We would like to thank both anonymous reviewers for their helpful reviews. Their comments have helped us to make changes which we believe significantly improve the manuscript. Below, we respond to each reviewer’s comments (italicised) in turn before providing a summary of the changes we have made to the manuscript.

Response to Referee 1

The paper deals with a very important question related to cloud aerosol analysis from satellite data. The authors seek to know what is the correct scale for a joint analysis of aerosol and clouds’ properties. On one hand in order to see if there is significant causality reflected by significant trends over the large natural variability, one would try to analyze datasets that are as large as possible and to increase the samples number. For that, one has to use long time series and preferably large spatial domains. On the other hand meteorology is a key player in the cloud-aerosol system and natural climatological trends (meteorology) may create apparent relations between aerosol and clouds that are not based on “cause and effect”. The authors compare two methods of trend analysis, one by using all the data in the box (region), collecting all times and all pixels together, and one that calculates regression (slope) for each pixel separately using the pixel’s time series and then average the slopes (weighted by their significance) in order to estimate the regional one. They compare the two methods and show how the results deviate as the box size gets larger. I think that the ideas presented in the paper are important and should be explored but I miss serious discussion on few key topics and found few inaccurate descriptions of references.

Thank you for your summary. Specific responses, including updated descriptions of the references, are detailed below.

Specifically: 1) The problem of meteorology vs. true aerosol effect is known to be important for quite a while. There are several approaches to tackle this problem. The approach that is described in this paper refers to gradients. If clouds and aerosols have significant gradients in their averages (meteorology) these gradients may create apparent correlations driven only by these gradients. If we will take as an example the Atlantic subtropics off the Saharan coast, we expect the aerosol loading to be higher near the African coast than downwind in the ocean. Gradients in cloud properties are also expected to occur as a function of the distance to Africa, as upwellung cold water in the east part defines the distribution of the marine stratocumulus. As the SST gets warmer and with the increase in the MBL height westwards the clouds gradually transform from close cells to open ones and later to sparse cumulus clouds near the American coast. This is all meteorology, but scatter plot between aerosol optical depth and cloud properties will show trends. In order to decouple meteorology form aerosol effect there were many attempts to slice the data per given meteorological conditions.

The reviewer is right to suggest that we neglected to point out some of the meteorolog-
A comment that meteorology has been ‘accounted for to differing extents by the aforementioned studies’ has been added to the paragraph beginning at 15419.21. In the Conclusions, it is now stated that ‘additional cloud type and meteorological constraints applied in several of the cited studies (which implicitly includes both those discussed in the Introduction and re-cited in the Conclusions and those cited in the Conclusions e.g. Koren et al. 2005, 2008) ‘may make their results less susceptible to the spatial gradient effects discussed here.’

2) By taking 10 years of data for all seasons (most of the paper’s figures) the authors do in time what they do not recommend to do in space. In theory seasonality may create very strong apparent correlations between clouds and aerosols. The authors should base their core analysis on shorter times within a given season.

Clarification has been added to the Results section and Fig. 2 caption that the analysis is done for individual seasons and the results are then averaged to find an annual mean. Of course, this will not completely remove temporal effects due to the annual cycle since climatological variation occurs within seasons. However, a detailed analysis of ‘temporal gradient’ effects is outside the scope of this study, although it presents similar problems to the spatial gradients focused on here. Mention of temporal gradients has been now been added to the end of the Conclusions: ‘One possible further approach would be to investigate the contribution of temporal climatological gradients, analogous to the spatial gradients discussed here. However, although seasonal time-scale choices should be considered, this would not fully account for meteorological effects. The development of more advanced methods to investigate the contribution of meteorology to observed aerosol–cloud relationships would be highly beneficial.’

3) Using shorter time intervals will reduce the significance of the pixel by pixel analysis dramatically. In such case most of the pixels may exhibit insignificant trends. In such case collecting data from larger boxes will be essential. And the discussion of how to reduce the meteorological variance mentioned in (1) will be relevant.
As can be seen in Fig. 2, statistical significance is not necessarily lost by calculating sensitivities at grid-box scales and then averaging to larger regional scales.

Mention of the potential prohibitions of analysing data at grid-box resolution has now been added to the Conclusions: ‘Of course, it may not always be possible to conduct analyses at the small spatial scales recommended here. Dataset limitations may prohibit this, particularly when extra temporal and meteorological constraints reduce dataset size. Potential spatial scale methodological errors should be considered alongside other considerations.’

The Conclusions conclude with a paragraph stating the importance of ultimately accounting for meteorological variation, and that this the basis for future work.

4) Following the previous points, the paper misses serious discussion on the one pixel analysis option. Is it always better? What are the weaknesses? Does the (weighted) average of the means of the trends reflect in the right way the regional trend? The authors compare the trends of the gridded analysis to the regional one assuming that the gridded is the correct one. Is it always true? Intuitively in places where there are no significant joined gradients, analysis of larger boxes should reflect the trends better.

Thank you for pointing this out. Further discussion about the potential weakness that the grid-method may bias sensitivities towards regions with a low error in the sensitivity has now been added to an Appendix and additional figure, as well as being mentioned in the updated Method and Conclusions sections.

The relevant section of the new Appendix reads, ‘Weighting by the one-sigma error when calculating \( b_{N_e|G} \) for regions larger than \( 1^\circ \times 1^\circ \) and multi-season means of both \( b_{N_e|G} \) and \( b_{N_e|R} \) has the potential to introduce a bias towards regions and seasons with a low one-sigma error. In order to demonstrate that this potential problem does not appear to be the major contributor to the region and grid method differences discussed in this paper, the final row of Fig. A1 shows annual mean \( 60^\circ \times 60^\circ \) \( N_e \) sensitivities calculated with no error weighting. As can be seen by comparison of the first and third rows of Fig. A1, the overall global picture remains similar.’

The updated Method section contains, ‘Although the grid-method has the obvious advantage of reducing spatial gradient methodological errors, the error weighting may lead to bias towards regions with a small error in the sensitivity. However, as will be discussed in Appendix A, this does not appear to be a major problem in this study.’

Mention that the existence of spatial scale errors ‘does not appear to be the result of a sampling bias due to error weighting’ has been added to the Conclusions.

Response to Referee 2

General comment:

In this article, the authors try to answer the following questions: What are sensible choices of spatial scale for aerosol-cloud interaction studies? What effect may spatial scale choices have on global estimates of radiative forcing due to the cloud albedo effect?

To answer those important questions, the authors used 10 years of Terra MODIS satellite product (cloud products and aerosol optical depth) to study the global cloud albedo effect. The authors calculated the cloud albedo effect for each season and using 2 different averaging methods within regions of different size to highlight the impact of the spatial scale. Maps of annually averaged cloud albedo effects for different spatial scales are shown in the paper. Variations of the methodological error due to spatial scale against the size of regions are also shown. Notably, the authors found that for regions larger than 4x4 degrees, methodological errors due to spatial scale become significant. For a region with a spatial scale of 60x60 degrees, the uncertainty is about 80%.

Thank you for summary and general comments.

While the effect of the spatial scale is clearly convincing, the results are only valid for satellite-based aerosol-cloud interaction studies without any constrain on cloud type,
The revised manuscript now contains more detailed consideration of constraints and limitations. Further details of these additions are provided in the responses to the specific comments below.

I recommend this paper for publication after addressing the comments below:

Specific comments:

In the introduction, the authors cited the results of previous studies on cloud-aerosol interactions and notably the size of the regions used. As mentioned by reviewer 1, the cited papers studied aerosol-cloud interactions on large regions but they reduced the uncertainty in those regions by selecting specific meteorological situations or specific type of clouds (shallow clouds, constant LWP). In this paper, the only constraint that I see, if I am right, is liquid water clouds. McComiskey et al. (2009) have shown that aerosol-cloud interactions is dependent on the spatial scale when inhomogeneous clouds are involved. The spatial scale dependence is reduced if the clouds are more homogeneous (by constraining the LWP for instance). So basically, all the results on the methodological uncertainty found in this paper are potentially incorrect for constrained data.

Further mention of constraints in the cited studies has now been added to the Introduction. It is now stated that Quaas et al. (2008) use clouds 'with a liquid water path \( w > 20 \text{ gm}^{-2} \)' and that McComiskey et al. (2009) 'consider different \( w \) and spatial resolution constraints'. As mentioned above, the possibility that adding meteorological constraints may make studies less susceptible to spatial gradient errors has been added to the Conclusions.

Further mention of constraints in the cited studies has now been added to the Introduction. It is now stated that Quaas et al. (2008) use clouds 'with a liquid water path \( w > 20 \text{ gm}^{-2} \) and that McComiskey et al. (2009) 'consider different \( w \) and spatial resolution constraints'. As mentioned above, the possibility that adding meteorological constraints may make studies less susceptible to spatial gradient errors has been added to the Conclusions.

The \( r_e \) values used in this study are from the quality assured dataset for all liquid water clouds. The \( N_e \) values have the further constraint that only clouds with \( r_e > 4 \mu m \) and \( \tau_c > 4 \) are considered. The \( \tau_c > 4 \) constraint implicitly removes clouds with a small liquid water path, discussion of which has now been added to the Method section:

'It is worth noting that, assuming vertically homogeneous \( r_e \), \( w \) is calculated as

\[
w = \frac{2}{3} \rho_l \tau_c r_e
\]

where \( \rho_l \) is the density of liquid water Platnick (2000). Therefore, removing clouds with \( \tau_c < 4 \) implicitly removes clouds with small \( w \) in the calculation of \( N_e \). For \( r_e = 10 \mu m \), this corresponds to excluding clouds with \( w < 27 \text{ gm}^{-2} \).

In an Appendix, the potential effect of adding a single-layer cloud constraint is now considered and found to be small: 'The \( N_e \) and \( r_e \) results presented in Sect. 3 and the first row of Fig. A1 are for all liquid clouds (i.e. no single-layer cloud constraint has been applied). For comparison, the annual mean \( 60^\circ \times 60^\circ \) \( N_e \) sensitivities shown in the second row of Fig. A1 use \( N_e \) values calculated from the single-layer cloud histogram. As can be seen by comparison of the first and second rows, although some of the details may change, the application of a single-layer constraint does not appear to have a significant effect here. In particular, the general global picture of \( b_N \mid |R - b_N| \mid G \) changes little.'

A second problem that has not been discussed in the introduction is the vertical distribution of aerosols. Within the same grid cell, an absorbing aerosol layer can have a different effect on a cloud if the layer is located far above, right above or mixed with a cloud. Depending on the vertical distribution, the cloud albedo effect can be negligible, positive or negative (respectively). Costantino and Breon (GRL 2010) showed that by restricting a statistical analysis to aerosol-cloud mixed cases, the cloud albedo effect is increased. Again, constraining the dataset affects the statistics.

A paragraph mentioning the problem of vertical distributions, with reference to
Costantino and Bréon (2010), has been added to the Introduction: ‘Satellite-observed aerosol and cloud may have different vertical distributions and may not actually mix. Using CALIPSO (Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observation) vertical profile data, Costantino and Bréon (2010) find that a much stronger correlation between PARASOL (Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar) \( r_e \) and MODIS aerosol index exists for mixed aerosol–cloud cases that for non-mixed layers in the Eastern South Atlantic stratocumulus region.’

Unfortunately, vertical resolution requires the use of active instruments (like CALIPSO) which currently lack the spatial coverage and accuracy of passive instruments (like MODIS and PARASOL).

The unconstrained method used in the paper makes the cloud albedo effect found in 1x1 degree regions (the reference scale in the paper) doubtful, as a given AOD value will be related to different vertical distributions and meteorological situations. It can potentially cause spurious relationships as aerosol (concentration and vertical distribution) and meteorology (transport pattern for instance) are potentially correlated, and cloud/meteorology are correlated. As the authors said (page 15426 line 10-13), aerosol types and cloud properties are known to vary spatially within large regions and variations may have a significant impact on observed aerosol indirect effects. I would like to add that in a 1x1 degree region (a small region), cloud properties and aerosol type/distribution vary throughout a season (the time scale used in the paper). So basically the uncertainty of the values within 1x1 degree regions could be very large. The authors should at least emphasize that the results in the paper are valid for unconstrained data only, and that constraining aerosol-cloud interaction statistics to a specific cloud type, meteorology or aerosol distribution would reduce the methodological uncertainty due to the spatial scale of the regions.

 Mention that ‘the additional cloud type and meteorological constraints applied in several of these studies may make [the cited studies'] results less susceptible to the spatial gradient effects discussed here’ has been added to the Conclusions.

Furthermore, the following sentence “For regions on the scale of 60x60 degrees, these methodological errors may lead to an overestimate in global cloud albedo effect radiative forcing of order 80%” is correct relatively to a 1x1 degree region only. We don’t know the uncertainty in the 1x1 degree region itself. Please, be more specific in the abstract and the paper.

Thanks for this. The abstract has been revised as recommended, by adding that the radiative forcing error is ‘relative to that calculated for regions on the scale of 1° × 1°’.

I suggest that the authors check if the relationship found between methodological uncertainty and region scale is robust by constraining, for instance, the LWP, cloud height, or some meteorological parameters within each region. The paper and its impact would benefit of it. The authors should at least constrain their data the same way as Quaas et al. (2008) (by excluding data with LWP < 20 gm⁻² and multilayered clouds) before inferring any methodological uncertainty in Quaas et al. (2008).

Thanks for this suggestion. As discussed above, further discussion of constraints used for \( \lambda_x \) has been added to the Method (implicit liquid water path filtering) and Appendix (single-layer clouds).

The application of detailed meteorological constraints is outside the scope of the current study and is the basis for future work, as mentioned in the concluding paragraph of the Conclusions.

**List of changes to revised manuscript**
Once again, the authors would like to thank both reviewers for their helpful comments on the manuscript. In light of their comments, the following changes have been made. Most of these have been quoted in the responses above.

15418.14: adding ‘relative to that calculated for regions on the scale of 1° × 1°’.
15418.16: hyphenating ‘present-day’.
adding 'for clouds with a liquid water path $w>20 \text{ gm}^{-2}$'.

They consider different $w$ and spatial resolution constraints.

They perform a multiple regression analysis to investigate the contribution of meteorology to this observed relationship.

Replacing '$r_e$ and $\tau_c$' with '$\tau_c$ and $r_e$' for clouds below 3 km'.

Satellite-observed aerosol and cloud may have different vertical distributions and may not actually mix. Using CALIPSO (Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observation) vertical profile data, Costantino and Bréon (2010) find that a much stronger correlation between PARASOL (Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar) $r_e$ and MODIS aerosol index exists for mixed aerosol–cloud cases that for non-mixed layers in the Eastern South Atlantic stratocumulus region.

Replacing ‘these observed relationships’ with ‘observed relationships between aerosol and cloud properties’; adding ‘accounted for to differing extents by the aforementioned studies’.

It is worth noting that, assuming vertically homogeneous $r_e$, $w$ is calculated as

$$w = \frac{2}{3} \rho_l \tau_c r_e$$

where $\rho_l$ is the density of liquid water Platnick (2000). Therefore, removing clouds with $\tau_c<4$ implicitly removes clouds with small $w$ in the calculation of $N_e$. For $r_e=10 \mu\text{m}$, this corresponds to excluding clouds with $w<27 \text{ gm}^{-2}$.

Identical sampling has not been applied to the quality-assured $r_e$ used in this study.

The results shown in Sect. 3 have not undergone a single-layer cloud constraint in the calculation of $N_e$. However, these results are relatively insensitive to the application of such a constraint, as will be shown in Appendix A.

In both methods, sensitivities with fewer than five contributing data points are excluded. Further significance testing is also applied, with sensitivities which are insignificant at the two-sigma level being shown as white in Figs. 2, 3, 4 and 5. Both the region-method and the grid-method assume that cloud and aerosol measurements for different grid boxes and days are independent, an assumption which may cause the one-sigma errors calculated in this study to be too small. The validity and effect of this assumption will be discussed further in Appendix A.

Although the grid-method has the obvious advantage of reducing spatial gradient methodological errors, the error weighting may lead to bias towards regions with a small error in the sensitivity. However, as will be discussed in Appendix A, this does not appear to be a major problem in this study.

The sensitivities are calculated for each season and then an error weighted annual mean is calculated.

These results for $b_{N_e|R}$, $b_{N_e|C}$ and $b_{N_e|R}-b_{N_e|C}$ are relatively insensitive to the the application of a single-layer cloud constraint, as will be discussed in Appendix A.

The existence of these spatial scale errors appears to be robust to the application of a single-layer cloud constraint and also does not appear to be the result of a sampling bias due to error weighting.

The possibility of analysing data at this higher resolution should be seriously considered.
Of course, it may not always be possible to conduct analyses at the small spatial scales recommended here. Dataset limitations may prohibit this, particularly when extra temporal and meteorological constraints reduce dataset size. Potential spatial scale methodological errors should be considered alongside other considerations.

However, the additional cloud type and meteorological constraints applied in several of these studies may make their results less susceptible to the spatial gradient effects discussed here.

Finally, one possible further approach would be to investigate the contribution of temporal climatological gradients, analogous to the spatial gradients discussed here. However, although seasonal temporal scale choices should be considered, this would not fully account for meteorological effects. The development of more advanced methods to investigate the contribution of meteorology to observed aerosol–cloud relationships would be highly beneficial.

(all seasons one-sigma error weighted).

Appendix and Fig. A1, briefly discussing the effect of applying a single-layer cloud constraint, error weighting choices and assumption of data independence.