EUMETSAT Satellite Application Facility on Climate Monitoring

Algorithm Theoretical Basis Document

CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 3 (CLARA-A3)

Cloud Products (level-1 to level-3)

DOI: 10.5676/EUM_SAF_CM/CLARA_AVHRR/V003

	CDR	ICDR
Fractional Cloud Cover	CM-11012	CM-6011
Joint Cloud property Histogram	CM-11022	CM-6021
Cloud Top Level	CM-11032	CM-6031
Cloud Phase	CM-11042	CM-6041
Liquid Water Path	CM-11052	CM-6051
Ice Water Path	CM-11062	CM-6061

Reference Number: SAF/CM/DWD/ATBD/CLARA/CLD
Issue/Revision Index: 3.3
Date: 06.02.2023

Document Signature Table

	Name	Function	Signature	Date
Author	Karl-Göran Karlsson Abhay Devasthale Martin Stengel Irina Solodovnik Jan Fokke Meirink Nikos Benas	CM SAF scientist (SMHI) CM SAF scientist (SMHI) CM SAF scientist (DWD) CM SAF scientist (DWD) CM SAF scientist (KNMI) CM SAF scientist (KNMI)		06.02.2023
Editor	Marc Schröder	CM SAF Science Coordinator		06.02.2023
Approval	CM SAF Steering Group			
Release	Rainer Hollmann	Project Manager		

Distribution List

Internal Distribution	
Name	**No. Copies**
Board & Steering Group members	electronically

External Distribution		
Company	**Name**	**No. Copies**
EUMETSAT Project Team		On request

Document Change Record

Issue/ Revision	Date	DCN No.	Changed Pages/Paragraphs
3.0	21.04.2021	SAF/CM/DWD/ATBD/CLARA/CLD	Initial version for CLARA-A3 PCR
3.1	28.06.2021	SAF/CM/DWD/ATBD/CLARA/CLD	Revised version after CLARA-A3 PCR
3.2	02.10.2022	SAF/CM/DWD/ATBD/CLARA/CLD	Updates for joint DRR3.2/ORR covering CDR and ICDR Layout revision
3.3	06.02.2023	SAF/CM/DWD/ATBD/CLARA/CLD	Updates following discussions at DRR3.2/ORR Adapted ICDR processing

Applicable documents

Reference	Title	Code
AD 1	EUMETSAT CM SAF CDOP 4 Product Requirements Document (PRD)	SAF/CM/DWD/PRD/ Issue 4.1

Reference documents

Reference	Title	Code
RD 1	Algorithm Theoretical Basis Document for the Cloud Probability of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/ CloudProbability Issue: 2.0, 26.04.2021

Reference	Title	Code
RD 2	Algorithm Theoretical Basis Document for Cloud Top Temperature, Pressure and Height of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/ CTTH Issue 3.0, 26.04.2021
RD 3	Algorithm Theoretical Basis Document for Cloud Micro Physics of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/ CMIC Issue 3, Rev. 0e, 26.04.2021
RD 4	Algorithm Theoretical Basis Document for the Cloud Mask of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/ATBD/ CloudMask, Issue 3, Rev. 0 26.04.2021
RD 5	Algorithm Theoretical Basis Document CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 3 (CLARA-A3) - Top-of-Atmosphere Radiation	SAF/CM/RMIB/ATBD/GAC/TOA

Table of Content

List of Tables

List of Figures

The EUMETSAT SAF on Climate Monitoring

The importance of climate monitoring with satellites was recognized in 2000 by EUMETSAT Member States when they amended the EUMETSAT Convention to affirm that the EUMETSAT mandate is also to “contribute to the operational monitoring of the climate and the detection of global climatic changes". Following this, EUMETSAT established within its Satellite Application Facility (SAF) network a dedicated centre, the SAF on Climate Monitoring (CM SAF, https://www.cmsaf.eu).

The consortium of CM SAF currently comprises the Deutscher Wetterdienst (DWD) as host institute, and the partners from the Royal Meteorological Institute of Belgium (RMIB), the Finnish Meteorological Institute (FMI), the Royal Meteorological Institute of the Netherlands (KNMI), the Swedish Meteorological and Hydrological Institute (SMHI), the Meteorological Service of Switzerland (MeteoSwiss), the Meteorological Service of the United Kingdom (UK MetOffice), and the Centre National de la Recherché Scientifique (CNRS). Since the beginning in 1999, the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) has developed and will continue to develop capabilities for a sustained generation and provision of Climate Data Records (CDR’s) derived from operational meteorological satellites.

In particular, the generation of long-term data records is pursued. The ultimate aim is to make the resulting data records suitable for the analysis of climate variability and potentially the detection of climate trends. CM SAF works in close collaboration with the EUMETSAT Central Facility and liaises with other satellite operators to advance the availability, quality and usability of Fundamental Climate Data Records (FCDRs) as defined by the Global Climate Observing System (GCOS). As a major task the CM SAF utilizes FCDRs to produce records of Essential Climate Variables (ECVs) as defined by GCOS. Thematically, the focus of CM SAF is on ECVs associated with the global energy and water cycle.

Another essential task of CM SAF is to produce data records that can serve applications related to the Global Framework of Climate Services initiated by the WMO World Climate Conference-3 in 2009. CM SAF is supporting climate services at national meteorological and hydrological services (NMHSs) with long-term data records but also with data records produced close to real time that can be used to prepare monthly/annual updates of the state of the climate. Both types of products together allow for a consistent description of mean values, anomalies, variability and potential trends for the chosen ECVs. CM SAF ECV data records also serve the improvement of climate models both at global and regional scale.

As an essential partner in the related international frameworks, in particular WMO SCOPE-CM (Sustained COordinated Processing of Environmental satellite data for Climate Monitoring), the CM SAF - together with the EUMETSAT Central Facility, assumes the role as main implementer of EUMETSAT’s commitments in support to global climate monitoring. This is achieved through:

- Application of highest standards and guidelines as outlined by GCOS for the satellite data processing,
- Processing of satellite data within a true international collaboration benefiting from developments at international level and pollinating the partnership with own ideas and standards,

- Intensive validation and improvement of the CM SAF climate data records,
- Taking a major role in data record assessments performed by research organisations such as WCRP (World Climate Research Program). This role provides the CM SAF with deep contacts to research organizations that form a substantial user group for the CM SAF CDRs,
- Maintaining and providing an operational and sustained infrastructure that can serve the community within the transition of mature CDR products from the research community into operational environments.

A catalogue of all available CM SAF products is accessible via the CM SAF webpage, https://www.cmsaf.eu. Here, detailed information about product ordering, add-on tools, sample programs and documentation is provided.

1 Introduction

This CM SAF Algorithm Theoretical Basis Document (ATBD) provides an overarching view on the cloud algorithms implemented for the CM SAF cLoud, Albedo and surface Radiation dataset from AVHRR data - Edition 3 (CLARA-A3), a successor of the CLARA-A2.1 record. The record contains retrieved geophysical parameters from inter-calibrated measurements of the Advanced Very High Resolution Radiometer (AVHRR) onboard the NOAA satellites (TIROS-N, NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11, NOAA-12, NOAA-14, NOAA-15, NOAA-16, NOAA-17, NOAA-18, NOAA-19) and the EUMETSAT Metop-A, Metop-B and Metop-C satellites.

The cloud algorithms applied are part of the NWC/PPS version v2021 software package (PPS hereafter; Note: cloud algorithms in PPS v2021 are identical to and PPSv2018-patch5). The PPS package includes retrieval algorithms for probabilistic cloud detection, cloud top properties (height, temperature, pressure) and cloud physical properties (cloud-top thermodynamic phase, cloud optical thickness, cloud particle effective radius, and liquid/ice water path).

This document seamlessly describes all elements of the production of the final CLARA-A3 cloud products which are structured in the following three topics.

1. A description of the data sources and a summary of AVHRR instrument characteristics are given, including a description of the inter-calibration applied to AVHRR GAC measurements.
2. A report on the derivation of the cloud products by applying PPS algorithms. Note, significant parts of the PPS algorithms have already been documented in external ATBDs ([RD 1], [RD 2], [RD 3], and [RD 4]), which will be referred to in this document when appropriate.
3. The elaboration of the production of the daily composites (level-2b), and daily and monthly means and histograms (level-3 data) based on the level-2 products provided by PPS.

Basic accuracy requirements are defined in the product requirements document [AD 1].

The CLARA-A3 data record contains multiple cloud parameters derived from AVHRR:

	CDR	ICDR
Fractional Cloud Cover (CFC, Section 3.1)	**CM-11012**	**CM-6011**
Joint Cloud property Histogram (JCH, Section 3.2)	**CM-11022**	**CM-6021**
Cloud Top Level (CTO, Section 3.3)	**CM-11032**	**CM-6031**
Cloud Phase (CPH, Section 3.4)	**CM-11042**	**CM-6041**
Liquid Water Path (LWP, Section 3.5)	**CM-11052**	**CM-6051**
Ice Water Path (IWP, Section 3.6)	**CM-11062**	**CM-6061**

2 Processing of measures AVHRR radiances (level-1)

2.1 Historical overview of the AVHRR GAC data record

Measurements from the Advanced Very High Resolution Radiometer (AVHRR) radiometer onboard the polar orbiting NOAA satellites and the EUMETSAT Metop satellites have been performed since 1978. Figure 2-1 gives an overview over the satellites carrying the AVHRR instrument, which are used in the production of CLARA-A3. The AVHRR instrument only measured in four spectral bands in the beginning (AVHRR/1) but from 1982 a fifth channel was added (AVHRR/2) and in 1998 even a sixth channel was made available (AVHRR/3), although only accessible if switched with the previous third channel at 3.7 micron.

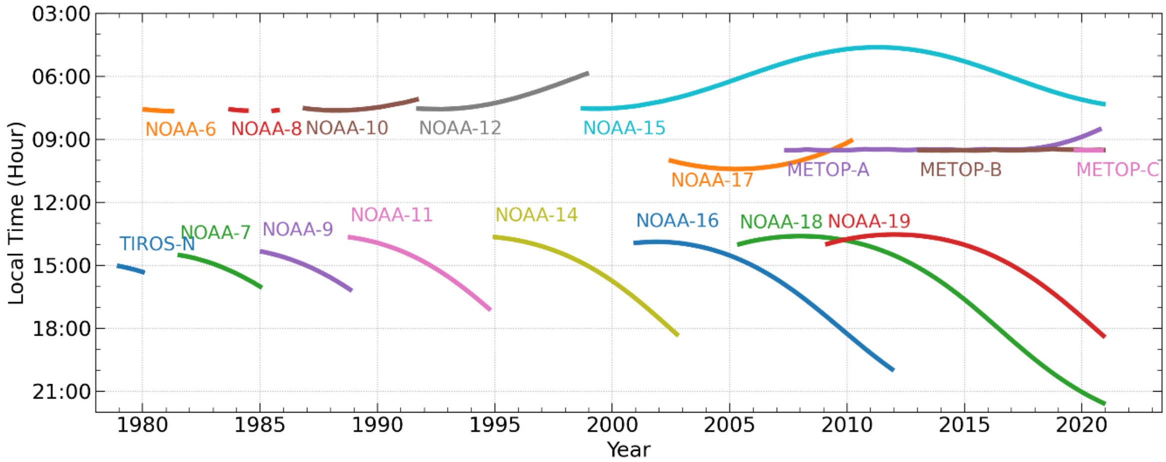

Figure 2-1: Local solar times at equator observations for all AVHRR-carrying NOAA satellites from TIROS-N to NOAA-19 and EUMETSAT's Metop A/B/C satellites. Shown are all data that are used for the CDR processing. Satellites shown for 2020 are also the baseline for the ICDR processing beyond 2020 except Metop-C where some calibration problems were encountered for years after 2020. The figure shows ascending (northbound) equator crossing times for afternoon satellites (NOAA-7 to NOAA-19) and descending (southbound) equator crossing times for morning satellites (NOAA-12 to NOAA-17 and Metop A/+B/C). Corresponding night-time observations take place 12 hours earlier/later. Some data gaps are present but only for some isolated dates.

Table 2.1 describes the AVHRR instrument, its various versions and the satellites carrying them. The AVHRR instrument measures at a horizontal resolution close to 1 km at nadir but only data at a reduced resolution (footprint size approximately 1km (along track) by 4 km (across track); sampled every third scanline) are permanently archived and available with global coverage since the beginning of measurements. This data record is denoted Global Area Coverage (GAC) AVHRR data.

2.2 Compilation of the CM SAF CLARA-A3 cloud data record

The CLARA-A3 data record of global cloud products retrieved by CM SAF cloud retrieval methods spans the time period 1979-present and is based on AVHRR generations AVHRR/1, AVHRR/2 and AVHRR/3 (see Table 2.1).

Table 2.1: Spectral channels of the Advanced Very High Resolution Radiometer (AVHRR). The three different versions of the instrument are described as well as the corresponding satellites. Notice that channel 3A was only used continuously on NOAA-17, Metop-A/B/C. For the other satellites with AVHRR/3 it was used only for shorter periods. The wavelength ranges given represent the full width of the spectral response function (SRF) at half maximum of the SRF (FWHM).

Channel Number	Wavelength (micrometers) AVHRR/1 TIROS-N, NOAA-6,8,10	Wavelength (micrometers) AVHRR/2 NOAA-7,9,11,12,14	Wavelength (micrometers) AVHRR/3 NOAA-15,16,17,18,19 Metop-A/B/C
1	0.58 - 0.68	0.58 - 0.68	0.58 - 0.68
2	0.725 - 1.10	0.725 - 1.10	0.725 - 1.10
3A	-	-	1.58 - 1.64
3B	3.55 - 3.93	3.55 - 3.93	3.55 - 3.93
4	10.50 - 11.50	10.50 - 11.50	10.50 - 11.50
5	Channel 4 repeated	11.5 – 12.5	11.5 – 12.5

Observations from polar orbiting sun synchronous satellites are made at the same local solar time at each latitude band. Normally, satellites are classified into observation nodes according to the local solar time when crossing the equator during daytime (illuminated conditions). For the NOAA satellite observations, a system with one morning observation node and one afternoon observation node has been utilized as the fundamental polar orbiting observation system. This guarantees four almost equally distributed observations per day (if including the complementary observation times at night and in the evening when the satellite passes again 12 hours later). Equator crossing times have varied slightly between satellites. Morning satellites have generally been confined to the local solar time interval 07:00-08:00 and afternoon satellites to the interval 13:30-14:30. However, a change was introduced for the morning satellites NOAA-17 and Metop-A, now being defined in a so-called mid-morning orbit with equator crossing times close to 10:00. A specific problem with the observation nodes for the NOAA satellites has been the difficulty to keep observation times stable for each individual satellite (e.g., as described by Ignatov et al., 2004). This is also illustrated in Figure 2-1 for all used satellites. Some compensation for this has been attempted in the CM SAF data record but not for all parameters.

2.3 AVHRR Level-1 input to CLARA-A3

An important aspect for any product-based climate data record (formally denoted Thematic Climate Data Records - TCDRs) is that retrieved products have been derived from accurately calibrated and harmonized radiances (formally denoted Fundamental Climate Data Records - FCDRs). So far, only the visible channels of the AVHRR have received FCDR status. Attempts were made in the FIDUCEO project to create an FCDR also for the infrared channels but this work could not be finished within the framework of that project. As a consequence, and awaiting an FCDR solution later on, the EUMETSAT Secretariat defined a project that in cooperation with CM SAF would compile an AVHRR FDR that could serve the needs for the CLARA-A3 TCDR and which also would release the FDR in parallel with CLARA-A3 for external use. The FDR is labelled the EUM AVHRR FDR. The FDR is basically compiled with the PyGAC tool, which was used also for CLARA-A2, but with some changes to output format and quality flag information. Thus, the AVHRR input to CLARA-A3 is the EUM AVHRR FDR with the following DOI number: 10.15770/EUM_SEC_CLM_0060. In the following the calibration is summarized, complemented with a short description of a channel 3b noise filtering which is applied to earlier satellites.

2.3.1 Calibration

The calibration of the AVHRR shortwave reflectances, which is applied in the generation of the FDR by EUMETSAT, is based on work initially performed by NOAA (updated version of Heidinger et al. (2010)). A similar FDR is prepared for the compilation of the "NOAA Pathfinder Atmospheres - Extended" (PATMOS-x) data record (for full description, see http://cimss.ssec.wisc.edu/patmosx/overview.html). The AVHRR reflectances are calibrated based on collocations with MODIS. This information is propagated back in time to all AVHRR-carrying satellites. A complete documentation of the calibration method is given by Heidinger (2018).

The used visible calibration information for both the CDR and the ICDR was delivered in 2017 by NOAA. It means that time-dependent calibration corrections after 2017 are based on extrapolated information. For some satellites (e.g. NOAA-19, Metop-A and Metop-B) a limited number of years could be used to estimate these calibration corrections. More serious, for Metop-C (launched in 2018) only a preliminary calibration could be used and the time-dependent corrections are therefore particularly uncertain. For this reason, it was decided to not use Metop-C data in the continuous ICDR production. Metop-C data will be included later in the ICDR when a new calibration update is available. This update will also be beneficial for the other mentioned satellites.

The calibration of infrared AVHRR channels is based on the onboard blackbody calibration. This has been found to provide stable and reliable results already. However, future upgrades of the AVHRR FCDR need to address remaining issues here also for the infrared channels (e.g., recognizing the work of Mittaz et al., 2009).

More details on the methods applied for calibration can be found in Section 2.1.1 of [RD 5].

2.3.2 Quality control

Based on meta-data information resulting from quality flags in the EUM AVHRR FDR, the quality control and orbit blacklisting is performed. Orbit blacklisting, which means removal of a particular orbit from processing, is applied when an orbit is so contaminated with corrupt or erroneous data (e.g., due to scan motor problems) that the product quality can be seriously affected. This is practically speaking very similar to what was done previously for CLARA-A2.1. Here we also utilise the results from the rigorous work to monitor data quality in the AVHRR GAC period 1980-2015 that was done for the previous CLARA-A2 and ESA-CLOUD-CCI data records.

However, two additional screening methods have been used for CLARA-A3 leading to partial removal of affected orbits. Now also parts of orbits with pronounced solar contamination of infrared information (especially problematic for the calibration of AVHRR channel 3b) have been removed. This was not done for CLARA-A2. These occurrences were quite frequent for some satellites and unfortunately led to the need of gap filling measures for some products (in particular, TOA radiation products). The second additional screening of data was made after discovering a diverging behaviour of AVHRR channel 3b at the extreme cold end for satellites NOAA-15, NOAA-16 and Metop-B. To be more specific, the new cloud masking method CMAPROB applies the brightness temperature difference between AVHRR channels 3b and 4 for improving the identification of low-level water clouds during night. However, this method has been trained with data from the two satellites NOAA-18 and NOAA-19 in the CALIPSO era 2006-2016. It turned out that this works well for all other previous satellites except the three mentioned satellites. Here, the channel 3b values deviate considerably when temperatures get very cold and this could lead to false detection of clouds. Consequently, it was decided to blacklist all night-time results for these satellites when ERA-5 surface temperatures fell below 255 K. This led to a limited sampling of e.g. the Arctic region in the polar winters in the period 2000-2002 before other satellites without any sign of this problem (e.g. NOAA-17) were introduced. This abnormal behaviour of Ch3b brightness temperatures has to be studied further in a future work by creating an AVHRR FCDR.

2.3.3 Noise filtering for channel 3b

A post-processing noise-filtering method was developed for the earlier CLARA-A2 version, and is applied in the same manner for CLARA-A3 as well. This method is applied for reducing the time-varying and sometimes high noise levels in AVHRR channel 3b at 3.7 micrometers (caused by internal interference problems on the satellite platform) for some of the earlier NOAA-satellites, i.e. all AVHRR/1 carrying satellites and AVHRR/2 carrying satellites until NOAA-14. Periods with high noise levels were previously found to lead to increased frequencies of falsely detected cloudy and clear areas, in particular over cold land surfaces. Cloud detection is partly based on a careful analysis of brightness temperature differences between infrared channels. This is particularly important at night when only infrared channels are available. Noise in AVHRR channel 3b may lead to artificially introduced brightness temperature differences and may thus also lead to false detection of clouds. Since the impact of noise is increasing for colder targets (because of the limited radiometric resolution of the

AVHRR instrument), most serious problems are seen over cold land surfaces at high latitudes and in the polar regions.

By use of the new method this has now been significantly reduced. The method is based on the use of a dynamic size median filter operating on channel 3b brightness temperatures. The size of the filter is a function of previously monitored noise levels. The method also includes a specific restoration method for small-scale cloudy pixels in warm environments (e.g. cumulus over warm land and ocean surfaces) to prevent the method from removing true clouds at the very finest scales. It is important to note that the method can only reduce the problems of noise contamination and not solve it. It also leads to the loss of some details at the finest resolvable resolution in the cloud fields that otherwise would have been observed. A full description of the noise filtering method is given by Karlsson et al. (2017).

3 Retrieval of swath-based cloud properties (level-2)

This Section provides basic information on the retrieval of cloud properties with PPS from inter-calibrated AVHRR observations. References to more in-depth information are given in each subsection. However, observe that in section 3.7we give a summary of the main limitations of each product.

3.1 Fractional Cloud Cover [CM-11012, CM-6011, CFC]

This product is a result of the PPS probabilistic cloud masking procedure based on Bayesian theory. Individual measurement features from all available AVHRR channels (e.g., reflectance at 0.63 µm), are linked to the likelihood of being cloudy, which is inferred from a reliable reference, i.e. the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) in this case. The impact of differences in observation geometry between AVHRR and CALIOP has been thoroughly examined and is discussed in detail by Karlsson and Håkansson (2018). The probabilities of being cloudy for all individual features are multiplied, assuming no correlations, which enables an analytic solution to infer the total probability of a pixel being cloudy. The assumption of no correlation between features is not always fulfilled but this approach has nevertheless been shown to give improved results compared to the previous method applied for CLARA-A2 (as reported by Karlsson et al., 2020). The changed methodology has also led to that the CFC product is now extended to include also an estimation of uncertainty.

One of the recently introduced key features is that the cloud probabilities are derived scene-dependently. The training of CMAPROB is based on AVHRR-collocations with CALIPSO-CALIOP where data has been collected and sub-divided over 28 different surface types, also separated between day and night conditions. Consequently, 28 different statistical distributions of image feature behaviour for cloudy and clear conditions for day and night, respectively, is then used when executing CMAPROB. This accounts for scene-dependent variation in detection skills of passive satellite sensors and supports the general application globally of a 50% probability threshold to distinguish cloudy and clear-sky scenes without scene-dependent biases in detection skill. It was found especially important to apply a dynamic definition of different surfaces which follows temperature regimes in a reasonable way. Thus, instead of geographically fixed coordinates (i.e., latitudes and longitudes) we used characteristic temperature intervals for tropical, sub-tropical, high-latitude and polar surfaces guided by surface temperatures from ERA-5. By doing this, a seamless transition of the results between different surfaces could be achieved.

All downstream retrievals of the other cloud properties described in the following subsections are made for cloudy pixels only, identified by applying the 50% threshold. All details on the cloud detection schemes are described in [RD 1] where also measures are described for how to handle varying viewing angles due to the CALIOP observation limitation to nadir viewing. The reader is also referred to Karlsson et al. (2015, 2020) for more examples (e.g., application of the methodology for other sensors, like SEVIRI).

The cloud fractional cover is defined as the fraction of cloudy pixels (applying 50% threshold, see above) per grid cell compared to the total number of analyzed pixels in the grid cell, i.e. all pixels that have a cloud probability assigned. Fractional cloud cover is expressed in percent.

3.2 Joint Cloud property Histogram [CM-11022, CM-6021, JCH]

The JCH product is a combined histogram of CTP (see Section 3.3), COT (see Sections 3.5 and 3.6), and CPH (see Section 3.4), covering the solution space of these parameters. This three-dimensional histogram gives the absolute numbers of occurrences for specific COT-CTP-CPH combinations defined by specific bins, which can be found in Section 4.4.2

3.3 Cloud Top level [CM-11032, CM-6031, CTO]

This product is based on the PPS neural network (of the type multilayer perceptron) trained offline with collocations of passive imager measurements (i.e., the three infrared channels of AVHRR) and CALIOP (version 4 data) cloud top pressure observations. Training was performed using collocations over the full globe, thus covering all variations caused by varying viewing angles. Collocated Reanalysis profiles from ERA-5 are used to convert cloud top pressure to cloud top temperature and height. The latter is expressed as altitude above ground topography. All three variants of this product are available for users.

In addition to the retrieved values (direct output of the neural network), an estimation of the upper and lower retrieval boundaries (at the 16^{th} and 84^{th} percentile level) are produced using a special version of Quantile Regression Neural Networks (see Pfreundschuh et al. 2018). The actually reported CTTH uncertainties are processed by averaging the distances of both boundaries to the retrieved value to present a 68% confidence interval around the retrieved value.

The new CTO product, based on a different methodology compared to the one previously used for CLARA-A2, gives significantly higher cloud top altitudes for high-level clouds and a reduction of altitudes for low-level boundary layer clouds trapped in inversions. Thus, the documented (from validation activities) CLARA-A2 underestimation of high-level cloud altitudes and overestimation of low-level cloud altitudes are largely being addressed and removed in the CLARA-A3 data record. Furthermore, since the CTO product is used as input to products CPH and JCH, improvements are also expected for those products.

More information on the PPS CTTH retrieval can be found in [RD 2]. The reader is also referred to Håkansson et al. (2018) for more information.

3.4 Cloud Phase [CM-11042, CM-6041, CPH]

The cloud phase retrieval is based on a number of threshold tests using AVHRR channels 3a, 3b, 4 and 5, and represents the thermodynamic phase of the particles near the cloud top, with shorter wavelengths penetrating slightly deeper into the cloud. The algorithm is run for cloudy pixels and initially yields one of the following cloud types: fog, liquid, supercooled, opaque ice, cirrus, and overlap. These initial cloud types are then further condensed to liquid (former three) and ice (latter three) phase. Separate retrieval schemes are applied during daytime and nighttime. Since the implementation of the algorithm uses information from additional channels during daytime, compared to the night time retrieval, continuity cannot be ensured. For this reason, the CPH data record is also provided separately for day and night. An extensive motivation for and description of the several spectral tests is given in Pavolonis and Heidinger (2004) and Pavolonis et al. (2005). Details can be found in [RD 3].

3.5 Liquid Water Path [CM-11052, CM-6051, LWP]

For pixels to which the liquid phase has been assigned, liquid water path (LWP) is derived from the cloud optical thickness (COT or τ) and particle effective radius (CRE or re). The τ - r_e retrieval scheme, developed at KNMI, uses a pair of satellite radiances at wavelengths in the non-absorbing (for clouds) visible channel at 0.6 µm and the moderately absorbing solar infrared part of the spectrum in either the 1.6 µm or 3.7 µm channels, following methods introduced by, e.g., Nakajima and King (1990). The observed radiances are iteratively matched with values simulated with the Doubling Adding KNMI (DAK, De Haan et al., 1987) radiative transfer model and stored in lookup tables (LUTs), yielding τ between 0.1 and 150 and r_e between 3 and 34 µm. Scattering properties were calculated with Mie theory for spherical droplets in size distributions with an effective variance of 0.1. Maximum solar and satellite zenith angles for which retrievals are performed are 84 degrees.

Liquid water path is then computed from the retrieved τ and r_e values by as in Stephens 1978): LWP = 2/3 ρ_l τ r_e, in which ρ_l represents the density of liquid water (1000 kg m-3). In addition, also the cloud droplet number concentration (CDNC) and the cloud geometrical thickness (CGT) are estimated from the determined COT and CRE, following a model termed Idealized Stratiform Boundary Layer Cloud (ISBLC) by Bennartz and Rausch (2017). This model assumes a linearly increasing Liquid Water Content (LWC) with height and a vertically constant CDNC. It yields formulas for the calculation of CDNC and CGT, which are based on COT, CRE, and parameters related to the model assumptions, including the rate of increase of LWC with height for moist adiabatic ascent, and the so-called adiabatic factor, to obtain the actual rate of LWC increase with height. Further details are provided in Section 4.1 of [RD 3]. It should be stressed that, since CDNC and CGT calculations are based on COT and CRE values, the two former parameters should not be treated as independent from the two latter.

The level-2b product files do not only contain LWP but also COT, CRE, CDNC, and CGT as additional layers. Moreover, uncertainty estimates of these parameters are given. The error sources taken into account include the sensor measurements, reflectance of the land/ocean surface, atmospheric absorption and thermal emission (for channel 3b). Details are provided in Section 4.4. of [RD 3].

3.6 Ice Water Path [CM-11062, CM-6061, IWP]

For pixels to which the ice phase has been assigned, ice water path (IWP) is retrieved in the same way as LWP described in Section 3.5. The effective radius for ice particles is defined by the ratio of the mean particle volume to the mean projected cross-sectional area, which applies for both spherical and non-spherical particles (e.g., Schumann et al., 2011). This equivalent definition ensures the validity of the cloud water path equation for both liquid and ice clouds. For ice particles, the scattering properties were calculated based on severely roughened compact aggregates of solid columns (Yang et al., 2013; Baum et al., 2011), with effective radii between 5 and 60 µm and size distributions with an effective variance of 0.1. IWP is then computed with the same formula as LWP but with the density of ice (930 kg m-3). More details on the retrieval scheme can be found in [RD 3]. As for LWP, COT and CRE are provided as additional layers in the product files, and uncertainty estimates of these parameters are given.

3.7 Summary of product limitations

The following table lists the main limitations of each product.

Table 3.1: Main limitations product by product.

Product	Limitation
CFC	- The efficiency of cloud detection depends on illumination conditions and the availability of spectral channels. Consequently, results are somewhat degraded (i.e., CFC is underestimated) in the twilight zone and for night time or winter conditions over high-latitudes and over the polar regions - The accuracy of geographical information (i.e., land mask and topography) influences the choice of correct surface categories for the CMa-prob method. This may lead to some unrealistic regional CFC values (best example is seen over the Aral Sea).
CTO	- As for all sub-sequent products, a dependency on the CFC (cloud mask) product exists. Falsely detected clouds will consequently have erroneous values and non-detected cloud will lack a proper cloud top determination. - In areas with very low CFC (e.g., some desert regions), the CTO L3 product may produce a noisy pattern (because of being based on on only a small number of valid observations).
CPH	- For CPH and all the following products, slightly different results are achieved depending on whether the 1.6 µm channel (only available on satellites NOAA-16, NOAA-17 and the Metop satellites) or the 3.7 µm channel has been used.
LWP	- Only available during daytime (i.e., for solar zenith angles below 84 degrees in level-2b products and below 70 degrees in level-3 products). - The product is difficult to retrieve over surfaces with high surface reflectance (e.g., over snow and ice) where values are often overestimated.
IWP	- Only available during daytime (i.e., for solar zenith angles below 84 degrees in level-2b products and below 70 degrees in level-3 products).

	- The product is difficult to retrieve over surfaces with high surface reflectance (e.g., over snow and ice) where values are often overestimated.
JCH	- Only available as level-3 product composed from daytime observations (i.e., for solar zenith angles below 70 degrees). - Since this product is based on accumulated statistics of pair-wise occurrences of CTP and COT for either liquid or ice clouds, the resolution is coarser (1-degree grid) than for the other products.

3.8 ICDR specific adaptions in ancillary, auxiliary and NWP data used

Supporting the retrieval algorithms, a selected set of external data is utilized spanning surface albedo, surface emissivity, NWP thermodynamic profiles and surface data, sea-ice/snow-cover information. Table 3.2 gives an overview of the external data used for the CDR and ICDR. More detailed information can be found in [RD 1], [RD 2] and [RD 3].

Table 3.2: Summary of external data used as input to the retrieval algorithms and how they deviate between CDR and ICDR. ERA5 is the ECMWF Re-Analysis 5 and ERA5T its low-latency version. OSI SAF is the Satellite Application Facility on Ocean and Sea Ice.

	CDR	ICDR
Surface albedo	based on snow-free gap-filled MODIS data + aggregated and complemented with ERA5 snow information	based on snow-free gap-filled MODIS data + aggregated and complemented with ERA5T snow information
Surface emissivity	Monthly climatologies based on MODIS	Monthly climatologies based on MODIS
NWP data	ERA5 skin temperature and profiles of temperature, moisture, geopotential height Snow cover information (interpreted from snow depth)	ERA5T skin temperature and profiles of temperature, moisture, geopotential height Snow cover information (interpreted from snow depth)
sea-ice	OSI-SAF-450 product and its extension OSI-430b and ERA5	near-realtime product OSI-401b and ERA5T
Level-1 FDR	EUMETSAT	Generated by CM SAF (using the same procedures as was used for creating the EUM FDR).

4 Generation of final products (level-2b & level-3)

Following the findings and the experiences obtained producing the earlier editions of the CLARA data record, the theory and implementation of level-3 generation plays a larger part within the processing for the current edition. This covers the transformation from level-2 data, native satellite projection, into global composites (level-2b) by (sub-) sampling, with level-2b data being the input to daily mean and histogram data (level-3), which are further on used as input to monthly products (level-3). An overview of level-2b and level-3 products is given in Table 4.1.

Table 4.1: The CLARA-A3 products feature a wide range of global composites data (level-2b), daily and monthly averages as well, as monthly histograms, with liquid and ice cloud separation done for some products. Day and night separation are indicated for products having such a distinction but it should be noted that LWP and IWP products (and as a consequence also JCH) are only retrieved during daytime. Level-2b refers to the non-averaged, pixel-based cloud retrievals sampled onto a global lat/lon grid. All products listed exist separately for each satellite, but also aggregated as so-called 'AVPOS' products. The AVPOS products combine all available satellites, while for CPH day, LWP, IWP and JCH only the satellites having channel 3b active during daytime are included.

	level 2b global[1] 0.05° lat/lon grid	**Daily mean global**[3] 0.25° lat/lon grid	**Monthly mean global**[2] 0.25° lat/lon grid	**Monthly histograms global**[2] 0.25° lat/lon grid	**Daily mean Polar**[3] 25 km EASE grid	**Monthly mean Polar**[3] 25 km EASE grid
CFC	✓ as CMA*	✓ day/night high/mid/low	✓ day/night high/mid/low	-	✓ day/night high/mid/low	✓ day
CTO (CTH, CTP, CTT)	✓	✓ day/night liquid/ice	✓ day/night liquid/ice	✓ liquid/ice	✓ day/night liquid/ice	✓ day liquid/ice
CPH	✓	✓ day/night	✓ day/night	-	-	-
LWP (+τ, r_e, CDNC, CGT)	✓ as CWP	✓	✓	✓ as CWP	-	-
IWP (+τ, r_e)	✓ as CWP	✓	✓	✓ as CWP	-	-
JCH	N/A	N/A	N/A	✓ liquid/ice (1.0° grid)	N/A	N/A

** CMA is a level-2b binary cloud mask that also contains a data layer the probabilistic cloud mask.*

[1] Products exist for individual satellites only

[2] Products exist for individual satellites and as AVPOS (all satellites combined) product

[3] Products exist as AVPOS (all satellites combined) product only

CFC and CPH monthly histograms are not produced, since these two parameters are given in binary form in level-2b (cloudy/clear and liquid/ice, respectively). Furthermore, liquid and ice water path are not included in polar grids; their day-only retrieval would lead to data sets mostly gapped due to polar winters. The same is true for CPH, which is divided in day/night products. Monthly mean products on polar grid contain day-time fields only.

Table 4-1 also indicated the grids of the products with more information on these given in Sections 4.3 & 4.4.

Different quality checks are applied before sampling swath-based level-2 data into level-2b products, including filtering valid numbers for each parameter. This is motivated by the fact that not in all cases, where a cloud has been detected, all cloud retrieval results of CPH, CTO, COT, CRE, LWP, IWP, CDNC and CGT are available. Also, some filtering was applied before aggregating data to level-3 and in the daily to monthly mean averaging step. The specifics for filtering are summarized in Table 4.2 and Table 4.3 and also detailed in the following subsections.

Table 4.2: Overview of filtering applied for derivation of level-2b from level-2 products.

Product	Filtering	Explanation
CPH day, LWP (and additional products), IWP (and additional products)	*Exclude* pixels satisfying ALL following conditions: - surface type is water - scattering angle differs less than 27° from direct glint angle - satellite viewing zenith angle > 30°	Avoid possibly sunglint-affected conditions, which mainly give a risk of erroneous cloud optical/microphysical retrievals at higher viewing zenith angles
CRE (liquid and ice), CDNC (liquid), CGT (liquid)	*Exclude* pixels with CRE retrieval outside look-up table. This concerns in the order of 10% of the cloudy pixels.	Avoid these flagged L2 retrievals to enter L3. Note: corresponding LWP, IWP and COT retrievals are not excluded.
All products for specific satellites: - NOAA-15 - NOAA-16 - Metop-B	*Exclude* pixels satisfying ALL following conditions: - solar zenith angle < 84° - surface type is land or sea ice - skin temperature < 260K - brightness temperature measured by channel 4 < 255K	An incorrect calibration of ch3b at the cold end of brightness temperatures (removing very cold, dark areas, mostly located at high latitudes in winter months). See section 2.3.3

Table 4.3: Overview of filtering applied for calculation of level-3 products.

Product	Filtering	Explanation
All daytime products (CFC day, CTH/CTP/CTT day, CPH day, LWP (and additional products), IWP (and additional products), JCH	solar zenith angle < 70°	Definition of daytime (avoid low solar elevations because of potential retrieval problems)
All night-time products (CFC night, CTH/CTP/CTT night, CPH night)	solar zenith angle > 95°	Definition of night-time (avoid twilight conditions)
CFC low	CTP > 680 hPa	ISCCP definition of low-level clouds
CFC middle	440 hPa < CTP < 680 hPa	ISCCP definition of mid-level clouds
CFC high	CTP < 440 hPa	ISCCP definition of high-level clouds
All global monthly mean products	number of valid daily means per grid box for aggregating to monthly mean >= 20	representative monthly average (avoid monthly means based on few observations)
CFC/CPH polar monthly mean products	*Exclude* night-time observations	representative monthly average (avoid monthly means based on few observations)

4.1 Definition of product specifications

The CM SAF CLARA-A3 cloud data record from AVHRR provides global coverage of a number of cloud parameters. AVHRR GAC retrievals at native swath projection are firstly subsampled onto the level-2b grid, which are then aggregated to the spatio-temporally averaged daily and monthly data and histograms. The products are available as daily and monthly averages for each satellite separately on a regular latitude/longitude grid with a spatial resolution of 0.25° × 0.25° degrees. For the polar region an EASE grid with spatial resolution of 25 km is used. The sampled level-2b product is defined with a spatial resolution of 0.05° x 0.05° for the global area on a daily basis for each satellite with separate layers for the ascending and descending node. All monthly and daily averages are also available in aggregated form (i.e., merging all satellites). Acknowledging the different observation capabilities during night and during day and also taking into account existing diurnal variations in cloudiness, a further separation of results into daytime and night-time portions has also been done for fractional cloud cover, cloud top phase and cloud top properties. For any level-3 product, daytime observations are defined by a solar zenith angle of less than 70°. Nighttime observations are defined by a solar zenith angle greater than 95°. Any observation with a solar zenith angle between them is considered as twilight observation. Only for the level-2b products, cloud physical properties are available for solar zenith angles with up to 84°.

In addition to the daily and monthly means, histograms are provided on monthly time scales. The Joint Cloud property Histograms are three-dimensional histograms of COT, CTP and CPH and are composed with a spatial resolution of 1° x 1° degrees. (See Section 4.4.2 for more technical details). For CTP, CTT, CWP (cloud liquid/ice water path), COT and REF additionally two-dimensional histograms (each product is separated into liquid and ice clouds) are constructed on a monthly basis with a spatial resolution of 0.25°x0.25° degrees (See Sections 4.4.3, 4.4.4 , 4.4.5, and 4.4.6 for more technical details.).

4.2 Definition of the grids

The cloud products are defined on three different grids. One is the traditional global equal angle grid, the second is a spatially high-resolution equal angle grid, and the third is the polar equal area grid. The elongated shape of Greenland suggested that the area should be expanded slightly so that the entire Greenland area could be included. This explains the slightly different sizes of the two polar grids.

4.2.1 Global equal angle grid

Level-3 daily and monthly means are presented on a global equal angle grid. Each grid box has the size of $(0.25°)^2$. Hence, the full global area has the dimension of 1440x720 grid boxes. The same grid is used for the two-dimensional histograms. For the three-dimensional histograms, this grid is coarser with a spatial resolution of $(1°)^2$.

4.2.2 Polar grid

The grid for the polar region is an equal area grid centered at the north or south pole with a spatial resolution of $(25\text{ km})^2$. The polar grid is defined on the basis of the existing “AVHRR Polar Pathfinder 25 km EASE-grid Composites”. EASE-Grid has a different area coverage for the northern and southern hemisphere. On the northern hemisphere the grid extends from 90°N to 48.4°N and on the southern hemisphere from 53.2°S to 90°S. The relation of the coordinates (latitude and longitude) and the field indices is for the northern hemisphere

$$x = -2\frac{R_{Earth}}{C_{Cell\,size}}\cos(longitude)\sin\left(\frac{\pi}{4}-\frac{latitude}{2}\right)+x_{max} \tag{1}$$

$$y = 2\frac{R_{Earth}}{C_{Cell\,size}}\sin(longitude)\sin\left(\frac{\pi}{4}-\frac{latitude}{2}\right)+y_{max} \tag{2}$$

and for the southern hemisphere

$$x = \quad 2\frac{R_{Earth}}{C_{Cell\,size}}\cos(longitude)\cos\left(\frac{\pi}{4}-\frac{latitude}{2}\right)+x_{max} \tag{3}$$

$$y = \quad 2\frac{R_{Earth}}{C_{Cell\,size}}\sin(longitude)\cos\left(\frac{\pi}{4}-\frac{latitude}{2}\right)+y_{max}\ . \tag{4}$$

R_{Earth} is the radius of the earth in km (here set to 6371.228 km), $C_{Cell\ size}$ is the size of one cell in km (here 25 km), and x_{max} and y_{max} denote the number of grid boxes in the x or y direction (321 for the southern and 361 for the northern hemisphere).

4.2.3 Level-2b

Since the second edition of the CLARA data record the cloud products are provided on an additional grid, level-2b. The level-2b projects the orbital observations onto a global grid with nearly nadir sensor resolution. It is available on a daily basis as global composites.

The level-2b data representation is motivated by the inhomogeneous global coverage of polar sun-synchronous satellite data. Each polar satellite offers 14 observations per day for each location near the pole (evenly distributed over the day) while when passing the equator each location is only observed twice (approximately 12 hours apart). The idea with the introduction of the level-2b data representation is to form a more homogeneous data record having only two observations per day per satellite for each location globally. The alternative to use all observations for level-3 products results in a very skew distribution of the observations because of the inhomogeneous observation frequency (increasing with latitude). By selecting only the observations which are made closest to zenith (the NADIR condition) we ensure that observations are made at almost the same viewing conditions and, most importantly, observations are made at nearly the same local time globally for each level-2b product. In this way, the restricted level-2b products are easier to deal with for certain applications compared to the full set of observations. This concerns in particular the use in COSP simulators aiming at reproducing satellite datasets from Climate Model data.

The level-2b approach leads obviously to a significant reduction of the amount of used observations. However, the high observation frequency near the poles is undoubtedly very valuable and consequently there are also polar products added which uses all available observations

Level-2b is defined on a global equal angle grid with a spatial resolution of $(0.05°)^2$. This is close to the AVHRR GAC pixel resolution, which is in nadir about $(4km)^2$. A further advantage of the resolution is that the level-3 grid resolution (0.25°) is a multiple of the level-2b grid resolution. This facilitates the generation of the level-3 products based on level-2b. In contrast to the level-3 products, level-2b separates the orbit into ascending and descending node. Thus, multiple observations of the same area, i. e. daylight and night, can be distinguished.

The sampling of the data onto the level-2b grid causes special handling of some effects to ensure assigning each grid box one observation. These effects are:

1) If more than just one pixel fall into most grid boxes, when observing close to satellite nadir. This concerns about 200 pixels per AVHRR scan line;
2) Mostly in the polar region, but also at lower latitudes between about 55°N/S and the pols, if AVHRR orbits are overlapping. Thus, for these regions several observations exist during each day and node for single satellites. For usual days each area can be observed by one satellite between 2 times at about 55°N/S and up to 14 times at the poles.
3) Towards the edge of the swath, when AVHRR the pixel size exceeds the size of the grid box. Thus, these pixels must be duplicated into all grid boxes that are covered by their FOV. Also, at high latitudes the width of a grid box decreases in size as an effect of projecting an equal angle grid onto a sphere. Here, the longitudes converge at the polar singularity of the grid.

To avoid that these effects result in errors or empty grid boxes in the level-2b grid special data handling is done:

1) The actual foot print size of each pixel FOV on the earth surface is considered. Therefore, the pixel size and the grid box size are calculated. The grid box size varies strongly with latitude and has the largest size near the equator. Thus, in polar areas a satellite pixel FOV covers more than one grid box, while the nadir resolution of AVHRR GAC at the equator is of nearly the same size as the level-2b grid box size. If one pixel covers more than one grid box, a duplication of the satellite pixel observation is performed to cover each grid box. To locate each covered grid box, that is observed by the satellite pixel, it is assumed, that only in the across track dimension the size of the pixel field of view changes. Within the half distance to both neighboring pixels, to the left and the right, all grid boxes are identified and filled with the pixel observation.
2) From multiple observations, i.e. around nadir or at overlapping swath edges or in case of multiple over passes, the nearest nadir pixel is chosen. This is done because of the better retrieval performance at lower viewing angles.

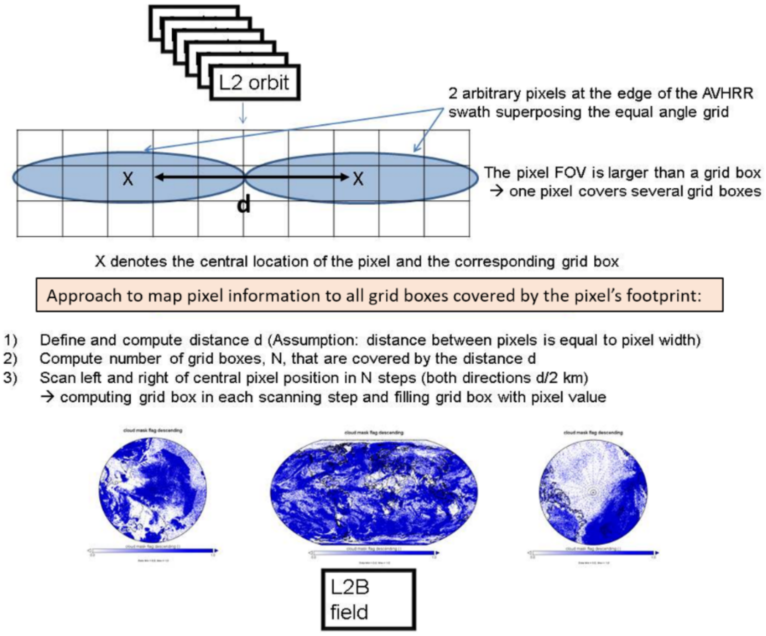

Figure 4-1: From level-2 to level-2b. How to deal with mapping effects

Figure 4-1 outlines how this is implemented. First, for each grid box only a single observation is used, sampled. So, in the level-2b no averaging takes place. The representation of the pixel field of view onto the spatial area of a grid box is done by comparing the distance between two neighboring pixels and the size of the collocated grid box. Both are a function of latitude. Very close to the poles a single AVHRR pixel can cover more than hundred grid boxes. This complex treatment becomes necessary as the grid resolution is close to the nadir sensor resolution and AVHRR observes the whole Earth, so no grid box should remain empty.

4.3 Calculation of level-3 products (daily and monthly averages and histograms)

The level-2b products are the starting point for the aggregation to daily means. Due to the level-2b sampling technique a globally nearly constant number of observations per grid box can be achieved in the daily mean products, which themselves are input for the aggregation to the monthly mean products. Monthly mean standard deviation is defined as the standard deviation over the daily means.

This method for defining level-3 products follows the GEWEX Cloud Assessment approach (Stubenrauch et al., 2012). It should be noticed that for LWP and IWP products the daily mean is calculated as both in-cloud and all-sky values which then leads to two different versions of the monthly mean. Since the monthly mean is calculated from the daily means it means that the in-cloud version will give a relatively larger weight to days with low cloud amounts while the all-sky version will give equal weight to all days in a month.

For the case of the polar grids it was decided to choose an alternative averaging. For two major reasons it was decided to average all orbits directly from the level-2 native satellite projection onto the final 25km polar equal area grid:

A) the overlapping of the different orbits at high latitudes is a clear strength of AVHRR
B) the equal area grid centered on the pole is not affected by the decreasing grid box area, described in level-2b Section.

The latter is only an issue for equal angle grids close to the poles. Thus, the polar grid acts as a beneficial supplement to the global product in terms of describing cloud parameters close to the poles.

4.4 Short overview on level-3 cloud properties

4.4.1 Fractional Cloud Cover [CFC]

The Fractional Cloud Cover product contains mean fractional cloud cover *CFC* and mean probabilistic cloud mask *CMAPROB*.

The daily mean probabilistic cloud mask is a result of the temporal averaging of probabilistic cloud mask values. All valid entries of the level-2b fields (both ascending and descending) are aggregated and then weighted by the number of cloudy and clear cases.

In more detail, the daily mean fractional cloud cover is calculated from the aggregation of the instances of the level-2b binary cloud mask information.as follows:

Equation 4-1
$$CFC(i,j) = \frac{N(i,j)_{Cloudy}}{N(i,j)_{Cloudy} + N(i,j)_{Clear}}$$

with i and j being the indices of the level-3 grid (0.25° resolution), $N(i,j)_{Cloudy}$ the number of level-2b cloudy cases and $N(i,j)_{Clear}$ the number of level-2b clear cases that fall into the (*i,j*) grid cell .

Acknowledging the different cloud detection capability during day and night time, an additional separation into day time and night time averages is done for CFC and CMAPROB. Here, the solar zenith angle of ≤70° and ≥95° are used to define day and night, respectively. Cases with solar zenith angles between 70° and 95° are included in the nominal daily mean but excluded when collecting data for day-only and night-only averages. Mean fractional cloud cover in pre-defined vertical layers is also provided separately. This is calculated in different pressure intervals, which are defined by Cloud Top Pressure. CFC_low for CTP > 680 hPa, CFC_middle for 440 hPa < CTP < 680 hPa, CFC_high for CTP < 440 hPa.

The monthly mean cloud fractional cover and mean probabilistic cloud mask are calculated as mean over the daily means with each day being weighted equally.

4.4.2 Joint Cloud property Histogram [JCH]

The JCH does not include a classical mean of a specific cloud property but covers the solution space for the 3 cloud parameters: COT, CTP and CPH. Hence, the JCH product is only available for daytime observations, defined as observations with solar zenith angle lower than 70°. This product is described in a five-dimensional field $JCH(i,j,t,p,ph)$. Indices i and j refer to location space, while t and p being the indices for specific bins of the range of occurring COT and CTP values, and *ph* denotes the cloud phase.

Each specific field entry contains the absolute number of cloud pixels with phase *ph* falling into the COT bin t and the CTP bin p. The following values bordering the bins of COT and CTP (given in hPa) have been defined. Note that, while in the L2 output, COT values are maximized at 150, this variable can in principle take arbitrarily large values, which is reflected by setting the upper bound in the histogram to infinity.

COT:{ 0, 0.3, 0.6, 1.3, 2.2, 3.6, 5.8, 9.4, 15, 23, 41, 60, 80, 149.99, inf} and

CTP:{ 1, 90, 180, 245, 310, 375, 440, 500, 560, 620, 680, 740, 800, 875, 950, 1100} [hPa].

These histograms are calculated for liquid and ice clouds separately, thus:

Equation 4-2
$$JCH(i,j,t,p)_{ice} = N(i,j)_{COT\in COTbin;CTP\in CTPbin;CPH=ice}$$

and

Equation 4-3
$$JCH(i,j,t,p)_{liquid} = N(i,j)_{COT\in COTbin;CTP\in CTPbin;CPH=liquid}$$

and merged into one field (JCH(i,j,t,p,ph)).

4.4.3 Cloud Phase [CPH]

Similarly to CFC, the daily averages of CPH are calculated by temporal averaging of retrieval results on original pixel basis and subsequent remapping. CPH is expressed as fraction of liquid water clouds by calculating the ratio of number of detected liquid clouds $N(i,j)_{Cloudy}$ with respect to the total number detected clouds $N(i,j)_{Cloudy}$:

Equation 4-4
$$CPH(i,j) = \frac{N(i,j)_{liquid}}{N(i,j)_{Cloudy}}$$

Cloud Phase is also provided as separate daytime and night-time averages.

The monthly mean CPH is calculated as mean over the daily means with each day being weighted equally.

4.4.4 Cloud Top Level [CTO]

The CTO product contains daily means for CTH, CTP, and CTT. For these parameters all valid entries of the original fields are aggregated and then weighted by the number of used entries.

Equation 4-5
$$< x(i,j) >= \frac{1}{N(i,j)_{Cloudy}} \sum_{k=1}^{N(i,j)_{Cloudy}} x_k(i,j)$$

with $x(i,j)$ being a general expression for CTH, CTP and CTT at a specific original grid cell.

After temporal averaging, the fields are remapped to the final resolution as described in Section 4.2.

For CTP, an alternative way of averaging is followed and additionally calculated and provided as geometrical mean where the variables are averaged in logarithm space:

Equation 4-6
$$< ctp(i,j) >_{\ln} = \exp(\frac{1}{N(i,j)_{Cloudy}} \sum_{k=1}^{N(i,j)_{Cloudy}} \ln(ctp_k(i,j)))$$

Geometrical mean is added to keep consistency between CTH and CTP. CTP depends logarithmically on CTH, so if CTH is averaged linearly, CTP has to be averaged logarithmically to preserve the relation.

An additional separation due to the cloud thermodynamic phase as well as between day and night leads to four further variables: $<x(i,j)>_{liq\_day}$, $<x(i,j)>_{ice\_day}$, $<x(i,j)>_{liq\_night}$, $<x(i,j)>_{ice\_night}$. The cloud thermodynamic phase is defined by CPH product, daytime corresponds to solar zenith angle lower than 70°, night-time - higher than 95°.

The monthly mean cloud top level parameters are calculated as mean over the daily means with each day being weighted equally.

One-dimensional monthly histograms are generated for CTP and CTT on the same spatial resolution, but only on monthly basis. The bin borders for these histograms are:

CTP:{ 1, 90, 180, 245, 310, 375, 440, 500, 560, 620, 680, 740, 800, 875, 950, 1100} [hPa].

CTT;{160, 200, 210, 220, 230, 235, 240, 245, 250, 255,260, 265, 270, 280, 290, 300, 310, 350} [K]

4.4.5 Liquid Water Path [LWP]

Daily mean LWP is calculated for each grid cell <LWP(i,j)> as given in Equation 4-5. Liquid clouds COT, CRE, CDNC and CGT are aggregated in the same way, and provided as additional data layers. Furthermore, for COT a logarithmic average is included, which is more consistent with the cloud radiative effect.

In addition to the in-cloud averages, all-sky LWP is calculated by a weighting of the LWP with the CFC including cloud-free and ice-cloud pixels as zeroes in the averaging.

The in-cloud and all-sky monthly mean LWP as well as the monthly mean of COT, CRE, CDNC and CGT of liquid clouds are calculated as mean over the daily means with each day being weighted equally.

One-dimensional histograms are generated for LWP and for COT, CRE, CDNC and CGT of liquid clouds on the same spatial resolution, but only on monthly basis. These histograms are described in a 4-dimensional field *LWP(i,j,bin,ph).* Indices *i* and *j* refer to location space, while *bin* is the index for specific bins of the range of occurring values, and *ph* denotes the cloud phase. Practically, these histograms are included in the liquid dimension of the CWP, COT, CRE, CDNC, CGT histograms with the following bin borders. Note that, as explained in Section 4.4.2, COT can in principle acquire arbitrarily large values and therefore the upper histogram bin was chosen to extend to infinity. The same is true for CWP, CDNC and CGT, since these are proportional to COT to some power.

CWP: {0, 5, 10, 20, 35, 50, 75, 100, 150, 200, 300, 500, 1000, 2000, inf} [g/m²]

COT: {0.0, 0.3, 0.6, 1.3, 2.2, 3.6, 5.8, 9.4, 15.0, 23.0, 41.0, 60.0, 80.0, 149.99, inf}

CRE: {3, 6, 9, 12, 15, 20, 25, 30, 40, 60} [µm].

CDNC: {0, 2, 5, 10, 20, 50, 100, 150, 200, 300, 500, inf} [cm-3]

CGT: {0, 50, 100, 150, 250, 350, 500, 700, 1000, 1500, 2000, inf} [m].

Note: CDNC and CGT are retrieved for liquid clouds only but contain a phase dimension for consistency.

4.4.6 Ice Water Path [IWP]

Daily mean IWP, as well as the ice cloud COT and CRE, are calculated in exactly the same way as LWP. This includes the computation of all-sky IWP and COT by including cloud-free and liquid-cloud pixels as zeroes in the averaging.

The in-cloud and all-sky monthly mean IWP as well as the monthly mean of COT and CRE of ice clouds are calculated as mean over the daily means with each day being weighted equally.

One-dimensional histograms are generated for IWP and for COT and CRE of ice clouds on the same spatial resolution, but only on monthly basis. Practically, these histograms are included in the ice dimension of the CWP, COT and CRE histograms with the following bin borders:

CWP: {0, 5, 10, 20, 35, 50, 75, 100, 150, 200, 300, 500, 1000, 2000, inf} [g/m²]

COT: {0.0, 0.3, 0.6, 1.3, 2.2, 3.6, 5.8, 9.4, 15.0, 23.0, 41.0, 60.0, 80.0, 149.99, inf}

REF: {3, 6, 9, 12, 15, 20, 25, 30, 40, 60} [µm].

4.4.7 Additional statistical parameters

In addition to the daily and monthly mean values, the standard deviations S over all valid and used values is calculated for CFC, CPH, CTO, LWP and IWP for each grid box with

Equation 4-7
$$\sigma_{\mathrm{std}}^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(x_i - \langle x\rangle)^2$$

4.4.8 Uncertainty propagation

The propagation of CTP, CTH, CTT, COT, CRE, LWP and IWP uncertainties into level-3 products is implemented following Stengel et al. (2017). To recall, the reported pixel-based uncertainties σ_x for a given variable x represent the 68% confidence interval that the true value is within $x \pm \sigma_x$ This confidence interval can be propagated into daily mean level-3 products. For this, the standard deviation (Equation 4-7), the daily mean uncertainty (Equation 4-8) and the daily mean squared uncertainty (Equation 4-9) were calculated using the same pixels used for the calculation of the daily mean (Number: N, pixel index: i). Using Equation 4-10 the natural variability of a property can be determined, and Equation 4-11 reports how the uncertainty of the mean ($\sigma_{\langle x\rangle}$) can then be calculated. Note, that the uncertainty correlation *c* needs to specified. As this is usually not known, we used *c*=0.1 and *c*=1.0 to reflect two rather extreme scenarios. Thus, the uncertainties of the daily means are given for both uncertainty correlation estimates and again presents the confidence interval $\langle x\rangle \pm \sigma_{\langle x\rangle}$ in which the true mean lies with a confidence of 68% around the mean.

Equation 4-8
$$\langle\sigma_i\rangle = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(\sigma_i)$$

Equation 4-9
$$\langle\sigma_i^2\rangle = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(\sigma_i^2)$$

Equation 4-10
$$\sigma_{natural}^2 = \sigma_{std}^2 - (1-c)\langle\sigma_i^2\rangle$$

Equation 4-11
$$\sigma_{\langle x\rangle}^2 = \frac{1}{N}\sigma_{natural}^2 + c\langle\sigma_i\rangle^2 + (1-c)\frac{1}{N}\langle\sigma_i^2\rangle$$

Based on the rationale above for daily mean products, the uncertainties for monthly mean products can be determined equivalently by replacing σ_i, which was the pixel-based (level-2) uncertainty, with $\sigma_{\langle x\rangle}$, which is the uncertainty of the daily mean, in Equation 4-7 to Equation 4-11 to express the uncertainty of the monthly mean ($\sigma_{\langle\langle x\rangle_j\rangle}^2$); see Equation 4-12, with M being the number of daily means available.

Equation 4-12
$$\sigma_{\langle\langle x\rangle_j\rangle}^2 = \frac{1}{M}(\sigma_{\langle x\rangle})_{std}^2 + c\langle(\sigma_{\langle x\rangle})_j\rangle^2 + (1-c)\frac{1}{M}\langle(\sigma_{\langle x\rangle})_j^2\rangle$$

An uncertainty correlation has to be specified for the calculation of the uncertainties of the monthly means as well, and this was done as for the daily means. Thus, two scenarios *c*=0.1 and *c*=1.0 were implemented with both using the daily mean uncertainty at *c*=0.1 as input. The choice to not base the daily-to-monthly uncertainty propagation on the daily mean uncertainty at *c*=1.0 has only practical reasons to keep files sizes low and to not overwhelm users with too many implementations and data fields. Furthermore, it is believed that an uncertainty correlation of 0.1 is more realistic than 1.0. Beyond this, all input data is available to redo the uncertainty propagation from level-2b to level-3 daily and monthly data with any uncertainty correlation using the equations above. The reader can contact the CM SAF helpdesk in case any further assistance is needed.

4.4.9 Requirements on the availability of measurements

Level-3 products are generated, as mentioned above, in two configurations. The first configuration is only based on observations from single AVHRR instruments. The second configuration uses all available AVHRRs at a specific day.

To guarantee that the daily mean product is not representing artificial retrieval errors or not representative observations, at least two observations in level-2b must available to generate a daily mean for that grid box.

In the level-2b representation of the products (which is the one used when averaging) only two observations per day per satellite are available separated 12 hours apart (reflecting ascending and descending nodes). Furthermore, for each 0.25-degree grid point it is possible then to get about 25 measurements at each observation time/node since the GAC resolution is about 0.05 degrees (approximating the AVHRR GAC resolution at nadir of 4 km). Thus, in the ideal case 50 measurements per day and grid point are possible. The limit of 2 measurements as mentioned above was set as a compromise and in order to retain as much as possible of real measurements. The limit might appear to be very low but one has to consider that the cases when only a few GAC pixels are used are connected to cases where the swath edge is passing through or close to the grid point. A swath edge pixel is actually covering a much larger area than the original 4 km resolution in nadir. Thus, just a few GAC pixels might in this case cover most of the grid point area.

For level-3 global monthly products there is a requirement of at least 20 valid daily means per grid box for averaging to monthly means.

5 References

Baum, B. A., P. Yang, A. J. Heymsfield, C. G. Schmitt, Y. Xie, A. Bansemer, Y.-X. Hu, and Z. Zhang, 2011: Improvements in shortwave bulk scattering and absorption models for the remote sensing of ice clouds.

Bennartz, R. and J. Rausch, 2017: Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations, Atmos. Chem. Phys., 17, 9815-9836, doi:10.5194/acp-17-9815-2017.

De Haan, J. F., P. Bosma, and J. W. Hovenier, 1987: The adding method for multiple scattering calculations of polarized light, Astron. Astrophys., 183, 371-391.

Heidinger, A. K., 2018: Climate Algorithm Theoretical Basis Document (C-ATBD): Fundamental Climate Data Record (CDR) of Reflectance from AVHRR Bands 1, 2 and 3a, CDR Program Document Number CDRP-ATBD-0184, Revision 2.

Heidinger, A.K., W.C. Straka, C.C. Molling, J.T. Sullivan and X.Q. Wu, 2010: Deriving an inter-sensor consistent calibration for the AVHRR solar reflectance data record. Int. J. Rem. Sens.,31(24), 6493-6517.

Hess, H, R. B. A. Koelemeijer, and P. Stammes, 1998: Scattering matrices of imperfect hexagonal crystals. *J. Quant. Spectrosc. Radiat. Transfer*, **60**, 301–308.

Ignatov, I.L., E.D. Harrod, K.B. Kidwell and G.P. Goodrum, 2004: Equator crossing times for NOAA, ERS and EOS sun-synchronous Satellites, Int. J. Rem. Sens.,25 (23), 5255–5266, DOI: 10.1080/0143116041000171298

Karlsson, K.-G., & Håkansson, N., 2018: Characterization of AVHRR global cloud detection sensitivity based on CALIPSO-CALIOP cloud optical thickness information : demonstration of results based on the CM SAF CLARA-A2 climate data record. *Atm. Mea. Tech.*, **11**(1), 633–649. https://doi.org/10.5194/amt-11-633-2018

Karlsson, K.-G., Håkansson, N., Mittaz, J. P. D., Hanschmann, T., & Devasthale, A., 2017: Impact of AVHRR Channel 3b Noise on Climate Data Records : Filtering Method Applied to the CM SAF CLARA-A2 Data Record. *Rem. Sens.*, **9**(6). https://doi.org/10.3390/rs9060568

Karlsson, K.-G., E. Johansson and A. Devasthale, 2015: Advancing the uncertainty characterisation of cloud masking in passive satellite imagery: Probabilistic formulations for NOAA AVHRR data, Rem. Sens. Env., 158, 126-139; doi:10.1016/j.rse.2014.10.028.

Karlsson, K.-G., E. Johansson, N. Håkansson, J. Sedlar, S. Eliasson, 2020: Probabilistic Cloud Masking for the Generation of CM SAF Cloud Climate Data Records from AVHRR and SEVIRI Sensors. Remote Sens.12, 713. https://doi.org/10.3390/rs12040713

Mittaz, P.D. and R. Harris, 2009: A Physical Method for the Calibration of the AVHRR/3 Thermal IR Channels 1: The Prelaunch Calibration Data. J. Atmos. Ocean. Tech., 26, 996-1019, doi: 10.1175/2008JTECHO636.1.

Nakajima, T., and M. D. King, 1990: Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part 1: Theory. *J. Atmos. Sci .*, **47**, 1878-1893.

Pavolonis, M. J. and A. K. Heidinger, 2004: Daytime cloud overlap detection from AVHRR and VIIRS, *J. Appl. Meteorol.*, **43**, 762-778.

Pavolonis, M. J., A. K. Heidinger, and T. Uttal, 2005: Daytime global cloud typing from AVHRR and VIIRS: Algorithm description, validation, and comparison, *J. Appl. Meteorol.*, **44**, 804-826.

Pfreundschuh, S. and Eriksson, P. and Duncan, D. and Rydberg, B. and Håkansson, N. and Thoss, A. (2018) A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems. Atmospheric Measurement Techniques. doi: 10.5194/amt-11-4627-2018. url: https://www.atmos-meas-tech.net/11/4627/2018/

Roebeling, R.A., A.J. Feijt and P. Stammes, 2006: Cloud property retrievals for climate monitoring: implications of differences between SEVIRI on METEOSAT-8 and AVHRR on NOAA-17, J. Geophys. Res., 111, D20210, doi:10.1029/2005JD006990.

Rossow, W.B. and R.A. Schiffer, 1991: ISCCP Cloud Data Products, Bull. Amer. Meteorol. Soc., 72, 1, 2-20.

Schumann, U., B. Mayer, K. Gierens, S. Unterstrasser, P. Jessberger, A. Petzold, C. Voigt, J.-F. Gayet, 2011: Effective radius of ice particles in cirrus and contrails. J. Atmos. Sci., 68, 300–321, doi:10.1175/2010JAS3562.1.

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, R. G., Meirink, J. F., Feofilov, A., Bennartz, R., Bojanowski, J. S., and Hollmann, R.: Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the Cloud_cci project, Earth Syst. Sci. Data, 9, 881-904, https://doi.org/10.5194/essd-9-881-2017, 2017.

Stephens, G. L., 1978: Radiation profiles in extended water clouds: II. Parameterization schemes. J. Atmos. Sci., 35, 2123-2132.

Stubenrauch, C. J., W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. Di Girolamo, B. Getzewich, A. Guignard, A. Heidinger, B. Maddux, P. Menzel, P. Minnis, C. Pearl, S. Platnick, C. Poulsen, J. Riedi, S. Sun-Mack, A. Walther, D. Winker, S. Zeng, G. Zhao, 2012: ASSESSMENT OF GLOBAL CLOUD DATASETS FROM SATELLITES: Project and Database initiated by the GEWEX Radiation Panel, Bull. Amer. Meteor. Soc., doi: 10.1175/BAMS-D-12-00117.

Yang, P., L. Bi, B. A. Baum, K.-N. Liou, G. W. Kattawar, M. I. Mishchenko, and B. Cole, 2013: Spectrally consistent scattering, absorption, and polarization properties of atmospheric ice crystals at wavelengths from 0.2 to 100 μm, J. Atmos. Sci., 70, 330-347, doi:10.1175/JAS-D-12-039.1.

6 Acronyms

ATBD	Algorithm Theoretical Baseline Document
AVHRR	Advanced Very High Resolution Radiometer
AVPOS	AVHRR Polar Satellites (L3 products based on all available AVHRR-carrying satellites)
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarisation
CDNC	Cloud Droplet Number Concentration
CDOP	Continuous Development and Operations Phase
CDR	Climate Data Record
CF	Climate and Forecast Metadata Convention ((http://cf-pcmdi.llnl.gov/))
CFC	Fractional Cloud Cover
CGT	Cloud Geometrical Thickness
CFOT	Cloud Feature Optical Depth
CLARA-A	CM SAF cLoud, Albedo and Radiation products, AVHRR-based
CLAAS	CM SAF cLoud dAtAset using SEVIRI
CM SAF	Satellite Application Facility on Climate Monitoring
CMAPROB	Probabilistic cloud masking method (PPS)
CNRS	Centre National de la Recherche Scientifique
COT	Cloud Optical Thickness
CPH	Cloud Phase
CPR	Cloud Profiling Radar
CRE	Cloud Particle Effective Radius
CTH	Cloud Top Height
CTO	Cloud Top product
CTP	Cloud Top Pressure
CTT	Cloud Top Temperature
CPP	Cloud Physical Properties

DAK	Doubling Adding KNMI (radiative transfer model)
DRR	Delivery Readiness Review
DWD	Deutscher Wetterdienst (German MetService)
ECMWF	European Centre for Medium Range Forecast
ECV	Essential Climate Variable
EDR	Environmental Data Record
ERA5	ECMWF Re-Analysis dataset 5
ERA5T	Low-latency version of ERA5
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
FCI	Flexible Combined Imager
FIDUCEO	Fidelity and uncertainty in climate data records from Earth Observation satellites
FMI	Finnish Meteorological Institute
FOV	Field of View
GAC	Global Area Coverage (AVHRR)
GEWEX	Global Energy and Water cycle EXperiment
ICDR	Interim Climate Data Record
ISCCP	International Satellite Cloud Climatology Project
IR	InfraRed
IWP	Ice Water Path
JCH	Joint Cloud properties Histogram
KNMI	Koninklijk Nederlands Meteorologisch Instituut
LUT	Lookup Table
LWP	Liquid Water Path
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
NIR	Near InfraRed
NOAA	National Oceanic & Atmospheric Administration

NWC SAF	SAF on Nowcasting and Very Short Range Forecasting
NWP	Numerical Weather Prediction
PATMOS-x	Pathfinder Atmospheres-Extended dataset (NOAA)
PPS	Polar Platform System (NWC SAF polar cloud software package)
PRD	Product Requirement Document
PUM	Product User Manual
RMIB	Royal Meteorological Institute of Belgium
RTTOV	Radiative Transfer model for TOVS
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SAF	Satellite Application Facility
SMHI	Swedish Meteorological and Hydrological Institute
SRF	Spectral Response Function
SYNOP	Synoptic observations
SZA	Solar Zenith Angle
CDR	Thematic Climate Data Record
VIS	Visible
VZA	Viewing Zenith Angle
WCRP	World Climate Research Programme