Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-276-RC1, 2019

Atmospheric Measurement Techniques Discussions

Interactive comment on "Comparison of the cloud top heights retrieved from MODIS and AHI satellite data with ground-based Ka-band radar" *by* Juan Huo et al.

Nina Håkansson (Referee)

nina.hakansson@smhi.se

Received and published: 12 September 2019

1 General comments

The paper compares cloud top height retrievals from two different algorithms on different sensors (MODIS and AHI) with ground based radar data over Beijing. Retrieval accuracy was found to be comparable for the two imagers/algorithms and better for thicker clouds. Results where presented in terms of bias, standard deviation and the percentage of retrieval errors within 0.25, 0.5 and 1km. Overall the paper is well organized and clear but some claims are not supported by the results and references to

previous validations of MODIS collection 6 using radar data is missing.

2 Special comment about a missing reference

I am the first author of paper Neural network cloud top pressure and height for MODIS Håkansson et al. 2018: https://www.atmos-meas-tech.net/11/3177/2018/amt-11-3177-2018-metrics.html. This is a recent paper (2018) evaluating MODIS collection 6 like this paper (not MODIS collection 5 as most of the currently referenced papers) using global space borne CloudSat (CPR) radar data. The results are comparable to the ones in this paper. Further it does discuss which statistical measures are appropriate for describing the resulting non-Gaussian CTH error distributions. I apologize in advance for bringing up my own research but I believe it is very relevant in this case and that this is an important reference that should be included in the article. I am sure the editor or other reviewers can help you decide on inclusion or not, as I am biased.

3 Specific comments

1. Line 12: *"Large differences were mainly occurring for the retrieval of thin clouds of CD < 1 km, especially clouds higher than 4 km"*. seem to be in contradiction to the results at line 13-14: *"MODIS CTHs greater than 6 km showed better agreement with the radar data than those less than 4 km"*. High clouds can not both have better agreement and larger differences? It would be clearer if you add a sentence detailing that radar high thin clouds with large differences will typically have low MODIS CTHs.

2. Line 23: *"Statistical analysis showed that the average AHI CTHs were lower than the average MODIS CTHs by -0.64 ± 2.36 km."* For me it was not immediately

clear that AHI CTH is statistically significant lower than MODIS CTH given the error distribution. As we are only looking at a sample and the true distribution might very well be centered at zero (as the SD is large and bias is small). Inclusion of a one sample two sided t-test showing that the true bias is not equal to zero would give more support to the claim. However assuming a sample size of 600 the result seem to be significant.

3. As seen in figure 4-b the error distributions are non-Gaussian. This makes the interpretation of bias/STD difficult for the reader (see Håkansson et al. 2018 for a longer discussion). Inclusion of medians (or modes) and mean absolute error (or interquartile range) would be helpful. At least medians should be included and discussed.

4. Line 29: change "is" to "can be" or reformulate to make it clear that the assumption that *"cloud is regarded as black bodies"* is not made by all algorithms. And at least it is not made for all cloud types. Clouds regarded as black bodies are also regarded as opaque. And many algorithms handle also semi-transparent cloud with some skill.

5. Line 30: Remove the word *"Surface"*. I agree that active sensors are ideal for accurately detection of CTH. Space borne active sensors have the benefit of global coverage. What is the benefit of ground based ones? Is there smaller problem with clutter for ground based radars compared to space borne radars?

6. Line 38-39: Many references to MODIS collection 5. Baum et al. is the only one using MODIS collection 6 data and it uses space borne lidar (not radar) for validation. I suggest updating the references list, and at least add more studies of MODIS collection 6 validated with radar data.

7. Line 40-44: It is better to include validation results for MODIS collection 6 as that is what you are using. And especially as Baum et al. 2012 shows that CTH is

much improved in Collection 6 compared to Collection 5.

8. Reformulate/ or remove *"Evaluation results from previous studies are not representative of specific regions".* The results in global investigations might or might not differ from specific regions, it can not be assumed to be different. It might be different though and that is one reason why your investigation is important.

9. Line 70: A remark: Note that as the CO2 slicing is used only for mid- and high-level clouds, height estimation is needed before the height retrieval can start. A bit of a hen and egg problem.

10. Line 80: *"Most published evaluation studies on the MODIS cloud top properties are from the Collection-5 version datasets"*. Include and relate your results to some more studies using MODIS Collection 6 CTH data.

11. What is the intercept method and what is the interpolation method? Please describe them.

12. What method/algorithm/product is used to determine which clouds are semi-transparent?

13. Line 107: Are any such *"short-term collaborative observations"* affecting the data used in the study.

14. Line 115: Did you use a lower threshold to include more clouds with weak returns? Or would a lower threshold include more clouds with weak returns that you do not wish to include? Is there a risk for non clouds contamination the results, like aerosols or insects? Please clarify.

15. Table 1. That is a high vertical resolution of 30m! Does this mean that the thinnest clouds detected (three cloudy bins) are only 90m thick? What is the horizontal resolution of the radar data? Please include it in the table.

16. Line127-130: *"For comparison with satellite data, for multilayer clouds in a period, the CTH is also the average CTH of all cloudy profiles even if the upper-level clouds do not cover the lower-level cloud, rather than the average CTH of the upper-level clouds."* This sentence is confusing. Is at any point the CTH of the second level of a multi-layer radar profile included in the averaging? I think the first part of the definition of the radar CTH is very clear as it is. Can the second part be reformulated more like: *Note that for multilayer clouds only the CTH of the highest cloud is used. In the case of a cloudy period/scan with a low cloud which is partly covered by a high second layer cloud the resulting CTH will be an average of the CTH of the upper layer (from the parts with multilayer clouds) and the CTH of the lower layer (from the single-layer parts of the period/scan).* Or include a figure to make it clear what is done.

17. . Line 133: Suggestion: replace *"data covers larger areas"* to *"have larger field of views*".

18. . Line 134: Reformulate (or at least remove *"Thus"*): *Thus, temporal and spatial collocation of the radar, MODIS and AHI data is critical to facilitate effective comparison and evaluation*. Note that temporal and spatial collocation is necessary regardless of data coverage, FOV (Field Of View) differences, or repetition frequencies. It might be more straightforward in the case of similar FOV and repetition frequencies, though.

19. Line 148-149: I do not understand the first part of this sentence: *"According to the climatological distribution of clouds, the ground-based CTH measurements from the Ka radar were averaged within 10 min of the MODIS observation time ($\pm$5min) in this study"*. Please reformulate.

20. Line 149-150: Note that averaging data in time or space is optional. (In Håkansson et al. we successfully used nearest neighbor matching between CloudSat

(CPR) radar data and MODIS collection 6 data). And note that your method, using all MODIS data within 5km and 10min of radar data instead of nearest MODIS 1km pixel and 1.7 minutes of radar data, might decrease the part of imager data actually seen by the radar. The effect would depend on the horizontal resolution of the radar which is unknown to me. A motivation for the averaging of the MODIS data should be included.

21. Line 144: *"These collocation methods were designed to match the research goal."* Reformulate to make it sound less as researchers are designing experiments to achieve before-hand determined results. Researchers will make best effort to choose as sound settings as they can. Averaging in time and/or space can improve the study (for example maybe decrease effect of outliers), but can also introduce new problems. For example in situations with clouds of two heights averaging will introduce new types of clouds not present in the original data.

22. In Figure 2: Please note with a different marker (for example x) which MODIS pixels where included in the averaging for each case. Also make the size of the radar dot match the field of view of the radar if it does not already.

23. Equation 4: Reformulate or split to two separate equations. Now, because of the parenthesis the reader get the impression the MODIS height H_m is a function of the AHI height H_a. A formulation like $D_{mr/ar} = H_{mr/ar} - H_r$ would be better.

24. Line 170: How was the poor quality data defined? Which data were of poor quality MODIS, radar or both? Please clarify.

25. Line 173: Nice to see these statistics suitable for non-Gaussian error distributions: *"Among all comparisons, about 14% differences were less than 0.25 km, 27% were less than 0.5 km and 49% were within 1.0 km."* In Håkansson et al. we found for, MODIS collection 6 compared to space borne CloudSat (CPR) radar data, the part of errors higher than (0.25km, 0.5km and 1km) to be (84%, 70%

and 48%). This would correspond to (16%, 30% and 52% of comparisons with in 0.25km, 0.5km and 1km. Considering that our investigation was global and with a space borne radar compared to this investigation using a ground based radar at a single point I think results are noticeably similar.

26. Line 181: First I was confused as I did not understand you were splitting results into high low with respect to MODIS CTHs in this sentence. I assumed you would use the radar CTH. But now I understand and the result makes sense as the difficult thin high clouds would have a high chance of ending up as MOIDS low clouds. Please make clearer.

27. Line 205: Cloud occurency frequency. Would *radar cloud fraction* be a better description? The definition is a bit unclear and could be improved. What does cloud time mean? What is observation time? Is it correct that COF = number of radar cloudy profiles divided by the total number of radar profiles (within the 10min time window)?

28. Figure7-b: Should it be bias in the legend of the red line? Please include also medians in Figure 7-b.

29. Figure 8 and Figure 9: Please note that it is AHI COTs that are used also in the figures.

30. There are many results in the text; one or more tables giving an overview of the results would help the reader.

31. Line 236 and Line 281: *"Statistically, the AHI retrieval algorithm showed better performance for multilayer clouds than single-layer clouds"*. Better performance for multi-layer clouds compared to single-layer clouds is the opposite of what would be expected for any CTH algorithm. This means that strong evidence is needed to support such a claim. Note that bias is better for multilayer clouds but

STD is higher. So there is no support for the claim that AHI would retrieve more accurate CTH for multi-layer clouds compared to single-layer clouds. Please reformulate.

32. Line 237: *Compared with MODIS, the AHI retrieval algorithm showed a slightly better performance for multilayer clouds.* Can these numbers really be used to say one algorithm is better than the other? If I understand correctly, in these investigations it is not only cases where all three instrument match (AHI, MODIS and the radar), instead there is one data set with all MODIS-radar matches and another with all AHI-radar matches. And there are only 210 multi-layer MODIS-radar matches which is a quite small sample size. And this means that the population bias of all multi-layer clouds actually can be quite different from the sample bias (-1.23km) calculated on 210 samples. A difference between sample bias and actual bias in the order of $2.6SD/\sqrt{210} = 0.5km$ is very well possible. Therefore I think there are not enough results to support the claim that AHI perform better than MODIS for multilayer clouds. In my opinion all three datasets should be matched, or/and the differences between the distributions (or the number of samples) need to be larger in order to form enough support for the claim. Please reformulate or update with a statistical test supporting the claim.

33. Line 237-238: Use COF defined on line 205.

34. Line 241: Is it as in the other investigations: MODIS data are averaged and for AHI the nearest pixel is used?

35. Line 243: Please include also median differences.

36. Line 256: As the error distributions are non-Gaussian I would be very cautious to recommend using bias and STD in any meteorological application.

37. At least add median to Figure 10. The difference between bias/and median at

least show if the distributions are skewed. Even if the high kurtosis of the error distributions would still be hidden.

38. Line 268: Note that all CTH algorithms do not include radiative transfer models. Or maybe they do, at least implicitly? Please motivate or reformulate.

39. Line 272: At least include also median.

40. Line 284: It is not certain that MODIS had the lowest accuracy in spring. It had the lowest bias (but bias is not the same as accuracy, especially not for non–Gaussian distributions!). Note that also standard deviations were low for MODIS in spring which would indicate better accuracy. If you want to use a single measure to evaluate the accuracy of the algorithms I would recommend mean absolute error. Please update.

41. Figures showing the distributions for D_{ar} and AHI-MOIDS should be included. Similar to the one for D_{mr} in Figure 4-b.

4 Technical corrections

- Line 231: “might due to” => “might be due to” or maybe better “could be caused by”

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-276, 2019.