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 #2

Title: The paper only deals with marine liquid water clouds. This should be reflected in the title.

The title has been changed to "The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals within marine liquid water clouds"

P304, L9-11: From Figs. 5 and 6 it appears that tau, r_e and N_d are all essentially constant up to 70 degrees rather than 65 degrees. The value of 65 degrees appears throughout the paper, and should be consistently changed. This is important, because the 'quick reader' will mainly remember this value, and may then distrust more data than justified from the results in this paper.

We feel that the optical depth and Nd results show increases above 65 degrees, but that SZA=70 degrees might be better suited for the re changes. Thus we have changed the figure to 65-70 throughout the manuscript.

P304, L14: I would say the changes in r_e are not 'somewhat' but 'an order of magnitude' smaller than in tau.

We have changed the word to "considerably".

P305, L22: It should be made clear at the beginning of the paper that the effective radius is retrieved at the cloud top.

"effective radius" has been replaced with "cloud top effective radius"

P314-315, Sect. 2.2.3: The authors choose the large-scale (5 km) variability of cloud top temperature as a measure of inhomogeneity. This is a valid choice, but I would very much like to see also the sub-pixel (250 m) variability in reflectances included as well. This metric was used in several previous studies (e.g. Zhang and Platnick, 2011) and it seems more relevant for explaining artifacts in 1-km cloud property retrievals. This would be especially interesting since some of the results obtained using the CTT-based metric are counter-intuitive, i.e. mainly the decrease in tau with increasing sigma_CTT in Fig. 11.

We agree that it would be good to have an analysis of the sub-1km (250m) variability as presented in Zhang and Platnick (2011) and other studies. However, our dataset was built from Level-2 data that did not include this parameter. In order to include it we would have to process the Level-1 data, which unfortunately has not been possible within the timeframe of this paper. Thus we leave this to a future study. However, we have repeated the analysis using the variability of 1km optical depth data and have added significant amount of discussion on this - please see the response to the comments from Referee #1 for more details on this..

P318, Sect. 2.3.2: More details should be provided on how the MODIS 1x1-degree dataset has been generated. First of all, what has been done differently compared to the official MODIS L3 product? I guess the fact that different overpasses are kept separate? Are there more differences? Please also briefly explain what the sub-sampling involves, and what 'joint-L2' means. State which collection has been used. Have MODIS quality flags been applied?

We have added some detail about the quality assurance flags used to compile the dataset and some information on how this varies from the official Level-3 product (as far as we can ascertain). Please see the response to Referee#1 for more details on this. We have explained what the joint-L2 product is and the collection used:-

In a similar manner to that used to create the MODIS L3 product (King et al., 1997; Oreopoulos, 2005), we processed MODIS collection 5.1 joint-L2 swaths for these times into $1° \times 1°$ grid boxes. Joint-L2 data is a sub-sampled version of the full L2 swaths (sampling every 5th 1 km pixel) that also contains fewer parameters. To confirm that there is no effect from the sub-

Following up on the previous point: I assume the authors have used MODIS collection 5.1. Recently, collection 6 has (partly) become available. A short statement in the outlook on whether results are expected to change in this collection would be welcome.

This has been added (as far as we can anticipate any changes) to the very end of the main body of the manuscript :-

Finally we note that MODIS Collection 6 datasets are now being released (the data used in this study came from Collection 5.1). One significant difference is that quality assurance flags are no longer assigned, but rather new pixel level uncertainty calculations are included that are intended to replace them. It is unclear whether this will account for θ_0 effects, although generally MODIS uncertainty calculations have only accounted for instrumental measurement error rather than forward modelling error and so this is perhaps unlikely. It seems unlikely that the new uncertainty calculations will lead to the results presented here changing significantly since only pixels with the highest confidence quality assurance flags were used to calculate cloud properties. There will also be a 1km resolution physical cloud height product that may be useful for assessing cloud top height variation at a higher resolution than that afforded by the 5 km resolution cloud top temperature product of Collection 5.1, if the technique used proves to be sufficiently accurate for low level stratocumulus clouds.

P320, L4: It seems the actual reason is not so much what is described on P320, but rather on P321 (i.e. the likely misidentification of phase at lower temperatures, or at least different phase identification between low and high SZA).

This has been changed to:-

5. The mean CTT is restricted to values warmer than 268 K. This is done both to avoid clouds containing ice and because there appear to be problems in identifying the phase of clouds at high θ_0 for temperatures colder than this. These factors are discussed shortly.

In order to reflect this.

P 323, L8-10: Why would the effect of higher cloud fraction at higher viewing angle

play a role at high SZA and not at low SZA?

The following sentence has been added to address this:-

Although the error associated with the high θ value in this θ_0 bin is fairly large, this might indicate a dependence of τ on θ at very high θ_0, although it is also possible that the tendency to observe a higher cloud fraction at high θ could also be having an influence on the identification of scenes with cloud fraction $> 90\%$. It is conceivable that the mis-diagnosis of lower cloud fraction scenes as overcast might affect higher θ_0 retrievals more strongly than low θ_0 ones due to a stronger influence of cloud heterogeneity at high θ_0. Heterogeneity effects are discussed in more detail in section 4.4.

P329, L5: I would like to see a reference for the statement that the reliability of 1.6-micron r_e retrievals is 'still a matter of debate'.

This sentence has been removed in lieu of a reference for the statement.

P339, L11: Could the authors clarify what they mean by this statement? Is it an encouragement to the MODIS team to alter their level-3 generation approach?

Yes, we would suggest that high SZA retrievals are not included within Level-3 daily averages. We have added words to this effect to the end of the sentence:-

affect very large regions of the globe. Given this, an operational solution to the problem would ideally be sought in order to avoid the inclusion of high θ_0 retrievals within daily L3 averages.

P341, L28: I don't see these values mentioned above.

The value "0.8" should read "0.7" (f parameter). The fact that PZ11 suggested a change to f=0.7 from f=1.0 has been made clearer in the text above.

P366, Fig. 6: The effective radii at low viewing zenith angles are significantly lower than those at high viewing angles for all NIR bands. This deserves mentioning in the text. It is consistent with the findings of Maddux et al. (2010, see their Fig. 2) at lower latitudes (and lower SZA).

We have added the following to the discussion of Fig. 6:-

It is also worth mentioning that at both low and high θ_0 the observed r_e values were higher for the lower range of θ than for the upper range. This is interesting because an increase in r_e with θ was also observed in Maddux et al. (2010, see their Fig. 2).

P371, Fig.11: In panel (b) the uppermost point is missing.

This will be rectified.

* Suggestions for shortening (and rearranging) the manuscript *

Abstract: Please only mention the key findings here. Suggest to reduce paragraphs 3 and 4 to a few sentences.

Paragraphs 3 and 4 have been shortened to remove excess detail.

Sections 2.2.1 and 2.2.2 are a literature review, whereas Sections 2.1, 2.2.3, and 2.3 (largely) describe the adopted methodology. I would propose to have the former two sections first (as 'literature review'), followed by the methodology sections.
The sections have been moved into a new section of their own discussing the literature for these issues.

Sections 2.2.1 and 2.2.2 are very long. Try to identify the main retrieval artefacts in a concise way.

These sections have been reduced considerably to remove excess unnecessary detail. The new text is below:-

2.1 Optical depth retrieval artifacts

Cahalan et al. (1994) showed that the non-linearity of the relationship between R_{nab} and τ causes a decrease in albedo for heterogeneous clouds compared to a plane-parallel cloud with the same mean τ. This is known as the plane parallel (PP) albedo bias and is likely to lead to τ underestimates made using the measured reflectances and PP LUTs. Also, at near-nadir viewing angles and for low θ_0, cloud variability is known to cause the mean reflectance of a region to be slightly reduced compared to a homogeneous cloud with the same mean τ via 3-D effects, due to the leakage of photons horizontally from the sides of the region and due to channeling of photons from regions of high extinction to regions of low extinction where they can be lost through downward transport (Loeb et al., 1997; Davies, 1978; Kobayashi, 1993; Varnai and Davies, 1999). However, these biases are generally small compared to those that have been reported at high θ_0.

Studies using data from the ERBE (Loeb and Davies, 1996, 1997) and AVHRR (Loeb and Coakley, 1998) satellites have demonstrated that at high θ_0 ($\theta_0 \gtrsim 65°$) the optical depth inferred from the observations increased with θ_0. This was attributed to the increasing (positive) difference in reflectances between the real observed clouds and those calculated from the PP model as θ_0 increased. The results were found to be very sensitive to the thickness of the cloud with higher biases reported for the more optically thick clouds; for $\tau > 12$ and nadir viewing the positive bias was present even at low θ_0.

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. (1997) showed that 3-D nadir reflectances increase with θ_0, whereas reflectances calculated using the PP approximation decrease. This was consistent with the above observational studies indicating that 3-D radiative transfer effects within a heterogeneous cloud environment were the cause. Sensitivity tests suggested a roughly equal contribution to the bias from cloud side illumination effects and cloud top height variability effects, with the latter effect attributed to changes in the slope of cloud elements at cloud top. Such effects occurred even for completely overcast scenes. It was also indicated that cloud top height variability was more important than extinction variability. Similar conclusions were found from the modeling results of Varnai and Davies (1999).

One limitation of these modelling studies is that only nadir views were tested and the observational data mentioned above indicated that both θ and ϕ viewing angles might modulate the τ bias at high θ_0. By examining differences between nadir and off-nadir MISR retrievals Liang and Girolamo (2013) also found that τ retrievals are likely to be affected by θ and ϕ, although, the effects were observed to be complicated and the sign and magnitude of the biases was suggested to be dependent upon many competing factors. However, significant τ biases were generally not seen until very high θ values of 70.5° were reached; biases within the MODIS θ range were much lower. It was also found that cloud heterogeneity tended to enhance the magnitude of the effects, particularly for low optical depth clouds and at high θ_0.

Finally, Seethala and Horvath (2010) found that MODIS derived LWP measurements increased significantly relative to co-located measurements from AMSRE (Advanced Microwave Scanning Radiometer-EOS) at high θ_0. A large part of this was attributed to unphysical increases in τ with θ_0. The increase was greater as the inhomogeneity of MODIS τ over the $0.25° \times 0.25°$ scenes increased, which is consistent with the above results.

2.2 Effective radius retrieval artifacts

Whilst there have been a number of studies examining the effects of cloud variability and viewing geometry on τ retrievals there have been far fewer studies on the r_e effect. Marshak et al. (2006, hereafter M06) was one of the first to do so and introduced a theoretical basis to at-

tempt to explain the effects of 3-D radiative transfer on r_e retrievals that were made using cloud fields from an LES model. M06 divided the effects into those due to resolved variability of reflectances (i.e. variability at scales larger than the satellite pixel size) and those due to sub-pixel scale variability.

For resolved scale variability the theory suggested that 3-D radiative transfer effects were expected to lead to a tendency for an overall increase in r_e and τ (relative to the true values) due to the non-linearity of the relationship between the reflectances and r_e and τ. M06 suggested that sub-pixel variability would lead to a low bias of both the τ and r_e values retrieved for that pixel due to averaging of the reflectances prior to the retrieval of τ and r_e (a satellite viewing the pixel would report the averaged reflectance). The theory stipulated a number of assumptions that are unlikely to hold true in all circumstances. Nevertheless, the results from the retrievals made from reflectances calculated from the LES cloud model fields corroborated the theoretical arguments, suggesting that, at least in this case, the assumptions may have been valid, or irrelevant.

However, also using retrievals performed on LES clouds Z12 found the opposite result for the effect of sub-pixel averaging of reflectances, with the retrieved r_e increasing above the true sub-pixel r_e mean. Within $800\,\mathrm{m} \times 800\,\mathrm{m}$ regions (close to the size of a $1\,\mathrm{km} \times 1\,\mathrm{km}$ MODIS pixel) it was found that the r_e of the $100\,\mathrm{m} \times 100\,\mathrm{m}$ model resolution elements was approximately constant, but that there was quite a wide spread in τ. This was also demonstrated for a limited sample of real clouds using MODIS observations. Z12 showed that for such variability the nature of the dual-band (i.e. 2-D) LUT used for MODIS retrievals would lead to increases in r_e (and decreases in τ) and that the increase would be greater as the sub-pixel heterogeneity of R_{nab} increased. For the cases considered, these results negated the assumption of independence of the τ and r_e retrievals made in M06 since the sub-pixel τ variability meant that the non-orthogonal regions of the LUT were utilized. Thus it remains to be explained why the results from the LES model simulations in M06 were consistent with that theoretical basis.

One major difference between the simulations of M06 and Z12 that might provide a potential explanation is that the radiative transfer on the cloud fields from the M06 simulations were performed at the moderately high θ_0 of 60°, whereas in Z12 radiative transfer was performed

at $\theta_0 = 20$ and 50° and on the whole results were reported for the combination of the two θ_0 values. It is likely that the result obtained will depend on the degree of sub-pixel variation of both R_{nab} and R_{ab}, the region of the LUT covered by the reflectance values and the influence on the sub-pixel reflectances of 3-D effects. Such factors are likely to be affected by the value of θ_0. Other factors that alter the orthogonality and non-linearity of the LUTs are also likely to affect this result, such as the near-infrared wavelength used, as also demonstrated in Z12. Their results showed that the increase of r_e due to sub-pixel averaging was substantially greater for the $2.1\,\mu\mathrm{m}$ band relative to the $3.7\,\mu\mathrm{m}$ band, and that this most likely because the LUT for the latter is more orthogonal than for the former.

There have been several attempts in the literature to use the differences between r_e from the different MODIS bands to infer information about the vertical structure of the cloud. This may be theoretically possible since the different wavelengths of light have different penetration depths into the cloud and thus produce a weighted mean r_e that is representative of different vertical regions of the cloud (Platnick, 2000). However, the heterogeneity effects just mentioned will clearly impact such attempts. Further discussion on this is deferred to Sects. 4.2.2 and 5.3.

Section 4 is again very long. Especially 4.1 reiterates a lot of what was written in Sections 2.2.1 and 2.2.2.

These sections have also been cut down, although removing a lot of the detail from Sections 2.2.1 and 2.2.2 cut down on the repetition.

Appendix B: Suggest to remove the last two paragraphs. This concerns observations made for other clouds / over other surfaces, and are thus not directly relevant for the present study.

These paragraphs have been removed.

Appendix C: Consider to remove last paragraph (seems not relevant for the present study).

Hallett Mossop ice initiation may be relevant because it can lead to large ice concentrations at temperatures warmer than -5 deg C (which was the temperature cut off for the MODIS dataset used). However, this paragraph has been considerably shortened in the new version.

* Technical corrections *
P304, L19: Remove 'which'.
P304, L26: separate 'r_e3.7' and 'r_e'.
P304, L15, L24; P308, L2: It is not clear what these 'processes' are.
P311, L7: Add hyphen between MODIS and derived.
P311, L8: AMSRE should be AMSR-E. P312, L16: 'valid, or irrelevant': these seem to be opposite characterizations; please
clarify.
P314, L13: PZ11 has not been introduced yet.
P315, L20: Suggest to write 'datasets are', since there are different instruments and collections.
P316, L11: Add '(earlier)' after 'later'.
P316, L25: Consider revising the phrase 'than are necessary'. Who determines 'what is necessary'?
P320, L5: Should 'also' be replaced by 'already'?
P320, L16: Do you mean 'will' instead of 'can'?
P323, L15: Tables should be introduced in order of appearance (i.e. Tables 1 and 2). Same holds for Figs. 6 and 7.
P325, L22: Replace 'because' by 'i.e.' (it is not a reason).
P326, L17: cloud top 'temperature' would be more precise than 'height'.
P356, Table 1 as well as Tables 2 and 3. These are hardly readable. Make sure that font size is considerably increased in the finally published manuscript.
P363, Fig.3: Add that this is for the selected region in Fig. 2.
P365, Fig. 5: Add that errors as discussed in the text are indicated by horizontal bars, and that for most points the errors are so small that they cannot be seen (at least that's what I assume).
P365, Fig. 5: Suggest to call 'sensor zenith angle' 'viewing zenith angle' throughout, and abbreviate it with VZA. Also the lower limit of viewing angle can be omitted (i.e. write VZA < 41.4). P367, Fig.7: In the caption change '>' to '<', or better: write 'theta < 41.4'. In the legend change 85 to 81 degrees (the upper limit of MODIS optical property retrievals).

The above technical corrections will be performed for the revised manuscript.