes in liquid water content. The latter is possible via changes in droplet scattering, semi-direct aerosol effects, etc. However, the main aim was to test the effect of the aerosols on the auto-conversion rate via changes in Nd to test the hypothesis that this was the main cause of the LWP and cloud fraction changes. Since the results showed that the increases in these two quantities between PI and PD was reduced dramatically in the constant auto-conversion runs we feel that this hypothesis was sufficiently proved with this test. The differences in Nd between land and ocean regions for the constant values used in the autoconversion equation should have little impact on this conclusion since the contrast is the same in both the PI and PD runs. The description given in Appendix D did not get across the fact that we used the constant Nd values in both the PI and PD and so we have changed it to :-*

*"Figures D1 and D2 show the percentage increases between PI and PD for LW c and fc (ΔLWPic and $\Delta f_c$), respectively, when aerosols are prevented from affecting the rain autoconversion process. This is done by setting Nd in the autoconversion process equation to a constant value of 300 $cm^{-3}$ over land and 100 $cm^{-3}$ over oceans in both the PI and PD runs."*

Clouds are a very variable component of the atmosphere, as evidenced by the noisy aspect of many figures, so I was surprised that the authors base their analysis on a single, 1-year simulation. There is year-to-year variability even in stratocumulus regions, and as acknowledged by the authors nudging will not suppress that variability, which is a good thing if one wants to isolate adjustments. So readers need to be told which results can be safely generalised. This is especially true of the decomposition of aerosol forcing and the analysis of aerosol-driven cloud regime transitions. I acknowledge that extending the simulated period represents a lot of work, so a discussion of relevant literature in the conclusion section may be a good alternative, although I could not identify specific papers. Perhaps some AeroCom papers looked at interannual variability in aci?

*We decided to run repeat the simulations for an additional year to determine how sensitive the results were to a different meteorological year. We found that the decomposition of the* *aerosol forcing was very similar to that obtained using the original meteorology with only small differences apparent. We added the following text and figures as a new appendix :-*

Fig. H1 shows that the pattern and magnitude of both $ERF_{ARI}$ and $ERF_{ACI}$ are very similar for the alternate year with only slight differences (compare to Fig. 12): the region of $ERF_{ARI}$ forcing off the coast of the USA extends further east across the Atlantic; the main region of $ERF_{ACI}$ around Newfoundland is still present, but extends further south and less to the southeast; the region of $ERF_{ACI}$ north of Scandinavia has a similar spatial pattern, but is enhanced somewhat in the alternate year.

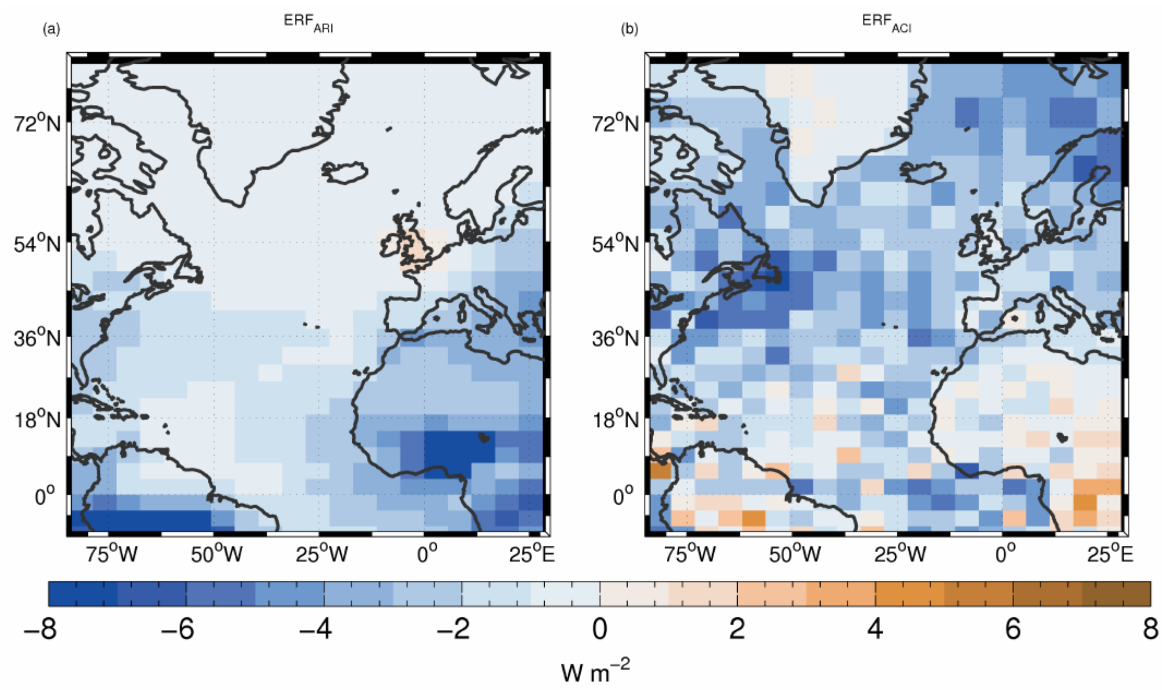


**Figure H1.** As for Fig. 12 except for the 2010-2011 period.

Fig. H2 (compare to Fig 15) shows the contributions to $ERF_{ACI}$ from the changes in $N_{\mathrm{d}}$, $LWP_{\mathrm{ic}}$ and $f_{\mathrm{c}}$ between the PI and PD for the alternate year. It shows that there are some small differences in the spatial patterns, but the overall patterns of the contributions remain very similar with $N_{\mathrm{d}}$ and $LWP_{\mathrm{ic}}$ changes dominating over the northern NA region and $f_{\mathrm{c}}$ changes dominating over the southern NA as was the case for the original chosen time period.

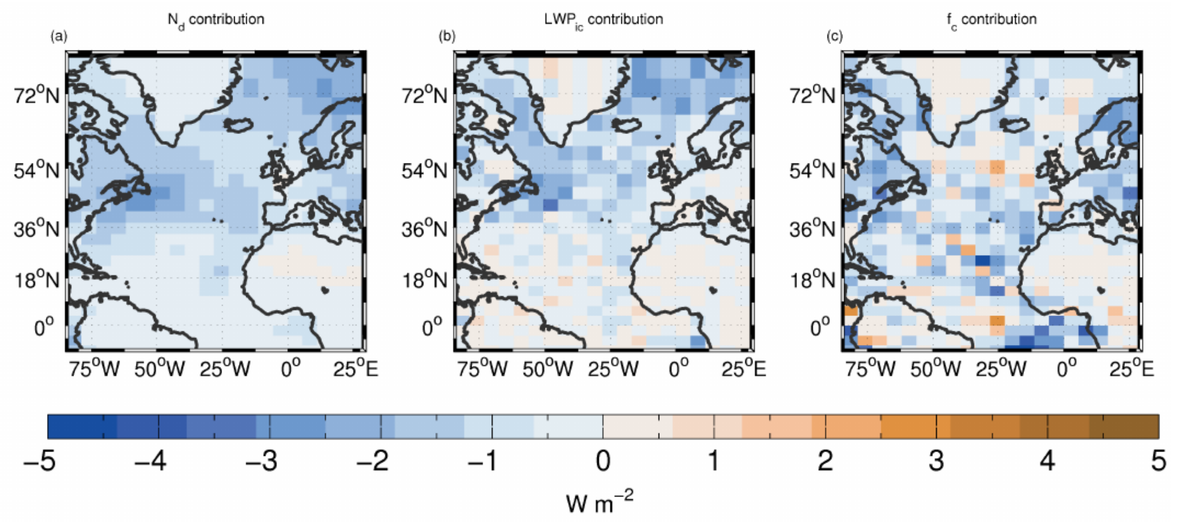


**Figure H2.** As for Fig. 15 except for the 2010-2011 period.

A figure equivalent to Fig. 18 for the alternative year showing the contributions from changes between different cloud states (not shown) reveals a very similar pattern with very similar magnitudes of contribution. Likewise, the alternative version of Fig. 19 (contributions from changes between cloud fractions; not shown) is again very similar to original version in terms of both the pattern and magnitude of the contributions.

*We have also added the following to the main body of the text :-*

The results in the main body of the paper are based on meteorology and emissions from the period 28th March 2009 to 28th March 2010. It is possible that the results presented vary depending on the chosen year since meteorology, cloud fields, etc. vary from year to year. To address this we have also run the PI and PD simulations for an additional year for the period 28th March 2010 to 28th March 2011. We have compared selected key results from Section 3.2, which are shown in Appendix G. Very similar results are found using the alternative year, which demonstrates that our results are robust and not sensitive to the chosen year of meteorology.

As I said above I very much like that the authors try to evaluate (or at least speculate) the potential impact of model biases on simulated aerosol forcing. However, that discussion could be more complete. Could biases in $f_c$ affect adjustments in that variable – for example allowing $f_c$ adjustments in sky that should be overcast, or vice-versa? Could that be significant? Also, ari is easily masked by even moderately thick clouds, so biases in $f_c$ or LWP could translate in the wrong masking of ari. Is that important? A similar comment could be made about high cloud biases, as those biases would affect the amount of aerosol forcing that is masked by clouds above. Finally, the authors dutifully restrict their analysis to ocean regions, but land-based biases (which seem much larger than over ocean according to section 3.1.1 figures) likely matter for aerosol transport to ocean regions and their biases.

*We have added further discussion on this in Section 4.1. However, it is difficult to quantify these effects, so we leave this for future more detailed work on this. As for biases over land we decided not to focus on these since satellite retrievals are less accurate over land (or not possible for AMSRE), which would make interpretation more difficult. Also, the focus of our study is the North Atlantic oceanic region. Here is the revised part of Section 4.1 :-*

Low-altitude cloud fraction biases have the potential to significantly impact aerosol forcing since they are closest in altitude to the aerosol sources. If we assume that the PI cloud cover is biased by a similar amount to the PD cloud cover and assume no cloud adjustments to aerosol then the bias in forcing will be similar to the bias in $f_c$. The results from this paper suggest that the second assumption is reasonable for the northern NA region because the Twomey effect dominates. Thus we might expect the 5.1% low-altitude $f_c$ bias there to make a small contribution to any error in forcing bias. Of course it is also possible that the $f_c$ increase between PI and PD is underestimated by the model for this region, but the PI low-altitude $f_c$ is too high in order to give an overall PD bias. As such it is difficult to make firm conclusions.

In the southern NA $f_c$ changes in response to aerosols (adjustments) were large, so the above assumptions are less valid and the effects on forcing even less clear. The negative present-day $f_c$ biases could indicate a cloud fraction response to aerosol that is too low, which would cause subsequent negative forcing biases. On the other hand, the too-low $f_c$ may mean that the model PI era is in a broken, precipitating cloud regime too often. Such regimes are thought to be more sensitive to aerosols and more prone to produce cloud adjustments (Ackerman et al., 2004a). In this case the model forcing values would be too large. The effect of this may be significant given the large bias here (-23.9%). Likewise, the too-large $f_c$ values in the northern NA (if also occurring in the PI) might prevent some instances of $f_c$ increase between PI and PD and lead to a forcing that is



too small, although the overall bias for the northern NA region was only 5.1%. The presence of clouds can also mask ARI forcing and hence the $f_c$ and $LWP_{ic}$ biases might therefore affect the predicted $ERF_{ARI}$ magnitude. The larger low- and mid-altitude biases (which are likely to be thicker and have a stronger masking effect) in the southern NA combined with the larger $ERF_{ARI}$ suggest that this effect would likely be more pronounced there. Here the $f_c$ biases are negative, which would produce a positive $ERF_{ARI}$ bias by this mechanism. Mid- and high-altitude clouds can also mask $ERF_{ACI}$ forcing from low-altitude cloud (which is likely to be the biggest contributor to forcing). For both the northern NA and southern NA mid-altitude clouds tend to have negative biases that are larger in magnitude than the positive biases of the high-altitude clouds suggesting the potential for an overall negative $ERF_{ACI}$ forcing from this mechanism. However, further work would be needed to quantify the effect of the model $f_c$ biases on the aerosol forcing.

Line 5: Caution is of course also needed when interpreting high-resolution, process-resolving models!

*Agreed, but it is difficult to mention this here in the abstract. However, we note that we include this sentence in the Recommendations section :-*
*"The assumption would be that the cloud responses to aerosol of this would be more accurate than the global model resolution simulations, although its performance should be tested using the observations."*

Line 14: "further large increases in $f_c$" implies that aerosols can only increase, and not decrease, cloud fraction, at least on average over the regions studied. This is true in the model used in the present study, but not in the real world, so I would suggest rephrasing here.

*We feel that it is clear that we are talking about the "world according to the model" throughout this part of the abstract and so would like to keep this sentence as it is to avoid adding lots of caveats and making it less succinct.*

Lines 28-29: Why the sudden focus on the north of Scandinavia?

*This was the main region where LWPic biases were seen and it is still part of the North Atlantic region, so we thought it should be mentioned.*

Line 75: HadGEM2-ES was a CMIP5 model, wasn't it?

*Yes, thanks, this has been fixed.*

Line 84: Could give the resolution here, as "coarse resolution" for a given model may be medium or even high resolution for another.

*This has been added.*

Line 113, lines 148-150 and lines 164-165: The paper should say early (and in the abstract) that it only considers aci with a subset of liquid (not ice) clouds, but determining what that subset is seems complicated by the distinction between convective and large-scale clouds. Does the model carry two sets of cloud variables (especially water content and cloud fraction)? Are those two sets considered separately for cloud fraction and radiative purposes? Or is $f_c$ based on both types of clouds? Would it then follow that aerosols only affect an unknown fraction of the cloud field? What would that mean for linking cloud biases to forcing?

*The model does carry two sets of liquid water content and cloud fraction values – one for large-scale cloud and one for convective cores. As mentioned in Appendix C :-*

*"LWP associated with the convection scheme is only used by the radiation scheme for shallow convective clouds (clouds with geometrical depths less than 500 m over land and 1500 m over ocean) meaning that the majority of liquid water in deep clouds has no effect on radiation"*

*The cloud fractions used in this paper come from the COSP satellite simulator, which likely only uses the cloud that is considered by the radiation scheme. Hence the model bias analysis likely does not apply to the deeper convective core regions. However, convective cores make up only a very small fraction of the total cloud area and so would have little radiative impact anyway. Also, as also mentioned in Appendix C :-*

*"Liquid water content and vapour are detrained to the environment from the convection scheme and incorporated into the large-scale cloud scheme (see UM Documentation Paper 030; hereafter UMDP030; \url{https://code.metoffice.gov.uk/doc/um/vn11.3/umdp.html#030})."*

*This means that cloud associated with the convection scheme (that directly detrained or created due to the convective detrainment of humidity) does have a radiative impact and should be included in the COSP cloud fraction.*

*However, aerosols are not considered by the convective parameterization and for convection can only affect the cloud after it has been detrained. This is a limitation in the representation of aerosol forcing in this model (and many others) and is mentioned in the paper when talking about convective parameterizations :-*

*"The parameterizations do not take into account aerosol, or droplet concentrations and they use their own simplified microphysics scheme."*

*We have also added the following to the abstract :-*
*"This model does not include aerosol effects within the convective parameterization (but aerosol does affect the cloud associated with detrainment) and so it should be noted that the representation of aerosol forcing for convection is incomplete."*

*We also add this line to the main text in Section 2.2 :-*
*"The parameterizations do not take into account aerosol, or droplet concentrations and they use their own simplified microphysics scheme. Therefore the representation of aerosol forcing is incomplete."*

Lines 201-205: Should say here that the analysis of biases is limited to ocean surfaces.

*Done.*

Line 220: To evaluate whether aerosol forcing contributes to biases, one could probably identify regions where aerosol impacts on $f_c$ go in the same direction of the bias. Or you could repeat your bias analysis with the PI simulations. If it looks better against observations than PI, then aerosols must be to blame.

*We feel that we have addressed this point above and in Section 4.1.*

Lines 238-240: This is an important observation if one wants to link the present paper to Booth et al. (2012). Because of the definition of ERF, one needs to assume that forcing decomposition and cloud regime transitions are not affected by the coupling with the ocean.

*We have added this sentence :-*

*"This also implies that the coupled model and the nudged model used here exhibit similar cloud regimes and gives more confidence that the results in this paper apply to coupled models."*

Line 269: Clear regions do not really contribute to LWP. I suggest rephrasing.

*Thanks. This has been changed to :-*

*"Note that LWP from AMSR-E is the all-sky LWP and so includes the zero values from clear regions and the LWP contributions from cloudy regions"*

Line 409: "usually larger" – should be "smaller" I think.

*Thanks – we meant ERF_ARI here.*

Lines 424-429: It would be useful to clarify here that the hypothesis is reasonable because in the model aerosols only affect autoconversion rates. Alternative mechanisms that could potentially decrease liquid water content and/or cloud fraction, for example easier evaporation of smaller droplets or changes in above cloud air entrainment, are not represented in the model.

*We added a sentence here to make this point.*

*"We note that in reality clouds can also respond to enhanced aerosol by increasing cloud top entrainment, which can reduce LWPic and fc (Ackerman et al., 2004; Bretherton et al., 2007; Hill et al., 2009). However, this mechanism is currently not included in the model."*

Line 332: "exclude" rather than "prevent"?
*Changed.*

Figures 9 and 10: It would help to apply the same colour scale for panels (c) of both figures.

*We have applied the same colourscale for panel c of Figs. 8 and 9 since they are directly comparable. However, Fig. 10c is the contribution from just the cloud fraction bias and we would prefer to keep the colour scale as it is to show some of the detail.*

Caption of Figure 17 could note that category 1 is clear sky.

*Fixed.*

# Referee #2

Major comments
In this study, the aerosol radiative forcing is estimated with one-year model simulation. However, the surface ACI forcing is quite noisy over some regions, even a smooth is applied. The estimated ACI forcing could be from model internal variability other than aerosol cloud interaction. With the large internal noise, it is difficult to tell whether further findings of the manuscript are correct. Ensemble simulations could be a useful and simple way to estimate the uncertainty from the internal noise (Liu et al. 2018). Only the point where the estimated ACI forcing is statistically significant could be analyzed. The authors made detailed evaluation of the model simulated cloud properties against the observation. However, the simulated aerosol properties were barely mentioned in the manuscript. Please compare PD AOD with the observation. The changes in AOD from PI to PD should be also shown. More details could be found in the comments below.

*The point about only using one year of data and the results potentially being due to noise was also mentioned by Referee #1. We have run an additional year of simulation (with a different year of meteorology) and found almost identical results. Please see our response above for more details. We have added a reference to Liu (2018) also. We have included an AOD evaluation and PI to PD changes in the revised manuscript.*

Other comments
Line 100: Does the UKESM1 has the similar performance on global scale? Please make a comparison with the results of Mülmenstädt et al. 2019.

*We have added some text to describe the comparison to Mulmenstadt 2019.*

Line 115: Is it done in any previous studies? Please provide references here.
*Yes, this method has been used before. We added a reference to Seethala, JGR, 2020 (doi:10.1029/2009JD012662).*

Line 185: Similar methods were applied in previous studies (e.g., Ghan et al. 2012; Jiang et al. 2016), which should be mentioned here.
*Thanks, we added references to these papers.*

Line 197: The surface forcing could be decomposed. How about the TOA forcing?
*TOA forcing can also be decomposed using this method. We now mention this in the paper. We focus on surface forcing because of the interest in ocean forcing in the North Atlantic.*

Line 200: There are two many figures for this part. Please consider move some figures (e.g. middle and high cloud fraction) to the supplement.
*We would prefer to keep these figures in we feel that clouds of different altitudes should be considered. Referee #1 has also referred to them, for example indicating their importance for shielding lower clouds and masking ARI and ACI effects.*

Line 400: Please show PD AOD values and make a comparison with the observation.
*This is now included in the new manuscript.*

Line 400: Please show changes in AOD from PI to PD. The contribution from different aerosol types (sulfate, BC, dust and POM) should be also shown.

*We now show the PI to PD changes. However, we don't have information on the direct contributions from the different aerosol types to AOD (except for dust). Instead we show the changes in the column integrated aerosol mass mixing ratios for the different types.*

Line 415: Please show the cloud condensation nuclei (CCN) change together with other cloud properties.
*Unfortunately, we did not output this information and so cannot show this.*

Line 465: The surface ACI forcing due to LWP and fc is very noisy over the southern NA region. It implies the change could be from model internal variability other than aerosol-cloud-interaction.
*We feel that the addition of the results from the extra year of data (see above) shows that this is not due to noise since a similar result is found.*

Line 490: Are the different states classified with the annual mean value or instantaneous value? Are the estimated forcing values statically significant?

*Yes, the states are classified with the instantaneous data. We have added a sentence to make this clearer in the revised manuscript. The fact that the same result is obtained using the alternative year of meteorology indicates that the result is robust.*