Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-409-RC2, 2019

Atmospheric Measurement Techniques Discussions
Open Access

Interactive comment on "A Machine Learning-Based Cloud Detection and Thermodynamic Phase Classification Algorithm using Passive Spectral Observations" *by* C. Wang et al.

Anonymous Referee #2

Received and published: 18 December 2019

This paper applies a machine learning (ML) approach to the problem of cloud detection and thermodynamic phase assignment from passive satellite measurements. This is potentially significant considering the challenges noted in the manuscript with the traditional methods currently being employed and the rapidly increasing interest in using ML for satellite analyses of clouds. The ML approach evaluates a number of models that are tested and evaluated using various combinations of passive sensor radiances and ancillary data products as inputs while CALIOP data are used to define the reference labels for cloud occurrence and phase. Two models are selected for evaluation,

one that employs solar and infrared radiances (daytime) and one that employs only infrared radiances (all day). The view angle, latitude, longitude and the surface skin temperature were found to be the most important ancillary data needed. In addition, the models are trained for 7 surface types. The two models are found to perform reasonably well and performance metrics generally exceed the current approaches employed on MODIS and VIIRS by the MODIS Science Team. However, the significance of the results are difficult to gauge for a variety of reasons. For example, the ML and current (referred to as traditional in the manuscript) approaches are designed much differently with regards to the targeted clouds, atmospheric correction, scene type dependencies and other factors. With respect to the clouds, the ML model development excludes the most difficult clouds which are pervasive over the Earth. In particular, clouds in polluted environments, broken clouds and single-layer and multi-layer ice overlapping water clouds are screened out of the training and validation dataset. The rationale for taking this approach is not well described. Part of the evaluation of the ML method against current methods with respect to CALIPSO (figs 6-9) could perhaps be considered an apples to apples comparison in that the same pixels are being evaluated. But, considering that the ML approach was developed using a particular subset of (screened) data while the current approaches were designed for application over a much wider range of conditions is possibly unfair, and the comparison are potentially misleading. I wish the authors had taken a more globally applicable ML approach to the problem. It seems to me that at best the results suggest that ML methods can perhaps perform at least as well as the current non-ML methods and that these can be developed for application to other satellites much easier (and cheaper). Despite all of these issues, the study is a reasonable initial step, the results are clearly presented and the manuscript is grammatically clean. Therefore, I find that the manuscript could be published after some revision. In particular, I recommend that the authors clarify the rationale for the approach, clarify the significance of the results, and temper suggestions regarding the potential for ML to improve the accuracies of global cloud analyses since in my view this is not adequately demonstrated here given the heavily restricted dataset that is used. Additional recommendations to improve the manuscript follow:

Line 23: Strongly suggest something like this: "It is shown using a conservative screening process that excludes the most challenging cloudy pixels for passive remote sensing, . . .

Line 35: 'will' need further attention

Line 62: Zhou reference may need updating

Line 79-80: This statement is too vague and possibly misleading. How is the uncertainty assessment more difficult for a cloud classification derived with the traditional methods vs the ML approach? It is true that in a Bayesian context, uncertainties in satellite retrievals associated with inversion are easy to extract, but these do not include uncertainties w.r.t ground truth data due to simplifying assumptions in the forward models and a host of other factors. Please elaborate to clarify and support your contention.

Line 195: should be Sayer et al 2017?

Line 221-223: not clear what you mean here

Line 231-234. Not sure what the relevance of this update is to the paper unless you used the older version. If this is the case, then you'll need to elaborate on the impact of the deficient version 1.0 algorithm on this study.

Line 249. Not sure what GOES-16/17 have to do with anything. Suggest 'which is now applied to VIIRS.'

Line 301-311: This is an important section with no rationalization for the decisions made to create the training/validation datasets. You should explain why each of these decisions were made and justified

Line 316: define complicated

Line 327: describe how the tuning and optimization were achieved

Line 334: It would be useful to elaborate on possible reasons for the importance of geolocation as an input and the lack of importance for Ts. Why use Ts instead of Tclr computed at TOA? Wouldn't the latter be more consistent with the traditional approaches?

Line 346: Not clear what you mean by similar tests. Consider elaborating further.

Line 348: change to 'IR bands used in the all-day model'

Line 353: Consider tabulatin the daytime results similar to table 2. I think this would be useful.

Line 378 and further: Figs 6-9 are fine but it would help the reader better understand the comparisons if these data could also be tabulated (unless of course you don't think that they are significant enough to further illuminate)

Line 387-389: Is this any surprise considering that you have eliminated the most difficult clouds?

Lines 406-412: the results in figures 8 and 9 are not very clear or well described. In a relative sense, which algorithms are overdetecting or underdetecting ice and water clouds and why?

Line 450: change to something like this "The above results indicate that for the screened data considered here, the two RF models have better and more consistent performance over different regions and surface types in comparison with the MODIS and VIIRS products suggesting the potential to improve the overall performance in more global operational applications

Line 457: It is good to drive home the point regarding the ease and cost savings of applying ML vs the traditional approaches which took years to develop. 'a few hours' seems vague tho. Consider elaborating further.

Line 459-462: Do they really use similar input? The channel complements are different, so if this in any way affects the phase determination, then what you are saying could be unfair and misleading since the two methods were not designed for continuity.

Line 465: In this section it should be emphasized again that a screened dataset is used to train and test the ML methods that excludes the more difficult pixels for passive sensor methods. While the ML methods appear to offer some advantages, the higher accuracies found here compared to the traditional approaches may not be representative of those found when applied to a more inclusive dataset

Lines 474-478: This is also vague and won't make much sense to most readers. What is the objective for your passive determination? Consider elaborating further on the definition and applications for cloud phase (cloud top or radiative), and the relative sensitivities of passive vs active. Maybe then it would be more clear what you mean when you say that a multi-layer clouds category could help

Lines 489-490. The screening process almost certainly impacts the comparisons with the traditional methods which were not developed with a similar screening process. Please make sure that you address this somewhere in the manuscript.

Line 518: why is this more impractical? It actually seems necessary

Line 534: using the collocated CALIOP products in 2017 and excluding the more difficult pixels associated with polluted, broken and mixed-phase cloud conditions

Line 553: should read " . . .phase detections in a limited set of conditions

Line 555: consider changing 'a few hours' to 'considerably more efficiently' ??

Line 562 and 563: change 'can' to 'could'

Line 564: Suggest adding this at the end: It remains as future work to determine how such an approach might lead to improved consistency in cloud properties derived from different satellite remote sensors

Line 607: reformat with last name first or change reference on line 150

Line 651: reformat with last name first or change reference on line 121

Line 829: Why is MODIS CLDPROP not shown in figure 12?

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-409, 2019.