Anonymous referee #1

RC: referee comment

AR: author response

AC: author's changes in manuscript

RC: 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-propertiesclimate-data-record).
AR: We will move p2 lines 97-106 to the very beginning of the introduction, and add some text to elaborate how our study fits into ESA CCI and the production of ECVs. We will now also mention CDR, but we do not elaborate on it, as it is not an essential part of this paper. We thus refer to Stengel et al., 2017, where this concept and resulting data have been published.
AC: "The European Space Agency has established the ESA Climate Change Initiative program (ESA CCI, 2015; Hollmann et al., 2013) in order to advance knowledge of the climate system through the generation of satellite based data records utilizing European and non-European assets. The CCI project's primary focus is the production of thirteen Essential Climate Variables (ECVs) covering ocean, atmosphere, and land geophysical variables. With these data records, CCI is aiming to fulfil highest climate requirements from the Global Climate Observing System (GCOS). This study presented here is part of the ESA CCI for clouds (ESA Cloud_cci), which has the objective to develop a state-of-the-art open-source community cloud retrieval algorithm which shall be capable of processing passive satellite imager data covering several decades. Both in part I and part II of this paper, we present the processing framework as developed within ESA Cloud_cci (CC4CL, part I), the detailed mechanisms of the optimal estimation retrieval (part II), and provide an initial assessment of the strengths and weaknesses of derived cloud parameters (part I). With CC4CL, several decades of passive imaging satellite data have been processed and are made available to the user. The resulting climate data records (CDR) are presented in Stengel et al., 2017."

AR: We will amend p2, line 106ff.
AC: p2, line 106ff: "In order to produce the cloud CDR presented here, we used satellite data from MODIS…"

RC: p1, l38: "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?
AR: We agree that shielding is superfluous, will remove.
AC: "Clouds considerably influence the global energy budget through direct radiative effects."

RC: p1, l38: "forcing": There is a difference between "radiative forcing" and "radiative effect"- which are the authors referring to? Probably the latter.
AR: Agreed, we are referring to radiative effect.
AC: See comment above.

RC: p1, l49: This sounds like the variables "propagate uncertainties" into the derived cloud properties, which would be incorrect.
AR: Will clarify.
AC: "Several secondary variables (state of surface and atmosphere, viewing geometry, sensor calibration and spectral response uncertainties) further complicate cloud retrievals, and insufficient knowledge on their state propagates uncertainties into the derived cloud properties."

RC: 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.
AR: Agreed, will replace auxiliary with ancillary throughout the text.

RC: 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?
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 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 theoretically 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.

See also the introduction in part II of this paper for a detailed definition a radiative consistency. We will not elaborate much on the issue here, as that already happened in part II.
AC: "However, the derived microphysical variables are not guaranteed to be radiatively consistent with independently derived cloud parameters, as most of the retrieval methods are separated into solar and thermal methods even though measurements in these spectral regions are not independent of parameters retrieved in the other."

RC: p2,l39: "sees" into the cloud: A retrieval is not animate. Replace colloquial "see" with more appropriate wording.
AR: Will rephrase.
AC: "and beyond a penetration depth into the cloud corresponding to > 1 cumulative optical depth."

RC: p2,l50: CONUS = contiguous US (conterminous is synonymous, but used much less frequently, also not by Sun et al., 2015).
AR: Will rephrase.
AC: "and contiguous United States"

RC: 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)?
AR: Here, we are referring to cloud cover as one of several other variables that were analysed for other retrieval frameworks in separate studies. In this study, we are only presenting a retrieval framework and an initial assessment of its data quality. We do not present trend data, but rather refer to other studies that assessed quality of other CDRs, including cloud cover but also CTH and CTP. The ANNCOD is a temporary retrieval product, from which we derive cloud cover. Cloud cover information is used to avoid processing cloud-free pixels and thus to reduce processing time. ANNCOD data are contained in ESA Cloud_cci L3U products (see Stengel et al., 2017).
AC:

RC: 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?
AR: The thresholds in table 2 have been quantified through iterative optimisation rather than by the ANN technique (p 6, l 68). They are fixed for all sensors and orbits, and thus, as is shown in Table 2, only vary as a function of illumination and surface condition. There are no sensor specific thresholds, but we apply a simple viewing angle correction on the input satellite data. The ANN weights themselves have been trained with NOAA18 data, and we linearly adjusted input data for other satellites to better match NOAA18. The text already states which ancillary data have been used when training (p 6, l 38 - 42), including surface conditions, and also that several ANNs were produced (p 6, l 32 - 34). We are using 3 different ANNs (day, twilight, night), which reduced the overfitting problem mentioned by the reviewer. Also, overfitting was minimised by comparison with an independent test dataset while training.
We did not quantify a cutoff optical thickness as asked by the reviewer. Instead, our approach involved quantifying those threshold values for which the fit between CC4CL and CALIPSO cloud cover is best.
AC: p6, l 65 ff: “The thresholds themselves vary depending on illumination and surface conditions, namely land, sea, and snow/ice cover (Table 2), and were quantified through iterative optimization. They are fixed for all sensors and orbits.”

RC: 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?
AR: It was never the intention of creating a COD retrieval that can also be used to extract a cloud mask. We aim at creating a binary cloud mask. For a COD retrieval, we would have needed to train the full range of CALIPSO COD (approx. 0-15), but we cut off at a COD of 1 (and set all COD > 1 = 1). We assumed that CALIPSO COD values > 1 are clouds that will always be correctly detected by passive sensors. Considering that, we do not think that a comparison of ANNCOD with CALIPSO COD makes sense and thus should not be included here.
How can the non-linearities of radiative transfer be emulated with a single hidden layer?

AR: We agree that the use of at least one more layer could have improved the retrieval. However, in the CC4CL framework we used an IDL based library who does not provide more than 1 hidden layer.
What is the result for ANNCOD for the training data set as opposed to the test data set?
AR: The training dataset is only a small part of the collocation dataset. When training, the dataset was divided by 90/10 percent into a training dataset and a test dataset, i.e. we trained on 90 percent and tested simultaneously on the rest of the data (10 percent). So, the test dataset has only been tested while training. To avoid the overtraining, the training has been stopped when both RMSE (train/test) started to differ.
How is the correction for viewing angle done?
AR: We found that only a part of the whole viewing angle geometry was trained (0-35° out of up to 70° for AVHRR). We created an averaged ANNCOD with respect to each viewing angle and found a cosine-shaped dependency, which we corrected with an empirical cosine function.
$ANNCOD_{corrected}$ = ANNCOD - (1. / 12. * (1. / cos(satellite_zenith_angle * degree_to_radians) - 1.))
How many inputs does the input layer have; what are they?
AR: This depends on the illumination (day/twilight/night) and the availability of channels. We will add the input variables to the text.
AC: p6 l 37: “For the input layer, input variables are surface temperature, snow/ice cover, and the land/sea mask for all three cloud masks. Regarding sensor data, input channels are Ch1, Ch2, Ch5, Ch6, and Ch5-Ch6 for the day ANN, Ch4, Ch5, Ch6, Ch5-Ch4, and Ch5-Ch6 for the night ANN, and Ch5, Ch6, and Ch5-Ch6 for the twilight ANN.”
What is the activation function? Are there bias perceptrons?
AR: Our activation function is the sigmoid function. We did include bias perceptrons.
What motivates the use of one single hidden layer, and why are there 50 neurons in it?
AR: Regarding the one hidden layer, see comment above. Will add text an number of neurons.
AC: p6 l37: “Through incremental testing, we found that 50 neurons was the value for which the trade-off between output quality and computing speed was optimal.”
Is the network re-trained for every new satellite data set, or are the weights fixed?
AR: The weights are fixed, see p6 l 68 – p7 l 6. We tried to overcome the problem of having different shapes of the spectral response functions by applying linear regression coefficients (see Table 3). In a later version of the cloud mask, we applied a more sophisticated approach through using multispectral observations of IASI and SCIAMACHI. In the version presented here however, the linear regression is based on a one month triple collocation between AVHRR NOAA18, AATSR-ENVISAT, and MODIS AQUA.
How exactly were the threshold values from table 2 determined that are applied to ANNCOD to translate into cloud mask?
AR: We determined these thresholds through incremental application of a skill score analysis of the ANN cloud mask with CALIPSO for the whole collocation dataset (as a reminder, the training dataset is only a small part of the collocation data set). See previous comment, including a text change at p6, l 65 ff.
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?
AR: We did not quantify this relationship in detail. However, Figure 6 shows retrieval uncertainties of CTP, COT, and CER together with cloud mask uncertainty. The patterns do not appear to be clearly related. A quantitative analysis, e.g. calculating correlations between relative uncertainties, would certainly provide more detailed answers, but was out of this paper’s scope.
AC: p18 l30: “It would also be worth investigating the relationship between the quality of retrieved variables (CTH, COT, CER, cloud phase) and cloud mask uncertainty.”

RC: 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).
AR: Please note that the ANNCOD is a *pseudo* optical depth. The threshold value, here 0.4 for daytime over snow/ice, is a relative value between 0 and 1. It does not provide any information on the absolute optical thickness value, but is rather a normalized optical thickness that attempts to fit CALIPSO measurements.

RC: 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.
AR: Agreed, will clarify. See also reviewer #2 comments on the same issue.
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: p2,l90: “MODIS provides”: is a partial repetition of material in the left column of the same page.
AR: Agreed, will delete that sentence, as it is redundant.

RC: p3,l18: “on other” > “over other”
AR: Will rephrase.
AC: “improvement over other established”

RC: 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.
AR: We are referring to macrophysical products such as CTT and CTH. Please see also our related answer on radiative inconsistency above, and the detailed description of the issue in part II of this paper.

RC: p3,l46: Quantify “very realistic”, or just use “realistic”
AR: Will rephrase.
AC: “provides realistic estimates”

RC: p4,l35: Auxiliary > Ancillary
AR: Agreed.
AC: “Ancillary”

RC: p4,l38: Neural Network not yet defined at this point. May need the NN section prior to this statement.
AR: Will add a reference to the ANN section here. Most readers probably have at least a vague idea what a neural network is.
AC: “and as input to a neural network cloud mask (see Section 3.1.1)”

RC: 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.
AR: Will add section references here. Again, these references should be sufficient, as the readers will have heard these terms before and do not require a detailed definition here to understand the following text.

AC: “The USGS data are used as a land sea mask within the optimal estimation retrieval (Section 3.3.3), as well as a land cover classificator within the cloud mask and the Pavolonis cloud typing scheme (Section 3.3.2).”

RC: p5,l1: “were” > “are”
AR: Cannot find “were” in that sentence.

RC: p5,l1: Reference and/or data source (link) needed for CALIPSO product
AR: Cannot find the CALIPSO product in that sentence. Maybe there is a linenumber mismatch?

RC: p5,l68-l72: multiple acronyms need to be introduced prior to first use.
AR: Most acronyms in that sentence were introduced in the first paragraph on page 2.
AC: “…Clouds from AVHRR Extended (CLAVR-X) (…) Global Cloud and Aerosol Dataset Produced by the Global Retrieval of ATSR Cloud Parameters and Evaluation (GRAPE) …”

RC: 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.
AR: Will rephrase.
AC: “was chosen from three competing algorithms in a “round-robin” (i.e. each algorithm is tested against all other algorithms) analysis. All algorithms have proven their maturity for deriving the considered cloud parameters (cloud cover, liquid and ice water path, cloud top height) from AVHRR and MODIS data (Stengel et al., 2015).”

RC: p5,l98: Do these channel numbers refer to the CC4CL IDs from table 1?
AR: Yes, will clarify.
AC: “The albedo of snow/ice covered pixels is set to globally constant values of 0.958 (Ch1, CC4CL ID as in Table 1), 0.868 (Ch2), 0.0364 (Ch3), and 0.0 (Ch4),”

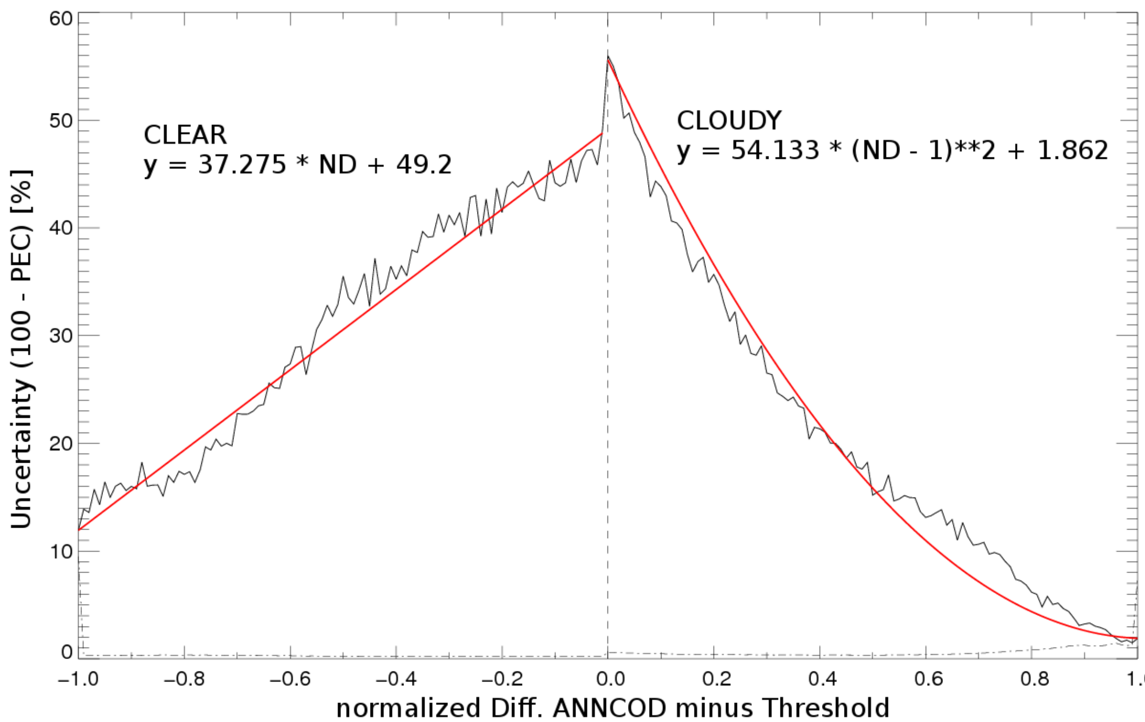
RC: p6: Cloud detection: See multiple comments above (following p2,l55-l57 comment). Also: Are there any convolutional layers included in the approach? This would have allowed capitalizing on the context of a pixel.
AR: We did not add any convolutional layers in the ANN. See above comments regarding cloud detection and the ANN.

RC: 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?
AR: The regression coefficients were calculated based on reflectance and brightness temperature data. The offsets might be a result of imperfect collocation, relative calibration differences, and mainly differences in spectral response functions. It is difficult to quantify the contribution of each, but spectral response probably explains most of the offset.

RC: 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).
AR: This paragraph is a brief summary of the Pavalonis algorithm performance. Not surprisingly, it performs better if more spectral channels are used for cloud typing. The sentence emphasizes the generic limitation of using only AVHRR heritage channels, which does not only affect cloud detection or optimal estimation, but also cloud typing. We do not think that other algorithms using other data than the heritage algorithms need to be elaborated or mentioned here (however, we do so elsewhere). It is obvious that they perform better if using more channels, but that is not the point here.

RC: 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?
AR: As mentioned in the text, the x-axis is normalized, i.e. the difference between ANNCOD and the threshold was divided by the threshold. Yes, x=0 means no difference between ANNCOD and the threshold. Again, please remember that this is a *pseudo* optical thickness. CLEAR means that the ANN cloud mask defined a pixel as cloud free. It shows that we need different equations to quantify uncertainty for clear and cloudy cases. The text also explains how uncertainty is calculated: 100 – PEC [%], with PEC = the ratio between all correctly classified pixels and the number of all pixels analysed. Also, if x=-0.2, CC4CL misclassifies cloudy pixel as 'clear' in 40% of the cases with respect to CALIPSO. The uncertainty defines the misclassification of CC4CL compared to CALIPSO for a certain combination of ANNCOD and the threshold used.

See below Figure 2 with larger labels and annotations. We also increased labels for Figures 3-7 (see responses to other reviewers).



RC: 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?
AR: We are referencing the wrong figures. Will correct.
AC: “COT uncertainties increase with COT magnitude, and largest uncertainties are found in cases of opaque cloud coverage (Figure 4 middle and Figure 6 topright).”

RC: p9,l3-5: :Validation is show for . . . rather than: Unclear. What is the difference between CTH and “its” retrieved value?
AR: CTH is a derived variable, i.e. derived from CTP, which is the retrieved value. Will clarify.

AC: “The validation is shown for comparisons of CTH (derived from CTP) rather than CTP (retrieved) to enable…”

RC: 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.
AR: Will clarify.
AC: “However, TOA radiation is the product of emission and scattering processes throughout the atmospheric column (Platnick, 2000).”

Platnick, S. (2000), Vertical photon transport in cloud remote sensing problems, J. Geophys. Res., 105(D18), 22919–22935, doi:10.1029/2000JD900333.

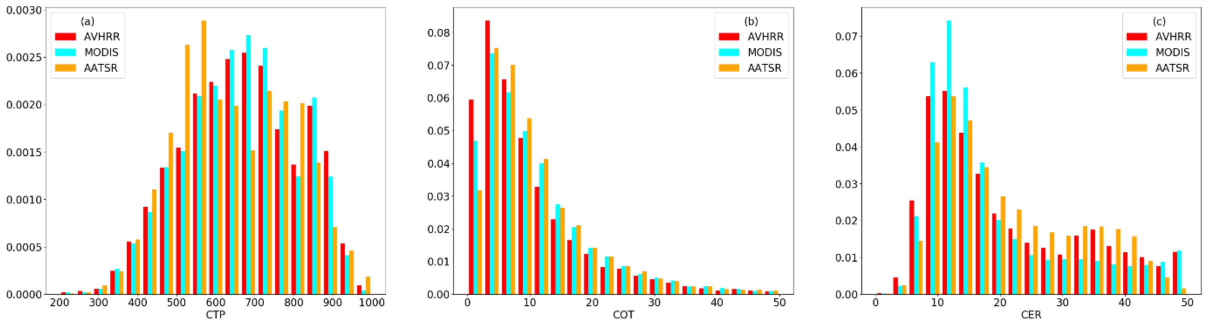
RC: p10,l1: How is the CTH adjustment done if the cloud base is not known? Where does cloud base (or cloud geometrical thickness) information come from?
AR: We approximate the observed temperature as emitted from one optical depth into the cloud. Assuming the cloud is vertically homogeneous with a constant lapse rate Γ, we can write the thickness-corrected CTT as,
$T_{cor} = BT(\lambda) + \Gamma / (\sigma N)$,
where BT is the observed brightness temperature, σ is the cloud particle cross-section, and N is the cloud particle number concentration. Using the observations at 11µm and 12µm provides two simultaneous equations in T_{cor} which can be solved, using σ values for a LUT.

RC: 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.
AR: Will add figure references.
AC: “CC4CL correctly classifies all pixels as cloud covered, with a few exceptions in sectors 3 and 4 (Figures 8 and 9).”

RC: p10,l14: Please define what is meant by “surface” in this case.
AR: Will clarify.
AC: “In the case of a (semi-)transparent cloud top layer, multiple surfaces (several cloud layers, Earth surface) contribute to the observed satellite data.”

RC: p10,l16: insert “a” before “single-layer”
AR: Agreed.
AC: “For a single layer, optically thick (COT > 1) cloud,…”

RC: 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.
AR: We will enlarge labels and histograms. However, we do think it is sufficient to show histograms for all surfaces combined to make our point that there are differences between retrievals, which is also supported by the statistics.



RC: 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.
AR: Will remove the subordinate clause.
AC: "In general, the quantitative and qualitative agreement between CC4CL and CALIOP CTH is impressive."

RC: p12,l88: "AVHRR" > "for AVHRR"
AR: Agreed.
AC: "and AATSR data than for AVHRR"

RC: 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 . . ."
AR: Agreed, we will rephrase as suggested.
AC: "The CC4CL phase identification does not agree with any of the three CALIOP cloud flags consistently, which is reasonable given . . ."

RC: 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).
AR: We will remove the statement about "merging" datasets, which was once foreseen in the project but has not been done at the moment this paper was written.
AC: "In general, the retrieved values are insensitive to the specific instrument evaluated. Absolute…"

RC: 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?
AR: Will rephrase, as these statements are definitively hard to understand. Yes, the last sentence refers to AATSR.
AC: "Also, even though spectral response differences are largest between MODIS and AVHRR (which results in a reflectance difference of up to 30–40 % (Trishchenko et al., 2002)), their retrieval values are much more similar. The difference between AATSR and both AVHRR and MODIS is largest for CER, so microphysical variables, which are derived from reflectance data only, appear to be most affected."

RC: 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.
AR: This is a very basic t-test, using a well-defined symbology. It is a test for significance of the difference between the mean values of two populations (i.e. $\mu 1$ = mean of population 1, $\mu 2$ = mean of population 2). The data were indeed re-gridded to a common grid, which is all explained in section 2.3.
AC: “The differences between mean values ($\mu 1$ and $\mu 2$) are almost always significant (t-Test p-value < 0.1, H_0: $\mu 1 = \mu 2$).”

RC: 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?
AR: The sensors are not collocated, but the data are. And the collocation should minimize differences due to observation times and observation area. The significance level is now mentioned in the correction above, but can also be found in the caption of Table 6. As said above, the data were re-gridded.
AC: “…when driven with spatiotemporally collocated satellite data obtained from three different sensors.”

RC: 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.
AR: We added references to give examples. We do not think that the AVHRR and MODIS cloud data record should be seen as one continuous, consistent data record. Rather, AVHRR provides the opportunity of long-term data coverage back to 1982, providing data that are at least comparable to MODIS. That certainly excludes local analyses, but rather refers to continental to global applications.
AC: “depending on the user’s application, 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: 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.
AR: As mentioned above, it is shown in Figure 4 middle and Figure 6 topright.
AC: “we see that COT uncertainty scales with COT itself (Figure 4 middle and Figure 6 topright)”

RC: 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).
AR: Will simplify and add references.
AC: “CC4CL COT values are at times unnaturally large, and the associated uncertainty reflects that. Also, it highlights under which conditions the optimal estimator converges to a solution with a relatively large divergence from the measurements, which here are associated with optically thick clouds or underlying snow/ice cover (see also Kahn et al., 2015, Wang et al., 2011). COT and CER

uncertainties are clearly largest, and reflect the limited information available with which to retrieve these values. For further possible explanations due to assumptions and limitations within the methodology applied, please see part II.”

Kahn, B. H., M. M. Schreier, Q. Yue, E. J. Fetzer, F. W. Irion, S. Platnick, C. Wang, S. L. Nasiri, and T. S. L'Ecuyer (2015), Pixel-scale assessment and uncertainty analysis of AIRS and MODIS ice cloud optical thickness and effective radius, J. Geophys. Res. Atmos., 120, 11,669–11,689, doi:10.1002/2015JD023950.

Wang, C., P. Yang, B.A. Baum, S. Platnick, A.K. Heidinger, Y. Hu, and R.E. Holz, 2011: Retrieval of Ice Cloud Optical Thickness and Effective Particle Size Using a Fast Infrared Radiative Transfer Model. J. Appl. Meteor. Climatol., 50, 2283–2297, https://doi.org/10.1175/JAMC-D-11-067.1

RC: p15, l11: “otherwise are” > “otherwise they are”
AR: Will rephrase.
AC: “otherwise they are”

RC: p15,l15: “may it stem” does not work in English; consider “whether it stems from. . . or”
AR: Will rephrase.
AC: “whether it stems from a cloud or the Earth’s surface”

RC: 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?
AR: Unfortunately, the table itself cannot be increased due to space limitations and a bug in the Python library applied to produce the table. The colours show cloud phase. Cloud type numbers are not as important, and we could have removed them as for Figure 15. At latitude 61°, we see that there are broken cloud fields in the area, which might have appeared in the sensor’s field of view but not in CALIPSO’s.

RC: p16,l8: consider “a conscious decision was made to [deliberately] trade. . .”
AR: Will rephrase.
AC: “For ESA Cloud_cci, a conscious decision was made to trade spectral information for time series continuity.”

RC: p16,l19: “on a first view” > “at first glance”
AR: Will rephrase.
AC: “At first glance, estimates of…”

RC: p18,l59: “synergic” > “synergistic”
AR: Will rephrase.
AC: “exploits synergistic capabilities of several EO missions”

RC: p18,l95: “accurate and precise”: These two were not discussed separately. Where was this done? If not, please clarify this statement.
AR: We will remove precise, which stands for a low standard deviation of errors (not shown here). The results are accurate due to the relatively low bias.
AC: “optically thick cloud retrievals are very accurate when compared against CALIOP (bias < 240 m)”

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). "

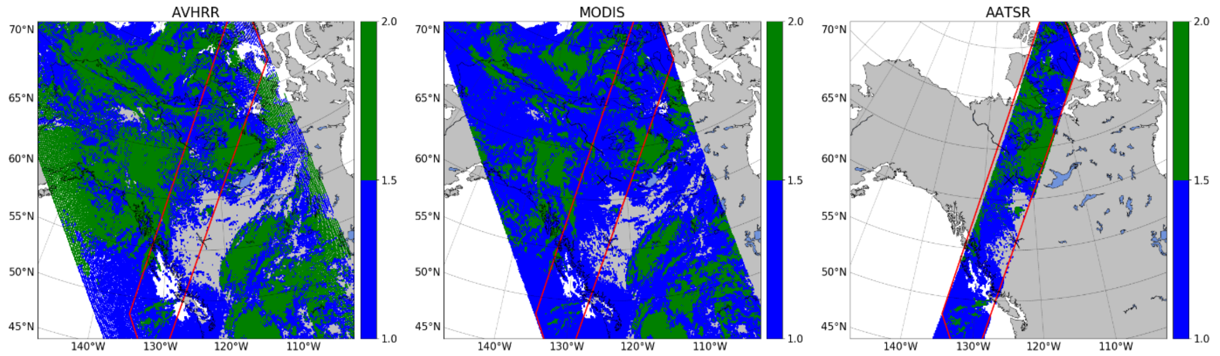


Figure 6. Cloud phase retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).

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."

Anonymous referee #3

RC: referee comment

AR: author response

AC: author's changes in manuscript

Received and published: 8 January 2018

Review of "The Community Cloud retrieval for Climate (CC4CL). Part I: A framework applied to multiple satellite imaging sensors", by Sus et al.

RC: The manuscript introduces a valuable approach to establish a common passive cloud retrieval applicable to a series of standard polar orbiters in order to create data sets usable for climatological studies. This would be an important step for the community and the usability of satellite products outside the satellite community. I also understand and acknowledge the need to base such an approach on well established methods instead of more experimental approaches as suggested by one of the other referees. The general presentation is of good/excellent quality. My two co-referees have elaborated on a number of specific technical and scientific details already. I want to focus on a more general weakness.

What exactly is the focus of this manuscript? If I missed important, clear, early statements in the existing text, I apologize. If not, the reader needs this guideline. In many places important details can not be given and are not explained owed to the shear extent of this project. In most cases the reader is then correctly referred to other publications where the methods of CC4CL are introduced. This way the purpose of the manuscript at hand becomes more and more unclear while reading through it. First impression is that the general method will be explained. But then the core retrieval techniques are explained elsewhere (McGarragh). Then a technical explanation of the ANN cloud mask is started, but it stays too short to be fully comprehensible. After the introduction of example cases Fig 3-5 and cross sections Fig 9-16, I expected an in-depth discussion of reason for differences and a quantitative validation (section titles containing "validation") or cross-comparison of all products, but the discussion stays very general and mostly describes differences. Proper validation is again shown elsewhere (Stengel).

The limited original content of this manuscript (correct me, if I'm wrong) is not reflected by the title and manuscript length (e.g. 8 figures 8-15 with very comparable content and not too surprising differences between active and passive sensor, but no quantitative validation). The authors should clarify the purpose of this manuscript and shorten parts published elsewhere even stronger. I suggest to consider these general points and a revision of the manuscript.

AR: We appreciate the comments of referee 3 and agree that some clarification is required to explain the purpose of the paper. Please note that this has also been pointed out by referee #2, so our answer here has been copied from our comments to reviewer #2.

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 "validation" with "examination" or "analysis" throughout the text. The reviewer is correct that no true validation study has been carried out here, and we rephrase in order to avoid misunderstandings.

Specific major issues:

RC: p3, line 27: "Moreover, the resulting time series are carefully validated ... (ISCCP, PATMOS-x, CM SAF, and MODIS Collection 6), reanalysis and model data (ERAInterim and EC-Earth), ground-truth synoptic observations, and CALIOP lidar data."

My understanding was that I would see that in this manuscript: You will only show CALIOP comparisons, will you? Could you please clarify.
AR: Yes, we only compared with CALIOP. We will add a reference here to the Stengel paper, and also a reference to our internal product validation report.
AC: "Moreover, the resulting time series were carefully validated against well-established climatologies (ISCCP, PATMOS-x, CM SAF, and MODIS Collection 6), reanalysis and model data (ERAInterim and EC-Earth), ground-truth synoptic observations, and CALIOP lidar data (Stengel et. Al, 2017, PVIR)."

RC: p8, section 4.3: I think you cannot call this chapter "validation". There is no systematic validation, only a few selected case studies, which mainly show the problems and no systematic quantitative validation. Four case studies of time height cross sections are shown only to present that lidar cth does not have much to do with passive cth? I also expected CER and COT validation somewhere.
AR: We agree with the reviewer and will, as mentioned above, replace "validation" with "comparison". As the reviewer mentions, this comparison shows the generic strengths and weaknesses of CC4CL, which certainly relates to the processing of passive imager data. However, the reader should appreciate the basic functionality of CC4CL and we find that these local comparisons are well suited for that purpose. Please also note that the CALIOP COT information is less reliable than CTH, which is why we did not compare with COT.
AC: "Comparison with CALIOP"

RC: p13, line 15: You mean, proper quantitative validation is shown in another paper ...Stengel et al 2017 ESSD? The retrieval method was shown in two other papers as well... McGarragh 2017 at JAS and AMT. Remind me about the reason for this manuscript?
AR: Please see our comments above.

Minor issues:

RC: p 2, line 33: AVHRR was not introduced before, was it?
AR: The reviewer is correct. Will rephrase.
AC: “Compared to the Advanced Very High Resolution Radiometer (AVHRR), MODIS has several…”

RC: p 2, line 53: What is r?
AR: The Pearson correlation coefficient.
AC: “(up to a Pearson correlation coefficient r = 0.94)”

RC: p 2, line 65: How can CTP and CTH be underestimated at the same time? Can you please comment?
AR: We agree that the use of the word “underestimates” twice suggests that CLARA-A2 is wrong in both cases. However, whereas CALIOP data are considered to be “truth” data, we will now simply state that CLARA-A2 has a *lower* CTP than the other retrievals, which is not an underestimation, just a different retrieval outcome.
AC: “Comparing CLARA-A2 to PATMOS-X, MODIS C6 and ISCCP, global CTP is lower by 4–90 hPa…”

RC: p 2, line 66: What is a "cloud phase bias ... of 9%"? Cloud phase? Liquid and ice? Or cloud cover?
AR: This refers to the fraction of liquid clouds.

RC: p 2, line 68: Low or high bias?
AR: We will specify.
AC: “+ 197 m”

RC: p 2, line 102: It would be nice to say at this early stage what the purpose of this particular manuscript is in ESA Cloud_cci? And what other parallel publications contribute? Later on, the reader gets the impression that everything relevant is introduced elsewhere.
AR: We agree that this needs clarification. We copied our answer to reviewer #1, who made a similar comment.
AC: “The European Space Agency has established the ESA Climate Change Initiative program (ESA CCI, 2015; Hollmann et al., 2013) in order to advance knowledge of the climate system through the generation of satellite based data records utilizing European and non-European assets. The CCI project’s primary focus is the production of thirteen Essential Climate Variables (ECVs) covering ocean, atmospheric, and land geophysical variables. With these data records CCI is aiming to fulfil highest climate requirements from the Global Climate Observing System (GCOS). The study presented here is part of the ESA CCI for clouds (ESA Cloud_cci), which has the objective to develop a state-of-the-art open-source community cloud retrieval algorithm being capable of processing passive satellite imager data for several decades. Both in part I and part II of this paper, we present the processing framework as developed within ESA Cloud_cci (CC4CL, part I), the detailed mechanisms of the optimal estimation retrieval (part II), and provide an initial assessment of the strengths and weaknesses of derived cloud parameters (part I). With CC4CL several decades of passive imaging satellite data have been processed and are made available to the user. The resulting climate data records (CDR) are presented in Stengel et al., 2017.”

RC: p 6, line 27: If this is the only description of ANNCOD available, you might at least want to cite Kox et al . 2014 (AMT, 7, doi:10.5194/amt-7-3233-2014) who introduced the idea and described in much more detail.
AR: Please also our answers to reviewer #1, who asked for a more detailed introduction of ANNCOD, which we will provide. Kox et al. developed an approach similar to ours for retrieving Cirrus COT and CTH, but we do not think that they introduced our idea for cloud masking.

RC: p6, line 51ff: This is all a slightly vague description, if it isn't detailed somewhere else. Why do you need ... after viewing angle dependency correction ... a whole set of thresholds? ANNCOD already gives an answer on the question cloud or no-cloud, doesn't it?

AR: That would mean that the ANNCOD perfectly reproduces CALIPSO data, which is not the case. The thresholds were necessary to avoid overestimation of cloud cover due to the sensitivity of the passive sensors. With the passive sensor we measure reflectance and temperatures, in contrast to CALIPSO which is independent of both. Strongly reflecting surfaces and/or difficult illumination conditions will create ambiguities. Especially under difficult illumination conditions such as twilight, and over ice/snow surfaces we needed to increase the thresholds to avoid overestimation and decrease the false alarm rate (knowing that we might miss some clouds). We made a skill analysis with CALIOP to find the most suitable thresholds. The viewing angle correction has nothing to do with this, but more or less you can see this as a sun-zenith and surface correction of the retrieval.

RC: p7, Figure 2: y-axis. It is PEC not 1-PEC shown, isn't it? Does the graph show that, at your threshold you are only correct by about 50%?? Please discuss.

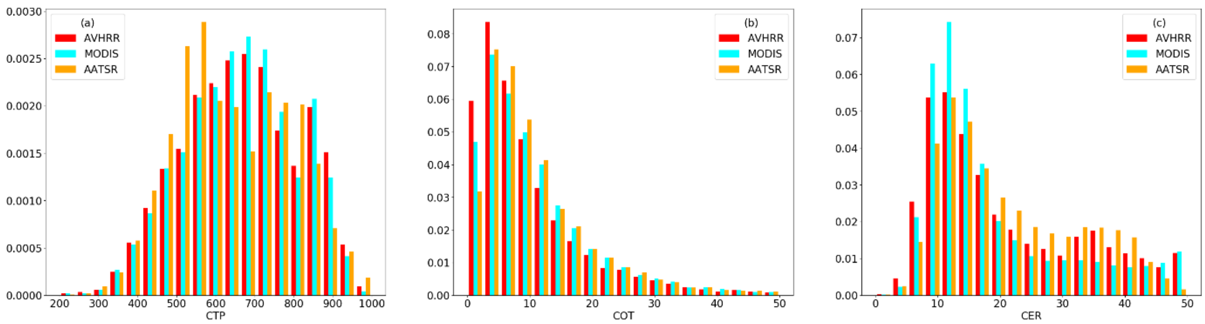
AR: The y-axis shows 100 – PEC [%]. The graph shows that the uncertainty increases to about 50 % at the threshold. This makes sense, as an ANNCOD value close to its threshold indicates that no clear distinction between cloud/no cloud can be made, thus the highest uncertainty. The larger the difference between ANNCOD and its threshold, the lower the associated cloud mask uncertainty.

RC: p10, line 10: "consistent". You could also say its all over the place, with different physical reasons in any single column. This is not a validation. You even tried to correct cth for cc4cl and still have big problems.

AR: We do find that Figure 9 shows very similar retrieval results of CTH for all three sensors, except in sector 2. We are referring here to the agreement *amongst* sensors, not between sensors and CALIOP data.

RC: p11, Figure 7: Please make the labels consistent with the rest of the manuscript: n18->avhrr, myd->modis ...

AR: We will modify labels accordingly.



The Community Cloud retrieval for Climate (CC4CL). Part I: A framework applied to multiple satellite imaging sensors

Oliver Sus[1], Martin Stengel[1], Stefan Stapelberg[1], Gregory McGarragh[2], Caroline Poulsen[3], Adam C. Povey[4], Cornelia Schlundt[1], Gareth Thomas[3], Matthew Christensen[2,3], Simon Proud[2], Matthias Jerg[1], Roy Grainger[4], and Rainer Hollmann[1]

[1]DWD - Deutscher Wetterdienst, Frankfurter Straße 135, 63067 Offenbach, Germany
[2]Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, U.K.
[3]RAL Space - Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, U.K.
[4]National Centre for Earth Observation, Atmospheric, Oceanic and Planetary Physics, University of Oxford, Parks Road, Oxford OX1 3PU, U.K.

Correspondence to: Martin Stengel
martin.stengel@dwd.de

Abstract. We present here the key features of the Community Cloud retrieval for CLimate (CC4CL) processing algorithm. We focus on the novel features of the framework: the optimal estimation approach in general, explicit uncertainty quantification through rigorous propagation of all known error sources into the final product, and the consistency of our long-term, multi-platform time-series provided at various resolutions, from 0.5° to 0.02°.

By describing all key input data and processing steps, we aim to inform the user about important features of this new retrieval framework, and its potential applicability to climate studies. We provide an overview of the retrieved and derived output variables. These are analysed for four, partly very challenging, scenes collocated with CALIOP (Cloud-Aerosol lidar with Orthogonal Polarization) observations in the high-latitudes and over the Gulf of Guinea/West Africa.

The results show that CC4CL provides very realistic estimates of cloud top height and cover for optically thick clouds but, where optically thin clouds overlap, returns a height between the two layers. CC4CL is a unique, coherent, multi-instrument cloud property retrieval framework applicable to passive sensor data of several EO missions. Through its flexibility, CC4CL offers the opportunity for combining a variety of historic and current EO missions into one data set, which, compared to single sensor retrievals, is improved in terms of accuracy and temporal sampling.

1 Introduction

The European Space Agency has established the ESA Climate Change Initiative program (ESA CCI, 2015; Hollmann et al., 2013) in order to advance knowledge of the climate system through the generation of satellite based data records utilizing European and non-European assets. The CCI project's primary focus is the production of thirteen Essential Climate Variables (ECVs) covering ocean, atmosphere, and land geophysical variables. With these data records, CCI is aiming to fulfil highest climate requirements from the Global Climate Observing System (GCOS). This study presented here is part of the ESA CCI for clouds (ESA Cloud_cci), which has the objective to develop a state-of-the-art open-source community cloud retrieval algorithm which shall be capable of processing passive satellite imager data covering several decades. Both in part I and part II of this paper, we present the processing framework as developed within ESA Cloud_cci (CC4CL, part I), the detailed mechanisms of the optimal estimation retrieval (part II), and provide an initial assessment of the strengths and weaknesses of derived cloud parameters (part I). With CC4CL, several decades of passive imaging satellite data have been processed and are made available to the user. The resulting climate data records (CDR) are presented in Stengel et al. (2017).

Satellite data are an essential source of information for understanding and predicting climate change. They provide global long-term observations from which geophysical parameters can be derived. These are used for time-series analysis of climate variables, and also for the assimilation into or validation of climate models (Comiso and Hall, 2014; Yang et al., 2013). A paramount goal of these efforts is the comprehensive characterization of the global energy and water budgets (Stephens et al., 2012).

Clouds considerably influence the global energy budget through [..[1]]direct forcing effects (Kiehl and Trenberth, 1997). However, clouds are difficult to quantify, having highly variable composition and spatiotemporal distributions, and produce the largest uncertainty in our understanding of climate change (Norris et al., 2016; IPCC, 2013). Observations from passive imagers do not sufficiently resolve several important cloud properties, such as vertical structure, sub-pixel heterogeneity, the cloud boundary, and the column-integrated ice or liquid water path. Several secondary variables (cloud forward model assumptions, state of surface and atmosphere, viewing geometry, sensor calibration and spectral response uncertainties) further complicate cloud retrievals, [..[2]]and insufficient knowledge on their state propagates uncertainties into the derived cloud properties (Hamann et al., 2014). Nonetheless, passive satellite imagers are the most widely used instruments for cloud retrievals as they provide long-term, global coverage at acceptable cost for the user.

There are several satellite-based retrieval frameworks. One of the earliest is the International Cloud Climatology Project (ISCCP) (Rossow and Schiffer, 1999). ISCCP provides data on cloud products for 1983–2009, and introduced a cloud type classification based on cloud optical thickness-cloud top pressure (COT-CTP) joint histograms that is still popular even today. Continuously reprocessed retrieval systems include Pathfinder Atmosphere Extended (PATMOS-x) (Heidinger and Pavolonis, 2009; Heidinger et al., 2012), EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) cLoud, Albedo and RAdiation (CLARA-A1) (Karlsson et al., 2013), and MODIS Collection 6 (MODIS C6) (Platnick et al., 2017) as well as the CERES-MODIS products (Minnis et al., 2011). These retrievals vary in their [..[3]]ancillary data sources, approaches, and complexity but generally use radiative transfer models and/or derived look-up tables (LUT) to provide a clear-sky reference and for simulating atmospheric and cloud contributions to top of atmosphere (TOA) radiances. Cloud properties are derived using decision trees and thresholding (PPS in CLARA-A1), LUT based inversions (MODIS C6), or optimal estimation theory (PATMOS-X). COT and CER (cloud effective radius) are usually calculated following Nakajima and King (1990). However, the derived microphysical variables are not guaranteed to be radiatively consistent with independently derived cloud parameters, as

[1]removed: shielding and
[2]removed: propagating
[3]removed: auxiliary

most of the retrieval methods are separated into solar and thermal methods even though measurements in these spectral regions are not independent of parameters retrieved in the other. For cloud masking, the retrieval frameworks apply various approaches such as Naïve Bayes (PATMOS-X), dynamic thresholding (CLARA-A1), or a [..[4]]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). Finally, cloud phase or type is determined as a function of a combined convergence/cloud top temperature (CTT)-test (CLARA-A1), the Pavolonis et al. (2005) threshold algorithm (PATMOS-X), or a bispectral decision tree considering channels at 8.5 and 11 µm (MODIS C6). Compared to [..[5]]the Advanced Very High Resolution Radiometer (AVHRR), MODIS has several additional spectral channels at shortwave infrared (SWIR) wavelengths that provide cloud microphysical information (Platnick et al., 2017), such that MODIS data provide more information for retrieving cloud products than AVHRR. Still, the MODIS C6 cloud top retrieval loses sensitivity for optically thinner clouds (COT < 2, Menzel et al. (2010); Christensen et al. (2013)), and [..[6]]beyond a penetration depth into the cloud [..[7]]corresponding to > 1 cumulative optical depth (Baum et al., 2012). This complicates validation against independent measurements such as those derived from lidar, which explicitly observe the cloud top. Despite some promising results, these studies show that current retrievals underestimate cloud top pressure for optically thin clouds [..[8]]due to the inherent limitation of the spectral information content of passive IR channels.

There are numerous studies that evaluate the performance of the aforementioned retrievals for cloud [..[9]]fraction with weather station data, such as over the Mediterranean (Sanchez-Lorenzo et al., 2017) and [..[10]]contiguous United States (Sun et al., 2015). The results are variable, but generally show that the inter-annual correlation is highest for PATMOS-X (up to a Pearson correlation coefficient $r = 0.94$) and lowest for CLARA-A1 ($r = 0.20 – 0.7$). More importantly, these studies emphasize the difficulty of deriving reliable cloud [..[11]]fraction trends from AVHRR time series, as the retrievals overestimate the change in cloud [..[12]]fraction by as much as an order of magnitude (Sun et al., 2015). There are also several evaluation or validation studies for individual retrieval algorithms. Differences between PATMOS-X microphysical retrievals using MODIS data and the collocated MYD06 product are within retrieval uncertainty (Walther and Heidinger, 2012). CLARA-A2 underestimates global cloud top height (CTH) by 840 m compared to CALIOP. Comparing CLARA-A2 to PATMOS-X, MODIS C6 and ISCCP, [..[13]]global CTP is lower by 4–90 hPa and has [..[14]]an absolute cloud phase bias of lower than 9 % (Karlsson et al., 2016). MODIS C6 CTH bias for low-level boundary layer water clouds is + 197 m compared to CALIOP, and the phase detection has been improved for [..[15]]liquid clouds. However, the detection of [..[16]]optically thin ice clouds over warm, bright

[4]removed: battery of threshold tests (MODIS C6)
[5]removed: AVHRR
[6]removed: sees
[7]removed: to an optical thickness of approximately unity
[8]removed: even when the full potential of MODIS spectral coverage is used.
[9]removed: cover
[10]removed: conterminous
[11]removed: cover
[12]removed: cover
[13]removed: it underestimates global CTP
[14]removed: a
[15]removed: optically thin ice
[16]removed: supercooled water clouds remains problematic (Baum et al., 2012)

surfaces remains problematic (Marchant et al., 2016). For MODIS C5, global CTH was underestimated relative to CALIOP by 1.4 km (Holz et al., 2008).

Satellite observations of clouds are available for the last 40 years. However, data need to be carefully processed and analysed in order to derive a consistent long-term data record from several inter-calibrated satellite platforms. Consistency in approach can be traded for continuity of results, and multi-platform algorithms could exploit additional data when newer sensors become available. Modern sensors provide improved spectral coverage and spatial resolutions and, thus, potentially better cloud retrievals. However, their data records are too short to produce climatologies of at least 30 years, and discontinuities are built into time series when higher resolution satellite data are input to the processing. Major complications of cloud retrievals include optically transparent clouds, multi-layer or overlapping clouds, and effective cloud top height determination. The degree to which these complications can be addressed depends on the nature of the retrieval and the type of input satellite data used. MODIS provides a much larger spectral sampling than the six AVHRR heritage channels. MODIS and atmospheric sounders are clearly superior when detecting cloud height through the application of the "CO2-slicing" technique. However, when consistent climatologies are to be built, time series length and spatiotemporal resolution limit the choice in retrieval type and input satellite data.

[..[17]]

In order to produce the cloud CDR presented here, we used satellite data from MODIS Aqua and Terra (2000–2014) (King et al., 1992), AVHRR on NOAA-7 to NOAA-19 and METOPA (1978–2014) (Jacobowitz et al., 2003), ATSR-2 on ERS-2 (1995–2003), and AATSR on ENVISAT (2002–2012). Only the AVHRR-equivalent channels from MODIS and AATSR are used. Hence, the resulting retrieval data are hereafter referred to as the "AVHRR heritage dataset". Moreover, the resulting time series [..[18]]were carefully validated against well-established climatologies (ISCCP, PATMOS-x, CM SAF, and MODIS Collection 6), reanalysis and model data (ERA-Interim and EC-Earth), ground-truth synoptic observations, and CALIOP lidar data (Stengel et al., 2017, 2018).

The CC4CL core algorithm was developed in a modular fashion and provides open-source access to support distribution and development within the scientific community. Particular attention was paid to allow processing of multiple instruments within a single framework, thus maximising the consistency of cloud products independent of the sensor source. The framework accounts for physical consistency amongst all output variables and radiative consistency amongst all input satellite radiances. This is an improvement [..[19]]over other established retrieval frameworks. These commonly derive COT and CER by adopting the Nakajima and King (1990) approach, but macrophysical products are estimated independently and are thus radiatively inconsistent with the former variables. [..[20]]Here, parameters are retrieved simultaneously, providing a retrieval that is

[17] removed: The European Space Agency has established the ESA Climate Change Initiative program (ESA CCI, 2015; Hollmann et al., 2013) in order to advance knowledge of the climate system. The project's primary focus is the production of thirteen Essential Climate Variables (ECVs) for ocean, atmosphere, and land. The main objective of ESA Cloud_cci is to develop a state-of-the-art open-source community cloud retrieval algorithm which is capable of processing passive imager data for a number of (non-)European satellites covering several decades. We

[18] removed: are

[19] removed: on

[20] removed: Another novel

radiatively consistent over the wavelengths of the observations, given that the instrument's noise characteristics are well known. 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. With these criteria in mind, the Optimal Retrieval of Aerosol and Cloud (ORAC) (Thomas et al., 2009a; Poulsen et al., 2012) was chosen from three competing algorithms [..[21]]in a "round-robin" (i.e. each algorithm is tested against all other algorithms) analysis. All algorithms have proven their maturity for deriving the considered cloud parameters (cloud cover, liquid and ice water path, cloud top height) from AVHRR and MODIS data (Stengel et al., 2015).

In this study, we present the key features of the CC4CL processing algorithm. We particularly focus on discussing the [..[22]]key features of the framework: the optimal estimation approach in general, the explicit uncertainty quantification through rigorous propagation of all known error sources to the final product, and the consistency of our long-term, multi-platform time-series provided at various resolutions, from 0.5° to 0.02°. By describing all key input data and processing steps, we inform the future user about important features of this new processing framework, and its potential applicability in climate studies. We provide an overview of the retrieved and derived output variables. These are initially [..[23]]examined in a detailed analysis of retrieval results that we collocated with CALIOP observations for three scenes in the Arctic and one scene in the Gulf of Guinea/West Africa. The results show that CC4CL produces mixed-layer estimates for cases where optically thin clouds overlap, but provides [..[24]]realistic estimates of cloud top height and cover for optically thick clouds.

2 Data and methods

2.1 Level-1 (L1) satellite data

2.1.1 AVHRR

The Advanced Very High Resolution Radiometer (AVHRR) is a cross-track scanner with a 2900 km swath width, providing almost daily global coverage. The sensor is equipped with six spectral channels (Table 01), out of which only five can be transmitted simultaneously so that either channel 3a or 3b is available. In-flight calibration is performed only for thermal channels, using a stable blackbody and a space view as references. AVHRR has been mounted on several NOAA platforms as well as on EUMETSAT's MetopA/B, all of which are sun-synchronous, polar orbiting satellites. Due to a lack of orbit control technology for all NOAA AVHRR's, there is considerable orbit drift in equatorial crossing times (ECT) both for morning (ECT $<$ 12:00 LST) and afternoon (ECT $>$ 12:00 Local Solar Time (LST)) satellites. To reduce drift-induced changes in retrieved cloud properties, any AVHRR is replaced with its corresponding successor once available ($=$ the AVHRR prime record). Typically, one morning and one afternoon NOAA satellite are in orbit at any time.

[21] removed: within a "Round Robin" selection process (Stengel et al., 2015).
[22] removed: novel
[23] removed: validated in a comprehensive and
[24] removed: very

For CC4CL, we use Global Area Coverage (GAC) L1c data on a reduced spatial resolution of 1.1 km × 4 km at nadir (Devasthale et al., 2017). The AVHRR GAC L1c data record, including advanced inter-calibration efforts, was produced for ESA Cloud_cci and CMSAF (Schulz et al., 2009; Karlsson et al., 2013). CC4CL processed AVHRR data from 08/1981 (NOAA-7) up to 12/2014 (MetopA + NOAA-19). We applied a filtering technique to channel 3b data, and a database algorithm for splitting midnight orbits and blacklisting.

2.1.2 MODIS

The Moderate Resolution Imaging Spectroradiometer (MODIS) is carried by NASA's Terra and Aqua satellite platforms in a near sun-synchronous polar orbit at 705 km altitude. Due to orbit control, ECT is a constant 10:30 LST for Terra, and 13:30 LST for Aqua. The Aqua satellite is a member of the "A-Train" constellation, which also includes the CALIOP and CloudSat satellites. MODIS is a cross-track scanner with a 2330 km swath width, producing a complete near-global coverage in less than two days (Xiong et al., 2009).

CC4CL is applied to Collection 6 MOD021km (Terra) and MYD021km (Aqua) L1b input data (NASA LP DAAC, 2015). For the AVHRR-heritage dataset produced here, the NASA Goddard [..[25]]Space Flight Center performed a spectral subsetting of the 36 MODIS channels available (see Table 01 for the channels extracted), and data were directly shipped to ECMWF (European Centre for Medium-Range Weather Forecasts) for archiving. The files are stored in HDF-EOS format at 1km spatial resolution, with the 250 m and 500 m channels having been aggregated to 1 km resolution. MODIS L1b data are organized in granules, each of which contains ~5 minutes of MODIS data or ~203 scan lines. Geolocation information is provided in separate files for Terra (MOD03) and Aqua (MYD03), containing geodetic latitude and longitude and solar/satellite zenith and azimuth angles. L1b data are corrected for all known instrumental effects through on-board calibrator data, and are organized into a viewing swath matching the geolocation file structure (MODIS Characterization Support Team, 2009). With CC4CL, we processed data from 02/2000 (Terra) or 08/2002 (Aqua) to 12/2014.

2.1.3 ATSR-2 and AATSR

The second and third generation Along Track Scanning Radiometers (ATSR-2 and Advanced ATSR, Merchant et al. (2012)) were launched on ESA's polar orbiting satellites ERS-2 and ENVISAT in 04/1995 and 03/2002, respectively. Both platforms were put into a sun-synchronous orbit at ~780 km altitude, with ECT = 10:30 LST for ERS-2 and ECT = 10:00 for ENVISAT. Both ATSRs are identical in their overall configuration except for data transfer bandwidth (Table 01). ATSR is [..[26]]is equipped with on-bard calibration capabilities, such as two black-body targets for the thermal channels and a sun-illuminated opal target for the visible/near-infrared channels. ATSR uses a dual-view system: a nadir view, and a forward view scanning the surface at an angle of 55°. The continuous scanning pattern produces a nadir resolution of approximately 1 km × 1 km with a swath width of 512 pixels or ~500 km, providing global coverage every six days.

[25] removed: space flight centre

[26] removed: designed to be self-calibrating, with two on-board

We used no forward view data for cloud retrievals, as the 3-dimensional cloud structure produces parallax effects which are not accounted for within the current forward model. With CC4CL, we processed ATSR data from launch date until 05/2003 (ERS-2) and 04/2012 (ENVISAT).

2.2 [..[27]]Ancillary data

2.2.1 ERA-Interim

We use ERA-Interim data as first-guess input for the retrieval of surface temperature, and as input [..[28]]to a neural network cloud mask (see Section 3.3.1). ERA-Interim is a reanalysis of the global atmosphere, and is available from 1979 until today (Berrisford et al., 2011; Dee et al., 2011). The atmospheric profile variables are defined at 60 vertical levels. The original horizontal resolution is defined through a T255 spherical-harmonic representation for the basic dynamical fields, and through a reduced Gaussian grid with ~79 km spacing for surface fields. We downloaded ERA-Interim data from the ECMWF's MARS archive at a spatial resolution of 0.72°(the default preprocessing grid resolution), and at a higher resolution of 0.1° for the neural network cloud mask input variables (Table A1). We acquired analysis (i.e. not forecast) data at 6-hourly timesteps. After download, all files were remapped to the CC4CL preprocessor grid through Climate Data Operators (CDO, 2015). This was necessary, as ERA-Interim coordinates are defined at the cell boundaries, whereas they are defined at the cell centres within CC4CL. The reanalysis data are temporally interpolated onto the satellite image's centre time by linearly weighting the files before and after.

ERA-Interim's land-surface model still needs to be improved in terms of its simulation of soil hydrology and snow cover. This affects the utilization of satellite data over land surfaces within ERA-Interim, which has negative effects on the representation of clouds and precipitation (Berrisford et al., 2011). The confidence in temperature trend estimates, however, has improved considerably so that ERA-Interim data have been used as an alternative to observational datasets to monitor climate change (Willett et al., 2010).

2.2.2 Land use

We downloaded United States Geological Service (USGS) Land Use/Land Cover raster data from the global land cover characteristics database (U.S. Geological Survey, 2016). This was necessary, as early AVHRR data are distributed without masking information. The USGS data are used as a land sea mask within the optimal estimation retrieval (Section 3.3.3), as well as a land cover classificator within the cloud mask and the Pavolonis cloud typing scheme (Section 3.3.2). The dataset is defined on a regular lat/lon grid with 0.05° resolution. The USGS land cover classification was primarily derived from 1 km AVHRR Normalized Difference Vegetation Index (NDVI) 10-day composites for April 1992 through March 1993 (U.S. Geological Survey, 2016).

[27] removed: Auxiliary

[28] removed: for the

2.2.3 Land surface BRDF

MODIS Collection 6 Bidirectional Reflectance Distribution Function (BRDF) data (MCD43C1, Schaaf and Wang (2015)), providing kernel weights for the Ross-Thick/Li-Sparse-Reciprocal BRDF model, are used within the retrieval scheme to set surface albedo and bidirectional reflectance distribution conditions. These data are available every 8 days derived from cloud-cleared 16-day Terra and Aqua measurements, and provided in HDF-EOS format at 0.05° spatial resolution. MCD43C1 data are classified as high-quality given sufficient observations, and otherwise a low quality estimate is produced based on climatology anisotropy models. Validation against albedo measurements made at Baseline Surface Radiation Network (BSRN) sites show that the black-sky and white-sky albedo computed from the single sensor MCD43A1 high-quality product are well within 5 % of the measured albedo, while the low-quality product is within 10 % (Lucht, 1998).

We regridded MCD43C1 data to instrument resolution through bilinear interpolation, and filled missing pixels within the time series with pixel values of the temporally closest 8-day composite file providing valid data. For the pre-MODIS era, we produced a BRDF climatology by averaging all data available for a particular 8-day time slot. MCD43C1 kernel weights are applied to all CC4CL sensors, neglecting differences in spectral response functions[..[29]]. This might result in retrieval biases, particularly in spectral regions that are sensitive to rapidly changing environmental processes such as vegetation growth (near-IR). 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.

2.2.4 Land surface emissivity

For land surface emissivity, we used the Cooperative Institute for Meteorological Satellite Studies (CIMSS) global land surface infrared emissivity database created by the Baseline Fit method (Seemann et al., 2008). These data are derived from the MODIS operational land surface emissivity product (MOD11), to which the fit method is applied for filling spectral gaps between channels. CIMSS emissivity data are available on a monthly basis at ten wavelengths with 0.05° spatial resolution.

As for BRDF, we produced a land surface emissivity climatology for the pre-MODIS era by averaging all data available for a particular month.

2.3 Collocating CC4CL Level-2 (L2) data and CALIOP

We resampled CC4CL L2 data to a regular latitude/longitude grid at 0.1° × 0.1° resolution. This resampling is required for an intercomparison of CC4CL L2 data on a common grid[..[30]]. However, note that differences in sensor spatial resolution [..[31]]can lead to significantly different PDFs within a grid box, the effect of which we did not analyse. CALIOP's Level 2 5 km Cloud Layer data were produced by averaging over ∼14 beams with 70 m diameter taken every 335 m within a 5 km along-track corridor. Thus, CALIOP data have a 70 m across-track × 5 km along-track spatial resolution (see also Holz et al. (2008)), and the size of the corresponding CC4CL grid box is approximately 11 km (meridional) × 2.9 to 5.6 km (zonal). As

[29] removed: as the surface is a relatively minor component of the observed signal

[30] removed: , as

[31] removed: are reduced when averaging all values available for each grid box

a consequence, the CC4CL grid boxes are larger than the reference CALIOP pixels, but are still small enough to resolve some of the cloud features that CALIOP observes. Note that AVHRR GAC data were produced by averaging 5 neighbouring pixels across-track, but CALIOP data were averaged along-track.

3 The CC4CL retrieval system

3.1 Heritage

In the early stages of the Cloud_cci project, a "Round Robin Exercise" evaluated three different algorithms regarding their applicability for retrieving cloud parameters from satellite data (Stengel et al., 2015), which were 1) the operational processing system of the CM SAF (2015), 2) the Clouds from AVHRR Extended (CLAVR-X) algorithm used to generate the PATMOS-x climatology (Heidinger et al., 2013), and 3) the ORAC retrieval which was previously used to produce the [..[32]]Global Cloud and Aerosol Dataset Produced by the Global Retrieval of ATSR Cloud Parameters and Evaluation (GRAPE) data set (Thomas et al., 2009b; Natural Environment Research Council et al., 2015). All three algorithms were driven with identical MODIS and AVHRR input data and ERA-Interim meteorological background information for five days in 2008. The results where analysed with respect to CloudSat, CALIOP and AMSR-E reference data.

Based on the outcomes of that study (Stengel et al., 2015), ORAC was selected to be the cloud retrieval scheme within CC4CL. Moreover, code modifications were identified and characterized to render ORAC fit for the purpose of ECV production.

3.2 Preprocessing

The CC4CL preprocessor initially defines the dimensions and content of the sensor and preprocessing grids (Figure 01).

The sensor grid has the same extent and resolution as the input orbit or granule. The sensor grid is filled with sensor radiances and angles, time, and geolocation data (section 2.1), whereas surface BRDF (section 2.2.3), snow/ice coverage (from ERA-Interim, section 2.2.1), and surface emissivity (section 2.2.4) are bilinearly interpolated onto that grid. We use BRDF data over land only. For sea pixels, the Cox and Munk ocean surface reflectance model calculates BRDF coefficients as a function of ERA-Interim wind speed. These coefficients also contain foam and underlight components (Sayer et al., 2010). The albedo of snow/ice covered pixels is set to globally constant values of 0.958 (Ch1, CC4CL ID as in Table 01), 0.868 (Ch2), 0.0364 (Ch3), and 0.0 (Ch4), and is area-weighted in the event of fractional sea/ice cover.

The preprocessing grid is a regular latitude/longitude grid that covers the extent of the sensor grid, but at a coarser resolution of 0.72°× 0.72°. It is used to store the average of all sensor angle and surface emissivity values falling within a grid box and spatially interpolated (nearest neighbour) land-use data (section 2.2.2). ERA-Interim variables were transformed before input to the preprocessing grid as described in section 2.2.1. For profile variables, vertical geopotential coordinates are calculated from pressure coordinates.

[32]removed: GRAPE

The preprocessor then calls the cloud mask (section 3.3.1) and cloud typing (section 3.3.2) algorithms. Finally, the Radiative Transfer for TOVS (RTTOV) model is executed on the preprocessing grid data as defined by ERA-Interim surface and profile variables. RTTOV outputs profiles of cloud transmittance both above and below cloud for the shortwave channels and emissivity for the longwave channels. For details on RTTOV and the forward model, see part II of this paper (McGarragh et al., 2017c).

All data are written to NetCDF files. In theory, the main processor would evaluate these inputs twice, assuming different cloud phases (e.g. ice and liquid). In practice, ORAC uses the preprocessed cloud mask and phase to select an appropriate method to reduce processing time.

3.3 CC4CL cloud retrieval

3.3.1 Cloud detection

The CC4CL cloud mask is produced by (1) estimating pseudo CALIOP cloud optical depth (ANNCOD) from L1 measurements with an artificial neural network (ANN), (2) correcting ANNCOD for viewing-angle dependencies, and (3) classifying ANNCOD into binary cloud mask information by thresholding.

CC4CL applies a set of ANN for cloud masking, one for each of the illumination conditions day (solar zenith angle $\theta_0 < 80°$), night ($\theta_0 \geq 90°$), and twilight ($80 \leq \theta_0 < 90°$). The ANNs are multilayer perceptrons with one input layer, one hidden layer with 50 neurons, and one output layer, which produces ANNCOD ranging from 0 to 1. Through incremental testing, we found that 50 neurons was the value for which the trade-off between output quality and computing speed was optimal. For the input layer, input variables are surface temperature, snow/ice cover, and the land/sea mask for all three cloud masks. Regarding sensor data, input channels are Ch1, Ch2, Ch5, Ch6, and Ch5-Ch6 for the day ANN, Ch4, Ch5, Ch6, Ch5-Ch4, and Ch5-Ch6 for the night ANN, and Ch5, Ch6, and Ch5-Ch6 for the twilight ANN.

The various ANNs were trained with NOAA-18 AVHRR L1c data, [..[33]]ancillary information (ECMWF land-sea mask, snow-ice mask, and surface temperature), and cloud optical depth (COD) "truth" data obtained from CALIOP'S 532 nm lidar product (CAL_LID_L2_05kmCLay-Prov-V3-01). AVHRR Ch3a data were generally excluded. We trained the day ANN with all remaining AVHRR channels, but also excluded Ch3b to be consistent with those NOAA platforms that switch between Ch3b transmission at night and Ch3a at day (NOAA-16, NOAA-17, MetopA) For night and twilight conditions, we produced ANNs both with and without Ch3b data input. This was necessary to avoid misclassification of very cold clouds and/or land surfaces due to Ch3b's very low signal-to-noise ratio. In addition to the days evaluated in the "Round Robin" comparison, we selected 12 further training days in 2008 that contain collocations between NOAA-18 and CALIOP, represent COD seasonality, and provide global coverage. Prior to training, all CALIOP COD values > 1 were set to unity. [..[34]]Ancillary data input are the ERA-Interim skin temperature, a snow/ice mask derived from ERA-Interim snow depth and sea ice concentration, and the USGS land/sea mask. Finally, we applied a simple correction algorithm to remove a cosine viewing-angle dependency of

[33] removed: auxiliary

[34] removed: Auxiliary

retrieved ANNCOD. This was necessary, as the maximum viewing angle in the AVHRR training dataset was just 35°. The binary cloud mask is estimated by classification of ANNCOD data into clear and cloudy through a set of threshold values. The thresholds themselves vary depending on illumination and surface conditions, namely land, sea, and snow/ice cover (Table 02), and were quantified [..[35]]through iterative optimization. They are fixed for all sensors and orbits. As the ANN was trained with AVHRR data only, differences in spectral response functions need to be considered before the ANN can be applied to MODIS and AATSR. 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). The resulting coefficients were applied to MODIS and AATSR satellite data before ANN input.

We estimate cloud mask uncertainty based on the assumption that this uncertainty is inversely proportional to the difference between retrieved ANNCOD and the threshold applied. As a first step, we generated a CALIOP cloud mask by application of a clear/cloudy threshold value of 0.05. The CALIOP cloud mask is then compared with the collocated ANN mask by quantification of a Percent Correct (PEC) score. PEC estimates the ratio between all correctly classified pixels and the number of all pixels analysed. Finally, the "truth" uncertainty is defined as $100 - \text{PEC}\ \%$. We then established the statistical relationship between this uncertainty and the ANNCOD difference to its threshold. Before application of the approach, we normalised differences (ND) to 1. We found a linear correlation between uncertainty and ND for clear cases given by

$$y = 37.275 \times \text{ND} + 49.2, \tag{1}$$

and a second order polynomial correlation for cloudy cases (Figure 02)

$$y = 54.133 \times (\text{ND} - 1)^2 + 1.862. \tag{2}$$

The equations of these regression fits are used within CC4CL to quantify cloud mask uncertainty as a function of ND.

3.3.2 Cloud typing

Cloud phase is determined by application of the Pavolonis cloud typing algorithm (Pavolonis et al., 2005). The Pavolonis algorithm outputs 6 cloud types (Table 04), which we then reclassified into water or ice clouds: liquid = fog/warm liquid/supercooled, ice = opaque ice/cirrus/overlap. For CC4CL, the fog type test was deactivated. The algorithm always uses the 0.65, 11, and 12 µm channel data. It reads 3.75 µm data whenever available, and 1.65 µm otherwise. These two different approaches produce nearly identical results, except for certain thin clouds and cloud edges (Pavolonis et al., 2005). In addition, we introduced two new cloud types within CC4CL. In response to validation studies, we decided to change the phase of ice clouds whose retrieved CTT is > 273.16 K, the freezing point of water (new cloud type = SWITCHED_TO_WATER), and of water clouds whose CTT < 233.16 K, the lower limit of supercooled water (SWITCHED_TO_ICE).

[35] removed: by trial and error

The Pavolonis algorithm has weaknesses in detecting cirrus clouds at high latitudes, which are often misclassified as opaque ice clouds. Performance is considerably better when the VIIRS algorithm is used, which provides additional channels and threshold tests. However, these cannot be applied to our AVHRR heritage dataset (Pavolonis et al., 2005).

3.3.3 Optimal estimation retrieval of COT, CER and CTP

The optimal estimation retrieval ORAC is a non-linear statistical inversion method based on Bayes' theorem (Rodgers, 2009). A state vector containing all variables to be retrieved is optimized to obtain the best fit between observed TOA radiances and radiances simulated by a forward model. The retrieval problem is that of finding the minimum value of a cost function. This function is based on a χ^2 distribution, which is a combination of the squared deviations between the measurements and the forward model and the retrieved state vector and the a priori state vector, each weighted by their associated uncertainties. The important benefits of ORAC, relative to more traditional retrieval methods, are that cloud parameters are retrieved using information in all satellite channels simultaneously, so that the retrieved parameters provide a robust representation of the short-wave and long-wave radiance effects of the observed cloud. The algorithm estimates the retrieval uncertainty, which [..[36]]quantifies the range of values that are feasible considering the uncertainty in the satellite measurements, ancillary data and ORAC forward model. For a more detailed description of the ORAC algorithm see part II of this publication (McGarragh et al., 2017c).

3.4 Post-processing

For each input pixel, the main processor [..[37]]evaluates these inputs twice, assuming different cloud phases (e.g. ice and liquid[..[38]]). In theory, ORAC could use the preprocessed cloud mask and phase to select an appropriate method to reduce processing time. The postprocessor will then select the appropriate output variables according to the Pavolonis cloud phase. As described in section 3.3.2, the postprocessor changes cloud phase in case retrieved CTT does not match the Pavolonis phase. Finally, output variables are written to primary and secondary NetCDF files (Table 05).

4 L2 data - [..[39]]initial [..[40]]analysis

We first examine CC4CL cloud properties for one sample scene that extends from approximately 100° W to 170° W and 45° N to 75° N over North America. We focus on the consistency of retrieval values derived from different sensors (AVHRR, MODIS, AATSR). This includes pixel-based uncertainties of the key variables (CTP, COT, CER, and cloud mask). We then perform

[36] removed: can be thought as a measure of the consistency between the retrieved cloud parameters and the satellite measurements(Poulsen et al., 2012)
[37] removed: produces retrieval values for both
[38] removed: clouds.
[39] removed: analysis and
[40] removed: validation

[..[41]]an analysis of retrieved cloud properties, for which CALIPSO data are our reference. This [..[42]]comparison is limited to three high-latitude scenes for which collocations for all sensors with CALIOP are available.

4.1 CC4CL cloud properties

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 ([..[43]]Figures 03 to 06). Visually, similar spatial patterns are observed in the three products. The data show that there are more cloud free AVHRR pixels, which is related to the coarser spatial resolution compared to MODIS and AATSR. The LST difference is $\leq$ 5 minutes, so there is little cloud displacement between observations.

CTP data are approximately normally distributed for all three sensors. Both COT and CER show positive kurtosis and skewness, as values close to 0 are common. CER data are somewhat bimodal, having a primary peak at $\sim$12 µm and a secondary peak at $\sim$35 µm (Figure 08 and Table 06). These peaks probably correspond to liquid and ice phase clouds, respectively. Mean value differences are not significant between AVHRR and MODIS for CTP, MODIS and AATSR for COT, and AVHRR and AATSR for CER. The standard deviation of differences between two sensors are always lowest for AVHRR minus MODIS (Table 06). Significance tests of mean differences and standard deviations of residuals between sensor retrievals are sensitive to outliers[..[44]]. Although cloud displacement due to observation time differences is probably small, we cannot discard its influence on such outliers. Even though we found no significant relationship between sensor retrieval residuals and observation time difference (not shown), residuals are likely to be smaller and thus possibly insignificant if sensor observation times were identical. Moreover, even modest relative radiometric calibration differences between sensors of a couple percent could cause large retrieval differences, particularly for COT.

4.2 Uncertainties

Median absolute uncertainties are CTP = 26.7 hPa, COT = 6.1, CER = 2.0 µm, and cloud mask = 13.7 % (Figure 07). The median relative retrieval uncertainty (not shown) is relatively low for [..[45]]CTP and CER, but considerably larger for COT (CTP = 4.7 %, COT = [..[46]]55.0 %, CER = [..[47]]13.6 %). COT uncertainties increase with COT magnitude, and [..[48]]largest uncertainties are found in cases of opaque cloud coverage and cloud over sea-ice surfaces (Figure 04 middle and Figure 07 topright). CER results are similar to COT, although relative uncertainties are somewhat lower. Cloud-free areas show increased cloud mask uncertainties, particularly over sea-ice surface areas. Note that the cloud mask uncertainties have been quantified as a function of the normalized difference to the cloud mask threshold, whereas relative retrieval uncertainties (100 $\times$ uncertainty $\div$ retrieved value) are shown for CTP, COT, and CER.

[41] removed: a validation
[42] removed: validation
[43] removed: Figures 03 to 05
[44] removed: , and are to some extent influenced by
[45] removed: all three retrieval variables
[46] removed: 6.1
[47] removed: 2.0
[48] removed: the RGB image (Figure 011) shows that the

4.3 [..[49]]

4.3 Comparison with CALIOP

We found collocations between CALIOP, AVHRR, MODIS, and AATSR for three study areas in the Arctic at 07/22/2008 19:15 LST (study area North America 1 = NA1, n = 120, Figure 09, RGB channels: red = Ch4 solar component, green = Ch2, blue = Ch1), 07/22/2008 20:58 LST (NA2, n = 163, Figure 011), and 07/27/2008 08:10 LST (Siberia = SIB, n = 116, Figure 013). These are located within 60° to 75° N latitude, and contain vegetated land, snow-covered land, open ocean, and sea-ice surfaces. For NA1 and SIB, all CALIOP pixels were classified as cloud covered, while for NA2 about half of the pixels are cloud free.

When including AATSR, collocations are restricted to high latitude areas and by the narrow swath of AATSR. We thus decided to include another scene without AATSR data in the Gulf of Guinea/West Africa between 7° S and 12° N at 24/10/2009 13:45 LST (Africa = AFR, n = 1181, Figure 015). There, about ten times more pixels are available than in the other scenes and cloud systems not contained in the Arctic data are observed, such as low-level stratocumulus and deep convection.

We divided all study areas into logical sectors, for each of which a characteristic pattern of cloud coverage and type predominates. The [..[50]]analysis is shown for comparisons of CTH [..[51]](derived from CTP) rather CTP (retrieved) to enable a more intuitive visualization and discussion. CTH is derived using the retrieval's atmospheric profile. An important caveat to note is the difference between physical and radiating cloud top. CALIOP uses an active sensor that is (roughly) sensitive to particle number. It identifies what we call the physical cloud top, denoted by the sharp increase in particle number. The passive radiometers analysed using CC4CL are (roughly) sensitive to the temperature of the cloud, from which the height is calculated. However, TOA radiation is [..[52]]product of emission and scattering processes throughout the atmospheric column [..[53]](Platnick, 2000). As there is no single height contributing, the retrieved CTT is more accurately described as an effective radiating pressure, being an average of the cloud's temperature profile weighted by the probability that a photon from each level can arrive at the detector. As a rule of thumb, the observed CTT represents the state one optical depth into the cloud. For the purposes of comparison to CALIOP, CC4CL computes a 'corrected' CTH, which adjusts the retrieved CTH to where the physical cloud top would be expected, assuming an adiabatic profile.

4.3.1 Case studies

Case study NA1

Study area NA1 is a completely cloud-covered scene over northern Canada containing clear and ice-covered land and open ocean surfaces (Figures 09 and 010). There are a variety of single and multi-layered clouds. CC4CL correctly classifies all

[49] removed: Validation with CALIOP
[50] removed: validation
[51] removed: rather than the retrievedvalue
[52] removed: the sum total
[53] removed: observed

pixels as cloud covered, with a few exceptions in sectors 3 and [..[54]]4 (Figures 09 and 010). CTH retrievals are consistent between the three sensors, only differing in sector 2. CTH is generally lower than CALIOP's top layer height, unless the latter is optically thick as in sector 4. In the case of a (semi-)transparent cloud top layer, multiple surfaces (several cloud layers, Earth surface) contribute to the observed satellite data. CC4CL CTH is then located closer to, at, or even below the underlying cloud layer (sectors 1, 3 and 2, respectively). For a single-layer, optically thick (COT > 1) cloud, CC4CL and CALIOP CTH agree very well (sector 4). Under such conditions, the retrieval is very accurate. Cloud phase agreement between CC4CL and CALIOP is very variable. It is best for optically thick high ice cloud coverage (sector 1), and worst for low water clouds (sector 4).

Case study NA2

Study area NA2 is located entirely over snow/ice free land in Western Canada. CALIOP cloud coverage is 4.5 %, spatially broken, and variable in height and phase. Clear-sky pixels are mostly identified by CC4CL (69.3 % correct), and cloudy pixels are occasionally missed (78.7 % correct). CC4CL retrievals of thin high clouds and false positive cloudy pixels have low CTH values (sector 1). Small-scale horizontal variability in CALIOP cloud phase is reflected by CC4CL data, which overestimate the fraction of liquid water clouds in sector 2. CC4CL reproduces CALIOP's spatial variability in CTH, which it slightly underestimates by 0.5–1 km in sector 2. In sector 3 CC4CL considerably underestimates CTH by up to 7 km. Most of these clouds are optically and geometrically thin.

Case study SIB

Study area SIB crosses the Novaya Zemlya islands north of Siberia and is defined by a mixture of open ocean and partially snow/ice covered land surfaces. According to both CALIOP and CC4CL, it is completely cloud covered.

In the event of single-layer cloudiness, CC4CL CTH agrees very well with CALIOP (sector 1 and, in particular, sector 3). The CTH difference between CC4CL retrievals increases in the presence of overlapping clouds (sector 2). There are optically thin but vertically thick ($\sim$4 km) clouds in sector 2. For these the retrieved CTH is considerably underestimated by $\sim$6 km, which is probably a result of lower layer contributions that "contaminate" the satellite signals. Overall, about 62.3 % of CC4CL pixels agree with CALIOP phase. Phase mismatch occurs in cases of single layer optically thin clouds (sector 2) and, less frequently, stratiform cloudiness (sector 3).

Case study AFR

Study area AFR is located over the Gulf of Guinea and Western Africa, containing open ocean and snow/ice free land surfaces. As we excluded AATSR data, about 10 times more pixel collocations with CALIOP are available (n = 1181) than for previous cases. Additionally, measurements contain tropical and coastal cloud systems such as extensive low-level stratiform cloudiness and continental deep convection (Figure 015).

[54] removed: 4.

In general, the quantitative and qualitative agreement between CC4CL and CALIOP CTH is impressive[..[55]]. CC4CL data track the spatial pattern of continental CTH very well (sector 3), which increases northwards and shows some small scale variability beyond 8° N. However, CC4CL underestimates CTH of vertically thick clouds and instead places the cloud top at a layer's vertical centre. The height of the stratiform cloud field is almost identical for CC4CL and CALIOP, although CTH of near-surface stratiform clouds is overestimated below thin high cirrus (sector 1). For the small layer located at 4 km height at $\sim$ 6.7° S and the thin high cirrus layer around 2° S, MODIS retrieval values differ somewhat from CC4CL AVHRR and CALIOP. Again, the phase of optically thick clouds is retrieved very well, which however is not the case for the thin ice cirrus clouds. Generally, CC4CL using the AVHRR heritage channel dataset are almost entirely insensitive to the very high, thin cloud layer in sector 1 (covering stratiform clouds) and 2 (covering the sea surface). Here, CC4CL is rather driven by contributions from very low clouds or the sea surface.

4.3.2 [..[56]]Summary of case studies

The four study areas clearly show that CC4CL retrievals of CTH are very close to CALIOP values for single layer, optically thick clouds. For significant extents of the regions presented, the CTH is accurate to within 240 m. For multi-layer clouds, CC4CL estimates are almost exclusively located in between CALIOP's top and bottom layer estimates. For these cases, the optimal estimation algorithm processes satellite signals that are likely to contain radiance contributions from multiple cloud layers. The OE then optimizes the fit between modelled and observed radiances by placing the cloud lower in the atmospheric profile, and so the mixed nature of the satellite data leads to an underestimation of CTH. The results shown here are a representative sample from an extensive validation performed within the Cloud_cci project [..[57]](Stengel et al., 2017).

There is no clear influence of the underlying land type or topography on retrieval values or the cloud mask. However, the limited sample size does not allow for generalizations. For site NA2, CALIOP identified cloud-free pixels, 69.3 % of which were also detected as cloud-free by CC4CL's neural network cloud mask, and with few exceptions as low level water clouds otherwise. In relatively few cases, CC4CL fails to detect clouds seen by CALIOP (% of missed clouds = 9.0 (NA1), 21.3 (NA2), 0.6 (SIB), 3.1 (AFR)). We did not account for fractional cloud coverage, as we set a grid box as cloud covered if any corresponding CC4CL pixel contains cloud information. As a consequence, there are slightly more cloud covered pixels for the spatially higher resolved MODIS and AATSR data than for AVHRR.

The CC4CL phase identification [..[58]]does not agree with any of the three CALIOP cloud flags [..[59]]consistently, which is [..[60]]reasonable given the differences between active and passive observations. After rounding CC4CL values to the nearest integer, the percentage of pixels with equal phase is lowest for the top layer at COD > 0 (NA1 = 49.9 %, NA2 = 32.2 %, SIB = 62.3 %, AFR = 53.0 %), but similar for the mid layer COD > 0.15 (NA1 = 53.1 %, NA2 = 46.5 %, SIB = 66.4 %, AFR =

[55] removed: , compared to the performance of existing algorithms

[56] removed: Validation summary

[57] removed: ..

[58] removed: agrees continually with none

[59] removed: shown

[60] removed: sensible

94.7 %), and the bottom layer COD > 1 (NA1 = 47.3 %, NA2 = 57.4 %, SIB = 66.3 %, AFR = 91.1 %). These values however do show that phase determination performs very well if optically thick clouds dominate, as is the case for study area AFR. When averaged over all layers, phase agreement is largest for site AFR (79.6 %), followed by SIB (65.0 %), and clearly lower for NA1 (50.1 %) and NA2 (45.4 %).

For ice clouds, the most frequently occurring cloud types are cirrus (ID=6) for CALIOP and overlap (ID=8) or cirrus (ID=7) for CC4CL. Water cloud types are more heterogeneous and for CALIOP predominantly low transparent (ID=0), but altostratus (ID=5) and altocumulus (ID=4) are also frequent. CC4CL water clouds are approximately equally distributed amongst water (ID=3) and supercooled (ID=4) cloud types.

The scenes investigated and discussed here are just a small subset of the large variety of global cloudiness. We could only find collocations with AATSR data at high latitudes, where multi-layer cloud coverage is common. Under such difficult conditions, the retrieved CTH is a mixture of all radiatively contributing cloud layers. Please refer to Stengel et al. (2017) for a quantitative, global validation of CC4CL cloud properties.

5 Discussion

5.1 The flexibility of the optimal estimation approach

In general, the retrieved values are insensitive to the specific instrument evaluated[..[61]]. Absolute mean differences are $\leq$ 21.9 hPa for CTP, $\leq$ 1.3 for COT, and $\leq$ 2.1 for CER. These are mostly smaller than the mean retrieval uncertainties themselves. Moreover, the RGB images show that all major patterns of cloud coverage and structure are resolved by all three sensors. However, AATSR data show larger deviations than the other sensors (Figure 08). It is unlikely that differences in spectral response functions are the reason. MODIS and AATSR heritage channels are relatively close in their spectral response but their retrieval values do differ considerably. Also, even though spectral response differences are largest between MODIS and AVHRR [..[62]](which results in a reflectance difference of up to 30–40 [..[63]]% (Trishchenko et al., 2002), [..[64]]their retrieval values are much more similar. The difference [..[65]]between AATSR and both AVHRR and MODIS is largest for CER, so microphysical variables, which are derived from reflectance data only, appear to be most affected."

The differences between mean values ($\mu1$ and $\mu2$) are almost always significant (t-Test p-value < 0.1, H_0: $\mu1 = \mu2$). Thus, from a statistical point of view, the samples we analysed for AVHRR, MODIS, and AATSR have been drawn from different populations and are thus statistically inconsistent. In other words, the retrieval system should not produce statistically consistent cloud parameters when driven with spatiotemporally collocated satellite data obtained from three [..[66]]different sensors. However, differences in cloud conditions at the various observation times and sensor spatial resolution explain part of

[61] removed: , such that the merged data set is sensible.

[62] removed: disagree nonetheless more in their spectral response,

[63] removed:

[64] removed: but

[65] removed: to

[66] removed: spatiotemporally collocated

these discrepancies. Moreover, a non-significant t-Test result is possibly too strict a metric for estimating the consistency of retrieval results. There is a range of confounding processes that affect each individual retrieval estimate, such as observation times, spectral responses, calibration deficiencies, and a varying number of cloudy pixels to be compared. 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. The case studies clearly show that, under optimal conditions for single layer cloud retrievals, CC4CL products are consistent with CALIOP and practically insensitive to sensor characteristics.

We suggest that AVHRR and MODIS data can be used interchangeably, depending on the user's application, such as model validation, data assimilation applications, or climate studies in general (Liu et al., 2017; Yang et al., 2016). AVHRR data provide long-term data records from 1982, however at a relatively coarse resolution of 5 km × 3 km. The MODIS data record started in 2000, and is thus too short to be considered a climatology. However, L1 data are available at 1 km resolution, and orbit control is guaranteed. With CC4CL, we also produced 0.05° lat/lon daily composites for Europe (data not shown), which is close to MODIS's original resolution in that area. These data provide a more detailed view on cloud features than AVHRR. In that sense, CC4CL products retrieved from AVHRR and MODIS are complementary. More detailed analysis is required to assess differences in CC4CL output data when applied to AATSR data.

5.2 The value of uncertainty quantification

The retrieval uncertainties prove to be a valuable source of information. 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). On the other hand, they allow for diagnosis of potential retrieval shortcomings. For example, we see that COT uncertainty scales with COT itself and is thus heteroscedastic (see also Poulsen et al. (2012)). CC4CL COT values are at times unnaturally large, and the associated uncertainty reflects that. Also, it highlights under which conditions the optimal estimator converges to a solution with a relatively large divergence from the measurements[..[67]], which here are associated with optically thick clouds or underlying snow/ice cover (see also Kahn et al. (2015); Wang et al. (2011)). COT and CER uncertainties are clearly largest, and reflect the limited information available with which to retrieve these values. For further possible explanations due to assumptions and limitations within the methodology applied, please see part II.

We applied an independent approach to quantify cloud mask uncertainty. It is valuable information, as a neural network does not provide output uncertainty. The approach we adopted here is straightforward. When the NN output, which is a pseudo CALIOP COD, approaches a defined threshold value for cloudiness, the uncertainty increases towards a maximum of 50 %. This maximum value indicates that a cloud mask value is basically random, as it is equally likely to be cloudy or cloud-free. With that in mind, the cloud mask uncertainty data are easy to interpret. For example, we see that sea-ice pixels classified as cloudy to the North of study area NA2 (Figure 07) show uncertainties of 40 - 50 %. This indicates that the NN is sensitive to bright ground cover, which may be confused with clouds. We suggest that users of ESA Cloud_cci data should routinely

[67] removed: . In the cases shown here, large uncertainties

consult cloud mask uncertainties. If a more conservative cloud mask is required, it can be easily built by setting a maximum value for an acceptable uncertainty level.

5.3 Strengths and weaknesses

The results clearly show that CC4CL retrieves CTH of single layer, optically thick clouds with high accuracy and precision. When compared to CALIOP, the mean deviation for these cases is as low as 10–240 m CTH. This is a promising result, and shows that the optimal estimation framework is robust and appropriate for retrieving cloud properties from AVHRR heritage channels.

In the case of multi-layer clouds, CC4CL is not able to retrieve CTH of the top layer if it is optically thin. The estimate is somewhere between the two cloud pressure values. This is an expected limitation of our framework, and also of other retrieval algorithms using passive sensor data (Holz et al., 2008; Karlsson and Dybbroe, 2010). Poulsen et al. (2012) found that ORAC CTP and CER estimates are robust when the top ice cloud layer is >5 optical depths, and otherwise they are the weighted average of several cloud layers. The AVHRR heritage channels do not provide sufficient information on retrieving cloud vertical structure. In the case of semi-transparent top-layer clouds, the upwelling signal, [..[68]]whether it stems from a cloud or the Earth's surface, contributes to the total TOA reflectance or brightness temperature. This mixed satellite measurement is input to CC4CL, which retrieves cloud parameters assuming a single cloud layer. As brightness temperatures and reflectances of, e.g., a cold, semi-transparent, bright top-layer cirrus cloud overlapping a warmer, opaque, and darker low-level water cloud, are a mix of several contributing surfaces, so will be the final retrieval value. Any CTH retrieved from AVHRR (heritage) data is the radiatively effective rather than physical cloud top (Karlsson et al., 2013). For CC4CL, we often see that the final CTH estimate is placed between top and lower levels and is thus an underestimate, which is a common problem amongst retrieval algorithms using passive sensor data (Watts et al., 2011; Holz et al., 2008; Karlsson et al., 2013).

Multi-layer cloud property retrievals have been developed (Watts et al., 2011), and we also implemented and tested such an approach within CC4CL (McGarragh et al., 2017b). However, this method requires MODIS channels beyond the AVHRR heritage set, and thus will not be applicable to a full AVHRR reprocessing. For ESA Cloud_cci, [..[69]]a conscious decision was made to [..[70]]trade spectral information for time series continuity. Thus, discontinuities due to changing spectral coverage within the entire dataset are avoided (Stengel et al., 2017). In addition, we introduced corrected estimates of CTP and the derived CTT and CTH to get closer to the physical or geometrical cloud top. The correction is based on a vertical displacement of CTP along the atmospheric profile based on optical thickness and the cloud's extinction coefficient, which is a function of CER (McGarragh et al., 2017a). The correction is only made for ice phase clouds.

[..[71]]At first glance, estimates of cloud phase appear reasonable when compared to CALIOP. However, we find the best overall agreement of ~ 65 % for the lower layers (cumulative COD >0.15 or >1). This is just slightly better than a random guess of cloud phase. Cloud phase is generally difficult to quantify, and estimates of various satellite derived products disagree

[68] removed: may it stem
[69] removed: the
[70] removed: deliberately
[71] removed: On a first view

considerably for that variable (Stengel et al., 2015). An evaluation of MODIS Collection 6 cloud phase yielded a total cloud phase agreement of over 90 % with CALIOP. However, as the study is exclusively based on single-phase cloudy pixels, the performance of MODIS C6 as applied to multi-phase pixels is still unknown (Marchant et al., 2016). We also find very high scores for cloud phase determination if restricting the analysis to optically thick, spatially extensive cloud fields such as in study site AFR. There, cloud phase agrees with lower layer CALIOP estimates by as much as 95 %.

The key problems for phase determination are vertical stratification and the lack of direct in-situ measurements of cloud phase. CALIOP observations, and also DARDAR (radar lidar, Ceccaldi et al. (2013)), are currently considered to be the most advanced estimates of cloud phase, relying on active measurement principles with depolarization and total attenuated backscatter at multiple wavelengths for additional constraints (Winker et al., 2009; Karlsson and Dybbroe, 2010). However, this assumption is primarily based on the physical theory underlying their retrievals, rather than on a comprehensive validation with independent observations of cloud phase.

Within CC4CL, we apply the Pavolonis algorithm for phase detection (Pavolonis et al., 2005). It was designed using simulated radiance data for varying phase, and further adjusted after analysis with real satellite data. The algorithm itself is a decision tree that contains a set of fixed threshold values for input reflectances and brightness temperatures, and was tuned to AVHRR. Even though we expect differences in phase determination between AVHRR vs. MODIS and AATSR due to varying spectral response functions, these were not large for the three study sites. Pavolonis et al. (2005) state that their product could not be validated due to the lack of direct observations, but rather underwent a consistency check with ground-based, independent estimates.

The relatively low degree of agreement between CC4CL and CALIOP is not satisfying if CALIOP is considered to be the truth. However, we refrain from concluding that the CC4CL phase estimate was unrealistic as, to date, no robust, spatially resolved in-situ observations are available and our comparisons included multi-layered cloud conditions. It is difficult to determine the representative CALIOP cloud layer when validating a passive sensor retrieval. For single layer, optically thick clouds, CC4CL can be compared with any layer exceeding a cumulative optical thickness of 0 or 1. If such a cloud layer was covered by optically and geometrically thin cirrus clouds, the satellite data are still dominated by lower cloud level reflectance and, in particular, emittance. Consequently, Pavolonis cloud phase is not a top layer estimate in such cases. For study area AFR, we also found situations where the NN cloud mask, which was trained with CALIOP data, correctly identifies thin high cirrus as cloud over ocean but the cloud type algorithm failed to identify its ice phase. One potential improvement would be to use the NN to provide an estimate of cloud phase. Initial tests indicated that this approach would indeed improve the (global) agreement with CALIOP, which is to be expected, as the NN is trained with CALIOP data. However, no estimate of cloud type would be provided. It would also be worth investigating the relationship between the quality of retrieved variables (CTH, COT, CER, cloud phase) and cloud mask uncertainty.

CALIOP data are considered to be the current benchmark of cloud detection, vertical structure, and phase (Winker et al., 2009; Karlsson and Johansson, 2013; Holz et al., 2008), and are – except for the cloud mask – a source of validation with absolute independence from CC4CL. The main limitation of CALIOP though is its narrow view, so that global coverage is very limited. Also, the instrument is only able to probe the full geometrical depth of clouds whose total optical thickness is not larger than about 3–5 (Karlsson and Johansson, 2013). We found no clear relationship between CC4CL CTP uncertainty and the difference between CC4CL CTP and CALIOP CTP (data not shown). This suggests that the AVHRR heritage channels provide independent information on cloud vertical structure that is not clearly related to CALIOP's CTP estimates. Retrieval uncertainty is [..[72]]estimated using only well-understood error sources (e.g. measurement and forward model error[..[73]]), neglecting errors due to model assumptions (e.g. the complex, real vertical [..[74]]structure). Such errors can be approximated through validation activities and are not currently believed to be significant in most circumstances.

6 Conclusions

We have shown that CC4CL is a robust and flexible framework for producing cloud products from passive satellite sensor data. Differences between retrieved values for collocated satellite data are smaller than estimated uncertainties for AVHRR, MODIS, and AATSR. ESA Cloud_cci data provide climatologies (AVHRR) as well as highly resolved snap-shots for selected regions (e.g. Europe, MODIS). The complete sensor set of CC4CL data forms a unique, coherent, long-term, multi-instrument cloud property product that exploits [..[75]]synergistic capabilities of several EO missions. Compared to single sensor retrievals, CC4CL data are improved in terms of accuracy and spatiotemporal sampling.

CC4CL explicitly estimates retrieval uncertainties according to the principles of error propagation through optimal estimation theory. These uncertainties are a valuable source for model validation, data assimilation, climate studies, or retrieval diagnosis. Cloud mask uncertainty is a novel feature that enables the user to assess product quality and to create individualized cloud masks.

We find that CC4CL is limited by weaknesses that are common to passive sensor cloud product retrievals. In general, an initial [..[76]]comparison against CALIOP data shows that the CTH of optically thin clouds is underestimated. In the case of multi-layer clouds, the retrieved CTH is a mixture of all radiatively contributing cloud layers. The AVHRR heritage channels do not provide sufficient physical information that would allow for detailed retrievals of cloud vertical structure. Moreover, the forward cloud model is structurally incomplete, as it assumes a single-plane cloud layer. A multi-layer cloud property retrieval has been added to CC4CL, but is only applicable to MODIS data.

To account for CTH underestimation, we implemented a correction for CTH that assumes that passive sensor data see beyond the top into the clouds up to a penetration depth of ~ 1 optical depth. Corrected cloud top values are stored as separate variables within CC4CL output files.

Similarly, we find that the cloud phase estimate is only accurate for optimal retrieval conditions (optically thick top clouds). In a subsequent reprocessing of the AVHRR data record, we replaced the Pavolonis et al. (2005) algorithm with a neural network cloud phase estimation with better performance scores.

[72]removed: driven by background
[73]removed: as well as the mixed-signal satellite radiances rather than by the
[74]removed: cloud structure.
[75]removed: synergic
[76]removed: validation with

Under optimal conditions for single layer cloud retrievals, CC4CL products show little sensitivity to sensor characteristics. Single layer, optically thick cloud retrievals are very accurate [..[77]]when compared against CALIOP (bias $<$ 240 m), which emphasizes the maturity and robustness of CC4CL. We thus recommend ESA Cloud_cci data to be used for multi-annual studies of cloud parameters and more detailed assessments of regional patterns and diurnal variability.

Acknowledgements. NASA kindly provided MODIS Collection 6 radiance data. ECMWF kindly provided support and assistance with their computing facilities for development and processing. This research/work was supported by the European Space Agency through the Cloud_cci project (contract No.: 4000109870/13/I-NB).

[77] removed: and precise

References

Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K., and Yang, P.: MODIS Cloud-Top Property Refinements for Collection 6, Journal of Applied Meteorology and Climatology, 51, 1145–1163, https://doi.org/10.1175/JAMC-D-11-0203.1, 2012.

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kållberg, P., Kobayashi, S., Uppala, S., and Simmons, A.: The ERA-Interim archive Version 2.0, Shinfield Park, Reading, 2011.

CDO: Climate Data Operators, http://www.mpimet.mpg.de/cdo, 2015.

Ceccaldi, M., Delanoë, J., Hogan, R. J., Pounder, N. L., Protat, A., and Pelon, J.: From CloudSat-CALIPSO to EarthCare: Evolution of the DARDAR cloud classification and its comparison to airborne radar-lidar observations, Journal of Geophysical Research: Atmospheres, 118, 7962–7981, https://doi.org/10.1002/jgrd.50579, 2013.

Christensen, M. W., Stephens, G. L., and Lebsock, M. D.: Exposing biases in retrieved low cloud properties from CloudSat: A guide for evaluating observations and climate data, Journal of Geophysical Research: Atmospheres, 118, 12,120–12,131, https://doi.org/10.1002/2013JD020224, 2013.

CM SAF: The Satellite Application Facility on Climate Monitoring, http://www.cmsaf.eu, accessed: Oct 13, 2015.

Comiso, J. C. and Hall, D. K.: Climate trends in the Arctic as observed from space, Wiley Interdisciplinary Reviews: Climate Change, 5, 389–409, https://doi.org/10.1002/wcc.277, 2014.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Devasthale, A., Raspaud, M., and Schlundt, C.: PyGAC: an open-source, community-driven Python interface to preprocess more than 30-year AVHRR Global Area Coverage (GAC) data, in preparation, 2017.

ESA CCI: The ESA Climate Change Initiative, http://cci.esa.int/, accessed: Oct 13, 2015.

Hamann, U., Walther, A., Baum, B., Bennartz, R., Bugliaro, L., Derrien, M., Francis, P. N., Heidinger, A., Joro, S., Kniffka, A., Le Gléau, H., Lockhoff, M., Lutz, H.-J., Meirink, J. F., Minnis, P., Palikonda, R., Roebeling, R., Thoss, A., Platnick, S., Watts, P., and Wind, G.: Remote sensing of cloud top pressure/height from SEVIRI: analysis of ten current retrieval algorithms, Atmospheric Measurement Techniques, 7, 2839–2867, https://doi.org/10.5194/amt-7-2839-2014, 2014.

Han, Q., Rossow, W. B., and Lacis, A. A.: Near-Global Survey of Effective Droplet Radii in Liquid Water Clouds Using ISCCP Data, Journal of Climate, 7, 465–497, https://doi.org/10.1175/1520-0442(1994)007<0465:NGSOED>2.0.CO;2, 1994.

Heidinger, A. K. and Pavolonis, M. J.: Gazing at Cirrus Clouds for 25 Years through a Split Window. Part I: Methodology, Journal of Applied Meteorology and Climatology, 48, 1100, https://doi.org/10.1175/2008JAMC1882.1, 2009.

Heidinger, A. K., Evan, A. T., Foster, M. J., and Walther, A.: A Naive Bayesian Cloud-Detection Scheme Derived from CALIPSO and Applied within PATMOS-x, Journal of Applied Meteorology and Climatology, 51, 1129–1144, https://doi.org/10.1175/JAMC-D-11-02.1, 2012.

Heidinger, A. K., Foster, M. J., Walther, A., and Zhao, X.: The Pathfinder Atmospheres–Extended AVHRR Climate Dataset, Bulletin of the American Meteorological Society, 95, 909–922, https://doi.org/10.1175/BAMS-D-12-00246.1, 2013.

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van Roozendael, M., and Wagner, W.: The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables, Bulletin of the American Meteorological Society, 94, 1541–1552, https://doi.org/10.1175/BAMS-D-11-00254.1, 2013.

Holz, R. E., Ackerman, S. A., Nagle, F. W., Frey, R., Dutcher, S., Kuehn, R. E., Vaughan, M. A., and Baum, B.: Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIOP, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD009837, 2008.

IPCC: Summary for Policymakers, book section SPM, p. 1–30, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.004, www.climatechange2013.org, 2013.

Jacobowitz, H., Stowe, L. L., Ohring, G., Heidinger, A., Knapp, K., and Nalli, N. R.: The Advanced Very High Resolution Radiometer Pathfinder Atmosphere (PATMOS) Climate Dataset: A Resource for Climate Research, Bulletin of the American Meteorological Society, 84, 785–793, https://doi.org/10.1175/BAMS-84-6-785, 2003.

Kahn, B. H., Schreier, M. M., Yue, Q., Fetzer, E. J., Irion, F. W., Platnick, S., Wang, C., Nasiri, S. L., and L'Ecuyer, T. S.: Pixel-scale assessment and uncertainty analysis of AIRS and MODIS ice cloud optical thickness and effective radius, Journal of Geophysical Research: Atmospheres, 120, 11,669–11,689, https://doi.org/10.1002/2015JD023950, 2015.

Karlsson, K.-G. and Dybbroe, A.: Evaluation of Arctic cloud products from the EUMETSAT Climate Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations, Atmospheric Chemistry and Physics, 10, 1789–1807, https://doi.org/10.5194/acp-10-1789-2010, 2010.

Karlsson, K.-G. and Johansson, E.: On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO-CALIOP data: example investigating the CM SAF CLARA-A1 dataset, Atmospheric Measurement Techniques, 6, 1271–1286, https://doi.org/10.5194/amt-6-1271-2013, 2013.

Karlsson, K.-G., Riihelä, A., Müller, R., Meirink, J. F., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann, J., Kaspar, F., Hollmann, R., and Wolters, E.: CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data, Atmospheric Chemistry and Physics, 13, 5351–5367, https://doi.org/10.5194/acp-13-5351-2013, 2013.

Karlsson, K.-G., Sedlar, J., Devasthale, A., Stengel, M., Hanschmann, T., Meirink, J. F., Benas, N., and van Zadelhoff, G.-J.: Validation Report – CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2) – Cloud Products, Tech. Rep. 2.3, EUMETSAT Satellite Application Facility on Climate Monitoring, https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002, 2016.

Kiehl, J. and Trenberth, K. E.: Earth's annual global mean energy budget, Bulletin of the American Meteorological Society, 78, 197–208, 1997.

King, M., Kaufman, Y., Menzel, W., and Tanre, D.: Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS), Geoscience and Remote Sensing, IEEE Transactions on, 30, 2–27, https://doi.org/10.1109/36.124212, 1992.

Laszlo, I., Stamnes, K., Wiscombe, W. J., and Tsay, S.-C.: The Discrete Ordinate Algorithm, DISORT for Radiative Transfer, pp. 3–65, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-49538-4_1, 2016.

Liu, C., Allan, R. P., Mayer, M., Hyder, P., Loeb, N. G., Roberts, C. D., Valdivieso, M., Edwards, J. M., and Vidale, P.: Evaluation of satellite and reanalysis-based global net surface energy flux and uncertainty estimates, Journal of Geophysical Research: Atmospheres, 122, 6250–6272, https://doi.org/10.1002/2017JD026616, 2017.

Lucht, W.: Expected retrieval accuracies of bidirectional reflectance and albedo from EOS-MODIS and MISR angular sampling, Journal of Geophysical Research: Atmospheres, 103, 8763–8778, https://doi.org/10.1029/98JD00089, 1998.

Marchant, B., Platnick, S., Meyer, K., Arnold, G. T., and Riedi, J.: MODIS Collection 6 shortwave-derived cloud phase classification algorithm and comparisons with CALIOP, Atmospheric Measurement Techniques, 9, 1587–1599, https://doi.org/10.5194/amt-9-1587-2016, 2016.

McGarragh, G., Poulsen, C., Christensen, M., Sus, O., Povey, A., Proud, S., and Grainger, R.: On the retrieval of cloud top pressure of semi-transparent clouds: Methods and radiative consistency, To be submitted to: Journal of the Atmospheric Sciences, 2017a.

McGarragh, G., Poulsen, C., Siddans, R., Povey, A., Proud, S., Thomas, G., and Grainger, R.: Optimal Retrieval of Aerosol and Cloud (ORAC): the multilayer cloud retrieval, To be submitted to: Atmospheric Measurement Techniques, 2017b.

McGarragh, G. R., Poulsen, C. A., Thomas, G. E., Povey, A. C., Sus, O., Schlundt, C., Stapelberg, S., Proud, S., Christensen, M. W., Stengel, M., Hollmann, R., and Grainger, R. G.: The Community Cloud retrieval for CLimate (CC4CL). Part II: The optimal estimation approach, To be submitted to: Atmospheric Measurement Techniques, 2017c.

Menzel, W. P., Frey, R. A., and Baum, B. A.: Cloud top properties and cloud phase algorithm theoretical basis document, 2010.

Merchant, C. J., Embury, O., Rayner, N. A., Berry, D. I., Corlett, G. K., Lean, K., Veal, K. L., Kent, E. C., Llewellyn-Jones, D. T., Remedios, J. J., and Saunders, R.: A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers, Journal of Geophysical Research: Oceans, 117, https://doi.org/10.1029/2012JC008400, c12013, 2012.

Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., Spangenberg, D. A., Arduini, R. F., Trepte, Q. Z., Smith, W. L., Ayers, J. K., Gibson, S. C., Miller, W. F., Hong, G., Chakrapani, V., Takano, Y., Liou, K. N., Xie, Y., and Yang, P.: CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data – Part I: Algorithms, IEEE Transactions on Geoscience and Remote Sensing, 49, 4374–4400, https://doi.org/10.1109/TGRS.2011.2144601, 2011.

MODIS Characterization Support Team: MODIS Level 1B Product User's Guide, NASA Goddard Space Flight Center, 2009.

Nakajima, T. and King, M. D.: Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory, Journal of the Atmospheric Sciences, 47, 1878–1893, https://doi.org/10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2, 1990.

NASA LP DAAC: MOD02 Level-1B Calibrated Geolocation Data Set. Version 6. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=02, https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD021KM.006, accessed: May 25, 2015.

Natural Environment Research Council, NERC CWVC GRAPE campaign participants, and Grainger, D.: Global Cloud and Aerosol Dataset Produced by the Global Retrieval of ATSR Cloud Parameters and Evaluation (GRAPE) Project as part of the Clouds, Water Vapour and Climate (CWVC) Programme. NCAS British Atmospheric Data Centre, http://catalogue.ceda.ac.uk/uuid/cf7ae349d16c067cd00d3d2d910bee89, accessed: Oct 13, 2015.

Norris, J. R., Allen, R. J., Evan, A. T., Zelinka, M. D., O'Dell, C. W., and Klein, S. A.: Evidence for climate change in the satellite cloud record, Nature, 536, 72–75, https://doi.org/10.1038/nature18273, 2016.

Pavolonis, M. J., Heidinger, A. K., and Uttal, T.: Daytime Global Cloud Typing from AVHRR and VIIRS: Algorithm Description, Validation, and Comparisons, Journal of Applied Meteorology, 44, 804–826, 2005.

Platnick, S.: Vertical photon transport in cloud remote sensing problems, Journal of Geophysical Research: Atmospheres, 105, 22 919–22 935, https://doi.org/10.1029/2000JD900333, 2000.

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W. L., and Riedi, J.: The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua, IEEE Transactions on Geoscience and Remote Sensing, 55, 502–525, https://doi.org/10.1109/TGRS.2016.2610522, 2017.

Poulsen, C. A., Siddans, R., Thomas, G. E., Sayer, A. M., Grainger, R. G., Campmany, E., Dean, S. M., Arnold, C., and Watts, P. D.: Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR, Atmospheric Measurement Techniques, 5, 1889–1910, https://doi.org/10.5194/amt-5-1889-2012, 2012.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World Scientific, Singapore, 2009.

Rossow, W. B. and Schiffer, R. A.: Advances in understanding clouds from ISCCP, Bulletin of the American Meteorological Society, 80, 2261–2287, 1999.

Sanchez-Lorenzo, A., Enriquez-Alonso, A., Calbó, J., González, J.-A., Wild, M., Folini, D., Norris, J. R., and Vicente-Serrano, S. M.: Fewer clouds in the Mediterranean: consistency of observations and climate simulations, Scientific Reports, 7, https://doi.org/10.1038/srep41475, 2017.

Sayer, A. M., Thomas, G. E., and Grainger, R. G.: A sea surface reflectance model for (A)ATSR, and application to aerosol retrievals, Atmospheric Measurement Techniques, 3, 813–838, https://doi.org/10.5194/amt-3-813-2010, 2010.

Schaaf, C. and Wang, Z.: MCD43C1 MODIS/Terra+Aqua BRDF/AlbedoModel Parameters Daily L3 Global 0.05Deg CMG V006, NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MCD43C1.006, 2015.

Schulz, J., Albert, P., Behr, H.-D., Caprion, D., Deneke, H., Dewitte, S., Dürr, B., Fuchs, P., Gratzki, A., Hechler, P., Hollmann, R., Johnston, S., Karlsson, K.-G., Manninen, T., Müller, R., Reuter, M., Riihelä, A., Roebeling, R., Selbach, N., Tetzlaff, A., Thomas, W., Werscheck, M., Wolters, E., and Zelenka, A.: Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF), Atmospheric Chemistry and Physics, 9, 1687–1709, https://doi.org/10.5194/acp-9-1687-2009, 2009.

Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L.: Development of a Global Infrared Land Surface Emissivity Database for Application to Clear Sky Sounding Retrievals from Multispectral Satellite Radiance Measurements, Journal of Applied Meteorology and Climatology, 47, 108–123, https://doi.org/10.1175/2007JAMC1590.1, 2008.

Stengel, M., Mieruch, S., Jerg, M., Karlsson, K.-G., Scheirer, R., Maddux, B., Meirink, J., Poulsen, C., Siddans, R., Walther, A., and Hollmann, R.: The Clouds Climate Change Initiative: Assessment of state-of-the-art cloud property retrieval schemes applied to AVHRR heritage measurements, Remote Sensing of Environment, 162, 363–379, https://doi.org/10.1016/j.rse.2013.10.035, 2015.

Stengel, M., Stapelberg, S., Sus, O., Schlundt, C., Poulsen, C., Thomas, G., Christensen, M., Carbajal Henken, C., Preusker, R., Fischer, J., Devasthale, A., Willén, U., Karlsson, K.-G., McGarragh, G. R., Proud, S., Povey, A. C., Grainger, D. G., Meirink, J. F., Feofilov, A., Bennartz, R., Bojanowski, J., and Hollmann, R.: Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the Cloud_cci project, Earth System Science Data Discussions, 2017, 1–34, https://doi.org/10.5194/essd-2017-48, 2017.

Stengel, M., Stapelberg, S., Schlundt, C., Karlsson, K.-G., Meirink, J. F., Poulsen, C., Bojanowski, J., Stöckli, R., and Hollmann, R.: ESA Cloud CCI - Product Validation and Intercomparison Report, Tech. Rep. 5.1, ESA Climate Change Initiative - Cloud CCI, 2018.

Stephens, G. L., Li, J., Wild, M., Clayson, C. A., Loeb, N., Kato, S., L'ecuyer, T., Stackhouse Jr, P. W., Lebsock, M., and Andrews, T.: An update on Earth's energy balance in light of the latest global observations, Nature Geoscience, 5, 691, 2012.

Sun, B., Free, M., Yoo, H. L., Foster, M. J., Heidinger, A., and Karlsson, K.-G.: Variability and Trends in U.S. Cloud Cover: ISCCP, PATMOS-x, and CLARA-A1 Compared to Homogeneity-Adjusted Weather Observations, Journal of Climate, 28, 4373–4389, https://doi.org/10.1175/JCLI-D-14-00805.1, 2015.

Thomas, G. E., Carboni, E., Sayer, A. M., Poulsen, C. A., Siddans, R., and Grainger, R. G.: Oxford-RAL Aerosol and Cloud (ORAC): aerosol retrievals from satellite radiometers, in: Satellite Aerosol Remote Sensing over Land, edited by Kokhanovsky, A. and de Leeuw, G., Springer Praxis Books, pp. 193–225, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-69397-0_7, 2009a.

Thomas, G. E., Poulsen, C. A., Sayer, A. M., Marsh, S. H., Dean, S. M., Carboni, E., Siddans, R., Grainger, R. G., and Lawrence, B. N.: The GRAPE aerosol retrieval algorithm, Atmospheric Measurement Techniques, 2, 679–701, https://doi.org/10.5194/amt-2-679-2009, 2009b.

Trishchenko, A. P., Cihlar, J., and Li, Z.: Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors, Remote Sensing of Environment, 81, 1–18, https://doi.org/10.1016/S0034-4257(01)00328-5, 2002.

U.S. Geological Survey: The ESA Climate Change Initiative, https://lta.cr.usgs.gov/glcc/globe_int/, accessed: Sep 2, 2016.

Walther, A. and Heidinger, A. K.: Implementation of the Daytime Cloud Optical and Microphysical Properties Algorithm (DCOMP) in PATMOS-x, Journal of Applied Meteorology and Climatology, 51, 1371–1390, https://doi.org/10.1175/JAMC-D-11-0108.1, 2012.

Wang, C., Yang, P., Baum, B. A., Platnick, S., Heidinger, A. K., Hu, Y., and Holz, R. E.: Retrieval of Ice Cloud Optical Thickness and Effective Particle Size Using a Fast Infrared Radiative Transfer Model, Journal of Applied Meteorology and Climatology, 50, 2283–2297, https://doi.org/10.1175/JAMC-D-11-067.1, 2011.

Watts, P. D., Bennartz, R., and Fell, F.: Retrieval of two-layer cloud properties from multispectral observations using optimal estimation, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2011JD015883, 2011.

Willett, K., Alexander, L., and Thorne, P.: Global climate, in: State of the Climate in 2009, edited by Arndt, D. S., Baringer, M. O., and Johnson, M. R., vol. 91, pp. 19–52, https://doi.org/10.1175/BAMS-91-7-StateoftheClimate, 2010.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO mission and CALIOP data processing algorithms, Journal of Atmospheric and Oceanic Technology, 26, 2310–2323, https://doi.org/10.1175/2009JTECHA1281.1, 2009.

Xiong, X., Wenny, B. N., and Barnes, W. L.: Overview of NASA Earth Observing Systems Terra and Aqua moderate resolution imaging spectroradiometer instrument calibration algorithms and on-orbit performance, Journal of Applied Remote Sensing, 3, 032 501–032 501–25, https://doi.org/10.1117/1.3180864, 2009.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R.: The role of satellite remote sensing in climate change studies, Nature Clim. Change, 3, 875–883, 2013.

Yang, Q., Losch, M., Losa, S. N., Jung, T., Nerger, L., and Lavergne, T.: 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, 10, 761–774, https://doi.org/10.5194/tc-10-761-2016, 2016.

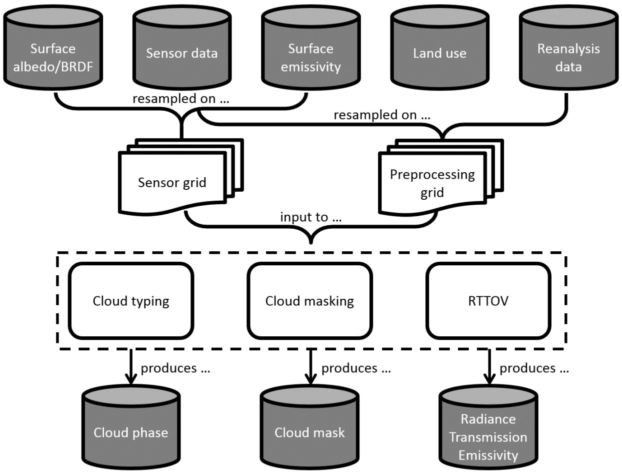


Figure 01. Schematic of the CC4CL preprocessor.

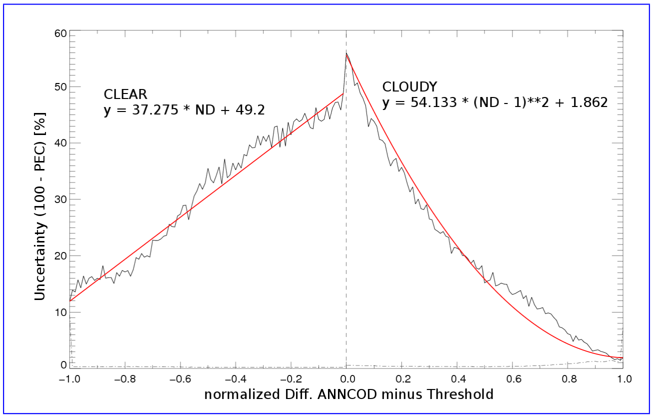


Figure 02. Neural network cloud mask uncertainty as derived from observations.

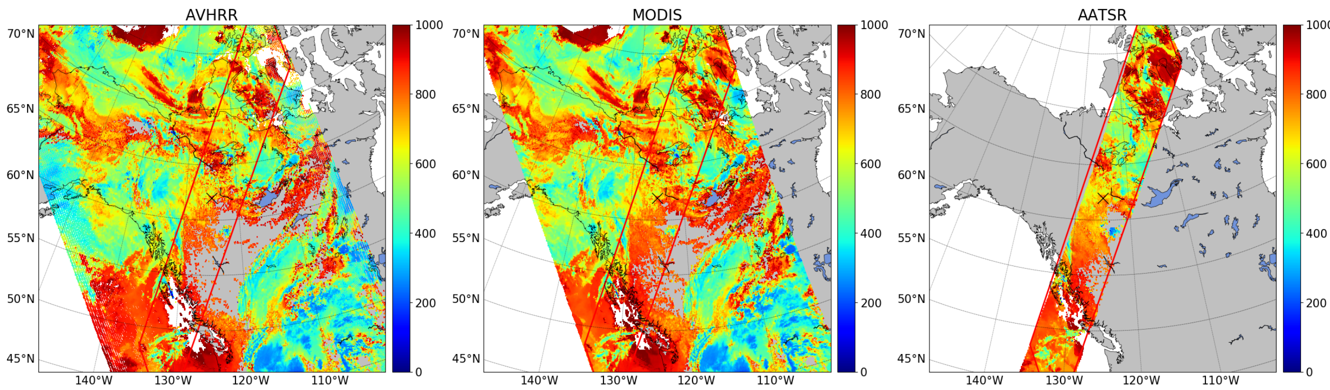


Figure 03. CTP retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).

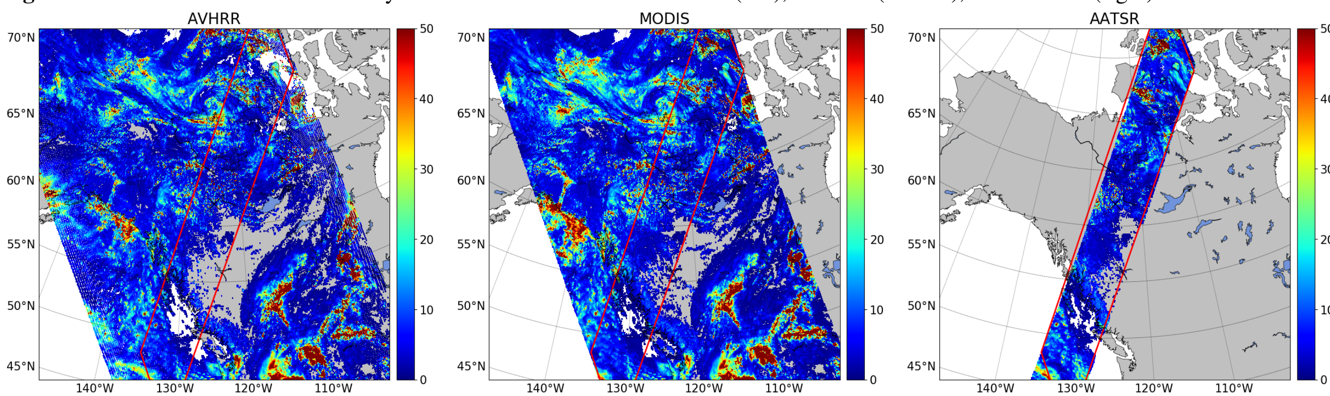


Figure 04. COT retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).

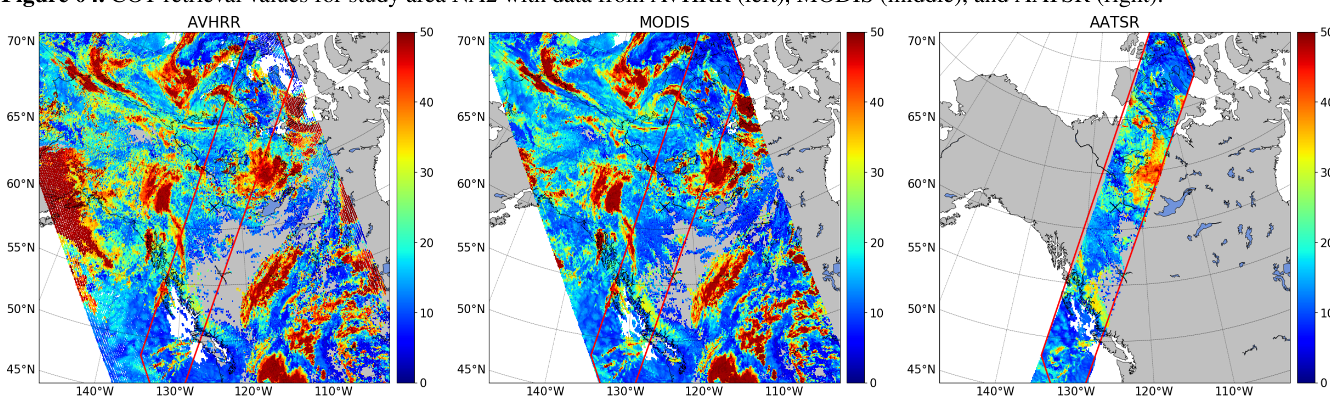


Figure 05. CER retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).

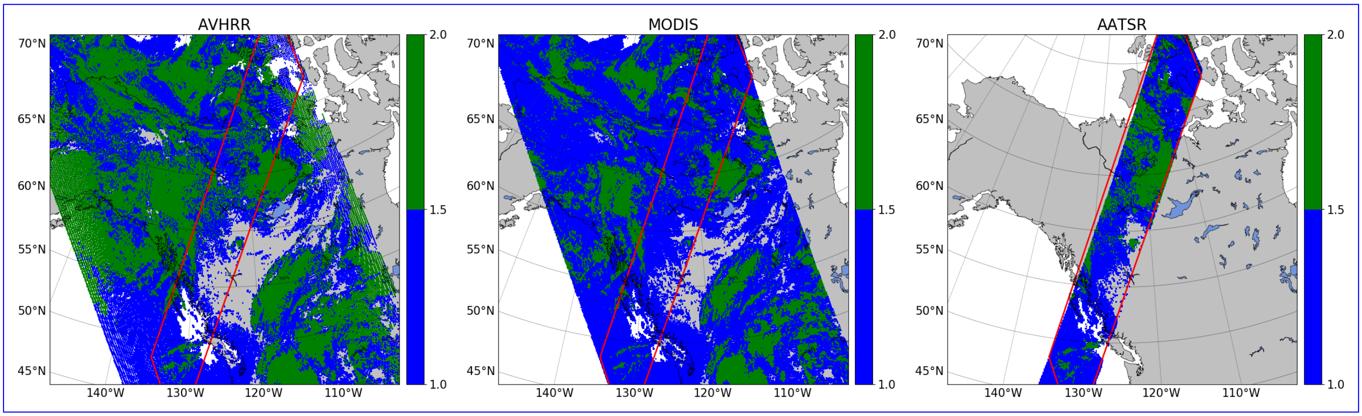


Figure 06. Cloud phase retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).

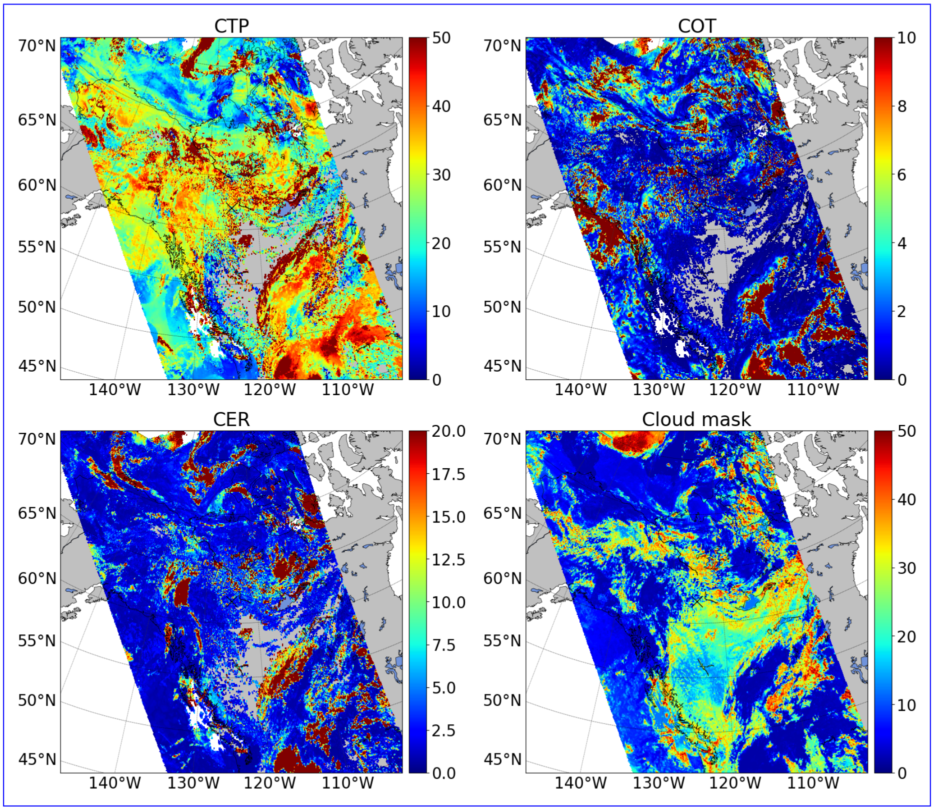


Figure 07. Absolute uncertainties of MODIS AQUA retrieval data for study area NA2 and CTP [hPa], COT, CER [μm], and Cloud mask [%].

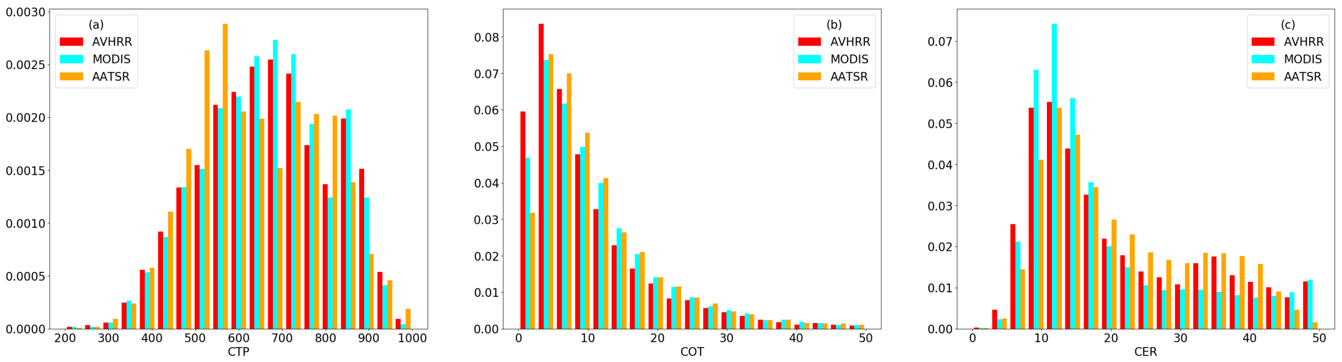


Figure 08. Density histograms of NOAA18 (N18), MODIS AQUA (MYD), and AATSR (ENV) retrieval data for study area NA2 and (a) CTP, (b) COT, and (c) CER.

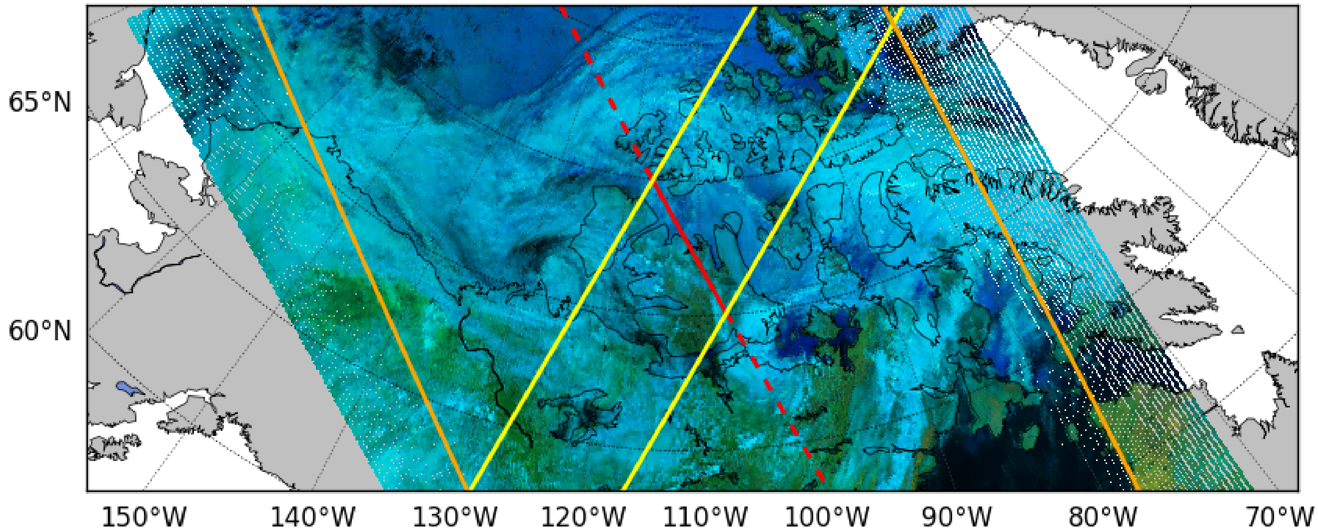


Figure 09. Study area NA1 (North America 1). Red (Ch1), green (Ch2), blue (Ch4 - Ch5) image derived from NOAA18 data resampled to 0.01°×0.01° resolution. Date of observation is 07/22/2008, 19:15 LST. Orange lines: extent of the collocated MODIS granule, yellow lines: extent of the collocated AATSR orbit, red line: CALIOP track outside (dashed) and within (solid) study area.

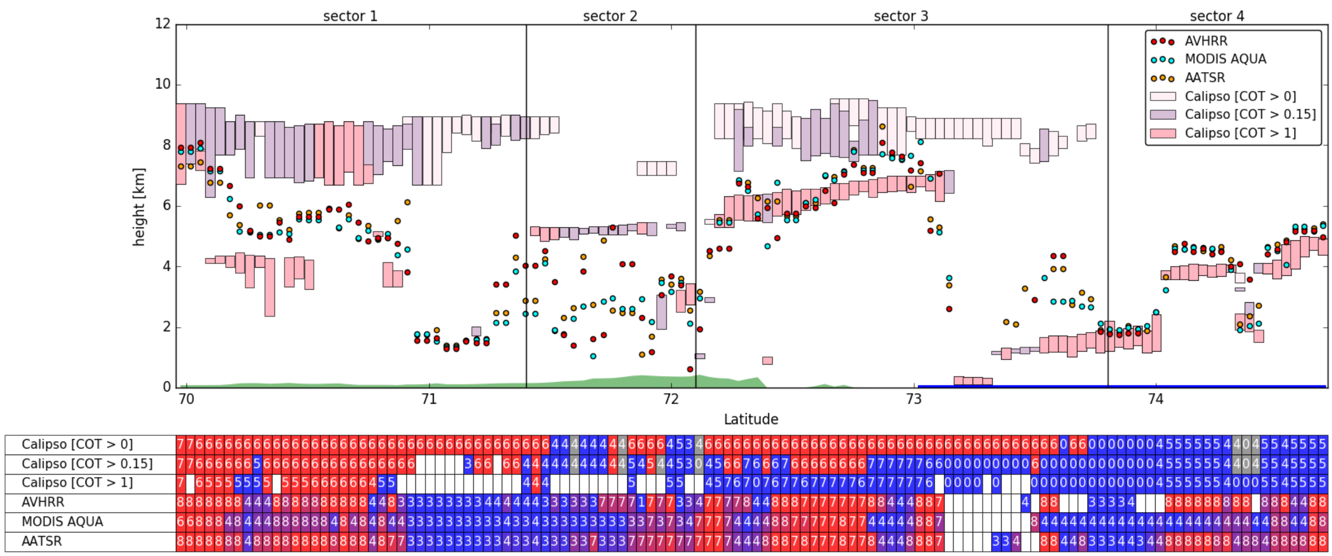


Figure 010. Vertical cross section of study area NA1 (North America 1) along the CALIOP track at 5 km horizontal resolution. Top: CTH for CC4CL retrievals (coloured points) and CALIOP measurements (vertical bars), and surface elevation and surface type (blue = open water, green = land, grey = snow/ice). The CALIOP data are shown for those pressure layers where the cumulative top-to-bottom COD exceeds a threshold value of 0 (top layer), 0.15 (mid layer), and 1 (bottom layer). Bottom: Cloud mask/phase (ice to water = red to blue, cloud free = white, not determined = grey) and type (see Table 04 for key/value pairs) for all three CALIOP layers and CC4CL retrievals. For CC4CL, cloud phase was averaged when resampling, and cloud type was assigned to the most frequent class per grid box. Sectors of characteristic cloud fields are separated by black vertical lines. Number of pixels n = 120

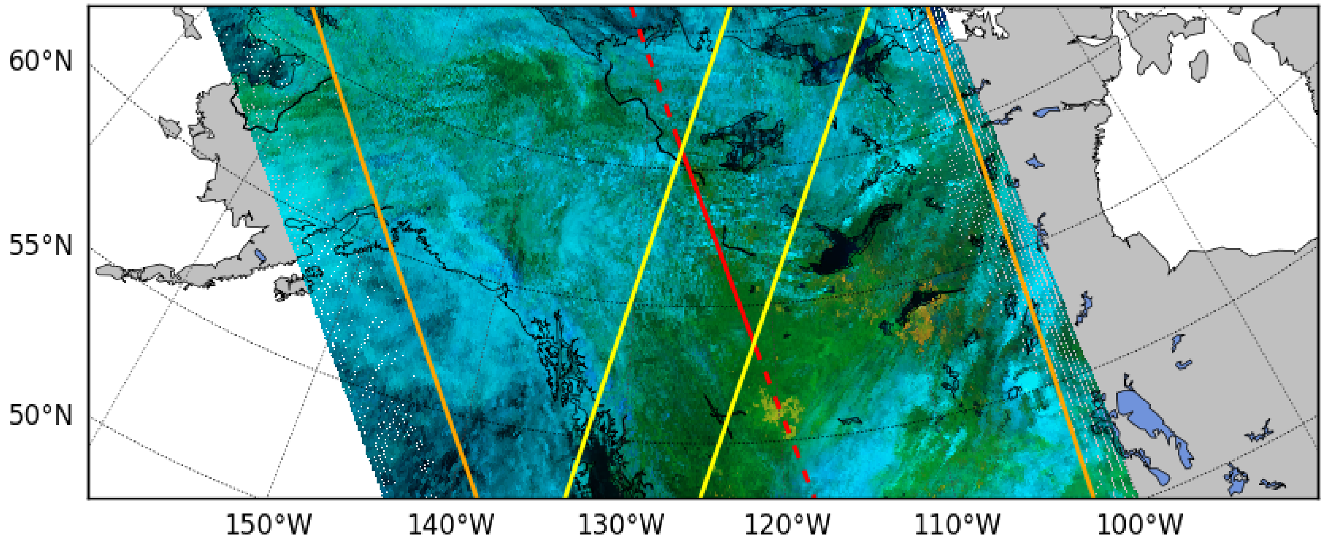


Figure 011. Study area NA2 (North America 2). As Figure 09, but at 07/22/2008, 20:58 LST.

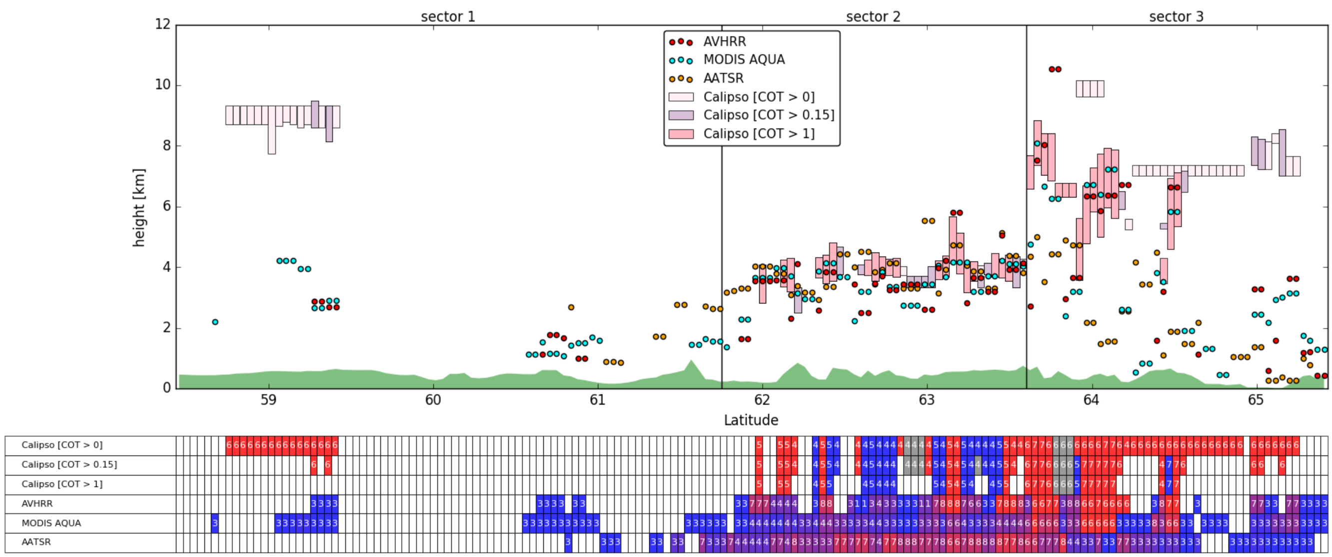


Figure 012. Study area NA2 (North America 2). As Figure 010, but at 07/22/2008, 20:58 LST (n = 163).

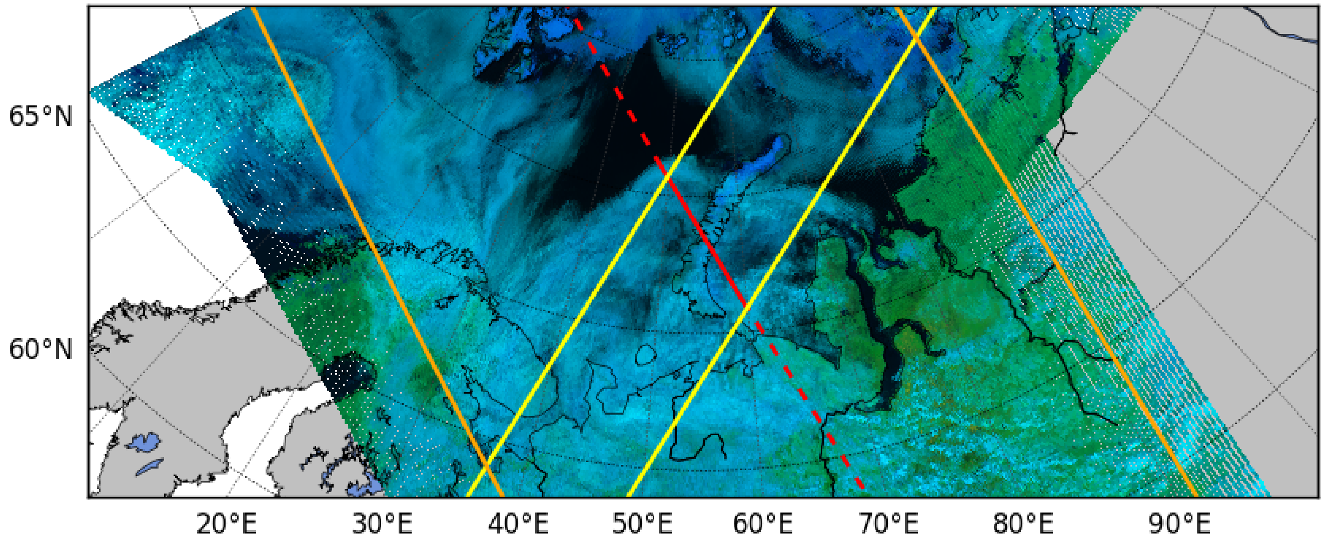


Figure 013. Study area SIB (Siberia). As Figure 09, but at 07/27/2008, 08:10 LST.

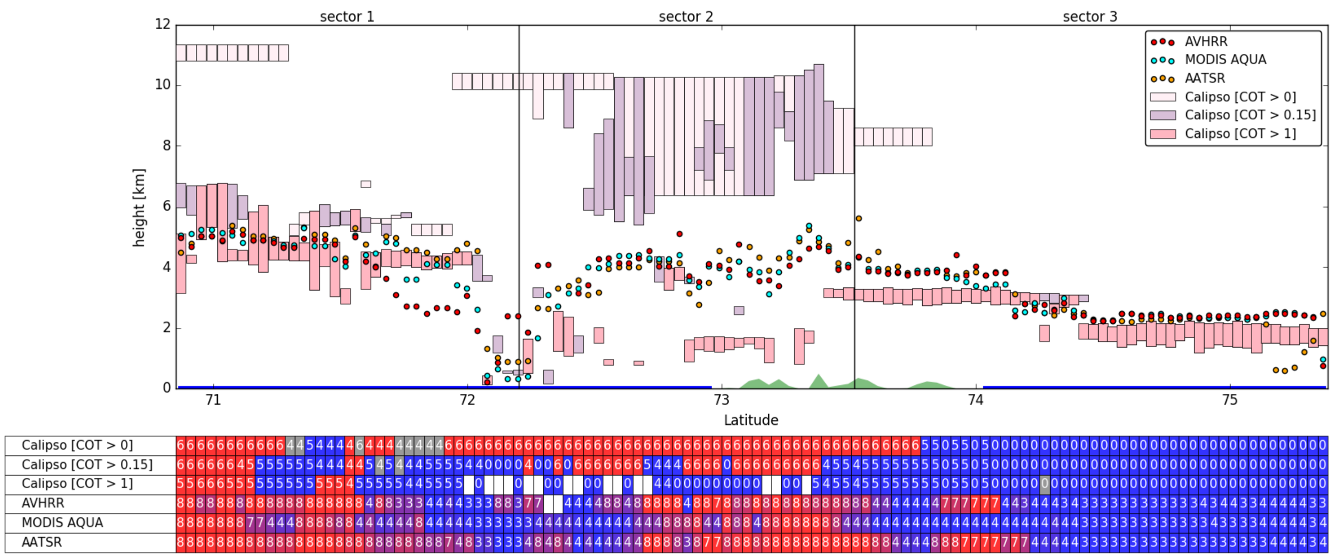


Figure 014. Study area SIB (Siberia). As Figure 010, but at 07/27/2008, 08:10 LST (n = 116).

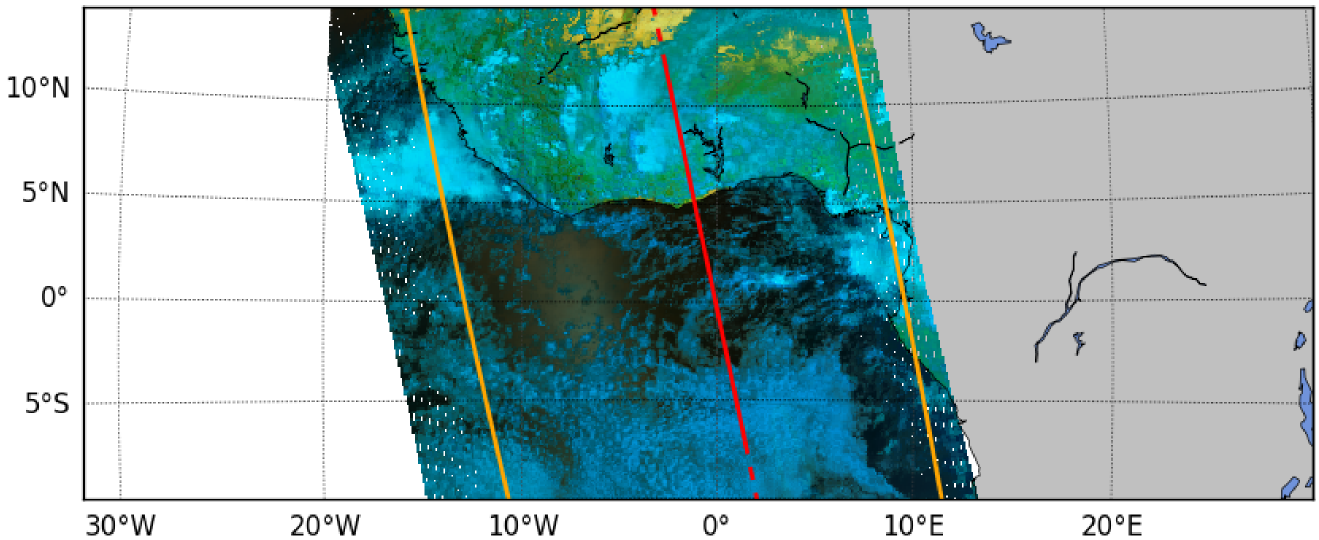


Figure 015. Study area AFR (Africa). As Figure 09, but at 10/24/2009, 13:45 LST.

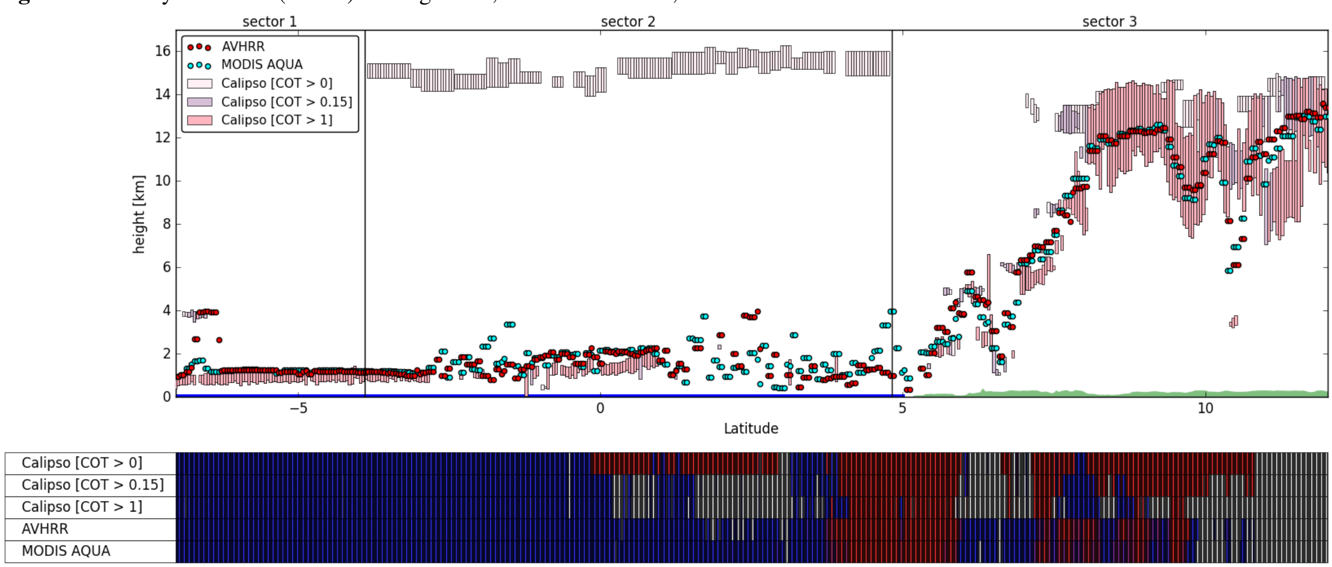


Figure 016. Study area AFR (Africa). As Figure 010, but at 10/24/2009, 13:45 LST. Due to space restrictions, no cloud type values are shown in table. n = 1181

Table 01. The CC4CL AVHRR-heritage dataset channel characteristics for AVHRR, AATSR, and MODIS. Instrument noise as applied within CC4CL is in reflectance for CC4CL channels 1-3, and in brightness temperature [K] for channels 4-6.

	CC4CL ID	sensor ID	channel width (μm)	noise
AVHRR	1	1	0.58 – 0.68	0.005
	2	2	0.725 – 1.10	0.005
	3	3a	1.58 – 1.64	0.005
	4	3b	3.55 – 3.93	0.25
	5	4	10.50 – 11.50	0.2
	6	5	11.5 – 12.5	0.2
MODIS	1	1	0.62 – 0.67	0.01
	2	2	0.841 – 0.876	0.01
	3	6	1.628 – 1.652	0.01
	4	20	3.66 – 3.84	0.2
	5	31	10.78 – 11.28	0.2
	6	32	11.77 – 12.27	0.2
AATSR	1	1	0.545 – 0.565	0.005
	2	2	0.649 – 0.669	0.005
	3	4	1.58 – 1.64	0.005
	4	5	3.51 – 3.89	0.25
	5	6	10.4 – 11.3	0.1
	6	7	11.5 – 12.5	0.1

Table 02. Threshold values applied to ANNCOD data for cloud mask classification.

day	night	twilight	land	sea	snow/ice	threshold
x			x			0.2
x			x		x	0.35
x				x		0.1
x				x	x	0.4
	x		x			0.3
	x		x		x	0.35
	x			x		0.2
	x			x	x	0.4
		x	x			0.3
		x	x		x	0.4
		x		x		0.35
		x		x	x	0.4

Table 03. Linear regression coefficients between collocated AVHRR and MODIS/AATSR channels.

CC4CL channel ID	sensor	regression coefficients
1	MODIS	$0.8945 \times ch1 + 2.217$
	AATSR	$0.8542 \times ch1$
2	MODIS	$0.8336 \times ch2 + 1.749$
	AATSR	$0.7787 \times ch2$
4	MODIS	$0.9944 \times ch4 + 1.152$
	AATSR	$1.0626 \times ch4 - 15.777$
5	MODIS	$0.9742 \times ch5 + 7.205$
	AATSR	$0.9793 \times ch5 + 5.366$
6	MODIS	$0.9676 \times ch6 + 8.408$
	AATSR	$0.9838 \times ch6 + 4.255$

Table 04. Cloud type classification for CC4CL and CALIOP.

ID	CC4CL	ID	CALIOP
0	clear	0	low transparent
1	switched to water	1	low opaque
2	fog	2	stratocumulus
3	water	3	low broken cumulus
4	supercooled	4	altocumulus
5	switched to ice	5	altostratus
6	opaque ice	6	cirrus
7	cirrus	7	deep convective
8	overlap	8	n/a

Table 05. CC4CL primary and secondary output. NN = neural network, SV = state vector, PP = postprocessed, PV = Pavolonis et al. (2005) algorithm, OE = optimal estimation.

variable name	abbrev.	unit	origin	description
		primary variables		
cloud mask	cldmask	1	NN	Binary cloud occurrence classification
cloud type	cldtype	1	PV	Categorical cloud type classification
cloud phase	phflag	1	PV	cloud phase classification
cloud top pressure	ctp	hPa	SV	OE retrieval estimate of cloud top pressure
cloud top pressure unc.	ctp_unc	hPa	SV	OE retrieval unc. of cloud top pressure
cloud effective radius	cer	µm	SV	OE retrieval estimate of cloud effective radius
cloud effective radius unc.	cer_unc	µm	SV	OE retrieval unc. of cloud effective radius
cloud optical thickness	cot	1	SV	OE retrieval estimate of cloud optical thickness
cloud optical thickness unc.	cot_unc	1	SV	OE retrieval unc. of cloud optical thickness
surface temperature	stemp	kelvin	SV	OE retrieval estimate of surface temperature
surface temperature unc.	stemp_unc	kelvin	SV	OE retrieval unc. of surface temperature
		secondary variables		
cloud mask unc.	cldmask_unc	1	PP	derived from NN output and threshold distance
cloud top height	cth	km	PP	derived from CTP and atmospheric profile
cloud top height unc.	cth_unc	km	PP	derived from retrieval unc. of CTP
cloud top temperature	ctt	kelvin	PP	derived from CTP and atmospheric profile
cloud top temperature unc.	ctt_unc	kelvin	PP	derived from retrieval unc. of CTP
cloud water path	cwp	g/m^2	PP	derived from CER and COT (Han et al., 1994)
cloud water path unc.	cwp_unc	g/m^2	PP	derived from retrieval unc. of CER and COT
cloud albedo at 0.06 µm	cla	1	PP	derived from CER and COT based on DISORT (Laszlo et al., 2016)
cloud albedo at 0.06 µm unc.	cla_unc	1	PP	derived from retrieval unc. of CER and COT
cloud albedo at 0.08 µm	cla	1	PP	derived from CER and COT based on DISORT (Laszlo et al., 2016)
cloud albedo at 0.08 µm unc.	cla_unc	1	PP	derived from retrieval unc. of CER and COT
cloud effective emissivity	cee	1	PP	derived from 10.8 and 12.0 µm data

Table 06. Statistics of CTP, COT, and CER retrieval values for study area NA2 and AVHRR (first value in each cell), MODIS (second value), and AATSR (third value). Δ values are given for AVHRR minus MODIS (first value in each cell), AVHRR minus AATSR (second value), and MODIS minus AATSR (third value). *t-Test p-value > 0.1, indicating that differences in mean values are not significant.

	mean	median	stddev	skewness	kurtosis
CTP	667.2, 665.0, 645.2	667.8, 668.1, 632.4	147.5, 142.7, 146.2	-0.2, -0.2, 0.1	-0.4, -0.4, -0.8
Δ CTP	2.2*, 21.9, 19.7	4.2, 22.3, 18.5	63.0, 138.7, 138.9	-0.4, -0.3, -0.3	8.2, 1.0, 0.7
COT	12.3, 13.6, 13.4	7.2, 8.6, 8.8	19.8, 19.7, 17.6	6.6, 5.7, 5.3	60.5, 46.2, 40.8
Δ COT	-1.3, -1.2, 0.2*	-0.6, -1.2, -0.5	16.5, 22.0, 21.3	0.7, 2.4, 1.8	59.6, 41.5, 33.1
CER	21.1, 19.2, 21.3	16.5, 14.4, 18.1	13.0, 12.1, 10.9	1.1, 1.4, 0.6	1.4, 1.2, -0.8
Δ CER	1.9, -0.2*, -2.1	0.5, -1.0, -1.9	7.0, 11.6, 11.3	0.8, 0.8, 0.5	7.9, 4.4, 2.3

Table A1. ERA-Interim variables used within CC4CL. Variables marked with * are available at 0.1°spatial resolution, all others default to 0.72°.

variable name	abbrev.	ID	unit
profile variables			
Geopotential	Z	129	$m^2\ s^{-2}$
Temperature	T	130	K
Specific humidity	Q	133	$kg\ kg^{-1}$
Log. surface pressure	LNSP	152	Pa
Ozone mass mixing ratio	O3	203	$kg\ kg^{-1}$
surface and single level variables			
Sea-ice cover*	CI	31	(0-1)
Snow albedo	ASN	32	(0-1)
Sea surface temperature	SSTK	34	K
Total column water vapour	TCWV	137	$kg\ m^{-2}$
Snow depth*	SD	141	m of water equivalent
10 metre U wind component	U10M	165	$m\ s^{-1}$
10 metre U wind component	V10M	166	$m\ s^{-1}$
2 metre temperature	T2M	167	K
Land/sea mask	LSM	172	(0,1)
Skin temperature*	SKT	235	K

Table A2. CC4CL L2 primary output variables. NN = neural network.

variable name	abbrev.	unit
latitude	lat	degree
longitude	lon	degree
solar zenith	solzen	degree
satellite zenith	satzen	degree
relative azimuth	relaz	degree
cloud top pressure	ctp	hPa
cloud top height	cth	kilometer
cloud top temperature	ctt	kelvin
cloud liquid water path	cwp	g/m^2
cloud effective radius	cer	µm
cloud optical thickness	cot	1
NN cloud optical thickness	cccot	1
cloud albedo	cla	1
cloud effective emissivity	cee	1
cloud fraction	cc_total	1
NN cloud mask	cldmask	(0,1)
cloud phase flag	phflag	1
Pavolonis cloud type	cldtype	1
retrieval convergence flag	conv	1
number of retrieval iterations	niter	1
a priori cost at solution	costja	1
measurement cost at solution	costjm	1
quality control flag	qcflag	1
land/sea flag	lsflag	(0,1)
snow/ice mask	siflag	(0,1)
illumination flag	ilflag	1
surface temperature	stemp	kelvin

Table A3. CC4CL L2 secondary output variables. NN = neural network.

variable name	abbrev.	unit
cloud optical thickness a priori	cot_ap	1
cloud optical thickness first guess	cot_fg	1
cloud effective radius a priori	cer_ap	µm
cloud effective radius first guess	cer_fg	µm
cloud top pressure a priori	ctp_ap	hPa
cloud top pressure first guess	ctp_fg	hPa
surface temperature a priori	stemp_ap	kelvin
surface temperature first guess	stemp_fg	kelvin
albedo in channel no X	alb_ch_X	1
reflectance in channel no X	ref_ch_X	1
brightness temperature in channel no X	bt_ch_X	kelvin
firstguess reflectance in channel no X	fg_ref_ch_X	1
firstguess brightness temperature in channel no X	fg_bt_ch_X	kelvin
reflectance residual in channel no X	ref_res_ch_X	1
brightness temperature residual in channel no X	bt_res_ch_X	kelvin
degrees of freedom signal	deg_free	1