Anonymous referee #2

RC: referee comment

AR: author response

AC: author's changes in manuscript

RC: p 2, line 5: I would add cloud forward model assumptions to the list of secondary confounding factors

AR: Will add „cloud forward model assumptions“ to list.

AC: „Several secondary variables (cloud forward model assumptions, state of surface and atmosphere, viewing geometry, sensor calibration and spectral response uncertainties) …“

RC: p 2, lines 12-14: The CERES-MODIS products (e.g., Minnis et al., 2011a,b, IEEE TGRS) should also be included here.

AR: Will add the CERES-MODIS products.

AC: „and MODIS Collection 6 (MODIS C6) (Platnick et al., 2017) as well as the CERES-MODIS products (Minnis et al., 2011).“

P. Minnis *et al.*, "CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part I: Algorithms," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4374-4400, Nov. 2011.

RC: p 2, lines 23-24: The MODIS C6 phase referred to here is the IR phase of Baum et al. (2012), which is in fact a quad-spectral algorithm (7.3, 8.5, 11, 12µm channels) using β ratios (the authors' description is more appropriate for the C5 algorithm). This IR phase algorithm is run in conjunction with, and is informed by, the cloud top property retrieval algorithm. The authors should be aware, and I believe that they are given the reference to Marchant et al. (2016) later in the paper, that this IR algorithm does not determine phase for the C6 cloud optical properties retrieval; phase for the optical retrieval is determined by the Marchant algorithm that uses the IR phase as one piece of information. Results from the IR and cloud optical properties phase algorithms are often at odds, specifically in cases where phase is more ambiguous.

AR: Will correct the text to reflect that the Marchant algorithm is applied.

AC: ", or a majority vote algorithm that combines four phase tests based on CTT, tri-spectral IR, 1.38 µm, and spectral CER data (Marchant et al., 2016)."

RC: p 2, line 24: Should probably specify that the additional spectral channels are at shortwave infrared (SWIR) wavelengths.

AR: Will add SWIR here.

AC: ",… MODIS has several additional spectral channels at shortwave infrared (SWIR) wavelengths that provide…"

RC: p 2, lines 29-30: Indeed, this is an inherent limitation of the spectral information content of passive IR channels!

AR: Will rephrase.

AC: "these studies show that current retrievals underestimate cloud top pressure for optically thin clouds due to the inherent limitation of the spectral information content of passive IR channels."

RC: p 2, lines 31-35: I assume from the references given that cloud cover refers to cloud fraction or related metrics, and not to geophysical retrievals.

AR: Yes, we are referring to cloud fraction. Will replace cloud cover with cloud fraction.

AC: "There are numerous studies that evaluate the performance of the aforementioned retrievals for cloud fraction with weather (…). More importantly, these studies emphasize the difficulty of deriving reliable cloud fraction trends from AVHRR time series, as the retrievals overestimate the change in cloud fraction by as much as an order of magnitude"

RC: p 3, line 5: Is the cloud phase bias positive or negative?

AR: The cited bias values were reported as absolute numbers.

AC: "and has an absolute cloud phase bias of lower than + 9 %"

RC: p 3, lines 6-7: See my p 2 comment above regarding MODIS phase algorithms; this statement again refers only to the IR phase.

AR: Will rephrase to refer to Marchant et al., 2016.

AC: "and the phase detection has been improved for liquid clouds. However, the detection of optically thin ice clouds over warm, bright surfaces remains problematic (Marchant et al., 2016)."

RC: p 3, lines 10-11: What is the difference between consistency and continuity? I can surmise that it is consistency in approach versus continuity of results, but it is not clear to the general reader.

AR: The reviewer's assumption is correct, will clarify.

AC: "Consistency in approach can be traded for continuity of results, and multi-platform algorithms could exploit additional data when newer sensors become available"

RC: p 3, lines 34-35: It's not initially clear why independent retrievals of COT/CER and macrophysical products are inherently radiatively inconsistent. I would guess that it depends on the approach, i.e., how (or if) one set of retrievals informs the retrieval of the other. Can the authors better explain?

AR: The effect of COT/CER/CTH on the top-of-atmosphere (TOA) radiances differs between the different sensing bands as a function of atmospheric state. For example if you used just the 11 or 12 micron measurement to estimate CTH then you must assume something about the COT (usually that it is thick) and something about the CER (typically a climatological value). If the COT assumption is incorrect (e.g. cloud is not thick) so that more upward radiance is transmitted through the cloud than expected, then the cloud top appears too warm and is located (incorrectly) lower in the atmosphere. On the other hand using an all channel fit, as we did here, will identify the cloud as optically thin (from the visible and near visible reflectance measurements) and will avoid this error.

We note that retrieving a specific cloud property from a specific channel is radiatively inconsistent (as example above) but it is generally possible to do a sequential optimal estimation retrieval. In this case one iterates through the channels improving the estimates of CTH/CER/COT with each step. The final result should be the same as an all channel optimal retrieval. This method is not adopted for our problem as it would be computationally less efficient.

AC: “but macrophysical products are estimated independently and are thus radiatively inconsistent with the former variables. Here, parameters are retrieved simultaneously, providing a retrieval that is radiatively consistent over the wavelengths of the observations, given that the instrument’s noise characteristics are well known.”

RC: p 4, line 1: Retrieval uncertainty estimates that propagate errors is not a novel feature of CC4CL. See, for instance, the MODIS C6 cloud optical properties (Platnick et al., 2017), which provide pixel-level retrieval uncertainties calculated in a manner that is mathematically consistent with that of optimal estimation (although the uncertainties are not part of the solution process).

AR: Agreed, will clarify.

AC: “Another key feature of CC4CL is the production of uncertainty estimates of retrieval parameters (see also Platnick et al., 2017) through explicit error propagation from input to output data.”

RC: p 4, line 6: Following on my comment above, neither the optimal estimation approach nor the uncertainty quantification are novel features of CC4CL. As the authors themselves state on p 2, PATMOS-x uses optimal estimation theory, and the MODIS C6 (and C5) cloud optical properties provide rigorous pixel-level uncertainties.

AR: Agreed, will remove novel here.

AC: “We particularly focus on discussing the key features of the framework: the optimal estimation approach in general, …”

RC: p 4, lines 5-13: Regarding statements about consistency of the long-term, multiplatform time series, and the potential of the framework for climate studies, I don’t think the authors make a convincing case for either in the text that follows. Four case studies hardly constitute a “comprehensive and detailed analysis of retrieval results,” and certainly do not provide enough evidence of the potential for climate studies. Such statements require detailed analyses of long-term and large-scale inter-sensor statistical comparisons, which it appears are actually presented in a companion paper in a different journal (Stengel et al., 2017). It’s thus not clear to me why the present paper was not instead a part of the Stengel paper, or vice versa. Given that the primary contributions are a brief discussion of the ancillary and data sources and a rather limited CTH analysis, I'm not convinced that this paper can or should stand on its own.

AR: This paper's main purpose is to present a new cloud retrieval framework (CC4CL). It is a two part publication that contains a detailed description of the retrieval algorithm in part II. Part I should not be seen as a validation paper, but rather contains a section that provides the reader with an overview of the functionality of CC4CL, including generic strengths and weaknesses. The goal is to inform the reader of potential applications of this data in future research. The four case studies aim to illustrate the strengths and weaknesses of CC4CL through detailed, direct (i.e. with very little averaging), and collocated comparisons with independent CALIOP data. The Stengel paper, as the reviewer correctly mentions, contains a true validation of CC4CL, but to include such an in-depth analysis here would have substantially increased the paper's length. We think that keeping part I concise and focused better serves its purpose as an introduction to the functionality and generic applicability of CC4CL. For readers who might be interested in a validation of CC4CL after reading part I, we refer to the Stengel paper in the text.

However, we will replace "validated" with "examined", as the former indeed suggests more than the paper intends to provide, and remove "comprehensive".

AC: p 4, line 10: "These are initially examined in a detailed analysis of …"

RC: p 4, line 15: Consider using Level-1 instead of L1, which for some readers implies a Lagrange point 1 orbit.

AR: Will clarify here that L1 stands for Level-1. L1 is standard terminology in this field.

AC: "Level-1 (L1) satellite data"

RC: p 4, lines 21-25: Yes, replacing any AVHRR once its successor becomes available will lessen the impacts of orbital drift (and thus sampling times), but drift impacts are likely still to exist. Are these accounted for in CC4CL, specifically when constructing long-term multi-sensor time series?

AR: Orbital drift effects are not accounted for within CC4CL, which is why we write to only reduce drift-induced changes, not to eliminate them.

RC: p 4, line 29: Regarding filtering channel 3b data, is this to include or exclude that channel?

AR: The filter removes noise artefacts from channel 3b data, which are used in the retrieval.

RC: p 5, lines 8-10: It should be NASA Goddard Space Flight Center.

AR: Will change text.

AC: "the NASA Goddard Space Flight Center performed"

RC: p 5, lines 21-23: "Self-calibrating" is I think a little misleading. MODIS, for instance, has a similar design (onboard black bodies and solar diffuser), yet requires a continual effort to monitor instrument stability and identify/correct calibration drifts, typically using fixed ground targets among others.

AR: Will clarify.

AC: “ATSR is equipped with on-board calibration capabilities, such as two black-body targets for the thermal channels and a sun-illuminated opal target for the visible/near-infrared channels.”

RC: p 7, lines 3-4: Has the “gap filling” of the MCD43C1 data been validated? Is the approach similar to what is used in the MCD43B3 gap-filled product (Schaaf et al., 2011, “Aqua and Terra MODIS albedo and reflectance anisotropy products,” in Land Remote Sensing and Global Environmental Change: NASA’s Earth Observing System and the Science of ASTER and MODIS)?

AR: We did not validate the “gap filling”, for which we applied a very basic approach to meet our requirements. The approach applied to gap-fill MCD43B3 data is certainly more sophisticated, but its application in our study was out of scope.

RC: p 6-7, Sections 2.2.3-2.2.4: Have the authors verified that there are not any trends in the land surface BRDF and emissivity time series during the MODIS era? If there are, wouldn’t the use of the climatology derived from all MODIS data introduce a discontinuity in the surface time series?

AR: We did not perform a trend analysis for these time series. We agree that a trend in the input data would indeed add an artefact to our retrieval output.

AC: p 7 l 14: “Note that the use of a climatology would add a discontinuity in the surface time series if there are trends in the surface BRDF and emissivity time series during the MODIS era.”

RC: p 7, lines 6-7: I disagree that the surface is a minor component of the observed signal, specifically for optically thinner clouds. Thus not accounting for the spectral response functions can introduce biases, particularly in spectral regions such as the near-IR (e.g., AVHRR channel 2, MODIS channel 2) where reflectance by vegetation can change rapidly.

AR: Agreed, will clarify.

AC: “in spectral response functions. Note that this might result in retrieval biases, particularly in spectral regions that are sensitive to rapidly changing environmental processes such as vegetation growth (near-IR).”

RC: p 7, line 16: Resampled or aggregated?

AR: Resampling is defined as the technique of manipulating a digital image and transforming it into another form. Thus the term is applicable here. As is aggregated.

RC: p 7, line 16-17: I would agree that differences in sensor spatial resolution are reduced when averaging radiances/reflectances. However, this is likely not the case when averaging L2 geophysical parameters, as is done here, since the retrievals can have significantly different PDFs within a grid box due to pixel size differences alone.

AR: Agreed, will clarify.

AC: “This resampling is required for an intercomparison of CC4CL Level-2 data on a common grid. However, note that differences in sensor spatial resolution can lead to significantly different PDFs within a grid box, the effect of which we did not analyse.”

RC: p 9, line 5: How much data was used to train the ANN? Was an observation time difference filter applied to the NOAA-18/CALIOP co-location?

AR: See p 9, line 10-11. Yes, the time difference filter was 15 minutes.

RC: p 9, lines 19-21: If I understand correctly, the reflectances/radiances were adjusted to account for spectral response differences? Were the co-located observations filtered for cases in which both satellites viewed the scene at the same sun-view angle geometry? Such angle matching is important when comparing solar channels where reflectance is strongly angularly dependent.

AR: Yes, we did account for sun-view angle geometry differences. We filtered all collocations with differences in satellite zenith angle > 0.5°, sun zenith angle > 1°, and observation time > 30 mins.

AC: "We derived appropriate coefficients through linear regression analysis between collocated satellite observations for each input channel pair (Table 03), applying a filter on differences in satellite zenith angle (> 0.5°), sun zenith angle (> 1°), and observation time (> 30 mins)."

RC: p 10, lines 21-22: My understanding is that the uncertainty obtained from the optimal estimation framework can be thought of as the sensitivity of the solution space at the point of the solution to the measurement uncertainty (which includes instrument, ancillary, etc., uncertainties).

AR: That is correct. The statement will be revised.

AC: "The algorithm estimates the retrieval uncertainty, which quantifies the range of values that are feasible considering the uncertainty in the satellite measurements, auxiliary data and ORAC forward model."

RC: p 10, line 25: This statement differs from the statement at the end of Section 3.2 (phase is determined first to reduce computation time resulting from retrieving assuming both phases).

AR: That was the original processing setup, but in the end we decided to process both phases for all pixels. That was required in order to swap retrieval output if phase needed to be switched due to mismatches with CTT. Will clarify.

AC: p 8, line 25-27: "The main processor evaluates these inputs twice, assuming different cloud phases (e.g. ice and liquid). In theory, ORAC could use the preprocessed cloud mask and phase to select an appropriate method to reduce processing time."

RC: p 11, Section 4.1, Figure 3-5. The observation date/times should be stated here. I see they are listed in Section 4.3, but it is better to include them at first reference. Also, a thermodynamic phase image would be useful.

AR: Will add observation date/times here. Will also add the thermodynamic phase image.

AC: "The sample scene (07/22/2008 20:58 LST) is characterized by various cloud types, and the CC4CL cloud mask defines a relatively small fraction as cloud free (Figures 3 to 6). "

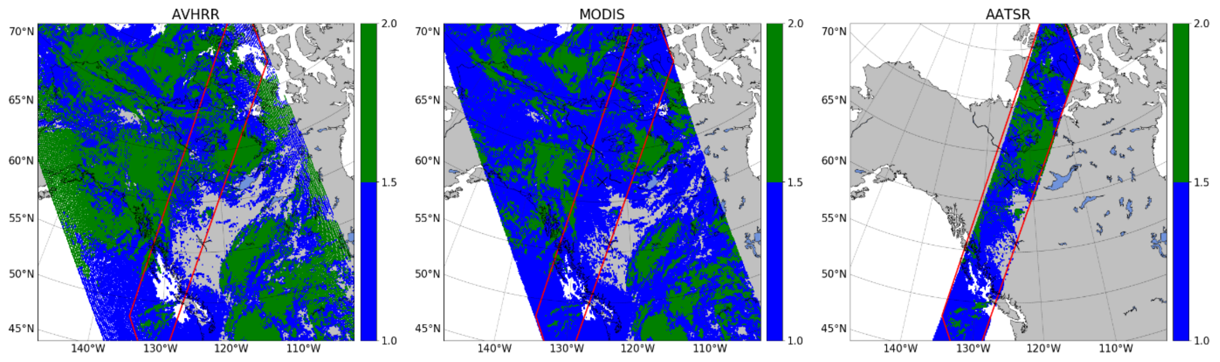

Figure 6. Cloud phase retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).

RC: p 11, lines 13-14: I'm guessing the peaks at 12 and 35 µm likely correspond to liquid and ice phase clouds, respectively.

AR: Agreed.

AC: "CER data are somewhat bimodal, having a primary peak at ~12 µm and a secondary peak at ~35 µm (Figure 07 and Table 06). These peaks probably correspond to liquid and ice phase clouds, respectively."

RC: p 11, line 17: The statement on cloud displacement here contradicts the statement in line 11.

AR: Although observation time difference is small, and thus cloud displacement, it cannot be discarded to contribute to the significance test, in particular to outliers.

AC: "Significance tests of mean differences and standard deviations of residuals between sensor retrievals are sensitive to outliers. Although cloud displacement due to observation time differences is probably small, we cannot discard its influence on such outliers."

RC: p 11, Section 4.1: What about relative radiometric calibration between the different sensors? Even minor differences of a couple percent could case large retrieval differences, particularly for COT.

AR: Agreed, will add a statement at the end of the paragraph.

AC: p 11, line 20: "Moreover, even modest relative radiometric calibration differences between sensors of a couple percent could cause large retrieval differences, particularly for COT."

RC: p 11, line 22: If median absolute CER uncertainty is 2µm, how does this correspond to a median relative uncertainty of 2% (line 24). Figure 10: What wavelengths are used for this RGB?

AR: The reviewer is correct, these statistics are wrong. Will correct. For the RGB, we used red = Ch4 solar component, green = Ch2, blue = Ch1.

AC: "Median absolute uncertainties are CTP = 26.7 hPa, COT = 6.1, CER = 2.0 µm, and cloud mask = 13.7 % (Figure 06). The median relative retrieval uncertainty (not shown) is relatively low for CTP and CER, but considerably larger for COT (CTP = 4.7 %, COT = 55.0 %, CER = 13.6 %). COT uncertainties increase with COT magnitude, and the RGB image (Figure 010, red = Ch4 solar component, green = Ch2, blue = Ch1) shows that the largest uncertainties are found in cases of opaque cloud coverage and cloud over sea-ice surfaces.”

RC: p 12, Section 4.3: Hard to call this “validation” without using a much larger dataset (e.g., months, seasons, years) for statistical analyses.

AR: Agreed, will rephrase.

AC: “Comparison with CALIOP”

RC: p 12, line 21: What assumptions are made other than adiabaticity (e.g., extinction profile, etc.)? Also, what does adiabaticity mean for an ice phase cloud?

AR: We assume that the cloud is vertically homogeneous with a constant lapse rate.

RC: p 12, Case Study NA1: Need to include the Figure number in the text.

AR: Agreed.

AC: “Study area NA1 is a completely cloud-covered scene over northern Canada containing clear and ice-covered land and open ocean surfaces (Figures 08 and 09).”

RC: p 13, lines 25-26: Which existing algorithms were compared to these results?

AR: Will remove the subordinate clause.

AC: “In general, the quantitative and qualitative agreement between CC4CL and CALIOP CTH is impressive.”

RC: p 14, lines 10-11: Why not show the extensive validation here?

AR: Will add a reference to the Stengel paper mentioned above.

AC: “The results shown here are a representative sample from an extensive validation performed within the Cloud_cci project (Stengel et al., 2017).”

RC: p 15, lines 6-11: For the optimal estimation retrieval, are the spectral response differences handled similar to the ANN cloud mask (i.e., adjustment factors), or are they explicitly included in the forward model? What about relative radiometric calibration, could that be playing a role in the large MODIS-AATSR retrieval differences?

AR: Spectral response difference are taken into account when producing LUTs applied within CC4CL and are thus included in the forward model.

RC: p 15, line 18: Here calibration deficiencies are acknowledged. Relative calibration should be explored as a cause of the retrieval differences.

AR: Although we acknowledge that there are calibration differences, and doubt that sensors give precisely the same results, they were found to be consistent over vicarious calibration sites. For example, a 3 % offset between AATSR and MODIS has been found for visible channels (Smith and Cox, 2013), and a bias of < 0.3 K between MODIS and AVHRR longwave infrared channels (Cao and Heidinger, 2002). We think that this difference is not large enough to account for all the retrieval differences we see here. Note that the LUTs do take spectral differences into account, with the limitation that they have been calculated for an average value and not the full spectral shape, so that non-linear effects remain.

AC: "We did not quantify the contribution of each of these processes to overall retrieval differences when using different sensor data. In particular it would be worth investigating the impact of spectral response differences, which was outside the scope of this paper and the ESA Cloud_cci project."

D. L. Smith and C. V. Cox, "(A)ATSR Solar Channel On-Orbit Radiometric Calibration," in IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 3, pp. 1370-1382, March 2013. doi: 10.1109/TGRS.2012.2230333

Changyong Cao, Andrew K. Heidinger, "Inter-comparison of the longwave infrared channels of MODIS and AVHRR/NOAA-16 using simultaneous nadir observations at orbit intersections", Proc. SPIE 4814, Earth Observing Systems VII, (24 September 2002); doi: 10.1117/12.451690; https://doi.org/10.1117/12.451690

RC: p 15, lines 29-30: Can the authors provide references for these user applications?

AR: Will add references.

AC: "On the one hand, they are useful for several user applications, such as model validation, data assimilation applications, or climate studies in general (Liu et al., 2017, Yang et al., 2016)."

Liu, C., R. P. Allan, M. Mayer, P. Hyder, N. G. Loeb, C. D. Roberts, M. Valdivieso, J. M. Edwards, and P.-L. Vidale (2017), Evaluation of satellite and reanalysis-based global net surface energy flux and uncertainty estimates, J. Geophys. Res. Atmos., 122, 6250–6272, doi:10.1002/2017JD026616.

Yang, Qinghua, et al. "Brief communication: The challenge and benefit of using sea ice concentration satellite data products with uncertainty estimates in summer sea ice data assimilation." The Cryosphere, vol. 10, no. 2, 2016, p. 761.

RC: p 16, line 29: "radiatively effective rather than physical cloud top"

AR: Will correct.

AC: "Any CTH retrieved from AVHRR (heritage) data is the radiatively effective rather than physical cloud top …"

RC: p 17, line 9: The MODIS C6 phase referred to here is that of the cloud optical properties algorithm, not the IR phase referred to earlier in the paper.

AR: The reviewer is correct. Our modifications above already account for that.

RC: p 18, lines 10-12: Perhaps this is worded poorly? I would imagine that real, complex vertical cloud structure is in fact a large source of retrieval errors, but the analytical approach to retrieval uncertainty used here (and in other retrievals) cannot account for this

AR: We think that, in the case of optically thick, i.e. opaque, clouds, the vertical cloud structure is not a major driver of TOA radiances and thus retrieval uncertainty. TOA radiances are mainly constrained by the cloud top layer, and also by lower layers until their influence becomes negligible due to vertical extinction.

AC: "Retrieval uncertainty is estimated using only well-understood error sources (e.g. measurement and forward model error), neglecting errors due to model assumptions (e.g. the complex, real vertical structure). Such errors can be approximated through validation activities and are not currently believed to be significant in most circumstances."