Reviewer 1

Dear Editor, dear authors,

The authors made an effort to improve the manuscript and to support their presented results. The previously raised points have been properly addressed. I recommend accepting the manuscript for publication after the following few minor/technical points have been addressed:

- L422-428 + L461-472 The presentation of the joint evaluation dataset is still a bit confusing to me. The amounts of samples between the two ML-based methods naturally differ as clearly clarified now. However, the dataset for which the scatter plots of Figure 2 are produced is not clear. I understand the authors compile colocated samples for the year 2017 and then randomly split these samples for training and validation. Is this validation set then used for all the evaluation (plots and metrics) presented? Overall, I would encourage the authors to centralize the information about the joint dataset in the section 3.4 to improve clarity (eg. move the beginning of section 4.1 to this section).

Answer: Thanks for your suggestion. In Figure 2, we used the exact same set of instances to test the four algorithms. As per your suggestion, I moved the beginning part of Section 4.1 to Section 3.4.

- "... physics-based methods ..." (L416)

Answer: Thanks for your nice advice. We have revised the sentence.

- L615 "Figure 5a and 5b..."

Answer: Thanks for your nice advice. We have revised it.

- L741 "... data from 2017 and 2019."

Answer: Thanks for your nice advice. We have revised it.

- Figure 1 "... by the four independent ..."

Answer: Thanks for your nice advice. We have revised it.

- Figure 3 There is one too many (a) indicators in the caption. Using the same colors

for the respective method retrievals as in all the other figures (5, 6, 7, 8) would be perfect.

Answer: Thanks for your nice advice. As per your suggestion, we have revised the Figure 3.

- Figure 7 Remove UTC in the caption as the local time is used.

Answer: Thanks for your nice advice. We have removed UTC.

- Extensive review of the use of acronyms was done by the authors in the revised manuscript, just very few cases remain (L326, L392, L666 in track changes).

Answer: Thanks for your nice advice. We have unified the description of all cloud base height in this paper.

Reviewer 2

Technical note: Applicability of physics-based and machine-learning-based algorithms of geostationary satellite in retrieving the diurnal cycle of cloud base height – Wang et al. (egusphere-2024-1516)

This paper addresses several interesting but somewhat diverse research topics, and I appreciate the authors' efforts in conducting this work. However, the manuscript gives the impression that multiple studies are being combined, which can make it difficult to discern whether the primary focus is on the development of the ML algorithms or the accuracy of the CBH products from each algorithm. I believe that the organization and clarity of the presentation could be improved to better highlight the main objectives.

If the focus is on the diurnal cycles of CBHs, some sections, such as the MODIS product evaluation, might not be necessary and may detract from the central message of the paper. A more concise reference to this information, perhaps by adding appropriate citations, could suffice. Streamlining the content in this way could help avoid the level of detail typically found in a graduate thesis and keep the reader's attention on the core contributions of the paper.

Another main concern is, regarding line 251 and others:
I find the lack of discussions or additional explanations of CTH aspects in the CBH eval comparisons, even though those are mentioned in the description part. Probably the evaluation results from Seaman et al. 2017 or Noh et al. 2017 cited in this paper (e.g., line 407-) were conducted under the "within-spec" condition when CTH is a 2-km error range compared against ground truth data, which aimed to isolate CBH eval, decreasing CTH effects as the physics-based algorithms are highly dependent on CTH accuracy.

No sufficient discussions on multilayers: Multilayer cases (either limitations or future plans) as well as nighttime cases should be addressed in the conclusion.
Detailed comments and questions for further clarification are below.

Answer: Thanks for your nice advice. We believe your suggestion is absolutely right; indeed, some parts of the text were overly verbose, making it hard for readers to grasp the key points. We have already made some revisions, such as cutting down unnecessary wording and rearranging sections, like moving Section 4.1 to Section 3.4, removing some descriptions in section 3.3.

Appendix A section has discussed and validated the accuracies of cloud products from H8 satellite. We have provided a detailed introduction in the second paragraph of Chapter 2 (Data) and made a thorough comparison in Appendix A. In fact, several early papers have already conducted detailed comparisons. In Section 4.1, we once again cited the paper by Min et al., 2020, to describe the accuracy of CTH used in this study "*According to previous CALIPSO validations (Min et al., 2020), the absolute bias of cloud top height retrieved by the H8 satellite is approximately 3 km, with an absolute bias of 1 to 2 km for samples below 5 km. The accuracy of CTH is crucial for estimating CBH in the subsequent algorithm.*".

Moreover, since it is not easy to exactly distinguish between multilayer and single-layer clouds solely based on geostationary satellites passive sensor, neither the early physical algorithms from NOAA and NASA nor the newly developed ML-based methods are effective at making this distinction (Baker, 2011; Noh et al., 2017). It assumes the single layer for cloud base heigh retrieval of all samples. The algorithm will retrieve primarily the cloud base of the uppermost layer cloud if lower clouds are

not well detected and the column-integrated cloud water path (CWP) retrieval is made for the topmost layer, which shows that the accuracy and representativeness of the upstream CWP retrieval is essential to the CBH estimate together with the CTH accuracy. It should be noted that multilayer cloud scenes are typically more challenging for CTH and CBH retrievals, and CLAVR-x CTH also often tends to be biased low for the multilayer situation (Noh et al., 2017). Therefore, we did not classify the clouds into multilayer and single-layer categories in this study. However, we have added some sentence in our manuscript to underscore this issue at the beginning of Section 3.1 "*It is important to note that multilayer cloud scenes remain a challenge for retrieving both CTH and CBH, especially when considering the column-integrated cloud water path (CWP) used in physics-based algorithms (Noh et al. 2017). In this study, we will simplify the scenario by assuming a single-layer cloud for all algorithms.*" .

Reference

Min Min, Jun Li*, Fu Wang, Zijing Liu, W. Paul Menzel, 2020. Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms [J]. Remote Sensing of Environment, 239: 111616, doi: 10.1016/j.rse.2019.111616

Baker, N.: Joint Polar Satellite System (JPSS) VIIRS Cloud Base Height Algorithm Theoretical Basis Document (ATBD), 2011.

Noh, Y.-J., Forsythe, J. M., Miller, S. D., Seaman, C. J., Li, Y., Heidinger, A. K., Lindsey, D. T., Rogers, M. A., and Partain, P. T.: Cloud-base height estimation from VIIRS. Part II: A statistical algorithm based on A-Train satellite data, Journal of Atmospheric and Oceanic Technology, 34, 585–598, 10.1175/JTECH-D-16-0110.1, 2017.

Comments/questions:

Line 50 "sensor may be attributed to utilizing the same dataset …" : Not clear.

Answer: Thanks for your nice advice. We have added "…of CloudSat/CALIOP… " to make it more clear at this Line.

Line 130 “These methods aforementioned are prominent in retrieving CBH … space-based remote” : Not clear, which methods you are referring to?

Answer: Thanks for your nice advice. We have revised it as “*These passive space-based remote sensing methods aforementioned, such as satellite imagery, play a key role in retrieving CBH.*” to make it more clear at this Line.

Line 132 …”The first method …”: It seems like a starting sentence is missing. Clarification needed. What these first and second methods about?

Answer: Thanks for your nice advice. We have revised it as “*In terms of detection principles, the first method…*” to make it more clear at this Line.

Line 161: “… algorithm, achieving a high correlation coefficient (R) of 0.92 and a low root mean square error (RMSE): add compared against which data?

Answer: Thanks for your nice advice. We have revised it as “*...compared with CloudSat/CALISPO data.*” to make it more clear at Line 164.

Line 167-168: It should be partially true, but Tana et al.'s study that the authors cited right above has used Himawari-8 data, which doesn't support this argument. It would be better to replace "mainly" with partially or something similar in line 166 and also better to rewrite

Answer: Thanks for your nice advice. We have revised it as “…. achieving a similar high..” at line 162.

line 164-168 to address the diurnal cycles haven't been well investigated in both GEO and LEO remote sensing research.

Answer: Thanks for your nice advice. Agree, we have added a sentence of “*The diurnal cycles of CBH have not been well investigated in both GEO and LEO remote sensing research.*” at lines 169-170.

Line 217: For MODIS data, please add a couple of additional explanations why the authors describe MODIS here with such details (including the MODIS product evaluation in Appendix), in order to help readers' understanding, even though the

reason appeared later but not clear yet here in this general Data section. Otherwise, it may distract the main topic of this paper.

Answer: Thanks for your nice advice. We have added a sentence of "*High-quality, long-term series MODIS data is often used as a validation reference to evaluate the products of new satellites.*" to make it more clear at lines 226-227.

Line 272-273: No need to address this here, out of the main focus of this study which adopts only the algorithm for the H8 application, not JPSS VIIRS, anyway.

Answer: Thanks for your nice advice. Agree, we have deleted this sentence.

Line 287: This "reliable" looks already quite deterministic. "Another" should be enough here.

Answer: Thanks for your nice advice. Agree, we have revised it.

Line 291-292: "… studies have also demonstrated a R of 0.569 and a RMSE of 2.3 km for the JPSS.. " -> This is also unnecessary here (no info how the error statistics were obtains won't be helpful to readers, too), slightly out of this study's scope.

Answer: Thanks for your nice advice. I still hope to keep this sentence, as it will help readers compare the accuracy of our work with previous studies.

Line 311-317: It would be good if this is a thesis, but somewhat too much extra information for the paper. It looks already enough by citing Breiman and Min et al. Tan et al.'s papers.

Answer: Thanks for your nice advice. We have deleted some descriptions about this part.

Line 328: "based on"-> using additional information from NWP model data or similar sentences may be considered. Need to rewrite. It may give the impression that the algorithms rely solely on NWP data.

Answer: Thanks for your nice advice. Agree, we have revised many "based on.." in our manuscript.

Line 336-337: but employs different view zenith angles and azimuth angles. -> Not clear. Need more clarification.

Answer: Thanks for your nice advice. IR-single model only uses view zenith and azimuth angles (no solar zenith and azimuth angles). Hence, we have revised it as "... *but employs only view zenith angles and azimuth angles.*" at line 334.

Line 338: "matching method" what about parallax corrections between two sensors? It seems some technical details are missing here.

Answer: Thanks for your nice advice. We think there is no parallax correction between two sensors. It only means that the nearest distance matching method is used to match two different product in space and time domains. We have revised this sentence at line 384 "... *ensuring that collocating the closest points and the observation time difference between the CloudSat/CALIPSO observation point and the matched Himwari-8 data is less than 5 minutes.*" to further explain this "matching method" here.

Line 407-412: Did the error statistics consider a similar factor for CTH eval with Seaman et al.'s "within spec" comparisons or just under all cloud conditions?

Answer: Thanks for your question. We use the similar factors for CBH evaluation as Seaman et al., 2017. We have described this at the end of section 4.1.1.

Line 421: "the CBHs lower than 2 km for"-> Is there any possibility of inversion in the low boundary layers as GFS NWP data may not have such high vertical resolution to resolve and thus CTH errors causing CBH errors in physics algorithms and also NWP input impact on ML algorithms maybe.

Answer: Thanks for your nice advice. The reason you mentioned is possible, but based on the inversion results of the cloud top height, it doesn't seem very likely. The main reason is probably that the lower cloud base signals are difficult to detect from the satellite using infrared or visible light channels.

Line 477-478: Not necessary to describe all the general lidar observation theory. The paragraph can be trimmed.

Answer: Thanks for your nice advice. We have deleted some sentences in this paragraph.

Line 493: As the authors addressed, ground lidar observations tend to be quickly attenuated near lowest cloud base especially for thick clouds. If using solely lidar data in comparisons, cloud characteristics (type, depth, etc) related weather conditions would be good to be discussed as well.

Answer: Thanks for your nice advice. We have added a sentence of "*These two days have been cloudy, with stratiform clouds at an altitude of around 5 km and no precipitation occurring.*" to make it more clear.

Line 527: Line 538 and below details about the radar data should be placed here.

Answer: Thanks for your nice advice. Agree, we have moved this paragraph forward as your suggestion.

Fig. 1: Additionally, it will be good to mention these comparisons for all cloud conditions including single and multilayer cloud scenes, and something like the CTH accuracy (or evaluation) is not considered.

Answer: Thanks for your nice advice. Agree, we have added a sentence "*...RF IR-single algorithm for all cloud conditions including single and multilayer cloud scenes.*" at Line 391 to further explain this issue. We have compared CTH in our supplementary documentation. Also we have explained the accuracy of CTH in Section 4.1 mentioned before.

Figure 2 caption : CBHs -> lowest CBHs, 2017 -> the statistics for all cloud scenes including both single and multilayers? If so, please specify it.

Answer: Thanks for your nice advice. Yes, the statistics are for all cloud scenes including both single and multilayer clouds. As we mentioned earlier, passive remote sensing satellites have difficulty distinguishing between single-layer and multi-layer clouds, especially at night. We have added an interpretation in this caption.

Fig. 3: What happened to the "No CloudSat obs" part on the right end in (b)?

Answer: Thanks for your nice advice. We believe there are clouds, just not clouds with very strong Radar echoes.

Line 534-538: It's not well organized, which seems like a jump in the context. It should be placed at the beginning of this ground-observation eval section.

Answer: Thanks for your nice advice. Agree, we have moved this paragraph forward to make it more clear.

Line 648-650: Specify what exactly the factor is.

Answer: Thanks for your nice advice. We have added a sentence "*Therefore, this factor induced by detection principle could contribute…*" at line 644 to make it more clear.

Line 655: Too early conclusive remark with limited comparisons and without intensive case analyses.

Answer: Thanks for your nice advice. We have changed this sentence as "*Ideally, we guess that including more spaceborne cloud profiling radars with varying passing times (covering the entire day) in the training dataset could improve the machine learning technique, potentially leading to a higher-quality CBH product with more comprehensive observations.*".

Line 565: "it is more reasonable to opt for physics-based cloud base height algorithms. " It seems like a too early conclusive remark with limited comparisons and without intensive case analyses.

Answer: Thanks for your nice advice. Agree, we have deleted this sentence.

Line 670: As well as "nighttime cases" , multilayer cases should be mentioned in the conclusion.

Answer: Thanks for your nice advice. We have added a sentence "*… CBHs from GEO H8/AHI data under the assumption of single layer cloud.*" at the second paragraph to explain this issue.

Minors:

Line 35: remove 'one'

Answer: Thanks for your nice advice. We have removed it.

Line 123: remove "As well known"

Answer: Thanks for your nice advice. We have removed it.

Line 134: References in Line 227 should be put here, too, which is the first place for MODIS.

Answer: Thanks for your nice advice. We have added the references in line 136.

Line 162: the random forest -> RF . The acronym was already defined. Found the same errors in several places.

Answer: Thanks for your nice advice. We have checked again and revised the corresponding words in the entire text.

Line 204: The validation "is"-> has been

Answer: Thanks for your nice advice. We have revised it.

Line 335: remove "Global Forecast System" which acronym has been already defined.

Answer: Thanks for your nice advice. We have removed it.

Line 375: Not good to use two denotes COT and D_COT in one paper. Please use one consistently.

Answer: Thanks for your suggestion. We replaced COT with D_{COT} have used one denote D_{COT} consistently.