\title{The Community Cloud retrieval for CLimate (CC4CL). Part I: A

framework applied to multiple satellite imaging sensors.}

\author[1]{Oliver Sus}

\author[1]{Martin Stengel}

\author[1]{Stefan Stapelberg}

\author[2]{Gregory McGarragh}

\author[3]{Caroline Poulsen}

\author[4]{Adam Povey}

\author[1]{Cornelia Schlundt}

\author[3]{Gareth Thomas}

\author[3]{Matthew Christensen}

\author[2]{Simon Proud}

\author[1]{Matthias Jerg}

\author[4]{Roy Grainger}

\author[1]{Rainer Hollmann}

\affil[1]{DWD - Deutscher Wetterdienst, Frankfurter Straße 135, 63067 Offenbach, Germany}

\affil[2]{Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, U.K.}

\affil[3]{RAL Space - Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, U.K.}

\affil[4]{National Centre for Earth Observation, Atmospheric, Oceanic and Planetary Physics, University of Oxford, Parks Road, Oxford OX1 3PU, U.K.}

\begin{abstract}

In this study, we present the key features of the Community Cloud retrieval for Climate (CC4CL) processing algorithm. We particularly focus on discussing the novel features of the framework, which set it apart from other approaches: 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 on various resolutions, from 0.5\textdegree\ up to 0.02\textdegree.

Through describing all key input data and processing steps, we inform the future user about important features of this new dataset, and its potential applicability in climate studies. We provide an overview of the retrieved and derived output variables. These are initially validated in a comprehensive and detailed analysis of retrieval results that we collocated with Calipso 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 very realistic estimates of cloud top height and cover for optically thick clouds. The complete sensor set of CC4CL data forms a unique, coherent, long-term, multi-instrument cloud property product that exploits synergetic capabilities of several EO missions. Compared to single sensor retrievals, CC4CL data are improved in terms of accuracy and spatiotemporal sampling.

\end{abstract}

\introduction

Satellite data are an essential source of information in order to better understand and predict climate change. They provide global long-term observations from which geophysical parameters can be derived. These data are used for time-series analysis of climate variables, and also for the assimilation into or validation of climate models \citep{Comiso14,Yang13}. A paramount goal of these efforts is the comprehensive characterization of the global energy and water budgets.

Clouds modify atmospheric windows and radiative forcings of major greenhouse gases \citep{Kiehl97}, and thus considerably constrain the global energy budget. However, clouds are difficult to quantify in terms of composition and temporal or spatial distribution, and possibly are the largest uncertainty in our understanding of climate change \citep{Norris16}. Observations of passive imagers do not sufficiently resolve several important cloud properties, such as vertical structure, sub-pixel heterogeneity, the physical rather than radiatively effective cloud boundary, and the column-integrated ice or liquid water path. Moreover, several background conditions (state of surface and atmosphere, viewing geometry, sensor calibration and spectral response uncertainties) complicate cloud retrievals. These complications propagate uncertainties into derived cloud properties themselves \citep{Hamann14}. Nonetheless, passive satellite imagers are the most widely used instruments for cloud retrievals, providing global coverage at acceptable cost.

There are several satellite-based retrieval frameworks that exploit these types of instruments. One of the earliest of such retrievals is the International Cloud Climatology Project (ISCCP) \citep{Rossow99}. ISCCP provides data on cloud products for 1983--2009, and introduced an approach of cloud type classification based on COT-CTP joint histograms that is popular until today. Current retrieval systems with continuous reprocessing are Pathfinder Atmosphere Extended (PATMOS-x) \citep{Heidinger09,Heidinger12}, EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) cLoud, Albedo and RAdiation (CLARA-A1) \citep{Karlsson13}, and MODIS Collection 6 (MODIS C6) \citep{Platnick17}. These retrievals are variable in their auxiliary data sources, approaches, and complexity. They generally apply radiative transfer models and derived look-up tables (LUT) for providing a clear-sky reference and for simulating atmospheric and cloud contributions to TOA radiances. Retrieval algorithms for cloud properties are variable and based on decision trees and thresholding (PPS in CLARA-A1), CO$\_{2}$-slicing (MODIS C6), or optimal estimation theory (PATMOS-X). COT and REF are generally derived following \citet{Nakajima90}. However, these microphysical variables are not guaranteed to be radiatively consistent with other cloud parameters, which are quantified independently. For cloud masking, the retrieval frameworks apply various approaches such as Na\"{i}ve Bayes (PATMOS-X), dynamic thresholding (CLARA-A1), or a battery of legacy threshold tests (MODIS C6). Finally, cloud phase or type is determined as a function of a combined convergence/CTT-test (CLARA-A1), the \citet{Pavolonis05} threshold algorithm (PATMOS-X), or a bispectral decision tree for channels at 8.5 and 11 µm (MODIS C6). Compared to AVHRR, MODIS has several additional spectral channels that provide cloud microphysical information \citep{Platnick17}. Given this enhanced spectral resolution, MODIS data provide a better setup for retrieving cloud products than AVHRR. Still, the MODIS C6 cloud top retrieval looses sensitivity for optically thinner clouds (COT $<$ 2, \citet{Menzel10}), and sees into the cloud to an optical thickness of approximately 1 \citep{Baum12}. Despite of some promising results, these studies show that current retrievals underestimate cloud top estimates for optically thin clouds even when the full potential of MODIS spectral resolution is used.

There are numerous studies that evaluate the performance of the aforementioned retrievals for cloud cover with weather station data, such as for the Mediterranean \citep{Sanchez17} and conterminous United States \citep{Sun15}. The results are variable, but generally show that the inter-annual correlation is best for PATMOS-X (up to r = 0.94) and worst for CLARA-A1 (r = 0.20 -- 0.7). More importantly, these studies emphasize the difficulty of deriving reliable cloud cover trends from AVHRR time series data, as the retrievals overestimate cloud cover decrease by as much as an order of magnitude \citep{Sun15}. 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 \citep{Walther12}. CLARA-A2 underestimates global CTH by 840 m compared to CALIPSO. For PATMOS-X, MODIS C6 and ISCCP, CLARA-A2 underestimates global CTP by 4--90 hPa and has a cloud phase bias of lower than 9\% \citep{CLARA\_A2\_PVIR}. 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 optically thin ice clouds. However, the detection of supercooled water clouds remains problematic \citep{Baum12}. For MODIS C5, global CTH was underestimated relative to CALIOP by 1.4 km \citep{Holz08}.

Satellite observations of clouds are available for the past 40 years. However, the production of climatologies and trend analyses is a complicated task. 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 can be traded for continuity, and multi-platform algorithms could exploit additional data when newer sensors become available. Modern sensors provide improved spectral and spatial resolutions, and thus potentially better cloud retrievals. However, their data records are too short to produce climatologies of $>$ 30 years, and discontinuities are built into time series when higher resolution satellite data are input to the processing. Major complications of cloud retrievals are 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 resolution than just 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.

The European Space Agency has established the ESA Climate Change Initiative program \citep{ESA\_CCI\_web,Hollmann13} in order to tackle the problems outlined above and to advance the knowledge of the climate system. The project's primary focus is the production of thirteen Essential Climate Variables (ECVs) for the three domains ocean, atmosphere, and land. ECVs are being produced for various climate drivers such as ozone, sea surface temperature, ice sheets, and clouds. This study has been financed as part of the cloud ECV component of ESA CCI \citep{ESA\_Cloud\_CCI\_web}. The main objective of ESA Cloud\textunderscore 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 \mbox{(non-)European} satellites covering several decades. We used satellite data as retrieval input from MODIS Aqua and Terra (2000--2014) \citep{King92}, AVHRR on NOAA-7 to NOAA-19 and METOPA (1978--2014) \citep{Jacobowitz03}, and AATSR on ENVISAT (?--?). Only the AVHRR-equivalent channels from MODIS and AATSR are used, thus the resulting retrieval data are henceforth referred to as the ``AVHRR heritage dataset''. Moreover, the resulting time series are carefully validated against well established existing climatologies (ISCCP, PATMOS-x, CM SAF, and MODIS Collection 6), reanalysis and model data (ERA-Interim and EC-Earth), ground-truth synoptic observations, and CALIPSO Lidar data.

The CC4CL core algorithm was developed in a modular fashion and provides open source access to support distribution and further development within the scientific community. Particular attention was paid to provide the flexibility of processing multiple instruments with one and the same framework, thus maximising the consistency of cloud products independent of the sensor source. Across the solar and thermal spectrum, the framework accounts for physical and radiative consistency amongst all output variables and with input satellite radiances. This is an improvement compared to other established retrieval frameworks. These commonly derive COT and CER by adopting the \citet{Nakajima90} approach, but macrophysical products are estimated independently and are thus radiatively inconsistent with the former variables. Another key novel feature of CC4CL is the production of uncertainty estimates of retrieval parameters through explicit error propagation from input to output data. With these criteria in mind, the Oxford and Rutherford Aerosol and Cloud (ORAC) retrieval \citep{Thomas09, Poulsen12} was chosen out of three competing algorithms within a ``Round Robin'' selection process.

In this study, we present the key features of the CC4CL processing algorithm. We particularly focus on discussing the novel features of the framework, which set it apart from other approaches: 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 on various resolutions, from 0.5\textdegree\ up to 0.02\textdegree. Through describing all key input data and processing steps, we inform the future user about important features of this new dataset, and its potential applicability in climate studies. We provide an overview of the retrieved and derived output variables. These are initially validated in a comprehensive and detailed analysis of retrieval results that we collocated with Calipso 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 very realistic estimates of cloud top height and cover for optically thick clouds.

\section{Data and methods}\label{input\_data}

\subsection{L1 satellite data}\label{sec:L1\_data}

\subsubsection{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 (\autoref{tab:channels}), 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 with a stable blackbody and a space view as references. AVHRR was/is 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 LST) satellites. To reduce drift induced changes in retrieved cloud properties, any AVHRR is replaced with its corresponding morning or afternoon successor once available ($=$ the AVHRR prime record). Typically, one morning and one afternoon NOAA satellite are in orbit simultaneously.

For CC4CL, we use Global Area Coverage (GAC) L1c data on a reduced spatial resolution of 1.1 km $\times$ 4 km at nadir \citep{Devasthale16}. The AVHRR GAC L1c data record, including advanced inter-calibration efforts, was produced for ESA Cloud\textunderscore cci and CMSAF \citep{Schulz09,Karlsson13}. CC4CL processed AVHRR data from 08/1981 (NOAA-7) up to 12/2014 (MetopA + NOAA-19). We applied a filtering technique to noisy channel 3b data (cite), and a database algorithm for splitting midnight orbits and blacklisting.

\begin{table}[h]

\caption{The CC4CL AVHRR-heritage dataset channel characteristics for AVHRR, AATSR, and MODIS. Instrument noise is in reflectance for CC4CL channels 1-3, and in brightness temperature [K] for channels 4-6.}

\begin{tabular}{l|p{0.7cm}p{0.7cm}p{1.8cm}p{0.5cm}}

\hline

& CC4CL ID & sensor ID & channel width ($\mu m$) & noise \\

\hline

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 \\

\hline

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 \\

\hline

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 \\

\hline

\end{tabular}

\label{tab:channels}

\end{table}

\subsubsection{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 constant at 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 CALIPSO 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 \citep{Xiong09}.

CC4CL is applied to Collection 6 MOD021km (Terra) and MYD021km (Aqua) L1b input data \citep{MODIS\_L1B}. For the AVHRR-heritage dataset produced here, the NASA Goddard space flight centre performed the spectral subsetting of the 36 MODIS channels available (see \autoref{tab:channels} for the channels extracted), and data were directly shipped to ECMWF 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 \texttildelow 5 minutes of MODIS data or \texttildelow 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 \citep{MODIS\_PUG}. With CC4CL, we processed data from 02/2000 (Terra) or 08/2002 (Aqua) to 12/2014.

\subsubsection{ATSR-2 and AATSR}

The second and third generation Along Track Scanning Radiometers (ATSR-2 and Advanced ATSR, \citet{Merchant12}) 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 \texttildelow 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 (\autoref{tab:channels}). ATSR is designed to be self calibrating, with two on-board black-body targets for calibration of the thermal channels, and a sun-illuminated opal visible calibration 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\textdegree. The continuous scanning pattern produces a nadir resolution of approximately 1 km $\times$ 1 km with a swath width of 512 pixels or \texttildelow 500 km, providing global coverage every six days.

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 (operational?) launch date until 05/2003 (ERS-2) and 04/2012 (ENVISAT).

\subsection{Auxiliary data}

\subsubsection{ERA-Interim}\label{sec:ERA-Interim}

We use ERA-Interim data as an apriori and first-guess input for the optimal estimation retrieval, and as input for the neural network cloud mask. ERA-Interim is a reanalysis of the global atmosphere, and is available from 1979 until today \citep{ERAInterim,Dee11}. 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 \texttildelow 79 km spacing fur surface fields. We downloaded ERA-Interim data from the ECMWF's MARS archive at a default spatial resolution of 0.72\textdegree, and at 0.1\textdegree\ for the neural network cloud mask input variables (\autoref{tab:ERA-Interim}). We acquired analysis (i.e. not forecast) data at 6-hourly timesteps. After download, all files were converted from GRIB to NetCDF format and remapped to the CC4CL preprocessor grid through CDO. 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. Two ERA-Interim files before and after this time are linearly weighted as a function of their relative time differences.

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 \citep{ERAInterim}. 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 \citep{Willett10}.

\subsubsection{Land use}\label{sec:USGS}

We downloaded USGS Land Use/Land Cover raster data from the global land cover characteristics database \citep{USGS}. The USGS data are used as a land sea mask within the optimal estimation retrieval, as well as a land cover classificator within the cloud mask and the Pavolonis cloud typing scheme (update on this?). The dataset is defined on a regular lat/lon grid with 0.05\textdegree\ resolution. The USGS land cover classification was primarily derived from 1 km AVHRR NDVI 10-day composites for April 1992 through March 1993 \citep{USGS}.

\subsubsection{Land surface BRDF}\label{sec:BRDF}

MODIS Collection 6 BRDF data (MCD43C1, \citet{MODIS\_BRDF}), 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\textdegree\ 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\% (copied from Greg, citation?).

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 time slot. MCD43C1 kernel weights are applied to all CC4CL sensors, assuming negligible differences in spectral response functions.

\subsubsection{Land surface emissivity}\label{sec:emissivity}

For land surface emissivity, we used the CIMSS global land surface IR emissivity database created by the Baseline Fit method \citep{Seemann08}. 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\textdegree\ 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.

\subsection{Collocating CC4CL L2 data and Calipso}

We resampled CC4CL L2 data to a regular latitude/longitude grid at 0.1\textdegree\ $\times$ 0.1\textdegree\ resolution (better 0.05?). This resampling is required for a intercomparison of CC4CL L2 data on a common grid, as differences in sensor spatial resolution are reduced when averaging all values available for each grid box. Calipso's Level 2 5 km Cloud Layer data were produced by averaging over $\approx$14 samples beams with 70 m diameter taken every 335 m within a 5 km along-track corridor. Thus, Calipso data have a 70 m across-track $\times$ 5 km along-track spatial resolution (see also \citet{Holz08}), and the size of corresponding CC4CL grid box is approximately 11 km (meridional) $\times$ 2.9 to 5.6 km (zonal). As a consequence, the CC4CL grid boxes are larger than the reference Calipso pixels, but are still small enough to resolve some of the cloud features that Calipso observes. Note that AVHRR GAC data were produced by averaging 5 neighbouring pixels across-track, but Calipso data were averaged along-track.

\section{The CC4CL retrieval system}\label{processing\_chain}

\subsection{Heritage}

In the early stages of the Cloud\textunderscore cci project, a comprehensive ``Round Robin Exercise'' evaluated three different algorithms regarding their applicability for retrieving cloud parameters from satellite data \citep{Stengel15}, which were 1) the operational processing system of the \citet{CMSAF\_web}, 2) the CLAVR-X algorithm used to generate the PATMOS-x climatology \citep{Heidinger13}, and 3) the Oxford and Rutherford Aerosol and Cloud (ORAC) retrieval \citep{Thomas09, Poulsen12} which was previously used to produce the GRAPE climatology \citep{Thomas09\_GRAPE, GRAPE\_web}. 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, Calipso and AMSR-E reference data.

Based on the outcomes of that study, ORAC was selected to be the cloud retrieval scheme within CC4CL. Moreover, code modifications were identified and characterized to render ORAC fit for purpose.

\subsection{Preprocessing}

\begin{figure}[h]

% \centering

\includegraphics[width=0.47\textwidth]{figures/outline\_preprocessor.png}

\caption{Schematic of the CC4CL preprocessor.}

\label{fig:CC4CL\_preprocessor}

\end{figure}

The CC4CL preprocessor initially defines the dimensions and content of the sensor, surface, and preprocessing grids (\autoref{fig:CC4CL\_preprocessor}).

The sensor and surface grids have the same extent and resolution as the input orbit or granule. The sensor grid is filled with sensor radiances and angles (section \ref{sec:L1\_data}), whereas data on surface BRDF (section \ref{sec:BRDF}), snow/ice coverage (from ERA-Interim, section \ref{sec:ERA-Interim}), and surface emissivity (section \ref{sec:emissivity}) are bilinearly interpolated onto the surface grid. We use BRDF data over land only. For sea pixels, the Cox/Munk ocean surface reflectance model calculates BRDF coefficients as a function of ERA-Interim wind speed. The albedo of snow/ice covered pixels is set to globally constant values of 0.958 (Ch1), 0.868 (Ch2), 0.0364 (Ch3), and 0.0 (Ch4), and is area weighted in case of fractional sea/ice cover.

The preprocessing grid is a regular latitude/longitude grid that covers the extent of the sensor grid, but is defined at a coarser resolution of 0.72\textdegree $\times$ 0.72\textdegree. The preprocessor calculates geolocation values for each grid box, and fills it with data: the average of all sensor angle and surface emissivity values falling into a grid box, and spatially interpolated (nearest neighbour) USGS data (section \ref{sec:USGS}). ERA-Interim variables are defined as surface and profile variables, and were transformed before input to the preprocessing grid as described in section \ref{sec:ERA-Interim}. For profile variables, vertical geopotential coordinates are calculated from pressure coordinates.

The preprocessor then calls the cloud mask (section \ref{sec:CloudMask}) and cloud typing (section \ref{sec:Pavolonis}) algorithms, and writes output data to the surface grid. Finally, RTTOV is executed on the preprocessing grid data as defined by ERA-Interim surface and profile variables. RTTOV outputs above and below cloud transmission for shortwave channels, and radiances (above cloud up- and downwelling, below cloud upwelling), transmission (above and below cloud), and emissivity for longwave channels.

All data are written to NetCDF files as subsequent input to the main processor ORAC. As we determine the cloud mask and phase in the preprocessor, ORAC can optionally be run only on cloudy pixels and one phase to reduce processing time.

\subsection{CC4CL cloud retrieval}

\subsubsection{Cloud detection}\label{sec:CloudMask}

CC4CL applies a set of artificial neural networks (ANN) for cloud masking, one for each of the illumination conditions day, night, and twilight. The ANNs are multilayer perceptrons with one input layer, one hidden layer with 50 neurons, and one output layer, which produces pseudo CALIOP cloud optical depth (ANNCOD) ranging from 0 to 1.

\begin{table}[h]

\caption{Threshold values applied to ANNCOD data for cloud mask classification.}

\begin{tabular}{llllll|l} %{P{0.3cm}P{0.3cm}P{0.7cm}P{0.3cm}P{0.2cm}P{1cm}|P{0.7cm}}

\hline

day & night & twilight & land & sea & snow/ice & threshold \\

\hline

% day

x & & & x & & & 0.2 \\

x & & & x & & x & 0.35 \\

x & & & & x & & 0.1 \\

x & & & & x & x & 0.4 \\ \hline

% night

& x & & x & & & 0.3 \\

& x & & x & & x & 0.35 \\

& x & & & x & & 0.2 \\

& x & & & x & x & 0.4 \\ \hline

% twilight

& & x & x & & & 0.3 \\

& & x & x & & x & 0.4 \\

& & x & & x & & 0.35 \\

& & x & & x & x & 0.4 \\ \hline

\end{tabular}

\label{tab:ANN\_thresholds}

\end{table}

The various ANNs were trained with NOAA-18 AVHRR L1c data, auxiliary information, and cloud optical depth (COD) ``truth'' data obtained from the CALIPSO satellite's LIDAR called CALIOP at 532 nm (\mbox{CAL\textunderscore LID\textunderscore L2\textunderscore 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. Next to the round robin days (?), we selected 12 additional representative training days in 2008 that satisfy basic requirements in terms of collocation between NOAA-18 and CALIPSO, representation of cloud optical depth seasonality, and global coverage. Prior to training, CALIPSO COD values were set to be $\le$ 1. Auxiliary 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 viewing-angle dependency of retrieved ANNCOD.

The binary cloud mask is finally 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 (\autoref{tab:ANN\_thresholds}). 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 (\autoref{tab:ANN\_coefficients}). The resulting coefficients were applied to MODIS and AATSR satellite data before ANN input.

\begin{table}[h]

\caption{Linear regression coefficients between collocated AVHRR and MODIS/AATSR channels.}

\begin{tabular}{l|l|l} %{P{0.3cm}P{0.3cm}P{0.7cm}P{0.3cm}P{0.2cm}P{1cm}|P{0.7cm}}

\hline

CC4CL channel ID & sensor & regression coefficients \\

\hline

1 & MODIS & 0.8945 $\times$ ch1 + 2.217 \\

& AATSR & 0.8542 $\times$ ch1 \\ \hline

2 & MODIS & 0.8336 $\times$ ch2 + 1.749 \\

& AATSR & 0.7787 $\times$ ch2 \\ \hline

4 & MODIS & 0.9944 $\times$ ch4 + 1.152 \\

& AATSR & 1.0626 $\times$ ch4 - 15.777 \\ \hline

5 & MODIS & 0.9742 $\times$ ch5 + 7.205 \\

& AATSR & 0.9793 $\times$ ch5 + 5.366 \\ \hline

6 & MODIS & 0.9676 $\times$ ch6 + 8.408 \\

& AATSR & 0.9838 $\times$ ch6 + 4.255 \\

\hline

\end{tabular}

\label{tab:ANN\_coefficients}

\end{table}

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 quantified a ``truth'' uncertainty or, rather, error. To do so, 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 quantifies the ratio between all correctly classified pixels and the number of all pixels analysed. Finally, the ``truth'' uncertainty is defined as 100 $-$ 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, and a second order polynomial correlation for cloudy cases (\autoref{fig:NN\_unc}):

clear:

\begin{equation}

y = 37.275 \times ND + 49.2

\end{equation}

cloudy:

\begin{equation}

y = 54.133 \times (ND-1)^2 + 1.862

\end{equation}

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

\begin{figure}[h]

% \centering

\includegraphics[width=0.47\textwidth]{figures/NN\_uncertainty\_EQ.png}

\caption{Neural network cloud mask uncertainty}

\label{fig:NN\_unc}

\end{figure}

\subsubsection{Cloud typing}\label{sec:Pavolonis}

Cloud phase is determined by application of the Pavolonis cloud typing algorithm \citep{Pavolonis05}. The Pavolonis algorithm outputs 6 cloud types (\cref{tab:cloud\_types}), 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-$\mu m$ channel data. It is further designed to read 3.75$\mu m$ data whenever available, and otherwise switches to a 1.65$\mu m$ mode. These two different approaches produce nearly identical results, except for certain thin clouds and cloud edges \citep{Pavolonis05}. In addition, we introduced two new cloud types within CC4CL. We decided to change phase for ice clouds whose retrieval CTT is to high (new cloud type = SWITCHED\textunderscore TO\textunderscore WATER if CTT $>$ 273.16 K, the freezing point of water), and for water clouds whose CTT is too low (SWITCHED\textunderscore TO\textunderscore ICE if CTT $<$ 233.16 K, the lower limit of supercooled water).

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 \citep{Pavolonis05}.

\begin{table}[h]

\caption{Cloud type classification for CC4CL and Calipso.}

\begin{tabular}{l|l||l|l}

\hline

ID & CC4CL & ID & Calipso \\

\hline

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 \\

% 9 & probably opaque ice & n/a \\

% 10 & probably clear & n/a \\

\hline

\end{tabular}

\label{tab:cloud\_types}

\end{table}

\subsubsection{Optimal estimation retrieval of COT, CER and CTP}

The optimal estimation retrieval, the Oxford RAL Aerosol and Cloud (ORAC) algorithm, is conceptually based on an inverse problem. The forward model simulates top of atmosphere satellite radiance as a function of atmospheric state and a cloud model. Then, an inverse model attempts to find the optimal fit between the simulated and observed TOA radiances by incrementally varying the parameters of the cloud model. Simulations and measurements are weighted according to measurement errors and uncertainties regarding prior knowledge and the forward model. The important benefits of ORAC 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. Moreover, the algorithm quantifies retrieval uncertainty, which is as a measure of the consistency between the retrieved cloud parameters and the satellite measurements \citep{Poulsen12}. For a more detailed description of the ORAC algorithm see part II of this publication (add citation).

\subsection{Post-processing}

For each input pixel, the main processor produces retrieval values for both ice and liquid clouds. The postprocessor will then select the appropriate output variables according to the Pavolonis cloud phase. As described in section \ref{sec:Pavolonis}, 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.

\begin{table\*}[h]

\caption{CC4CL primary and secondary output. NN = neural network, CV = control vector, PP = postprocessed, PV = \citet{Pavolonis05} algorithm.}

\begin{tabular}{l|l|l|l|l}

\hline

variable name & abbrev. & unit & origin & description \\

\hline

\multicolumn{5}{c}{primary variables} \\

\hline

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 & CV & OE retrieval estimate of cloud top pressure \\

cloud top pressure unc. & ctp\\_unc & hPa & CV & OE retrieval unc. of cloud top pressure \\

cloud effective radius & cer & mic & CV & OE retrieval estimate of cloud effective radius \\

cloud effective radius unc. & cer\\_unc & mic & CV & OE retrieval unc. of cloud effective radius\\

cloud optical thickness & cot & 1 & CV & OE retrieval estimate of cloud optical thickness \\

cloud optical thickness unc. & cot\\_unc & 1 & CV & OE retrieval unc. of cloud optical thickness \\

surface temperature & stemp & kelvin & CV & OE retrieval estimate of surface temperature \\

surface temperature unc. & stemp\\_unc & kelvin & CV & OE retrieval unc. of surface temperature\\

\hline

\multicolumn{5}{c}{secondary variables} \\

\hline

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 \citep{Nakajima90} \\

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 ? \\

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 ? \\

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 mic ? \\

\hline

\end{tabular}

\label{tab:L2\_variables}

\end{table\*}

\section{L2 data - analysis and initial validation}\label{L2\_data}

The analysis and initial validation of L2 data is twofold. We first examine CC4CL cloud properties for one sample scene that extends approximately from 100\textdegree\ W to 170\textdegree\ W and 45\textdegree\ N to 75\textdegree\ N over North America. We focus on the consistency of retrieval values derived from different sensors (AVHRR, MODIS, AATSR). This also includes pixel-based uncertainties of the key variables ctp, cot, cer, and cldmask. We then perform an initial validation of retrieved cloud properties, for which CALIOP-CALIPSO data are our truth reference. This validation is limited to three high-latitude scenes for which collocations for all sensors with CALIOP are available.

\subsection{CC4CL cloud properties}

The sample scene is characterized by various cloud types, and the CC4CL cloud mask defines only a relatively small fraction as cloud free (\cref{fig:CTP\_intercomparison,fig:COT\_intercomparison,fig:CER\_intercomparison}). When visually inspecting CC4CL output for CTP, COT, and CER, the retrieval data appear highly similar for the three different sensors. Main spatial patterns are resolved in all 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 ranges from XY to XY minutes, and thus there is little cloud displacement between observation times.

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 most frequent. CER data are somewhat bimodal, having a primary peak at $\approx$12 micron and a secondary peak at $\approx$35 micron (\cref{fig:histograms,tab:retrieval\_statistics}). 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 (\cref{tab:retrieval\_statistics}). Significance tests of mean differences and standard deviations of residuals between sensor retrievals are sensitive to outliers. These are to some extent influenced by cloud displacement due to observation time differences. Even though we found no significant relationship between sensor retrieval residuals and observation time difference (data not shown), residuals are likely to be smaller and thus possibly insignificant if sensor observation times were identical.

\subsection{Uncertainties}

Median absolute uncertainties are CTP = \load{NA2\_ctp\_unc\_median}26.7 hPa, COT = \load{NA2\_cot\_unc\_median}6.1, CER = \load{NA2\_cer\_unc\_median}2.0 mic, and cloud mask = \load{NA2\_cmask\_unc\_median}13.7 \% (\cref{fig:uncertainties}). The median relative retrieval uncertainty (not shown) is relatively low for all three retrieval variables (CTP = \load{NA2\_ctp\_unc\_median\_relative}4.7 \%, COT = \load{NA2\_cot\_unc\_median\_relative}6.1 \%, CER = \load{NA2\_cer\_unc\_median\_relative}2.0 \%). COT uncertainties increase with COT magnitude, and the RGB image (\cref{fig:RGB\_07222058}) shows that largest uncertainties are found in cases of opaque cloud coverage and probably cloud over sea-ice surfaces. CER results are similar to COT, although relative uncertainties are somewhat lower. Cloud free areas show increased cloud mask uncertainties, and in particular 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.

\begin{table\*}[t]

\caption{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). $\Delta$ values are given for AVHRR minus MODIS (first value in each cell), AVHRR minus AATSR (second value), and MODIS minus AATSR (third value). $^{\ast}$t-Test p-value $>$ 0.1, indicating that differences in mean values are not significant.}

\begin{tabular}{l|lllll}

\hline

& mean & median & stddev & skewness & kurtosis \\

\hline

CTP & \load{ctpMeanN18}667.2, \load{ctpMeanMYD}665.0, \load{ctpMeanENV}645.2 & \load{ctpMedN18}667.8, \load{ctpMedMYD}668.1, \load{ctpMedENV}632.4 & \load{ctpStdN18}147.5, \load{ctpStdMYD}142.7, \load{ctpStdENV}146.2 & \load{ctpSkewN18}-0.2, \load{ctpSkewMYD}-0.2, \load{ctpSkewENV}0.1 & \load{ctpKurtN18}-0.4, \load{ctpKurtMYD}-0.4, \load{ctpKurtENV}-0.8 \\

$\Delta$ CTP & \load{ctpdMeanN18}2.2$^{\ast}$, \load{ctpdMeanMYD}21.9, \load{ctpdMeanENV}19.7 & \load{ctpdMedN18}4.2, \load{ctpdMedMYD}22.3, \load{ctpdMedENV}18.5 & \load{ctpdStdN18}63.0, \load{ctpdStdMYD}138.7, \load{ctpdStdENV}138.9 & \load{ctpdSkewN18}-0.4, \load{ctpdSkewMYD}-0.3, \load{ctpdSkewENV}-0.3 & \load{ctpdKurtN18}8.2, \load{ctpdKurtMYD}1.0, \load{ctpdKurtENV}0.7 \\

COT & \load{cotMeanN18}12.3, \load{cotMeanMYD}13.6, \load{cotMeanENV}13.4 & \load{cotMedN18}7.2, \load{cotMedMYD}8.6, \load{cotMedENV}8.8 & \load{cotStdN18}19.8, \load{cotStdMYD}19.7, \load{cotStdENV}17.6 & \load{cotSkewN18}6.6, \load{cotSkewMYD}5.7, \load{cotSkewENV}5.3 & \load{cotKurtN18}60.5, \load{cotKurtMYD}46.2, \load{cotKurtENV}40.8 \\

$\Delta$ COT & \load{cotdMeanN18}-1.3, \load{cotdMeanMYD}-1.2, \load{cotdMeanENV}0.2$^{\ast}$ & \load{cotdMedN18}-0.6, \load{cotdMedMYD}-1.2, \load{cotdMedENV}-0.5 & \load{cotdStdN18}16.5, \load{cotdStdMYD}22.0, \load{cotdStdENV}21.3 & \load{cotdSkewN18}0.7, \load{cotdSkewMYD}2.4, \load{cotdSkewENV}1.8 & \load{cotdKurtN18}59.6, \load{cotdKurtMYD}41.5, \load{cotdKurtENV}33.1 \\

CER & \load{cerMeanN18}21.1, \load{cerMeanMYD}19.2, \load{cerMeanENV}21.3 & \load{cerMedN18}16.5, \load{cerMedMYD}14.4, \load{cerMedENV}18.1 & \load{cerStdN18}13.0, \load{cerStdMYD}12.1, \load{cerStdENV}10.9 & \load{cerSkewN18}1.1, \load{cerSkewMYD}1.4, \load{cerSkewENV}0.6 & \load{cerKurtN18}1.4, \load{cerKurtMYD}1.2, \load{cerKurtENV}-0.8 \\

$\Delta$ CER & \load{cerdMeanN18}1.9, \load{cerdMeanMYD}-0.2$^{\ast}$, \load{cerdMeanENV}-2.1 & \load{cerdMedN18}0.5, \load{cerdMedMYD}-1.0, \load{cerdMedENV}-1.9 & \load{cerdStdN18}7.0, \load{cerdStdMYD}11.6, \load{cerdStdENV}11.3 & \load{cerdSkewN18}0.8, \load{cerdSkewMYD}0.8, \load{cerdSkewENV}0.5 & \load{cerdKurtN18}7.9, \load{cerdKurtMYD}4.4, \load{cerdKurtENV}2.3 \\

\hline

\end{tabular}

\label{tab:retrieval\_statistics}

\end{table\*}

\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/07222058\_ctp\_multi.png}

\caption{CTP retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).}

\label{fig:CTP\_intercomparison}

%\end{figure\*}

%\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/07222058\_cot\_multi.png}

\caption{COT retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).}

\label{fig:COT\_intercomparison}

%\end{figure\*}

%\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/07222058\_cer\_multi.png}

\caption{CER retrieval values for study area NA2 with data from AVHRR (left), MODIS (middle), and AATSR (right).}

\label{fig:CER\_intercomparison}

\end{figure\*}

\begin{figure\*}[h]

%\centering

\includegraphics[width=0.68\textwidth]{figures/07222058\_uncertainties\_absolute.png} %{figures/07222058\_uncertainties\_percent\_goodspacing.png}

\caption{Absolute uncertainties of MODIS AQUA retrieval data for study area NA2 and CTP [hPa], COT, CER [mic], and Cloud mask [\%].} %{Relative uncertainties [\%] of MODIS AQUA retrieval data for study area NA2 and CTP, COT, CER, and Cloud mask.}

\label{fig:uncertainties}

\end{figure\*}

\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/07222058\_histograms.png}

\caption{Density histograms of NOAA18, MODIS AQUA, and AATSR retrieval data for study area NA2 and (a) CTP, (b) COT, and (c) CER.}

\label{fig:histograms}

\end{figure\*}

\subsection{Validation with Calipso}

We found collocations between Calipso, 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 = \load{NA1\_length}120, \cref{fig:RGB\_07221915}), 07/22/2008 20:58 LST (NA2, n = \load{NA2\_length}163, \cref{fig:RGB\_07222058}), and 07/27/2008 08.10 LST (Siberia = SIB, n = \load{SIB\_length}116, \cref{fig:RGB\_07270810}). These are located within 60\textdegree\ to 75\textdegree\ N latitude, and contain vegetated land, snow-covered land, open ocean, and sea-ice surfaces. For NA1 and SIB, all Calipso pixels were classified as cloud covered, but for NA2, about half of the pixels are cloud free.

When including AATSR, collocations are restricted to high latitude areas and the few pixels located at the intersection between Calipso and AATSR. We thus decided to extend the validation analysis by another scene without AATSR data, which we found in the Gulf of Guinea/West Africa between 7\textdegree\ S and 12\textdegree\ N at 24/10/2009 13.45 LST (Africa = AFR, n = \load{AFR\_length}1181, \cref{fig:RGB\_10241345}). For this scene, about ten times more pixel values are available than for the other scenes, displaying cloud systems such as low-level stratocumulus and deep convection that are not contained in the Arctic data shown here.

We divided all study areas into logical sectors, for each of which a characteristic pattern of cloud coverage and type predominates. Moreover, the validation is shown for comparisons of CTH rather than CTP to enable a more intuitive visualization and discussion. CTH is derived from CTP and the retrieval's atmospheric profile and thus physically consistent.

\subsubsection{Case studies}

\vspace{5mm}\underline{\textit{Case study NA1}}\vspace{2mm}

Study area NA1 is a completely cloud covered scene over northern Canada containing land, ice-covered land, and open ocean surfaces. There is a mix of single and multi-layered clouds of variable optical thickness and height. CC4CL correctly classifies pixels as cloud covered, except for a few cases in sectors 3 and 4. CTH retrievals are very consistent among the three sensors, and only differ in sector 2. Clouds are generally retrieved too low compared to Calipso's top layer, unless the latter is optically thick as in sector 4. In the case of a (semi-)transparent cloud top layer, multiple surfaces contribute to the observed satellite data. CC4CL CTH is then located closer to (sector 1), at (sector 3), or even below the underlying cloud layer (sector 2). For single layer, optically thick (COT $>$ 1) cloud coverage however, CC4CL and Calipso CTH agree very well (sector 4). Under such conditions, the retrieval of CTP and derived products using AVHRR heritage channel data is very accurate. Cloud phase agreement between CC4CL and Calipso is very variable. It is best for optically thick high ice cloud coverage (sector 1), and worst for low water clouds (sector 4).

\vspace{5mm}\underline{\textit{Case study NA2}}\vspace{2mm}

Study area NA2 is located entirely over snow/ice free land in Western Canada. Calipso cloud coverage is \load{cfree\_calipso}4.5\%, spatially broken, and variable in height and phase. Clear sky pixels are mostly identified by CC4CL (\load{cfree\_cc4cl\_and\_calipso\_NA2}69.3\% correct), and cloudy pixels are rarely missed (\load{cloudy\_cc4cl\_and\_calipso\_NA2}78.7\% correct). CC4CL retrievals of thin high clouds and false positive cloudy pixels have low CTH values (sector 1). Calipso's small-scale horizontal variability in cloud phase is reflected by CC4CL data, which however overestimate the fraction of liquid water clouds in sector 2. CC4CL reproduces Calipso's spatial variability in CTH, which it only slightly underestimates by 0.5--1 km in sector 2. In sector 3 however, CC4CL considerably underestimates CTH by up to 7 km. Most of these clouds are optically and geometrically thin.

\vspace{5mm}\underline{\textit{Case study SIB}}\vspace{2mm}

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 Calipso and CC4CL, it is completely cloud covered.

In case of single-layer cloudiness, CC4CL CTH agrees very well with Calipso (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 ($\approx$4 km) clouds in sector 2. For these the retrieved CTH is considerably underestimated by $\approx$6 km, which is probably a result of lower layer contributions that ``contaminate'' the satellite signals. Overall, about \load{phase\_match\_calipso\_CC4CL\_SIB\_layer0}62.3 \% of CC4CL pixels agree with Calipso phase. Phase mismatch occurs in cases of single layer optically thin clouds (sector 2) and, less frequently, stratiform cloudiness (sector 3).

\vspace{5mm}\underline{\textit{Case study AFR}}\vspace{2mm}

Study area AFR is located over the Gulf of Guinea and Western Africa, thus containing open ocean and snow/ice free land surfaces. As we excluded AATSR data, about 10 times more pixel collocations with Calipso are available (n = 1181) than for the Arctic cases. Additionally, measurements contain tropical and coastal cloud systems such as extensive low-level stratiform cloudiness and continental deep convection (\cref{fig:RGB\_10241345}).

In general, the quantitative and qualitative agreement between CC4CL and Calipso CTH is remarkable. CC4CL data track the spatial pattern of continental CTH very well (sector 3), which increases northwards and shows some small scale variability beyond 8\textdegree\ 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 Calipso, 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 $\approx$ 6.7\textdegree\ S and the thin high cirrus layer around 2\textdegree\ S, MODIS retrieval values differ somewhat from CC4CL AVHRR and Calipso. 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 and 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.

\subsubsection{Validation summary}

The four study areas clearly show that CC4CL retrievals of CTH are very close to Calipso values for single layer, optically thick clouds. The mean CTH bias (CC4CL minus Calipso for layer COT $>$ 1) is just \load{NA1\_ctp\_bias\_part}0.24 km for NA1 and pixels at $>$ 73.7\textdegree\ N, \load{SIB\_ctp\_bias\_part}0.05 km for SIB at $>$ 74\textdegree\ N, and \load{AFR\_ctp\_bias\_part}-0.01 km for AFR at $<$ 3\textdegree\ S. For multi-layer clouds, CC4CL estimates are almost exclusively located in between Calipso'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. %In general, underestimation of CTP is rare (\% of pixels whose CTP is CC4CL $<$ Calipso COT $>$ 1 = \load{NA1\_cth\_cci\_lt\_cal}25.3 (NA1), \load{NA2\_cth\_cci\_lt\_cal}20.2 (NA2), \load{SIB\_cth\_cci\_lt\_cal}25.6 (SIB), and \load{AFR\_cth\_cci\_lt\_cal}43.7 (AFR)).

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. Only for site NA2, Calipso identified cloud-free pixels, \load{cfree\_cc4cl\_and\_calipso\_NA2}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 Calipso (\% of missed clouds = \load{cloudfree\_calipso\_cloudy\_CC4CL\_NA1}9.0 (NA1), \load{cloudfree\_calipso\_cloudy\_CC4CL\_NA2}21.3 (NA2), \load{cloudfree\_calipso\_cloudy\_CC4CL\_SIB}0.6 (SIB), \load{cloudfree\_calipso\_cloudy\_CC4CL\_AFR}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 AVHRR.

There is no Calipso cloud layer for which the comparison with CC4CL phase generally agrees best. After having rounded CC4CL values to the nearest integer, the percentage of pixels with equal phase is lowest for the top layer at COD $>$ 0 (NA1 = \load{phase\_match\_calipso\_CC4CL\_NA1\_layer0}49.9 \%, NA2 = \load{phase\_match\_calipso\_CC4CL\_NA2\_layer0}32.2 \%, SIB = \load{phase\_match\_calipso\_CC4CL\_SIB\_layer0}62.3 \%, AFR = \load{phase\_match\_calipso\_CC4CL\_AFR\_layer0}53.0 \%), but similar for the mid layer COD $>$ 0.15 (NA1 = \load{phase\_match\_calipso\_CC4CL\_NA1\_layer1}53.1 \%, NA2 = \load{phase\_match\_calipso\_CC4CL\_NA2\_layer1}46.5 \%, SIB = \load{phase\_match\_calipso\_CC4CL\_SIB\_layer1}66.4 \%, AFR = \load{phase\_match\_calipso\_CC4CL\_AFR\_layer1}94.7 \%), and the bottom layer COD $>$ 1 (NA1 = \load{phase\_match\_calipso\_CC4CL\_NA1\_layer2}47.3 \%, NA2 = \load{phase\_match\_calipso\_CC4CL\_NA2\_layer2}57.4 \%, SIB = \load{phase\_match\_calipso\_CC4CL\_SIB\_layer2}66.3 \%, AFR = \load{phase\_match\_calipso\_CC4CL\_AFR\_layer2}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 (\load{cph\_match\_calipso\_CC4CL\_AFR\_layermean}79.6 \%), followed by SIB (\load{cph\_match\_calipso\_CC4CL\_SIB\_layermean}65.0 \%), and clearly lower for NA1 (\load{cph\_match\_calipso\_CC4CL\_NA1\_layermean}50.1 \%) and NA2 (\load{cph\_match\_calipso\_CC4CL\_NA2\_layermean}45.4 \%).

For ice clouds, the most frequently occurring cloud types are cirrus (ID=6) for Calipso and overlap (ID=8) or cirrus (ID=7) for CC4CL. Water cloud types are more heterogeneous and for Calipso 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. %\textcolor{red}{Statistics: percentages for most frequent cloud types. Also: Which cloud type is dominant if CC4CL phase is different to Calipso phase?}

\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/RGB\_N18\_01x01\_07221915.png} %RGB\_multi\_01x01\_07221915.png}

\caption{Study area NA1 (North America 1). Red (Ch1), green (Ch2), blue (Ch4 - Ch5) image derived from NOAA18 data resampled to 0.01\textdegree$\times$0.01\textdegree\ 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: Calipso track outside (dashed) and within (solid) study area.}

\label{fig:RGB\_07221915}

\includegraphics[width=\textwidth]{figures/calipsoVsCci\_07221915\_nocot\_uncorrectedCtp.png}

\caption{Vertical cross section of study area NA1 (North America 1) along the Calipso track at 5 km horizontal resolution. Top: CTH for CC4CL retrievals (coloured points) and Calipso measurements (vertical bars), and surface elevation and surface type (blue = open water, green = land, grey = snow/ice). The Calipso 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 \cref{tab:cloud\_types} for key/value pairs) for all three Calipso 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 = \load{NA1\_length}120}

\label{fig:calipso\_07221915}

\end{figure\*}

\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/RGB\_N18\_01x01\_07222058.png}

%\includegraphics[width=\textwidth]{figures/RGB\_multi\_01x01\_07222058.png}

\caption{Study area NA2 (North America 2). As \autoref{fig:RGB\_07221915}, but at 07/22/2008, 20.58 LST.}

\label{fig:RGB\_07222058}

\includegraphics[width=\textwidth]{figures/calipsoVsCci\_07222058\_nocot\_uncorrectedCtp.png}

\caption{Study area NA2 (North America 2). As \autoref{fig:calipso\_07221915}, but at 07/22/2008, 20.58 LST. n = \load{NA2\_length}163}

\label{fig:calipso\_07222058}

\end{figure\*}

\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/RGB\_N18\_01x01\_07270810.png} %RGB\_multi\_01x01\_07270810.png}

\caption{Study area SIB (Siberia). As \autoref{fig:RGB\_07221915}, but at 07/27/2008, 08.10 LST.}

\label{fig:RGB\_07270810}

\includegraphics[width=\textwidth]{figures/calipsoVsCci\_07270810\_nocot\_uncorrectedCtp.png}

\caption{Study area SIB (Siberia). As \autoref{fig:calipso\_07221915}, but at 07/27/2008, 08.10 LST. n = \load{SIB\_length}116}

\label{fig:calipso\_07271915}

\end{figure\*}

\begin{figure\*}[h]

%\centering

\includegraphics[width=\textwidth]{figures/RGB\_N18\_01x01\_10241345.png} %RGB\_multi\_01x01\_07270810.png}

\caption{Study area AFR (Africa). As \autoref{fig:RGB\_07221915}, but at 10/24/2009, 13.45 LST.}

\label{fig:RGB\_10241345}

\includegraphics[width=\textwidth]{figures/calipsoVsCci\_10241345\_nocot\_uncorrectedCtp.png}

\caption{Study area AFR (Africa). As \autoref{fig:calipso\_07221915}, but at 10/24/2009, 13.45 LST. Due to space restrictions, no cloud type values are shown in table. n = \load{AFR\_length}1181}

\label{fig:calipso\_10241345}

\end{figure\*}

\conclusions[Discussion]\label{discussion}

\subsection{The flexibility of the optimal estimation approach}

In general, the retrieval values of all control vector variables for AVHRR, MODIS, and AATSR agree well enough to be used interchangeably. Absolute mean differences are small and $\leq$ 21.9 hPa for CTP, $\leq$ 1.3 for COT, and $\leq$ 2.1 for CER, and generally smaller than the mean retrieval uncertainties themselves \textcolor{red}{(calculate)}. Moreover, the RGB images show that all major patterns of cloud coverage and structure are resolved by all three sensors. However, AATSR data show largest deviations to the other sensors (\cref{fig:histograms}). 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, MODIS and AVHRR disagree much more in their spectral response, which results in a reflectance difference of up to 30--40 \% \citep{Trishchenko02}, but their retrieval values are much more similar nonetheless. The difference to 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 are almost exclusively significant. 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. However, differences in cloud conditions at the various observation times and sensor spatial resolution explain part of these discrepancies. Moreover, a non-significant t-Test result is possibly a too strict metric for estimating the comparability 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 amount of cloudy pixels to be compared. The case studies clearly show that, under optimal conditions for single layer cloud retrievals, CC4CL products are consistent with Calipso and practically insensitive to sensor characteristics.

We suggest that AVHRR and MODIS data can be used interchangeably, depending on the user's application. AVHRR data provide long-term data records from 1982, however at a relatively coarse resolution of 5 km $\times$ 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\textdegree\ 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.

\subsection{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. 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 \citet{Poulsen12}). CC4CL COT values are at times unnaturally large, and the associated uncertainty reflects that. Also, it shows under which conditions the optimal estimator does converge at a relatively large cost value. In the cases shown here, large uncertainties are associated with optically thick clouds or underlying snow/ice cover. COT and CER uncertainties are clearly largest, and reflect remaining shortcomings in retrieving these values. Possible explanations are errors in look-up-tables of forward model cloud properties and difficulties to retrieve optical properties for large solar zenith angles.

We applied an independent approach to quantify cloud mask uncertainty. It is a 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 Calipso cloud optical depth, approaches a defined threshold value for cloudiness, the uncertainty increases towards a maximum of 50 \%. This maximum value expresses 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 (\cref{fig:uncertainties}) show uncertainties of 40 - 50 \%. This indicates that the NN is sensitive to bright ground cover, which it might confuse with clouds. We suggest that users of ESA Cloud\textunderscore cci data should routinely 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.

\subsection{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 Calipso, the mean deviation for these cases is as low as 10--240 m CTH. This is a promising result, and shows that the general optimal estimation framework is robust and appropriate for retrieving cloud properties from AVHRR heritage channels.

In case of multi-layer clouds, CC4CL is not able to retrieve CTH of the top layer if the latter is optically thin. The estimate is rather a mid-layer cloud pressure value. This is an expected limitation of our framework, and also of other retrieval algorithms using passive sensor data \citep{Holz08,Karlsson10}. \citet{Poulsen12} found that ORAC CTP and CER estimates are robust when the top ice cloud layer is $>$ 5 optical depths, and otherwise are the weighted average of several cloud layers. The AVHRR heritage channels do not provide sufficient information on retrieving cloud vertical structure. In case of semi-transparent top layer clouds, the underground signal, may it stem from a cloud or the Earth's surface, contributes to the total TOA reflectance or brightness temperature to variable extent. 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 rather the radiatively efficient than physical or geometrical cloud top \citep{Karlsson13}. 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 \citep{Watts11,Holz08,Karlsson13}.

Multi layer cloud property retrievals have been developed \citep{Watts11}, and we also implemented and tested such an approach within CC4CL. However, this method requires additional MODIS channels other than those already contained within the AVHRR heritage set, and thus will not be applicable to a full AVHRR reprocessing. For ESA Cloud\textunderscore cci, the decision was made to deliberately trade spectral information provided by MODIS and AATSR for time series continuity. Thus, discontinuities due to changing spectral resolutions within the entire dataset are avoided (\textcolor{red}{cite Martins paper}). 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 (\textcolor{red}{cite Greg's paper, where this is probably shown in detail}). The correction is only made for ice phase clouds.

On a first view, estimates of cloud phase appear reasonable when compared to Calipso. However, we find a best overall agreement of $\approx$ 65 \% for the lower layers (cumulative COD $>$ 0.15 or $>$ 1). This is just slightly better than a random guess of cloud phase, which would converge towards a 50 \% level of agreement with Calipso as the sample size increases. Cloud phase is generally difficult to quantify, and estimates of various satellite derived products disagree considerably for that variable \citep{Stengel15}. 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 \citep{Marchant16}. 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. Calipso observations, and also DARDAR \citep{Ceccaldi13}, are currently considered to be the most advanced estimates of cloud phase, relying on active measurement principles \citep{Winker09,Karlsson10}. 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 \citep{Pavolonis05}. It was built on 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 (\textcolor{red}{include statistic on mutual agreement}). \citet{Pavolonis05} 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 overall relatively low degree of agreement between CC4CL and Calipso is not satisfying if Calipso was considered 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 however 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 also provide an estimate of cloud phase. Initial tests indicated that this approach would indeed improve the (global) agreement with Calipso (\textcolor{red}{cite Stefan or run tests}), which is to be expected, as the NN is trained with Calipso data. However, no estimate of cloud type would be provided.

Calipso data are considered to be the current benchmark of cloud detection, vertical structure, and phase \citep{Winker09,Karlsson13,Holz08}, and are -- except for the cloud mask -- a source of validation with absolute independence from CC4CL. The main limitation of CALIOP though is its nadir-only 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 probably 3--5 \citep{Karlsson13}. We found no clear relationship between CC4CL CTP uncertainty and the difference between CC4CL CTP and Calipso CTP (data not shown). This suggests that the AVHRR heritage channels provide independent information on cloud vertical structure that is not clearly related to Calipso's CTP estimates. Retrieval uncertainty is driven by background and forward model error as well as the mixed-signal satellite radiances rather than by the complex, real vertical cloud structure.

\conclusions\label{conclusions}

We show that CC4CL is a robust and flexible framework for producing cloud products from passive satellite sensor data. Differences between retrieval values for collocated satellite data are negligible for AVHRR, MODIS, and AATSR. ESA Cloud\textunderscore 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 synergic capabilities of several EO missions. Compared to single sensor retrievals, CC4CL data are improved in terms of accuracy and spatiotemporal sampling.

CC4CL explicitly quantifies retrieval uncertainties according to rules of error propagation and optimal estimation theory. These uncertainties are a valuable source for model validation, data assimilation, climate studies, or retrieval diagnosis. Neural network cloud mask uncertainty is a novel feature that enables the user to assess product quality and to create individualized cloud masks. %to build cloud masks that meet individual requirements.

We find that CC4CL is limited by weaknesses that are common to passive sensor cloud product retrievals. In general, an initial validation with CALIOP data shows that CTH of optically thin clouds is underestimated. In case of multi-layer clouds, the retrieved CTH is then rather a mix product 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, which however 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 $\approx$ 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). For a next reprocessing of the entire AVHRR data record, we replaced the \citet{Pavolonis05} algorithm with a neural network cloud phase estimation with better performance scores.

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 and precise when compared against CALIPSO (bias $<$ 240 m), which emphasizes the maturity and robustness of CC4CL. We thus recommend ESA Cloud\textunderscore cci data to be used for multi-annual studies of cloud parameters and more detailed assessments of regional patterns and diurnal variability.

\begin{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\textunderscore cci project (contract No.: 4000109870/13/I-NB).

\end{acknowledgements}

\appendix

\section{ERA-Interim input fields}

\begin{table}[ht]

\caption{ERA-Interim variables used within CC4CL. Variables marked with \* are available at 0.1\textdegree spatial resolution, all others default to 0.72\textdegree.}

\begin{tabular}{p{3.8cm}|l|r|p{1.5cm}}

\hline

variable name & abbrev. & ID & unit \\

\hline

\multicolumn{4}{c}{profile variables} \\

\hline

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$^{-1}$ \\

\hline

\multicolumn{4}{c}{surface and single level variables} \\

\hline

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 \\

\hline

\end{tabular}

\label{tab:ERA-Interim}

\end{table}

\section{Full list of variables in CC4CL Level-2 files}

\begin{table}[ht]

\caption{CC4CL L2 output variables. NN = neural network.}

\begin{tabular}{l|l|l}

\hline

variable name & abbrev. & unit \\

\hline

\multicolumn{3}{c}{primary variables} \\

\hline

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 & micrometer \\

cloud optical thickness & cot & 1 \\

NN cloud optical thickness & cccot & 1 \\

cloud albedo & cla & 1 \\

cloud effective emissivity & cee & 1 \\

cloud fraction & cc\textunderscore total & 1 \\

NN cloud mask & cldmask & 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 & 1 \\

snow/ice mask & siflag & 1 \\

illumination flag & ilflag & 1 \\

surface temperature & stemp & kelvin \\

\hline

\multicolumn{3}{c}{secondary variables} \\

\hline

cloud optical thickness a priori & cot\textunderscore ap & 1 \\

\hline

\end{tabular}

\label{tab:L2\_variables\_appendix}

\end{table}