Authors' answers to referee comments on the manuscript:
Aerosol indirect effects on continental low-level clouds over Sweden and Finland

By: Moa K. Sporre, Erik Swietlicki, Paul Glantz and Markku Kulmala

We would like to thank the reviewers for their helpful comments which we believe will greatly improve this manuscript. The referees' comments are presented in normal font and our responses are then presented in italic.

During the reprocessing of the data, owing to comments from the reviewers, we realized that we had used two minor criteria to screen out cases for the investigation that we had not mentioned in the paper. These two criteria are now included in Section 2.2 and 2.4. We also discovered a programming error in the criteria regarding the cloud fraction (Section 2.2) and the number of cases included in the study has changed from 116 to 122 for Vavihill and 252 to 261 for Hyytiälä. This does not alter the conclusions of the results but the values in the figures and tables are somewhat altered. Therefore, the text has in some places been slightly altered to fit the new figures.

Anonymous Referee #1

Sporre et al. have studied aerosol indirect effects on low-level clouds using ground based aerosol measurements and remote sensing data. The research topic is very interesting but, unfortunately, this manuscript does not bring anything new to it. All the conclusions of the manuscript have been published previously in other papers, some of them by the same authors. Furthermore, the methods used are not completely sound. For example, the authors base a number of their conclusions on correlations that are almost non-existent. Therefore, I suggests that the editor rejects this manuscript.

Although we certainly respect the opinion of the referee, recommending rejection, we obviously do not agree. Neither does the other referee, who recommends only minor changes to the manuscript.

The remark that "all the conclusions of the manuscript have been published previously in other papers, some of them by the same authors" is not valid, in our opinion.

There are a limited number of experimental studies on the effect of boundary layer aerosol loadings on low-level cloud microphysical properties. Those relevant here are mentioned in the manuscript, and all of them differ in some respect from our study. Our study combines ground-based aerosol data with satellite data of clouds and weather radar data of precipitation to investigate continental low-level clouds. Twohy et al. (2005) use in-situ aircraft measurements while Costantino and Breon (2013) use satellite measurements for both aerosol and cloud data, and both these studies assess marine clouds. The study by Lihavainen et al. (2010) use similar cloud and aerosol dataset as our study, but their focus is the ACI and no precipitation data is used.

.

The aerosol indirect effect on climate is of considerable importance and large uncertainties still remain, as expressed clearly in the recent IPCC AR5 WG1 report. The scientific community needs to make use of all possible methods to reduce these uncertainties, and to examine as many types of clouds in as many locations on Earth as we possibly can. If we consistently fail to observe the indirect effects of aerosol on clouds or continue having ambiguous results, can we then rely on the predictions made by GCM on the future cooling effect of aerosols? Are we starting to believe that the models are more reliable than observations?

Even though our present study makes use of the same four open access data sets (ACTRIS, MODIS, BALTEX, ECMWF) as in the previous Sporre et al (2014) study dealing with convective clouds, it is now focused on low-level clouds. Needless to say, these are two distinctly different types of clouds and certainly warrant separate studies. Our Sporre et al (2012) study examined marine Sc clouds, which are also markedly different from continental Sc clouds. Having examined one type of clouds surely does not mean that we have learnt all we need to know about every other cloud type. We have strived to be consistent in our scientific approach and use our tools to examine a range of study objects, in our case cloud types. This is definitely not uncommon practice in science. Indeed, systematic and careful examination of various phenomena using a given set of available tools (methodologies, models, data etc) is fundamental to scientific progress.

To the best of our knowledge, there is no other study that uses the same methodology as ours, applied on the same type of clouds.

When starting our work on the continental Sc clouds, we expected to see an aerosol effect on, not only the effective radius, but also on COT. Failure to observe an indirect effect of aerosols on COT is worth communicating to the scientific community, since the current understanding is that increased boundary layer aerosol particle concentrations will result in higher cloud droplet number concentrations and more reflective clouds, hence a climate cooling effect. Sc clouds are also considered to be more important for the global radiative balance than convective clouds, which display other responses (than COT) to aerosol loadings of consequence for climate.

The new results presented in this paper include:
Higher aerosol number concentrations are associated with smaller droplet effective radius and reduced precipitation intensity in low-level clouds. The COT on the other hand seems to be unaffected by aerosol number concentrations. We use a unique method where four open access dataset are combined for 10 years. This method has now also been applied to low-level clouds.

The other remark by the referee is that "the methods used are not completely sound. For example, the authors base a number of their conclusions on correlations that are almost non-existent."

One major results is that, given our data, apparently <u>no correlation</u> can be observed between the boundary layer aerosol particle concentrations (N_{130}) and COT. This is not the same as stating that the first indirect (Twomey) effect is non-existing, but rather that COT – and hence also cloud albedo – is affected by a number of other concurrent processes that effectively obscure the single effect (Twomey) under real-life conditions. What we observe is the net effect of all processes working on the continental Sc clouds, of which the first indirect effect is *just one. This does not mean that we have failed in our study. On the contrary, the net effect should be what the GCMs need to capture and properly account for in order to gain predictive capability. It is beyond the scope of this (observational) paper to elucidate the multitude of processes that affect the Sc clouds, but we certainly will continue our efforts to understand the whole picture, and describe the interlinked processes in models across a range of scales, from detailed cloud-resolved models to their GCM parameterizations and climate feedback mechanisms.*

More on the matter of correlations is found in reply to the same referee´s detailed comment regarding "significant correlations" following directly below.

Here are the detailed comments:
1) Significant correlation. This phrase is used a lot in the paper as a basis for the conclusions. However, when you look at the correlation coefficients they are in many cases below 0.35 which basically means that there is no correlation. Even though the p-values indicate significance, it doesn't mean that there is correlation between the parameters. The p-value just tells you how unlikely a given correlation coefficient will occur given no relationship in the population. The authors seem to have forgotten this. Futhermore, I'm not even sure if correlation is the best metric for this kind study. Maybe more advanced statistical methods should be used.

We are of course well aware of what the p-values indicate. As is common in inferential statistical analysis, we test the null hypothesis that there is no correlation between the dependent and independent variables, and the p-value indicates the likelihood that the variables are indeed uncorrelated, given our sampled data and assumed functional relationship (linear, logarithmic etc). The test provides no information as to the cause-effect relationship between our variables. Such information have to be derived from other observations or previous experience.

*If the sample size is large (many data points), then the correlation coefficient can be quite small and still indicate that the variables are very unlikely to be uncorrelated. True enough, this is not the same as showing that there is indeed a significant correlation between the dependent and independent variables, but the hypothesis that they are uncorrelated is shown to be very unlikely (for instance at the ** or *** confidence level).*

In all environmental data, the observed variables are influenced by a multitude of processes, many of which we are either unaware of or at least unable to control. This gives a large spread in the data, and large samples are required in order to be able to test for correlations. Clouds are no exception to this rule, and there are numerous examples in the literature of statistical inferences that have been drawn from noisy data sets. See for example Fig 3. McComiskey et al. (2009) and Fig 2. Garrett et al. (2004).

The referee comments further that "I'm not even sure if correlation is the best metric for this kind study. Maybe more advanced statistical methods should be used."

We have actually performed a multivariate statistical analysis to see whether we could gain additional more information on the covariances in our data that were somehow hidden. We carried out a PCA (principal components analysis) on the separate data sets for Vavihill and Hyytiälä. However, we could not observe any dependencies that were not seen in our analysis *of the correlation coefficients, as presented in the paper. Basically, the PCA results are evident from Tables 4 and 5 showing all correlation coefficients. These matrices also form the basis for the PCA analysis.*

In these tables, we see strong correlations (>0.9) between COT and LWP, as well as between CTT and SH. The third largest correlation is found between N_{130} and r_e. These are also evident from the three first PCA that have high loadings on the same combinations of variables, and in that order. When removing LWP and SH in an attempt to also remove correlations that are expected, then we RH and LTSS appear on PC1 while N_{130} and r_e are now found on PC2. These results are the same for both Vavihill and Hyytiälä. Our conclusion was that a PCA did not reveal any additional information on our data, but rather confirmed our interpretations. Therefore, we chose not to include this analysis in the paper.

We are careful when using the phrase significant correlation, and we do not see that this "is used a lot in the paper as a basis for the conclusions" as claimed by the referee. In most cases, we use the word "significant" when stating that there are in fact <u>no significant correlation</u> between two parameters, such is the case for N_{130} and COT.

The paper basically claims that correlations are high only between the following parameters (quotes from the paper):

- *"The r_e (at 3.7 µm) is best correlated with LWP and second best with N_{130} at both stations. The high correlation coefficients with LWP are expected since LWP is calculated from the r_e at 2.1 µm (section 2.2) which have similar values to the r_e at 3.7 µm."*

- *"LWP is also well correlated with COT since the LWP is calculated from the COT."*

We do not explicitly mention the high correlation between CTT and SH in the text (only seen in Tables 4 and 5), since this correlation is to be expected. It is nevertheless remarkable considering that two completely independent data sets were used to derive CTT and SH (MODIS and ECMWF respectively).

Both the LWP correlations are a result of the way that LWP is calculated from MODIS data. MODIS basically only retrieves two (nearly) independent (orthogonal) cloud microphysical parameters from the wave-length reflectance. Fortunately, these are two fundamental parameters: (i) effective radius, and (ii) COT. From these, one can estimate LWP and CDNC, but they are then dependent on r_e and COT, and will hence correlate strongly with both of them. Various assumptions also need to be made when estimating LWP and CDNC.

Other instances when we use "statistically significant correlations" are when discussing Figure 6 ("The solid markers in the figure denote statistically significant correlations at a 95% confidence interval.") and Figure 8 between precipitation rates (dbcz) and N_{130} ("Moreover, statistically significant correlations are obtained for Hyytiälä in both Fig. 8 a and b but only in Fig. 8b for Vavihill."). . "Significant correlation" is also used in the caption for Table 4: "The stars indicate at what confidence level (=95% **=99% ***=99.9%) the correlation coefficients are significant."*

We believe that these statements are valid as they are phrased now.

The following sentence has been changed:

"The only other parameter that is significantly correlated with COT at both stations is the RH (at 1000 hPa)." to "The only other parameter that may affect the COT is the RH (at 1000 hPa) (Tables 4 and 5)."

2) The data are from two sites (Vavihill and Hyytiälä) but at the end of several sections it is mentioned that the number of measurements from Vavihill are so small that you shouldn't trust those results. This raises the question, that why are those results presented in the first place?
We did this investigation for two separate sites (Vavihill and Hyytiälä) and have included both datasets because, on the whole, the results from the two stations agree quite well. For one part of the investigation (ACI Section 3.3) where the data is binned, the amount of data in some bins for Vavihill are not enough for the results to be completely reliable. We do not think this is reason enough to exclude the Vavihill dataset from the paper. Instead, we think it is useful to include Vavihill data since it shows that a large amount of data is required for a study of this kind.

3) The particle number concentrations were averaged for 5 hours and satellite data for 90x65 km2. Why did you choose to average so much data? Typically, when comparing ground based aerosol measurements with satellite retrievals you use 1 h and 50x50 km2 averages,respectively (see Ichoku et al. (2002) for more details). The basic idea is that the different instruments should have measured the same air masses. I'm not sure if that is the case in this study.
The areas were chosen so that we have a large amount of clouds to study and that the ground would be homogeneous, thus, it would not affect the satellite retrievals. We also wanted the areas for both stations to be of the same size. We chose to average the DMPS (Differental Mobility Particle Sizer) measurements over 5 hours because we believed that this would give aerosol number concentrations representative for the current areas. Since we are not comparing the ground based aerosol measurements to satellite measurements of aerosols (as is done in Ichoku et al. (2002)) the criteria for areas or time averaging do not need to be as strict. In previous studies (Freud et al., 2008;Ahmad et al., 2013;Janssen et al., 2011), also combining ground based aerosol and satellite cloud data, different sizes of investigation areas, both larger and smaller than the areas in our study, have been used. The authors in those studies also averaged over time intervals longer and shorter than ours. Moreover, we have also investigated how the chosen time interval affects the results. The results were not significantly altered if averaging over 1 hour or 3 hours around the time of the satellite passage was chosen. Averaging over even longer time intervals, such as 8-18 or 24 hours did not alter the results considerably either. We have added a sentence regarding this to Section 2.1: "Using longer (up to 24 hours) or shorter (down to 1 hour) time intervals did not alter the results considerably."

4) The results are compared to other studies but, usually, the comparison limits to a statement that others have got different kind of results. It would more useful for the readers if you could explain why the results are different.
We have added explanations to why the results in our study differ from previous studies. See comments p 12940, l 30, p 12942, l 10/p12943, l 27 *and p 12945, l 13.*

Technical comments: page 12933, line 5: Do "natural" aerosols have indirect effects?
The words "man-made" has been removed from the sentence.

p 12935, l 15: You mention in the inroduction that low-level clouds are affected by anthropogenic pollution but then you use background stations to study it. That sounds contradictory.
The current background stations are periodically affected by anthropogenic aerosols. In Asmi et al. (2011) they write about Vavihill: "It is well-suited for studies of the influx of polluted air from continental Europe to the Nordic countries along a south-north transect" and concerning Hyytiälä the following is written "The air masses are influenced by European pollution but at times very clean Arctic air is observed.".

p 12936, l 11: What was the temperature limit for the cloud screening?
The temperature limit was the lowest temperature below 1500m according to a temperature profile from the ECMWF for each day investigated. This is described in Section 2.4.

p 12936, l13: Distortion of what?
The satellite pixels. The sentence has been changed to "Images when the stations were situated as close to the middle of the satellite scene as possible were chosen to minimize the distortion of the pixels that occurs at the edge of the swath"

p 12936, l 17: What are the spatial resolutions of the used products?
The following sentence has been added to Section 2.2 "All MODIS products used here have a horizontal resolution of 1 km at nadir, with the exception for the CTT that has a resolution of 5 km at nadir".

p 12937, l 4: If Re (3.7) is 10 µm larger than Re(2.1), why does it mean that Re(2.1) is overestimated?
This sentence has been changed to "Pixels where r_e at 3.7 µm is more than 10 µm less than the r_e at 2.1 µm were also removed since this indicates an overestimation of the r_e at 2.1 µm and uncertainties in the r_e retrievals."

p 12937, l 5: What is the size and maximum amount of pixels in the scene?
The pixels have a 1 km horizontal resolution and the maximum amount of pixels in one scene is 5450 pixels at Vavihill and 5425 pixels at Hyytiälä.

p 12937, l 27: What is the maximum amount of pixels in the scene?
The maximum amount of pixels in the precipitation data is 1372 for Vavihill and 1354 for Hyytiälä.

p 12939, l 10: "only the ACI is calculated" –> the ACI is calculated only
This has been changed according to the reviewer's suggestion.

p 12940, l 30: What could cause the difference between these studies?
That the seasonal variation is missing in the current study is most likely because the study has less temporal coverage than the Asmi et al (2011) study. The following sentence has been added to the manuscript: "That the seasonal variation is missing in the current study is most likely because the study has less temporal coverage than the Asmi et al (2011) study."

p 12941, l 17: I wouldn't say that Re and RH are positively correlated if the correlation coefficient in only 0.25.
We refer to our answer to detailed comment no 1.

p 12941, l 19: Are the Re values also logarithmic in the figure?
No they are not.

p 12941, l 24: "The r2 values obtained ... are low due to large scatter.." Isn't this quite self evident?
We have removed the sentence.

p 12942, l 10/p12943, l 27: How did Jansen et al. (2011) measure Re and COT? Could different methods explain the differencies in the results?
Jansen et al. (2011) used Re and COT from MODIS just like we have done in this study. They have however used the Re at 2.1 µm, while we have used the one at 3.7 µm. The effective radius at the different wavelengths are usually well correlated though so we do not believe that this explains the difference in the results. Instead we believe that the different cloud screening methods discussed in paragraph 3 of Section 3 is causing the dissimilar results. We have added a sentence to Section 3.2 regarding this: "The dissimilar results are most likely due to the different data selection used in the two studies (discussed in Section 3). "

p 12943, l 1: The correlation coefficients between COT and Re vary between 0.24-0.34. The numbers are positive but I wouldn't call it correlation.
We refer to our answer to detailed comment no 1.

p 12943, l 3/p 12944, l 3: The correlation coefficients between COT and RH vary between 0.18-0.31. These parameters are not "significantly correlated", they are not correlated at all.
We have rewritten or removed these sentences.

p 12943, l 24: The correlation coefficients between LWP and N130 vary between -0.1 and -0.17. That does not indicate weak correlation, it indicates no correlation at all.
We have rewritten this sentence. It now reads: "There is however, only a weak negative correlation between N_{130} and LWP (Tables 4 and 5 and Fig. 3c) obtained here for Hyytiälä."

p 12944/Fig. 6: It would be easier to compare to the results by Lihavainen et al., if you would also have a line that is not binned by LWP.
We have added a line not binned by LWP to make the comparison easier.

p 12944, l 27: Shouldn't the number of data points affect the signicance of the correlation? In the Hyytiälä plot, the correlation is significant for the dark blue line (LWP:155-200) but not for the green (65-110) and red (20-65) lines when looking at the smaller cut-off diameters. Despite that you indicate that the red and green lines are more trustworthy than the dark blue line. Seems a bit contradictory.
When calculating the p-value, both the value of the correlation and the number of samples are used. Since the correlation for this LWP subset (LWP:155-200) is high (see Fig 6e) the p-value become lower than 0.05 even though the number of samples is low. Three subfigures showing the uncertainty of the ACI values has also been added to Fig 6 and error bars to Fig. 7. These show that the ACI uncertainty is higher for the intervals with higher LWP. The Section has been rewritten somewhat due to the changes in Fig. 6 and Fig. 7.

p 12945, l 9: Why doesn't the low number of measurements affect the significance in Hyytiälä?
See the answer to previous comment.

p 12945, l 13: What could explain the different results?
We have added a sentence regarding this: " Different cloud types (marine/continental), different cloud remote sensing (ground-based/satellite), different aerosol measurements (nephelometer/DMPS) or that (McComiskey et al., 2009) have calculated the ACI from COT may explain why the results are dissimilar."

p 12945, l 21: The highest correlation coefficient for dbcz is 0.28 which basically means that it is not correlated with other parameters.
We refer to our answer to detailed comment no 1.

p 12945, l 23: Please give the correlation values in the text.
We have rewritten this Section and now only correlations for the precipitating cases are included in Tables 4 and 5. See comment p.12945, L22 by Reviewer #2.

p 12947, p 1: How is this a surprise if you have already mentioned that Twohy et al. (2005) and Constantino and Breon (2013) found the same thing?
We have removed the word surprising from the sentence.

References:
Ichoku, C., D. A. Chu, S. Mattoo, et al., A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophys. Res. Lett., 29 (12), 2002
Interactive comment on Atmos. Chem. Phys. Discuss., 14, 12931, 2014.

Reviewer # 2

1. Line by line comments of scientific issues

p.12933, L10 - "Low-level stratiform clouds generally have low droplet number concentrations and are hence sensitive to changes in aerosol number concentrations."
This is a very general statement. The number of droplets in low level stratocumulus etc. is mainly dependent on where those clouds are and what aerosol is affecting them. The major stratocumulus decks on earth are next to landmasses, some of which are quite polluted. And so near the coasts their droplet concentrations are high and further offshore they generally get lower (e.g. SE Pacific). The only physical reason for stratocumulus having lower droplet concentrations than say deep convective clouds are their lower updrafts at cloud base, although I think it is a generalization to say that this leads to generally lower Nd, unless you can provide evidence for this from the literature.
This sentence has been removed and replaced by the two following sentences: "That different types of clouds have diverse effects on climate also introduces uncertainties. Low-level stratiform clouds mainly reflect incoming short wave solar radiation but do not affect the outgoing longwave radiation to any great degree (Boucher et al., 2013)."

p.12933, L26 - It would also be good to add the Bretherton (2007) paper after Ackermann (2004) since this paper expands upon the mechanism of how droplet size can effect LWP in stratocumulus. Citation:-

Bretherton, C. S., P. N. Blossey, and J. Uchida (2007), Cloud droplet sedimentation, entrainment efficiency, and subtropical stratocumulus albedo, Geophys. Res. Lett., 34, L03813, doi:10.1029/2006GL027648.
The reference has been added.

Aerosol measurements

p.12935, L19 – What is the end date for the data used?
We have rewritten the first sentence of Section 2.1 such that it now reads: "Data from ground-based aerosol measurement stations at Vavihill and Hyytiälä (Fig. 1) for the periods 2001-2009 and 2000-2009 respectively, have been used in this investigation."

p.12935, L27 – What is the justification of the value of 2.5 hours either side of the satellite overpass? Is there any sensitivity to this choice?
We chose 5 hours because we believed this would be a good time interval to get a representative aerosol loading for the entire area investigated in the satellite scenes. We have tested averaging over 1 hour, 3 hours, daytime (8-18) and for the 24 hour period but this did not alter the results to any great degree. We have added a sentence regarding this to the text: "Using longer (diurnal, daytime) or shorter (1hour, 3 hours) time intervals did not alter the results considerably."

L28 – What was the criteria used to decide whether there was too much variability in aerosol? This seems dubious to me as it seems likely that high variability would be associated with higher aerosol concentrations and thus this might bias the sampling. How many such days were removed and what happens if you put them back in?
The criteria used was that the standard deviation of N_{130} was to be smaller than the average N_{130} and this criteria was only applied to cases with N_{130} greater than 200 cm^{-3}. For Vavihill 5 days were removed due to this criterion and for Hyytiälä it was 3 days. The results are hardly affected at all if these days are included.

Satellite measurements

Some description of whether the quality flags for individual pixels were used to filter data or not - MODIS gives confidence flags for the optical depth and re retrievals – were these used?
No they have not been used.
Was any filtering for high solar zenith angle retrievals (SZA) done? For SZAs above ~65o COD and re retrievals are likely to be biased and should not be used – e.g. see Grosvenor, D. P. and Wood, R.: The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals, Atmos. Chem. Phys. Discuss., 14, 303-375, doi:10.5194/acpd-14-303-2014, 2014.
At such high latitudes SZA>65o would start being sampled from around October through to early March for the overpasses at the usual local time. And SZA would get much higher than 65 towards mid-winter (depending on the overpass time, etc.).
Therefore this has the potential to bias the results, especially when looking at seasonal cycles. At the very least this potential for bias should be mentioned, along with some investigation into how much of the dataset may be affected. Ideally, the results would be re-calculated without the high SZA data to gauge the effect.
We have not filtered away high solar zenith angles. It is true that such high zenith angles occur from approximately October to early March for the areas in this investigation. Since low-level clouds are more common during this time of the year 58% (Vavihill) and 57% (Hyytiälä) of the data used in the study have SZA >65°. Before submitting the paper we investigated the COT and the r_e and found none of them to systematically change with SZA. We have now also recalculated the results using only the cases with SZA lower than 65° and

found that the results regarding N_{130} and r_e and COT were not altered significantly (see Table comparing correlation coefficients (r) below). A paragraph has been added to Section 2.2 that discuss how the SZA can affect COT and r_e.

All SZA / Only SZA>65	*r: ln(N_{130}) to ln(r_e)*	*r : ln(N_{130}) to ln(COT)*
Vavihill	*-0.43 / -0.41*	*0.08 / 0.01*
Hyytiälä	*-0.56 / -0.56*	*0.06 / 0.06*

It would also be useful to examine whether there is any systematic change in the heterogentiy of the clouds as function of N or time of year. This is because cloud heterogeneity can also cause retrieval biases, which may affect the conclusions. One way to look at this could be through CTT variability or COT variability. See the Grosvenor paper above for discussion on this and for references.
We have calculated the CTT and COT variability according to the method presented in Grosvenor and Wood (2014). None of heterogeneity parameters were found to be affected by N_{130} or time of the year. This is also now briefly discussed in section 2.2.

p.12936, L28 – Was any CTT filtering done in order to remove the possibility of significant ice being present below cloud top (which may be possible even in cases where MODIS indicates a liquid cloud)?
No we did not do any filtering of CTT to exclude clouds which may have significant ice present below cloud top. Since MODIS only classifies the clouds as liquid or ice we have assumed that mixed phase clouds are a part of this investigation.

p.12937, L 4 – Should it be "less than re at 2.1um" and not greater? Also, why not use the absolute difference? A large difference of either sign would likely mean a large retrieval bias.
Yes it should be less than. This sentence has been changed to: "Pixels where r_e at 3.7 µm is more than 10 µm less than the r_e at 2.1 µm were also removed since this indicate an overestimation of the r_e at 2.1 µm and uncertainties in the r_e retrievals."

Results

p.12940, RE Fig. 2 :- There appear to be fewer samples when the air is from the south. Could this be due to the restriction on the aerosol variability that is imposed? Might this preferentially throw away some of the data when the air is from the south? The variability of N certainly looks higher for the southerly direction.
It would be good to show a graph of no. datapoints vs air mass origin and day or year. Or error bars maybe (but might look messy). Also, the sampling error will likely be higher for southerly direction due to the extra variability - is it too high to be meaningful?
We have added the number of cases from different directions and time of year as subfigures in Fig. 2. That there are fewer samples from the south is not due to the restriction on aerosol variability (only 8 cases were removed due to this criteria). Instead we think that the lower number of samples from the south could be due to that the meteorological conditions that exist when the air is arriving from that direction are not favourable for formation of low-level clouds. Since we are not singling out any particular wind direction we don't think that the higher variability from the south is a problem.

p.12940, L28 – Could the lack of seasonal variation in N (in contrast to the Asmi study) be due to a lack of samples in summer due to the lack of stratiform cloud? Could there be an

issue with the removal of scenes for which N is highly variable, since variability in N is likely to be associated with higher N?
Yes we believe that the lack of samples in summer could be the reason for the lack of seasonal variation in the current study. The following sentence regarding this has been added to the text: "That the seasonal variation is missing in the current study is most likely because the study has less temporal resolution than the Asmi et al (2011) study." Only 8 scenes were removed due to aerosol variability so we do not believe this to be the problem.

p. 12941 – RE Fig. 3a.
Since there is no restriction to constant LWP here it is hard to decide what to make of the relationships between re and N100. LWP changes with N are apparent and will affect the value of re and the change of re with N. Thus, much of the differences between the different observations could be due to this. This should be mentioned more prominently. A plot of LWP vs N should also be shown to go with the re and tau plots.
Since we have no independent measure of LWP we have not restricted the relationship between re and N100. None of the studies we are comparing with have however restricted their comparison according to LWP either. We have added a sentence regarding that different results between the studies could be caused by different LWPs: "Dissimilar LWP of the clouds in the different studies could also be a factor explaining why the slopes of the relationships differ." A plot of LWP vs N has also been added to Fig 3.

p. 12942, L8 – "re does not vary much with season"
Although, again, some of that could be due to LWP variation (and potentially a lack of samples in the summer). There may therefore still be a seasonal cycle in droplet concentration.
We do not see any variation in r_e with season in this dataset and we do not want to speculate whether there is a seasonal variation in droplet concentration.

L21 – Should refer the reader to tables 4&5 as evidence that higher N is correlated with lower re in the present study.
This paragraph was written as a summary of Section 3.2, but we have decided to remove it and only summarize the results in the conclusions.

L26 – "In addition, Fig. 3b shows that COT is more or less independent of N130." + associated discussion in Section 3.2.
I don't agree entirely with this. Fig. 3b shows that there is a small increase in tau with N130 and that the variability of COT increases with N130. Fig. 3b shows that higher N130 values are more likely to be associated with higher COT values (e.g. this would be seen in PDFs at low and high N130). It almost looks like there are two branches in the scatter points – one with little COT variation and one with a reasonable amount. Anyway, since the data is not separated by LWP we would expect some of the change in COT with N to be due to the reduction in LWP that is evident from the reduction in re (and from the correlations in the tables). This is alluded to later, but needs a bit more work.
Since the increase in COT with N_{130} is non-significant and the correlations are very low we have drawn the conclusion that COT is more or less independent of N_{130} Reviewer #1 has cautioned us from drawing conclusions from low correlations that are significant. We have added changed the sentence such that it now reads: "In addition, Fig. 3b shows that COT is more or less independent of N_{130} but the variability of COT seem to increase with N_{130}."
The expected increase in COT for the observed re decrease over the range of N should be calculated for an idealized cloud (e.g. a cloud for which the LWC content is assumed

adiabatic and for which Nd is constant in height – see for example Bennartz, JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, D02201, doi:10.1029/2006JD007547, 2007). The observed re range will indicate that a large increase in COT is expected. However, this behavior is very sensitive to the slope of the re vs Nd curve. A fairly modest increase in the slope (say to one slightly larger than that seen in the Twohy study) would be enough to reduce the expected change in COT over the N range to something close to that seen in Fig. 3b. Given that the measurement of re is quite uncertain and the r2 values for the fits are fairly low, such a discrepancy in the slope is perhaps not unreasonable. This would lead to the conclusion that the tau vs N curve is within the realm of what might be expected.
To calculate the COT from the observed re from an idealised cloud we could rearrange eq 4 from Bennartz (2007) $LWP = \frac{2}{3}\rho_l * COT * r_e$ *and we would need to use the LWP from MODIS. The LWP from MODIS is calculated using the formula* $LWP = \frac{4}{3*Q(r_e)} * COT * r_e$ *and for most pixels Q(*r_e*) is close to 2. Since* ρ_l *is set to 1 these formulas are almost identical. Hence, using the LWP calculated from the latter equation to calculate a COT from the former equation seem like circular argument to us. We therefore do not want include a discussion regarding expected COT in the manuscript.*
It would be good if this was investigated more thoroughly in the paper – e.g. can you estimate the sampling uncertainty for each N100 bin and the retrieval uncertainty from MODIS re and follow that through to give a range of likely COT responses to N100? It is likely that this range is very large meaning that estimation of the COT indirect effect through observations is very difficult and uncertain.
The relative uncertainty in the aerosol number concentrations are in the order of 10%. There is no retrieval uncertainty for the re3.7 provided by MODIS but the average retrieval uncertainty for re2.1 is never greater than 17% for the cases in this study. This range is not very large and we believe that an estimation of the COT indirect effect can be done using observational data.

What are the uncertainties in the fits for Fig. 3a and 3b (you give them for Fig.8, why not for Fig. 3)?
We have added the uncertainties for the fits in Fig 3.

This discussion could be placed alongside the discussion on p. 12943, L25. Considering the uncertainty, the reduction in LWP might wholly explain the lack on change of COT. p.12944 L1-5 – this paragraph overstates the case, as described above, and so should be toned down or removed.
This paragraph was written as a summary of Section 3.3 but we have decided to remove it and only summarize the results in the conclusions. We have also removed the conclusion at the end of Section 3.3 and 3.4 to be consistent.

p.12944, L7-10 - just because re2.1 and COT are used to calculate LWP it does not mean that studying the variation of these quantities with N100 (which is fully independent) within constant LWP bins is a bad idea. Besides re3.7 and re2.1 are likely to be very well correlated (except when the retrievals are dubious). Therefore I don't see why you don't examine the COT effect for constant LWP bins. It would help to sidestep some of the issues listed above RE LWP variation with N.
We have in a review process with a previous manuscript been asked to remove analysis of ACI values in constant LWP bins for re2.1 and COT because LWP is calculated from these two parameters. Hence, we don't have an independent measure of LWP and we have therefore decided to only perform the LWP binning on the re3.7, since this is not directly

related to the LWP. We know that there are different opinions in the scientific community regarding LWP binning of satellite data to calculate ACI, but binning the 3.7 r_e is as far as we are willing to go. We do however agree with the referee that the re3.7 and re2.1 are very well correlated.

Figs. 6 & 7 – Can you please calculate the uncertainty in the ACI values and put as error bars on the plot (or show some other way)? You give R2, but the uncertainty would be very useful to know also as this would take into account low numbers of samples, etc.
We have added error bars to Fig. 7. In Fig. 6 however error bars made the figures too messy and we instead added 3 more subfigures that show the uncertainty of the ACI. The test in Section 3.3 has also been slightly rewritten to account for this change.

p.12945, L22 – "If only precipitating cases are included"
It seems to me that this should be the case anyway. The dbcz values for non-precipitating cases are likely to be very low and highly uncertain. It would seem more sensible to restrict to above a threshold dbcz for all correlations.
We have now only included precipitating cases in Tables 4 and 5. Section 3.4 and the figure captions have been changed to clarify this. Section 3.4 has also been somewhat rewritten to account for the changes in the correlations in Tables 4 and 5.
Also, have the authours investigated whether using calculating a precipitation rate from the dbz would be better than using dbcz directly? Maybe this would be more directly related to the phenomena of interest.
We have tried to calculate a precipitation rate from the dbcz. However, since low-level clouds usually produce quite low precipitation rates we found that this only condensed the data and we therefore prefer using dbcz.
p.12945 and Fig. 8 – Have you tried doing the same plot for dbz vs re? Since droplet size is likely a better indicator of whether a cloud is precipitating or not.
We tried to do this and dbcz increases with increasing r_e for both stations, but the correlations are not significant. In order to keep the manuscript from becoming lengthy we have chosen not to include this in the paper.
Or, why not partition the data into constant LWP bins as for Figs. 6 & 7, since LWP will also be a key determinant of whether a cloud is precipitating.
We have partitioned the data into constant LWP bins, but the fits for the different LWP bins were not separated from each other and the correlations were all insignificant.
p.12946 and Fig. 9 – It should be reiterated here what the definition used to dividie precipitation and non-precipitating cases is. It is also necessary to try out other thresholds. -30dbz is very low – what happens if you use say -15, or -10?
We have written the definition for precipitation in the caption of Fig 9. Pixels that are not precipitating are set to – 30 in the BALTEX dataset and therefore we used this threshold. We agree with the reviewer that this might be low and have therefore changed it to -10 instead. The results in Fig. 8 are somewhat altered and we have removed the line for the combined dataset and added uncertainty to the Vavihill and Hyytiälä dataset to make the layout the same as for Fig. 3. The last row in Table 6 that concerned the relationship for the combined data has also been removed. The combined dataset no longer decrease with increasing N_{130}, since the Vavihill and Hyytiälä relationships are shifted with respect to each other. The text in Section 3.4 has also been somewhat rewritten.

Conclusions – these should be altered bearing in mind the above discussion.

The conclusion has been somewhat altered due to the changes in the results regarding precipitation. We have however not changed the discussions to such a degree that the conclusions should be altered.

2. Technical corrections

p.12934, L9 – "clouds" _ "cloud"
This has been changed according to the reviewer's suggestion.
L14 – "to find whether" _ "to find out whether"
This has been changed according to the reviewer's suggestion.
p.12937, L21 – "do not" _ "does not"
This has been changed according to the reviewer's suggestion.
L28 – "values is set to" _ "values are set to"
This has been changed to "value is set to"
p.12938, L14 – "1500" and "sort out" _ "select"
The text has been changed to "1500 m to select clouds".
p.12940, L3 – "particle" _ "particles"
This has been changed according to the reviewer's suggestion.
L9 – "correspond" _ "corresponds"
This has been changed according to the reviewer's suggestion.
p.12941, L1 – "amount" _ "number"
This has been changed according to the reviewer's suggestion.
p. 12843, L17 – geometrically thinner than what? Comma should be after "while" instead of "study".
This has been changed according to the reviewer's suggestion.
L 18 - "Small cloud droplets and suppressed precipitation associated with higher CCN concentrations has also in cloud simulations been shown to enhance entrainment into the clouds, leading to a reduction in the LWP (Ackerman et al., 2004)."
_ "In cloud simulations, small cloud droplets and suppressed precipitation associated with higher CCN concentrations have also been shown to enhance entrainment into the clouds, leading to a reduction in the LWP (Ackerman et al., 2004)."
This has been changed according to the reviewer's suggestion.
p.12944, L8 - "hence not" _ "hence is not"
This has been changed according to the reviewer's suggestion.
p.12945, L8 – "also have peak" _ "also peak"
This has been changed according to the reviewer's suggestion.
L9 – "by low number" _ "by the low number"
This has been changed according to the reviewer's suggestion.
p. 12946, L1 – Split the sentence at "however".
This has been changed according to the reviewer's suggestion.
L3-4 – "concentrations is" _ "concentrations are" + comma after "clouds"
This has been changed according to the reviewer's suggestion.
Table 4 – "The logarithm has not been applied to the CTT and dbcz parameters since these contain mainly negative values." - although there seem to be values listed?
The values listed for the CTT and dbcz are the correlation between the logarithm of the other parameters and the CTT and dbcz. We have however removed the correlation between the CTT and dbcz since this value is the same as that in the upper-right part of the Tables.

Fig. 6 – needs letter markers (a-f).
We have added the markers

Fig. 7 – you need to explain what the filled symbols mean.
An explanation concerning the filled symbols have been added to the figure caption.

References

Ahmad, I., Mielonen, T., Grosvenor, D. P., Portin, H. J., Arola, A., Mikkonen, S., Kühn, T., Leskinen, A., Juotsensaari, J., and Komppula, M.: Long-term measurements of cloud droplet concentrations and aerosol-cloud interactions in continental boundary layer clouds, Tellus B, 65, 2013.
Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A. M., Sellegri, K., Birmili, W., Weingartner, E., Baltensperger, U., Zdimal, V., and Zikova, N.: Number size distributions and seasonality of submicron particles in Europe 2008–2009, Atmos. Chem. Phys, 11, 5505-5538, 2011.
Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, J. Geophys. Res, 112, D02201, 2007.
Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S. K., Sherwood, S., Stevens, B., and Zhang, X. Y.: Clouds and Aerosols. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Presss, Cambridge, United Kingdom and New York, NY, USA., 2013.
Costantino, L., and Breon, F. M.: Aerosol indirect effect on warm clouds over South-East Atlantic, from co-located MODIS and CALIPSO observations, Atmos Chem Phys, 13, 69-88, 2013.
Freud, E., Ström, J., Rosenfeld, D., Tunved, P., and Swietlicki, E.: Anthropogenic aerosol effects on convective cloud microphysical properties in southern Sweden, Tellus B, 60, 286-297, 2008.
Garrett, T. J., Zhao, C., Dong, X., Mace, G. G., and Hobbs, P. V.: Effects of varying aerosol regimes on low-level Arctic stratus, Geophys Res Lett, 31, -, doi: 10.1029/2004gl019928, 2004.
Grosvenor, D., and Wood, R.: The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals, Atmospheric Chemistry and Physics Discussions, 14, 303-375, 2014.
Ichoku, C., Chu, D. A., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanre, D., Slutsker, I., and Holben, B. N.: A spatio-temporal approach for global validation and analysis of MODIS aerosol products, Geophys Res Lett, 29, 2002.
Janssen, R. H. H., Ganzeveld, L. N., Kabat, P., Kulmala, M., Nieminen, T., and Roebeling, R. A.: Estimating seasonal variations in cloud droplet number concentration over the boreal forest from satellite observations, Atmos Chem Phys, 11, 7701-7713, 2011.
Lihavainen, H., Kerminen, V. M., and Remer, L. A.: Aerosol-cloud interaction determined by both in situ and satellite data over a northern high-latitude site, Atmos Chem Phys, 10, 10987-10995, 2010.
McComiskey, A., Feingold, G., Frisch, A. S., Turner, D. D., Miller, M. A., Chiu, J. C., Min, Q. L., and Ogren, J. A.: An assessment of aerosol-cloud interactions in marine stratus clouds based on surface remote sensing, J Geophys Res-Atmos, 114, D09203, 10.1029/2008JD011006, 2009.

Twohy, C. H., Petters, M. D., Snider, J. R., Stevens, B., Tahnk, W., Wetzel, M., Russell, L., and Burnet, F.: Evaluation of the aerosol indirect effect in marine stratocumulus clouds: Droplet number, size, liquid water path, and radiative impact, J Geophys Res-Atmos, 110, D08203, doi: 10.1029/2004jd005116, 2005.