Review of Sporre et al. (2014), ACPD.

This paper attempts to identify aerosol effects upon stratiform cloud over two land sites using the combination of ground based aerosol measurements, cloud properties retrieved from satellite and ground based radar. It is useful to attempt to observe the predicted effects of aerosols upon clouds from observations alone in order to help evaluate models, for example. As such this study makes a somewhat useful contribution. However, it also demonstrates the difficulty in observing such processes due to sampling uncertainties and evident negative correlations of cloud Liquid Water Path (LWP) with aerosol concentration. I feel that these are a little downplayed in this paper and need to be looked at more closely. It is never really discussed what the reason for the reduction in LWP with N is. Correlation with relative humidity seems to be ruled out, which leaves other meteorological factors, or the possibility that there is an aerosol related effect on LWP. Has this been looked at for this region before? A bit more discussion on this would be nice.

A decrease in effective radius (re) occurs with increasing aerosol, but only a small increase in optical depth (COD). However, it is never demonstrated how much optical depth effect would be expected from an "idealized" cloud, especially given that there is a clear decrease in LWP with N (which also should be plotted). There also needs to be a more thorough evaluations of the uncertainties in the approach, and whether they can account for the observed lack of increase in COD with N. Satellite observations can have large uncertainties associated with them, which could strongly affect, for example, the slope of the observed re curve in the re vs N plot. This uncertainty, combined with those in the retrieval of COD, could allow for a consistent picture between the observed changes of re and COD with N, and this needs to be investigated and discussed in more detail. I can also see little reason why the change of COD with N is not investigated within constant LWP bins as it is for re, since this would help to sidestep some of these issues. The reason for not doing so given in the paper is that re2.1 and COD are used to calculate LWP and LWP is not independent of these parameters. This does not make sense to me, especially since you do look at re3.7 within LWP bins; re at 3.7um is likely to be highly correlated with re at 2.1um.

Calculation and discussion of the sampling uncertainties also needs to be done, since some seasons and air sectors appear to be less well sampled than others. The same goes for the calculations of ACI for which it is stated that some LWP bins have low numbers of samples.

My other major comment is that there is little discussion on the uncertainties and problems inherent in MODIS satellite cloud retrievals. It needs to be discussed that there are known biases at high solar zenith angles (SZA) that are likely to affect this high latitude dataset. It would be good to test whether removing high SZA retrievals affects the conclusions made. Similarly, cloud heterogeneity can also cause biases, which should at least be mentioned if they are not going to be investigated. Whether or not any quality assurance flags were used to discard pixels based on the COD and re retrieval confidence should also be mentioned.

These points are all discussed in more detail within the comments below:-

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.

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.***

Aerosol measurements

p.12935, L19 – What is the end date for the data used?

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?

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?

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?

Was any filtering for high solar zenith angle retrievals (SZA) done? For SZAs above ~65^o 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>65^o 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.

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.

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)?

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.

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?

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?

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.

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.

L21 – Should refer the reader to tables 4&5 as evidence that higher N is correlated with lower re in the present study.

L26 – "In addition, Fig. 3b shows that COT is more or less independent of N_{130}." + 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.

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 r^2 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.

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.

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

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.

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 R^2, but the uncertainty would be very useful to know also as this would take into account low numbers of samples, etc.

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. 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.

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. 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.

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?

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

2. Technical corrections

p.12934, L9 – "clouds" → "cloud"

L14 – "to find whether" → "to find out whether"

p.12937, L21 – "do not" → "does not"

L28 – "values is set to" → "values are set to" p.12938, L14 – “1500” and “sort out” → “select”

p.12940, L3 – “particle” → “particles”

L9 – “correspond” → “corresponds”

p.12941, L1 – “amount” → “number”

p. 12843, L17 – geometrically thinner than what? Comma should be after “while” instead of “study”.

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).”

p.12944, L8 - “hence not” → “hence is not”

p.12945, L8 – “also have peak” → “also peak”

L9 – “by low number” → “by the low number”

p. 12946, L1 – Split the sentence at “however”.

L3-4 – “concentrations is” → “concentrations are” + comma after “clouds”

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?

Fig. 6 – needs letter markers (a-f).

Fig. 7 – you need to explain what the filled symbols mean.