Answer to referee #2 James Campbell

Dear J. Campbell:

Thanks a lot for your useful comments, technical notes, editing recommendations and questions. We appreciate it very much; it will definitely improve our manuscript. Below you will find our replies and short descriptions of the changes we've made in the text. Your comments are in red and start with **"R:"** and our replies are in black and start with **"A:".** Original manuscript text is shown in **blue**, with new text highlighted in **yellow**.

We would like to warn you, however that we are still working on our data analysis to accommodate all suggestions from both referees and hence the final numbers might still change. You urged us to change our definition of cirrus clouds, but what is being more challenging is the request by Referee #1 to do a full multiple-scattering correction, instead of the approximate correction we have now. The new manuscript will be uploaded as soon as we finish all these changes.

R1: My primary scientific concern relates to the definition of "cirrus" clouds in the manuscript. (…) In a recent paper that I authored (Campbell et al. 2015), we went to significant length to demonstrate a practical and viable definition for cirrus clouds in autonomous long-term datasets like this one, and in particular for those that lack a polarized backscatter measurement. (…) This absolutely has to be revised. I recognize that this is a serious request, and I raise this point very respectful of the work that has been put into the manuscript, the statistics and the analysis. However, I question every number you have in here, again respectfully, because of such a simple and non-physical definition applied for discriminating these clouds.

I would respect if the authors were to disagree with our conclusions/recommendations in Campbell et al. (2015). But, in response, they'd better come up with a physically- based reason for doing so. Cloud top temperatures of -37°C makes physical sense for the class of clouds that we call cirrus. It's a practical and defendable threshold.

A: We agree with the reviewer that cirrus is a phenomenological classification based on surface visual observations. We also agree that such classification (and alternatives to make it more practical) has been debated in the literature. For instance, the -25 degC threshold that we applied has been used in previous papers (e.g. SEIFERT et al., 2007; GOLDFARB et al., 2011). However, we don't agree that this criterion allows our numbers to be, although respectfully, strongly questioned. In the tropics, our threshold corresponds to a minimum cloud-base altitude of about 8 km. Likewise, the -37 degC cloud-top threshold suggested by the reviewer corresponds to roughly 11 km. The cloud-top histogram in figure 5 shows that less than 6% of what we've called cirrus would not fit the reviewer's suggested criteria. Therefore, our numbers cannot be that far off from what would be obtained following Campbell et al (2015).

Nonetheless, a physically based definition such as proposed by Campbell et al (2015) is indeed preferred. We will use this definition in the new version of the manuscript. We are already reprocessing all our data.

R2: Its unclear what the authors are saying about the presence of SNR > 3 in the upper troposphere with respect to cloud observation. Do they mean clear-sky? Or, do they mean within particulate scattering layers?

A: We apologize for not having stated this clearly in the manuscript. What we meant is that the molecular lidar signal just below the cirrus cloud-base should have a SNR of at least 3, in a single bin of 7.5m (our raw resolution). For 30m bins, for instance, the SNR will increase to about 6, and one should remember that the molecular fitting involves many points, which further reduce the noise. For the typical cirrus optical depths, SNR>3 means that the laser was not attenuated and hence we will most likely still have good enough SNR above the cloud top, which is needed for the retrieval of the optical depth and lidar ratio by the transmittance method. We also have to evaluate the SNR in cirrus-free profiles in order to count all the profiles in which cirrus could have been detected if they were present. This is necessary to compute frequency of occurrence correctly.

The first paragraph of section 2.2 was modified as follows around lines 115-118:

A total of 36,597 5-minute profiles were analyzed and only 20,752 had a signal to noise ratio (SNR) higher than 3, for a single 7.5 m bin just below the cirrus base. Given the typical cirrus cloud optical depths, this threshold means we will also have a SNR at cloud top that is good enough for estimating the optical depth with the transmittance method (see section 2.4). Statistical tests with the transmittance method based on simulations for various SNR, COD and cloud thickness (not shown) were conducted to obtain the SNR threshold. The number of 5-min lidar profiles and number of profiles with good SNR during each month of the studied period were analyzed.

R3: Its unclear how the authors define the tropopause, and thus accommodate the potential for resolving the bottom/top of the tropopause transition layer, in Section 2. This hurts the discussion later on where context is necessary for understanding where the clouds are with respect to this boundary.

A: We apologize for not having stated this clearly. As we wrote in the manuscript, we are using the WMO definition (International Meteorological Vocabulary, 1966). In this technical document the Tropopause is defined as: "*the boundary between the troposphere and the stratosphere, where an abrupt change in lapse rate usually occurs. It is defined as the lowest level at which the lapse rate decreases to 2°C/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2°C/km*".

One should be careful when applying this definition, however, as the number of vertical levels in the sounding, reanalysis or model data might be too coarse. To overcome this issue, we follow the methodology suggested by the National Meteorological Center (McCalla, 1981). The lapse-rate is assumed to vary linearly with pressure, and the exact altitude where Γ=2°*C/km* (i.e. the tropopause) is found by linearly interpolating between the closest available pressure levels.

We modified section 2.2 around lines 130-147 to better explain the calculation of the tropopause altitude:

This dataset was used to obtain the mean high level winds, near to the cirrus clouds habits (200 hPa). The tropopause altitudes were obtained from vertical profiles over the site using the definition of the World Meteorological Organization (IMV WMO, 1966), i.e. "*the lowest level at which the lapse rate decreases to 2°C/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2°C/km*". We further assumed the lapse rate to vary linearly with pressure (McCalla, 1981), and the exact altitude where Γ=2°C/km (i.e. the tropopause) was found by linearly interpolating between the closest available pressure levels. A precipitation dataset for the same period was acquired from TRMM (Tropical Rainfall Measuring Mission) version 7 product 3B42 (Huffman et al., 2007) with 0.25° and 3 h of spatial and temporal resolution, respectively.

And we include these two references:

International meteorological vocabulary. WMO, No. 182. TP. 91. Geneva (Secretariat of the World Meteorological Organization) 1966. Pp. xvi, 276. Sw. fr. 40. Q.J.R. Meteorol. Soc., 93: 148. doi:10.1002/qj.49709339524

McCalla, C., 1981: Objective Determination of the Tropopause Using WMO Operational Definitions, Office Note 246, U.S. Department of Commerce, NOAA, NWS, NMC, 18pp, October 1981.

R4: Since the sample size is stated to relative to the ability to measure SNR > 3 in the upper troposphere, all of the samples appear to be relative occurrence frequencies and not absolute ones. This is HIGHLY confusing. There is no way that you're resolving an absolute cloud frequency of 67%. In a new paper that we have in Early Online Release in JAMC (Campbell et al. 2016), we show in a year's worth of MPLNET observations at Greenbelt, MD an absolute frequency near 16%, which owes to the attenuation of the beam from low-level clouds and undersampling of the upper troposphere. There are multiple places in the narrative where serious confusion arises and the speculative discussion becomes meaningless because of this confusion.

A: We do not agree with the referee in this point.

We know, based on satellite studies that the cirrus absolute frequency is much higher than the 16% found by the referee for Greenbelt. For instance, based on Calipso and CloudSat, Sassen et al., JGR 2001 show for US east coast a frequency about 25% (fig.1 of that paper). They also show a frequency about 50% at our site in the Amazon (3S, 60W). The only way one could have such a low frequency, as suggested by the reviewer, is if one divides the total number of cirrus detected by the total number of possible observations (i.e. including low-level clouds, etc…). We argue, however, that such number would not have a physical meaning. It would just reflect your sampling issues.

On the other hand, it is very usual to report the cloud occurrence the way we do (e.g. Erika Kienast-Sjögren et al. 2016; Nazaryan et al., 2008; and references therein). There is a broad but valid assumption behind, which is justifying this approach. The lifetime of cirrus is much longer than for other clouds. At the same time, the presence of cirrus clouds in the sky is rather independent of low-level water clouds that can fully attenuate the laser beam. Hence, you can estimate the absolute cirrus frequency simply by dividing: the number of lidar profiles with a cirrus, by the number of profiles where you could have detected a cirrus cloud.

We agree this is not the true cirrus frequency, but it is the best estimate one can make. Besides, our cirrus frequencies are in agreement with values obtained from CALIPSO, if we consider the same time of the satellite overpass. See, for instance, the values reported by NAZARYAN et al. 2008 or SASSEN et al 2008 and compare with the values in our paper.

We want to give a very naïve example, and we do so very respectfully with the aim of making our point very clear. The approach we follow is the same as doing a pool to figure out who will win the next election for president. You take a small sample of 1000 people from a population of 150 million voters and you can still tell the outcome of the election (given that your sample was randomly selected). But please note that we don't have to worry about the "random selection" in the case of cirrus frequency because: 1) the presence of cirrus and low-level clouds are independent (and also independent from our sampling failures); and 2) we sampled 37k profiles of 5-min, which is 1/3 of the maximum possible number of profiles during 1 year.

But how do we know which profiles we could have detected a cirrus (if they were there)? We do that by looking at the SNR at the typical cirrus altitudes and knowing the efficiency of our algorithm as a function of the SNR. Based on the analysis of simulated profiles (GOUVEIA, 2014, Msc Thesis, U. of Sao Paulo), we know that our algorithm can detect 99% of cirrus clouds with COD > 0.005 if the SNR is at least 3 below cloud base. With this strategy, for example, profiles with low water clouds that kill the laser beam are not added in the denominator. The total number of profiles measured in each month and the number of profiles with good SNR was shown in figure S.1 in the supplement and discussed it in the text:

Lines 119-125

July, August and September, the driest months show the higher fraction of profiles with good SNR, while the wettest months have the lowest fraction of lidar profiles with good SNR (see figure S.1). The cloud fraction of low, optically thick clouds increases during this season, thereby attenuating the signal and reaching the cirrus clouds altitudes with a low SNR. The frequency was then defined as the ratio between the number of lidar profiles of 5 min with good SNR containing cirrus clouds and the total number of profiles with good SNR.

R5: Speaking of this issue, nothing is said of the work of Thorsen et al. (2011) and Protat et al. (2014) and undersampling issues relating to ground-based profiling, attenuation, and the relative cloud samples that we have to analyze. This is a serious weakness that leads to three other points of concern.

Now that the reviewer mentioned Thorsen and Protat, we believe to have understood his concerns. We probably did not explain very well how we count the profiles for our statistics and that might have lead to his confusion here and in the previous comment.

As we explained above, properly counting how many profiles you have in the denominator of your cirrus cloud fraction is the key point (think of the election pool). You cannot include attenuated profiles otherwise you will introduce a bias. You also cannot average over a long period of time by simply averaging your data, unless it is uniformly distributed.

All of these points are considered in our approach. For instance, how do we calculate the year average if we measured a different amount of days in each month (different sample sizes)? We just average the fractions of each month (weighted by the number of days), and the fraction in each month was calculated including in the denominator only the profiles for which you could have detected a cirrus. If our sampling varied too much within a month, we could've broken it up into weeks. The same strategy is applied when we calculate the diurnal cycle. The cirrus fraction in a given hour-bin is the number of profiles with cirrus in that hour divided by the number of profiles for which we could have detected a cirrus in that hour. Therefore, we can still have a good estimative of the true diurnal cycle even if we have different sample sizes for each hour.

To our understanding, our approach (election pool) is the same as used by Thorsen et al JGR (2011) and Protat et al JAMC (2014), however, they've called it "conditional sampling". See, for instance, what Protat says in section 3 of his paper:

> "*Fortunately, conditional sampling (for instance excluding profiles where low-level obscuration occurs, as in Thorsen et al. 2011) can be carefully designed for sake of model and satellite product evaluation using data collected at the ground-based sites.*"

To make our approach more clear, we will modify section 2.2, removing the discussion about the SNR. That will be included in a new section called "sampling issues" where we will explain how we applied the conditional sampling of Thorsen (2011) and Protat (2014). That will be a summary of our replies to your comments R4 and R5.

R6: It is discussed that the lowest cloud observations occur around solar noon (10-12 LT). This leads me to believe that your instrument is suffering from issues with SNR from the bright background, even at 355 nm. Whereas it is introduced that this is potentially a real artifact, I see no reason to take such a claim at face value. As I cannot evaluate your algorithm or its performance, and with the practical understanding that you are willing to deal with cloud samples in the algorithm at an SNR as low as 3, I cannot help but conclude that you're dealing with sampling issues due to background noise.

We thank the reviewer for carefully looking at all details of our results. We should say, however, that we also have looked into this minimum around solar noon to be sure that it was not a problem with the solar background. Our conclusion is that it is real for the reasons below.

About the SNR of 3 -- We should emphasize that this is for a single bin of 7.5m in the molecular range below the cloud base, as explained in the reply to comment R2. The SNR of the cloud it self is, of course, always much larger than that. Besides, the molecular fitting (below and above the cloud) works as an averaging procedure and hence the effective molecular SNR is also much larger.

About the algorithm performance -- We have done an extensive simulation study to validate the methods we use, which was not included in the original submission. We were planning to have a separate manuscript on AMT about the accuracy and precision of our cloud detection algorithm and of the transmission method for the retrieval of COD and LR from elastic lidars. However, as both referees have questioned about this, we believe that some of that needs to be included in the supplement material. We will consult the editor to see if he/she agrees with this approach.

In the simulations, we varied the cloud thickness (from 15m to 4.5km), the cloud extinction coefficient (from 0.02 to 0.1 km^{-1}) and the SNR from 3 to 50. We verified that our cloud detection algorithm can identify 99% of clouds with COD > 0.005 if the molecular SNR is at least 3 below cloud base. This is evidence that we are not suffering from SNR issues from the bright background.

The second evidence is that other studies have also reported a minimum in the cirrus occurrence around noon. Hong et al JGR (2006) used the TOGA radar during the TRMM-WETAMC campaign in the Amazon, and also the PR and VIRS instruments onboard of TRMM. They showed that the diurnal cycle of thick anvils (hence no SNR x bright Background issue) has minimum around 8-12LT. This is similar to what we found and fits perfectly with the diurnal march of tropical convection (Machado JGR 2002). There are examples in other regions as well. Throsen et al JGR (2013) used ARM data from SGP together with CALIPSO and showed that the thinnest cirrus occurs around 12h LT. Liu et al Adv. Atmos. Sci. (2015) used an MPL in southeastern China and showed the diurnal cycle of total cloud fraction also has a minimum around noon.

The third evidence is the diurnal cycle separated for sub-visual, thin and thick cirrus (see below). The left panel uses the column total COD, hence the sum of the 3 curves will give exactly the black line in fig. 4 of the paper. In the right panel, we used the layer COD and hence the sum might be more than the total occurrence (because of multiple cirrus layers). There is a well marked diurnal cycle for the thick cirrus clouds. These clouds have COD > 0.3, hence there can't be any artifacts from SNR versus bright background issues. Their maximum occurs just after the maximum of precipitation and leads to the conclusion that they are actually formed from detrainment of the anvils, what fits nicely with our argument in the paper. The figure also shows that the maximum of thin cirrus occurs about 12h after the peak in thick cirrus. The COD of these clouds is way larger than the detection limit of our algorithm, and hence their diurnal cycle cannot be an artifact of low SNR during daytime. Together, thin and thick cirrus accounts for 60% of the total amount of cirrus and they both have a minimum around noon. The sub-visual cirrus (individual layers) show a prevalence during night-time. We are currently running more simulations to be sure that this is not an artifact. However, if there is one, it would be that we are missing day-time SVC and hence, that their true fraction is even larger than 40%.

Last but not least, we thank the reviewer again for raising this point. This forced us to further explore our dataset producing the figure/analysis below. We will modify the manuscript to make our argument more clear and self-evident, following this discussion.

Column COD

COD of individual layers

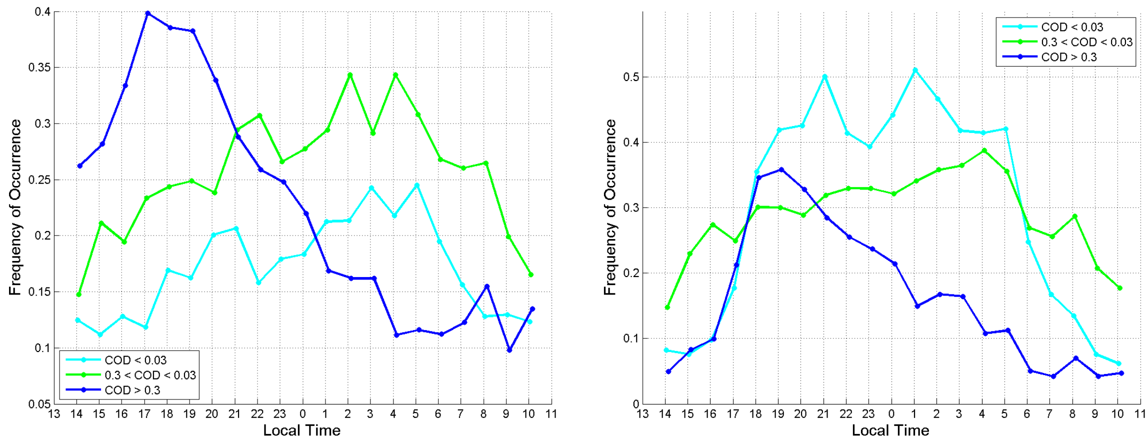

Fig – Diurnal cycle of cirrus frequency for each cirrus type for the year average. The sum of the three curves gives exactly the black curve shown in figure 4 of the manuscript.

R7: Furthermore, all of the speculation about the transport of clouds vs. near-source convective generation is very weak. The authors are forgetting that if the clouds are being generated at/near or on top of them, the lidar will not be profiling the clouds. You are *always* dealing with transport of some kind, as such. I recognize what they are trying to say, but recommend they be much more circumspect about how they are delineating source/transport with respect to the limited information that they have.

First of all let us clarify that when we talk about transported cirrus clouds we are specifically talking about long-range transport (Fortuin et al 2007). By the way, this was explained in the manuscript, at lines 256-258:

As the tropical cirrus can be transported by advection thousands of kilometers (Fortuin et al., 2007), we speculate that during the wet period, the cirrus clouds observed in central Amazonia are a mixture of locally produced and clouds transported by advection from other regions.

Nonetheless, we agree with the reviewer that our discussion about the sources of the cirrus we measured is weak, which was also noted by reviewer #1. He/she suggested that we could use back-trajectories to give further quantitative evidence that our cirrus clouds originate from deep convection. Hence, we did back-trajectory analysis using Hysplit forced by GDAS winds (1deg resolution), starting every 6h from 14.5 km over the site during the dry season period. Each of the 480 back-trajectories were integrated for 7 days. Figure below shows the result of this analysis. In the top panel, we show the individual trajectories just for the 0:00 of each day and there are so many lines that clutter the plot. The lower panel shows the trajectory density, i.e., the number of trajectories in a point divided by the total number of trajectories (a number [0-1]). In this case we used a log-scale because the density will obviously be much higher closer to the trajectory start point. The result is quite interesting as it reveals that many trajectories actually don't follow the average wind pattern (fig. 3 in the manuscript, top panel). On the other hand, many trajectories come from Colombia and Venezuela, exactly where precipitation from deep convection is found (also shown in fig. 3, top), and some even reach towards the ITCZ, far to the east. This comparison could be improved if we select only the trajectories starting at times when we detected a cirrus clouds (yet do be done).

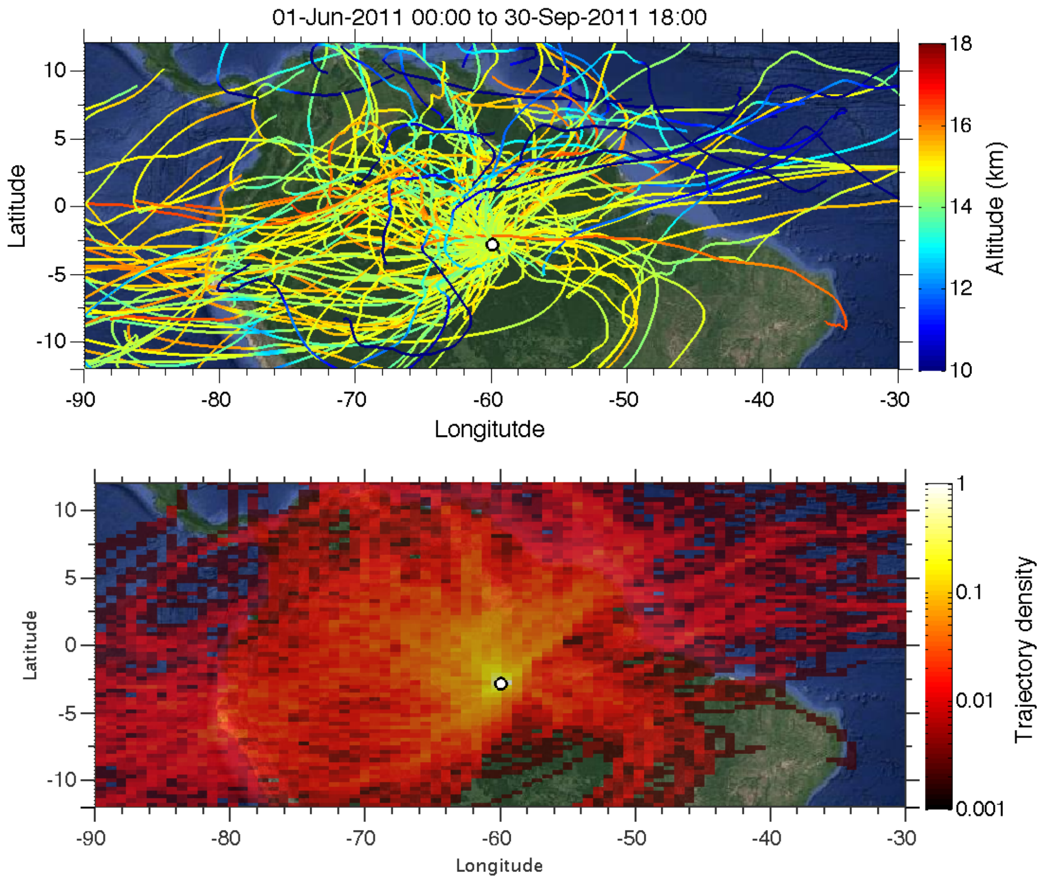

Figure C.1 – Hysplit 7day backward trajectories starting 14.5km above the site every 6h for the four months of the dry season. It should be compared to the top panel of figure 3 in the manuscript.

A way we could make this analysis more quantitative would be to run the back trajectories for each cloud layer detected and use GOES images to locate deep convective cells, and then calculate the distance between each trajectories and the surrounding precipitation (as a function of backward time). This is a huge effort and, we believe, deserves its own paper. Another possibility would be to do that, but just for one case study in each season (e.g. as Fourtin et al., 2007JGR). This, however, would not be very representative of the full dataset.

If the reviewer thinks these plots/analysis are interesting, we would be happy to extend it to the other seasons and include this discussion in the manuscript.

R8: The distribution of clouds as a function of COD also relates to sample bias and attenuation effects.

It is unclear which figure the reviewer is refereeing to. Moreover, we believe that after our reply to comment R6, it is now well explained how we are counting the profiles and how the method takes into account (and corrects for) sample bias and attenuation effects.

R8 (cont.) Yes, there is an exponential distribution of cirrus cloud occurrence with respect to COD (again, see what we have in Campbell et al. 2015). However, the distributions that you have with respect to subvisible, optically-thin and opaque clouds is absolutely not consistent with other studies.

We do not agree with the reviewer. Our distribution is very consistent with other studies! We have listed 7 papers in table 1, including one co-authored by the reviewer, which report distributions similar to ours. The fraction of SVC from these studies varies from 15 to 65% (but also vary the latitude), and we have found 40% of SVC. To mention the specific values: 15% (Seifert et al., 2007), 25% (Antuna and Barja, 2006), 38% (Goldfarb et al, 2001; Hoareau et al., 2013), 50% (Sassen and Campbel, 2000), 52% (Pandit et al., 2015), and 65% (Cadet et al., 2013).

R8 (cont.) There should rougly be a 50-60%/40-50% distribution between transluscent and opaque clouds. In Campbell et al. (2016), we see a very similar distribution as yours that we fully attribute to sampling bias. I see no reason to think this sample is not subject to the same effects.

We believe that after our reply to comment R6, it is now well explained how we are counting the profiles and how the method takes into account (and corrects for) sample bias and attenuation effects. Moreover, the proportion of 60%/40% mentioned by the reviewer cannot be taken as absolute. Firstly, it will not be the same in different locations. Particularly, there is no physical reason why it would be same over the Amazon (i.e. cirrus formed by deep convection, year precip > 2200m) and Greenbelt (frontal systems, year precip < 1100mm).

Secondly, the proportion will be different depending on the algorithm. Let's say the same dataset is analyzed by two different cloud-detection algorithms, one that can see clouds with very low COD (e.g. down to 0.001) and another that can detect only COD > 0.01. Of course, the amount of SVC detected will be very different, and hence the proportion of transluscent / opaque will be different!

R9: Although there is a point where the authors show a correlation between COD and cloud base, cloud base is a nearly useless parameter for such vigorous study. As myriad Sassen papers discuss and describe, cloud top is the most important layer because this is where cirrus cloud nucleate, grow and begin falling. Cloud base, as such, is redundant. Its simply the boundary where evaporation/sublimation is complete in falling crystals. So much effort in the narrative is spent on cloud base and drawing physical correlation, where it seems to have no physical meaning. Cloud top should be the focal point

We thank the reviewer for the suggestion. We will change the manuscript to focus on the cloud top.

R10: As such, there is absolutely no physical basis for evaluting lidar ratio versus mid- cloud temperature. It makes absolutely no physical sense. Now, I recognize that the CALIPSO team has done this very thing with their analyses. I don't agree with them either. But, they are dealing with a downward looking dataset, at least, and this offers other challenges that the authors are not dealing with in the zenith. Whereas I would accept if the authors referenced Garnier et al. (2015) and wanted to leave this as is, I still wouldn't think that it made much physical sense. In particular, as with CALIPSO, you're never actually going to know for certain what the mid-cloud temperature is (or unfortunately the cloud top temperature is) because of attenuation. For CALIOP, this is actually a bigger issues, since they can attenuate working downward with clouds that ground-based lidars would likely never reach. But, the comment still remains. I recommend sticking with what you can physically interpret, and particle effective size and habit are likelier in the long run to relate with available water vapor and temperature found at cloud top than somewhere within the cloud

We thank the reviewer for the suggestion. We will change the manuscript to evaluate the LR as a function of the cloud top temperature.

R11: No uncertainty analysis is provided for the lidar ratio analysis. This concerns me, again, because of the low SNR environments that you claim to be working with. As such, its unclear to me that you can actually develop meaningful correlative relationships, like Garnier, with a relatively low number of cases that the SNR would be sufficient and uncertainty suppressed. The uncertainty term presented appears to me to be a standard deviation, which again seems misrepresentative in context.

In the paper, we did not show any individual retrieved quantity. The tables and figures show average values, sometimes for the whole year, or season, or hours in a day. Hence, the uncertainties of individual retrieval are not given. Depending on the discussion, we reported either the error in the mean value, or the variability of the values.

However, we recognize that we should better explain how we evaluated the uncertainties in the LR and COD obtained by combination of the transmittance and Klett methods. Indeed, this was a request made be reviewer #1 as well.

To make it clear, we have calculated the uncertainties in the optical depth and LR. That comes from a simulation study we performed to access the accuracy and precision of our algorithms. In the simulations, we varied the cloud thickness (from 15m to 4.5km), the cloud extinction coefficient (from 0.02 to 0.1 km^{-1}) and the SNR from 3 to 50. Even for SNR=3, the difference between the true and retrieved LR was < 5 sr for COD = 0.01.

More details about the simulation can be found in the file attached with this answer. This is a draft of the material that we are preparing for the supplement.

R12: Please see my note about how you interpreted Chew et al. (2011). Its not correct. 34% of Level 2 AERONET observations were found biased by unscreened cirrus.

Thanks for pointing that out. We changed that in the manuscript.

R13: I recognize that this is a lot of stuff. I offer this with full respect to what you are trying to do, because its in my direct interest working so many years with MPLNET to see this sort of work get published. I present these thoughts in detail with the sincere hope of helping resolve what I believe to be significant scientific shortcomings in the narrative

We really appreciate your suggestions and, particularly, the time you dedicated for doing such a careful review of our manuscript. Your constructive criticism helped a lot to improve our work.

ATTACHED DOCUMENT – Hand written notes with many suggestions for improving the manuscript text.

We thank the referee for carefully reading. The suggestions for improving the English writing will definitely make the paper easier to follow. We have accepted all suggestions and made the changes in the manuscript.

TO BE INCLUDED IN THE SUPPLEMENT MATERIAL

S.1 Simulations for evaluating the retrieval methods

It is important to know the uncertainties in the retrieved cloud optical depth and lidar ratio, particularly because we are using the transmission method (Chen et al., 2002), which becomes very sensitive to signal noise for low optical depths. To estimate the effect of random signal noise in our retrievals and evaluate the errors for different COD, we did numerical simulations of lidar profiles having cirrus clouds with fixed LR of 20 sr. Cloud-base was fixed at 12 km and eight cloud-thickness were simulated: 15, 30, 45, 90, 150, 450, 1200, and 4500 m. For the cloud extinction coefficient, two values were simulated: 0.02 and 0.1 km^{-1}, thus the COD ranged from 3 x 10^{-4} to 0.45. Random noise following a Poisson distribution was added to the simulated photon-count signal to get signal to noise ratios of 50, 10, 5 and 3 in a single bin just below the cloud base. For each combination of COD and S/N, 100 simulations were performed. The simulated profiles were processed with the same algorithm used for atmospheric data. Therefore, we can evaluate the uncertainty in the COD and LR as a function of the S/N by calculating the mean, the standard deviation, and the standard deviation of the mean over these 100 realizations. The standard deviation will give how the signal random noise might affect the retrievals, while the mean and the standard deviation of the mean will show if the retrieved values converge to the expected values, after many observations.

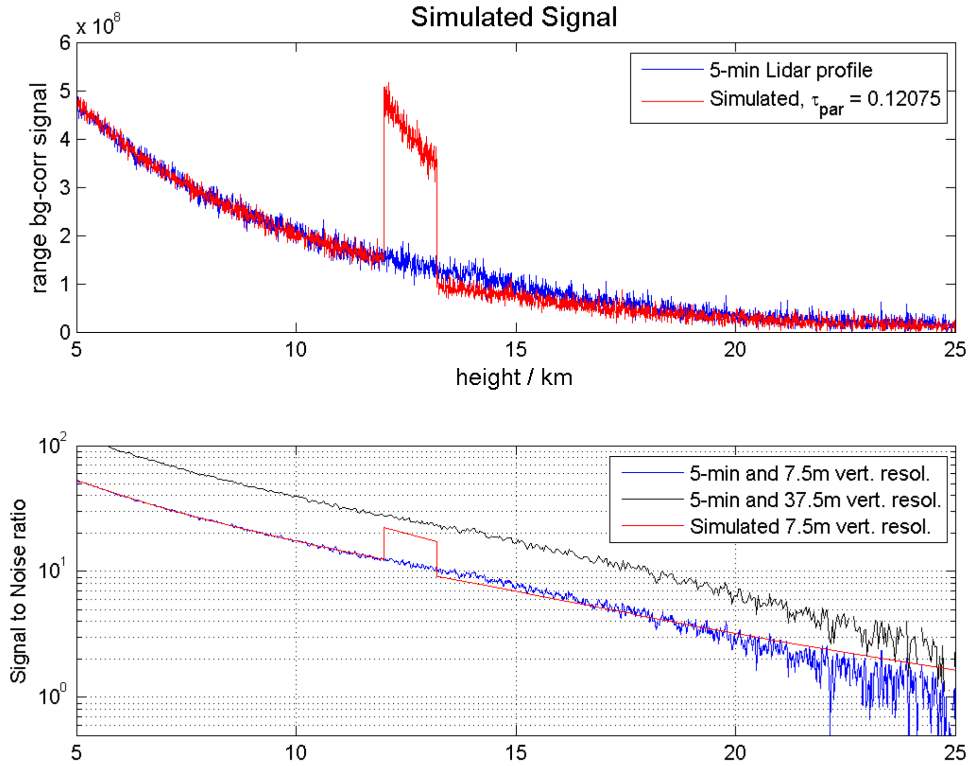

Figure S1: Example of background and range corrected signals (top) and the corresponding S/N ratio (bottom) are shown. The blue curve is a measured lidar profile with 5-min temporal average and 7.5 m vertical resolution, while the red curve is a simulated profile with similar S/N ratio at cloud base and a cirrus cloud of optical depth 0.12. The black curve is the same measured profile but with a 5-bin vertical binning (37.5 m), and thus a higher S/N ratio.

Figure S1 shows an example of a measured profile with 5 min average and original 7.5 m vertical resolution, from some day in July 2011. The system shows a good performance. Typical S/N ratio for the molecular backscatter at 12 km of altitude, for this temporal and spatial resolution, varies from 6 to 20, depending on the presence of low clouds and the solar background. This S/N ratio can be improved, for instance, by reducing the vertical resolution as shown in the lower panel (black curve).

S1.1 Uncertainty in the retrieval of COD

As discussed in section 2.4, for the calculation of the optical depth with the transmittance method, it is necessary to fit the molecular part of the signal below and above the cloud. Considering a large region for those fits, in our case 1 to 10 km, helps to reduce the effect of the noise. The difference between the two fits gives the cirrus transmittance and the optical depth is half the natural logarithm of that value (eq. 2 and 3). Figure S2 shows the mean COD and the standard deviation of that mean value, for the 100 simulations. These results show that the magnitude of absolute mean error (mean COD – truth) is independent of the true COD. The root mean square error (RMSE) is 2.5 x 10^{-3}, for S/N = 3, and only 2.3 x 10^{-4}, for S/N = 50. That is for the averages over 100 simulations, for single profiles it is 10 times larger. The relative error is smaller for large COD values. This error is less than 20% for COD > 0.005 and S/N = 3, and less than 6% for COD > 4.5 x 10^{-4} (minimum value) and S/N = 50 (largest value). We note that even for very low S/N ratio and small COD the method still find a mean value compatible with the true COD.

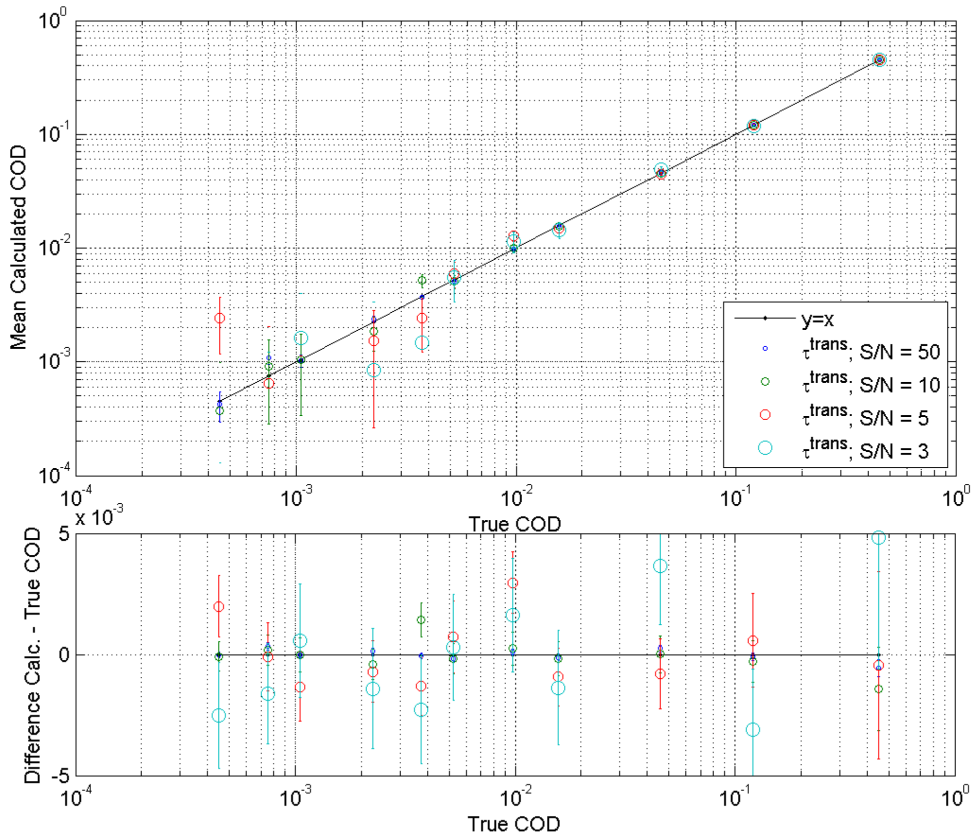

Figure S2: COD calculated by the transmittance method as a function of the true COD for different S/N ratios. The error bars are the standard deviation of the mean values. The absolute differences (lower panel) are all compatible with zero (i.e. mean calculated COD is compatible with true COD).

The cloud optical depth can also be calculated by integrating the extinction coefficient obtained with the Klett method, however an a-priori LR is required. We use this method in two cases. First, when there is more than one cloud layer. In this situation, the transmittance method gives the total cloud optical depth (all layers combined) and that is used to obtain an average LR (all layers combined), the same way as explained in section 2.4. The extinction profile from the Klett method (with that average LR) is then used to divide the total optical depth into contributions from each layer. The second case is when the interactive method described in section 2.4 fails to converge, i.e., when it cannot find a reasonable LR value that makes the Klett inversion give the same optical depth as the transmission method. This happens for XX % of our profiles with clouds and they have very low optical depth, about YY, and only ZZ m of thickness. These profiles are usually those near the edges of the clouds. In these cases, the cloud optical depth is obtained with the Klett method by assuming a LR equal to the average value obtained from all the other profiles (i.e. the ones when we could determine the LR, for the current version of the manuscript it is about 20 sr).

Figure S3 shows the COD obtained by this method in the best scenario, i.e. when imposing the true LR for the simulations (20 sr). We can see that the Klett method is much less sensitive to the S/N ratio. The RMSE is 2.8×10^{-4}, for S/N = 3, and 5.5×10^{-5}, for S/N = 50, both much smaller than the mean errors obtained with the transmittance method, but also closer to each other. As for the transmittance method, the Klett method also finds a mean value compatible with the true COD even for very low S/N ratios and small CODs.

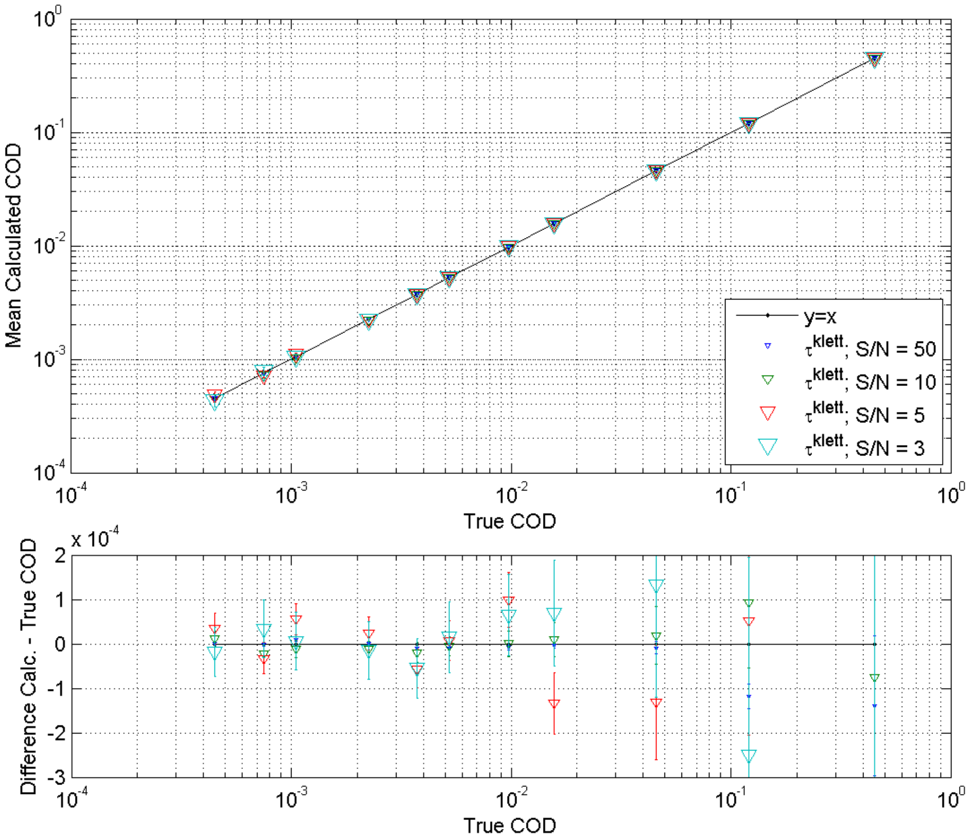

Figure S3: COD calculated by the Klett method, assuming the true LR = 20 sr, as a function of the true COD for different S/N ratios. The error bars are the standard deviation of the mean values. The absolute differences (lower panel) are all compatible with zero (i.e. mean calculated COD is compatible with true COD).

It should be noted, however, that a wrong guess about the LR would bias the retrieved CODs obtained with the Klett method. To quantify that effect, we applied the Klett method assuming a LR value 50% higher and lower than the true value (i.e. 10 and 30 sr) and S/N of 50 (so that it can be disregarded). The result is shown in figure S4 together with the result for the transmission method with S/N ratio of 10. It is clear that the COD retrieved by the Klett method is only as good as the estimative of the LR.

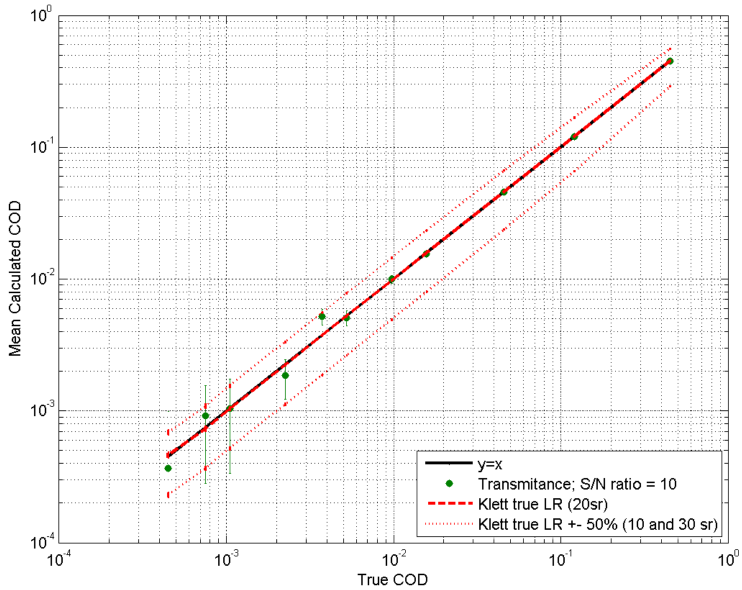

Figure S4: COD calculated by the Klett method, for LR = 10, 20 (true) and 30 sr, is shown as a function of the true COD for S/N = 50. Points in green are the transmittance method for S/N = 10. The error bars are the standard deviation of the mean values.

Figure S5 shows the relative errors for both methods. This is defined as RMSE/True_COD. As expected, the lower the S/N ratio the higher the error. In the worst case, i.e. S/N = 3, the relative error from a single retrieval using the transmittance method is below 20% only for COD > 0.1. For S/N = 10, this limit is COD > 0.025. This error, however, is random and fluctuates around zero as shown previously. By averaging over 100 profiles (right panel, Fig. S5), the relative errors decrease by a factor of 10. Under these circumstances, i.e. with many profiles, or if the S/N is high or if the COD is not very small, it is advantageous to use the transmittance method because it does not depend on an a-priori LR. In our study, we analyzed about 37k 5-min profiles, where 21k had S/N > 3 at 12km and in 14k of these we found a cirrus cloud. Thus, the error in the mean cloud optical depth reported in Table 1 is indeed much lower than shown in the right panel of Fig. S5.

As expected, the relative errors for the Klett method with the true LR are always smaller than those from the transmittance method for the same COD (Fig. S5, compare the respective lines with triangles and circles). However, there is a large uncertainty from the value of choice for the LR. The dashed black line shows the relative error from choosing a LR of 10 or 30 sr. The induced bias in the retrieved COD is proportional to the change in LR, hence the relative error is approximately constant. The 50% change in the LR translates in a relative error of about 50% for small COD, and 30% for large COD.

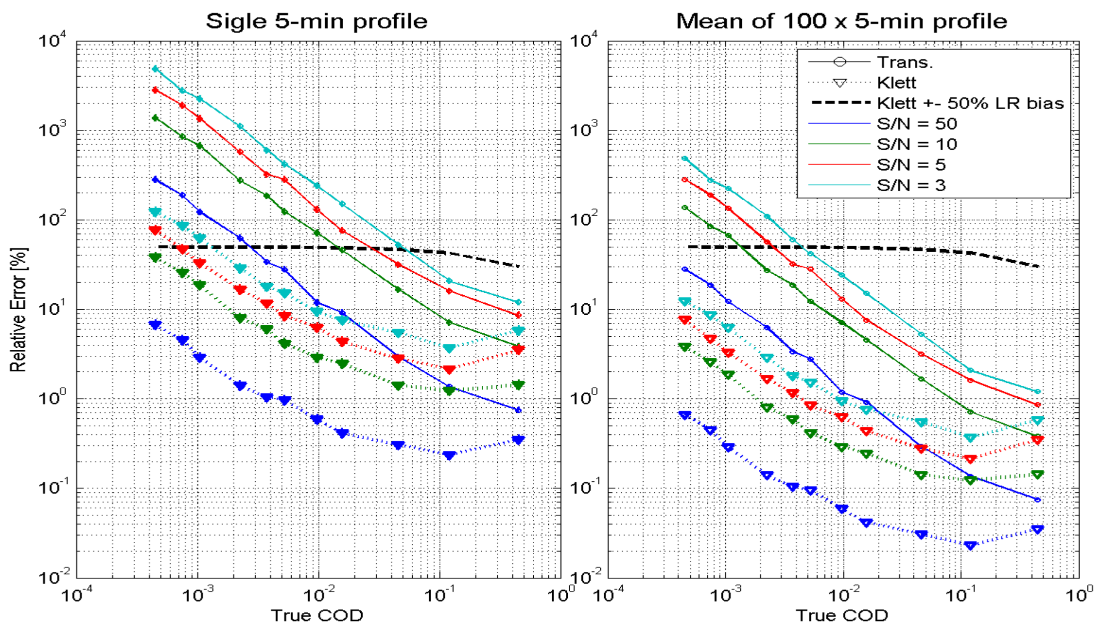

Figure S5: Relative error (in %) in determining the COD for both methods as a function of the true COD for the different signal to noise ratios are shown.

S1.2 Uncertainty in the retrieval of LR

As explained in section 2.4, the LR is estimated by a minimization procedure in which the LR is allowed to vary from 2 to 50 sr. The optimal LR is the one making the cloud optical depth from the Klett algorithm equal to that from the transmittance method. Typically, we are able to estimate the LR for clouds with COD > 0.01, which is about 91% of our observations. Below that threshold, the COD is not very sensitive to changes in the LR and the method does not converge. That is why we estimate the COD, in these cases, with the Klett method and a fixed LR = 20sr.

Figure S6 shows the results of the LR estimated for the same simulated profiles used for the evaluation of the COD retrievals but only for the cases when the LR algorithm converged. When the S/N is low or when the COD is small, there is a tendency of overestimating the LR (all deviations are positive). However, all retrieved values are still compatible with the true value (t-score < 3) and the maximum deviation is just 4.7 sr. Moreover, the variability of the calculated LR (standard deviation, i.e. 10 times the error bars in Fig. S6) decreases with increasing COD. For S/N of 5, it is 12 sr for COD = 0.02 and 3 sr for COD = 0.45. The variability also decreases with increasing S/N ratio. This is shown in the histograms in figure S7, for COD > 0.02 and S/N of 5 and 10. The variability was reduced from 5 to 2.5 sr, respectively. For these cases with somewhat larger COD, it is clear that there is no bias in the mean retrieved LR.

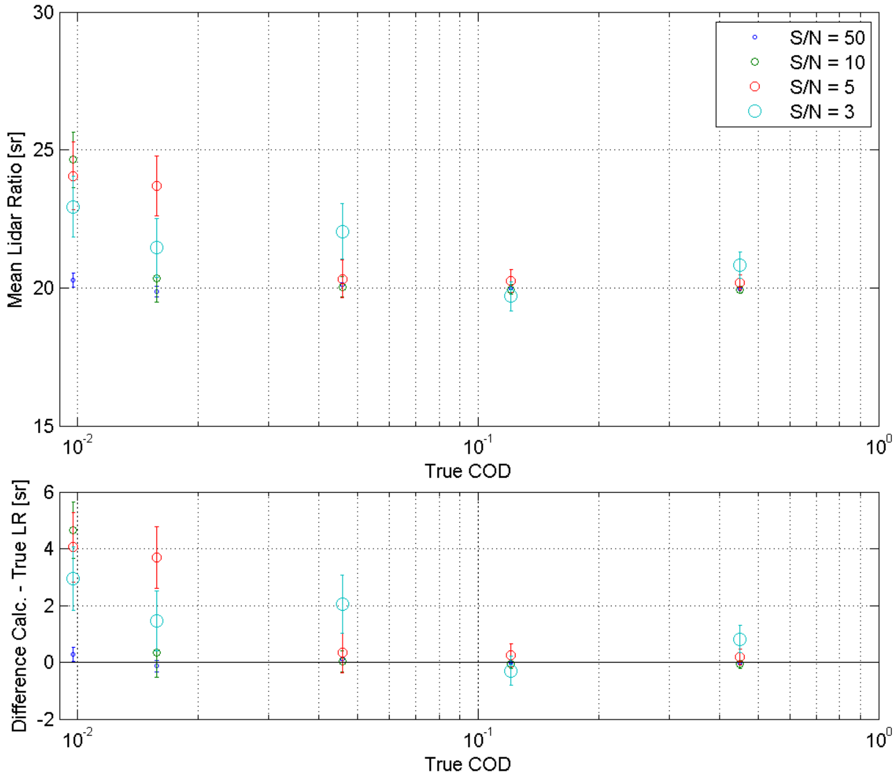

Figure S6: LR calculated from the combination of the transmittance and Klett methods, as a function of the true COD for different S/N ratios. The error bars are the standard deviation of the mean values. The absolute differences (lower panel) are all compatible with zero (i.e. mean calculated LR is compatible with true LR).

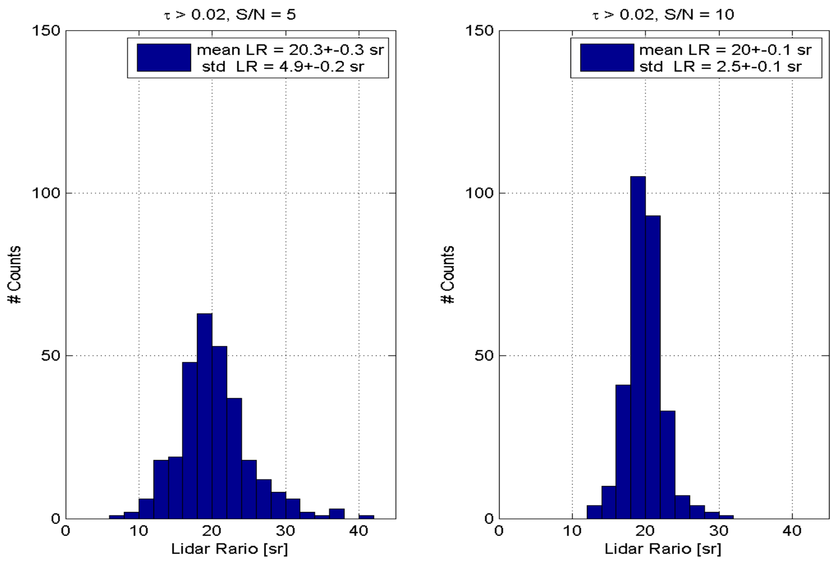

Figure S7: Histograms of LR calculated from the combination of the transmittance and Klett methods for S/N = 5 (left) and S/N = 10 (right) are shown just for COD > 0.02.