We thank the referees for their constructive comments that have helped to improve the manuscript. The issues raised are addressed individually below, including revised text where required.

Responses to Referee#1.

P 335, L 6-9: I share the authors' concern that the variability of 5 km-resolution cloud top temperature product may not be a good indicator of radiatively important small scale cloud variability. This seems especially important, as it may have led to the puzzling observation that the solar elevation dependence does not increase with cloud variability (which seems to weaken the variability-hypothesis considered throughout the paper). Therefore my main suggestion for improving the paper is to test different indicators of cloud variability to capture small-scale cloud variability, for example using the variability in 250 m reflectance (e.g., Di Girolamo et al. 2010, Zhang and Platnick 2011) or in 1 km brightness temperature (e.g., Varnai and Marshak 2002). If this wasn't possible, I recommend prominently pointing out this issue as soon as the first puzzling results appear in Figure 11, and mentioning it wherever the findings of Figures 11 and 12 are discussed.

Unfortunately, the assessment of the heterogeneity using the sub-1km reflectances is not possible with our current dataset since the information required is only available from Level-1 data. The Varnai and Marshak (2002) method is quite involved and cannot be implemented in time for the revisions to this paper. We hope to look at these different methods of assessing heterogeneity in a future paper.

However, we have examined what happens if we use the heterogeneity factor (Υ_τ) described in Cahalan (1994), which uses the variability of 1km optical depth (τ) and thus assesses variability at a smaller scale than the 5km resolution cloud top temperature data used in our study. We find that for effective radius the results are very similar to those presented in our Fig. 12. for both low and high SZA. The results are also similar for optical depth at low SZA. However, for high SZA optical depth shows a different response to Υ_τ than it did for σ_{CTT}; at low Υ_τ(indicating more τ homogeneity) there is a small difference in τ between low and high SZA, with the difference increasing with Υ_τ (larger τ for higher SZA). This is what we expect if the SZA dependence increases with heterogeneity.

We have added 3 new figures that relate to this issue along with corresponding discussion to section 3.4.1 of the ACPD paper. Extra discussion has also been added to sections 3.4.2 and 3.4.3 and new results are also referred to where necessary in the discussion. The new figures are included below, after the responses to the referees. Here is the amended text (highlighted in yellow):-

4.4.1 Cloud heterogeneity effects on optical depth

Figure 11a shows mean τ as a function of σ_{CTT}, at low θ values of $< 41.4°$ for both low and high θ_0. Figure 11b shows the τ difference between high and low θ_0 vs σ_{CTT}. In the lower range of σ_{CTT} ($<\sim 0.625$–$0.875\,\mathrm{K}$) τ increases as σ_{CTT} decreases for both low and high θ_0. The increase is much larger for high θ_0 (58 % increase between $\sigma_{CTT} = 0.875$ and $\sigma_{CTT} = 0.125\,\mathrm{K}$) than for low θ_0 (an increase of 27 % over the same range). At higher σ_{CTT}, τ is approximately constant within the error range. It is evident that the increase in τ between low and high θ_0 occurs at all values of σ_{CTT}. However, the increase is greatest at low values of σ_{CTT}, i.e. when the cloud tops are more homogeneous.

These results are surprising as previous work (Loeb et al., 1997; Varnai and Davies, 1999) has suggested that a "bumpy" cloud top was the most likely explanation for the increase in τ with increasing θ_0. If that were the case then it might be expected that τ would increase with increasing σ_{CTT} at high θ_0, that the τ increase with θ_0 would be greater at higher σ_{CTT}, and that at low σ_{CTT} there would be little difference in τ between low and high θ_0 cases.

One possible explanation is that sub-pixel variability is causing τ decreases, as suggested by M06 and Z12, and so this may be counteracting the expected increase due to resolved scale heterogeneity. Another possible explanation is that the actual (i.e. as opposed to the retrieved) τ of the clouds was higher at lower σ_{CTT}. Physically higher τ values at low σ_{CTT} might be expected to lead to a greater τ bias between low and high θ_0 (Loeb and Davies, 1996, 1997; Loeb and Coakley, 1998), as seen in Fig. 11. This seems likely to be factor given that an increase of τ with decreasing σ_{CTT} was observed at low θ_0, where our results indicate that θ_0 related biases should be small.

However, other factors are also likely at play and are now discussed through the examination of the effect of using γ_τ as a measure of cloud heterogeneity (see section 3.2). This parameter has the advantage that it is calculated using 1 km resolution τ data and so can capture variability at smaller scales than σ_{CTT}, which uses 5 km data. The disadvantage is that τ is a retrieved quantity and so γ_τ is subject to heterogeneity that is introduced through retrieval errors rather than representing solely physical cloud heterogeneity. CTT values are also retrieved and so may also suffer some heterogeneity biases. However, these are likely to be significantly less than those for τ retrievals.

Fig. 12 shows that at low θ_0 τ varies with γ_τ in a similar way to how it varies with σ_{CTT}. However, in contrast to when σ_{CTT} was used as a measure of heterogeneity, there is little increase in τ between low and high θ_0 for the lowest heterogeneity values. For high θ_0 there is also a fairly monotonic increase in τ with γ_τ over the lower range of the γ_τ values sampled. This is interesting since for γ_τ θ_0 biases therefore increase with heterogeneity, which would be the expected result if 3D radiative effects played a role in causing the θ_0 biases.

We now examine the relationship between γ_τ and σ_{CTT}. Fig. 13 shows the 2D histogram for these two parameters for both low and high θ_0. It shows that at low θ_0 (Fig. 13a) there is a lot of scatter with both low and high γ_τ values occurring for the higher σ_{CTT} range. The correlation coefficient in this case is only 0.24. From the figure it appears that there are two branches in the scatter of the data; one for which γ_τ increases rapidly with increasing σ_{CTT} and one for which there are only small increases in γ_τ. We have examined this plot for smaller ranges of viewing angles and relative azimuth angles and found broadly the same result, indicating that the scatter is not caused by variation in viewing geometry. Thus the results are suggestive that, at low θ_0, there is variability in the 1 km resolution radiative field (as captured by γ_τ) that is not predicted well by the physical cloud top height variability from 5 km resolution data (as captured by σ_{CTT}).

Fig. 13b shows the same result at high θ_0. This broadly shows only a single relationship between σ_{CTT} and γ_τ with considerably larger values of γ_τ for a given σ_{CTT}. Thus there is less scatter and a higher correlation coefficient of 0.45. Fig. 14 shows the mean γ_τ values for each bin of σ_{CTT}. The results are binned by σ_{CTT} since it was shown in Fig. ?? that this does not change much between low and high θ_0. In general there is an increase in γ_τ with increasing σ_{CTT} at both low and high θ_0. However, for a given σ_{CTT}, γ_τ is larger at high θ_0 showing that the increase in θ_0 has induced an increase in radiative heterogeneity. The greater degree of correlation between σ_{CTT} and γ_τ at high θ_0 indicates that physical cloud top variability as diagnosed from 5 km data is more representative of 1 km resolution radiative variability than at low θ_0.

However, considerable scatter still remains suggesting that other factors, such as physical cloud top variability at smaller scales than those captured using 5 km data are important. Extinction variations inside the cloud (without cloud top height variability) could also play a role, although this was found to have a small effect in Loeb et al. (1997) and Varnai and Davies (1999). Further work is needed to elucidate the relative merits of these explanations, which is beyond the scope of the observational dataset used in this study.

4.4.2 Cloud heterogeneity effects on effective radius

Figure 15a and b show r_e for the different wavelengths vs. σ_{CTT} at low and high θ_0, respectively. Note that the results shown here for r_e are very similar whether σ_{CTT} or γ_τ is used as a measure of heterogeneity. The figure shows r_e values that decrease with increasing σ_{CTT} (i.e. increasing cloud top heterogeneity) for all wavelengths. However, $r_{\mathrm{e}3.7}$ experiences the largest decrease and $r_{\mathrm{e}1.6}$ experiences only small changes. At low σ_{CTT}, $r_{\mathrm{e}3.7} > r_{\mathrm{e}2.1} > r_{\mathrm{e}1.6}$, which is actually what would be expected given the increased penetration depth of the shorter wavelength bands relative to the longer wavelength ones and an assumed increase of droplet size with height (e.g. see Platnick, 2000). The contrast to the usual MODIS observation of $r_{\mathrm{e}3.7} < r_{\mathrm{e}2.1} < r_{\mathrm{e}1.6}$ (e.g. Zhang and Plantnick, 2011) raises the possibility that the latter is caused by cloud top heterogeneity and that for homogenous cloud tops (at low θ_0) the r_e retrievals are more reliable and less prone to artifacts. Again, though, we have to bear in mind the possibility of physical cloud changes with σ_{CTT}.

The high θ_0 results follow a similar pattern with a larger r_e decrease with increasing σ_{CTT} for $r_{\mathrm{e}3.7}$ and $r_{\mathrm{e}2.1}$ compared to $r_{\mathrm{e}1.6}$. In fact, in the lower range of σ_{CTT} ($< 0.6\,\mathrm{K}$) $r_{\mathrm{e}1.6}$ actually increases slightly with σ_{CTT}. The convergence of $r_{\mathrm{e}1.6}$, $r_{\mathrm{e}2.1}$ and $r_{\mathrm{e}3.7}$ at the lowest σ_{CTT} value is probably fortuitous and likely due to the trends with σ_{CTT} of the different wavelength r_e values. Such convergence also occurs in Fig. 15a, although at a higher σ_{CTT} value. The difference can likely be put down to the effect of θ_0 since Fig. 3c suggests that the low and high θ_0 clouds would be physically similar at a given σ_{CTT}.

Additionally, the r_e values at high θ_0 are generally lower than, or similar to, those at low θ_0

for any given σ_{CTT}, with the differences being considerably greater for $r_{e3.7}$ and $r_{e2.1}$ than for $r_{e1.6}$. The relative lack of change of $r_{e1.6}$ with θ_0 and σ_{CTT} again raises the possibility that this wavelength might be less susceptible to r_e artifacts caused by cloud top heterogeneity at high θ_0. It also might be an argument against physical droplet size variations with σ_{CTT}. For the other wavelengths, the decreases in r_e between low and high σ_{CTT} are large, with the maximum decrease being 4.3 µm (35 %) in the case of $r_{e3.7}$ at high θ_0. Given the sensitivity of N_d to r_e this is likely to have a large impact on the retrieved N_d.

Earlier it was mentioned that the changes in r_e with heterogeneity were similar at both low and high θ_0 whether measured by σ_{CTT} or γ_τ. This is likely to only be possible if the two parameters are correlated and if r_e changes with one parameter generally act in the same direction as with the other. Therefore it seems that γ_τ explains little extra variability in r_e compared to σ_{CTT}. This in contrast to the situation with τ for high θ_0 (but not for low θ_0).

4.4.3 Cloud heterogeneity effects on droplet concentration

Similar plots to Fig. 15, but for N_d, are shown in Fig. 16a and b. Interestingly, in the low θ_0 case, at low σ_{CTT}, N_d values for all 3 wavelengths are very similar and there is little variation with σ_{CTT}. There is an increase and divergence amongst the wavelengths at higher σ_{CTT}, although the error bars also get larger. The increases from the lowest to highest σ_{CTT} value are 25, 40 and 71 % in the $r_{e1.6}$, $r_{e2.1}$ $r_{e3.7}$ cases, respectively.

For the high θ_0 case, N_d values are higher than for low θ_0 for any given σ_{CTT} value as expected from the τ and r_e results and from the results of Sect. 4.2.3. As for at low θ_0, though, N_d is similar for the three wavelengths at low σ_{CTT} and there is little variation of N_d with σ_{CTT}. However, compared to at low θ_0, N_d from the different wavelengths diverge at a lower σ_{CTT} and at high σ_{CTT} they diverge more widely and produce much higher N_d values. Although, again, the error bars are large at high σ_{CTT} due to a lack of samples. The increases in N_d between the lowest σ_{CTT} value and $\sigma_{CTT} = 2.6$, where the maximum N_d occurs, are 19, 69, 117 % for the $r_{e1.6}$, $r_{e2.1}$ $r_{e3.7}$ cases, respectively. Thus at both low and high θ_0 the changes in N_d are smaller for $r_{e1.6}$.

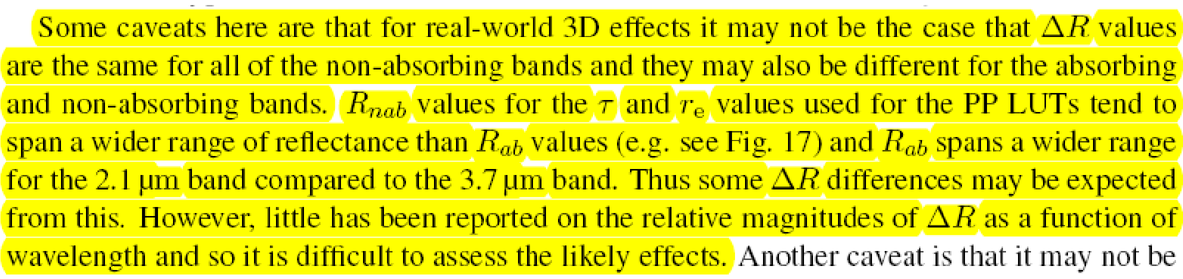
It is interesting that at both low and high θ_0 there is little change in N_d with σ_{CTT} for low σ_{CTT}, as well as little difference between N_d from the different wavelengths. The constant N_d is due to the cancellation of an increasing τ and increasing r_e as σ_{CTT} decreases. Since we might expect retrievals to be less prone to retrieval artifacts at low σ_{CTT}, the increase in τ with decreasing σ_{CTT} might suggest that the more homogeneous clouds are actually physically thicker with a corresponding higher τ and higher r_e, and thus that the τ and r_e changes are physical rather than due to retrieval artifacts. Also, it is feasible that N_d might be the same for homogeneous and heterogeneous clouds if the aerosol supply was similar for both cases, which would be consistent with the above result. However, heterogeneity is also known to be associated with increased precipitation and thus an increased CCN sink and might also be associated with altered updraft speeds, which would alter N_d activation. Shedding further light on this is difficult, however, without further observations of the clouds in question.

For low θ_0, when using γ_τ as the heterogeneity parameter the results are similar to those using σ_{CTT}, as would be expected from the similar variation of τ and r_e with both σ_{CTT} and γ_τ. At high θ_0 the lower τ values at low γ_τ (and high τ at high γ_τ) cause N_d to increase monotonically with γ_τ (not shown).

P 334, L 15-16: I fully agree with the authors’ statement, and would even guess that 3-D effects absorbing and non-absorbing wavelengths are more likely to have different than identical magnitudes. For example, the relative effects may be larger at absorbing wavelengths, while the absolute effects may be larger at non-absorbing wavelengths.

It would help to expand this discussion and include some references. If needed, the assumption and discussion could be expanded to other scenarios (e.g., larger 3-D effects at absorbing or non-absorbing wavelengths).

We agree that the changes in reflectance due to 3D effects may be different for absorbing and non-absorbing wavelengths and have added this to the text too. Unfortunately, there is little in the literature that has assessed the differences in dR between absorbing and non-absorbing bands, nor for different absorbing wavelengths. Thus we cannot provide references to aid the discussion and this also makes it difficult to justify ranges of dR to test using our approach. Therefore we will leave this to future work, but have noted this problem in the revised text:-



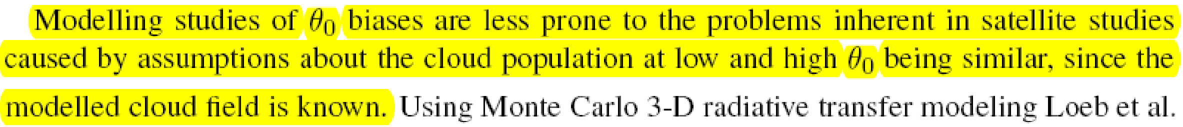
Some caveats here are that for real-world 3D effects it may not be the case that ΔR values are the same for all of the non-absorbing bands and they may also be different for the absorbing and non-absorbing bands. R_{nab} values for the τ and r_e values used for the PP LUTs tend to span a wider range of reflectance than R_{ab} values (e.g. see Fig. 17) and R_{ab} spans a wider range for the 2.1 µm band compared to the 3.7 µm band. Thus some ΔR differences may be expected from this. However, little has been reported on the relative magnitudes of ΔR as a function of wavelength and so it is difficult to assess the likely effects. Another caveat is that it may not be

Appendix D: The paper does a very good job at presenting thorough discussions about a wide range of considerations, but this results in a fairly long article. I believe some shortening would benefit the manuscript. For example it may be sufficient to mention Latin Hypercube Sampling only briefly, as Appendix D concludes that its results were not too different from a simple analysis of mean values.

We have shortened Appendix D considerably and have removed a lot of the detail, including table A1. We have also shortened the manuscript in other places following the suggestions of Referee #2 – please see the responses to Ref #2 for details on this.

P 310, L 6-8: The reasoning or wording here is not clear to me, as plane-parallel relationships are based on modeling, not on empirical correlations.

This section has been shortened and this part has been changed to:-



Modelling studies of θ_0 biases are less prone to the problems inherent in satellite studies caused by assumptions about the cloud population at low and high θ_0 being similar, since the modelled cloud field is known. Using Monte Carlo 3-D radiative transfer modeling Loeb et al.

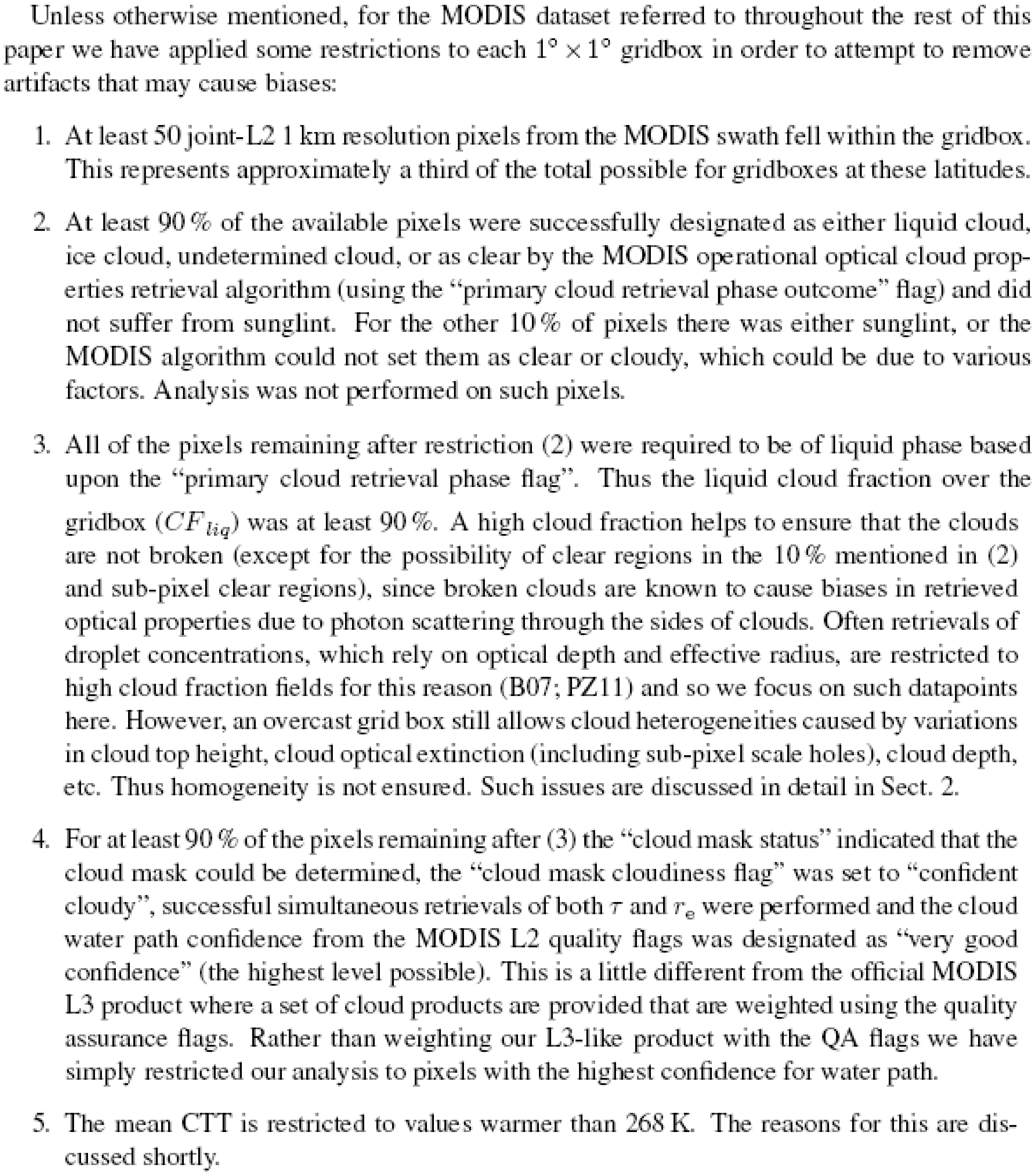
P 319, L 5: It would help to clarify whether the analysis used quality assessment flags included in the MODIS cloud product. (For example the multi-layer cloud flag may help reduce the effects of overlying ice clouds.)

When considering all pixels, we used the sunglint flag (p319, line 11) to avoid pixels that may be affected by this. Liquid phase pixels were selected using for pixels for which the "primary cloud retrieval phase outcome" indicated that a successful phase determination was made and for which the "primary cloud retrieval phase flag" indicated liquid water cloud.

For optical depth and reff retrievals the "cloud mask status" was used to select only pixels for which the cloud mask could be determined. The "cloud mask cloudiness flag" was also used to select only pixels that were designated as "confident cloudy". As mentioned at the top of p. 320 we use the Water path confidence QA flag to select

only pixels with “very good confidence”. The water path calculation depends on both optical depth and effective radius and therefore accounts for QA in both quantities. MODIS L3 provides a L3 cloud retrieval products that use weighting based upon the QA flags and a retrieval that does not use them. Rather than weighting our L3-like product with the QA flags we have simply restricted our analysis to pixels with the highest confidence. We did not use the multi-layer cloud flag and unfortunately it would require re-processing of the data to include this, which was not possible in time for this response. However, we note that, as explained in the text, a large number of other steps were taken to help avoid situations which could bias the retrievals.

These details have been added to the text to clarify which flags were used in the analysis:-



Unless otherwise mentioned, for the MODIS dataset referred to throughout the rest of this paper we have applied some restrictions to each 1° × 1° gridbox in order to attempt to remove artifacts that may cause biases:

1. At least 50 joint-L2 1 km resolution pixels from the MODIS swath fell within the gridbox. This represents approximately a third of the total possible for gridboxes at these latitudes.

2. At least 90 % of the available pixels were successfully designated as either liquid cloud, ice cloud, undetermined cloud, or as clear by the MODIS operational optical cloud properties retrieval algorithm (using the “primary cloud retrieval phase outcome” flag) and did not suffer from sunglint. For the other 10 % of pixels there was either sunglint, or the MODIS algorithm could not set them as clear or cloudy, which could be due to various factors. Analysis was not performed on such pixels.

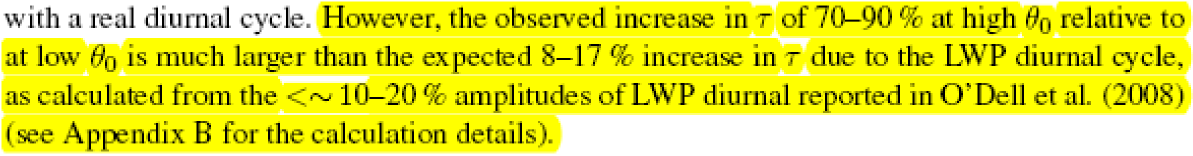
3. All of the pixels remaining after restriction (2) were required to be of liquid phase based upon the “primary cloud retrieval phase flag”. Thus the liquid cloud fraction over the gridbox (CF_{liq}) was at least 90 %. A high cloud fraction helps to ensure that the clouds are not broken (except for the possibility of clear regions in the 10 % mentioned in (2) and sub-pixel clear regions), since broken clouds are known to cause biases in retrieved optical properties due to photon scattering through the sides of clouds. Often retrievals of droplet concentrations, which rely on optical depth and effective radius, are restricted to high cloud fraction fields for this reason (B07; PZ11) and so we focus on such datapoints here. However, an overcast grid box still allows cloud heterogeneities caused by variations in cloud top height, cloud optical extinction (including sub-pixel scale holes), cloud depth, etc. Thus homogeneity is not ensured. Such issues are discussed in detail in Sect. 2.

4. For at least 90 % of the pixels remaining after (3) the “cloud mask status” indicated that the cloud mask could be determined, the “cloud mask cloudiness flag” was set to “confident cloudy”, successful simultaneous retrievals of both τ and r_e were performed and the cloud water path confidence from the MODIS L2 quality flags was designated as “very good confidence” (the highest level possible). This is a little different from the official MODIS L3 product where a set of cloud products are provided that are weighted using the quality assurance flags. Rather than weighting our L3-like product with the QA flags we have simply restricted our analysis to pixels with the highest confidence for water path.

5. The mean CTT is restricted to values warmer than 268 K. The reasons for this are discussed shortly.

P 326, L 6: A range of 10-20% may sound more realistic. Also, mentioning the reference for this expectation here could help readers even if it was mentioned earlier.

We decided to quote 8-17% because this was calculated from the 10-20% LWP diurnal cycle observed in O'Dell (2008). We feel that rounding up to 10-20% would cause confusion as it would be the same as for the LWP range. We have re-written the text here to read:-

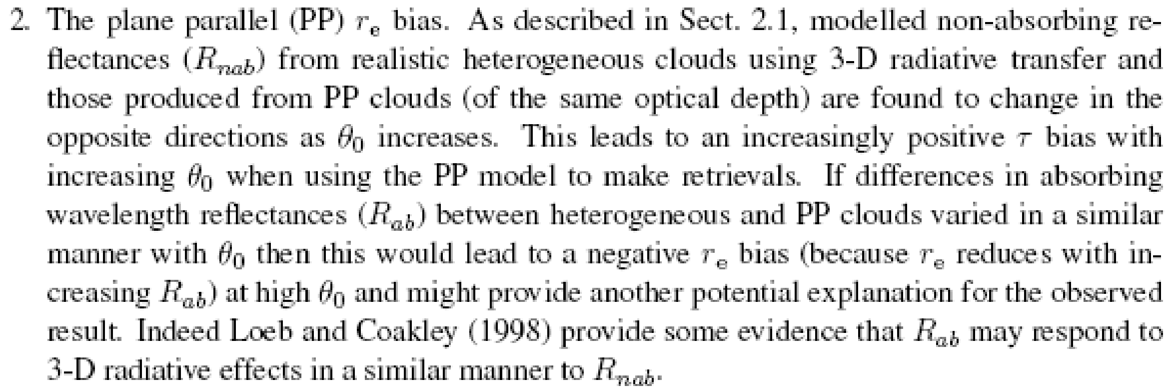


with a real diurnal cycle. However, the observed increase in τ of 70–90 % at high θ_0 relative to at low θ_0 is much larger than the expected 8–17 % increase in τ due to the LWP diurnal cycle, as calculated from the $<\sim$ 10–20 % amplitudes of LWP diurnal reported in O'Dell et al. (2008) (see Appendix B for the calculation details).

We have also removed the description of this from p.325, lines 18-20 to avoid repetition.

P332, L 17-25: This paragraph appears to combine plane-parallel bias (that is, variability within a 1-D framework) with 3-D issues. It would help to make the wording clear or to change the paragraph heading.

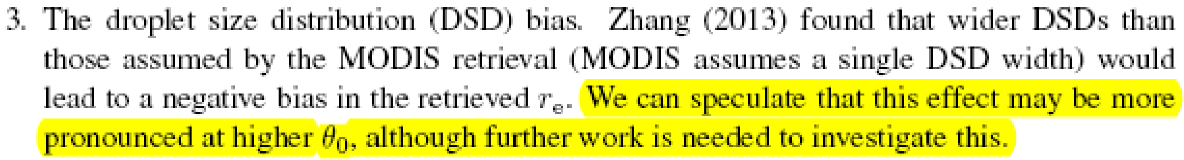
We have altered the text to make this more clear:-



2. The plane parallel (PP) r_e bias. As described in Sect. 2.1, modelled non-absorbing reflectances (R_{nab}) from realistic heterogeneous clouds using 3-D radiative transfer and those produced from PP clouds (of the same optical depth) are found to change in the opposite directions as θ_0 increases. This leads to an increasingly positive τ bias with increasing θ_0 when using the PP model to make retrievals. If differences in absorbing wavelength reflectances (R_{ab}) between heterogeneous and PP clouds varied in a similar manner with θ_0 then this would lead to a negative r_e bias (because r_e reduces with increasing R_{ab}) at high θ_0 and might provide another potential explanation for the observed result. Indeed Loeb and Coakley (1998) provide some evidence that R_{ab} may respond to 3-D radiative effects in a similar manner to R_{nab}.

P 332, L 27-29: To make the argument complete, it would help to mention what changes in the width of drop size distributions could cause the solar elevation dependent changes in retrieved re.

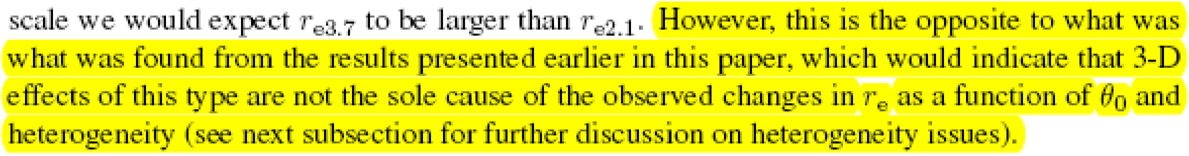
We have added a sentence to mention possible SZA effects:-



3. The droplet size distribution (DSD) bias. Zhang (2013) found that wider DSDs than those assumed by the MODIS retrieval (MODIS assumes a single DSD width) would lead to a negative bias in the retrieved r_e. We can speculate that this effect may be more pronounced at higher θ_0, although further work is needed to investigate this.

P 334, L 13: I suggest changing the wording "observed retrieved values", as it sounds awkward. Also, I suspect some typos or wording mix-ups in this sentence, as 3-D effects cannot cause retrieved values.

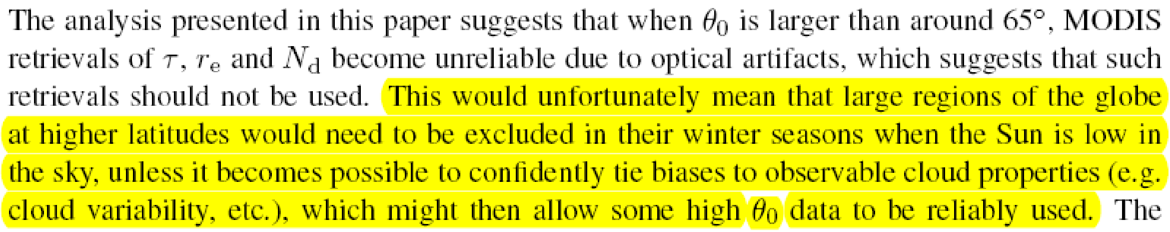
This has been changed to:-



scale we would expect $r_{e3.7}$ to be larger than $r_{e2.1}$. However, this is the opposite to what was what was found from the results presented earlier in this paper, which would indicate that 3-D effects of this type are not the sole cause of the observed changes in r_e as a function of θ_0 and heterogeneity (see next subsection for further discussion on heterogeneity issues).

P 338, L21 and P 339, L 11-16: While excluding suspicious data (at high solar zenith angles) may be a very good approach for eliminating retrieval biases at high latitudes, it seems worth mentioning that other approaches might also become possible in the future if the biases could be tied to cloud variability (or other factors) in a definite manner. Finally, it may also be worth mentioning whether the findings are relevant only to MODIS or to other datasets as well.

This has been added here:-



The analysis presented in this paper suggests that when θ_0 is larger than around 65°, MODIS retrievals of τ, r_e and N_d become unreliable due to optical artifacts, which suggests that such retrievals should not be used. This would unfortunately mean that large regions of the globe at higher latitudes would need to be excluded in their winter seasons when the Sun is low in the sky, unless it becomes possible to confidently tie biases to observable cloud properties (e.g. cloud variability, etc.), which might then allow some high θ_0 data to be reliably used. The

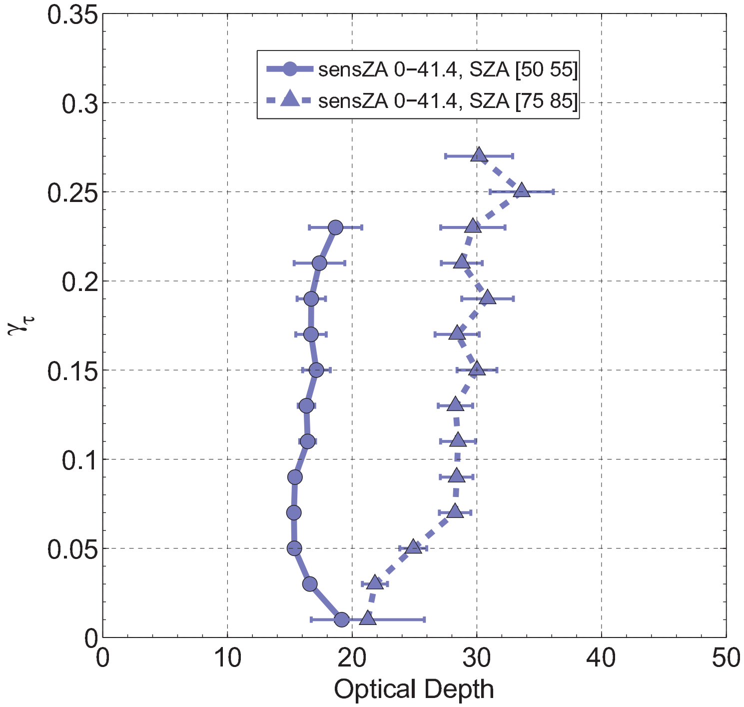
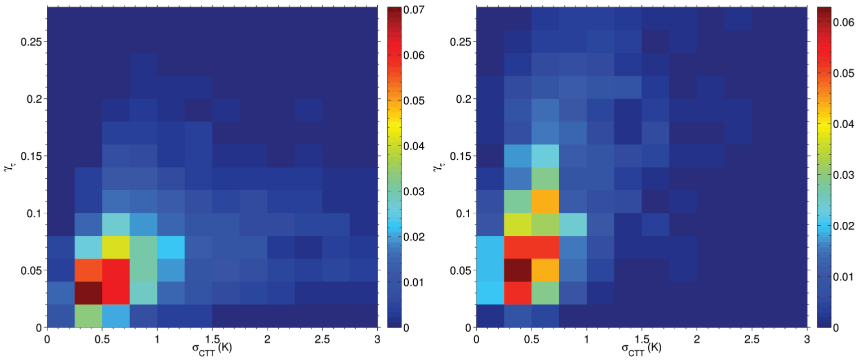


Fig. 12. As for Fig. 11a except for optical depth vs. γ_τ, where γ_τ is a measure of cloud heterogeneity based on the variability of the retrieved 1 km cloud optical depth. Low values of γ_τ indicate more homogeneity.



(a) SZA 50-55°.

(b) SZA 75-81.4°.

Fig. 13. 2D histogram of γ_τ vs. σ_{CTT} for low (a) and high (b) θ_0 ranges.

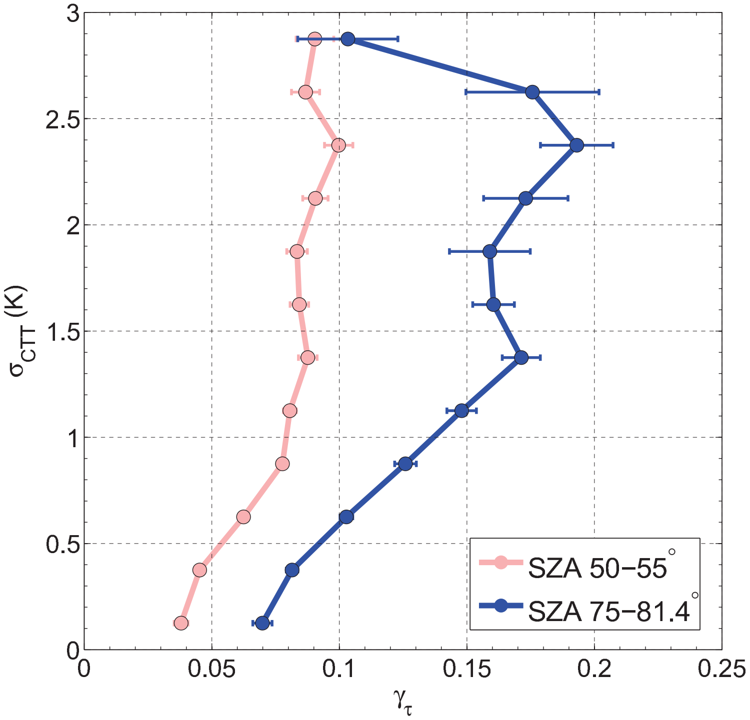


Fig. 14. Mean γ_τ for each σ_{CTT} bin from Fig. 13.