EUMETSAT Satellite Application Facility on Climate Monitoring

The EUMETSAT Network of Satellite Application Facilities

CM SAF Climate Monitoring

Algorithm Theoretical Basis Document

CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2)

Cloud Products (level-1 to level-3)

DOI: 10.5676/EUM_SAF_CM/CLARA_AVHRR/V002

Fractional Cloud Cover	**CM-11011**
Joint Cloud property Histogram	**CM-11021**
Cloud Top level	**CM-11031**
Cloud Phase	**CM-11041**
Liquid Water Path	**CM-11051**
Ice Water Path	**CM-11061**

Reference Number:	SAF/CM/DWD/ATBD/CLARA/CLD
Issue/Revision Index:	2.3
Date:	19.08.2016

Document Signature Table

	Name	Function	Signature	Date
Author	Karl-Göran Karlsson Timo Hanschmann Martin Stengel Jan Fokke Meirink			27.05.2016, 19.08.2016
Editor	Rainer Hollmann	Science Coordinator		19.08.2016
Approval	Steering Group			
Release	Martin Werscheck	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
1.0	24.04.2015	SAF/CM/DWD/CDOP/ATBD/22	
2.1	23.06.2015	SAF/CM/DWD/CDOP/ATBD/CLD/2.1	
2.2	27.05.2016	SAF/CM/DWD/ATBD/CLD	Version presented to DRR 2.2
2.3	19.08.2016	SAF/CM/DWD/ATBD/CLD	Implementation of RIDs from DRR 2.2

Applicable documents

Reference	Title	Code
AD 1	CM SAF Product Requirements Document	SAF/CM/DWD/PRD/2.7

Reference documents

Reference	Title	Code
RD 1	Algorithm Theoretical Basis Document SAFNWC/PPS "Cloud mask", PPS version 2014 patch 20150327 (OBS: in the text referred to as patch 1).	SAF/NWC/CDOP2/PPS/SCI/ATBD/1, Issue 1.1, 13 March 2015
RD 2	Algorithm Theoretical Basis Document "Cloud Top Temperature, Pressure and Height", PPS version 2014	SAF/NWC/CDOP2/PPS/SCI/ATBD/3, Issue 1, 15 September 2014
RD 3	Algorithm Theoretical Basis Document "Cloud Physical Properties", PPS version 2014	SAF/NWC/CDOP2/PPS/SCI/ATBD/5, Issue 1, 15 September 2014
RD 4	Algorithm Theoretical Basis Document AVHRR GAC cloud products extension: Probabilistic cloud masks – CMA-prob	SAF/CM/SMHI/ATBD/GAC/PBCM, Issue 1.1, 27 May 2016
RD 5	Algorithm Theoretical Basis Document Joint Cloud property Histogram AVHRR/SEVIRI	SAF/CM/SMHI/ATBD/GAC/JCH Issue 2.2
RD 6	Product User Manual CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2) Cloud Products	SAF/CM/DWD/PUM/GAC/CLD, issue 2.1

TABLE of CONTENT

LIST of FIGURES

LIST of TABLES

1 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, http://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), and the Meteorological Service of the United Kingdom (UK MetOffice). 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 lined out 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, http:/www.cmsaf.eu/. Here, detailed information about product ordering, add-on tools, sample programs and documentation is provided.

2 Introduction

This CM SAF Algorithm Theoretical Basis Document (ATBD) provides information on the processing chain implemented for the CM SAF AVHRR GAC data record to retrieve geophysical parameters from inter-calibrated measurements of the Advanced Very High Resolution Radiometer (AVHRR). This processing chain covers all steps from level-1b to the final level-3. It also provides information on the CM SAF retrieval schemes used to construct the data record employing the measurements of AVHRR instruments onboard the NOAA satellites (NOAA-07, NOAA-09, NOAA-11, NOAA-12, NOAA-13, NOAA-14, NOAA-15, NOAA-16, NOAA-17, NOAA-18, NOAA-19) and the EUMETSAT MetOp-A and MetOp-B satellites. The algorithms applied are based on an upgraded version (patch 1, see [RD 1]) of the PPS (Polar Platform System) Version 2014 cloud processing package (Dybbroe et al., 2005a and Dybbroe et al., 2005b), which is used to determine cloud fraction and cloud top properties. PPS 2014 includes the CPP (Cloud Physical Properties) algorithm (Roebeling et al. 2006), which retrieves cloud thermodynamic phase, cloud optical thickness, cloud particle effective radius, and liquid/ice water path.

The document seamlessly describes all elements of the production of the final CM SAF AVHRR GAC 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 [.evel-1b → level-1c].
2. A report on the derivation of the cloud products by applying CPP and PPS algorithms. Note, significant parts of the PPS and CPP algorithms have already been documented in previous ATBDs, which will be referred to in this document when appropriate [level-1c → level-2].
3. The elaboration of the production of the daily and monthly means (level-3 data) based on the level-2 products provided by CPP and PPS [level-2 → level-2b/level-3].

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

The CM SAF AVHRR GAC data record contains multiple cloud parameters derived from AVHRR. The CM SAF release of AVHRR GAC edition 2 contains the following cloud variables:

Fractional Cloud Cover [CM-11011, CFC, section 4.1]

Joint Cloud property histogram [CM-11021, JCH, section 4.2]

Cloud Top level [CM-11031, CTO, section 4.3]

Cloud Phase [CM-11041, CPH, section 4.4]

Liquid Water Path [CM-11051, LWP, section 4.5]

Ice Water Path [CM-11051, IWP, section 4.6]

3 Processing of measures AVHRR radiances (level-1)

3.1 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 3-1 gives an overview over the satellites carrying the AVHRR instrument until 2016, which are used within the current processing. In contrast to the first edition of the data record, now also Metop-B has been used. . The 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.

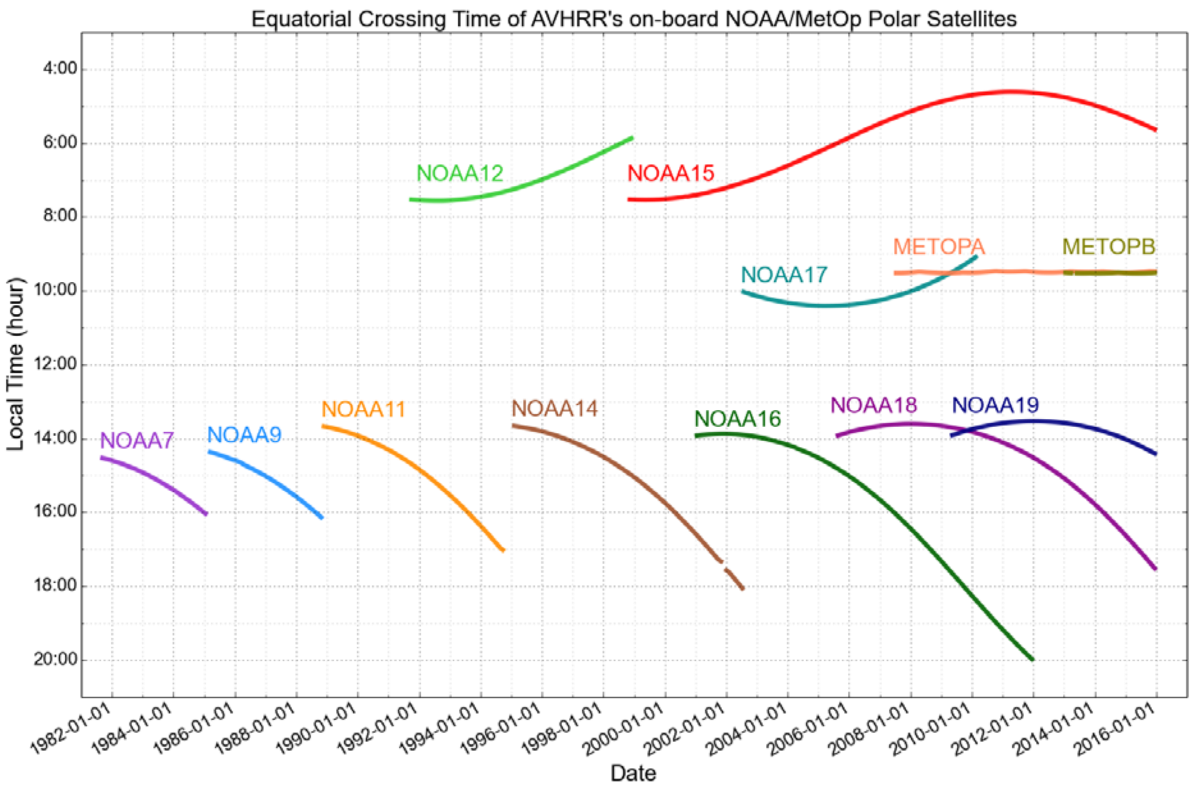

Figure 3-1*: Local solar times at equator observations for all satellites from NOAA-7 to NOAA-19 and METOP A/B. Shown are all data that are used for the processing. The figure shows ascending (northbound) equator crossing times for afternoon satellites (NOAA-7 TO NOAA-19) and descending (southbound) equator crossing times for morning satellites (NOAA-12 to NOAA-17 and METOP A+B). Corresponding night-time observations take place 12 hours earlier/later. Some data gaps are present but only for some isolated dates (Courtesy of Cornelia Schlundt, DWD).*

Table 3-1 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 of approximately 4 km are permanently archived and available with global coverage since the beginning of measurements. This data record is denoted Global Area Coverage (GAC) AVHRR data

3.2 Compilation of the CM SAF GAC cloud data record

The GAC data record of global cloud products retrieved by CM SAF cloud retrieval methods spans the time period 1982-2015. Retrieval methods have been dependent on the access to two infrared (split-window) channels at 11 and 12 microns meaning that only data from satellites carrying the AVHRR/2 or AVHRR/3 instruments have been used (see Table 3-1). Figure 3-1 further describes the coverage of observations from each individual satellite over the entire period. Notice that the limitations to the use of AVHRR/2 and AVHRR/3 instruments leads to poorer time sampling (i.e., only one satellite available for daily observations) between 1982 and 1991. On the other hand, from 2001 and onwards more than two satellites are available for daily observations. The second edition data record covers the period from 1982 to 2015, in total 34 years.

Table 3-1: *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 and Metop-B. For the other satellites with AVHRR/3 it was used only for shorter periods.*

Channel Number	**Wavelength (micrometers) AVHRR/1 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**
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 utilised 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 3-1 for all used satellites. Some compensation for this has been attempted in the CM SAF data record but not for all parameters.

3.3 Defining the AVHRR FCDR for CLARA-A2

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 homogenized radiances (formally denoted Fundamental Climate Data Records - FCDRs). For the CM SAF CLARA-A2 data record an AVHRR FCDR (as intermediate, non-official product) has been generated for which the calibration of the AVHRR shortwave reflectances is based on work performed by NOAA (updated version of Heidinger et al. (2010)). A similar FCDR 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 updated calibration data, compared to the FCDR used for CLARA-A1 and earlier versions of PATMOS-x, took advantage of the new MODIS Collection 6 data as its main calibration reference (Heidinger et al.; publication currently in preparation).

It is important to note, that the used calibration coefficients were calculated for AVHRR measurements until 2014. The extrapolation of the calibration curves into the year 2015 might introduce some additional uncertainties to the CLARA-A2 clouds products of 2015.

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., recognising the work of Mittaz et al., 2009). In addition to the calibration, the used FCDR underwent several quality control and data screening procedures before being used in the processing of CLARA-A2. All the details on how data have been screened and monitored will be reported in a specific SCOPE-CM project (project SCM-05 on the AVHRR GAC FCDR development).

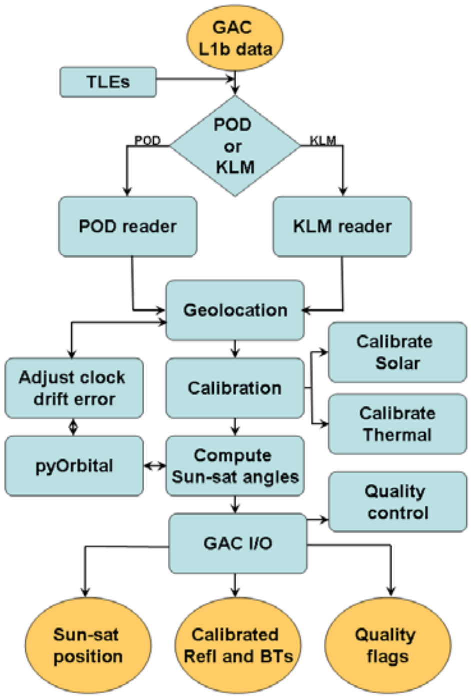

Figure 3-2 *Schematic representation of the components of PyGAC interface. POD represents the satellite family that carried the second generation AVHRR instruments (i.e., AVHRR/2, up till and including NOAA-14), while KLM represents satellite family carrying the third generation instruments (AVHRR/3). PyOrbital is a part of PyTroll family of Python interfaces designed to process meteorological satellite data. (Courtesy of Abhay Devasthale, SMHI)*

The first task to carry out prior to starting the processing of CLARA-A2 geophysical end products is to transfer the basic information in original NOAA level-1b to calibrated, inter-calibrated and navigated radiances, i.e., to formally define an AVHRR level-1c data record. The latter would then define the actual AVHRR FCDR used when deriving the CLARA-A2 data record. In contrast to the CLARA-A1 data record, where the actual AVHRR FCDR was produced within the framework of the official PPS software, a stand-alone dedicated Python tool - named PyGAC – has been used for this purpose. PyGAC is an open-source community-driven Python interface to read and calibrate raw level-1b AVHRR GAC data (https://github.com/adybbroe/pygac/tree/feature-clock). The interface was developed jointly under the frameworks of CM SAF and ESA's Cloud Climate Change Initiative (i.e., the ESA Cloud_cci project). For each GAC orbit, PyGAC provides calibrated reflectances (Earth-Sun distance corrected) and brightness temperatures, sun and satellite zenith and azimuth angles and scanline quality information.

Figure 3-2 gives an overview of the different steps of preparing AVHRR radiances and associated meta data records. The output files have HDF5 format and follow international Climate and Forecast Conventions. PyGAC employs state-of-the-art inter-calibration of visible channels provided by NOAA (Heidinger, 2014, personal communication). The visible channel intercalibration is of highest quality conforming to the requirements for fundamental climate data records. The geolocation is improved for the historical second generation of AVHRR data (AVHRR/2) using clock-drift corrections and the interface is easily adaptable to be able to compute independent navigation information. PyGAC is currently able to preprocess all GAC data from 1982 onwards.

During development of PyGAC a large number of technical issues and deficiencies of the original AVHRR GAC level-1b data record has been revealed and taken care of. One of the most important issues here is that the final level-1c data record is defined without any overlap in the beginning or in the end of each orbit for sub-sequent GAC orbits (i.e, no duplication of data is allowed as in the original level-1b data record). Also, GAC orbits with corrupt or in-correct data in any form in the original data is prohibited from further processing. From this follows that the logged quality information offers a way of describing the detailed status of the old level-1b data record as well as of the finally realised FCDR in the level-1c data record.

It is believed that this together with the standardized representation of level-1c data and the improved stability of infrared calibration would be very attractive for external users. Thus, requests from external users can be foreseen to use the level-1c data record and therefore the level-1c data record will be listed among the auxiliary data records accessible by users of CLARA-A2. A separate document (tentatively named "AVHRR level-1c metadata document") will be prepared describing (as complete as possible) the character of the level-1c data record. This will be complementary to the product user manual (PUM) since the level of details would be too high for the PUM.

Finally, it is worth to mention that in the generation of the final AVHRR FCDR for CLARA-A2 a post-processing method has been applied for reducing the high noise levels in AVHRR channel 3b at 3.7 micrometers for some of the earlier NOAA-satellites (in particular NOAA-7, NOAA-9 and NOAA-12). 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. 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. A publication of the detailed filtering procedure is under preparation.

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

This section provides information on the processing of PPS and CPP to retrieve cloud parameters from inter-calibrated AVHRR observations. Each parameter is briefly introduced in the following with the respective detailed ATBD being referred to.

4.1 Fractional Cloud Cover [CM-11011, CFC]

This product is derived directly from results of a cloud screening or cloud masking method. The cloud fractional cover is defined as the fraction of cloudy pixels per grid square compared to the total number of analysed pixels in the grid square. Fractional cloud cover is expressed in percent. The cloud screening and cloud masking is performed using the upgraded NWC SAF PPS 2014 version (with patch 1), which is described in more detail in [RD 1].

For future issues of the CLARA data record the PPS cloud mask product is likely to be replaced by a cloud probability product. In this way the uncertainty in cloud masking can be taken into account. A first prototype version of a probabilistic cloud mask has been developed in the CM SAF project which can be attached as a post-processing module to the official PPS software. The methodology is described in [RD 4]. A restricted data record with probabilistic cloud masks (in level-2 and level-2b formats) for a limited period will be prepared and made available to CLARA-A2 users for demonstration purposes. It is hoped that this could lead to important user feedback for the planning of the successor of CLARA-A2.

4.2 Joint Cloud property Histogram [CM-11021, JCH]

The JCH product is a combined histogram of cloud-top pressure (CTP, see Section 4.2), cloud optical thickness (COT, see Sections 4.5 and 4.6), and cloud phase (CPH, see Section 4.4), 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 5.3.2. Some further historical background and description of this product is given in [RD 5].

4.3 Cloud Top level [CM-11031, CTO]

Three versions of the CM SAF Cloud Top product exist: 1. Cloud Top Temperature (CTT), expressed in Kelvin; 2. Cloud Top Height (CTH), expressed as altitude over ground topography (m); 3. Cloud Top Pressure (CTP), expressed in pressure co-ordinates (hPa). The CTO product is derived using two approaches, one for opaque and one for fractional and semitransparent clouds, and it is applied to all cloudy pixels as identified by the PPS cloud mask product.

The opaque algorithm use simulated cloud free and cloudy TOA 11 µm radiances which are compared and matched to measured radiances. Cloudy radiances are simulated from ECMWF ERA profiles using RTTOV-11 assuming “black-body”-clouds at various levels.

The semi-transparent algorithm is applied to all pixels classified as semi-transparent cirrus or fractional water cloud. A histogram technique is applied based on the construction of two dimensional histograms using AVHRR channel 4 and 5 brightness temperatures composed over larger segments or boxes. The studied box comprises12x12 GAC pixels (approximately 60x60 km2=3600 km2). The method means that for CTO the horizontal variability of thin clouds is much lower than for other (opaque) clouds. This is, of course, a limitation but better than the alternative to leave the pixel without any CTO estimation (which e.g. would make the compilation of the JCH product problematic). By an iterative procedure a polynomial curve (simulating the arc shape) is fitted to the histogram-plotted values from which the cloud top temperature and pressure (taken from ERA-Interim profiles) is retrieved.

Both schemes are part of NWC SAF PPS 2014 patch 1. Details can be found in [RD 2].

4.4 Cloud Phase [CM-11041, CPH]

The cloud phase product is meant to represent the thermodynamic phase of the particles near the cloud top. The cloud phase retrieval is based on a number of threshold tests using AVHRR channels 3a, 3b, 4 and 5. 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 are then further condensed to liquid (former three) and ice (latter three) phase. Separate retrieval schemes are applied during daytime and nighttime. 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]

4.5 Liquid Water Path [CM-11051, LWP]

For daytime pixels to which the liquid phase has been assigned, liquid water path (LWP, units kg m^{-2}) is derived from the cloud optical thickness (COT or τ) and particle effective radius (REF or r_e). The τ - r_e retrieval scheme, developed at KNMI (Roebeling et al., 2006), uses a pair of satellite radiances at wavelengths in the non-absorbing (for clouds) visible and the moderately absorbing solar infrared part of the spectrum, 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 100 and r_e between 3 and 34 μm. Scattering properties were calculated with Mie theory for spherical droplets.

The AVHRR channels 1 (0.6 μm) and 3a/3b (1.6/3.7 μm) are used for the retrieval. For which time periods and satellites channel 3a or 3b was active is outlined in Table 1-2 in the Product User Manual [RD 6]. Additionally, each (swath-based) level-2 file contains information which shortwave infrared channel was used for that particular orbit.

Liquid water path is then computed from the retrieved τ and r_e (values by (Stephens, 1978): LWP = 2/3 $\rho_l \tau r_e$, in which ρ_l represents the density of liquid water (1000 kg m-3). More details on the retrieval scheme can be found in [RD 3].

In addition to LWP, COT and REF are provided as additional layers in the product files. Moreover, error estimates of these parameters are given.

4.6 Ice Water Path [CM-11061, IWP]

For daytime pixels to which the ice phase has been assigned, ice water path (IWP, units kg m^{-2}) is retrieved in the same way as LWP described in Section 4.5. Here, the scattering properties were calculated with a raytracing code (Hess et al. 1998) using the geometric optics approximation for randomly oriented, roughened, hexagonal ice crystals with effective radii between 5 and 80 μm. 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 REF are provided as additional layers in the product files, and error estimates of these parameters are given.

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

Following the findings and the experiences obtained producing the first edition 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 the final level-3 averages on a latitude/longitude grid and also the introduction of the new level-2b data format.

The level-3 processing uses pixel level retrievals from the PPS cloud processing software (Version 2014, patch 1), . The final outputs are produced on a global regular equal angle grid and the polar "Equal-Area Scalable Earth Grid" (EASE-Grid) as daily and monthly averages (level-3) as well as daily samples on an global equal area grid (level-2b). The related specifications for these level-3 data are summarized in section 5.1. Section 5.2 gives a detailed definition of the final grids. Level-2b is a new format which is introduced in the section 5.2.3. A detailed description of averaging methodology of level-2 data to level-3 data (daily and monthly means) on the final grid is given in section 5.3 & 5.4.

In addition to the first edition CLARA data record, the second edition provides results of all products on an equal-area polar grids with 25 km resolution for the Arctic and Antarctic regions, respectively. This is done to facilitate usage of results over the poles, where the converging longitudes make the use of regular latitude-longitude grids problematic.

5.1 Definition of product specifications

The CM SAF GAC cloud data record from AVHRR provides global coverage of a number of cloud parameters. AVHRR GAC retrievals are used at original swath level, re-projected on a final grid and aggregated to the spatio-temporally averaged data records. Also a remapped level-2b version of each product is processed in this second edition. 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 and the cloud top phase. For any level-3 product, daytime observations are defined by a solar zenith angle of less than 75°. 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, CPP data 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 5.4.2 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 5.4.3, 5.4.4, 5.4.5, and 5.4.6 for more technical details.).

The temporal coverage of the data records ranges from January 1982 to December 2014. Notice again (as mentioned in Section 3 and visualised in Figure 3-1) that for the first years in the series (1982-1991) only the afternoon satellites are included.

5.2 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 polar grid was used in CLARA A1 already for the fractional cloud cover. 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.

5.2.1 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 more coarse with a spatial resolution of $(1°)^2$.

5.2.2 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

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

and for the southern hemisphere

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

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).

5.2.3 Level-2b

With 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.

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 (see further down in Section 5.3).

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) 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, 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 pols.
3) Towards the edge of the swath 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.

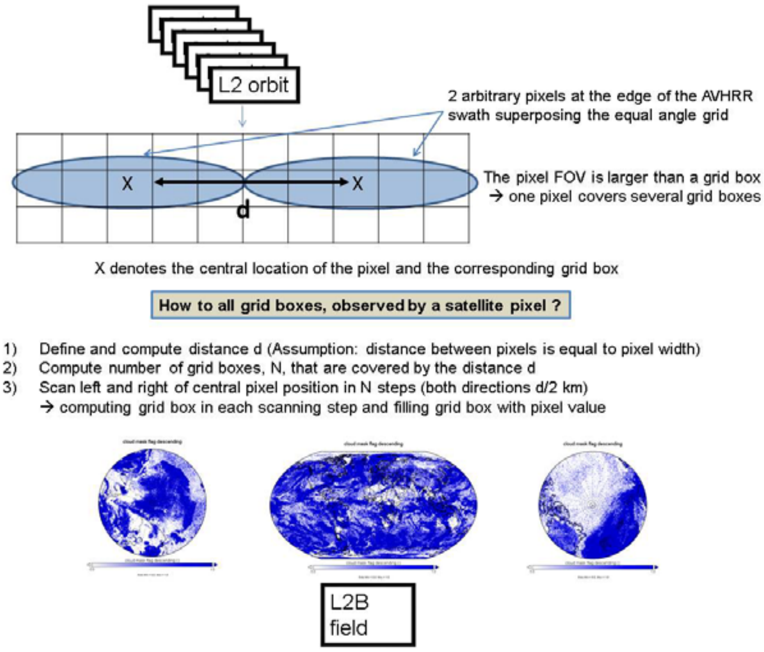

Figure 5-1 *From level-2 to level-2b. How to deal with mapping effects.*

Figure 5-1 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.

5.3 Calculation of final level-3 - products

For the official level-3 product, cloud parameters for all GAC orbits in level-2 native satellite projection, $x(i,j)$, could be considered for aggregation into daily and monthly means. However, a method has been applied that uses a sub-set of all observations for the calculation of level-3 products. Here, first the level-2b file with the highest spatial and temporal resolution has been defined. The level-2b file is then the starting point for the aggregation of the daily mean. In contrast to the first CLARA edition, in which all original level-2 files were used in the averaging, the project team sees here one major advantage with the proposed method: Due to the sampling technique a globally nearly constant number of observations per grid box can be achieved.. The daily mean is then used as the starting point for the aggregation of the monthly mean. Here the project team sees the advantage of a) a meaningful weighting of single days in order to not overestimate, i.e. single rainy days in dry areas in the monthly means and b) of a clear separation of the spatial and temporal variability. 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 and
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.

5.4 Short overview on level-3 cloud properties

5.4.1 Fractional Cloud Cover [CFC]

The daily mean fractional cloud cover is calculated from the aggregation of the instances of the binary cloud mask information, sampled on the level-2b grid, as follows:

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

with i and j being the indices of the final field projection, $N(i,j)_{Cloudy}$ the number of cloudy cases and $N(i,j)_{Clear}$ the number of clear cases. Acknowledging the different cloud detection capability during day and night time (the different methodologies described in RD 1), an additional separation is done leading to $CFC(i,j)_{Night}$ and $CFC(i,j)_{Day}$ separate day time and night time averages. Here, the solar zenith angle of <75° and ≥95° are used to define day and night, respectively. Cases with solar zenith angles between 75° and 95° are excluded due to specific problems occurring in twilight conditions. Daily means of each day of the month are arithmetically averaged to the monthly mean.

5.4.2 Joint Cloud property Histogram [JCH]

Differing from CFC, CPH, LWP and IWP product derivation, 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 75°. Notice that this solar zenith angle threshold approach is different from the one used for the SEVIRI data record. The product is also computed in a coarser grid (1°x1°) than the standard products in order to improve the statistical significance of the cloud distributions.

This product is described in five-dimensional fields $JCH(i,j,t,p,ph)$. Indices *i* and *j* again 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. An alternative way to describe it is that for each grid point, a 3-dimensional histogram with frequencies of simultaneous occurrences of cloud top pressure, cloud optical thickness and cloud phase categories is given. However, since cloud phase is only given with two alternatives (liquid or frozen), in practice two 2-dimensional COT-CTP histograms per grid point are available; one for water clouds and another one for ice clouds.

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*. COT and CTP bins are strategically defined – e.g. for COT using approximately equal spacing in terms of cloud albedo – following guidance from the International Satellite Cloud Climatology Project (ISCCP) JCH bin distributions (Rossow and Schiffer, 1991) but with a larger number of bins. Specifically, the bin borders have been chosen as follows (CTP in hPa):

COT: {0, 0.3, 0.6, 1.3, 2.2, 3.6, 5.8, 9.4, 15, 23, 41, 60, 80, 100}
*CTP: {*1, 90, 180, 245, 310, 375, 440, 500, 560, 620, 680, 740, 800, 875, 950, 1100*}*

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

$$JCH(i,j,t,p) = N(i,j)_{COT\epsilon COTbin;CTP\epsilon CTPbin;CPH=ice} \quad (2)$$

and

$$JCH(i,j,t,p) = N(i,j)_{COT\epsilon COTbin;CTP\epsilon CTPbin;CPH=liquid} \quad (3)$$

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

5.4.3 Cloud Top Level [CTO]

The CTO product contains daily and monthly means for CTH, CTP, and CTT. For these parameters all valid entries of the resampled level-2b fields are aggregated and then weighted by the number of used entries to the daily mean.

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

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

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:

$$< CTP(i,j) >_{ln} = \exp\left(\frac{1}{N(i,j)_{Cloudy}} \sum_{k=1}^{N(i,j)_{Cloudy}} \ln(x_k(i,j))\right) \quad (5)$$

Daily means of each day of the month are arithmetically averaged to the monthly mean

5.4.4 Cloud Phase [CPH]

Similarly to CFC, the daily means result from the aggregation over the different data points in the resampled level-2b field based on the cloud phase retrievals. CPH is expressed as fraction of liquid water clouds by calculating the ratio of number of detected liquid clouds $N(i,j)_{liquid}$ with respect to the total number detected $N(i,j)_{Cloudy}$:

$$CPH(i,j) = \frac{N(i,j)_{liquid}}{N(i,j)_{Cloudy}} \quad (6)$$

The cloud phase is presented as all day product and also as daytime only product. The latter is included as an additional layer.
Daily means of each day of the month are arithmetically averaged to the monthly mean

5.4.5 Liquid Water Path [LWP]

Daily mean LWP is calculated for each grid cell $< LWP(i,j) >$ as given in Equation (4). COT and REF 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 shortwave radiative effect.

In addition to the above means for the liquid-cloud portion of the sky, all-sky LWP and COT are calculated by including cloud-free and ice-cloud pixels as zeroes in the averaging. The daily mean all-sky LWP and COT are then arithmetically averaged to the monthly mean.

5.4.6 Ice Water Path [IWP]

Daily and monthly mean IWP, as well as the ice cloud COT and REF, 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.

5.4.7 Two-dimensional histograms

Additional to the 3D histogram (JCH) also monthly 2D histograms are generated for the parameters CTP, CTT, CWP, COT, REF. Each histogram shows the solution space of its variable with the cloud phase as additional dimension. Thus, with only two values for CPH (liquid and ice), it means that for every grid point two 1-dimensional histograms, one for water clouds and one for ice clouds are available. These histograms are provided on the spatial resolution of the level-3 products, $(0.25)^2$. The used bins are:

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, 100}

REF: {3, 6, 9, 12, 15, 20, 25, 30, 40, 60, 80} [µm]

CTP:{ 1, 90, 180, 245, 310, 375, 440, 500, 560, 620, 680, 740, 800, 875, 950, 1100} [hPa]
CTT:{ 200, 210, 220, 230, 235, 240, 245, 250, 255, 260, 265, 270, 280, 290, 300, 310, 350} [K].
Note that the COT and CTP 2D histograms are different from what would be obtained by collapsing the 3D JCH histogram. The latter includes only pixels for which both COT and CTP have valid values, meaning for instance that nighttime pixels are excluded because COT is not retrieved, and some daytime pixels are excluded because no valid CTP may have been retrieved.

5.4.8 Additional statistical parameters

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

$$s\big(x(i,j)\big) = \sqrt{< x^2(i,j) > - < x(i,j) >^2} \quad (7)$$

For cloud microphysical properties, CWP, COT and REF, the retrieval defines an uncertainty which is also presented in the daily mean and monthly mean as well as its standard deviation.
In the daily mean the standard deviation is based on all available observations in the level-2b file for the related grid box. In the monthly mean it is defined as standard deviation over all available daily means.

5.4.9 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. 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.

6 References

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., 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

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.

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.

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