

Review of Bennartz and Rausch – “Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations”

This paper describes a climatology of monthly mean cloud droplet concentrations (CDNCs) using the MODIS-Aqua satellite instrument. Some of the potential retrieval issues that can affect CDNC retrievals are discussed along with some examples of maps and seasonal cycles from the final climatology. This data set is likely to be a very useful commodity to many researchers and I commend the authors for making it available online. Thus I recommend the publication of the paper once certain issues have been addressed and additional discussion points added. The major issues are summarized next with more specific comments and typos below that.

Major issues

One key point on which I disagree with the authors is their restriction to situations where $re_{3.7} > re_{2.1} > re_{1.6}$. Painemal and Zuidema (2011) showed that this criteria was often violated for MODIS retrievals of stratocumulus (generally $re_{1.6}$, $re_{2.1}$ and $re_{3.7}$ were very close) and yet the aircraft observed LWC was found to increase with height. This is likely due to other errors in the retrieval of re_{eff} that affect the different MODIS channels in different ways. E.g. sub-pixel heterogeneity can increase $re_{2.1}$ more than $re_{3.7}$ even in relatively homogeneous stratocumulus (Zhang, 2012), and Grosvenor (2014) showed that resolved heterogeneity is likely to affect the $re_{3.7}$ more than $re_{2.1}$. Also, King (2012) show that the information content from the different MODIS channels is not enough to be able to retrieve vertical height variation in re_{eff} and so it seems to me that this restriction might be causing some spurious filtering of the data. It is also not clear from Fig. 3 that $re_{3.7} > re_{2.1}$ would select the cloud covered part of the curve as stated since $re_{3.7}$ and $re_{2.1}$ are very close, as was also seen in the Painemal and Zuidema (2012) aircraft observations. Differences between the stratified and non-stratified CDNC values are reported in Section 4.2.1, but can we be sure which is more realistic? I think that there should be more discussion about and acknowledgement of these issues.

Consideration of scattering angle effects and the effect of sunglint is made. However, the applied filtering does not take into account the likelihood of retrieval biases at high solar zenith angles. A little more detail should be given on this issue along with consideration on how this bias will affect your data set. For example, this could potentially cause large-scale biases in the winter months at high latitudes.

It can't be said for sure whether some of the scattering angle and sunglint effects in Section 4.1 are due to variations in region – certain regions or seasons may be sampled more often at certain scattering angles and with certain sunglint angle. And there can be strong natural regional changes in re_{eff} , τ , CDNC, etc. Is there a way to rule this out? E.g. can you look at the differences from within the same swath relative to the swath mean and then average these, so that the locations of samples are roughly the same? This is possible for sensor angle (as in Maddux, 2010) and sunglint angle, but I'm not sure about scattering angle. There is also no direct examination of the effect of at sensor angle. This would be

useful, or at least you could directly refer to this from the results of Maddux (2010)? Also, the swath edges may also sample a different range of scattering angles than elsewhere in the swath, which may also cause artifacts. Have you looked into this?

RE “uncertainties” – using the variability of the daily CDNC values means that a lot of what is termed “uncertainty” will actually be natural variation in CDNC, which, even with perfect retrievals, would not be expected to be constant throughout a cloud field, nor over the course of a month, nor over the climatology. I understand that given the fact that the uncertainties are not well characterized it is hard to separate the real variability from the noise introduced by retrieval errors. However, perhaps the quoted term should just be renamed “standard deviation” with it being made clear that some of this will be actual variability and some uncertainty. Also, since you can calculate the propagated CDNC uncertainty from the pixel level uncertainties in tau and reff and I think it would be useful if these were provided to get a sense of how much they contribute. However, it should also be made clear that these would not account for the entirety of the uncertainty since there will be a lot of error introduced by the forward model plane parallel independent pixel approximation that will not be captured.

Fig. 5 - it looks like there is very little data in box X12 – the numbers of samples going into each box should be quoted for this figure. There is also overlap between box R03 and X12, yet box X12 shows quite different results, perhaps indicating that the statistics for this box are too poor to be robust?

Specific comments

Abstract – “Resulting CDNC uncertainties for the climatology are in the order of 30% in the stratocumulus regions and 60% to 80% elsewhere” – as discussed in the other comments is it appropriate to call this “uncertainty”. A more careful description is needed here.

p.2, L27 – “1. The cloud is assumed to be horizontally homogeneous.” – This seems a little vague. It would be good to mention over what scales the cloud needs to be horizontally homogeneous and why. E.g. 1km, 1x1 degree or both? Or larger even? Perhaps it would be worth mentioning the application of the independent pixel approximation for reff and optical depth retrievals such that inhomogeneity over scales larger than the 1km pixel size is likely to violate this assumption with the scale being set by the degree to which net horizontal photon transport occurs (the scale of any shadows/bright spots, etc., or rather deviations from the plane-parallel reflectances, as defined in Marshak, 2006). This could be separated from the requirement of homogeneity within the 1km pixel to satisfy the plane-parallel retrieval assumption. In terms of the CDNC derivation itself only the sub-pixel homogeneity is explicitly required once we have the correct 1km reff and optical depth since 1km resolution CDNC values can be calculated.

p.3, L16 – “The true three-dimensional variability of clouds poses significant challenges to any remote sensing algorithm and a growing body work has been devoted toward understand the impact of this variability on remote sensing estimates of CDNC.” – these works should be cited here, or else you should refer to them in another section in the paper.

p.3, L19 – “as ultimately one would be interested in the number of cloud droplets activated at cloud base and not the number of cloud droplets observed” – I think this would depend on the application. Some studies may be interested in how the cloud top CDNC might change due to lateral mixing, evaporation, etc., or removal of CDNC by precipitation and not necessarily just the cloud base CDNC. I can see that the cloud base CDNC would be of interest for comparing to model processes, but I think that the statement here generalizes too much.

p. 4, Eqns. 2 & 4 – You should describe how c_w (condensation rate) was calculated – did you use the MODIS cloud top temperature? What pressure did you use? Did you use the full adiabatic value, or 80% of this as mentioned on p.3. An issue with the calculation used in Bennartz (2007) is mentioned in the acknowledgments - it would be good to add more detail in the main part of the paper about this issue along with a reference for the correct formula – e.g. Ahmad (2013). A discussion of the compensation of errors in k , fraction of adiabatic condensation rate, reff errors found by Painemal and Zuidema (2012) should also be included.

p.6 – “3. The cloud mask had to indicate the observation to be cloudy but not over ice or land.” – does this include the filtering of sea-ice covered regions? How robust is the detection of sea-ice? This is important since this is likely to lead to a poor retrieval.

p.6 – “6. Observations were only considered, if the three MODIS-retrieved effective radii stacked up as $\text{re}_{3.7} > \text{re}_{2.1} > \text{re}_{1.6}$, as observations violating this criterion will also violate the key assumption of a vertically increasing LWC in the ISBLC.” – this is a key point on which I disagree with the authors – Painemal and Zuidema (2011) showed that this criteria was often violated for MODIS retrievals of stratocumulus (generally $\text{re}_{1.6}$, $\text{re}_{2.1}$ and $\text{re}_{3.7}$ were very close) and yet the aircraft observed LWC was found to increase with height. This is likely due to other errors in the retrieval of reff that affect the different MODIS channels in different ways. E.g. sub-pixel heterogeneity can increase $\text{re}_{2.1}$ more than $\text{re}_{3.7}$ even in relatively homogeneous stratocumulus (Zhang, 2012), and Grosvenor (2014) showed that resolved heterogeneity is likely to affect the $\text{re}_{3.7}$ more than $\text{re}_{2.1}$. Also, King (2012) show that the information content from the different MODIS channels is not enough to be able to retrieve vertical height variation in reff and so it seems to me that this restriction might be causing some spurious filtering of the data.

p.6, L24 – “a solar zenith angle of 56 degrees. The resulting scattering angle of 124 degrees” – it would be good to define what you mean by the scattering angle. I.e. that it is the angle between the sun and the satellite within the plane of the sun and the satellite and that 180 degrees is backscatter.

p.7, L23 – “Typically, most of these issues affect the effective radius at 1.6 μm and 2.1 μm more strongly than the effective radius at 3.7 μm (Zhang et al., 2012b).” – this was true for the sub-pixel optical depth heterogeneity effect, but are there other examples? Grosvenor (2014) showed that the 3.7 μm channel is likely to be more strongly affected by resolved 3D radiative effects, so this is not always the case. Although the latter effect is more likely at higher solar zenith angles.

p.7, L33 – “observations (i.e. if cloud thickness and liquid water path varies within the field-of-view).”

– Strictly, Zhang (2012) talk about the effect of having sub-pixel heterogeneity in optical depth (which cloud arise through combinations of LWC, CDNC, etc.) changes, rather than cloud thickness and LWP.

p. 8, L 4 – “so that in the example shown in Figure 3 (right panels) the largest part of the retrievals would be (correctly) rejected.”

– it is not clear from Fig. 3 that $re_{3.7} > re_{2.1}$ would select the cloud covered part of the curve since $re_{3.7}$ and $re_{2.1}$ are very close, as was also seen in the Painemal and Zuidema (2012) aircraft observations. Why not just filter based on the cloud fraction?

p.9, L4 – “Grosvenor and Wood (2014) address issues the dependency of cloud microphysical retrievals on solar zenith angle.”

– However, it seems that the filtering applied does not take into account the likelihood of retrieval biases at high solar zenith angles and this issue is not brought up again. A little more detail should be given on this issue along with consideration on how this bias will affect your data set.

p.9, L14 & Fig. 4 – sunglint angle is not defined. I’m guessing that you mean the absolute difference between the sensor zenith angle of a given pixel and the sensor zenith angle of the middle of the sunglint band within the swath, but it is not obvious.

Section 4.1 – It can’t be said for sure whether some of the effects seen are not due to variations in region – certain regions or seasons may be sampled more often at certain scattering angles and with certain sunglint angle. And there can be strong natural regional changes in re_{eff} , τ , CDNC, etc. Is there a way to rule this out? E.g. can you look at the differences from within the same swath relative to the swath mean and then average these, so that the locations of samples are roughly the same? This is possible for sensor angle (as in Maddux, 2010) and sunglint angle, but I’m not sure about scattering angle.

p.10, L10 – presumably high sunglint angles (if my definition of this is correct) will be associated with high sensor angles since they will have to be near the edges of swaths. So, this could be contribute to the artifact. Why not also look at sensor angle directly? Or at least directly refer to it from the results of Maddux (2010)? However, the above interpretation seem inconsistent with Maddux (2010). They showed that the swath edge liquid optical depth values were lower and the re higher than those near nadir. Are the results consistent with Maddux (2010) as stated at line 15?

Also, the swath edges may also sample a different range of scattering angles than elsewhere in the swath, which may also cause artifacts. Have you looked into this?

Section 4.2.1 – differences between the stratified and non-stratified CDNC values are reported, but can we be sure which is more realistic? Also, it looks like there is very little data in box X12 – the numbers of samples going into each box should be quoted for Fig. 5. There is also overlap between box R03 and X12, yet box X12 shows quite different results, perhaps indicating that the statistics for this box are too poor to be robust?

Fig. 6 – the title uses the term “unflagged”, but “stratified” would be more consistent with the rest of the paper.

Also, the frequent positive values seem inconsistent with that expected from Fig. 4 where the main effect appears to be a decrease in CDNC at high scattering angles. Or is the increase at low sun glint angles dominating for these regions? Why not also exclude the very high sun glint angles for the flagged pixels?

p.12, L 5 – “PZ11. Various effects, including cloud top entrainment or the representativeness of the chosen condensation rate, can potentially cause the somewhat larger bias for the stratified cases.” – It’s hard to see how this might be the case? Can you explain? Could it also be that the stratified cases might be further from the aircraft observation?

p.12, L10 – “However, for both 51x51 and 21x21 neighbourhoods, the mean uncertainty (size of error bars in Figure 7 and last two columns in Table 1) is reduced for stratified cases over the un-stratified cases. ... Effective radius stacking is likely decreasing the selection of pixels of inhomogeneous clouds or those subject to sub-pixel effects, which, as discussed in Section 3, can result in retrievals with a wide spread of unphysical effective radii.” - Could this also be due to there being fewer samples in each region due to being more selective (since this would reduce the std. deviation)? Or were all boxes chosen to have all possible pixels present?

p.13, L3 – RE “uncertainties” – using the variability of the daily CDNC values means that a lot of what is termed “uncertainty” will actually be natural variation in CDNC, which, even with perfect retrievals, would not be expected to be constant throughout a cloud field, nor over the course of a month, nor over the climatology. I understand that given the fact that the uncertainties are not well characterized it is hard to separate the real variability from the noise introduced by retrieval errors. However, perhaps the quoted term should just be renamed “standard deviation” with it being made clear that some of this will be actual variability and some uncertainty. Also, since you can calculate the propagated CDNC uncertainty from the pixel level uncertainties in tau and reff and I think it would be useful if these were provided to get a sense of how much they contribute. However, it should also be made clear that these would not account for the entirety of the uncertainty since there will be a lot of error introduced by the forward model plane parallel independent pixel approximation that will not be captured.

p. 13 & Fig. 9 – “magnitude of the annual cycle” – can you state somewhere what you actually mean by this? Is it the peak-to-peak amplitude of the annual cycle (i.e. max monthly mean minus min monthly mean), or is it calculated based on the cosine fits? “Amplitude” would be a better word than “magnitude” since it relates more directly to the definition.

Fig. 10 – the bottom right panel needs the y-axis range changing to show the negative values.

p. 14, L31 – “We further show that neglecting this screening does not only lead to moderate biases in the annually averaged CDNC” – Are these biases or just differences? How do we know which one is correct due to the issues raised earlier in this review?

p.15, L32 – “We found some remaining retrieval artefacts in the MODIS-retrieved effective radius and optical depth that propagate through into artefacts of the CDNC climatology as well.” – it would be useful to mention the issues that you are talking about here.

p.15, L6 – “However, feel that it might be beneficial in future work to re-create retrievals working directly on the Level-1 reflectances that address some of the issues we have identified.” – although, perhaps it should be mentioned that it may not be possible to resolve many of these issues even if re-processing Level-1 data – e.g. it would be hard to correct biases for 3D radiative effects, sunglint effects, etc.

Tables – it would make sense to alter the order of these to align with the order of mention in the paper – i.e. first Table 3, then Table 2 and then Table 1.

Typos

p.9, L4 – “Grosvenor and Wood (2014) address issues the dependency of” – remove “issues”.

p.9, L20 – “Fro Figure 4”

p.9, L22 – “vary in a manner the roughly resembles”

p.10, L3 – “Since Aqua’s local equator crossing time around 14:30,”

p.10, L13 - “with an strong increase retrieved optical depth” -> “with a strong increase in retrieved optical depth”

p. 10, L19 – “4.2. Impact Stratification and retrieval artefacts on climatology” -> “Impact of Stratification...”

p.10, L24 - “boxes are shown were all 12 months”

p.12, L2 – “in the order of” -> “on the order of”

p.12, L4 – “For stratified results, magnitude of the RMSE and bias increase” -> “For the stratified results, the magnitude ...”

p.14, L25 – “a better traceable of results” -> “a better traceability of results”

p.15, L6 – “However, feel that”

p.20, L20 – “cover have discussed” -> “have been discussed”

References (only ones not cited in the paper under review are listed)

Ahmad, I., Mielonen, T., Portin, H. J., Arola, A., Grosvenor, D. P., Mikkonen, S., Leskinen, A., Komppula, M., Lehtinen, K. E. J., Laaksonen, A., and Romakkaniemi, S.: Long term measurements of cloud droplet concentrations and aerosol-cloud interactions in boreal boundary layer clouds, *Tellus B*, 65, 20138, doi:10.3402/tellusb.v65i0.20138, 2013.

Marshak, A., Platnick, S., Varnai, T., Wen, G., and Cahalan, R.: Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes, *J. Geophys. Res.-Atmos.*, 111, D09207, doi:10.1029/2005JD006686, 2006.

King, N. J. and Vaughan, G.: Using passive remote sensing to retrieve the vertical variation of cloud droplet size in marine stratocumulus: an assessment of information content and the potential for improved retrievals from hyperspectral measurements, *J. Geophys. Res.-Atmos.*, 117, D15206, doi:10.1029/2012JD017896, 2012.