Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-317-AC5, 2019

Atmospheric Measurement Techniques
Open Access
Discussions

Interactive comment on "Cloud base height retrieval from multi-angle satellite data" *by* Christoph Böhm et al.

Christoph Böhm et al.

c.boehm@uni-koeln.de

Received and published: 1 February 2019

"The authors describe a novel algorithm for the retrieval of cloud base height (CBH) from MISR satellite measurements. Global information on cloud base height is important for many applications and the retrieval approach is interesting and promising. However, the manuscript is not sufficiently convincing in demonstrating the reliability of the new CBH product. Below are a number of major issues to be addressed before this manuscript may be suitable for publication."

We thank the reviewer for the constructive feedback. We have implemented your comments in the following way:

"The algorithm is tuned with METAR observations over the U.S., i.e. extratropical land surface. How representative is this for ocean surfaces and other climate zones, where different cloud types prevail? To show the skill in other regions, some comparisons with independent measurements elsewhere would be required."

Our study includes cloud height retrievals over the continental U.S. over the course of two years (one year for calibration, one for validation). Therefore, various cloud types should be included in the analysis. Clouds within Arctic air masses which typically occur in the northern U.S. during winter, as well as tropical like deep convective clouds which typically occur during the summer in the southeastern U.S. should be included as well as stratocumulus clouds which usually occur at the coast of California. The METAR data set includes maritime island stations in the Gulf of Mexico and near the west coast of the U.S.

The utilized MISR product does not distinguish between land and ocean, but only between cloud and surface by a geometric technique. Furthermore, for each retrieval scene a particular configuration of the stereo-derived cloud mask is provided which characterizes the scene structure. Therefore, if a similar scene structure is found outside the continental U.S. region, for which the calibration and validation has been carried out, MIBase should perform similarly. However, we agree, that additional validation in other regions would be beneficial to backup this statement.

To further investigate MIBase limitations in respect to cloud types, we introduced two new sections "3.4 Scene limitations" and "4.1 Scene structure influence" into the manuscript and modified abstract and conclusions accordingly. In short: The statistics are rather robust with regard to the configuration of the stereo-derived cloud mask. The bias depends on the number of z retrievals marked high confidence cloud. This indicates that a bias correction might be feasible. Since the origin of the bias is not fully understood yet, we like to leave such potential improvements to future studies.

Furthermore, the new Section 4.1 includes an investigation of the influence of z_{top}

and Δz on the MIBase performance. Such parameters are also important for cloud type classification. High correlation coefficients (supplement Fig. 2h) and low RMSE (supplement Fig. 2f) for cloud thicknesses up to 1000 m indicate that the algorithm works particularly well for thinner clouds. For cirrus clouds or high clouds in general, we cannot make a robust statement as we tax the accuracy of the METAR reports insufficient for this particular height range.

"More information on the success rate of the retrieval algorithm is required to evaluate how useful it is. Statistics of the number of samples n_s are given but these are only absolute numbers, not (fractional) success rates. For example, Table 3 indicates that n_s is between 3059 to 7772 depending on R_fv. A rough calculation based on 1510 ceilometers, a MISR revisit time of 6 days, and a cloud fraction of 50% would potentially yield around 45,000 cloudy collocations. This suggests that in only 10% of the cloudy cases, a valid CBH retrieval is obtained. Is this correct? Such statistics, accompanied by the relative occurrence of different causes of retrieval failure, need to be provided, also for the global plots, to evaluate the applicability of the method."

We added statistics regarding the success rate of the algorithm to the manuscript (Section 3.4 and Section 5.1 of the revised manuscript). For the comparison with the retrievals from METAR, we combined the numbers for all sites and present the resulting numbers for the years 2008 (calibration) and 2007 (validation) individually in the new Section 3.4 including new Table 5. For the year 2008, we downloaded data which provided 80454 overpasses. Only 65 % of those contained valid z retrievals at the METAR sites with a corresponding METAR message. Out of those potential cases, about 30 % do not include z retrievals marked high confidence surface. Please, see the revised manuscript for more details. For the global application, the calculations are carried out for each grid cell, so that the spatial distribution of the numbers can be studied (Fig. 10 of the revised manuscript). We added a discussion on the retrieval failure statistics to Section 5.1.

"The calibration of the algorithm is done for z_base smaller than 3000m, because of the limited range of the ceilometers. However, for the global composites an upper threshold of 5000m is used. It is unclear whether this extrapolation outside the range for which the product has been trained, is valid."

One of the questions this manuscript tries to answer is whether the cloud base height can be derived from the MISR cloud top height product. The comparison with the ceilometer demonstrates that this is possible in case some preconditions are met. The reason this works is because typically the cloud top height is heterogeneous leading to geometrically thinner and thicker parts of the cloud. As far as we understand, this concept should not change for different heights within the troposphere. This means, if it can be validated within one height region, it should work in other height regions as well. However, we do agree that the algorithm might perform differently for different cloud types. Then of course, for different heights the distribution of cloud types varies. We discussed the dependence on cloud types in the reply to the first comment.

"The global maps in Figs. 9 and 10 are hard to interpret because upper limits of z_base and z_top have been applied. What does the median of a distribution cut-off at some value tell us?"

Applying a threshold is necessary to exclude high clouds from the analysis. This is appropriate, because we are focusing the study on clouds which occur in the lower troposphere. In our opinion, the median of z_{base} and z_{top} provides less valuable information if low and high clouds are mixed together. From our best judgement, 5000 m seems like a good choice for a threshold to ensure that the algorithm works properly. The resulting product is not highly sensitive to this threshold as can be seen in Fig. 4 (supplement). Inspired by reviewer 6 (RC4) who questioned the height limit for z_{top}, we changed the calculation of the median z_{top} height. We reproduced the figures by calculating the median only for those z_{top} values for which the respective z_{base} is below the 5000 m threshold. We updated Fig. 9 and Fig. 11 of the revised manuscript and their respective captions accordingly.

"I'm also confused by the description of Fig. 9, which says that the ITCZ is clearly visible with higher z_base and z_top. In the plots a brown band can be seen, but these are lower rather than higher values. Can you explain?"

In Fig. 9b and Fig. 10a, 10b (Fig. 11a, 11b in the revised manuscript), the ITCZ is revealed by the light turquoise band slightly north of the equator, indicating higher z_{base} and z_{top} compared to the immediate surroundings to the north and south. This band is most pronounced in the Pacific ocean. Over the Atlantic it can be seen most clearly in the manuscript's Fig. 9c, which shows a band of increased cloud vertical extent in that region. As stated in the manuscript, over continents the diurnal cycle should be kept in mind. MISR has a morning overpass which means, the three year median heights provided here represent the morning heights around 10 a.m. local time. For the Congo Basin, Taylor et al. (2007) investigated the diurnal cycle of cloud top temperature (CTT) retrieved via satellite remote sensing (SEVIRI). According to them, the CTT is lowest around the MISR overpass time with a mean value of about $290\,\mathrm{K}$ during late morning hours. If we take the observed z_{top} of about $1200\,\mathrm{m}$ and assume a lapse rate of $0.6\,\frac{\mathrm{K}}{100\mathrm{m}}$, the extrapolated surface temperature would be $297\,\mathrm{K}$ ($\approx 24°\mathrm{C}$) which seems very plausible.

"Is it also possible that the results in these multi-year median are biased to certain cloud types? For example, in the stratocumulus (Sc) areas west of the continents, cases with closed Sc will probably not yield a valid retrieval, while for open Sc z_base can be retrieved, so that the end result will be biased to the latter."

An inherent bias of the method results by the necessary condition of a MISR z retrieval which is marked high confidence surface by the stereo-derived cloud mask. In other words, a broken cloud scene is required. Therefore, cloud base heights for situations

with overcast are not included in the calculation of the three-year median. So yes, the statistic is biased towards particular cloud types. We stated this limitation more clearly in the manuscript by mentioning the limitation in Section 3.1 and additionally in the conclusion. Apparent overcast situations which prevent a valid MIBase retrieval occur mainly in the mid latitudes over ocean and in the subtropical stratocumulus areas which you mentioned as an example (Fig. 10d of the revised manuscript).

"The authors define percentile (P) values of the MISR lowest cloud layer z distribution to obtain z_base and z_top. For z_base one would expect P=0 because z_base should be lower than any MISR-derived cloud-top height. The chosen value P=15 is motivated by the noise in MISR z, which makes sense. However, for z_top I do not understand the chosen value P=95. All MISR z values are actual estimated cloud top heights. The logical way to aggregate these is to average the individual z observations or take the median. In other words, a value P=50 seems natural. The choice of P=95 should thus be motivated."

Without further validation, we apply the 95th percentile rather than the median, as we do not want a height which might be representative for the whole area, but rather an estimate of the highest top of the cloud especially for a heterogeneous cloud top height. The focus of this study is on the z_{base} retrieval method and its validation. The use of z_{top} retrievals serves only auxiliary purposes. It is a measure to describe the individual cloud scenes better. For instance, it allows a qualitative assessment of the algorithm's performance in dependence on cloud vertical extent. We extended the motivation of the 95th percentile at the end of Section 3.1.

"Cloudsat is, especially in combination with Calipso, arguably the most accurate source of cloud base height (as well as cloud top height) information from space. Surprisingly, Cloudsat is not mentioned at all (except for two remarks in the context of the Desmons et al. paper) in the manuscript. At the least, Cloudsat should be discussed, and it

would also be good to make some comparisons with this instrument, even if direct collocations with MISR supposedly only occur at high latitudes."

We agree, that it would be very beneficial to have further data sets to compare the method to. However, CloudSat has limitations in estimating the cloud base height, in particular for low liquid clouds. It does not detect small droplets at the base of the cloud (Sassen and Wang, 2008) due to its detection limit of ≈ -28 dBz. Furthermore, retrievals are degraded in the ground clutter reigon (Tanelli et al., 2008; Marchand et al., 2008). Mülmenstädt et al. (2018) evaluate the 2B-GEOPROF-LIDAR product (Mace and Zhang, 2014) which uses a combination of CloudSat and CALIOP retrievals. From this product, they extracted the LIDAR only and RADAR only subsets and compared the cloud base height retrievals with ceilometer measurements similar to the reference data utilized in this study (their Fig. 9). Within their study, the RADAR does not perform as well as the LIDAR. In fact, they find a correlation of 0.265 and an RMSE of 782 m for the RADAR only subset. Therefore, we believe CloudSat would not be suitable as reference data for our study.

A comparison with Cloudsat would require the identification of collocated measurements, which you mentioned would occur at high latitudes. CloudSat has an afternoon overpass, while MISR on Terra has a morning overpass. Therefore, a comparison would also require a discussion on the impact of the temporal difference. We believe, that this paper should focus on the introduction of this new z_{base} retrieval method, its calibration and validation, statistics on the success rate, and a discussion on its global application. Further comparisons go beyond the scope of this paper.

"P1, Abstract: The abstract does not include any information on the cloud types the algorithm is applicable to. In the manuscript this information is also too limited. Does the method work for cirrus, or for deep convective clouds?"

We added that overcast cloud scenes are not included in the statistics to the abstract.

Further, we added "The impacts of the cloud scene structure and macrophysical cloud properties discussed", to alert the reader, that this is an issue.

Our study does not focus on cirrus clouds. Since the MISR z retrieval method is in the visible light range, thin cirrus clouds are probably not included, since they are almost transparent. Further, we cannot validate cirrus clouds, as the METAR data does not include reliable retrievals for high clouds. Deep convection might be problematic, as the thinner cloud edge might not be as pronounced or hidden by the towering cloud. However, deep convection should not be a major issue, even in tropical regions, because of the morning overpass time of MISR on Terra.

"P2, L8: Stephens et al. (2002) is mainly about Cloudsat. It's not the appropriate reference for CALIPSO."

Thank you for pointing this out. We changed the reference by citing Winker et al., 2010.

"P4, L4: Is 'aftward' correct English?"

We think aftward is correct English. It is also used in the MISR product documentation. However, we will follow whatever guideline the editor suggests.

"P5, L6: 'measurements': what is measured?"

The signal return is measured. We rephrased the sentence.

"P5, L9: Is the value 5000 ft correct? It seems such a big jump from 100 ft and 200 ft in the two lower height categories, respectively."

Sorry, that is a typo. According to the "Automated Surface Observing System User's Guide"(National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, and United States Navy, 1998) it is 500 ft (≈ 150 m).

"P5, L10: This suggests that bins is the same as clusters, which is not the case, I assume."

We delete "bins" from the sentence.

"P8, L1-3: This paragraph looks out of place here. Suggest to move it somewhere else."

We moved it to the end of Section 2.1.

"P8, L7: Suggest to replace 'the estimated' by 'a typical'."

Done.

"P8, L15: what are 'z pixels'?"

It refers to the z retrievals from the MISR cloud product. We replaced "pixel" by "retrieval".

"P8, L25-26: Does this mean that the case in Fig. 4 is not included in the statistics? Isn't it a bit strange to present a case study that is not part of the selection applied furtheron?"

This case was studied before we decided which years we would use for the comparison. The main reason why it is shown, is because it illustrates the way the algorithm works and the parameters which are used. Preferentially, the presented case study should be a multi-layer case so the applied layer distinction can be illustrated as well. However, as mentioned in the manuscript, any multi-layer case will be masked out and not be included in the selection for further processing in order to calibrate the algorithm.

"P10, L5: Fig. 4 includes only one 'h_gap'."

It is only illustrated once. However, h_{gap} is calculated for any height gap between two z retrievals and then tested against the threshold. If it is greater than 500 m, the retrievals below and above the gap are treated as separate layers. In the case study, an apparent third layer which is about 1000 m above the top of the middle layer, is revealed in the density plot (Fig. 4 in the manuscript).

"P10, L13-14: The second layer detected by MISR has a base height between 5000 and 5500 m a.s.l. The ceilometer detects layer base heights at 853 m, 2286 m, and 7010 m a.g.l. None of these seems to match with the second MISR layer. Can you explain?"

Thank you for pointing this out. From the distribution of the z retrievals (Fig. 4 in the manuscript), we can distinguish three apparent cloud layers. The highest ceilometer retrieval seems to correspond well with the top layer (between 7000 m and 8000 m). The lowest ceilometer retrieval corresponds well with the bottom layer. We corrected the last sentence of Section 3.2 accordingly. The second ceilometer retrieval roughly matches the top of the lowest layer detected by MISR. The connection to the bottom layer detected by MISR might be an indication of a varying cloud base height within this cloud field. It could also be due to the temporal mismatch between the measurements.

"P12, Fig. 6: I assume this figure is for N=10 and R_fv=10 km. Can you add this to the caption for completeness?"

Yes. Done.

"P12, L6: N=10 seems a relatively low number and one could wonder whether P=15

makes sense for such a small N. Can you comment?"

As discussed in the manuscript, $N = 10$ is a compromise. If a higher N is chosen, the performance improves only slightly. At the same time, the algorithm would neglect more potential retrieval scenes. As addressed above, the bias increases for an increasing number of z retrievals marked high confidence cloud N_{HCC} (Fig. 8k of the revised manuscript). This indicates the potential for a bias correction. Another way to decrease the bias could be carried out by defining the selected percentile as a function of N_{HCC}. However, as mentioned above, the origin of the bias is not fully understood yet. Therefore, we like to leave such potential improvements to future studies.

"P13, Fig. 7: Can you add a color bar? Is the scale linear or logarithmic?"

The color indicates a normalized density. It is a linear scale. Contour lines are shown with the corresponding values on them. We modified the caption to point this out.

"P14, Table 4: Do you have any explanation why 2007 has 30% more valid retrievals than 2008?"

We found about $18\,\%$ more overpasses with valid z retrievals in the fields around the METAR sites. And out of those we did not have to neglect as many apparent clear sky cases. See the new section 3.4 in the manuscript for more details.

"P14, L5: Certainly the different measurement geometry (point over time versus circular area instantaneous) can cause differences. But why would this lead to a bias, and why to a negative bias of MISR in particular? Can't you tune the overall bias to near zero by increasing P?"

As we added to the manuscript, "the bias obtained in this study can have different sources: the different sample volumes of the defined MIBase cell and the ceilometer,

biased MISR z retrievals, various scene characteristics." As of now, we found that the bias seems to depend strongly on N_{HCC}. Simply modifying P to tune the bias to zero is overlooking this relationship. As mentioned earlier, an adaptive P depending on N_{HCC} or an appropriate bias correction would improve the algorithm. However, this goes beyond the scope of this study.

"P15, Fig. 8: I'm not sure how useful this distinction in two geometrical thickness classes is, in particular because this thickness is based upon the MISR retrieval itself."

We substituted this distinction by a more in depth discussion about the influence of the scene structure (new Section 4.1). The usage of z_{top} and Δz provides additional information which characterizes the scene structure beyond just the value of z_{base}. Therefore, it is justified, to use this information to further study the performance of the algorithm.

"P15, L3: The termination of the z_base range by the threshold height relates mostly to the lower thickness class, so it would be better to write: 'The smaller E for clouds with a smaller Delta z ..'."

We cut this part out.

"P16, L6: Sentence ends unexpectedly."

Fixed. Thank you.

"P17, L1: The sampling size is in Fig. 9c."

Yes. Fixed.

"P19, L3: effect should be affect."

Fixed

"P19, L32: A mean difference of 500 m is quite large relative to the retrieved z_base in Fig. 11, which appears to vary between 800 and 1200 m for the selected region. Is it reasonable to assume that the model can simulate a reasonable seasonal variation of z_base if it has such a large bias?"

Yes, we think, it is reasonable. The processes responsible for defining the height of the cloud base are different from the processes which produce the seasonal cycle. Models generally underestimate the maritime boundary layer height in the stratocumulus regions. However, the radiation forcing, and the strength of the subsidence which follow an annual cycle can be represented in the model with higher accuracy leading to a realistic seasonal cycle, despite the revealed bias.

Please also note the supplement to this comment:
https://www.atmos-meas-tech-discuss.net/amt-2018-317/amt-2018-317-AC5-supplement.pdf

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-317, 2018.