Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2019-409-RC1, 2019

Atmospheric Measurement Techniques Discussions

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 #1

Received and published: 15 December 2019

The authors describe a machine learning (ML) based approach to first detect clouds and second to assign cloud thermodynamic phase (liquid versus ice). The ML algorithm is trained using CALIOP detected liquid and ice clouds but is limited to the most straightforward single phase and single layer cloud configurations (or multilayer with the same phase), thus mixed phase and multi-layered clouds of different phases are not included in this study. The approach is tested against existing MODIS Collection 6 (C6) and MODIS/VIIRS continuity products (both the cloud mask and cloud phase). The ML approach is shown to improve the phase characterization over the existing MODIS

Printer-friendly version

Discussion paper

and MODIS/VIIRS continuity algorithms, with greater improvements over certain surface types including snow and ice. Cloud characterization efforts from satellite remote sensing platforms are increasingly utilizing ML algorithms and this paper is timely and a useful exploration of the potential of ML for passive cloud imagery characterization. Parts of the methodology are not as well detailed as need be and the results need to be placed into a broader context. After addressing the comments below and suggestions for straightforward revisions, this paper would be a nice addition to the literature.

Abstract: I found it to be a bit too detailed and meandering. Would suggest tightening it up and focusing on the main points rather than the details.

Line 59: 'having radiometric stability issues' is colloquial and not specific enough to be useful

Lines 79-80: There are two issues here that need to be raised and appear elsewhere. First issue, is this even true? There are many Bayesian methods in the literature that assign uncertainties as a part of the retrieval methodology. Furthermore, using the look up table methodology of MODIS C6, the reported uncertainties for the optical properties appear to be quite useful and rooted in physics. I don't know about uncertainties regarding phase so this could be a different issue. For the cloud mask, the raw Q values are quite useful for an estimate of cloud detection uncertainty. Second issue, calling one set of algorithms 'traditional' is confusing at best. Machine Learning (ML) research dates back to the 1950s and outdates many satellite retrieval algorithm approaches that are currently used. Wording along the lines of "in contrast to most operational and research methods," and similar changes elsewhere, will help make your points clearer. Then you could stick to "ML" as a separate algorithm branch.

Lines 138-139: Now the random forest (RF) model is mentioned and apparently it has a proven record, yet is "not traditional"? The "author classification" of algorithms needs to be reworked throughout the paper.

Lines 233-234: Need to be more explicit as to what "the fix" was to the thermodynamic phase algorithm.

Section 4.2: This is a long paragraph with a lot of information and should be make clearer than currently written. First recommendation: it would be very helpful to list the pixel count and relative percentages of each of the cloud/no cloud, aerosol/no aerosol, phase and cloud configuration categories that are kept for ML training or are discarded, and should be denoted clearly. It took me a while to figure out that some multilayer clouds are included but only for the same phases. How many multilayer clouds of the same phase occur relative to multiple phases? Second recommendation: say more clearly up front what is in the ML training rather than what is tossed out. Then follow with detail of what is tossed out. It is really hard to keep track of what goes into the sausage. First additional comment: why not attempt to address the ambiguous/mixed phase categories? Some advances in detection and characterization could be made with these types using ML. Do you have plans to do this in follow-on work? Second additional comment: why only use clouds with at least five consecutive labels that are the same? Doesn't this limit the number of cases greatly? Also doesn't this bias the ML training to larger-scale cloud behavior even though the classification is (presumably) done on a pixel-by-pixel basis? Small scale clouds might behave differently (with respect to phase sensitivity) than large scale clouds. Third additional comment: why are cloud structures in the ITCZ any more "complicated" than other geographical regions? What makes a cloud structure "complex"?

Lines 346-353: Is this a description of other experiments tried that are not shown in figures or tables? Or is this paragraph part of the methodology?

Lines 401-405: It would be really helpful to report what total percentage of all pixels considered these represent. The crux of the matter: does ML greatly help for a large percentage of cloudy pixels, or does it help for a small percentage of cloudy pixels? Also, in figures 6-9 showing the true versus false positive rates, it would greatly enhance the presentation of the results by including percentages for each subpanel of the total number of pixels considered.

Lines 418-423: There is a disconnect between this discussion and the earlier discussion on lines 308-310. How are inhomogeneous clouds being considered when earlier the authors state that they are “discarded”? These may be different issues but it is worth making clearer how inhomogeneous clouds are (or are not) considered and dealt with in this study.

Lines 456-457: “A few hours” doesn’t really mean anything scientifically. And without describing what is calculated and on what kind of computing platform, this also doesn’t convey any information.

Lines 457-459: While not written directly in this way, reading between the lines written by the authors, one could deduce that ML approaches could render instrument calibration efforts and algorithm continuity efforts pointless and irrelevant. Will ML have the potential to address discontinuous satellite observational records by a thorough and accurate labeling of training data for a ML algorithm? I don’t think this is what you intended to say, but it does raise the point – can ML methods be used in lieu of a properly calibrated and characterized satellite instrument? Same point applies to lines 467-468.

Lines 470-478: Regarding the use of CALIOP for labeling, one could make the argument that CALIOP is a distinctly different observation and should in fact see something different than a VIS/SWIR observation (e.g., MODIS and VIIRS). Doesn’t CALIOP labeling essentially “force” MODIS and VIIRS to observe like a lidar even though they do not contain the same physical sensitivity to clouds as the lidar? Will differences in instrument sensitivity (e.g., CALIOP vs. VIIRS) to a given cloud ultimately lead to poorer performing ML algorithms because one is made to “look like” the other? It is an interesting question to consider. For some clouds, the lidar and passive spectrometer could provide a lot of valuable complementary information, and that is basically “thrown out” in a ML algorithm when one is forced to behave like the other.

Lines 489-490: not sure what is meant by “screening process”

Lines 518-519: why is it more impractical to consider aerosol and cloud together?

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