Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-334-RC2, 2017

Atmospheric Measurement Techniques Discussions

Interactive comment on "The Community Cloud retrieval for Climate (CC4CL). Part I: A framework applied to multiple satellite imaging sensors" *by* Oliver Sus et al.

Anonymous Referee #1

Received and published: 31 December 2017

This is a mature manuscript that describes a new framework for retrieving cloud properties from passive imaging sensors, which are partially validated by / trained with active techniques (CALIPSO). The main thrust of the paper is to ensure that one single technique be applicable to a range of sensors (mainly AVHRR/MODIS). The stated goal is to maximize the length of the time series available from the collection of the various sensors in low earth orbit (the approach is only applied to polar orbiters, not to geostationary satellites). For this reason, only "heritage" channels are used, which has the clear disadvantage that newer developments (such as the 1.38 micron channel for the detection of thin ice clouds or the 2.1/2.25 micron channel for better phase discrimination in MODIS/VIIRS) cannot be taken advantage of. The authors acknowledge that

Printer-friendly version

Discussion paper

they made a conscious decision to do so, but this also begs the question about the distinct benefit of this particular manuscript over already existing Climate Data Records such as that based on PATMOS-x. The authors do discuss prior efforts extensively, but don't really answer the question why we need yet another multi-satellite retrieval framework, except to say that ESA solicited the creation of Essential Climate Variables within their Climate Change Initiative. There are a few novel aspects in this approach, which are not always sufficiently explained (details see below, sequential comments). Beyond those, however, the manuscript does not conceptually take us beyond ISCCP and more modern climatologies. A truly novel retrieval would move beyond a single-pixel approach and consider the context and geographic region for increasing the information content in cloud retrievals. It would also not apply traditional optimal estimation without careful consideration of non-linearities (this has be done with Bayesian approaches such as Markov Chain Monte Carlo sampling - see manuscripts by Posselt). The authors did replace the cloud mask with a Neural Network approach tied to CALIPSO, and this seems meritorious because it seeks to objectivize passive imagery based retrievals by using active techniques as independent data source. Other than this important innovation, it remains unclear whether CC4CL truly is an improvement over existing techniques (such as PATMOS-X), or simply re-creates such efforts with slight modifications. Despite this concern, the collocation and cross-comparison of multiple instruments is convincing, and the fact that the code is presented as an open-source development makes the work very compelling. The next manuscript version should include information where the code and documentation can be downloaded.

The manuscript is rather heterogeneous in terms of the language quality. Details are given along with the sequential comments below. The most important (major) comment is that about the use and interpretation of ANN for the cloud mask. Generally, more specificity will be required in multiple aspects. Most of them will be possible to implement through minor revisions, but it would be good to have the manuscript go through another brief review after implementing them.

Sequential comments (minor and major comments are mixed):

p1: “Climate data record” is not discussed. It is also not clear until page 2 that the manuscript does indeed seek to develop a Essential Climate Variable (although it is unclear without reading Hollmann 2013 which of the 13 ECVs CC4CL will contribute to). A discussion of “essential climate variable”, “CDR”, and how this work fits in should be better discussed. It should also be discussed how it distinguishes itself from existing efforts in this regard (e.g., https://www.ncdc.noaa.gov/news/new-cloud-properties-climate-data-record).

p1l38: “shielding” is not a radiative transfer term - what is it in this context? Isn’t it the same as “forcing”? Why are both terms used?

p1l38: “forcing”: There is a difference between “radiative forcing” and “radiative effect” - which are the authors referring to? Probably the latter.

p1l49: This sounds like the variables “propagate uncertainties” into the derived cloud properties, which would be incorrect.

p2,l14: While “auxiliary” instead of “ancillary” data have become almost interchangeable, the latter is more correct; “auxiliary” has the connotation of only being a replacement in case the “primary data” is not available (compare: auxiliary power, not ancillary). For satellite retrievals, ancillary expresses more accurately that data from other sources are ingested within the operational algorithm.

p2,l25: “. . .not guaranteed to be radiatively consistent with. . .” It is unclear what that means (although the reviewer agrees with the statement). Please provide references. Also, does CC4CL perform “better” in terms of radiative consistency?

p2,l39: “sees” into the cloud: A retrieval is not animate. Replace colloquial “see” with more appropriate wording.

p2,l50: CONUS = contiguous US (conterminous is synonymous, but used much less frequently, also not by Sun et al., 2015).

p2,l55-l57: This is an important statement: Cloud cover is not a good observable for trend detection because it depends on its definition (optical thickness threshold and/or reflectance threshold, sensor resolution) and instrument performance or calibration drifts. Even the CALIPSO-derived cloud information depends on which resolution is considered (because of sensitivity and SNR). A better observable would be the optical thickness itself (or better still, the cloud radiative effect). Have the authors considered a different primary variable that is more amenable to trend detection than cloud cover? In fact, their approach of retrieving "pseudo CALIPSO optical thickness" seems to be going exactly in this direction - and in the reviewer's opinion, this would be the right way to proceed. But why then go a step backwards and convert ANNCOD into a binary cloud mask? Why isn't the retrieved ANNCOD not reported directly (in addition to the binary cloud mask outcome)?

Related to the above [and also to material on p6]: Since CC4CL does keep cloud cover as primary variable, it should be explained whether the thresholds (table 2) vary (for example, with the specific sensor or orbit), or whether they are fixed once and for all, now that they have been optimized via the ANN technique. More importantly, do the weights as established during the ANN learning process vary? Are they a function of orbit, instrument, illumination, surface, topography...? Or else, are all of these dependencies incorporated in one single ANN? If so, how are commonly known problems with ANN (such as overfitting) avoided here? Using this cloud masking and thresholding technique, what is the (minimum) cutoff optical thickness, below which cloud are no longer detected? How do optical thickness detection thesholds vary with surface type and sun-sensor geometry?

Related to the above [and also to material on p6]: The three elements of the ANN need to be described better. How well is the pseudo-CALIOP optical depth itself estimated with the ANN? Figure 2 illustrates the performance of the cloud mask after ANNCOD has be converted into a binary cloud mask. Since the ANN predicts ANNCOD and not the cloud mask itself, it should be the performance of the ANN with respect to

ANNCOD that should be demonstrated here. In this context again: How is overfitting avoided? How can the non-linearities of radiative transfer be emulated with a single hidden layer? What is the result for ANNCOD for the training data set as opposed to the test data set? How is the correction for viewing angle done? How many inputs does the input layer have; what are they? What is the activation function? Are there bias perceptrons? What motivates the use of one single hidden layer, and why are there 50 neurons in it? Is the network re-trained for every new satellite data set, or are the weights fixed? How exactly were the threshold values from table 2 determined that are applied to ANNCOD to translate into cloud mask?

Finally, what is the quality of the thermodynamic phase retrieval, optical thickness and effective radius, depending on how close ANNCOD is to the cloud detection threshold? Essentially, the paper claims that a cloud retrieval is attempted if the optical thickness exceeds 0.4 over snow/ice during day light conditions. This would be a remarkable improvement over existing retrievals. MODIS usually does not detect clouds over snow-covered areas in the Arctic unless they have an optical thickness significantly larger than 0.4 (around 7). CC4CL would be an improvement of an order of magnitude, and the question is whether the cloud retrievals would be of practical use, especially when applying them to AVHRR instead of MODIS. The reviewer strongly believes that the only way to achieve detection thresholds on the order of 0.4 in optical thickness in snow/ice covered regions in the Arctic, one would need to use convolutional layers (i.e., use multi-pixel retrieval approaches).

p2,l78-80: “Consistency can be traded for continuity” needs clarification. Perhaps this can be done while elaborating on CDR (see comment above). This discussion will contribute to a better motivation of this study.

p2,l90: “MODIS provides”: is a partial repetition of material in the left column of the same page.

p3,l18: “on other” > “over other”

p3,l20: Which “macrophysical” product do the authors have in mind here? What exactly does “radiative inconsistent” mean (supposedly, macroscopical products are inconsistent with microphysical products, but this is different from “radiatively inconsistent”; the reader is currently left to guess here). How exactly does the CC4CL approach ensure radiative consistency amongst all input satellite radiances (and all output products)? Indeed, other approaches have a cloud mask that may be independently derived from the microphysics products. Simply stating that CC4CL is “different” in this regard does not support the statement that it is more “consistent”. More details are needed to add specificity.

p3,l46: Quantify “very realistic”, or just use “realistic”

p4,l35: Auxiliary > Ancillary

p4,l38: Neural Network not yet defined at this point. May need the NN section prior to this statement.

p4,l73/l75: “optimal estimation”, “cloud typing scheme”. None of these have been described at this point in the manuscript. Sequence needs to be re-shuffled.

p5,l1: “were” > “are”

p5,l1: Reference and/or data source (link) needed for CALIPSO product

p5,l68-l72: multiple acronyms need to be introduced prior to first use.

p5,l79: The outcomes of the study should be at least summarized here. Also, the use of “round robin” may not be ideal for an international readership as it is a cultural reference (British/American) that may not be commonly known. Consider paraphrasing the technique instead.

p5,l98: Do these channel numbers refer to the CC4CL IDs from table 1?

p6: Cloud detection: See multiple comments above (following p2,l55-l57 comment) Also: Are there any convolutional layers included in the approach? This would have

allowed capitalizing on the context of a pixel.

p7,table 3: How was the regression done - based on radiance or irradiance, based on counts? Based on brightness temperature (for IR channels)? The offsets seem rather large; what is the explanation for significant offsets?

p7,l49: VIIRS algorithm is used: What is the purpose of this statement? If it is kept, this needs to be elaborated (what does the VIIRS algorithm do differently). Also, there are various other algorithms that are improved over the heritage algorithms, which would probably all need to be mentioned here (or at least a subset thereof).

Figure 2: This is just one example where labels are too small, and are too pixelated. Generally improve the figure quality and enlarge labels. About the content: It is rather hard to interpret this figure. The x-axis is "normalized". Does that mean that the difference of the ANNCOD-retrieved value and the threshold from table 2 is divided by the threshold value itself? Does "x=0" mean that the retrieved optical thickness equals the threshold per table 2? Does the "CLEAR" label refer to CALIPSO? For x=-0.2, we find an uncertainty of 40%. Does that mean that CC4CL misclassifies clear pixels as "cloudy" in 40% of cases?

p8,l70: Why are largest uncertainties found for opaque clouds? Also, figure 10 does not show quantitative evidence for this statement - colors are harder to interpret than numbers on a graph. Can this somewhat counterintuitive statement be supported by a more succinct graph?

p9,l3-5: :Validation is show for . . . rather than: Unclear. What is the difference between CTH and "its" retrieved value?

p9,l13: "TOA radiation is the *sum total* of emission and scattering throughout the atmospheric column" - please formulate this more accurately: What is a "sum total" of two processes? Also, the next paragraph more or less paraphrases Platnick's vertical weighting function paper where this is formulated more accurately, and where the

concept of a weighting function is well explained. Please cite that paper and use similar terminology here. As for multi-layer clouds, there is a fairly new paper by Wind, Platnick et al. (http://journals.ametsoc.org/doi/abs/10.1175/2010JAMC2364.1), but it is probably not applicable to this paper here because of the channel selection.

p10,l1: How is the CTH adjustment done if the cloud base is not known? Where does cloud base (or cloud geometrical thickness) information come from?

p10,l9: Does this statement about sectors refer to figure 9? Please match figures and text, otherwise figures become "orphans" that are not tied to the manuscript.

p10,l14: Please define what is meant by "surface" in this case.

p10,l16: insert "a" before "single-layer"

Figure 7: please enlarge labels, as well as histograms; it is hard to compare the retrievals quantitatively otherwise. Also: It would really help if histograms were shown separately for snow-covered areas as opposed to dark surfaces. It is expected that retrieval quality would differ significantly depending on the surface conditions.

p12,l41: "performance of existing algorithms" What are the "existing algorithms" that CC4CL? Has the manuscript shown that these existing algorithms perform less well than CC4CL.

p12,l88: "AVHRR" > "for AVHRR"

p12,l89: Should "continually" be replaced with "consistently"? Unclear what this statement means. If it were "consistently" it would be more clear, but the word order should be fixed: "The CC4CL phase identification does not agree with any of the three CALIOP cloud flags consistently, which is reasonable given . . ."

p13,l19/20: ". . .insensitive to the specific instrument evaluated, such that the merged data set is sensible". What does this statement mean? The paper does not actually present a *merged* data set, or was that the actual intent of the paper? It does evaluate

collocated overpasses from different satellites, but these are not merged in the sense of a CDR. Please remove the statement about “merging” data sets unless this was the actual intent of the paper (in which case it would need to be modified considerably).

p13,l31: “disagree nonetheless”: They disagree despite their channels are fairly close? Can this be re-phrased? The whole paragraph is a bit roundabout. There’s a 30-40% difference in reflectance, but “their” retrieval values are “much more similar”? Please make this statement more precise. “The difference to AVHRR and MODIS is largest for CER” - does this statement refer to AATSR again?

p13,l39: The t-test needs to be explained in much more detail. What is H0, what is mu1, what is mu2? Are we talking about the covariance between two data sets, which is assessed using the t-test approach? If so, are the data from the two different data sets (supposedly this is what “mu1” and “mu2” refer to) re-gridded to one common grid before comparing them? The premise of this statement deserves at least one paragraph, if not half a page.

p13,l45: “spatiotemporally collocated sensors”: The sensors are not collocated - is that the point of the statement? Or is this an explanation why the t-test “fails? What does “non-significant” t-test mean? Could the strictness of the comparison be relaxed by gridding the retrievals to a coarser common grid before making the inter-comparison?

p13,l57: “depending on the user’s application” - this needs to be clarified. For which applications can they be used interchangeably? Could a combined AVHRR and MODIS cloud data record constitute a CDR (would it meet the requirements)? As stated above, the manuscript does not actually “merge” data sets in this way, but more specificity would be helpful here.

p13,l77: “we see that COT uncertainty scales with COT itself”: this is not shown in the manuscript. If it is, please refer to a figure or section.

p13,l79-l88: Consider re-writing this section; simplify and use literature references;

most of these observations have been documented before (large COT uncertainty as reflectance approaches asymptotic value; large uncertainties for bright surfaces).

p15, l11: “otherwise are” > “otherwise they are”

p15,l15: “may it stem” does not work in English; consider “whether it stems from. . . or”

p15,figure 11: The table below the cross section is too small. Also, what happened at lat=61? Why do the active imagers pick up a cloud where CALIPSO does not?

p16,l8: consider “a conscious decision was made to [deliberately] trade. . .”

p16,l19: “on a first view” > “at first glance”

p18,l59: “synergic” > “synergistic”

p18,l95: “accurate and precise”: These two were not discussed separately. Where was this done? If not, please clarify this statement.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-334, 2017.

Printer-friendly version

Discussion paper