ESA Cloud_cci
Product Validation and Intercomparison Report (PVIR)

(Applicable to Cloud_cci version 2.0 products)

Issue 5 Revision 1

06 March 2018

Deliverable No.:	D-4.1
ESRIN/Contract No.:	4000109870/13/I-NB

Project Coordinator:	Dr. Rainer Hollmann Deutscher Wetterdienst rainer.hollmann@dwd.de
Technical Officer:	Dr. Simon Pinnock European Space Agency Simon.Pinnock@esa.int

DOIs of Cloud_cci datasets:	DOI: 10.5676/DWD/ESA_Cloud_cci/AVHRR-PM/V002 DOI: 10.5676/DWD/ESA_Cloud_cci/AVHRR-AM/V002 DOI: 10.5676/DWD/ESA_Cloud_cci/MODIS-Aqua/V002 DOI: 10.5676/DWD/ESA_Cloud_cci/MODIS-Terra/V002 DOI: 10.5676/DWD/ESA_Cloud_cci/ATSR2-AATSR/V002 DOI: 10.5676/DWD/ESA_Cloud_cci/MERIS+AATSR/V002

Document Change Record

Document, Version	Date	Changes	Originator
PVIR_v4.0_draft	31/01/2016	Initial draft	Stefan Stapelberg, Martin Stengel, Cornelia Schlundt, Karl-Göran Karlsson, Jan Fokke Meirink, Caroline Poulsen, Jed Bojanowski, Reto Stöckli
PVIR_v4.0_submitted	06/02/2017	Final revision	Martin Stengel, Rainer Hollmann
PVIR_v4.1 track change	31/03/2017	Revised document after ESA review	Stefan Stapelberg, Martin Stengel
PVIR_v4.1	18/04/2017	Approved and issued version	Martin Stengel
PVIR v5.0		Adding uncertainty validation in Section 6.	Martin Stengel
PVIR v5.1 approved and released version	06/03/2018	Modifications after ESA review Adding validation for AATSR for uncorrected CTH in Table 3.8. Complete revision of uncertainty validation (Section 6) Adding GCOS compliance matrix and discussion to the Summary section (Section 7)	Martin Stengel

Purpose

This document is the Product Validation and Intercomparison Report (PVIR) for ESA Cloud_cci. It covers the validation results for all generated data sets, namely Cloud_cci AVHRR-PM, AVHRR-AM, MODIS-Aqua, MODIS-Terra, ATSR2-AATSR and MERIS+AATSR, together with outcomes from a comprehensive inter-comparison study. The application of the Cloud_cci datasets in the framework of climate research is covered in the Climate assessment report (CARv3) of Cloud_cci.

Table of Contents

1. Executive summary

In the ESA Cloud_cci project, comprehensive efforts have been undertaken to optimize the utilization of passive imaging sensors of European and non-European satellite mission for the generation of enhanced cloud property climate datasets. The Cloud_cci datasets are innovative in a variety of aspects, which are summarized in Section 2.2. In this report the main results of the evaluation process of Cloud_cci version 2.0 datasets are summarized. The evaluation is divided into 3 tasks:

(1) Validation of Cloud_cci data against high quality ground- and satellite-based reference observations, i.e. Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), human ground observations (SYNOP), passive microwave (MW) observations,

(2) Comparisons to well-established, satellite-based cloud property datasets of similar kind (also based on passive imager satellite sensors)

(3) A homogeneity analysis that specifically focuses on the stability of the cloud fraction data in those Cloud_cci datasets which consist of more than one sensor.

Within the evaluation process we use the metrics: Bias (alias: accuracy), the bias-corrected root mean square error (bc-RMSE, alias: precision) and stability (the ability to preserve the same bias throughout the time series).

The results of Task 1 are considered to be an objective measure of the quality of the Cloud_cci data, since the reference observations used are assumed to be of higher accuracy and precision than the Cloud_cci data. Here we specifically validated pixel-based cloud mask (CMA), cloud phase (CPH) and cloud top height (CTH) against CALIOP, a space based lidar. For CMA, we find that at least 80% of all clouds are correctly identified and usually also 80% of all clear scenes, with the exception of the high latitudes, where snow and ice covered surface still cause problems for cloud detection. These problems are amplified in night time conditions. Cloud detection in twilight conditions remains challenging as well. With respect to CTH, optically thick clouds can be vertically places with high accuracy. For low-level clouds this mean biases of partly smaller than 150m. Optically thin clouds, which are mostly high-level clouds, are still difficult to handle, resulting in partly large biases (large underestimation of CTH by Cloud_cci). The scores improve significantly when the CALIOP reference CTH is taken from a lower cloud layer in the profile for which a certain level-to-cloud-top cloud optical depth is exceeded. For CPH, in more than 70% of all cloudy cases an agreement in cloud phase determination between Cloud_cci datasets and CALIOP is found. This increases to almost 80% when the reference CALIOP phase is taken at a level of 0.15 optical depth below the geometrical cloud top.

Furthermore we used SYNOP observations of cloud cover, converted to monthly cloud fraction (CFC), to validate the Cloud_cci cloud fraction with the full data record. As a long-term data availability requirement is applied to the SYNOP, mainly only European and North-American stations remain. Corresponding validation of CFC reveals a very good quality and stability of nearly all Cloud_cci data. Only the early-morning satellites NOAA-12 and NOAA-15 (part of the AVHRR-AM dataset) and to a smaller extent also the afternoon satellites NOAA-7, NOAA-9 (part of AVHRR-PM datasets) as well as the ESA satellite ERS2 (part of ATSR2-AATSR2 datasets) are more difficult to handle. The biases are usually around zero, only lightly positive for the afternoon satellites and show as the bc-RMSE a small seasonal cycle with larger values during boreal winter.

As last point, we used monthly mean, satellite-based MW observations of liquid water path (LWP) (O'Dell et al., 2008) to validate the Cloud_cci equivalents in three selected regions, which are dominated by liquid cloud occurrence. Basically, all Cloud_cci data shows reasonable agreement with the reference data, but there are considerable differences among the Cloud_cci datasets or sensors used. Worst results are again found for NOAA-12 and NOAA-15, which, due to their very difficult illumination conditions (very high solar zenith angle due to their early-morning orbit), show a strong over estimation. Besides these problems, the biases are usually around ± 10 g/m^2 and the bc-RMSE is overall between 5 and 10 g/m^2. Linear correlation coefficients of monthly mean Cloud_cci and UWisc LWP range from 0.7 to 0.95.

Results of Task 2 are considered to be a consistency analysis among all Cloud_cci datasets but also in comparisons to reference datasets, which already exist and are well established, but can not necessarily serve as absolute references. Here we used CLARA-A2 (Karlsson et al., 2016), PATMOS-x (Heidinger et al., 2014) and MODIS Collection 6 datasets (Baum et al., 2012; Platnick et al., 2014; Marchant et al., 2016). The comparison include maps of mutli-annual mean values as well as zonal mean value plots both for a common period for which all datasets are available. Also time series plots of 60S-60N average are shown for the full records and discussed. For CFC, CPH and CTP we find reasonable agreement among all datasets of consideration, with usually only small difference, which can mainly be explained by sensitivities difference caused by different spectral bands used (e.g. MODIS Collection 6) as well as differences in algorithm approaches. However, in polar regions the spread in CFC is very large, probably due to difficulties to correctly account for snow and ice surface condition and a related limited signal to information ratio under in these conditions, avoiding any firm conclusion. For optical thickness (COT) we found a relative good agreement among all datasets for both liquid and ice clouds. Usually, the datasets agree within 5 to 10 optical thicknesses wrt. the average values for the common period (2003-2011).Interpretations for the polar regions are not easy, because the spread is large as well as many datasets tend to have very high optical thicknesses in the polar regions which seem counter intuitive. For cloud effective radius (CER) we find all datasets agreeing on reflecting the same global pattern. On the other hand the absolute values differ partly by 5 to 10µm for liquid clouds. For ice clouds the spread is even larger. This is firstly due to high CER_{ice} values reported in MODIS Collection 6, but secondly also by very high values in all CC4CL-based datasets (all except Cloud_cci MERIS+AATSR). The latter could be tracked down to two bugs, one being related to miscalculation of BRDF components for high solar zenith angle and/or snow/ice covered surfaces, and the other one begin error in the pre-calculated LUT for ice clouds that map ice cloud properties to radiances. Similar to CER, the cloud water path comparison show a reasonable agreement for all datasets for liquid clouds (LWP) but less agreement for ice clouds (IWP); the latter due to the same reason mentioned above. The LWP and IWP agreement is better once allsky values are considered, because here the impact of differences in cloud detection and phase determination can be reduced.

In Task 3 the results of a conducted homogeneity analysis are presented for which more than 800 SYNOP sites were used to monitor the spatially and temporally collocated time series of Cloud_cci CFC. The analysis was done by means of applying the Standard Normal Homogeneity Test (SNHT) to identify break points. Furthermore, linear trend analysis was performed using Theil-Sen estimates and Mann-Kendall significance estimations. For AVHRR-AM, a break in the time series is found for the transitions of NOAA-12 to NOAA-15 in all considered regions except Tropics and another one for 1994-1995. The inhomogeneity between NOAA12 and NOAA15 is largely responsible for a negative trend in CFC monthly anomalies of Cloud_cci AVHRR-AM. For AVHRR-PM, an evident inhomogeneity is only found for the transition of NOAA-7 to NOAA-9, which is a major driver for a negative trend found in the time series which is significant in all regions except tropics. For ATSR2-AATSR a major inhomogeneity is found between the two sensors which is more prominent for daytime than for night time observations.

Considering the results of all three tasks we have composed a table (Table 1-1) summarizing our interpretation of the evaluation results in three confidence levels, which should give the user a very first impression on the data quality. The confidence levels are:

high	for those products for which the validation could prove a high quality when compared to validation data and/or the comparisons to other datasets in section 2 suggested a high level of agreement to, but also among the reference dataset
moderate	for those products for which only a moderate agreement to the validation data could be found or, in case no validation data was available, a larger spread in the comparison datasets avoid any firm conclusions about the actual quality of the Cloud_cci data
low	for those products for which either a significant disagreement to the validation data was found or recently found bugs suggest significant systematic and/or random errors

It is important to note that the confidence table is a subjective measure and all users are advised to check the individual sections of this report for comparing specific accuracy and/or precision requirements that may exist with the actually validation and comparison results. Please also note that the confidence table sometimes separates time periods of a certain datasets in case the quality undergoes a significant change.

Table 1-1 *Subjective confidence table composed by the project team and based on the evaluation results documented in this report. See text for interpretation.*

Cloud variable	Criterion	Cloud_cci dataset					
		AVHRR-AM	AVHRR-PM	MODIS-Terra	MODIS-Aqua	ATSR2-AATSR	MERIS-AATSR
CFC	Accuracy (bias)	moderate[1]	moderate[1]	moderate[1]	moderate[1]	moderate[1]	moderate[1]
	Precision	high[1]	high[1]	high[1]	high[1]	high[1]	moderate[1]
	Stability	low[2] high after 2001	low[2] high after 2001	high	high	low high after 2003	low
CPH	Accuracy (bias)	moderate	moderate	moderate	moderate	moderate	moderate
	Precision	high	high	moderate	high	moderate	moderate
	Stability	high moderate[2] before 1999	high moderate[2] before 2001	high	high	high moderate before 2003	high
CTP	Accuracy (bias)	moderate	moderate	moderate	moderate	moderate	moderate
	Precision	high	high	high	high	high	high
	Stability	moderate[2] high after 2001	moderate[2] high after 2001	high	high	low high after 2003	moderate
COT_{liq}	Accuracy (bias)	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat
	Precision	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat	high in 60S-60N low at high-lat
	Stability	high moderate[2] before 1999	high moderate[2] before 2003	high	moderate	high moderate before 2003	high
COT_{ice}	Accuracy (bias)	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat
	Precision	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat	high (60S-60N) low at high-lat
	Stability	high after 2003 low[2] before 2003	high after 2003 low[2] before 2003	high	low	high	high
CER_{liq}	Accuracy (bias)	moderate	moderate	moderate	moderate	moderate	moderate
	Precision	moderate	moderate	moderate	moderate	moderate	moderate
	Stability	high low[2] before 2003	low[2]	high	high	high moderate before 2003	high
CER_{ice}	Accuracy (bias)	moderate low before 2003	low low at high-lat	low low at high-lat	low low at high-lat	low low at high-lat	moderate low at high-lat
	Precision	high low before 2003	moderate low before 2003	moderate low before 2003	moderate low before 2003	moderate low before 2003	moderate low before 2003
	Stability	high after 2003 moderate[2] before 2003	moderate[2] low[2] in 2000-2003	high	high	high moderate before 2003	high
LWP	Accuracy (bias)	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat
	Precision	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat
	Stability	high low[2] before 2003	low[2]	high	moderate	high moderate before 2003	high
IWP	Accuracy (bias)	moderate low at high-lat	low	low	low	low	moderate low at high-lat
	Precision	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat	moderate low at high-lat
	Stability	high low[2] before 2003	medium[2] low[2] before 2003	high	low	high moderate before 2003	high

[1] Significant quality reduction in the very high latitudes, especially over snow and ice surface.
[2] Diurnal cycle /drift correction necessary to increase the stability

2. Introduction

2.1 The ESA Cloud_cci project

The ESA Cloud_cci project covers the cloud component in the European Space Agency's (ESA) Climate Change Initiative (CCI) programme (Hollmann et al., 2013). In the ESA Cloud_cci project, long-term and coherent cloud property datasets have been generated exploiting the synergic capabilities of different Earth observation missions (European and non-European) allowing for improved accuracies and enhanced temporal and spatial sampling better than those provided by the single sources.

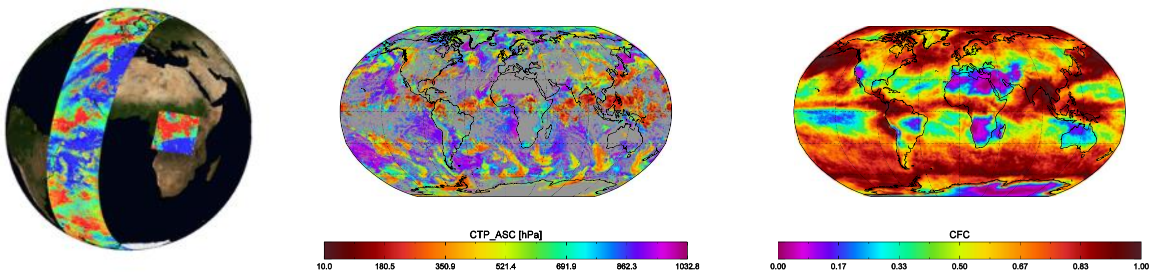


Figure 2-1 *Examples of Cloud_cci cloud products. Left: Pixel-based (Level 2), middle: daily composite on a global grid (Level 3U), right: monthly averaged on a global grid (Level 3C)*

To make the Cloud_cci datasets improved compared to existing ones, the following two essential steps were undertaken:

1) Revisit the measurement data (Level-1) and corresponding calibration performance and development of a carefully inter-calibrated and rigorously quality checked radiance data sets for AVHRR, so called Fundamental Climate Data Record (FCDR). Within this effort the calibration of AVHRR, MODIS and AATSR was compared and characterized. Please see the ATBDv5 for more information about all sensors used and their imaging characteristics. More information on the AVHRR FCDR produced and used is available in RAFCDRv1.0.

2) Development of two state-of-the-art physical retrieval systems that use the optimal estimation technique for a simultaneous, spectrally consistent retrieval of cloud properties including pixel-based uncertainty measures. The first retrieval framework is the Community Cloud retrieval for Climate (CC4CL; Sus et al., 2017; McGarragh et al., 2017) which is applied to AVHRR and AVHRR-heritage channels (i.e. channels which are available from all sensors) of MODIS and AATSR. The second retrieval framework is the Freie Universität Berlin AATSR MERIS Cloud retrieval (FAME-C; Carbajal Henken et al., 2014) and is applied to synergistic MERIS and AATSR measurements on-board of ENVISAT.

Based on these developments, six multi-annual, global datasets of cloud properties were generated using the passive imager satellite sensors AVHRR, MODIS, (A)ATSR and MERIS. These datasets were comprehensively evaluated (1) by using accurate reference observations of ground stations and space-based Lidar measurements and (2) by comparisons to existing and well-established global cloud property datasets.

All parts of the datasets generation effort were properly documented with the major components being the Algorithm Theoretical Baseline Documents (ATBD; ATBDv5, ATBD-FAME-Cv5, ATBD-CC4CLv5), the Product User Guide (PUGv3.1) and this Product Validation and Intercomparisons Report (PVIR; PVIRv4.1).

Furthermore, to facilitate the utilization for evaluation of regional and global atmospheric models, the development of a satellite simulator package for Cloud_cci datasets were fostered, which is planned to be part of one of the upcoming releases of the CFMIP Observation Simulator Package (COSP, Bodas-Salcedo et al. 2011).

2.2 The Cloud_cci datasets

In Cloud_cci two families of global cloud property datasets have been generated. The first family comprises datasets for individual sensor groups such as AVHRR, MODIS, ATSR2/AATSR, for which the AVHRR-heritage channels (0.6, 0.8, 1.6/3.7, 10.8, 12.0 μm) were utilized to retrieve cloud properties using the CC4CL algorithm. The second family comprises a dataset of cloud properties retrieved from simultaneous usage of AATSR and MERIS sensors (both mounted on ENVISAT) by applying the FAME-C algorithm. Since MODIS and AVHRR sensors are separated into morning and afternoon orbits, 6 distinct Cloud_cci datasets exist, which can be seen in Figure 2-2. In addition,

Table 2-1 summarizes the algorithms, sensors and satellites used for each dataset. The official versions of the datasets, as released under the issued Digital Object Identifies (DOIs, see Table 2-2), do not contain any diurnal cycle or satellite drift correction. Potential methods for such a drift correction were investigated for AVHRR and were documented in RODCv1.0. In Figure 2-3 the local observation time of each individual sensor considered are visualized. This information is often essential for properly characterizing time series of cloud properties derived from the satellite-based climate datasets. Other important aspects are the imaging properties. The sensors differ in terms of native footprint resolution (1x1km² for ATSR2, AATSR, MERIS, MODIS; 5x1km² for AVHRR). This, together with the sensor swath width, lead to very different observation frequency and spatial coverage. The latter is visualized in Figure A-18 in PUGv3.1. While MODIS and AVHRR have a complete global coverage within a day, the AATSR sensor needs about 3 days to accomplish this, however, with a higher spatial resolution compared to AVHRR.

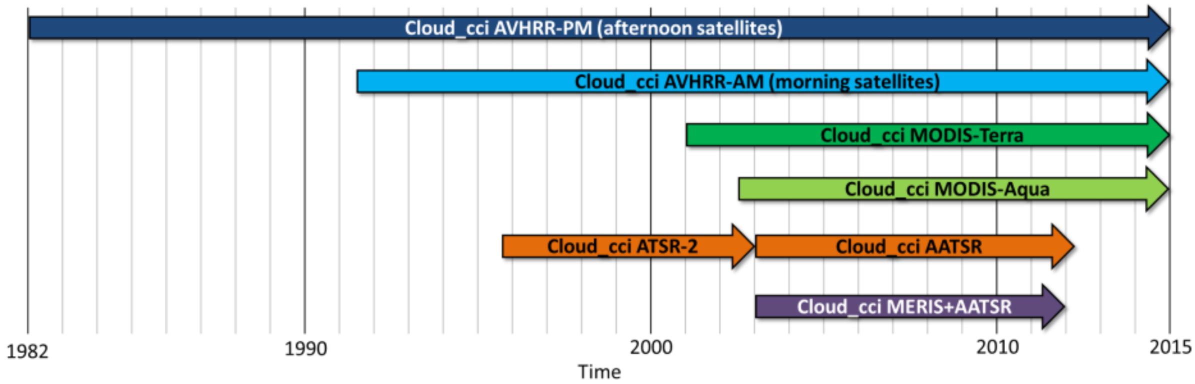


Figure 2-2 *Overview of Cloud_cci datasets and the time periods they cover.*

All datasets contain identical sets of cloud properties: cloud mask/fraction (CMA/CFC), cloud phase/liquid cloud fraction (CPH), cloud top pressure/height/temperature (CTP/CTH/CTT), cloud effective radius (CER), cloud optical thickness (COT), spectral cloud albedo at two wave lengths (CLA) and liquid/ice water path (LWP/IWP). The data is presented at different processing levels ranging from pixel-based retrieval products (Level-2), which are additionally projected (sampling – no averaging) onto a global Latitude-Longitude grid of 0.05° resolution (global composite, Level-3U), to monthly data summarizes including averages, standard deviation and histograms - all defined on a global Latitude-Longitude grid of 0.5° resolution (Level-3C). See Section 2.3 for more details.

All cloud properties (except CPH) are accompanied by uncertainty measures at all processing levels, which range from optimal estimation based uncertainty on pixel level (Level-2 and Level-3U) to propagated uncertainties in the monthly Level-3C products. See Section 2.2 for more information.

In addition to the passive imager based datasets mentioned so far, in Cloud_cci an IASI-based demonstrator dataset has been created, with more details to be found in Feofilov et al. (2017) and Stubenrauch et al. (2017).

Key strengths of Cloud_cci datasets:

- The Cloud_cci datasets are based on two newly-developed, state-of-the art retrieval systems named CC4CL and FAME-C that use the optimal estimation (OE) technique and are applied to passive imager sensors of current and past European and non-European satellite missions.
- The measurement records of the utilized sensors have been revisited, re-characterized and, in case of AVHRR, re-calibrated.
- Two special features of CC4CL and FAME-C are, among others, their applicability to multiple sensors: ATSR2, AATSR, MODIS, AVHRR (CC4CL) and the simultaneous utilization of AATSR and MERIS measurements (FAME-C, i.e. utilizing the O2-A band of MERIS) down to spatial footprint resolutions of 1km.
- Radiative consistency of derived cloud parameters is achieved by the OE-based, iterative fitting of a physically consistent cloud model (and radiative transfer simulations therefrom) to the sensor measurements in the visible and thermal infrared spectral range.
- Pixel-level uncertainty characterization is facilitated by the OE technique, which is physically consistent (1) with the uncertainties of the input data (e.g. measurements, a-priori) and (2) among the retrieved variables. These pixel-level uncertainties are further propagated into the monthly products using a developed sound mathematical framework.
- Potential to combine AVHRR-heritage datasets to achieve increased temporal resolution by including multiple polar-orbiting satellite instruments, which also allows for mature cloud property histograms on 0.5° resolution due to highly increased sampling rate.
- Comprehensive assessment and documentation of the retrieval schemes and the derived cloud property datasets, including possibilities of drift- and diurnal cycle corrections.
- Availability of a developed Cloud_cci satellite simulator facilitating the applicability of Cloud_cci data in regional and global climate models evaluation efforts.
- All datasets are available in netcdf (v4) format and fulfil high CCI-internal and external data standards (e.g. Climate and Forecast – CF conventions).

Table 2-1 *Cloud_cci datasets with the algorithms, sensor(s) and satellite(s) used and the time periods they cover. The Digital Object Identifiers (DOI) of all datasets are also listed.*

Dataset name	Sensor(s)	Satellite(s)	Time period	Algorithm
Cloud_cci AVHRR-PM DOI:10.5676/DWD/ESA_Cloud_cci/AVHRR-PM/V002	AVHRR-2/-3	NOAA-7,-9,-11,-14,-16,-18,-19	1982-2014	CC4CL
Cloud_cci AVHRR-AM DOI:10.5676/DWD/ESA_Cloud_cci/AVHRR-AM/V002	AVHRR-2/-3	NOAA-12,-15,-17, Metop-A	1991-2014	CC4CL
Cloud_cci MODIS-Terra DOI:10.5676/DWD/ESA_Cloud_cci/MODIS-Terra/V002	MODIS	Terra	2000-2014	CC4CL
Cloud_cci MODIS-Aqua DOI:10.5676/DWD/ESA_Cloud_cci/MODIS-Aqua/V002	MODIS	Aqua	2002-2014	CC4CL
Cloud_cci ATSR2-AATSR DOI:10.5676/DWD/ESA_Cloud_cci/ATSR2-AATSR/V002	ATSR2, AATSR	ERS2, ENVISAT	1995-2012	CC4CL
Cloud_cci MERIS+AATSR DOI:10.5676/DWD/ESA_Cloud_cci/MERIS+AATSR/V002	MERIS, AATSR	ENVISAT	2003-2011	FAME-C

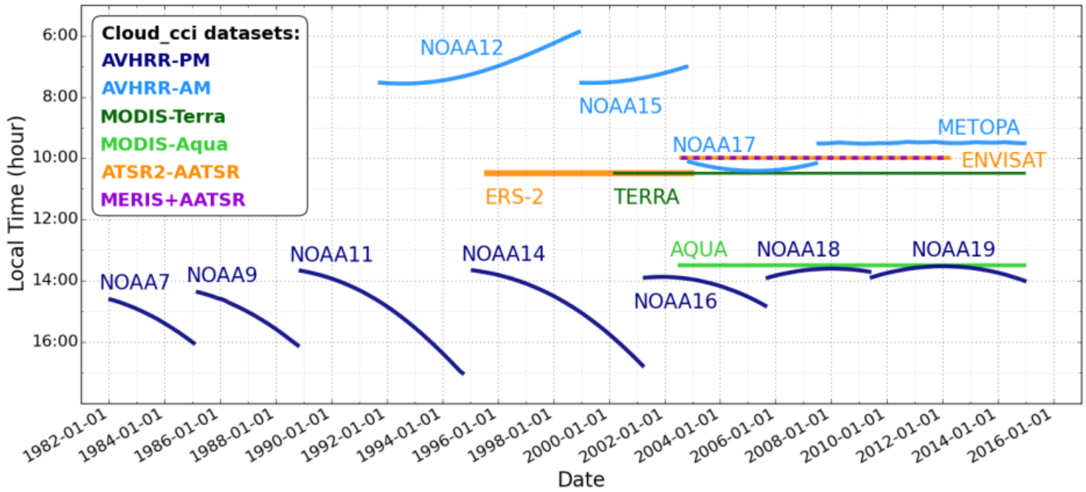


Figure 2-3 *Time periods and local observation times (equator crossing times) of each satellite sensor considered in Cloud_cci. Figure is taken from Stengel et al. (2017).*

2.3 Cloud_cci cloud products

The cloud properties derived on pixel level of each utilized sensor are listed in Table 2-2. It is important to note that the properties CLA, LWP, IWP are not directly retrieved, but rather determined from retrieved COT and CER in a post processing step. The same applies to CTH and CTT, which are inferred from the retrieved CTP. Based on these pixel level retrievals the data is further processed into different processing levels as summarized in Table 2-3. Level-3U denotes a composite on a global Latitude-Longitude grid (of 0.05° resolution) onto which the Level-2 data is sampled (see ATBDv5 for more details on Level-3U sampling). Level-3C products are also defined on Latitude-Longitude grid (here 0.5° resolution) onto which the properties are averaged or their frequency collected (histograms). Further separation of cloud properties in Level-3C in e.g. day/night, liquid/ice, were made wherever suitable (see Table 2-4). Level-3S products are generated merging the Level-3C of all individual sensors. Using Level-3S products requires careful consideration of the partly large and time-varying discrepancies between the used sensors. Please contact the Cloud_cci team for more information (http://www.esa-cloud-cci.org/?q=support).

Table 2-2 *List of generated cloud properties. CMA/CFC and CPH are derived in a pre-processing step. In the next step, COT, CER and CTP are retrieved simultaneously by fitting a physically consistent cloud/atmosphere/surface model to the satellite observations using optimal estimation (OE). Moreover, LWP and IWP are obtained from COT and CER. In addition, spectral cloud albedo (CLA) for two visible channels are derived.*

Variable	Abbrev.	Definition
Cloud mask / Cloud fraction	CMA/ CFC	A binary cloud mask per pixel (L2, L3U) and therefrom derived monthly total cloud fractional coverage (L3C, L3S) and separation into 3 vertical classes (high, mid-level, low clouds) following ISCCP classification (Rossow and Schiffer, 1999).
Cloud phase	CPH	The thermodynamic phase of the retrieved cloud (binary: liquid or ice; in L2, L3U) and the therefrom derived monthly liquid cloud fraction (L3C, L3S).

Variable	Abbrev.	Definition
Cloud optical thickness	COT	The line integral of the absorption coefficient and the scattering coefficient (at 0.55µm wavelength) along the vertical in cloudy pixels.
Cloud effective radius	CER	The area-weighted radius of the cloud drop and crystal particles, respectively.
Cloud top pressure/ height/ temperature	CTP/ CTH/ CTT	The air pressure [hPa] /height [m] /temperature [K] of the uppermost cloud layer that could be identified by the retrieval system.
Cloud lLiquid water path/ Ice water path	LWP/ IWP	The vertical integrated liquid/ice water content of existing cloud layers; derived from CER and COT. LWP and IWP together represent the cloud water path (CWP)
Joint cloud property histogram	JCH	This product is a spatially resolved two-dimensional histogram of combinations of COT and CTP for each spatial grid box.
Spectral cloud albedo	CLA	The blacksky cloud albedo derived for channel 1 (0.67 µm) and 2 (0.87 µm), respectively (experimental product)

Table 2-3 *Processing levels of Cloud_cci data products. Level-3U, Level-3C and Level-3S are each directly derived from Level-2.*

Processing level	Spatial resolution	Description
Level-2 (L2)	MODIS: 1km AATSR: 1km AVHRR: 5 km MERIS+ AATSR: 1km	Retrieved cloud variables at satellite sensor pixel level, thus with the same resolution and location as the sensor measurements (Level-1)
Level-3U (L3U)	Latitude-Longitude grid at 0.05° res. (MODIS-Europe: 0.02°)	Cloud properties of Level-2 orbits projected onto a global space grid without combining any observations of overlapping orbits. Only subsampling is done. Common notation for this processing level is also L2b. Temporal coverage is 24 hours (0-23:59 UTC).
Level-3C (L3C)	Latitude-Longitude grid at 0.5° res.	Cloud properties of Level-2 orbits of one single sensor combined (averaged / sampled for histograms) on a global space grid. Temporal coverage of this product is 1 month.
Level-3S (L3S)	Latitude-Longitude grid at 0.5° res.	Cloud properties of Level-2 orbits of all available single sensors combined (averaged / sampled for histograms) on a global space grid. Temporal coverage of this product is 1 month.

Table 2-4 *Cloud_cci product features incl. day and night separation, liquid water and ice as well as histogram representation. Level-3U refers to the non averaged, pixel-based cloud retrievals sampled onto a global Latitude-Longitude (lat/lon) grid. [1]CMA in Level-2 and Level-3U is a binary cloud mask. All products listed exist in each dataset listed above.*

	Level 2 **swath based** 1km/5km	**Level-3U** **daily sampled** **global** 0.05° lat/lon grid	**Level-3C** **monthly averages** **global** 0.5° lat/lon grid	**Level-3C** **monthly histograms** **global** 0.5° lat/lon grid
CMA/CFC	✓ as CMA[1]	✓ as CMA[1]	✓day/night/high/mid/low	-
CTP, CTH, CTT	✓	✓	✓	✓ liquid/ice
CPH	✓	✓	✓ day/night	-
COT	✓	✓	✓ liquid/ice	✓ liquid/ice
CER	✓	✓	✓ liquid/ice	✓ liquid/ice
LWP	✓ as CWP	✓ as CWP	✓	✓ as CWP
IWP			✓	
CLA	✓ 0.6/0.8µm	✓ 0.6/0.8µm	✓ 0.6/0.8µm	✓ 0.6/0.8µm/liquid/ice
JCH	-	-	-	✓ liquid/ice

2.4 Uncertainties

The retrieved cloud properties CMA, CTP, CTT, CTH, COT, CER, LWP and IWP (for CC4CL also CLA) are accompanied by pixel-based (Level-2) uncertainties, which are output of the OE technique and represent a rigorous propagation of the uncertainties in the input data, e.g. a-priori information, measurements, radiative transfer. These uncertainties values represent the 68% confidence interval of the true value being within the retrieved value ± uncertainty. These Level-2 uncertainties are also given in Level3U and further propagated into Level-3C. For this a sound mathematical framework has been developed and implemented taking into account the retrieval uncertainties but also the uncertainty correlations. The framework allows an estimation of both the real variability of the observed property and the uncertainty of the calculated mean. Determine and utilizing the uncertainty correlation is a particular key point for an appropriate propagation of Level-2 uncertainties into higher-level products (e.g. Level-3C). Please see the Comprehensive Error Characterization Report (CECRv3) and Stengel et al. (2017) for further details on the uncertainty measures provided. Results of uncertainty validation are given Section 6.

2.5 Validation strategy in this report

2.5.1 Evaluation measures

For geophysical quantities at Level-2 and Level-3U, such as cloud top height, and for aggregated products (Level-3C), we use the bias, i.e. mean difference between Cloud_cci and reference data as the metric for accuracy. In addition, the bias corrected root mean squared error (bc-RMSE) is used to express the precision of Cloud_cci compared to a reference data record.

Bias (accuracy):	Mean difference between Cloud_cci and reference data
bc-RMSE (precision):	Bias corrected root mean squared error to express the precision of Cloud_cci compared to a reference data record
Stability:	The ability of preserving the same bias throughout the time period.

In case of discrete Level-2 and Level-3U variables with only two possible events, e.g. cloud mask (*clear* or *cloudy*) and cloud phase (*liquid* or *ice*), we use the following scores which can be derived from the contingency table (Table 2-5).

Score	Description	Example
POD:	Probability of Detection: The fraction of correct Cloud_cci reports of a particular category relative to all reference reports of this category.	POD for event 1: $\frac{n_{11}}{n_{11}+n_{12}}$
FAR:	False Alarm Rate: The fraction of incorrect Cloud_cci reports of a particular category relative to all Cloud_cci reports of this category.	FAR for event 1: $\frac{n_{12}}{n_{11}+n_{12}}$
Hit Rate:	The total fraction of all correct Cloud_cci reports (i.e., summing n11 and n22 in Table 2-5) relative to all reference reports.	Hit rate: $\frac{n_{11}+n_{22}}{n_{11}+n_{12}+n_{21}+n_{22}}$
KSS:	Hanssen-Kuipers Skill Score: This is a measure of correct Cloud_cci reports, with random correct and unbiased reports subtracted out.	KSS: $\frac{n_{11}n_{22}-n_{21}n_{12}}{(n_{11}+n_{21})(n_{12}+n_{22})} \in [-1,1]$

Table 2-5: *Contingency table for the 2x2 problem. n_{ij} is the number of cases where Cloud_cci reports event i and the reference reports event j. For example event 1 may be clear and event 2 may be cloudy.*

	Reference reports 1	Reference reports 2
Cloud_cci reports 1	n_{11}	n_{12}
Cloud_cci reports 2	n_{21}	n_{22}

2.5.2 Datasets used for evaluation

In Table 2-6 the reference data used in this report is listed together with the cloud properties considered and the type of evaluation they are used for in this context. Data of CALIPSO-CALIOP, SYNOP and UWISC are considered as validation reference, while CLARA-A2, PATMOS-x and MODIS Collection 6 are considered for comparison purposes (grey shaded). While for UWisc the reference publication provides an error estimate of 15% - 30% for LWP (O'Dell et al., 2008); the uncertainty of SYNOP and CALIOP data used in this evaluation effort is not known, but assumed to be very small, smaller than the differences to the Cloud_cci data found.

Table 2-6: Reference data used in this report including a description of the type of data, the cloud properties available and the type of reference the data is considered as (validation or comparison).

Reference data	Type of data	Variables	Type of evaluation
CALIPSO-CALIOP	Retrieval based on space-based (active) Lidar measurements (See Section A.2 for details)	Level-2/Level-3U CMA, CTH, CPH	Validation
SYNOP	Ground-based, human observations of cloud cover (See Section A.1 for details)	Level-3C CFC	Validation
UWISC	Retrieval based on space-based (passive) microwave measurement (See Section A.5 for details)	Level-3C LWP	Validation
CLARA-A2	Retrievals based on space-based (passive) visible and infrared measurements of AVHRR (See Section A.6 for details)	Level-3C CFC, CPH, CTP, CER, COT, LWP, IWP, JCH	Comparison
PATMOS-x	Retrievals based on space-based (passive) visible and infrared measurements of AVHRR (See Section A.3 for details)	Level-3C CFC, CPH, CTP, CER, COT, LWP, IWP, JCH	Comparison
MODIS Collection 6	Retrievals based on space-based (passive) visible and infrared measurements of MODIS (See Section A.4 for details)	Level-3C CFC, CPH, CTP, CER, COT, LWP, IWP, JCH	Comparison

3. Validation of Cloud_cci products

3.1 Validation of cloud mask (CMA) and cloud top height (CTH) against CALIOP

An extensive evaluation of CC4CL cloud mask and cloud top height Level-2 products have been carried out based on CALIPSO-CALIOP data from the period 2006-2010, i.e., more than four years of observations (see Section A.2 for details on CALIPSO-CALIOP data). The validation effort did not include the cloud optical thickness parameter since CALIOP cloud optical thickness estimations above approximately the value range 5-8 are not possible (due to saturated signal, i.e., true cloud base is not detected). AVHRR-based results from both morning orbits (for satellites NOAA-17 and METOP-A) and afternoon orbits (for satellites NOAA-18 and NOAA-19) have been studied. In addition, also CC4CL ENVISAT products from AATSR and the synergy FAME-C cloud products have been evaluated for the same period. The collected data is more or less equally spread over the 5 year period mentioned above and over season.

The validation effort was based on CALIPSO-CALIOP cloud observations which were simultaneously (i.e., within 3 minutes) observing the same spot on Earth as the other satellites. For afternoon satellites (here exclusively AVHRR) this means that comparisons were made in near-nadir observation conditions, which is very close to the nadir observation condition for CALIOP, while for morning satellites all possible viewing angles were included because matchups then occurred across the swath and not along-track as for afternoon satellites. This also means that global matchups exist for afternoon satellites while only high-latitude matchups close to +/- 72 degrees latitude exist for morning satellites. A more detailed description of the validation method is given by Karlsson and Johansson (2013).

Regarding the study of ENVISAT products it should be emphasized that while AVHRR products are defined in GAC resolution (5 km) the AATSR/MERIS case are defined in a finer 1 km resolution. This means that it is appropriate to match with CALIPSO 1 km cloud products which unfortunately means that we have no estimations of the cloud optical thicknesses of individual cloud layers (these are exclusively prepared for CALIPSO 5 km products). Another consequence of this is that the amount of very thin clouds in the CALIPSO 1 km product will be lower which then falsely may improve results somewhat when compared to the 5 km results. All these differences must be kept in mind when comparing results between afternoon and morning satellites and between NOAA/METOP satellites and ENVISAT.

3.1.1 Cloud products from AVHRR-AM and AVHRR-PM

All in all, data from 1710 orbits were compared for afternoon satellites, generating more than 9 million individual FOV matchups, and data from 932 orbits were compared for morning satellites, generating half a million matchups. A number of commonly used statistical scores were calculated (described in more details in the previous section 3 and by Karlsson and Johansson (2013) and results were also stratified according to illumination, surface and geographical conditions (e.g., day, night, twilight, tropics, sub-tropics, high-latitudes, polar, snow-cover, ice-cover, mountain, low-land, etc.). Results were also sub-sampled based on CALIPSO-retrieved cloud optical thicknesses in order to account for the fact that the thinnest CALIOP-detected clouds cannot be detected from passive imagery. A sub-set of all these results are visualised and tabulated in the following. Results are also compared to corresponding validation results for the CM SAF CLARA-A2 data record which is released in the autumn of 2016.

<u>**Results for cloud fraction**</u>

A summary of the overall validation scores is given in Table 3-1 (first two columns) which also shows corresponding scores for the CLARA-A2 data record for the same period in time. The small discrepancy in number of orbits and matched FOVs between the two datasets is explained by a different treatment of a small number of orbits passing the date-line (e.g., 00 UTC). The part occurring after midnight (and added as a separate orbit for CLARA-A2) was unfortunately left out in the CC4CL dataset. However, this has only very marginal effect on results, i.e., the difference in matched FOVs is only about 4 %.

Table 3-1 shows generally that the scores for the two data records are very close or sometimes actually identical. However, we notice a lower bias for CC4CL (-11.3 %) compared to CLARA-A2 (-15.1 %). This explains a slightly higher Hitrate (81.4 % vs 79.7 %) for CC4CL. On the other hand, no difference is seen for the Kuipers score which is a better score for evaluating the success of cloud detection in cases when one of the categories cloudy and clear dominates (in this case we have average cloud cover above 70 % according to CALIPSO). Consequently, even if CC4CL detects more clouds there is also an over-prediction of clouds which is seen on the higher false alarm rate for cloudy conditions (5.9 % vs 4.4 %). This explains why we don't see a clear difference in the Kuipers score.

Because of the large number of realised matchups it is now possible to display the global geographical distribution of the validation scores. Figure 3-1 shows the overall fraction of correct cloudy and cloud-free cloud masks (Hitrate) for all afternoon satellites (NOAA-18 and NOAA-19) based on all available CALIPSO-CALIOP cloud masks (same data as being summarised in in columns 1 and 2 in Table 3-1). Here we have used a global equal-area Fibonacci grid (Gonzáles, 2009) with 150 km resolution to display results. Grid cells with a too low number of matchups are displayed in white colour (occurring close to the poles and sporadically close to the equator).

Table 3-1: *Summary of validation scores for afternoon (NOAA-18 and NOAA-19) and morning (NOAA-17 and Metop-A) satellites 2006-2010 after optical thickness filtering with threshold 0.15. Data used spans 51 months. N stands for NOAA and M for Metop.*

	Afternoon satellites no COD threshold		Afternoon satellites with COD threshold		Morning satellites with COD threshold	
	CC4CL (N-18,N-19)	CLARA-A2 (N-18,N-19)	CC4CL (N-18,N-19)	CLARA-A2 (N-18,N-19)	CC4CL (N-17,M-A)	CLARA-A2 (N-17, M-A)
Number of orbits	1710	1842	1710	1842	932	966
Matched FOVs	9 423 820	9 840 323	9 423 820	9 840 323	538 294	548 044
Bias	-11.3 %	-15.1 %	-0.8 %	-3.1 %	3.5 %	-8.2 %
RMSE	41.6%	42.5 %	39.1%	39.8 %	49.5%	44.0 %
POD cloudy	79.5 %	76.3 %	87.1 %	84.8 %	82.9 %	77.3 %
POD clear	86.7%	89.7 %	80.8%	82.9 %	63.2%	84.3 %
FAR cloudy	5.9 %	4.4 %	11.8 %	10.8 %	21.5 %	11.3 %
FAR clear	38.8 %	43.6 %	20.9 %	23.5 %	30.3 %	30.6 %
Kuipers score	0.66	0.66	0.68	0.68	0.46	0.62
Hit Rate	81.4%	79.7 %	84.7%	84.1 %	75.4%	79.9 %

We notice that scores are generally good everywhere except in the Tropical Region, over the eastern part of the Eurasian continent and over the Polar Regions. However, we also know that CALIPSO-CALIOP has much higher cloud detection sensitivity than what can be achieved from methods based on passive imagery (explaining a large part of the negative bias found in first two columns of Table 3-1). This means that we cannot determine from Figure 3-1a where we truly miss clouds due to real method weaknesses or due to theoretical limits. One way of reducing the impact of these theoretical limits of detecting very thin clouds is to remove them from the comparison or, in this case, treat them as non-existing (i.e., changed from cloudy to cloud-free observations). By doing this, cloud-detection problems related to optically thicker clouds will stand out more clearly. If we make this filtering of results with increasing CALIOP-determined cloud optical depths, the Hitrate score will increase until it peaks at an optical thickness value where 50 % of the clouds having this optical thickness value are detected. A filtering with a higher optical thickness value will degrade results since too many correctly detected clouds will then be interpreted as false clouds. Analysis of filtered results from CC4CL and CLARA-A2 showed that the peak in Hitrate occurs at an optical thickness value of 0.15 for both datasets. Figure 3-1b shows the same results as in Figure 3-1a but now with an optical thickness filtering based on this threshold.

We notice that results improve considerably over the Tropical Region and we conclude that the lower score here in Figure 3-1a was mainly related to sub-visible cirrus clouds. Remaining problems are seen mainly over the Polar Regions and in particular over the snow-covered land portions (i.e., Greenland and Antarctica). Here it is clear that also optically thick clouds are missed at a higher frequency. Lower Hitrates can be seen over sub-tropical ocean areas and over mountainous and high-latitude regions of the Eurasian continent and North-America. Regarding the decreasing skill over sub-tropical ocean, we believe that this is an effect of a larger dominance of small sub-pixel scale cloud elements. This will have two effects: 1. Increasing the risk of mismatches between CALIPSO and AVHRR due to very different sampling of the true 5 km GAC FOV for each pixel. In areas where clouds are more generally occurring at larger scales than the 5 km GAC FOV scale the matching agreement is much better (e.g. over high-latitude storm tracks). 2. Increasing risk of truly missing clouds in the cloud mask due to their sub-pixel nature.

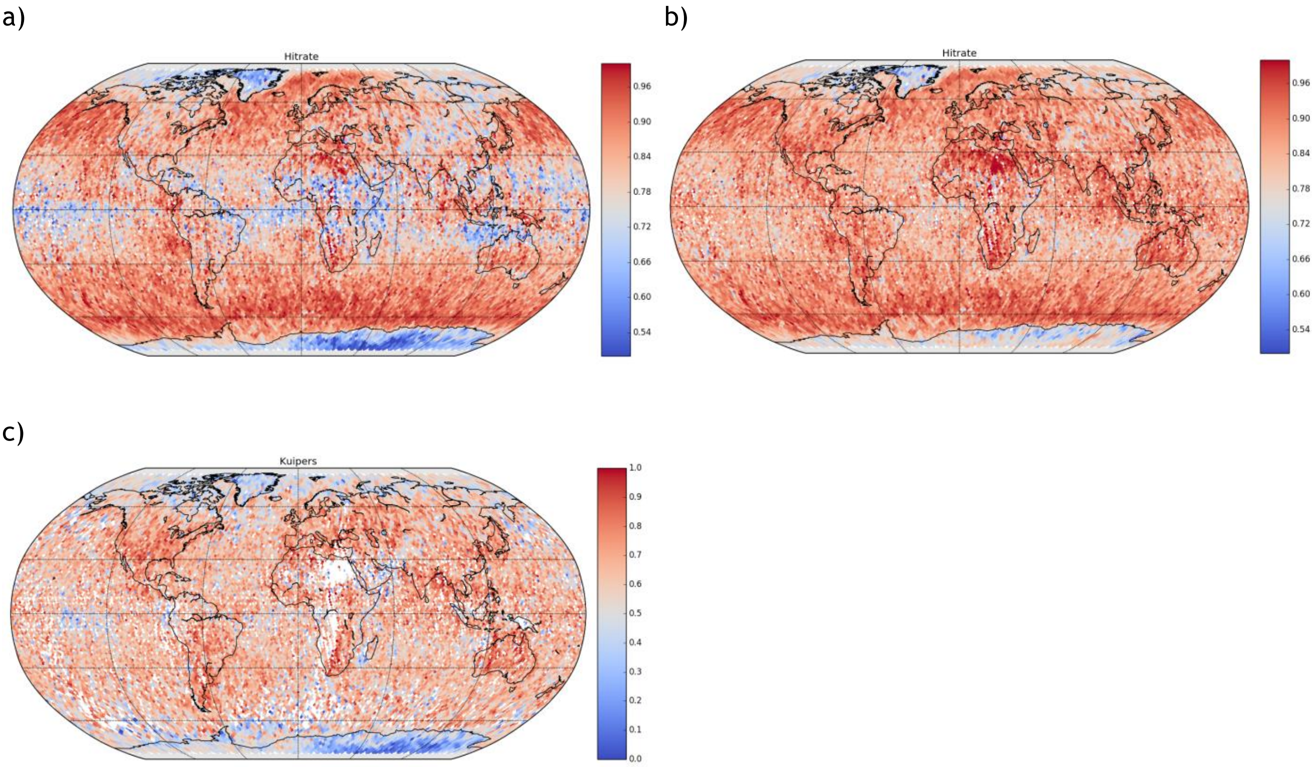


Figure 3-1: a) *Global plot of Hitrate for the CC4CL AVHRR cloud mask for afternoon satellites (NOAA-18 + NOAA-19) in the period October 2006-December 2010. b) Same as panel (a) but here all CALIPSO-CALIOP detected clouds with optical thicknesses below 0.15 are ignored and treated as being cloud-free. c) Same as panel (b) but for the Kuipers score.*

A closer scrutiny of the efficiency of the separation of cloud-free and cloudy conditions can be made by plotting the same results after optical thickness filtering for the Kuipers score (Figure 3-1c). Here we see that over the Polar Regions scores go in some places down to almost a zero level which practically means that we have no correlation at all in our results (i.e., random results). Some new problem areas appear like over the Arabian Peninsula and close to or slightly south of the equator in the Pacific Ocean. However, since the estimation of the Kuipers score requires presence of both cloudy and clear cases the coverage of valid results in each grid square is less good compared to the Hitrate (explaining also the increasing number of white areas in Figure 3-1c). Thus, more studies with more data seem necessary to pin point these features in more detail.

Columns 3 and 4 of Table 3-1 give the updated results after having performed optical thickness filtering with an optical thickness threshold of 0.15. Results are generally improving for both data records but conclusions remain basically the same concerning agreements and differences between the two data records.

Columns 5 and 6 of Table 3-1 give the corresponding (with applied optical thickness filtering) results for all morning satellites (NOAA-17 and METOP-A) in the same period. To be remembered here is that collocations with CALIPSO-CALIOP observations for morning satellites can only be made at high latitudes since CALIPSO-CALIOP is placed in an afternoon orbit. Also, the number of collocations will be smaller even at high latitudes since matchups will take place across the AVHRR swath and not in the along-track direction. Furthermore, matchups will take place at the full range of satellite viewing angles and not close to nadir as for afternoon satellites.

We notice that the overall results are not as good as for the afternoon satellites which are mainly explained by the large influence of polar conditions with sometimes difficult cloud detection conditions. However, results for CLARA-A2 seem to be markedly better here. Especially, the difference in the Kuipers score indicates particular problems for CC4CL in the correct separation of cloudy and cloud-free conditions. The positive bias and the high false alarm rate for cloudy conditions indicate that too many false clouds are produced. A closer scrutiny of results shows that this problem is mainly a problem occurring during the Polar Day or Polar Summer for snow- and ice-covered surfaces. It is further illustrated in Figure 3-2 showing the false alarm rate for cloudy conditions during the Polar Day for the two data records. We clearly see a more frequent appearance for CC4CL of false alarm rates above 25 % occurring primarily over the snow- and ice-covered parts of the Arctic. This problem is not present during the Polar Night when both methods have low scores (with probabilities of detecting cloudy conditions below 70 %).

The same difference between CC4CL and CLARA-A2 results for Polar Summer conditions is seen also for afternoon satellites (not shown here). The main difference regarding the data usage between the two methods during daytime conditions is that CLARA-A2 uses shortwave infrared channels (3.7 micron for afternoon satellites and 1.6 micron for morning satellites) in addition to the two visible and the two thermal infrared channels while CC4CL only uses the visible and thermal infrared channels. This emphasizes the great importance of using shortwave infrared channels for cloud detection over snow- and ice-covered surfaces. For next versions of the Cloud_cci datasets, the shortwave IR channels will be used in CC4CL.

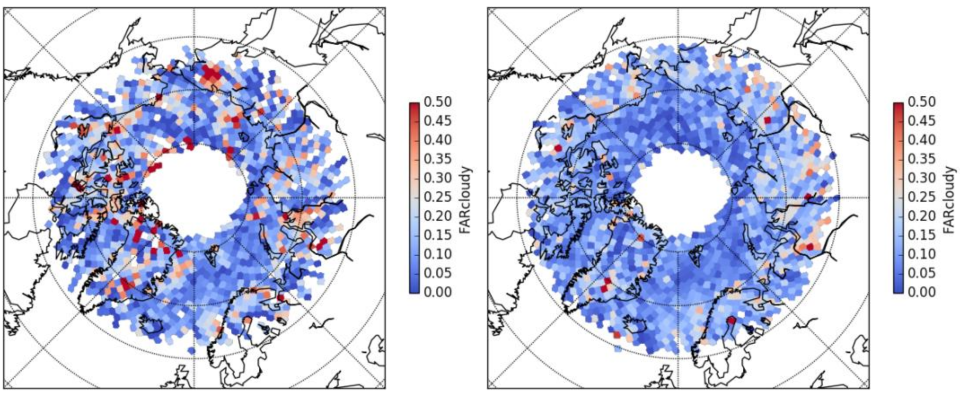


Figure 3-2: *False alarm rate for cloudy conditions over the Arctic region for morning satellites (NOAA-17 + METOP-A) during the Polar Day (here defined as for solar zenith angles below 80 degrees). Results for CC4CL are shown to the left and for CLARA-A2 to the right.*

As stated initially, this work included also separate studies for different times of day (daytime, twilight, night), latitude bands (tropical, sub-tropical, high-latitudes and polar) and surfaces (ocean, land, snow-cover, ice-cover). It is not possible to present the detailed results here but the following features can be highlighted:

- CC4CL results for twilight conditions (solar zenith angles between 80-95 degrees) are inferior to CLARA-A2 results (e.g., false alarm rate for cloudy conditions is twice as large as for CLARA-A2 afternoon satellites). Since results for morning satellites are more exposed to twilight conditions this could also explain the slightly poorer CC4CL performance for morning satellites compared to CLARA-A2.
- CC4CL is generally slightly better than CLARA-A2 over ice-free ocean surfaces at night.
- The two previous points plus the problematic conditions over snow- and ice-covered surfaces for CC4CL explain largely why overall global results for afternoon satellites are still very comparable (i.e., evened out by compensating factors).
- Both methods show decreased scores over sub-tropical ocean regions (as indicated in Figure 3-1). Also some adjacent land areas (e.g. eastern South-America and eastern Africa) show this feature. We suspect that this is mainly due to collocation problems and not to real cloud detection problems. In these regions small-scale (and sub-pixel scale) cumulus cloudiness is frequent which increases the risk of failure in finding the same cloud elements in collocated AVHRR and CALIOP FOVs.
- Concerning the global figures for afternoon satellites it is interesting to notice that neither of the two methods is capable of providing a probability of detection of cloudy conditions substantially larger than 90 % for an optical thickness filtering value of 1.0. Thus, about 10 % of all clouds remain undetected even if they are optically thick. A majority of those missed clouds appear to belong to the Polar Regions during the Polar Winter. However, a certain fraction of this mis-match comes from collocation errors and some further studies are necessary here for estimating exactly this portion of the missing clouds.

Results for cloud top height

Exactly the same collocated AVHRR/CALIOP dataset as for cloud fraction has been used to evaluate the CC4CL product. Also here comparisons have been made with the CLARA-A2 dataset. Comparisons have been performed against the highest CALIOP-detected cloud layer but also here some filtering of the thinnest clouds has been done (Table 3-2). Results shown are for afternoon satellites. Also shown are the results for the corrected cloud top height product of CC4CL. Results for morning satellites are excluded here due to the limited coverage and the specific cloud detection problems encountered near the Poles.

Results have also been further subdivided into low-level, mid-level and high-level clouds following the cloud type classification provided by the CALIPSO-CALIOP product (Table 3-3).

We notice that the standard CC4CL product does not differ very much from corresponding CLARA-A2 products but with one exception: Results for low-level clouds are considerably better (bias of 138 m compared to 500 m). Otherwise we see very large underestimations of high-level cloud heights. However, Table 3-2 also shows that the recently implemented correction of cloud top heights has a relatively large impact on especially high-level clouds (reducing bias from -4120 m to -3005 m) although at the expense of increasing bias-corrected RMSE.

Table 3-2: *Summary of corrected^T and uncorrected CC4CL (AVHRR-PM) and CLARA-A2 validation results for cloud top height based on afternoon satellites (NOAA-18 + NOAA-19).*

Cloud category	CC4CL^T (AVHRR-PM) Bias (m)	CC4CL^T (AVHRR-PM) Bias-corrected RMSE (m)	CC4CL (AVHRR-PM) Bias (m)	CC4CL (AVHRR-PM) Bias-corrected RMSE (m)	CLARA-A2 Bias (m)	CLARA-A2 Bias-corrected RMSE (m)
All clouds	-1688	3853	-2441	3674	-2386	3577
Low-level	222	1155	138	963	500	1259
Medium-level	-149	2409	-759	1887	-494	1611
High-level	-3005	4427	-4120	3918	-4268	3509

The high negative bias for high clouds is mainly explained by a higher occurrence of optically thin clouds at high levels. Cloud top height retrievals for passive imagery (with no CO_2-channels) have problems to cope with these semi-transparent clouds. To compensate for this one can remove very thin cloud layers in the same way as was done for the cloud fraction study. In this case, we have removed the uppermost cloud layers with a total integrated cloud optical depth less than 1.0. The corresponding results for the high cloud layers are shown in Table 3-3.

We notice that the filtering operation gives the best results for the uncorrected CC4CL results. However, since the applied cloud top height correction aims at finding the uppermost cloud layer rather than the radiatively efficient height we should expect a considerable overestimation here.

Table 3-3: *Summary of CC4CL unfiltered and filtered validation results for corrected^T and uncorrected cloud top height retrievals for high-level clouds based on afternoon satellites (NOAA-18 + NOAA-19). Filtering used the optical thickness threshold 1.0 for removing the influence of the uppermost thin cloud layers.*

Cloud category	CC4CL^T (AVHRR-PM) Bias (m)	CC4CL^T (AVHRR-PM) Bias-corrected RMSE (m)	CC4CL (AVHRR-PM) Bias (m)	CC4CL (AVHRR-PM) Bias-corrected RMSE (m)
Unfiltered high-level clouds	-3005	4427	-4120	3674
Filtered high-level clouds	1474	3370	51	2964

3.1.2 Cloud products from MODIS-Aqua and MODIS-Terra

The collocations carried out for MODIS are, in contrast to the other sensors, based on L3U data. This has been done to reduce the amount of data that needs to be handled. The advantage is, that we can now span a time period of 2006 to 2014 in this validation. The disadvantage is the introduction of a slightly increased spatial mismatch between MODIS and CALIOP. This however is a random effect and should not affect the scores significantly.

Results for cloud fraction

As for the other sensors, the cloud mask validation is carried out two times: (1) including all CALIOP clouds as reference and (2) removing the thinnest clouds (COT below 0.15) from CALIOP data. The calculated scores are reported in Table 3-4.

While MODIS-Aqua cloud mask shows a very well balanced behaviour, meaning similar POD scores for cloudy and clear, the scores POD scores for MODIS-Terra show a significant skewness towards cloudy (PODcloudy=94.4%; PODclear=60.8%), even though the bias is relatively small MODIS-Terra. This is also reflected in the Kuipers scores. The agreement to CALIOP generally improves when removing the thinnest clouds by applying the COT threshold to CALIOP data. Biases are now slightly positive and the RMSE values reduce significantly compared to unfiltered data. The False alarm rate for clear is nearly halved while the cloudy values double.

Table 3-4 *Summary of validation scores for MODIS-Aqua and MODIS-Terra data for 2006-2014 with and without optical thickness filtering with threshold 0.15.*

	no COT threshold		with COT threshold	
	CC4CL MODIS-Aqua	CC4CL MODIS-Terra	CC4CL MODIS-Aqua	CC4CL MODIS-Terra
Matched FOVs	15 224 843	18009	15 224 843	18009
Bias	-9.7 %	-1.1 %	2.3 %	4.7 %
RMSE	63 %	67 %	45 %	51 %
POD cloudy	81.0 %	94.4 %	88.6 %	96.5 %
POD clear	82.7 %	60.8 %	74.8 %	51.6 %
FAR cloudy	6.8 %	4.4 %	14.6 %	8.7 %
FAR clear	40.1 %	45.3 %	20.3 %	26.5 %
Kuipers score	0.64	0.55	0.63	0.48
Hit Rate	81.5 %	91.0 %	84.3 %	89.4 %

Results for cloud top height

Based on the cloudy subset of all inferred collocations the Cloud_cci MODIS CTH has been validated against CALIOP equivalents for both the uppermost cloud layer found in CALIOP data ('not filtered') as well as for taking the CALIOP CTH at the level where the COT above exceeds 1.0 ('Filtered COD>1.0'). Table 3-5 presents the results for MODIS-Terra. Considering all clouds, the bias is -3320m, thus a systematic underestimation of the CTH by MODIS, and the RMSE is 2430m. These values seem comparable ti AVHRR-PM, although the bias is slightly smaller and RMSE larger for MODIS-Terra. Applying the COD threshold removes the thin clouds and thin cloud-top layers. Consequently the bias is much improved (-1190m) as well as the RMSE (1930m). For low level clouds the bias is very small, similar to AVHRR-PM, accompanied by a small RMSE. Biases get more negative when going to mid-level and high clouds. For high clouds, the scores have also been calculated when filtering the thin clouds and cloud layers. The Bias and RMSE significantly improve.

Table 3-6 reports the same validation scores for MODIS-Aqua. The results look generally worse than for MODIS-Terra. The general Bias is -4010m even though the low-level cloud CTH is overestimated (MODIS-Terra has nearly no bias for low-level clouds. The high clouds seem particularly worse that MODIS-Terra, which dominates the overall scores. Introducing the filtering improves the agreement to CALIOP also here significantly. Biases for all and high clouds are even smaller than for MODIS-Terra, however RMSE values are still larger.

Table 3-5 *Summary of CC4CL unfiltered and filtered validation results for uncorrected cloud top height retrievals for Cloud_cci MODIS-Terra for all clouds as well as separated into low, mid-level and high-level clouds. Filtering used the optical thickness threshold 1.0 for removing the influence of the uppermost thin cloud layers.*

	Not filtered		Filtered COD>1	
Cloud category	CC4CL (MODIS-Terra) Bias (m)	CC4CL (MODIS-Terra) Bias-corrected RMSE (m)	CC4CL (MODIS-Terra) Bias (m)	CC4CL (MODIS-Terra) Bias-corrected RMSE (m)
All clouds	-3320	2430	-1190	1930
Low-level	70	900	-	-
Medium-level	-1720	1620	--	-
High-level	-3660	2390	-1940	1450

Table 3-6 *Summary of CC4CL unfiltered and filtered validation results for uncorrected cloud top height retrievals for Cloud_cci MODIS-Aqua for all clouds as well as separated into low, mid-level and high-level clouds. Filtering used the optical thickness threshold 1.0 for removing the influence of the uppermost thin cloud layers.*

	Not filtered		Filtered COD>1	
Cloud category	CC4CL (MODIS-Aqua) Bias (m)	CC4CL (MODIS-Aqua) Bias-corrected RMSE (m)	CC4CL (MODIS-Aqua) Bias (m)	CC4CL (MODIS-Aqua) Bias-corrected RMSE (m)
All clouds	-4010	3880	-900	2300
Low-level	680	1600	-	-
Medium-level	-1340	1780	-	-
High-level	-4320	3890	-1430	2040

3.1.3 Cloud products from ATSR2-AATSR

We have compared CC4CL AATSR cloud mask and cloud top products (as contained in the ATSR2-AATSR dataset) with CALIPSO 1 km cloud layer datasets for the 2007-2010 period (almost identical to the inter-comparison period for CC4CL AVHRR products). We managed to get quite a large number of matched orbits (975 orbits) which is quite comparable to the number of matched orbits for AVHRR data. Results include matchups during daytime for both hemispheres (Arctic and Antarctic matchups at +/- 73 degrees latitude) while matchups during night are exclusively made over the northern hemisphere (Arctic at + 73 degrees). Antarctic night time data are missing due to some data transfer issues.

Results for cloud fraction

Table 3-7 below shows a summary of the results for different illumination categories (daytime, twilight and night). In this compilation, results for twilight are collected for the solar zenith angle interval 80-95 degrees.

Comparing with previous overall results for AVHRR (e.g. Table 3-1) we get about the same results as for the CC4CL AVHRR results. Again, we see degraded results for the twilight category (as was pointed out in the end of the previous section on cloud fraction from AVHRR) where the fraction of false clouds is high (26.5 %). Some other cloud separability issues are also noticed, best shown by the Kuipers score. The AATSR-based Kuipers score for the three illumination categories are 0.57, 0.29 and 0.39, respectively, which can be compared with the CC4CL AVHRR scores 0.55, 0.32 and 0.41 and the CLARA-A2 AVHRR scores 0.64, 0.57 and 0.50. Besides the twilight-problems it is clear that daytime cloud detection over snow-covered surfaces remains problematic. For these surfaces, we have AATSR Kuipers daytime scores of 0.32 to be compared with 0.36 for CC4CL AVHRR and 0.51 for CLARA-A2 AVHRR.

Table 3-7: *Summary of CC4CL AATSR cloud mask validation results, sub-divided into categories daytime, twilight and night.*

	CC4CL AATSR daytime	CC4CL AATSR twilight	CC4CL AATSR night
Number of orbits	975	975	255
Matched FOVs	320 149	258 537	48 777
Bias	-0.7 %	9.6 %	-2.2 %
RMSE	42.7 %	53.6 %	50.6 %
POD cloudy	86.7 %	84.5 %	80.6 %
POD clear	69.9%	44.6%	58.5%
FAR cloudy	12.5 %	26.5 %	16.9 %
FAR clear	31.6 %	38.8 %	45.6 %
Kuipers score	0.57	0.29	0.39
Hit Rate	81.8%	70.3 %	74.4%

Results for cloud top height

Table 3-8 shows a summary of validation results for the CC4CL cloud top height products from AATSR. The right part of the table shows the corresponding CC4CL results from AVHRR afternoon satellites for comparisons. Notice that we only show results for the corrected CC4CL cloud top algorithm here and that corresponding results from AVHRR morning satellites have not yet been evaluated.

Results basically confirm previous results given in Table 3-3 (corresponding to the category "Unfiltered but corrected"). Results are considerably improved compared to original results yielding reduced biases for all vertical cloud levels. However, it is worth pointing out that the seemingly drastically improved results for AATSR high-level clouds compared to AVHRR in Table 3-2 is mostly explained by the differences in horizontal resolution in the examined datasets. The AVHRR products were compared against CALIPSO 5 km cloud layer products including much more of thin high clouds which especially increased the bias for high-level clouds.

Table 3-8: *Summary of CC4CL AATSR cloud top height validation results compared to corresponding CC4CL results for AVHRR afternoon orbit data (see text for explanation). The table shows the results of the uncorrected and the corrected^T cloud top height algorithm.*

Cloud category	CC4CL (ATSR2-AATSR) Bias (m)	CC4CL (ATSR2-AATSR) Bias-corrected RMSE (m)	CC4CL^T (ATSR2-AATSR) Bias (m)	CC4CL^T (ATSR2-AATSR) Bias-corrected RMSE (m)	CC4CL^T (AVHRR-PM) Bias (m)	CC4CL^T (AVHRR-PM) Bias-corrected RMSE (m)
All clouds	-959	2044	-315	2062	-1688	3853
Low-level	304	1062	426	1202	222	1155
Medium-level	-771	1566	-21	2012	-149	2409
High-level	-2683	2079	-1467	2417	-3005	4427

3.1.4 Cloud products from MERIS+AATSR

We have compared FAME-C cloud mask and cloud top products (as contained in the MERIS+AATSR dataset) with CALIPSO 1 km cloud layer datasets for the period 2007-2010 which is almost identical to the inter-comparison period for CC4CL AVHRR products.

Results for cloud fraction

Since the FAME-C cloud products are only retrieved during daytime (and also requiring a certain maximum solar zenith angle, i.e., no twilight conditions) we have a much smaller number of successful inter-comparisons compared to previous studies based on CC4CL AVHRR and AATSR products. In total, 163 ENVISAT orbits where evaluated compared to more than 900 orbits for the two NOAA-17 and METOP-A morning orbit satellites according to Table 3-1. A summary of results is given in Table 3-9, comparing also with previous daytime results of AVHRR from CC4CL and CLARA-A2.

Overall, FAME-C cloud mask results appear quite comparable with CLARA-A2 AVHRR results and clearly better than the CC4CL AVHRR results. However, the difference is probably to some extent explained by the differences in resolution (i.e., the 5 km CALIPSO datasets for the AVHRR study include more of very thin cirrus clouds which are more difficult to detect).

Table 3-9: *Summary of FAME-C cloud mask validation results compared to corresponding results for daytime CC4CL AVHRR and daytime CLARA-A2 AVHRR results.*

	FAME-C (AATSR+MERIS)	CC4CL (AVHRR-AM) (daytime)	CLARA-A2 (AVHRR_AM) (daytime)
Number of orbits	163	932	966
Matched FOVs	88 160	171 353	171 487
Bias	-4.3 %	-8.0 %	-11.0 %
RMSE	41.9%	45.3%	42.0%
POD cloudy	83.8 %	80.5 %	80.3 %
POD clear	78.9%	74.0 %	83.8%
FAR cloudy	10.6 %	9.9 %	6.0 %
FAR clear	30.4 %	43.7 %	42.3 %
Kuipers score	0.63	0.55	0.64
Hit Rate	82.3%	78.9 %	81.2%

Results for cloud top height

Corresponding results for cloud top height is shown in Table 3-10 below. The table inter-compares results for the two possible FAME-C cloud top retrieval schemes (AATSR-based or MERIS-based). These results can in turn be compared to corresponding CC4CL and CLARA-A2 results in the previous Tables Table 3-2 and Table 3-8.

Table 3-10: *Summary of FAME-C cloud top height validation results sub-divided into results from the infrared-only retrieval (AATSR, middle column) and the O_2 A-band retrieval (MERIS, right column).*

Cloud category	FAME-C AATSR Bias (m)	FAME-C AATSR Bias-corrected RMSE (m)	FAME-C MERIS Bias (m)	FAME-C MERIS Bias-corrected RMSE (m)
All clouds	-1143	2267	-1465	2317
Low-level	747	1261	-65	1048
Medium-level	-1323	1290	-1014	2007
High-level	-3408	1774	-2581	2505

The results based on AATSR infrared data and the MERIS results for Medium- and High-level clouds look very similar to previously achieved results for CC4CL AATSR + AVHRR) and CLARA-A2 (AVHRR). However, for the MERIS-based retrieval results are clearly better for the category of Low-level clouds. Instead of a large overestimation we get here only a small underestimation of cloud top heights.

3.2 Validation of cloud phase (CPH) against CALIOP

Based on Cloud_cci Level-3U data and CALIOP 5km Level2 data (see Section A.2 for details on CALIPSO-CALIOP data) spatiotemporal collocation of Cloud_cci and CALIOP cloud phase retrievals were collected in the time frame July 2006 to December 2014. For AVHRR-AM, AVHRR-PM a time window of ±3 minutes was used, resulting in sample size of up to 15 million CALIOP footprints for which a Cloud_cci pixel was found. Due to orbital characteristics of the AVHRR-AM satellites, i.e. NOAA-17 and Metop, and of Terra and ENVISAT, the collocation time window had to be increased to ±15 minutes to infer a decent sample size: (approx. 23.000 for AATSR/MERIS-ENVISAT, 18.000 for MODIS-Terra and 43.000 for AVHRR on NOAA-17 and Metop).

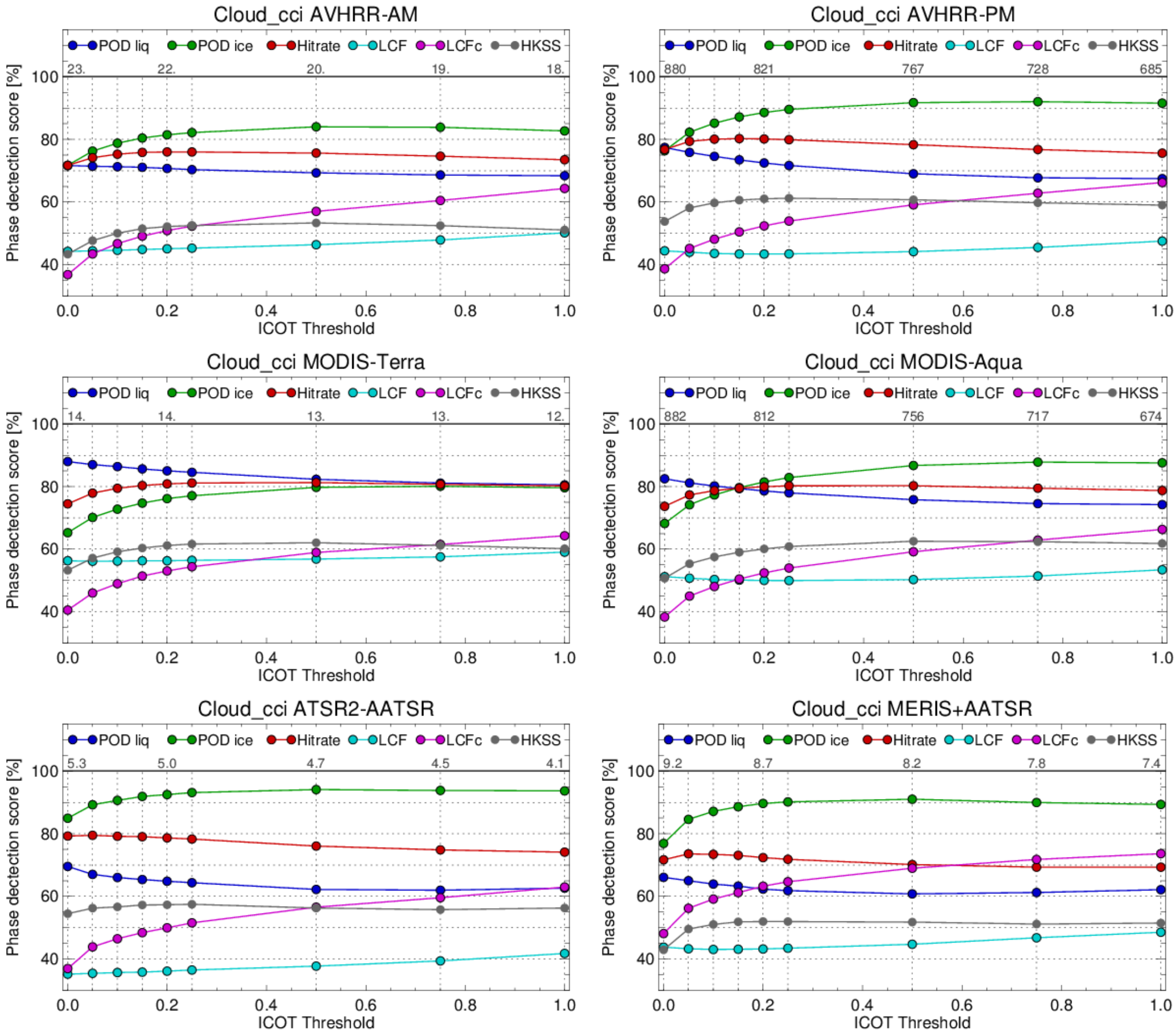


Figure 3-3 *Cloud phase scores for Cloud_cci dataset using CALIOP as reference. Left column: Probability of detecting liquid/ice cloud correctly ($POD_{liq/ice}$), general hitrate for detecting the correct phase, Hansen-Kuipers Skill Score (HKSS) and liquid cloud fraction for Cloud_cci (LCF) and CALIOP (LCFc) based on all collocation found. Grey numbers indicate sample size in thousands. Definition of used scores can be found in Section 2.5.1.*

These collocations are not only smaller in size compared to AVHRR-PM and MODIS-Aqua, they are also mainly located in the high latitude, which biases the results towards the cloud phase retrieval performance in these regions. Note that usually approximately only 50% of all collocations contain clouds, which again reduces the number of pixels that can be used for evaluation of the cloud phase. Figure 3-3 shows various kinds of phase detection scores for all Cloud_cci datasets as a function of iCOT threshold. The iCOT value stands for the level-to-cloud-top cloud optical thickness and represent the level at which the CALIOP reference phase is taken from. For example iCOT=0 means that the CALIOP phase is taken from the uppermost cloud level in the CALIOP cloud profiles; iCOT=0.5 means that the CALIOP phase is taken from lower levels in the CALIOP cloud profiles, i.e. the level at which the level-to-cloud-top optical thickness exceeded 0.5.

Exemplarily discussing the Hitrate score, increasing the iCOT threshold leads to increased Hitrate scores thus to increased agreement between CALIOP and all Cloud_cci datasets. Depending on the dataset, the maximum Hitrate score is found for iCOT values between 0.15 and 0.5. Very similar behaviour is found for the Hansen-Kuipers score and reflects the little sensitivity of the passive imaging sensors, and the algorithms applied to them, to very thin cloud (top) layers.

Table 3-11 presents the most important numbers of the evaluations shown in Figure 3-3, i.e. for iCOT values of 0.0 and 0.15. Using these values one can further analyse that Cloud_cci datasets usually have a liquid bias when comparing their cloud phase to the phase detected at CALIOPs uppermost cloud level, which agrees with the argumentation above that the uppermost thin cloud layers are not detected and ice clouds layers are usually on top of liquid cloud layers if both phased exist in a column. The liquid bias decreases with increasing iCOT values and is partly already negative (= ice bias) for iCOT values of 0.15. An exception from this argumentation are the ATSR2-AATSR and MERIS+AATSR datasets, which have already ice bias at iCOT=0. Another conclusion from Table 3-11 is that in more than 70% of all cases the cloud phase is correct for all datasets (the values vary a bit among the datasets). This is increased to almost 80% when going to iCOT=0.15.

Table 3-11 *Evaluation scores for Cloud_cci cloud phase retrievals using CALIOP phase as reference. The scores are calculated for the CALIOP phase at the uppermost cloud layer (iCOT=0.0) and for CALIOP phase taken at a lower level where the level-to-cloud-top cloud optical depth exceeded a threshold of 0.15 (iCOT=0.15). Definition of used scores can be found in Section 2.5.1.*

	Score	Cloud_cci dataset					
		AVHRR-AM	AVHRR-PM	MODIS-Terra	MODIS-Aqua	ATSR2-AATSR	MERIS+AATSR
iCOT=0.0	POD-liq	71.66	77.43	87.99	82.50	69.56	66.08
	POD-ice	71.72	76.37	65.28	68.22	84.94	76.88
	Hitrate	71.70	76.78	74.49	73.69	79.25	71.68
	HKSkill	0.43	0.54	0.53	0.51	0.54	0.43
	Bias	7.40	5.75	15.77	12.88	-1.76	-4.35
	Number	23.391	8.804.820	14.679	8.825.035	5.393	9.294
iCOT=0.15	POD-liq	71.10	73.46	85.64	79.36	65.38	63.28
	POD-ice	80.43	87.15	74.73	79.67	91.89	88.61
	Hitrate	75.85	80.24	80.34	79.51	79.05	73.11
	HKSkill	0.52	0.61	0.60	0.59	0.57	0.52
	Bias	-4.24	-7.03	4.89	-0.35	-12.59	-18.06
	Number	22.451	8.336.678	14.331	8.251.097	5.116	8.850

In Figure 3-4 and Figure 3-5 the phase detection scores are broken down into the cloud types that come along with CALIOP and Cloud_cci data, respectively. Drawing general conclusion from Figure 3-4, which presents the scores for the CALIOP cloud types, mid-level cloud are the most difficult ones in terms of correct phase detection. Hitrate scores are partly below 50%, depending on the dataset, however also conditions with transparent cirrus are characterized by low Hitrate scores at least for AVHRR-PM, AVHRR-AM, MODIS-Aqua and MODIS-Terra. Since this cloud type is very frequent (30-45% depending on collocation locations) it significantly affected the overall scores. The scores for ATSR2-AATSR and MERIS+AATSR are significantly higher for cirrus conditions compared to the other datasets, which is not entirely understood yet, but could be related to the regions in which CALIPSO-ENVISAT collocation can only be found.

Figure 3-5 shows the phase detection scores as function of Cloud_cci cloud type. As the Cloud_cci cloud type is used to derive the cloud phase (each cloud type corresponds to a certain phase), this analysis can be used to identify certain parts of the cloud typing working better or worse. First thing to notice is the high false alarm rate for the cloud type switched_to_liquid in the AVHRR-AM, AVHRR-PM and MODIS-Aqua datasets, which is for the latter two actually higher than the phase Hitrate. This cloud type is only chosen in Cloud_cci if the phase was detected ice but the cloud top temperature exceeded 0°C. Obviously this approach is suboptimal as it indeed increases the consistency between phase and CTT but this at the cost of correct phase determination. On the other hand, the relative potion of this cloud type is below 2%, thus not causing significant effects on the overall phase detection capabilities in Cloud_cci datasets. It is still unclear why no switched_to_water/ice cloud types are found for MODIS-Terra. This needs further investigation.

Apart from this, there is no other cloud type giving high false alarm rate for the corresponding phase, although the score vary among all datasets.

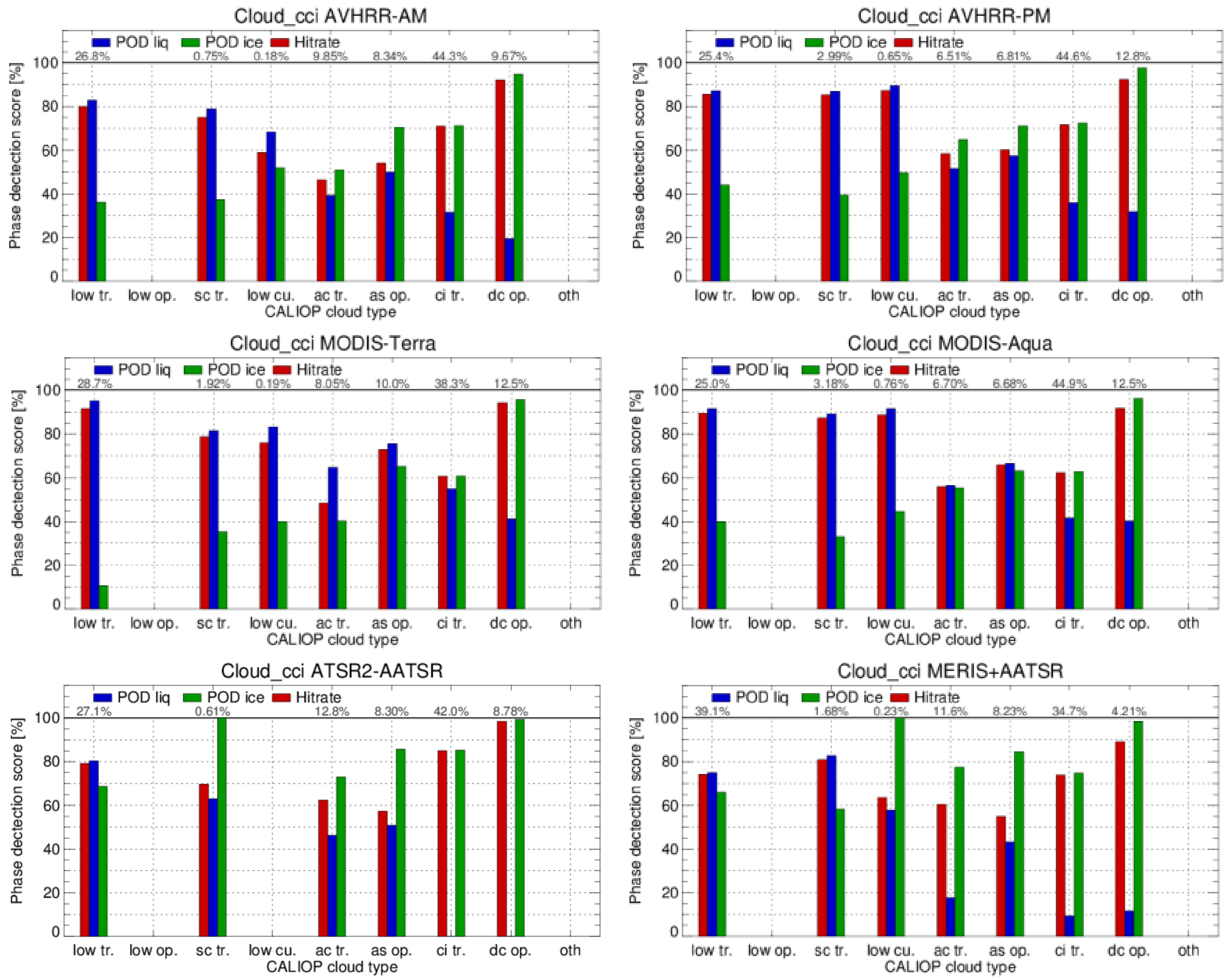


Figure 3-4: *Probability of correctly detecting liquid clouds (PODliq), ice clouds (PODice) and Hitrate scores (using CALIOP as reference) are shown broken down into CALIOP cloud types with low tr.: low overcast transparent, sc tr.: transition stratocumulus, ac.tr.: altocumulus transparent, as.op.: altostratus opaque, ci.tr.: cirrus transparent, dc.op.: deep convective opaque. Grey numbers indicate relative portion of cloud type wrt. to all clouds in collocations. Definition of used scores can be found in Section 2.5.1.*

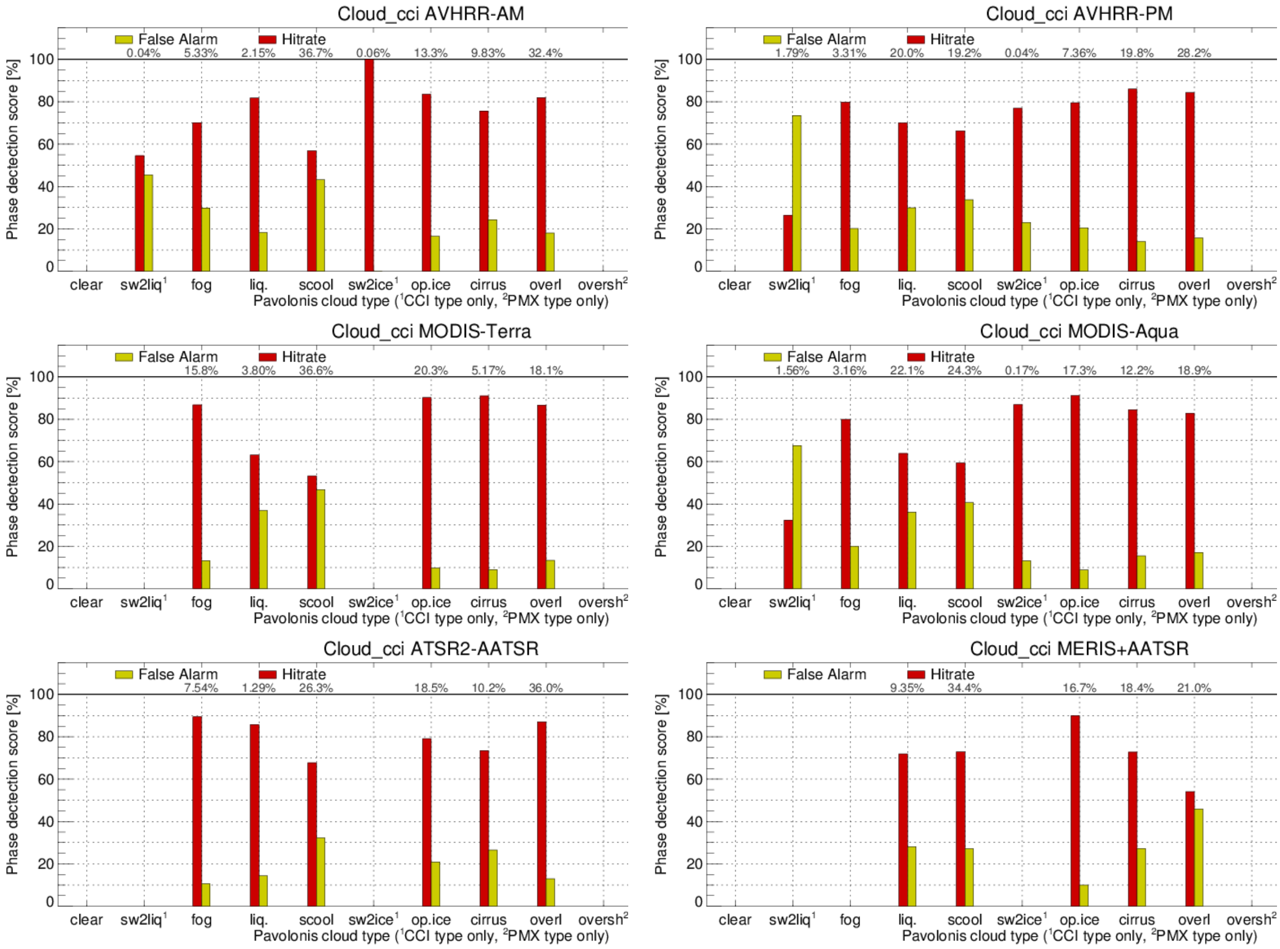


Figure 3-5 *False alarm and Hitrate scores for cloud phase detection (using CALIOP as reference) are shown broken down into cloud types analysed in CC4CL and FAME-C with sw2liq: switched_to_liquid, fog: fog, liq.: warm water clouds, scool: super-cooled water clouds, sw2ice: switched_to_ice, op.ice: opaque ice clouds, cirrus: cirrus, overl.: overlapping ice cloud. Grey numbers indicate relative portion of cloud type wrt. to all clouds in collocations. Definition of used scores can be found in Section 2.5.1.*

3.3 Validation of cloud fraction (CFC) against SYNOP

In this section we investigate the monthly mean cloud fraction of all datasets wrt. their agreement to ground based SYNOP observation (see Section A.1 for details on SYNOP data). For each SYNOP station the climatological monthly mean of the cloud fraction has been calculated from all available cloudiness observations (given in octa) during the day. For each SYNOP site, the Cloud_cci CFC of nearest Level-3C grid cell has been selected. For each dataset (Figure 3-6 to Figure 3-11) we present time series of CFC for Cloud_cci and SYNOP averaged over all SYNOP site locations as well as a time series of the corresponding monthly Bias and monthly standard deviation. In addition, we show the mean bias for each time period for each station as map, separated into the four seasons.

It is important to note that the cloud mask validation against CALIOP, given in Section 3.1, are based on spatiotemporally collocated observations and are thus not affected by any sampling errors. For the validation against SYNOP shown in this section Level-3 data was utilized without accounting for the different temporal sampling of satellite sensors and SYNOP. Thus, the results shown in this section give the total error of the Cloud_cci monthly cloud fraction, including the sampling error.

Generally speaking, all datasets show a good agreement to SYNOP cloudiness. However, a few features are found that need to be mentioned:

(1) AVHRR-PM, MODIS-Aqua and MODIS-Terra have a clear seasonal cycle in their agreement to SYNOP with a nearly zero bias in Northern Hemispheric summer while showing slight overestimations of 5 to 10% during winter. The latter seems to be caused by mid-latitude land regions in Europe and Asia (see winter maps in Figure 3-6, Figure 3-8 and Figure 3-9).

(2) AVHRR-AM shows strong overestimation of cloudiness for the period before 1999 with highest amplitude in Northern Hemispheric winter. This is apparently a problem with NOAA-12 which is not fully understood yet. NOAA-12 flies in a twilight orbit, which are characterized by very difficult illumination conditions. However, so is NOAA-15, for which the same feature cannot be found.

(3) AVHRR-AM (1999 and onwards), ATSR2-AATSR and MERIS+AATSR have very small biases throughout their respective time periods which show only very small season variations.

(4) Two distinct regions are visible in nearly all comparisons: a) the USA for which we have an underestimation of CFC against SYNOP nearly all the time and in all datasets, b) the Sahel zone at the southern end of the SAHARA desert, for which an underestimation of cloudiness is visible, again, nearly all the time for all datasets. The reasons for these two features are not known yet.

(5) Standard deviation against SYNOP seems relative small for all datasets including a small seasonal cycle with higher values again for Northern Hemispheric winter time.

(6) All datasets (except AVHRR-AM before 1999) exhibit a very stable behaviour in the comparisons shown, and this even though the number of SYNOP sites considered is not constant throughout the decades.

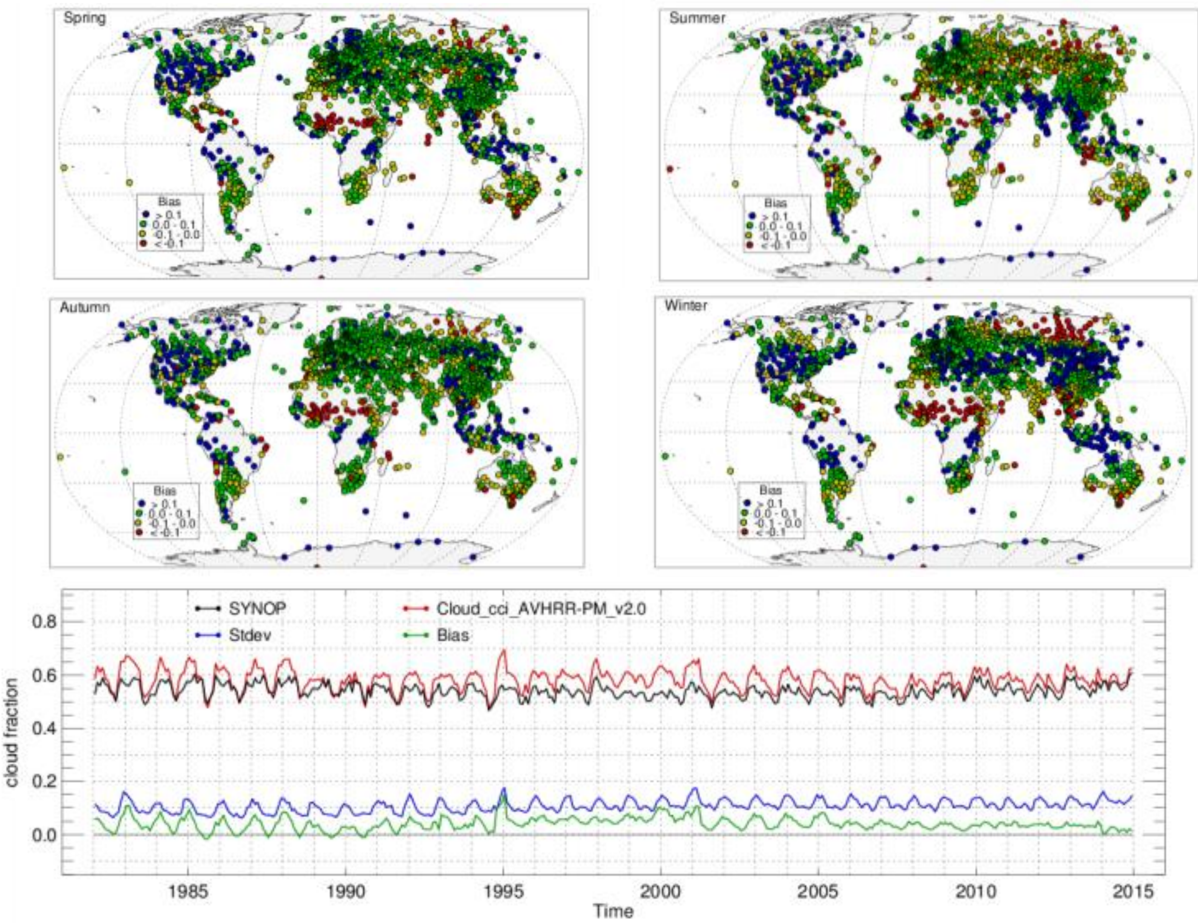


Figure 3-6 Top panels: Seasonal map of CFC Bias (Cloud_cci AVHRR-PM minus SYNOP) at each SYNOP site, averaged over the time period of the dataset (1982-2014) for each SYNOP comparisons for AVHRR-PM. Bottom panel: Time series of Cloud_cci AVHRR-PM and SYNOP CFC averaged over all SYNOP sites, as well as corresponding monthly Bias and standard deviation.

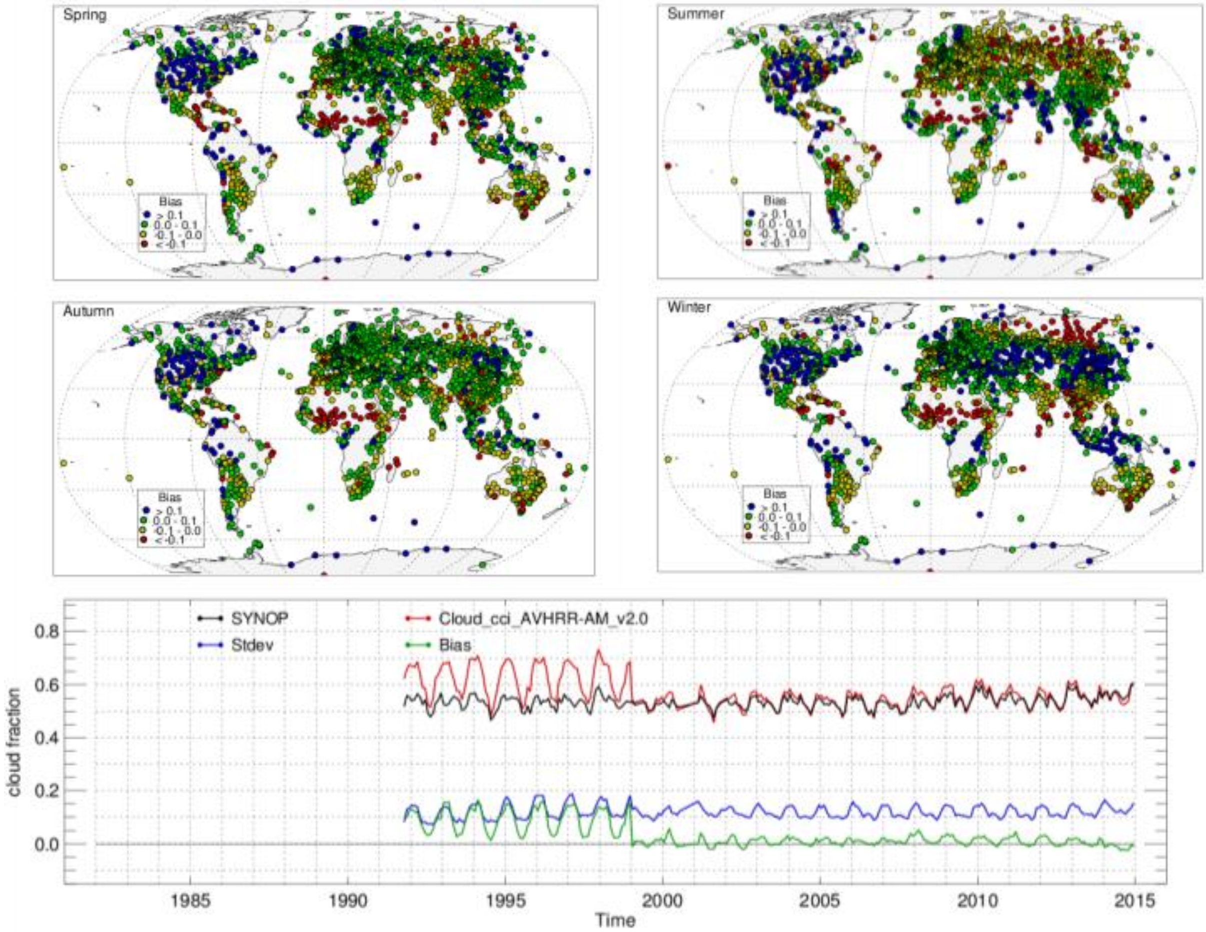


Figure 3-7 As Figure 3-6 but for Cloud_cci AVHRR-AM (1991-2014).

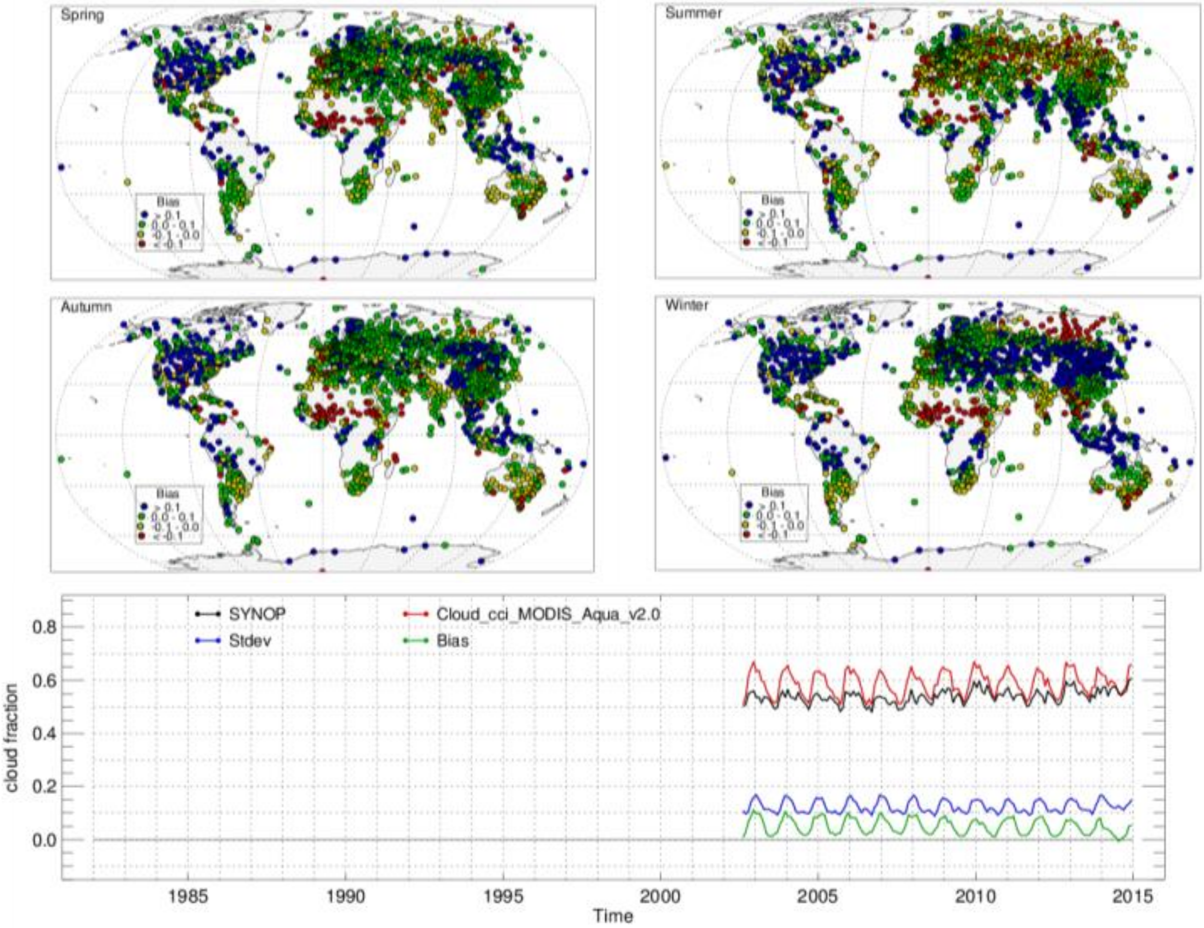


Figure 3-8 As Figure 3-6 but for Cloud_cci MODIS-Aqua (2002-2014).

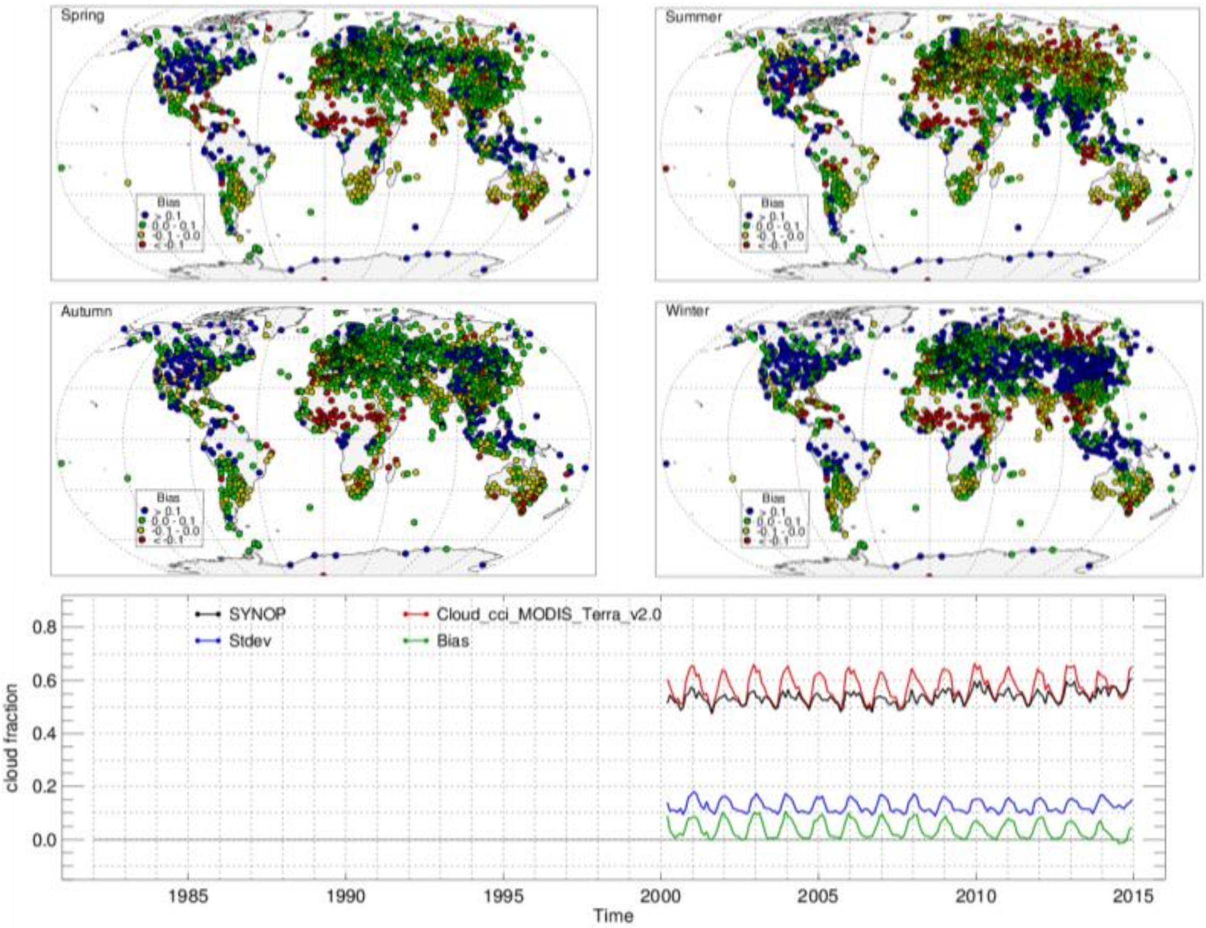


Figure 3-9 As Figure 3-6 but for Cloud_cci MODIS-Terra (2000-2014).

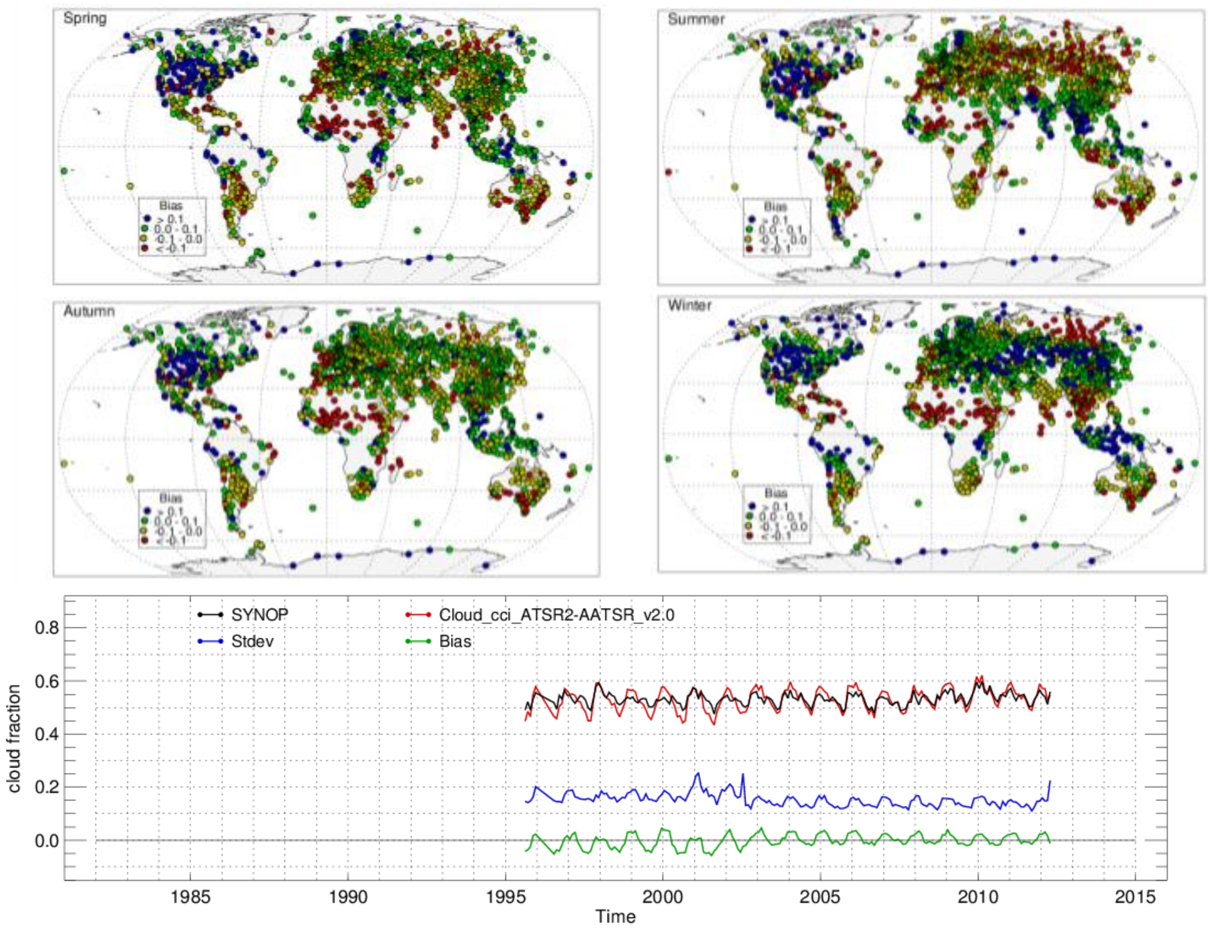


Figure 3-10 As Figure 3-6 but for Cloud_cci ATSR2-AATSR (1995-2012).

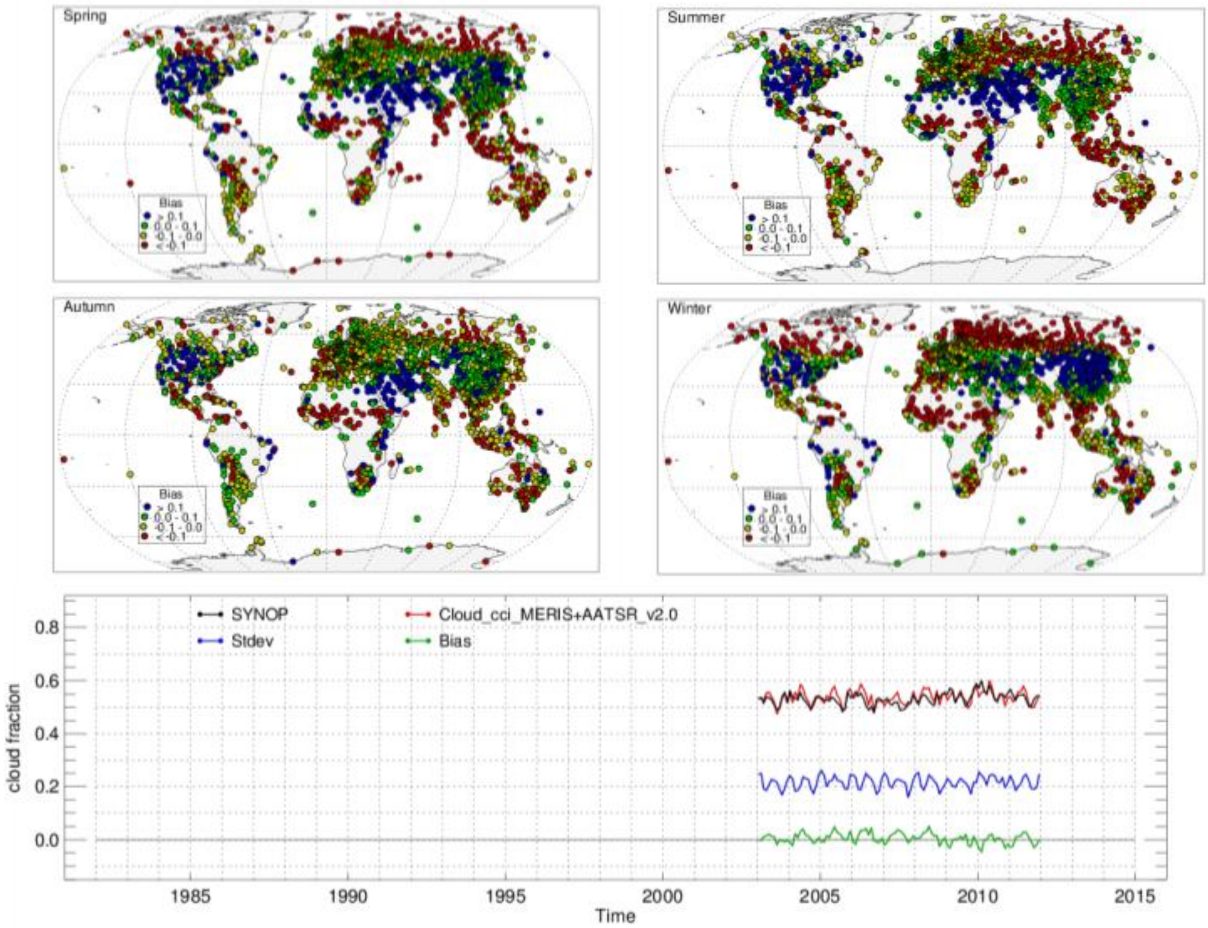


Figure 3-11 As Figure 3-6 but for Cloud_cci MERIS+AATSR (2003-2011).

3.4 Validation of liquid water path (LWP) against passive microwave retrievals

In this section the liquid water path of Cloud_cci datasets is evaluated against the UWISC microwave-based LWP dataset over ocean. The focus is given to region in which liquid clouds are dominant, i.e. three stratocumulus regions: the oceanic area west of Africa at 10°-20°S, 0°-10°E (SAF), the area west of South America at 16°-26°S, 76°-86°W (SAM), and the area west of California at 20°-30°N, 120°-130°W (NAM). See also Figure 3-12 for their locations. More details on the passive Microwave LWP reference dataset as well as on the validation strategy are given in Annex A.5.

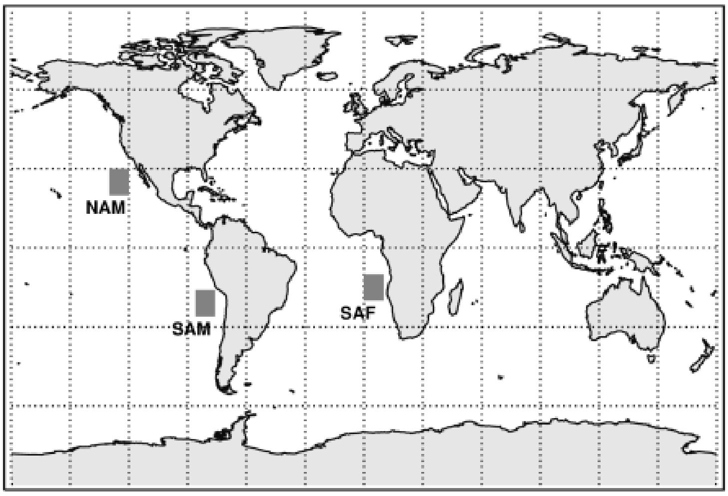


Figure 3-12 *The locations of the three validation areas.*

<u>Results</u>

Figure 3-13 and Figure 3-14 show the comparison of Cloud_cci LWP time series for the three selected regions for morning and afternoon satellites, respectively. The time series show overall reasonable agreement with the UWisc data in terms of seasonal variability, but there are considerable differences in average LWP among the different Cloud_cci instruments and the UWisc dataset. The retrievals from NOAA-12 and NOAA-15 are very problematic, with extreme deviations from the UWisc dataset. The reason is that these are twilight satellites which have a very difficult viewing geometry with a low elevation of the sun above the horizon. In general, the afternoon satellites appear to agree somewhat better with UWisc than the morning satellites. This may be partly related to the fact that the morning satellites NOAA-17 and Metop-A have the 1.6 micron channel active on AVHRR rather than the 3.7 micron channel. Cloud optical thickness and effective radius retrievals using the 1.6 micron channel are more sensitive to aerosol above cloud, which particularly occurs in the SAF region between July and October. The retrievals from AATSR yield rather high LWP, in particular much higher than from its predecessor ATSR2. From the afternoon satellites NOAA-14 LWP tends to have the largest deviation (underestimation) from UWisc.

Figure 3-15 shows some statistical measures of the Cloud_cci – UWisc inter-comparisons, separated by satellite and region. The bias in LWP is typically around +/- 10 g/m^2, although larger for some instruments and much larger for NOAA-12 and NOAA-15. The bias-corrected RMSE is overall between 5 and 10 g/m^2, again with exception of the twilight satellites, which illustrates that the seasonal variability is quite well captured. Linear correlation coefficients of monthly mean Cloud_cci and UWisc LWP range from 0.7 to 0.95.

In Figure 3-16, the average of the monthly mean Cloud_cci estimated LWP uncertainty is plotted, again per satellite and per region. Comparing this with the middle panel of Figure 3-15, it appears that the variations in estimated uncertainty do resemble the observed mismatch (bc-rmse) with UWisc to some extent. Most clearly, the very large deviations noticed for the twilight satellites are consistent with very large estimated uncertainties. Figure 3-17 illustrates this resemblance with a scatter plot. On average, the estimated uncertainty is about a factor 2 larger than the observed bc-rmse with UWisc, indicating that the estimates may be slightly conservative for this type of clouds.

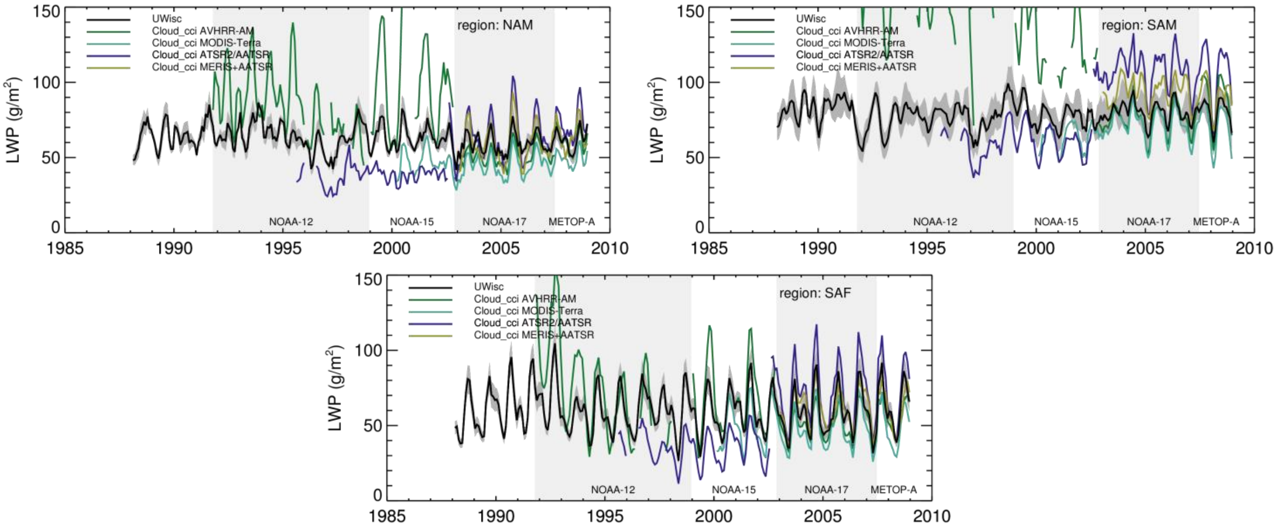


Figure 3-13 *Comparison of Cloud_cci LWP time series (colours) with UWisc (black) for the morning satellites. The grey shading represents the UWisc dataset sampled one hour earlier/later than 9:30 local time. The lines show 3-month running means.*

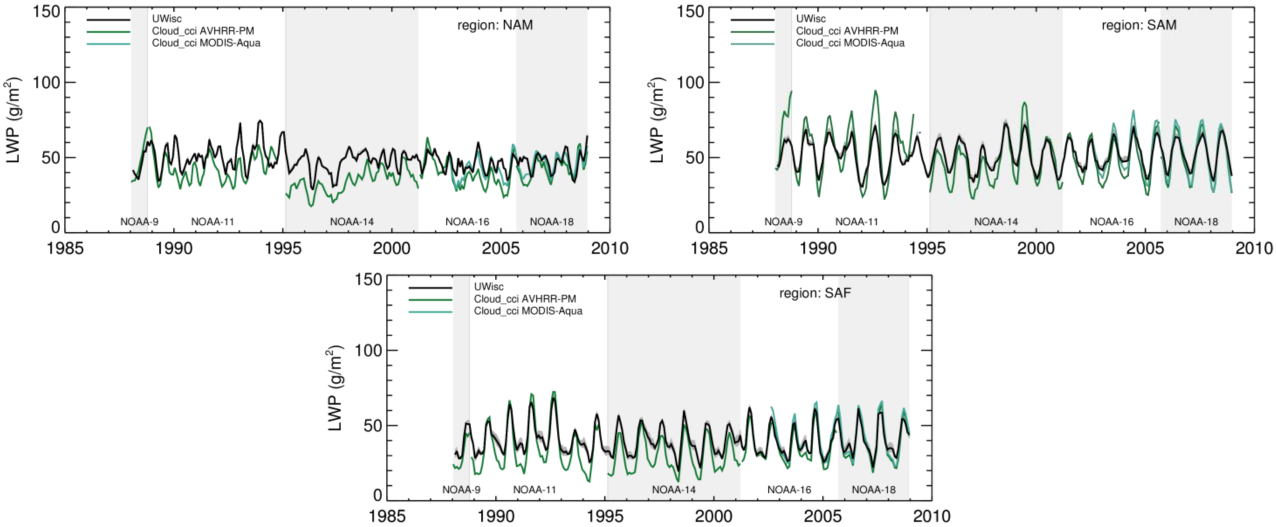


Figure 3-14 *Comparison of Cloud_cci LWP time series (colours) with UWisc (black) for the afternoon satellites. The grey shading represents the UWisc dataset sampled one hour earlier/later than 14:00 local time. The lines show 3-month running means.*

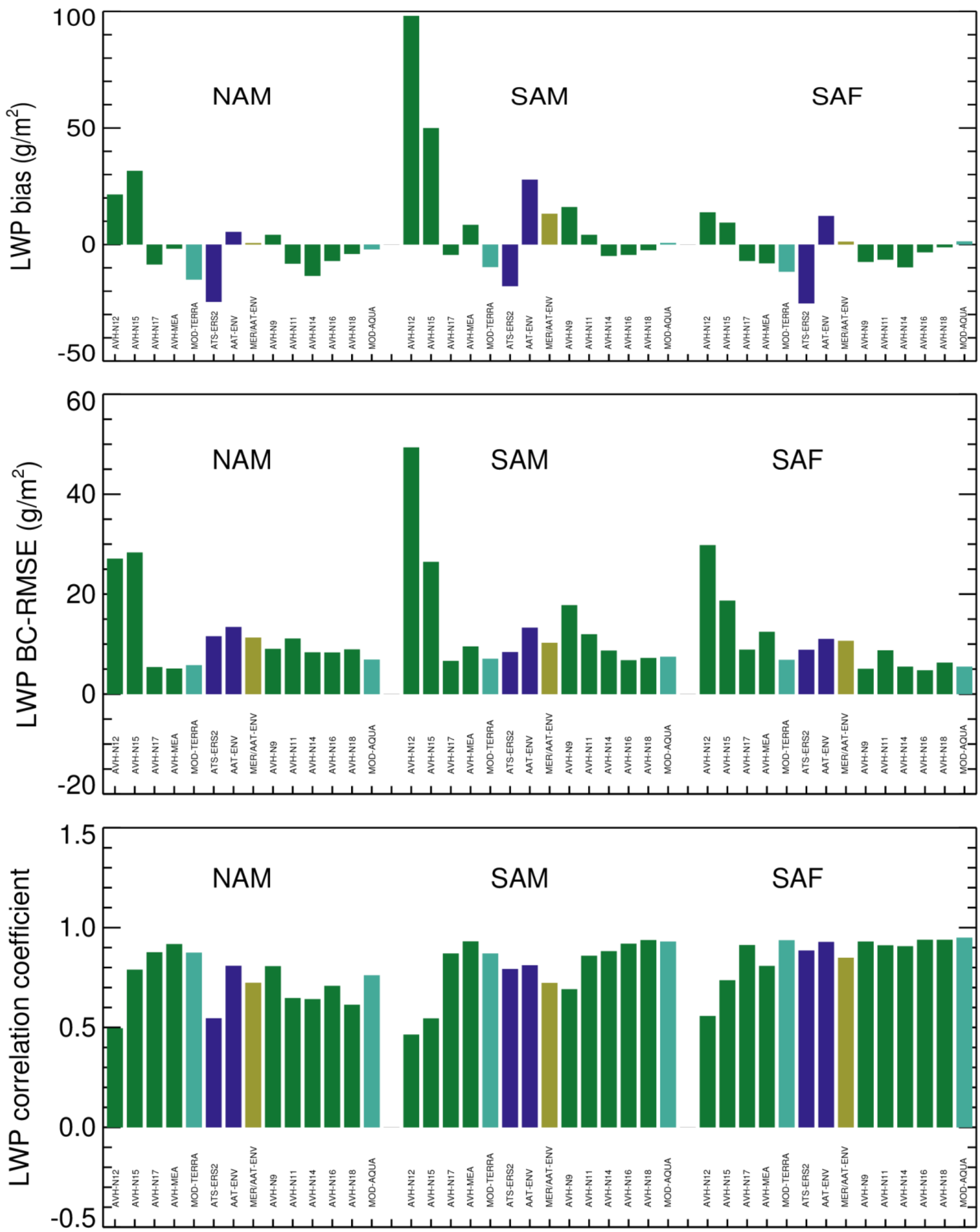


Figure 3-15 *Statistics of Cloud_cci monthly mean LWP with respect to UWisc: bias (top), bc-RMSE (middle), and linear correlation coefficient (bottom). Results are shown separately for the 14 satellite instruments and 3 regions analysed.*

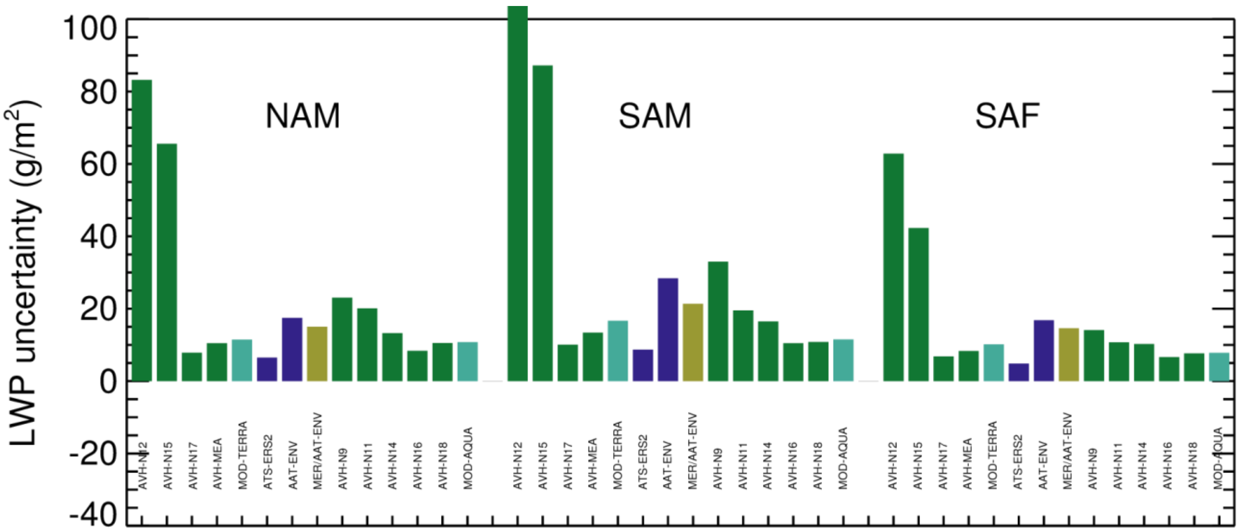


Figure 3-16 *Average of Cloud_cci monthly mean LWP uncertainty for the 14 satellite instruments and 3 regions analysed.*

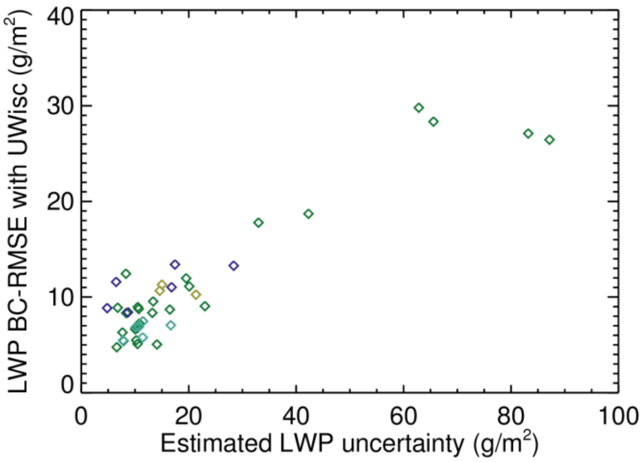


Figure 3-17 *Scatter plot of Cloud_cci LWP bc-RMSE with UWisc versus Cloud_cci estimated LWP uncertainty. Each symbol represents a satellite-area combination, and the same colours have been used as in Figures 3-11 and 3-12.*

4. Intercomparison of Cloud_cci products with other satellite-based datasets

In this section the Cloud_cci datasets are compared to other, well-established, satellite-based datasets. The bases for all comparisons are monthly mean values (Level-3C in ESA CCI notation).

The comparisons are separated into two classes for morning (Section 4.1) and afternoon (Section 4.2) datasets, according to the local (daytime) observation time of the individual, underlying satellites:

- Morning satellites: Terra, NOAA-12, NOAA-15, NOAA-17, Metop-A, ENVISAT, ERS2
- Afternoon satellites: Aqua, NOAA-7, NOAA-9, NOAA-11, NOAA-14, NOAA-16, NOAA-18, NOAA-19

The comparisons are further stratified into the individual cloud properties. For each of these subsections, Global maps of multi-annual means (of a common time period) and multi-annual (monthly) standard deviations are shown, together with zonal mean plots. In addition, time series plots of latitude-weighted global mean values (for a latitude band of 60S-60N) are given without restriction to the common time period. Using MODIS Collection 6 data as reference, Level 3C scores (Bias, bc-RMSE) are calculated for each Cloud_cci dataset for the common period and 60S-60N. The stability of the Cloud_cci dataset is investigated by calculating the linear trends and comparing these again to MODIS Collection 6 data.

4.1 Morning Satellites

In this subsection the morning satellite datasets are compared, these are:

- Cloud_cci AVHRR-AM (see Section 2.2 for details)
- Cloud_cci ATSR2-AATSR (see Section 2.2 for details)
- Cloud_cci MERIS+AATSR (see Section 2.2 for details)
- Cloud_cci MODIS-Terra (see Section 2.2 for details)
- CLARA-A2 (only subset of morning satellites used; see Section A.6 for details)
- MODIS Terra Collection 6 (see Section A.6 for details)

Figure 2-3 shows the local solar observation times of all satellites mentioned. For interpreting the following comparisons of cloud optical thickness, cloud effective radius, liquid water path and ice water path correctly it is necessary to revisit the channel settings of all sensors/datasets:

For Cloud_cci ATSR2-AATSR the 3.7µm channel is used as near-infrared channel. The same applies to the first half of Cloud_cci AVHRR-AM for which NOAA-12 and NOAA15 was used. For the second half of AVHRR-AM (NOAA-17 and Metop) the 1.6µm channel was used as near-infrared channel; the same applies to MERIS-AATSR. For CLARA-A2 the same channels as in AVHRR-AM were used. MODIS-Terra has three NIR channels and MODIS Collection 6 data usually has optical property products for each of them. In our comparisons, the maps always show the 2.1µm product, while in the time series and zonal mean plots 1.6µm and 3.7µm MODIS collection 6 products are visualized if both exist.

Table 4-1 *Near-infrared channels utilized for morning satellite datasets compared in this section. This information is needed for correctly interpreting the comparison results shown for cloud optical thickness, cloud effective radius, liquid water path and cloud water path. Time periods given are approximated.*

Channel used	Cloud_cci AVHRR-AM	Cloud_cci ATSR2-AATSR	Cloud_cci MERIS+AATSR	Cloud_cci MODIS-Terra	CLARA-A2	MODIS Terra Collection 6
1.6µm	2002-2014	-	2003-2011	-	2002-2014	(2000-2014)^{T}
2.1µm	-	-	-	-	-	(2000-2014)^{M}
3.7µm	1991-2002	1995-2012	-	2000-2014	1991-2002	2000-2014^{T}

4.1.1 Cloud Fraction

In this subsection the cloud fraction of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-1), zonal mean (Figure 4-2) and standard deviation (Figure 4-3), all for a common time period, and time series plots (Figure 4-4).

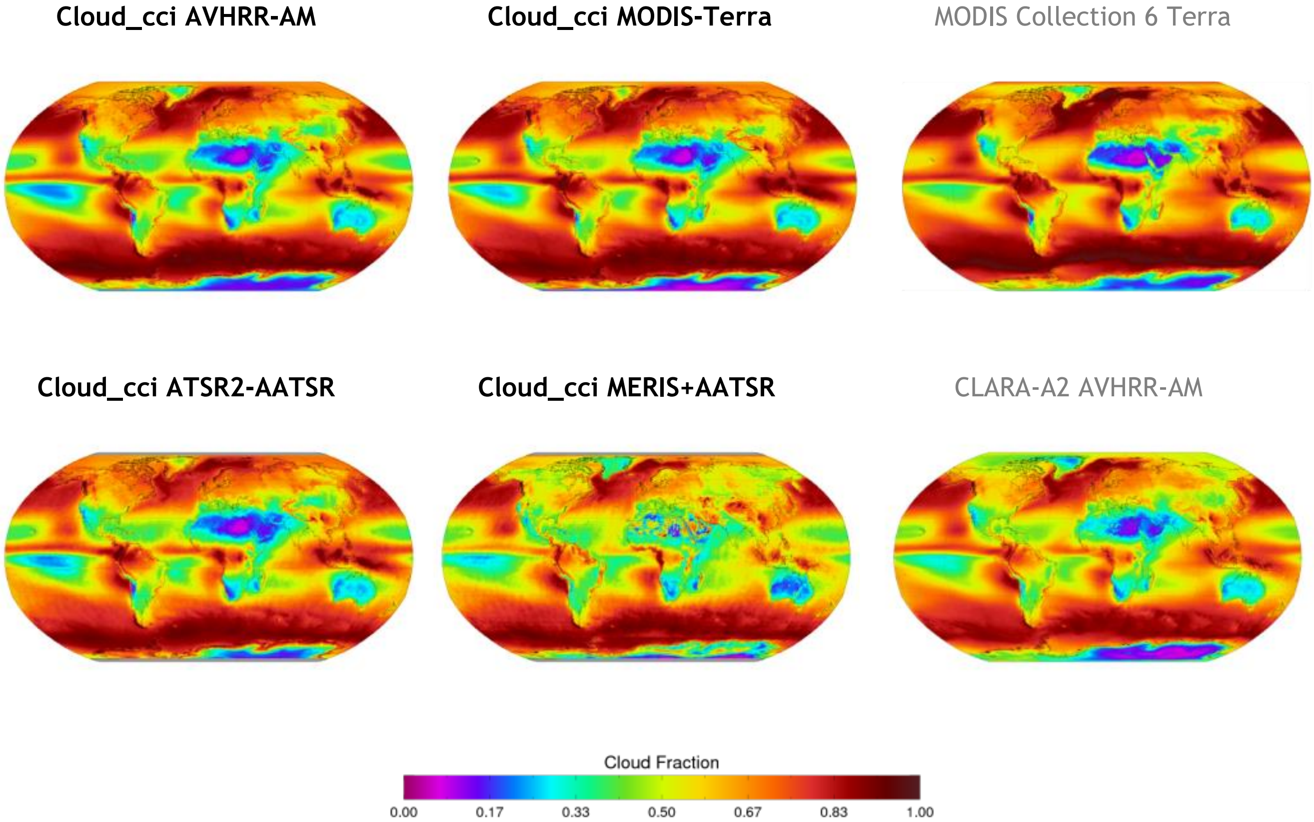
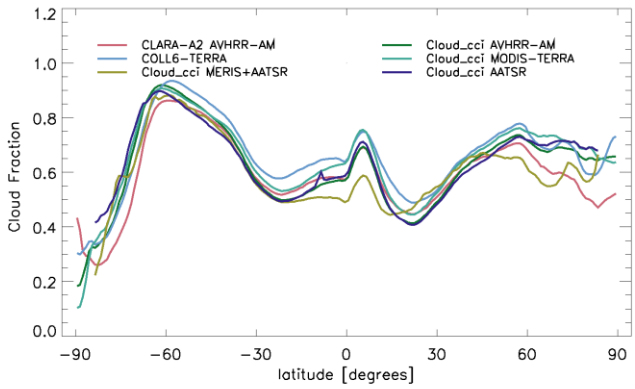


Figure 4-1 *Globally gridded means of cloud fraction for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*



Latitude weighted Means of Cloud Fraction [%]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	60.82	61.04	59.36	75.70	54.89	63.19	59.38	48.01	66.06
Cloud_cci AVHRR-AM	61.48	60.34	69.60	77.16	52.90	63.86	68.13	50.35	66.01
COLL6-TERRA	66.79	66.67	66.58	81.55	60.79	67.86	68.61	53.05	72.39
Cloud_cci MODIS-TERRA	64.72	63.69	70.83	77.98	57.37	66.67	71.91	53.47	69.29
Cloud_cci MERIS+AATSR	58.55	57.99	71.37	74.16	50.42	63.31	58.35	49.57	62.13
Cloud_cci AATSR	61.31	59.64	74.53	74.07	53.39	62.31	70.53	51.27	65.37

Figure 4-2 *Cloud Fraction from 2003 to 2011. Left: Zonal means. Right: Latitude weighted means for different regions. Highest and lowest values of each region are highlighted.*

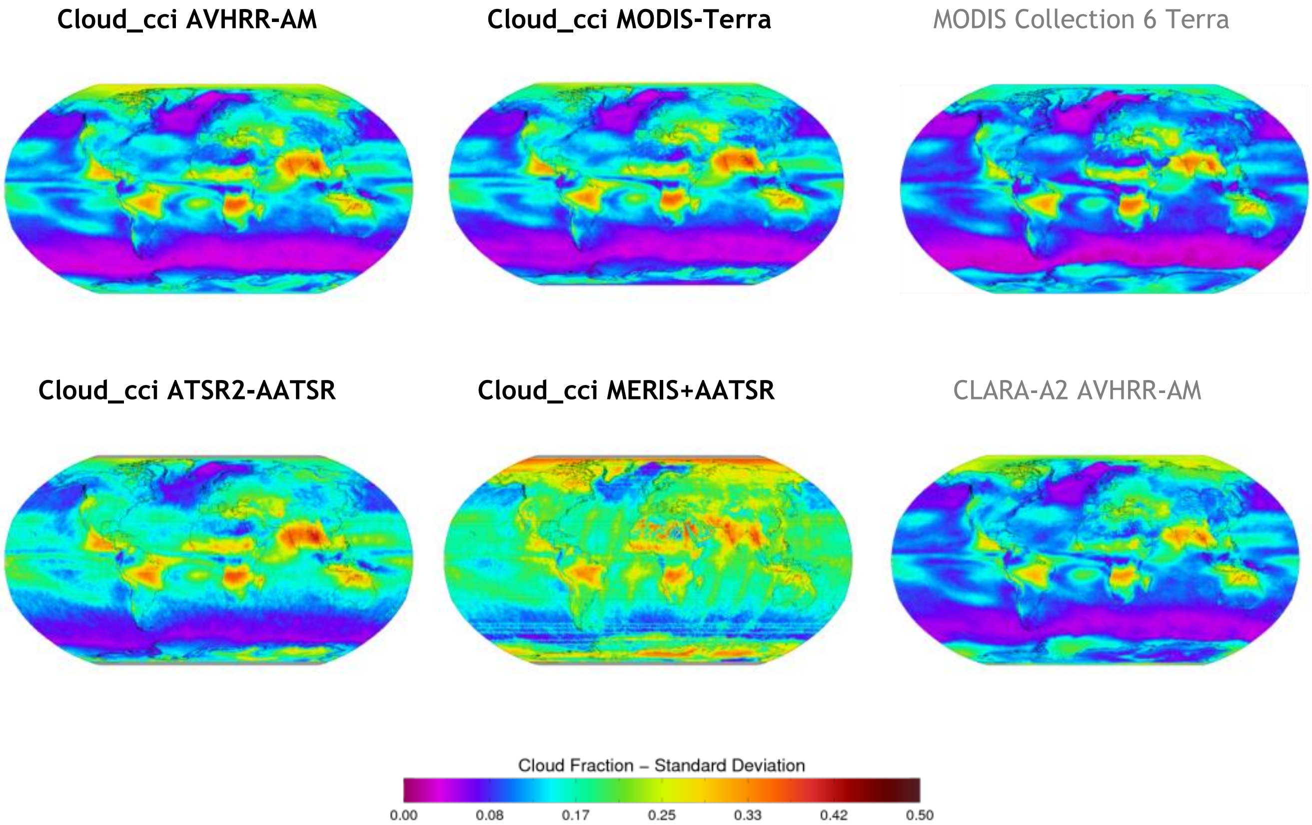


Figure 4-3 *Globally gridded means of cloud fraction standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*

General findings

- Multi-annual averages of all datasets generally compare well with each other; regionally higher disagreement is visible.
- Best agreement in Mid-Latitudes and Tropics with all datasets being with approx. 10% cloud fraction. Exception is MERIS+AATSR dataset which has much less cloud fraction in the Tropics.
- Largest spread among datasets is found in Polar Regions, mainly arising from polar night conditions.
- Temporal variability is highest for MERIS+AATSR and lowest for MODIS-Terra C6. Other datasets lie in between and seem relative consistent.
- Time series plots reveal jumps in the time series for (1) AVHRR-AM and CLARA-A2 at the transition from NOAA-12 to NOAA-15 around 1999, (2) for ATSR2-AATSR at the transition from ATSR2 (onboard ERS2) to AATSR (onboard ENVISAT) around 2003 and (3) for MERIS+AATSR in 2004 and in 2006.
- The time series of MODIS-Terra C6 also reveals a drifting to higher values from 2010 onwards. This is most likely due to the 8.6 µm channel of MODIS-TERRA is degrading resulting into a significant trend to higher cloud fraction in Coll6-Terra. Since the 8.6 µm channel is not included in the Cloud_cci Cloud Mask retrieval no trend is shown in Cloud_cci MODIS-TERRA

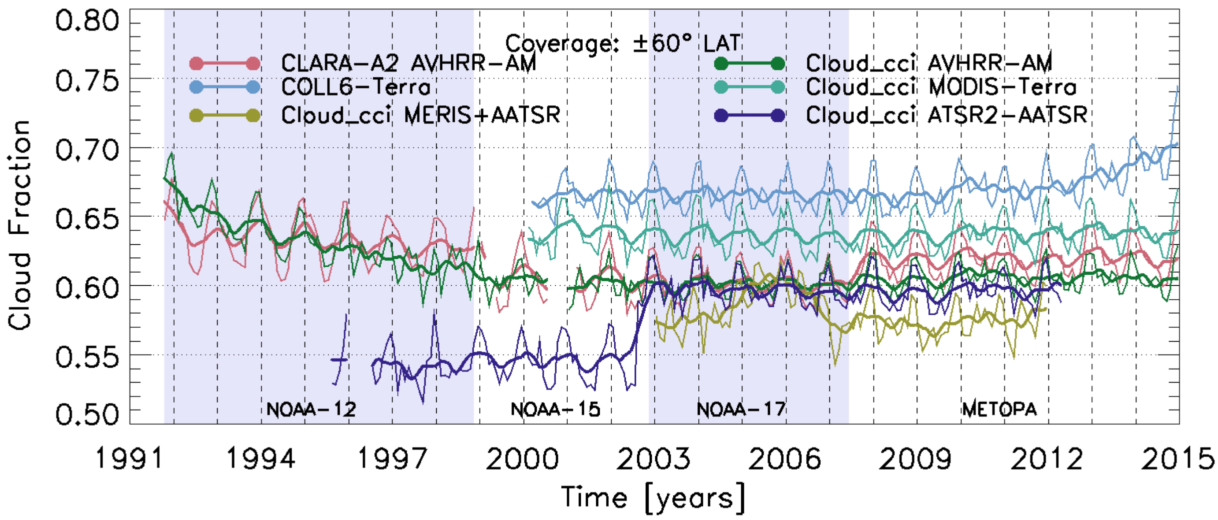


Figure 4-4 *Time series of cloud fraction for all morning satellite retrievals.*

Discussion

All Cloud_cci datasets are comparable to the reference datasets for most parts of the globe. An exception here is Cloud_cci MERIS+AATSR which has significantly lower CFC in the tropics. In general, MODIS-Terra C6 gives highest CFC for large parts of the globe, which might be due to a combination of relative fine spatial resolution and using more spectral bands than Cloud_cci. The zonal mean plots exhibits very high agreement among all datasets. The large spread in the polar regions might mainly due to polar night conditions in which no visible information is available and in which the surface temperature can be very low, both leading to only little or no contrast between signals from surfaces and clouds. This is a well-known problem for cloud detection based on passive imaging sensors.

In the time series of AVHRR-AM problems of the early AVHRRs (in particular NOAA-12) become evident. The NOAA-12 orbit is steadily drifting leading to a drift in local observation time, moving from late evening/twilight to night observations in case of the descending orbit node. As the cloud mask used in Cloud_cci AVHRR-AM has different branches for different illumination condition with significantly varying accuracies (twilight cloud detection is worst – large overestimation of cloudiness), the mean cloud fraction reduces with time as the observation time moves away from twilight condition.

Another point to mention is that an overestimation of cloudiness for CC4CL is found in condition of high aerosol loadings in the atmosphere. As discussed later for the afternoon satellites, the Pinatubo eruption in 1991 creates a significant positive anomaly in the cloud detection. This effect decreases as the aerosol loading is decreasing after some time after the eruption. Both, the reduction in twilight condition and the decreasing aerosol loadings after the Pinatubo outbreak result into strong negative trends in cloud fraction in the early years of Cloud_cci AVHRR-AM. CLARA-A2 shows similar behaviour with slightly decreased trend as CLARA-A2 cloud detection seems less impacted by the Pinatubo eruption. It should be noted that the gap in the Cloud_cci AVHRR-AM and CLARA-A2 datasets around the year 2000 is due to blacklisting of a large number of AVHRR-NOAA15 orbits due to AVHRR scan motor errors.

After the year 2000 no trends are seen in all Cloud_cci datasets and Cloud_cci AVHRR-AM agrees well with CLARA-A2. CLARA-A2 shows a positive jump of about 2 percentage points when switching from NOAA-17 to METOP-A. This is also seen in the Cloud_cci dataset but not that pronounced. Possible reasons for that can be a different equator crossing time of half an hour and slightly different viewing angle geometry.

MERIS+AATSR CFC increases cloud fraction in the year 2005 to 2007, which is not shown in the other datasets. This reason for this is still unclear, but the stability of this datasets is significantly impacted. The Cloud_cci MODIS-Terra cloud fraction is about 3-4 percentage points higher compared to all other Cloud_cci datasets. AATSR is a bit lower than AVHRR. The reason for the differences between MODIS, AVHRR and AATSR in Cloud_cci could be due to:

1) Passive sensors with higher viewing angles see more clouds: At higher viewing angles radiation has a longer path through the atmosphere, thus the impact of optically thin clouds on radiation is larger.
2) The major differences between all sensors are their spectral response functions of the AVHRR heritage channels. In particular the differences of the visible channels are very high between AVHRR and MODIS/AATSR. Within the CC4CL cloud detection all channels of MODIS and AATSR are corrected to mimic AVHRR by using slope and offset found by comparing collocated measurements. For AATSR, in contrast to MODIS, the offset of the reflectance channel correction is not applied to avoid a too strong increase for very low reflectance, which would lead to an increase in falsely detected clouds.

Table 4-2 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. All Cloud_cci datasets show negative CFC biases compared to MODIS C6 up to -10%. The bc-RMSE ranges from 6 to 18% with lower values for Cloud_cci AVHRR-AM and MODIS-Terra and higher values for Cloud_cci ATSR2-AATSR and MERIS+AATSR. The trend in 2003-2011 for is relative small for all Cloud_cci datasets and in relative good agreement to MODIS C6. Δ

Table 4-2: *Evaluation scores for Cloud_cci Level-3C cloud fraction (morning satellites) based on comparison to C6-Terra from 2003-2011, in case of MERIS+AATSR the C6-Terra cloud fraction is daytime only. The scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	AVHRR–AM	MODIS–Terra	ATSR2–AATSR	MERIS+AATSR
Bias [%]	–6.33	–2.98	–7.03	–8.15
bc–RMSD [%]	+7.11	+6.76	+10.03	+17.73
Cloud_cci trend [% / decade]	+0.79	+0.26	–0.11	–1.18
C6–Terra trend [% / decade]	+0.45	+0.45	+0.45	+0.60
Δ trend [% / decade]	+0.34	–0.19	–0.56	–1.78

4.1.2 Cloud Top Pressure

In this subsection the cloud top pressure of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-5), zonal mean (Figure 4-6) and standard deviation (Figure 4-7), all for a common time period, and time series plots (Figure 4-8).

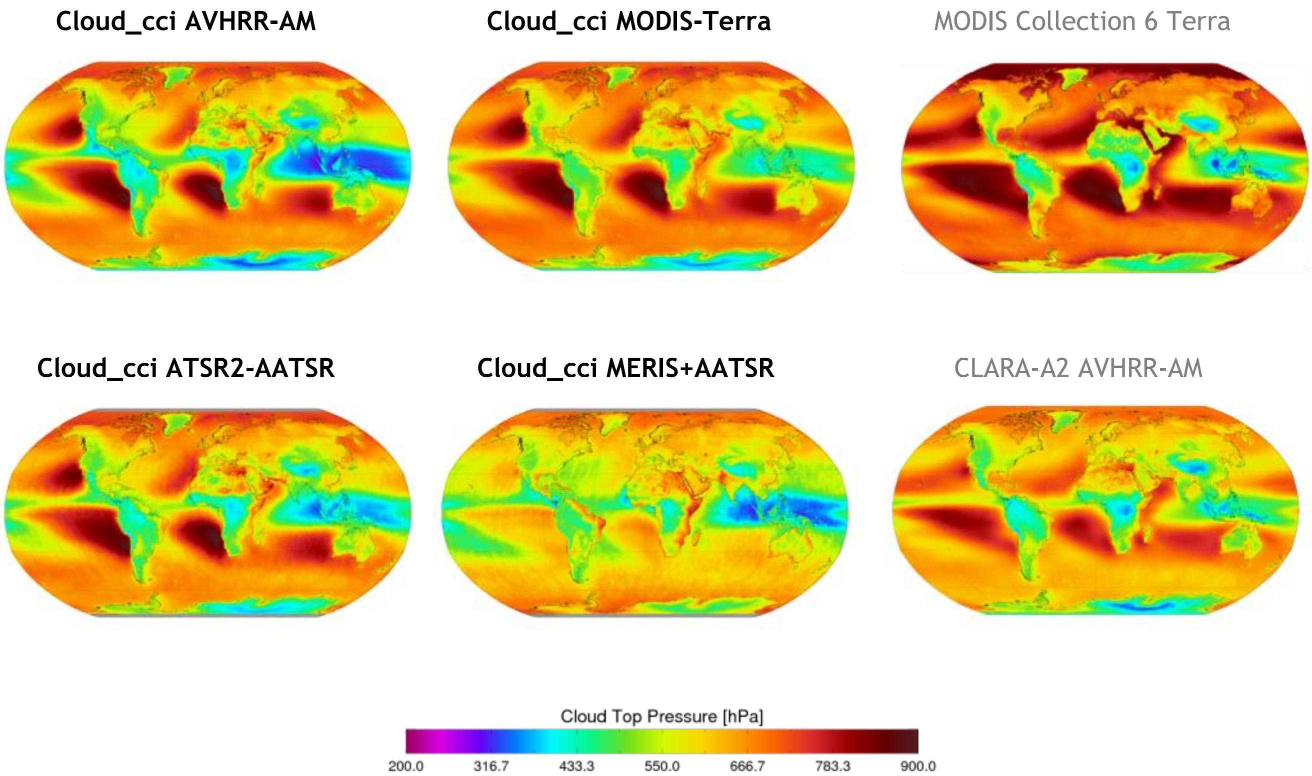
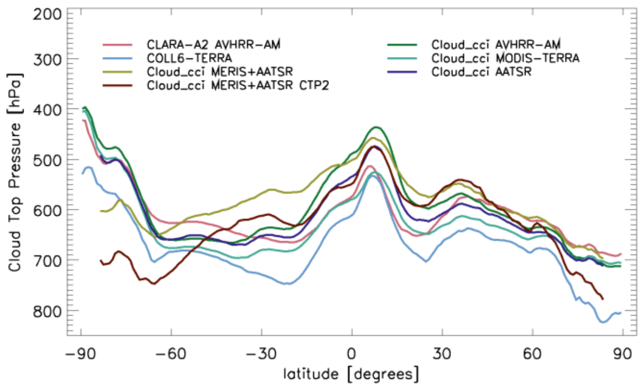


Figure 4-5 *Globally gridded means of cloud top pressure for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. The MERIS+AATSR map shown refers to a CTP based on AATSR measurements only.*



Latitude weighted Means of Cloud Top Pressure [hPa]

Dataset−name	Global	± 60° Lat.	Ant−arctica	Mid−Lat. South	Tropics	Mid−Lat. North	Arctic	Land	Sea
CLARA−A2 AVHRR−AM	614.43	615.09	571.55	638.32	614.71	593.03	647.78	542.53	643.67
Cloud_cci AVHRR−AM	588.57	583.82	574.67	658.47	552.29	595.33	663.69	527.26	613.47
COLL6−TERRA	669.50	667.09	644.38	701.44	658.79	655.49	726.22	580.53	705.79
Cloud_cci MODIS−TERRA	634.68	634.82	593.83	684.36	618.33	630.34	673.48	565.17	662.91
Cloud_cci MERIS+AATSR	558.53	551.96	631.44	597.06	529.87	575.57	644.12	567.40	555.14
Cloud_cci AATSR	608.53	605.07	593.17	660.05	583.35	609.26	670.14	543.41	634.66

Figure 4-6 *Cloud Top Pressure from 2003 to 2011. Left: Zonal means. Right: Latitude weighted means for different regions. Highest and lowest values of each region are highlighted. The suffix 'CTP2' for MERIS+AATSR refers to the MERIS-only CTP, while no suffix indicates the CTP based on AATSR only.*

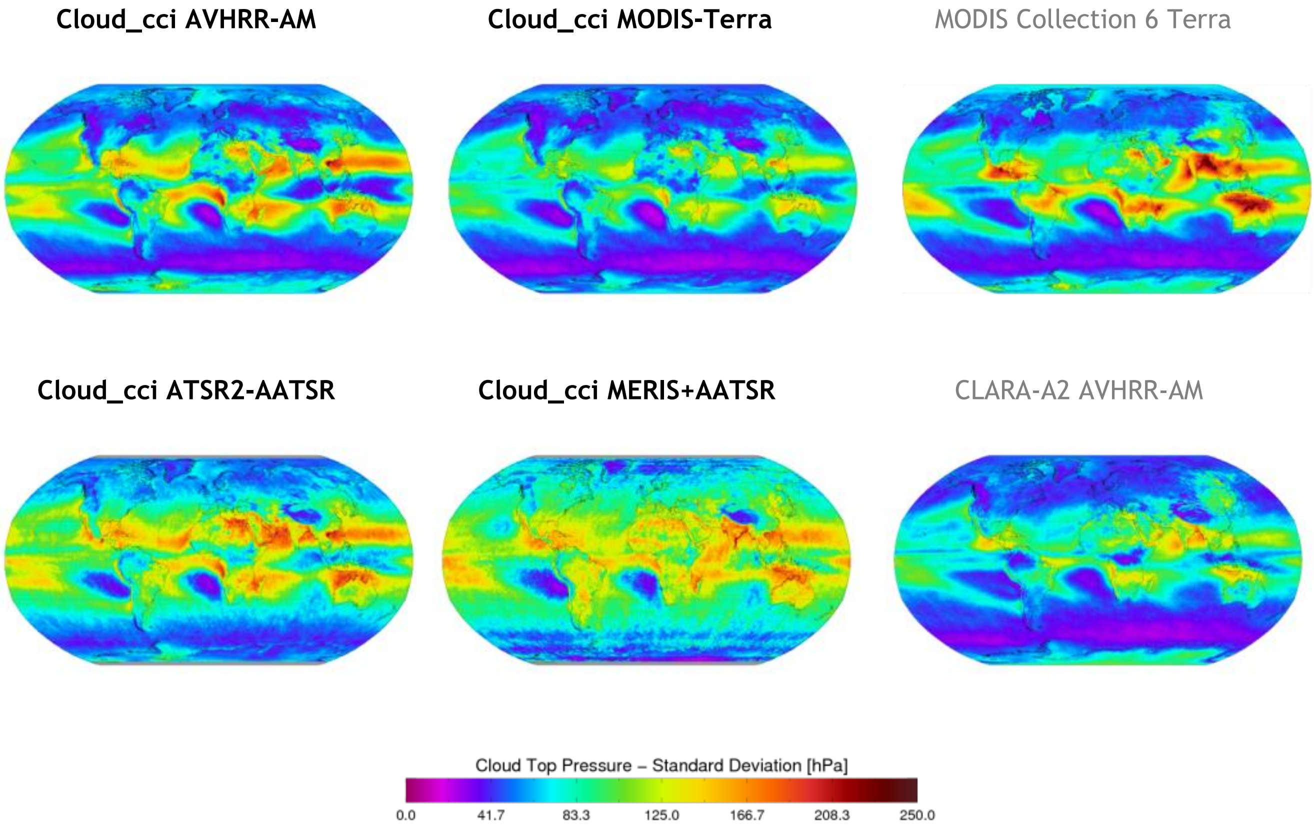


Figure 4-7 *Globally gridded means of cloud top pressure standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. The MERIS+AATSR map shown refers to a CTP based on AATSR measurements only.*

General findings

- All datasets reflect the main characteristic of global cloud top distribution. Lowest mean CTP are found in the Tropic, highest mean CTPs in the stratocumulus regions. Mean mid-level CTPs are found for the storm track region in the Mid-Latitudes of both hemispheres.
- Lowest mean CTP is found for Cloud_cci MERIS+AATSR, highest for MODIS-Terra C6 especially in the tropics over sea. Despite the agreement on mean distributions, the spread among the datasets is partly large in nearly all regions of the globe.
- Largest temporal variability is found for the sub-tropic regions, which is reflected in all datasets (to a smaller or larger extent) and most likely connected to the seasonal oscillation of the ITCZ.
- The time series plots confirm the large spread in mean CTP among all datasets, while most datasets for themselves remain rather stable throughout the time period covered. An exception here is a jump found for CLARA-A2 at the transition from NOAA-12 to NOAA-15. While this jump is not so pronounced in Cloud_cci AVHRR-AM, even though the same satellites were used as in CLARA-A2, there seems to be a small trend occurring for NOAA-12 and NOA15 periods. Cloud_cci ATSR2-AATSR shows some instability for ATSR2 (before 2003).

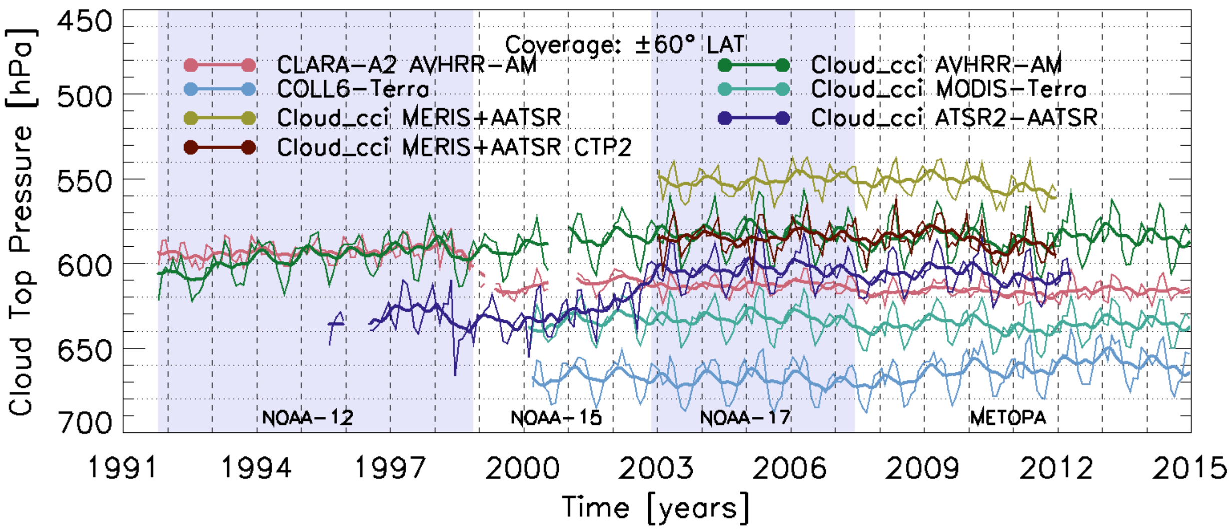


Figure 4-8 *Time series of cloud top pressure for all morning satellite retrievals. The suffix 'CTP2' for MERIS+AATSR refers to the MERIS-only CTP, while no suffix indicates the CTP based on AATSR only.*

Discussion

Cloud_cci L3C CTP data show characteristic global patterns as seen in the other datasets. However, partly large systematic deviations to the reference datasets but also among the Cloud_cci datasets are found. The most significant of the latter are the low values for Cloud_cci MERIS+AATSR and the high CTP values for Cloud_cci MODIS-Terra. The latter is more similar to MODIS-Terra C6, which might be explained by more clouds being detected for these two datasets (see higher cloud fraction for these two dataset in Section 4.1.1) assuming these additional clouds are low, small scale clouds. The low CTP for Cloud_cci MERIS+AATSR is not entirely understood yet, but is seems (at least partly) erroneous, since also the well-known stratocumulus fields over the parts of the oceanic subtropical regions are not well represented by a maximum in CTP compared to, for example, the mid-latitudes.

In Table 4-3 monthly 1-dimensional CTP histograms (also included in Cloud_cci and all other considered datasets) have been used to calculate the relative fraction of low, mid-level and high clouds on the total cloud amount using CTP thresholds of 440hPa and 680hPa according to the ISCCP definition. For most datasets we find a relative good agreement within 10 to 15%, but there are some outliers: MODIS-Terra C6 has 57% low clouds while all others have below 50%; Very few mid-level clouds in MODIS-Terra C6 of only 9.5%, in contrast many mid-level clouds in Cloud_cci MERIS+AATSR.

Table 4-3 *Relative fraction of low (CTP > 680hPa), mid-level (680hPa > CTP > 440hPa) and high (CTP < 440hPa) clouds calculated from 1-dimensional histograms for a common time period of 2003 to 2011 (excluding the Polar Regions higher/lower than +/-60°latitude). Highest values are in red, while lowest values are shown blue.*

Dataset	Low clouds	Mid-level clouds	High clouds
CLARA-A2 AVHRR-AM	35.30%	24.71%	39.98%
Cloud_cci AVHRR-AM	45.51%	20.38%	34.11%
Coll6 MODIS-TERRA	57.53%	9.57%	32.90%
Cloud_cci MODIS-TERRA	49.40%	25.92%	24.68%
Cloud_cci MERIS+AATSR	32.95%	34.28%	32.77%
Cloud_cci ATSR2-AATSR	46.71%	22.48%	30.81%

A general conclusion is that MODIS-Terra C6 includes many more small-scale clouds over ocean with high pressure values and the utilization of more than just the AVHRR-heritage channel may also lead to a better vertical placement of thin, high clouds in multi-layer cloud situation, while for AVHRR-heritage channel (without a absorption channel) clouds are often placed in the middle of two layers leading to more mid-level clouds.

In terms of stability, all Cloud_cci datasets seem relative stable, except small trends existing for Cloud_cci AVHRR-AM before 2001 (due to the drift of NOAA-12, but note that there is no significant jump between NOAA12 and NOAA15 as seen for CLARA-A2), for Cloud_cci ATSR2-AATSR before 2003 (some remaining problems with ATSR2) and a small decreasing tendency in mean CTP of Cloud_cci MERIS+AATSR in the last3 years of the datasets.

Table 4-4 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. Biases are below 100hPa for all Cloud_cci datasets except for Cloud_cci MERIS+AATSR. For the latter the bc-RMSE is also highest. There is a slight positive trend in all Cloud_cci datasets, while C6 shows a negative trend of about -8hPa/decade. We suspect the latter to be caused by the channel 8.7µm degradation on MODIS-Terra (channel not used for Cloud_cci MODIS-Terra) which prevents using MODIS-Terra C6 as reference for stability.

Table 4-4: *Evaluation scores for Cloud_cci Level-3C Cloud Top Pressure (morning satellites) based on comparison to C6-Terra from 2003-2011, in case of MERIS+AATSR the C6-Terra Cloud Top Pressure is daytime only. The scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	AVHRR–AM	MODIS–Terra	ATSR2–AATSR	MERIS+AATSR
Bias [hPa]	–83.28	–32.29	–61.83	–131.83
bc–RMSD [hPa]	+91.09	+71.10	+93.43	+129.04
Cloud_cci trend [hPa / decade]	+8.27	+8.82	+6.79	+4.07
C6–Terra trend [hPa / decade]	–7.95	–7.94	–8.01	–8.85
∆ trend [hPa / decade]	+16.22	+16.76	+14.80	+12.92

4.1.3 Cloud Optical Thickness – Liquid clouds

In this subsection the cloud optical thickness of liquid clouds (COT_{liq}) of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-9), zonal mean (Figure 4-10) and standard deviation (Figure 4-11), all for a common time period, and time series plots (Figure 4-12). The COTs of the products refer to slightly different wavelengths[1]

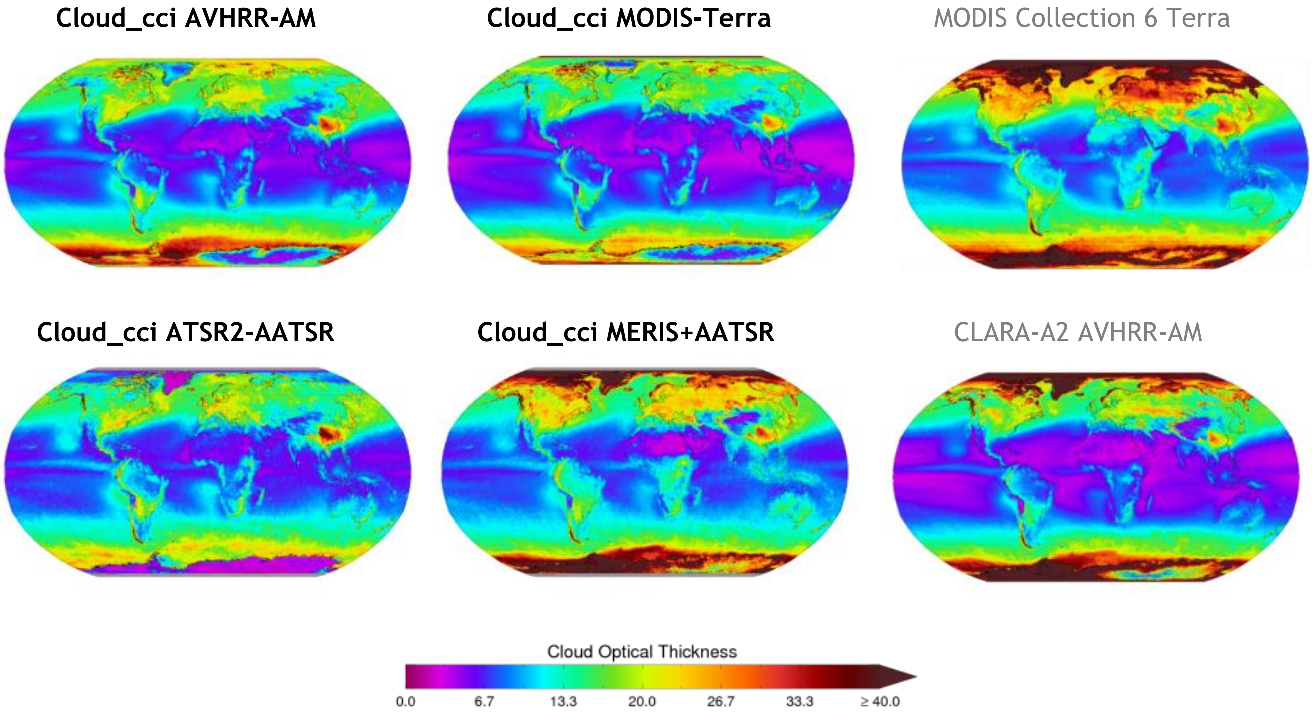
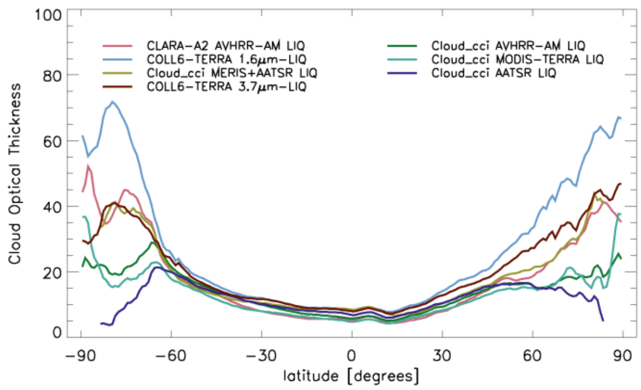


Figure 4-9 *Globally gridded means liquid cloud optical thickness for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*



Latitude weighted Means of Liquid Cloud Optical Thickness

Dataset-name	Global	± 60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	10.51	8.82	33.72	12.42	5.91	13.52	25.89	13.39	9.40
Cloud_cci AVHRR-AM	10.76	9.81	25.08	13.91	6.99	13.68	16.98	11.35	10.54
COLL6-TERRA	14.94	12.83	31.84	15.62	9.38	19.45	34.80	19.68	13.09
Cloud_cci MODIS-TERRA	9.29	8.38	20.47	11.40	5.97	12.13	16.94	10.77	8.71
Cloud_cci MERIS+AATSR	12.04	10.75	31.77	12.75	8.61	15.24	26.30	14.82	10.98
Cloud_cci AATSR	10.76	10.37	16.54	13.42	7.96	14.25	14.13	12.00	10.28

Figure 4-10 *Liquid Cloud Optical Thickness from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Terra the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

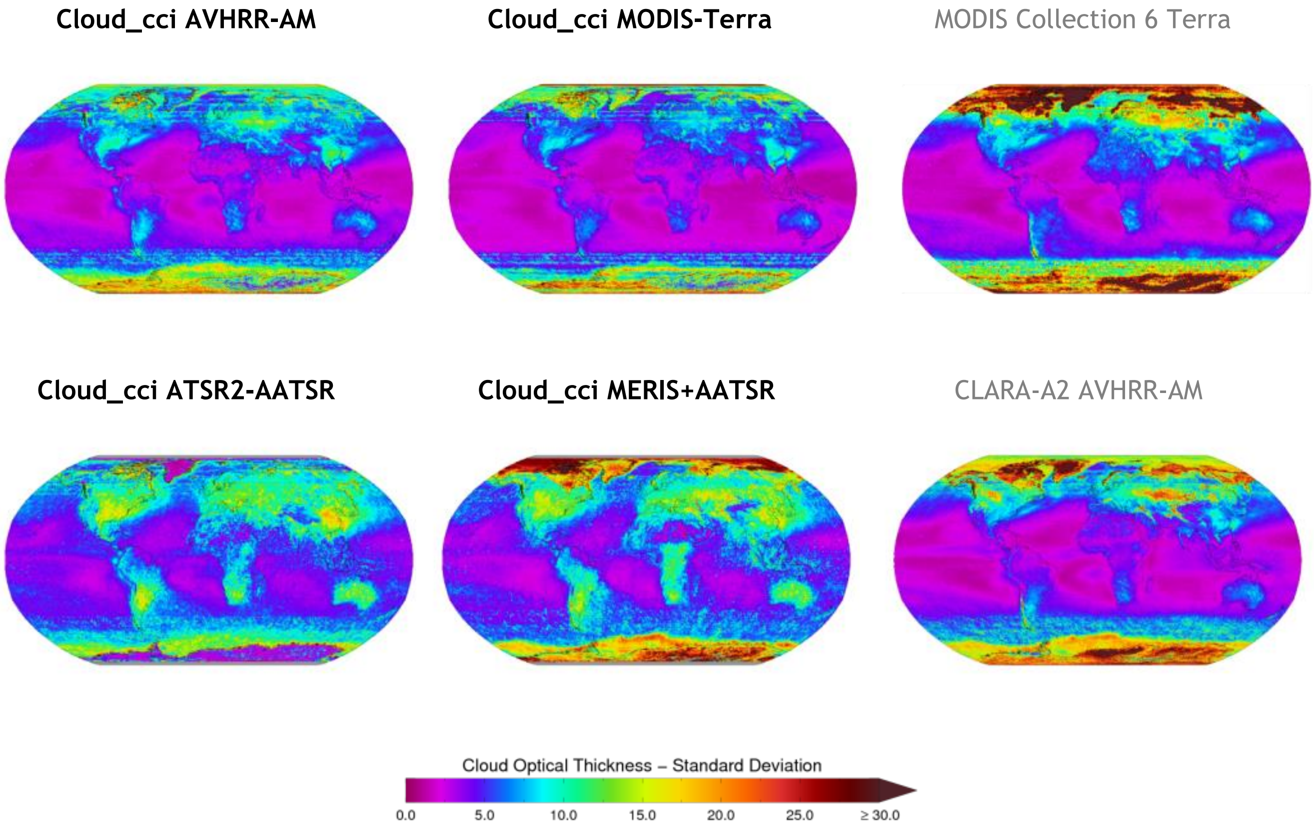


Figure 4-11 *Globally gridded means of liquid cloud optical thickness standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*

General findings

- For large parts of the globe the mean COT_{liq} looks similar among the datasets with higher values for MODIS Terra C6, Cloud_cci MERIS+AATSR and CLARA-A2.
- By far largest deviations are found for Polar Regions where the above mentioned datasets have extremely high values. To a smaller extend this can also be seen for AVHRR-AM over the Antarctic sea ice. The spread at the poles is nearly covering one order of magnitude.
- Lowest mean COT_{liq} are generally found in the subtropical regions with lowest values for CLARA-A2 and Cloud_cci MODIS-Terra and Cloud_cci AVHRR-AM
- In the Polar Regions, also the highest temporal variability is found, again most pronounced in MODIS Terra C6, Cloud_cci MERIS+AATSR and CLARA-A2. Outside the polar regions, largest temporal deviation is found over land regions in South-America, South-Africa, Australia, Eastern North-America, Eastern Europe and South-East Asia.
- Global mean values (between 60S and 60N) are generally between 5 and 15. Seasonal variations of these values are similar with higher values in boreal summer and lower in boreal winter.
- Time series plots show all datasets being relative stable in time. Small jumps occur for Cloud_cci AVHRR-AM and CLARA-A2 for the transitions from NOAA-12 to NOAA-15 and NOAA-15 to NOAA-17.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

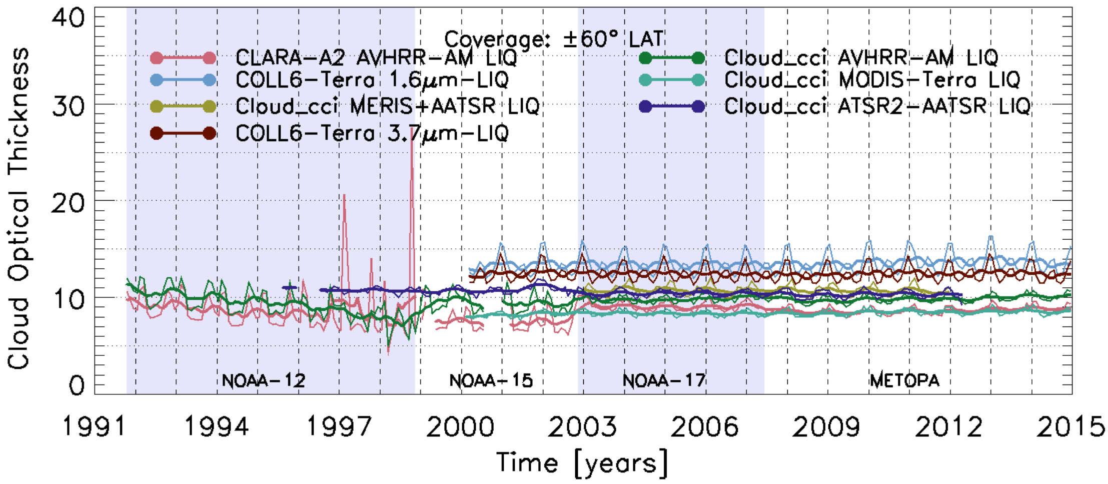


Figure 4-12 *Time series of liquid cloud optical thickness for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown.*

Discussion

For large parts of the globe the Cloud_cci COT_{liq} is in good agreement with the reference datasets. Mean COT_{liq} values around 10 in the tropics, slowly increasing with latitude to about 20 in the mid-latitudes, with only little spreads among the datasets and only small zonal variability. In contrast to that, in the high latitudes the spread becomes enormous, with Cloud_cci ATSR2-AATSR being at the lower end and MODIS-Terra C6 at the upper end and one order of magnitude between them. The very high COT_{liq} values in the Polar Regions seem not realistic since there is no reason to believe that clouds on the Polar Regions should be specifically optically thick. A retrieval artefact due to improper handling of snow and ice covered surfaces which usually causes a high reflectance in the visible seems more likely.

In case of the Cloud_cci datasets for which CC4CL was used (not MERIS+AATSR); a related bug has been identified (wrong treatment of the sun zenith angle in the calculation of the BRDF over snow and ice surfaces) after processing. The larger impact of this for Cloud_cci AVHRR-AM and Cloud_cci MODIS-Terra compared to Cloud_cci ATSR2-AATSR is likely due to the observation time differences and the narrower swath of ATSR2-AATSR, which could lead to a smaller number of conditions with high solar zenith angles. Also, wrongly detected clouds over snow and ice surfaces might play a role here. However, all Cloud_cci datasets seem relative stable in time, except a small decreasing trend for Cloud_cci AVHRR-AM for NOAA-12, thus before 1999, leading also to a small jump at the transition to NOAA-15. Reasons for that are likely to be the trend in cloud fraction seen before and the fact that drifting away from daytime leads to fewer daytime pixels for which COT_{liq} could be retrieved.

Table 4-5 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. The biases with respect to MODIS-Terra C6 are low (between -4 and -1 optical thickness, thus slightly lower mean values for Cloud_cci). The bc-RMSE are very similar for all Cloud_cci dataset (between 4 and 7 optical thickness). The stability for all datasets is very high.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

Table 4-5: *Evaluation scores for Cloud_cci Level-3C Optical Thickness - Liquid (morning satellites) based on comparisons to MODIS Terra C6 from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	$\text{AVHRR–AM}_{1.6}$	$\text{MODIS–Terra}_{3.7}$	$\text{ATSR2–AATSR}_{3.7}$	$\text{MERIS+AATSR}_{1.6}$
Bias	–3.30	–3.75	–1.59	–1.71
bc–RMSD	+6.47	+4.23	+6.38	+6.45
Cloud_cci trend [/ decade]	+0.03	+0.23	–0.05	–0.16
C6–Terra trend [/ decade]	+0.15	–0.04	+0.01	+0.27
Δ trend [/ decade]	–0.12	+0.27	–0.05	–0.43

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

4.1.4 Cloud Optical Thickness – Ice clouds

In this subsection the cloud optical thickness of ice clouds (COT_{ice}) of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-13), zonal mean (Figure 4-14) and standard deviation (Figure 4-15), all for a common time period, and time series plots (Figure 4-16). The COTs of the products refer to slightly different wavelengths[1].

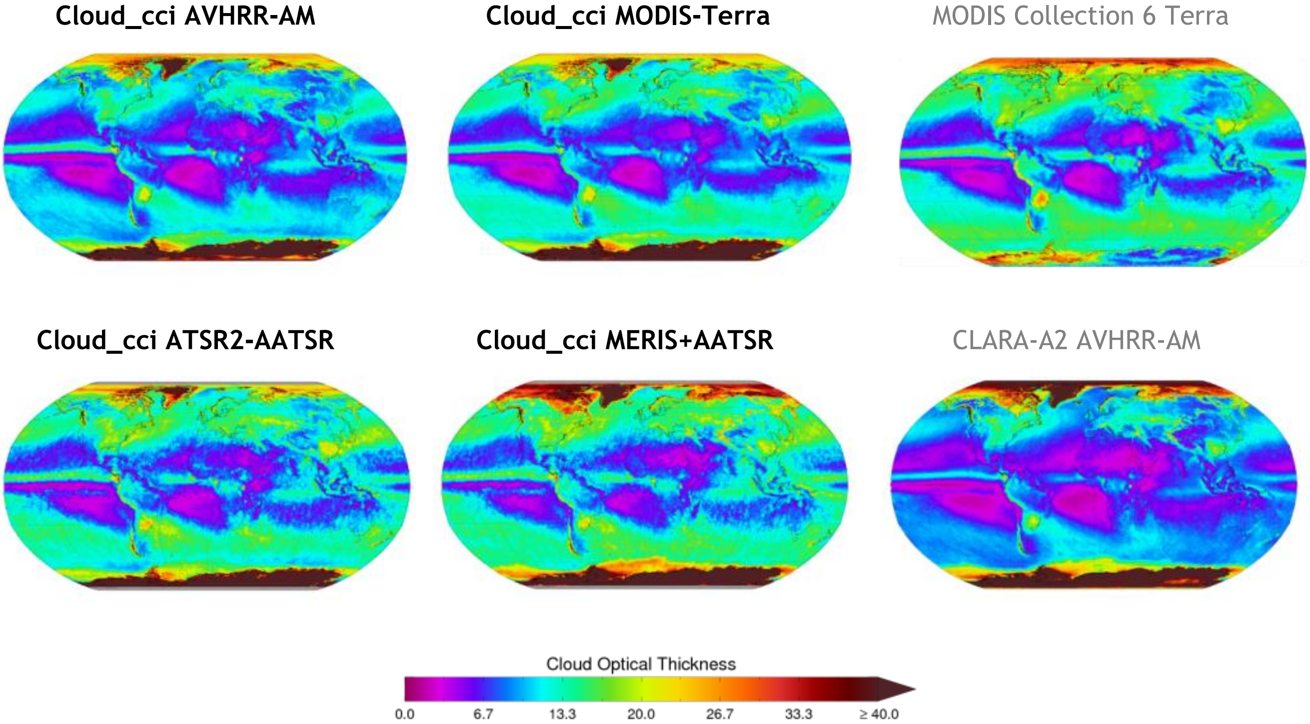
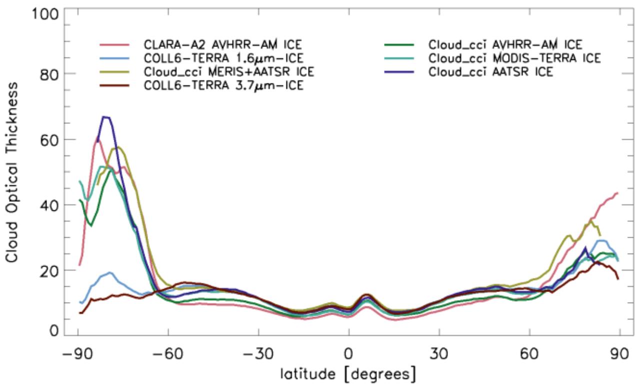


Figure 4-13 *Globally gridded means of ice cloud optical thickness for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*



Latitude weighted Means of Ice Cloud Optical Thickness

Dataset-name	Global	± 60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	9.47	7.88	31.35	9.35	6.44	10.21	21.13	11.90	8.50
Cloud_cci AVHRR-AM	9.95	8.87	25.18	10.64	7.50	10.83	16.54	11.47	9.36
COLL6-TERRA	11.51	10.92	14.87	14.01	8.83	13.38	18.02	11.67	11.44
Cloud_cci MODIS-TERRA	10.91	9.88	25.74	12.85	7.77	12.69	16.70	12.44	10.31
Cloud_cci MERIS+AATSR	12.76	11.41	32.59	13.60	9.60	13.95	23.46	13.42	12.48
Cloud_cci AATSR	11.86	10.87	27.99	13.32	8.87	13.47	16.85	12.98	11.41

Figure 4-14 *Ice Cloud Optical Thickness from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Terra the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

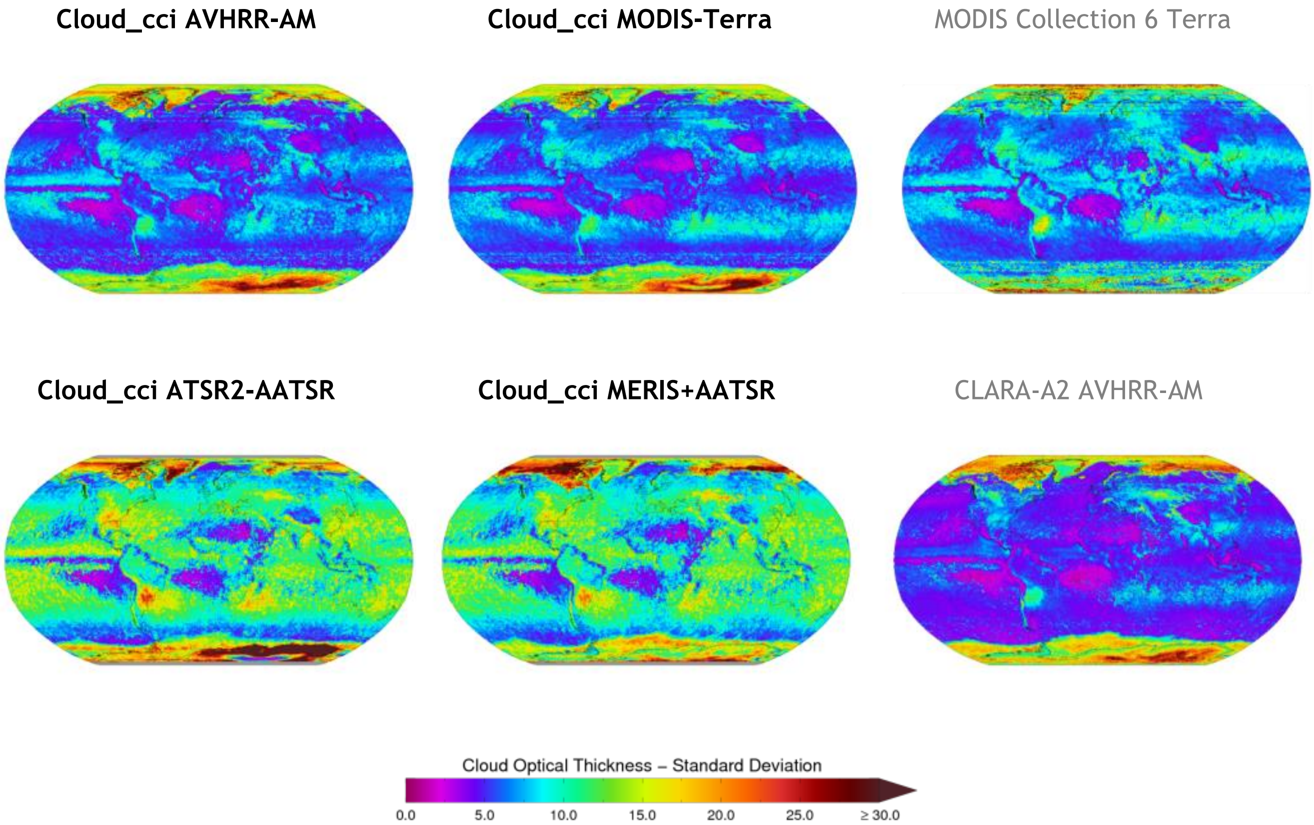


Figure 4-15 *Globally gridded means of ice cloud optical thickness standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*

General findings

- Comparable to COT_{liq}, the mean COT_{ice} is highest in Polar Regions for all datasets in particular in the Antarctic (above 30 optical thicknesses), with an exception for MODIS Terra C6. Lowest values are found in the subtropical regions, i.e. the stratocumulus regions, with mean value below 5. The general global distribution patterns are very similar for all datasets.
- The temporal variability is highest for MERIS+AATSR and ATSR2-AATSR, which seems to be two times as high as for the other datasets.
- The time series plots show a relative stable series for all datasets beyond 2003. Before 2003 the Cloud_cci AVHRR-AM and CLARA-A2 series suffer from difficulties in correctly treating the early morning satellites NOAA-12 and NOAA-15.
- The mid-morning satellites of AVHRR-AM that are using the 1.6µm channel instead of the 3.7µm (NOAA-17 and Metop-A) show a very stable time series with low variability.
- Cloud_cci AVHRR-AM with a constant positive bias compare to CLARA-A2.
- AATSR although using the 3.7 µm channel is higher than Cloud_cci AVHRR-AM using the 1.6µm channel after 2003.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

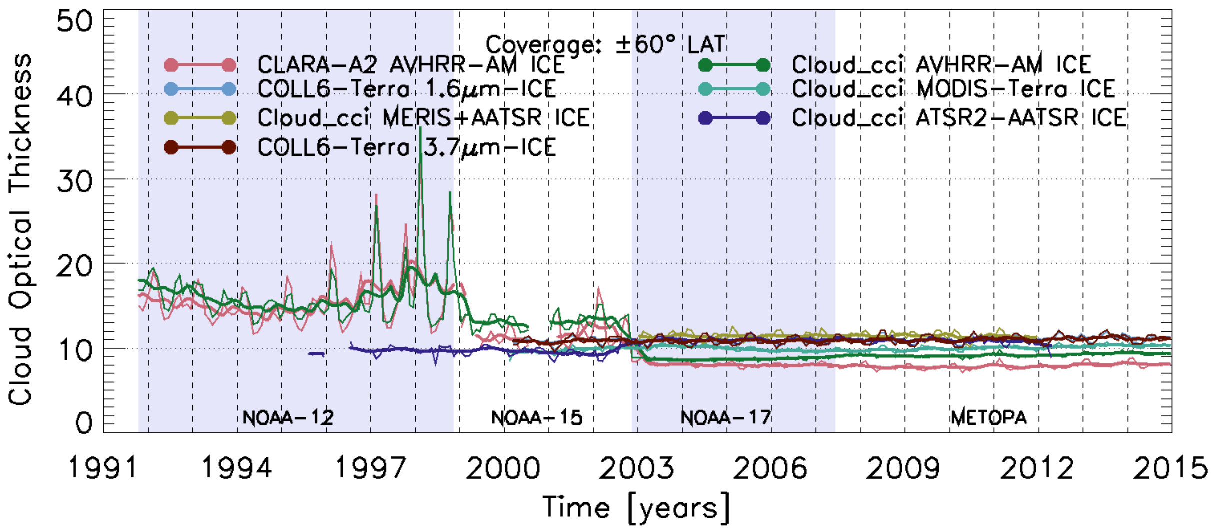


Figure 4-16 *Time series of ice cloud optical thickness for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown.*

Discussion

As seen for COT_{liq}, in the 60S-60N part of the globe the COT_{ice} of the Cloud_cci datasets is very similar among themselves and in comparisons to the reference datasets. All of them represent the local COT_{ice} maximum in the ITCZ, the minimum in the subtropics and show slightly increasing mean COT_{ice} values with increasing latitude, with Cloud_cci MERIS+AATSR and CLARA-A2 presenting max and min values. Thus, the mentioned global features are well represented in all Cloud_cci datasets.

However, In the Polar Regions the spread among the Cloud_cci datasets and the reference datasets is large. This is most likely due to the presence of snow/ice surfaces and the correspondingly difficult cloud detection, which could lead to pixel that contain snow/ice covered surfaces being erroneously classified as cloudy. In these cases, or in cases for which an optically thin cloud overlays a snow/ice surface, the retrieval schemes might attribute much of the measured reflection to the cloud optical thickness. In case of CC4CL-based Cloud_cci datasets, i.e. Cloud_cci AVHRR-AM and Cloud_cci MODIS-Terra, the results are additional affected by a bug in the calculation of the BRDF components over snow and ice surfaces and high solar zenith angle condition, which leads to a significant overestimation of COT_{ice} at these conditions.

The early-morning orbit is very close to twilight condition, thus solar zenith angles are very high, which makes the retrieval of optical cloud properties, e.g. optical thickness, extremely difficult. Differences to the liquid phase can be found in the behaviour of NOAA-12 which is not showing the strong negative trend as seen in the liquid phase.

Table 4-6 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. The biases of COT_{ice} are relatively low with value between -2 (Cloud_cci AVHRR-AM) and +0.3 (Cloud_cci MERIS+AATSR). The bc-RMSE for Cloud_cci AVHRR-AM and Cloud_cci MODIS-Terra is with approx. 5 optical thicknesses relative low; while the bc-RMSE for Cloud_cci ATSR2-AATSR and MERIS+AATSR is twice as large. Decadal trends remain below 1 optical thickness for all datasets - with highest values for Cloud_cci AVHRR-AM (0.7 / decade) – while only positive trends are found for Cloud_cci AVHRR-AM and MODIS-Terra C6.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

Table 4-6: *Evaluation scores for Cloud_cci Level-3C Optical Thickness - Ice (morning satellites) based on comparison to MODIS Terra C6 from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	AVHRR–AM$_{1.6}$	MODIS–Terra$_{3.7}$	ATSR2–AATSR$_{3.7}$	MERIS+AATSR$_{1.6}$
Bias	–1.96	–0.84	–0.29	+0.22
bc–RMSD	+4.91	+4.15	+9.07	+9.31
Cloud_cci trend [/ decade]	+0.70	–0.17	–0.07	–0.08
C6–Terra trend [/ decade]	+0.32	+0.30	+0.30	+0.42
Δ trend [/ decade]	+0.38	–0.46	–0.38	–0.50

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP (unknown). The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

4.1.5 Cloud Effective Radius – Liquid clouds

In this subsection the cloud effective radius of liquid clouds (CER_{liq}) of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-17), zonal mean (Figure 4-18) and standard deviation (Figure 4-19), all for a common time period, and time series plots (Figure 4-20).

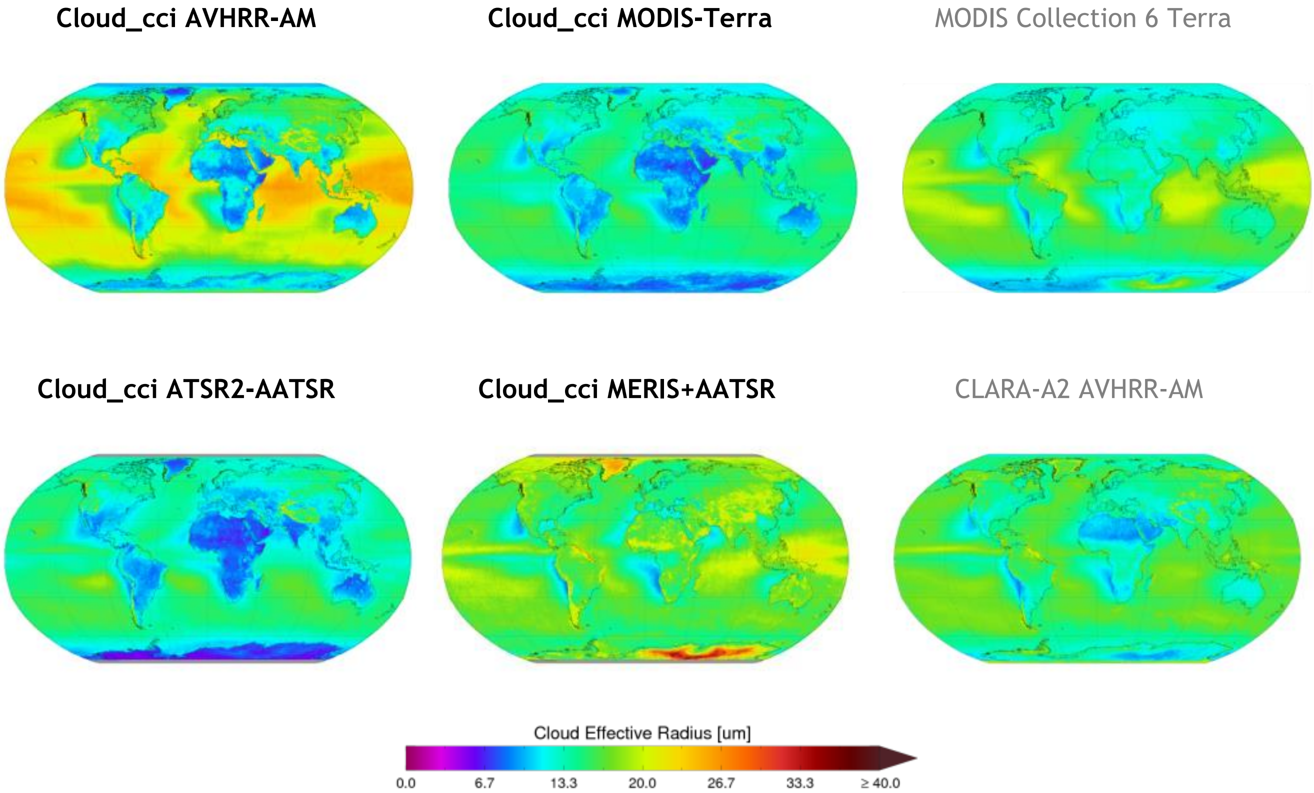
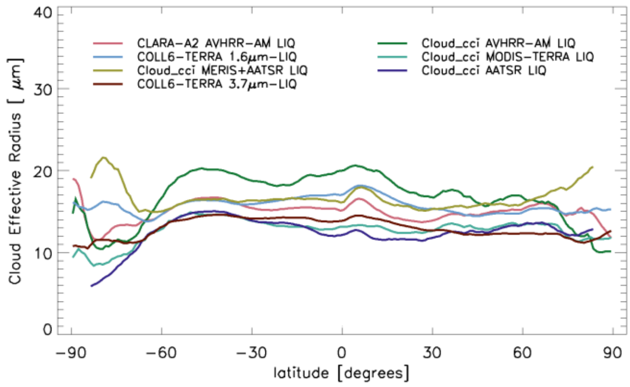


Figure 4-17 *Globally gridded means of liquid cloud effective radius for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*



Latitude weighted Means of Liquid Cloud Effective Radius [um]

Dataset-name	Global	± 60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	15.21	15.26	13.70	16.17	15.15	14.71	15.41	13.79	15.76
Cloud_cci AVHRR-AM	18.23	18.61	13.28	19.52	18.94	16.79	15.17	12.84	20.32
COLL6-TERRA	15.33	15.64	12.09	15.54	16.35	13.81	13.09	13.46	16.07
Cloud_cci MODIS-TERRA	13.23	13.34	11.07	14.40	13.03	13.18	13.00	10.95	14.12
Cloud_cci MERIS+AATSR	16.26	16.22	16.05	16.26	16.40	15.65	17.20	16.77	16.06
Cloud_cci AATSR	12.87	12.94	10.68	14.58	12.52	12.57	13.14	10.56	13.75

Figure 4-18 *Liquid Cloud Effective Radius from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Terra the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

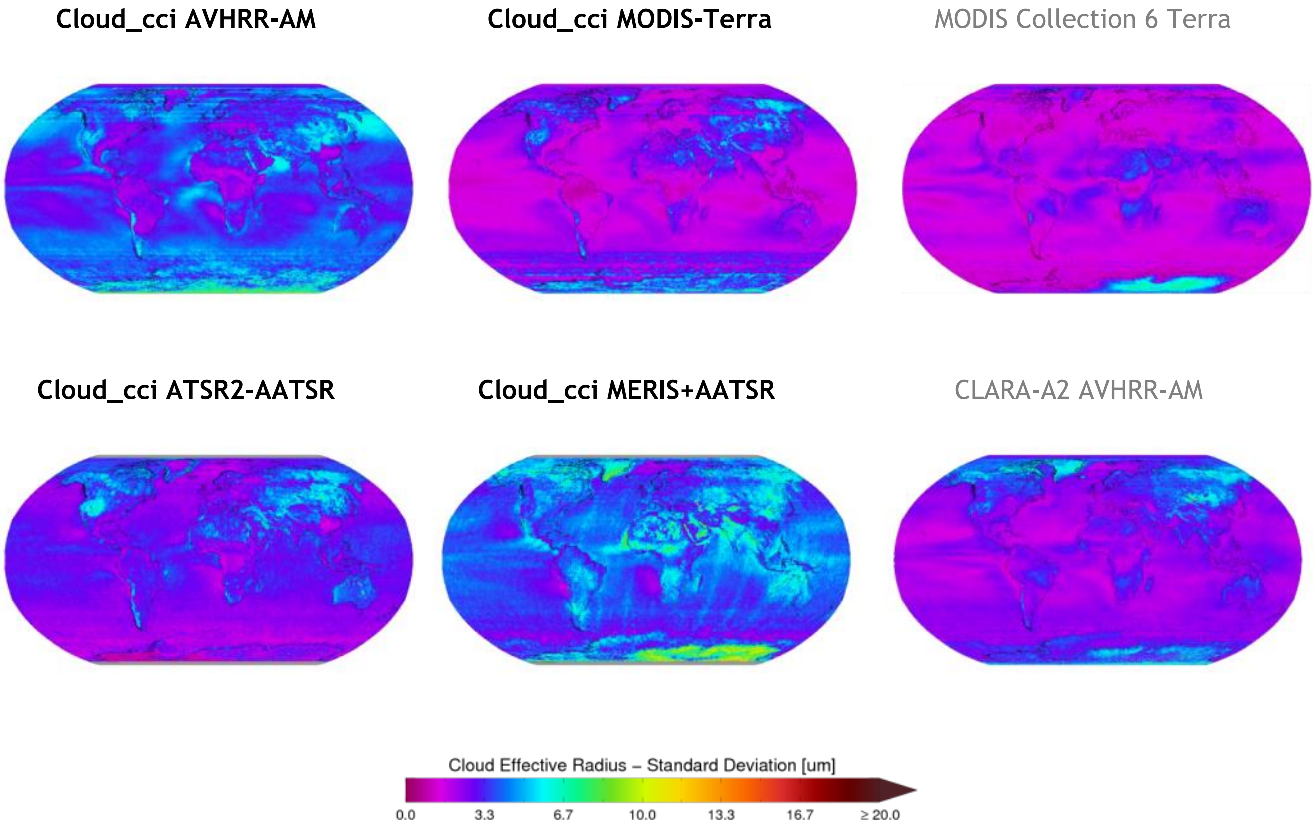


Figure 4-19 *Globally gridded means of liquid cloud effective radius standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*

General findings

- Mean CER_{liq} are lowest over land and highest over ocean in all datasets; however some deviations are found in the actual mean values, which are highest in Cloud_cci AVHRR-AM and lowest in Cloud_cci MODIS-Terra and Cloud_cci ATSR2-AATSR.
- In zonal mean plots it is seen that the mean CER_{liq} is between 10 and 20µm for all latitude band.
- The temporal variability is very low, with slightly higher values in Cloud_cci AVHRR-AM and Cloud_cci AATSR-MERIS compared to the others.
- All datasets are stable beyond 2003; AVHRR-AM reveals jumps in the time series of the global mean (60S-60N) values of CER_{liq} between the satellites NOAA-12 and NOAA-15. The jump from NOAA-15 to NOAA-17 is overplayed by a switch of the NIR channel available at daytime for the specific satellites. Which is 3.7µm for NOAA-12 and NOAA-15, and 1.6 µm for NOAA-17 and Metop-A.
- A small jump is also visible at the transition of the two sensors in ATSR2-AATSR dataset.

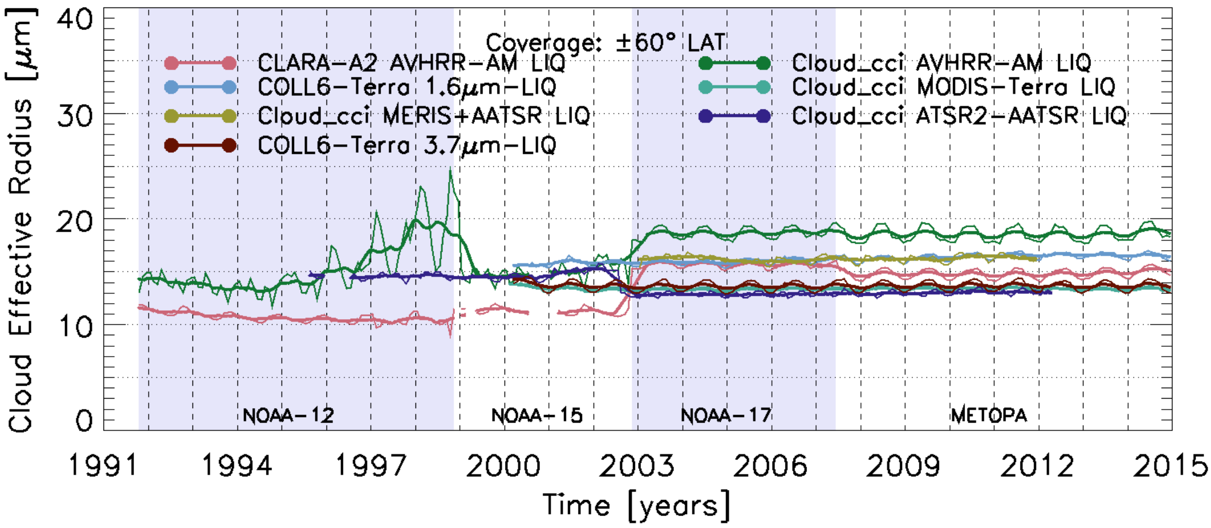


Figure 4-20 *Time series of liquid cloud effective radius for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown.*

Discussion

All Cloud_cci dataset agree to the reference data for CER_{liq} for most parts of the globe, except Cloud_cci AVHRR-PM for which the values are up to 5µm higher nearly everywhere over the ocean. As MODIS-Terra C6 and CLARA-A2, the Cloud_cci dataset do also show smaller mean CER_{liq} for land than over ocean. One could speculate if this potentially speaks for a higher number of cloud droplets over land due to a potentially higher aerosol concentration. Cloud_cci MERIS+AATSR deviates from this general finding by presenting rather similar numbers for land and ocean. What all datasets agree on is the smaller mean CER_{liq} for stratocumulus regions, compared to for example the trade cumulus regions. The spread among the datasets is highest in the tropics and seems to converge a little towards the mid-latitude and is larger again over the poles.

There is only little year-to-year variability of the 60S-60N mean values in all datasets beyond 2003. All datasets seem to have a similar seasonal cycle with a minimum in the boreal winter and a maximum in the boreal summer, the amplitudes being a bit diverse. The seasonal cycle of Cloud_cci ATSR2-AATSR is a little shifted. The transitions from ATSR2 to AATSR and from NOAA15 to NOAA17 can clearly be identifies. For the NOAA-AVHRR based datasets this is mainly due to the switch of the available daytime NIR channel, which is 3.7µm for NOAA12 and NOAA15 while 1.6µm for NOAA-17 and Metop-A. As the signal at 1.6µm can be assumed to come from lower parts of the cloud compared to measurements at 3.7µm, we can infer information about the vertical profiles of CER_{liq} in liquid clouds. Cloud_cci AVHRR-AM (specifically looking at NOAA15 and NOAA17) agrees well with CLARA-A2 and MODIS-Terra C6 (1.6 and 3.7µm products) with respect to showing the 1.6µm CER_{liq} being about 5µm larger than the 3.7µm CER_{liq} on average for 60S-60N; in other words smaller cloud droplets at the cloud top compared to below. Regionally, i.e. in regions dominated by stratocumulus clouds, one can find the opposite vertical structure in Cloud_cci AVHRR-AM (not shown).

Noticeable again is the problem with NOAA-12 increasingly during its lifetime. However, as this is only seen for Cloud_cci AVHRR-AM and not CLARA-A2, we assume some shortcomings in the CC4CL retrieval either for CER, cloud detection or cloud phase determination.

Table 4-7 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. The bias of +2.4 µm for AVHRR-AM sticks out a bit compared to the bias of the other Cloud_cci datasets. Also the bc-RMSE is highest for Cloud_cci AVHRR-AM with values of 4.2µm; while the bc-RMSE for the other Cloud_cci datasets wrt. MODIS-Terra C6 is a bit lower. Decadal trends vary between small negative and small positive values. MODIS-Terra C6 trend for the 1.6 CER_{liq} is 0.85µm / decade and a bit suspicious because no other dataset showing this to a similar extent.

Table 4-7: *Evaluation scores for Cloud_cci Level-3C Effective Radius - Liquid (morning satellites) based on comparison to MODIS Terra C6 from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	$AVHRR\text{−}AM_{1.6}$	$MODIS\text{−}Terra_{3.7}$	$ATSR2\text{−}AATSR_{3.7}$	$MERIS\text{+}AATSR_{1.6}$
Bias [µm]	+2.44	−0.30	−0.73	+0.04
bc−RMSD [µm]	+4.24	+1.91	+2.77	+3.48
Cloud_cci trend [µm / decade]	−0.18	+0.15	+0.33	+0.15
C6−Terra trend [µm / decade]	+0.85	+0.11	+0.11	+0.86
Δ trend [µm / decade]	−1.03	+0.04	+0.23	−0.71

4.1.6 Cloud Effective Radius – Ice clouds

In this subsection the cloud effective radius of ice clouds ($CER_{ice)}$ of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-21), zonal mean (Figure 4-22) and standard deviation (Figure 4-23), all for a common time period, and time series plots (Figure 4-24).

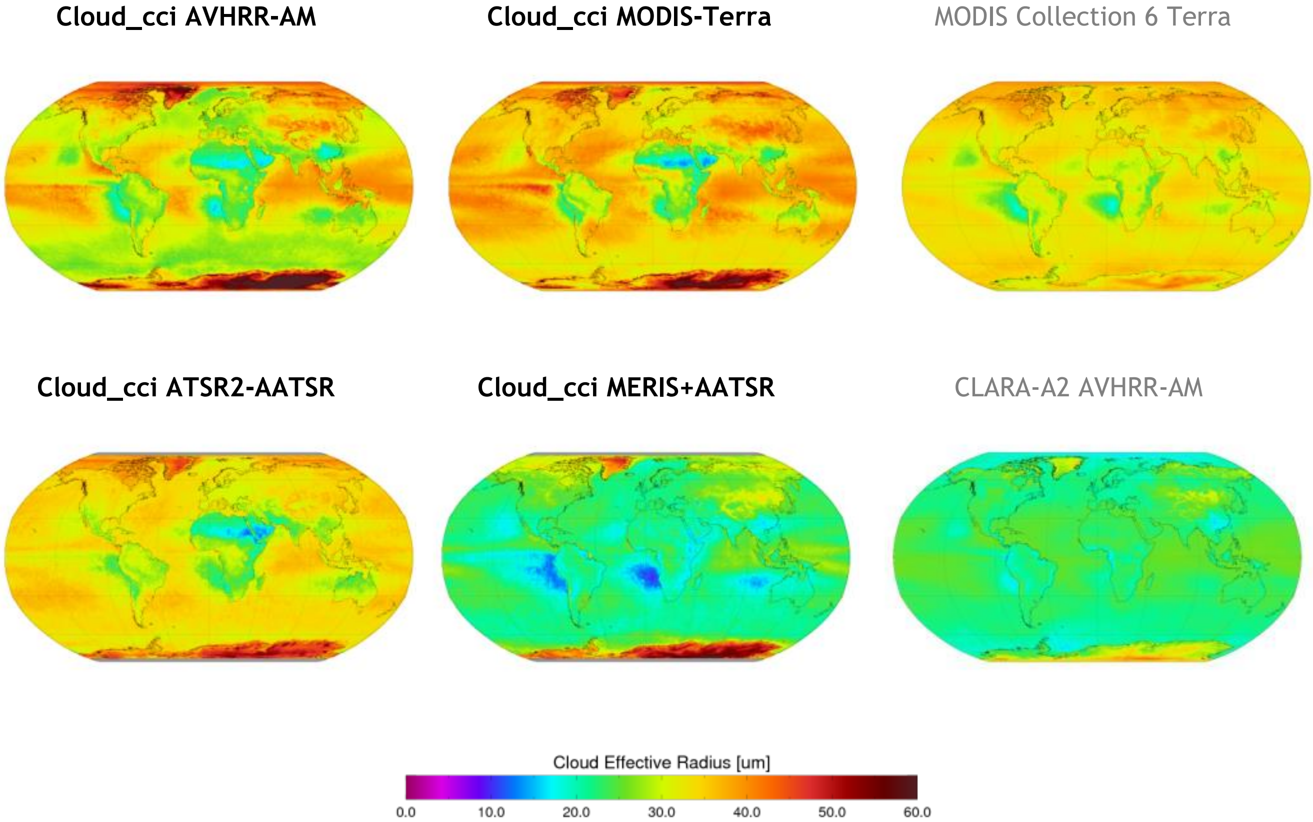
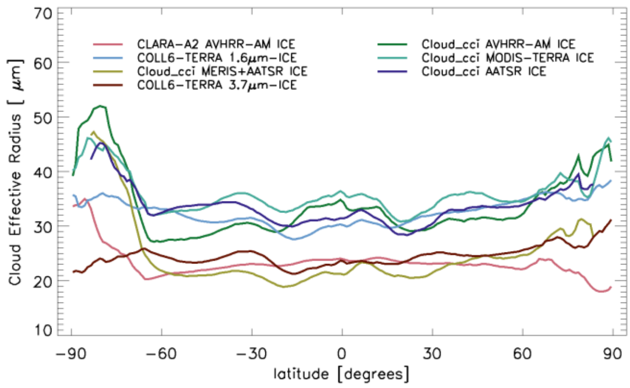


Figure 4-21 *Globally gridded means of ice cloud effective radius for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*



Latitude weighted Means of Ice Cloud Effective Radius [um]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	23.10	23.13	23.00	22.23	23.56	22.91	22.61	23.13	23.09
Cloud_cci AVHRR-AM	31.48	31.03	35.19	28.96	31.73	31.21	36.62	29.86	32.12
COLL6-TERRA	32.78	32.47	34.43	33.07	31.69	33.93	36.42	32.45	32.92
Cloud_cci MODIS-TERRA	34.60	34.36	36.96	34.30	34.04	35.31	36.94	31.80	35.69
Cloud_cci MERIS+AATSR	22.47	21.92	30.08	21.40	21.44	23.59	27.21	23.37	22.09
Cloud_cci AATSR	32.49	32.15	36.20	33.91	31.08	33.10	35.62	30.12	33.46

Figure 4-22 *Ice Cloud Effective Radius from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Terra the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

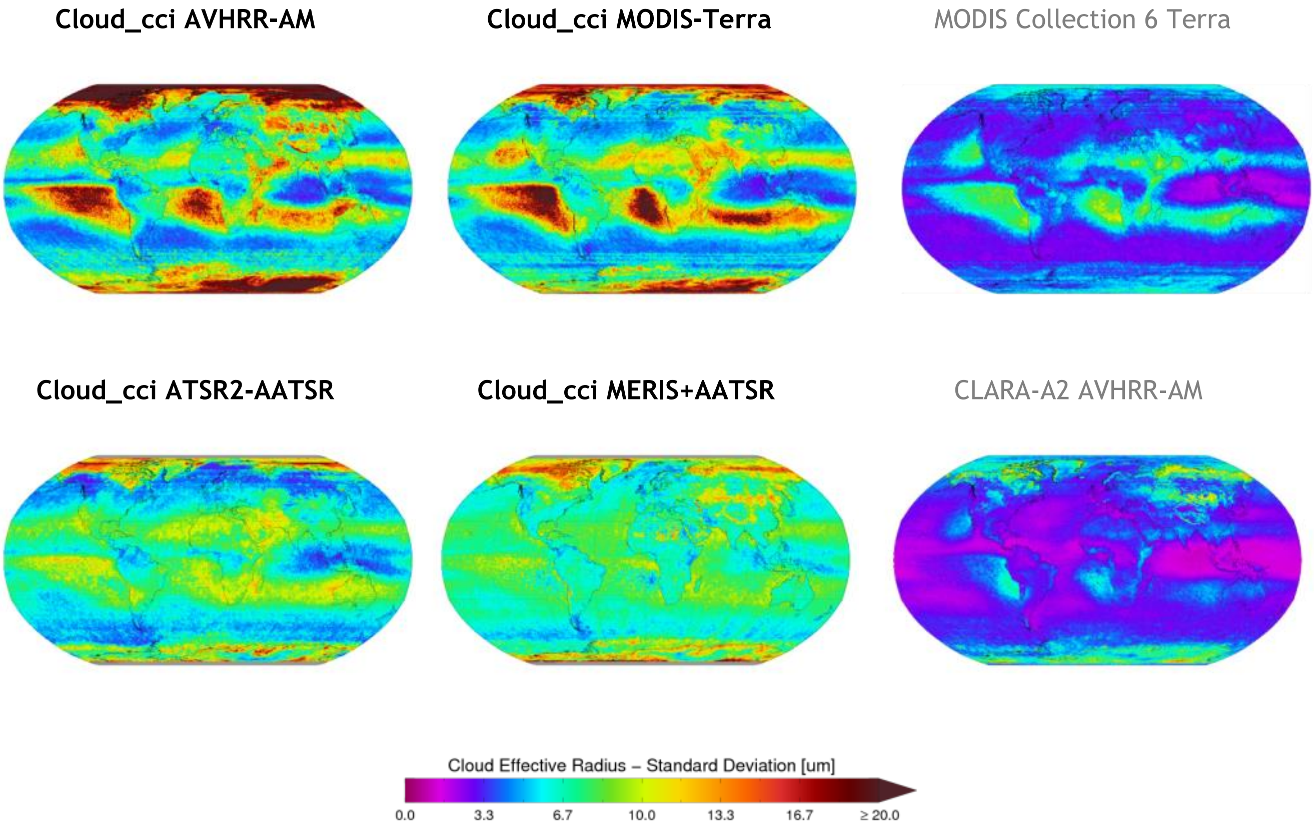


Figure 4-23 *Globally gridded means of ice cloud effective radius standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*

General findings

- Global maps of mean CER_{ice} show quite some systematic deviation between the CC4CL-based datasets together with MODIS C6 compared to Cloud_cci MERIS+AATSR and CLARA-A2. The deviation is nearly a factor of two. However, major global pattern remains similar in all datasets with lower values in the stratocumulus regions and higher values in the trade-cumulus regions and Tropics. Extremely higher values are found over Polar land regions in all Cloud_cci datasets.
- In terms of temporal variability, all datasets show highest variability in subtropical and Polar Regions. However, the variability spread among the datasets is, similar to the mean values, partly deviating by a factor of two, with highest variability values. For Cloud_cci AVHRR-PM and Cloud_cci MODIS-Terra.
- Wrt. the long-term stability, one can again see the switch of available (and used) near-infrared channel from NOAA-15 to NOAA-17 (from 3.7 to 1.6µm) in Cloud_cci AVHRR-AM and CLARA-A2, however, with different implications.

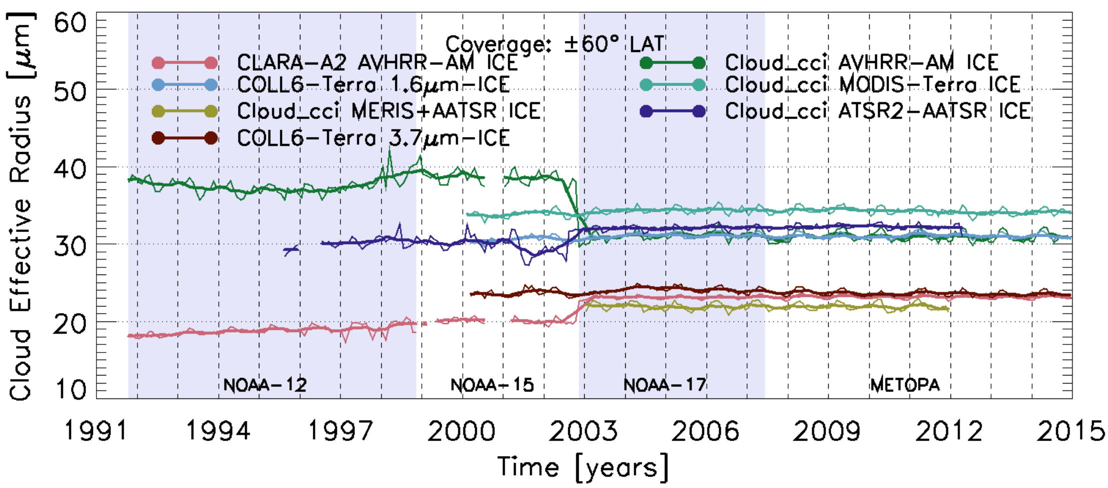


Figure 4-24 *Time series of ice cloud effective radius for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown.*

Discussion

The time series if CER_{ice} reveals clearly a problem found in the CC4CL (all Cloud_cci except Cloud_cci MERIS+AATSR) 3.7µm ice retrieval. Esp. the early AVHRR satellites seem unreasonable high. After switching to the 1.6µm channel for NOAA-17 and METOP-A, Cloud_cci AVHRR-AM is very much in line with MODIS-Terra C6 1.6µm retrieval. As for NOAA12 and NOAA15 in the Cloud_cci AVHRR-AM time data, the ATSR2-AATSR datasets (using only the 3.7µm) also shows a significant overestimation of CER_{ice} when compared against the Coll6 MODIS-Terra C6 product for 3.7µm. The main reason for the found high biases for CC4CL-based datasets has been narrowed down to erroneous 3.7µm channel look-up-tables used in CC4CL. This has been found after processing and will be fixed in a future new, reprocessed version.

Besides the positive biases for 3.7µm based CC4CL datasets, and the therewith partly introduced jumps in the time series of Cloud_cci AVHRR-AM, all Cloud_cci time series seem rather stable with nearly no year-to-year variability for the 60S-60N average nor long-term trend. A small season cycle is present, similar for all datasets, with a maximum in the boreal spring and a minimum at boreal autumn.

As for liquid clouds, the difference between 1.6µm and 3.7µm retrievals of CER_{ice} can give information about the vertical structure of particle sizes in ice clouds. When the new AVHRR-AM version is processed (including the corrected LUTs) it can be investigated if it will agree to what can be seen for CLARA-A2 and MODIS C6: higher CER_{ice} for 1.6µm than for 3.7µm on average for 60S-60N; which in other words means smaller ice particles found at the cloud top than below.

Table 4-8 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. Compared to liquid clouds, the agreement between Cloud_cci and MODIS-Terra C6 datasets is less. Except from Cloud_cci AVHRR-AM, we find large biases against the corresponding MODIS product (see also Table caption), which show a large overestimation of CER_{ice} for Cloud_cci MODIS-Terra and ATSR2-AATSR, while a underestimation for MERIS+AATSR. The mean Cloud_cci AVHRR-AM data seems to agree well with MODIS-Terra C6 for the 2003-2011 period, for which Cloud_cci AVHRR-AM used the 1.6µm channel of NOAA-17 and Metop-A. The large biases for Cloud_cci ATSR2-AATSR and MODIS-Terra are due to the 3.7µm LUT bug as mentioned before. The bc-RMSE against MODIS-Terra C6 is very similar for all Cloud_cci datasets with value between 6 and 8 µm. There is only little trend in all Cloud_cci dataset (below 0.2µm / decade), while MODIS-Terra C6 3.7µm product shows a clear negative trend.

Table 4-8: *Evaluation scores for Cloud_cci Level-3C Effective Radius - Ice (morning satellites) based on comparison to MODIS Terra C6 from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	AVHRR–AM$_{1.6}$	MODIS–Terra$_{3.7}$	ATSR2–AATSR$_{3.7}$	MERIS+AATSR$_{1.6}$
Bias [µm]	+0.09	+10.43	+7.98	–9.55
bc–RMSD [µm]	+7.94	+6.21	+6.91	+6.81
Cloud_cci trend [µm / decade]	–0.17	–0.10	+0.35	–0.11
C6–Terra trend [µm / decade]	+0.03	–0.79	–0.77	+0.01
Δ trend [µm / decade]	–0.20	+0.69	+1.12	–0.12

4.1.7 Liquid Water Path

In this subsection the cloud liquid water path of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-25), zonal mean (Figure 4-26) and standard deviation (Figure 4-27), all for a common time period, and time series plots (Figure 4-28 and Figure 4-29).

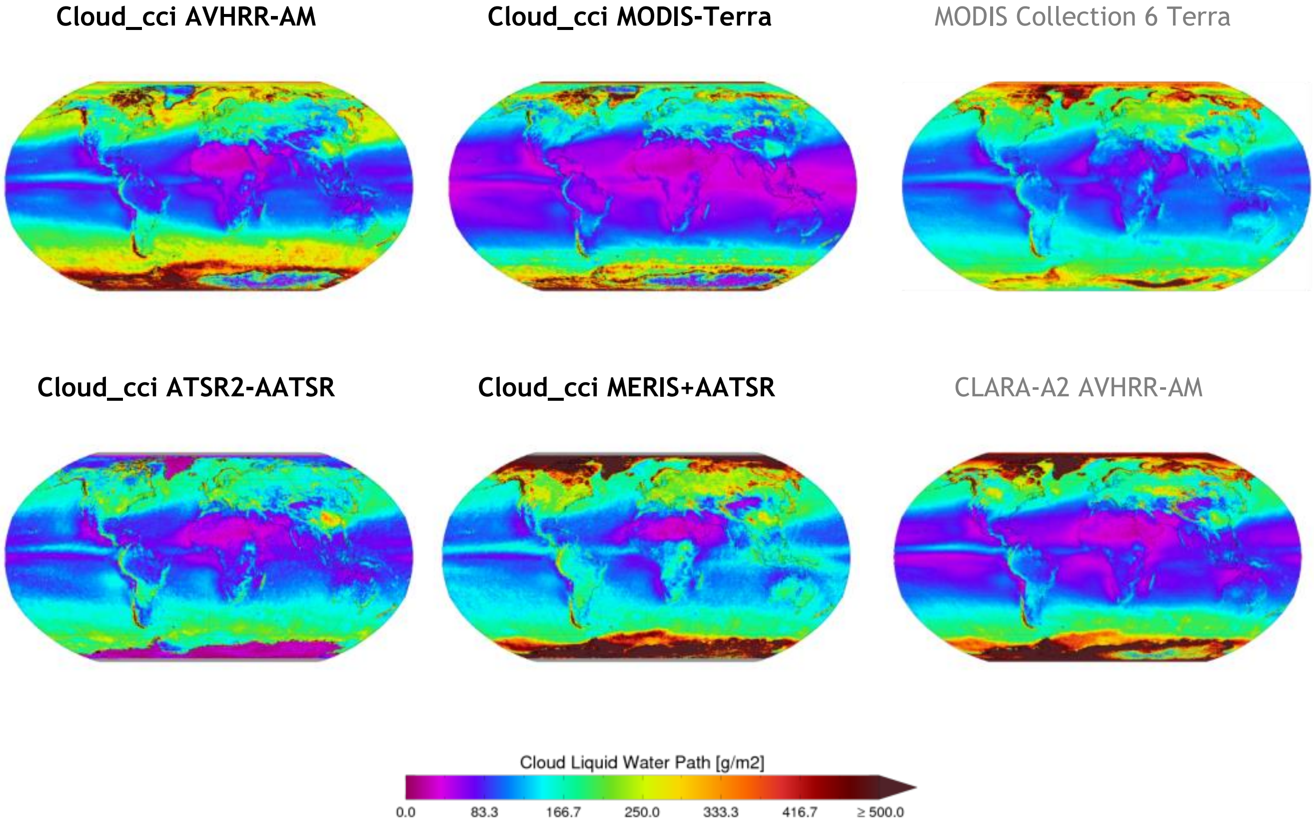
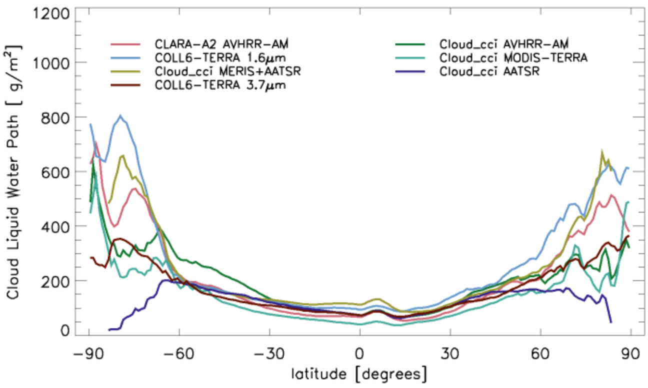


Figure 4-25 *Globally gridded means of cloud liquid water path for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1μm product is plotted.*



Latitude weighted Means of Cloud Liquid Water Path [g/m2]

Dataset–name	Global	±60° Lat.	Ant–arctica	Mid–Lat. South	Tropics	Mid–Lat. North	Arctic	Land	Sea
CLARA–A2 AVHRR–AM	125.37	104.90	383.00	150.25	73.24	150.30	328.96	151.29	115.34
Cloud_cci AVHRR–AM	143.84	129.86	345.55	207.91	87.29	172.58	241.74	124.49	151.34
COLL6–TERRA	136.22	122.71	232.61	148.55	98.74	162.26	274.92	161.77	126.24
Cloud_cci MODIS–TERRA	93.99	80.56	254.77	121.84	53.06	116.49	211.83	103.20	90.42
Cloud_cci MERIS+AATSR	147.41	130.22	394.63	152.32	110.98	166.61	350.60	171.35	138.29
Cloud_cci AATSR	116.72	113.50	152.77	153.02	88.09	147.46	154.29	118.37	116.09

Figure 4-26 *Cloud Liquid Water Path from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Terra the 1.6μm and 3.7μm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Terra the 2.1μm product is used. Highest and lowest values of each region are highlighted.*

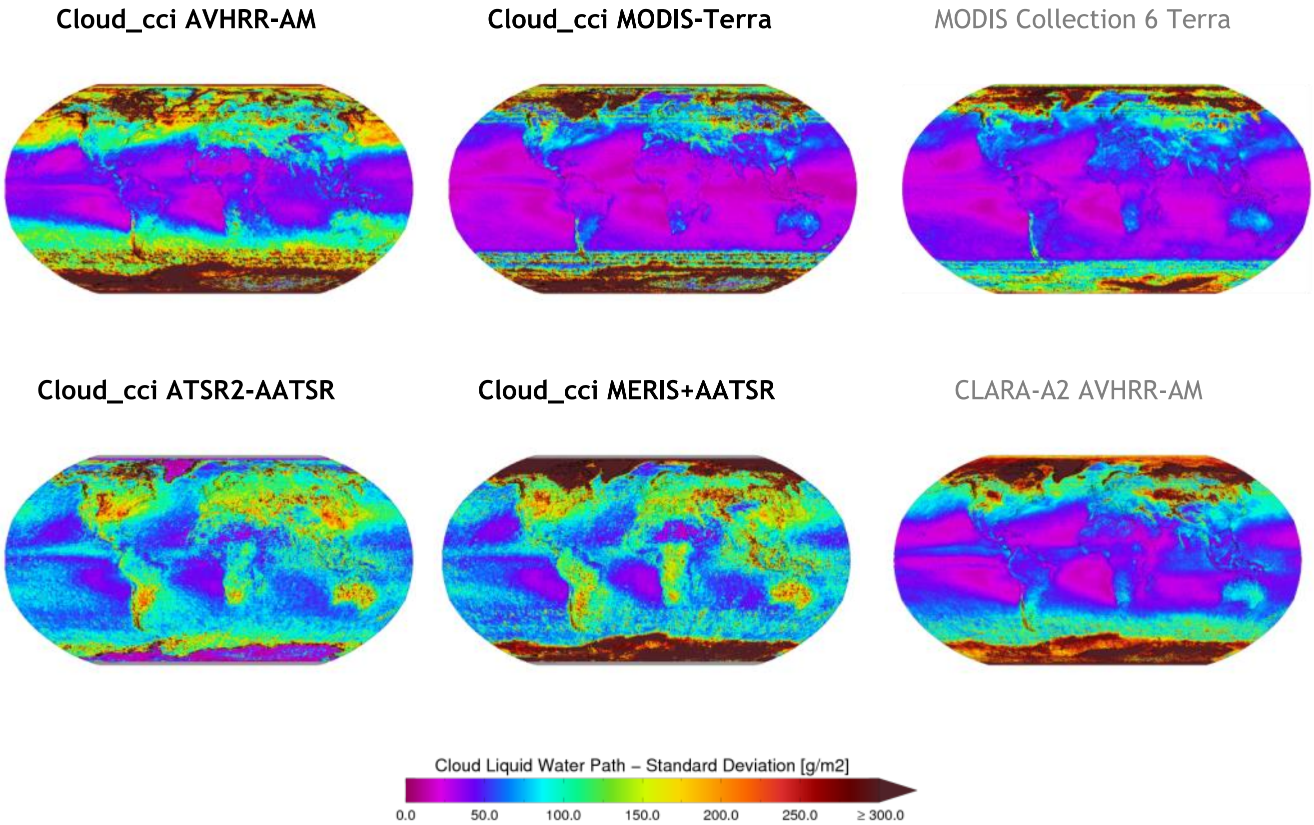


Figure 4-27 *Globally gridded means of cloud liquid water path standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*

General findings

- For large parts of the globe the liquid water content looks comparable among the datasets, with a small maximum in the inner tropics, somewhat less LWP in the subtropics (in particular in the stratocumulus regions) and increased LWP again for the mid-latitudes. In nearly all datasets very high mean LWP values are shown for the Polar Regions, except in Cloud_cci ATSR2-AATSR. The mean LWP in these regions goes up to 800 g/m², while the mean LWP is below 200 g/m² in all datasets for the rest of the globe.
- The temporal variability of LWP is strongly correlated to the mean LWP values, meaning the regions with low LWP exhibit low temporal variability and vice versa. Again, Polar Regions show highest variability.
- Global (60S-60N) mean values are lowest for Cloud_cci MODIS-Terra (approx. 80g/m²) and highest for MODIS C6 Terra with approx. 140g/m², but all datasets are stable beyond 2003. Transitions from one sensor to another caused jumps in the time series for Cloud_cci AVHRR-AM, CLARA-A2 and ATSR2-AATSR. In addition, LWP of Cloud_cci AVHRR-AM and CLARA-A2 is significantly impacted by the satellite drift of NOAA-12 between 1992 and 1999.
- Time series of the all-sky LWP highlight a good agreement among all datasets, except Cloud_cci AVHRR-AM, with nearly identical mean values of LWP for the 60S-60N region.

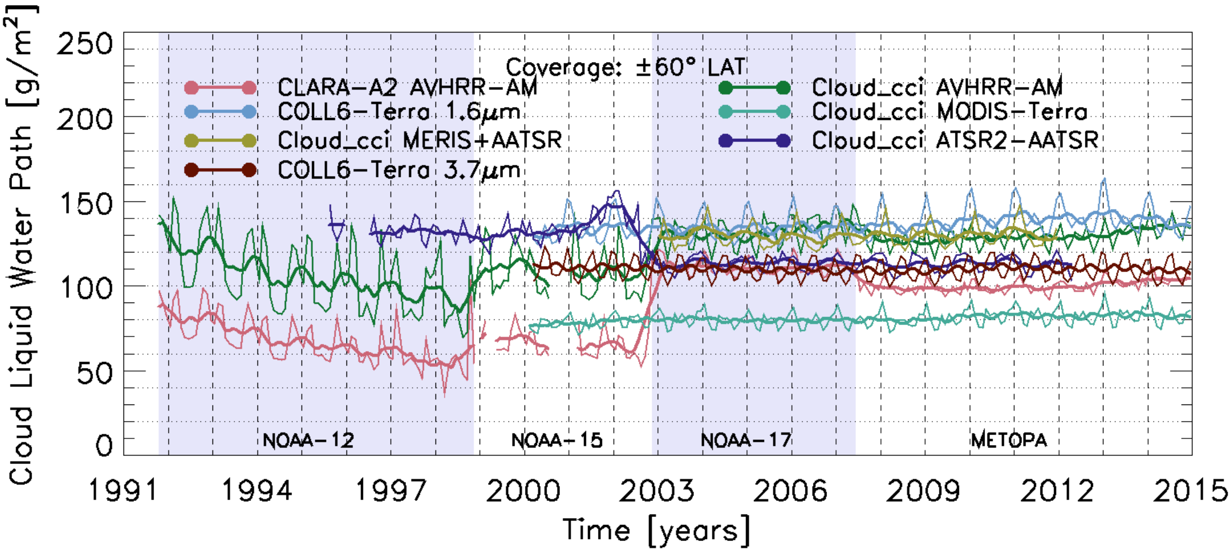


Figure 4-28 *Time series of "in-cloud" cloud liquid water path for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown.*

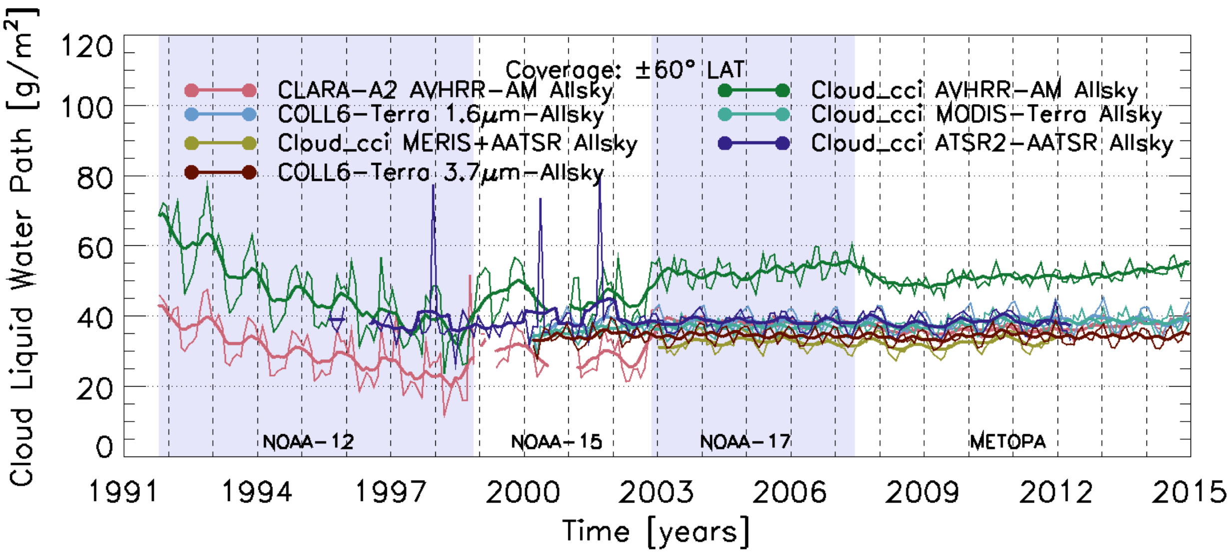


Figure 4-29 *Time series of "all-sky" cloud liquid water path for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown.*

Discussion

All Cloud_cci dataset reflect the typical meridional distribution of LWP, with a small local maximum in the inner tropic, smaller values in the outer tropics and slight increasing mean LWP values towards the mid and higher latitudes. Cloud_cci MODIS-Terra generally presents the lowest values while Cloud_cci MERIS+AATSR the maximum values for nearly all latitude bands, with the exception that Cloud_cci AVHRR-AM increases dramatically in the southern mid-latitudes. Evaluating the time series plots, from NOAA-15 onwards Cloud_cci AVHRR-AM agrees very well with the respective MODIS-Terra C6 retrievals for 1.6µm. Also, Cloud_cci ATSR2-AATSR (using 3.7µm) is very similar to MODIS Terra C6 for 3.7µm. However Cloud_cci MODIS-TERRA is very low compared to all other datasets, with shortcomings in cloud fraction and cloud phase determination potentially being the driving reasons for this. Cloud_cci AVHRR-AM has significantly lower values for NOAA-15 and NOAA-12 which is mainly due to the 3.7µm channel being switched on and used. A resulting jump can be seen at the transition from NOAA15 to NOAA17. The LWP of ATSR2 also see very high compared to the other 3.7µm data, in particular visible when compared against AATSR. The found seasonal cycle of 60S-60N time series is similar for the Cloud_cci datasets compared to the reference data, although with deviating amplitude.

Looking at the time series of the allsky-means, the difference appear to be vanishing. Cloud_cci datasets are very close together with the reference data after 2003 except Cloud_cci AVHRR-AM, which again points to shortcomings of either cloud fraction, cloud phase determination or both.

Table 4-9 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. Although the biases are relative small (only a bit amplified for Cloud_cci MODIS-Terra), the bc-RMSE are relative large (values between 75 and 107 g/m²). Decadal trends are divers among the Cloud_cci datasets, but none is as large as the trend of MODIS-Terra C6 1.6µm with +6 g/m² /decade.

Table 4-9: *Evaluation scores for Cloud_cci Level-3C Liquid Water Path (morning satellites) based on comparison to MODIS Terra C6 from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	AVHRR–AM$_{1.6}$	MODIS–Terra$_{3.7}$	ATSR2–AATSR$_{3.7}$	MERIS+AATSR$_{1.6}$
Bias [g/m^2]	–3.28	–26.45	+7.44	+2.00
bc–RMSD [g/m^2]	+106.22	+60.39	+75.94	+86.44
Cloud_cci trend [g/m^2 / decade]	–4.25	+3.33	+0.64	+0.21
C6–Terra trend [g/m^2 / decade]	+6.12	+0.56	+0.92	+6.77
Δ trend [g/m^2 / decade]	–10.37	+2.77	–0.28	–6.56

4.1.8 Ice Water Path

In this subsection the cloud ice water path of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-30), zonal mean (Figure 4-31) and standard deviation (Figure 4-32), all for a common time period, and time series plots (Figure 4-33 and Figure 4-34).

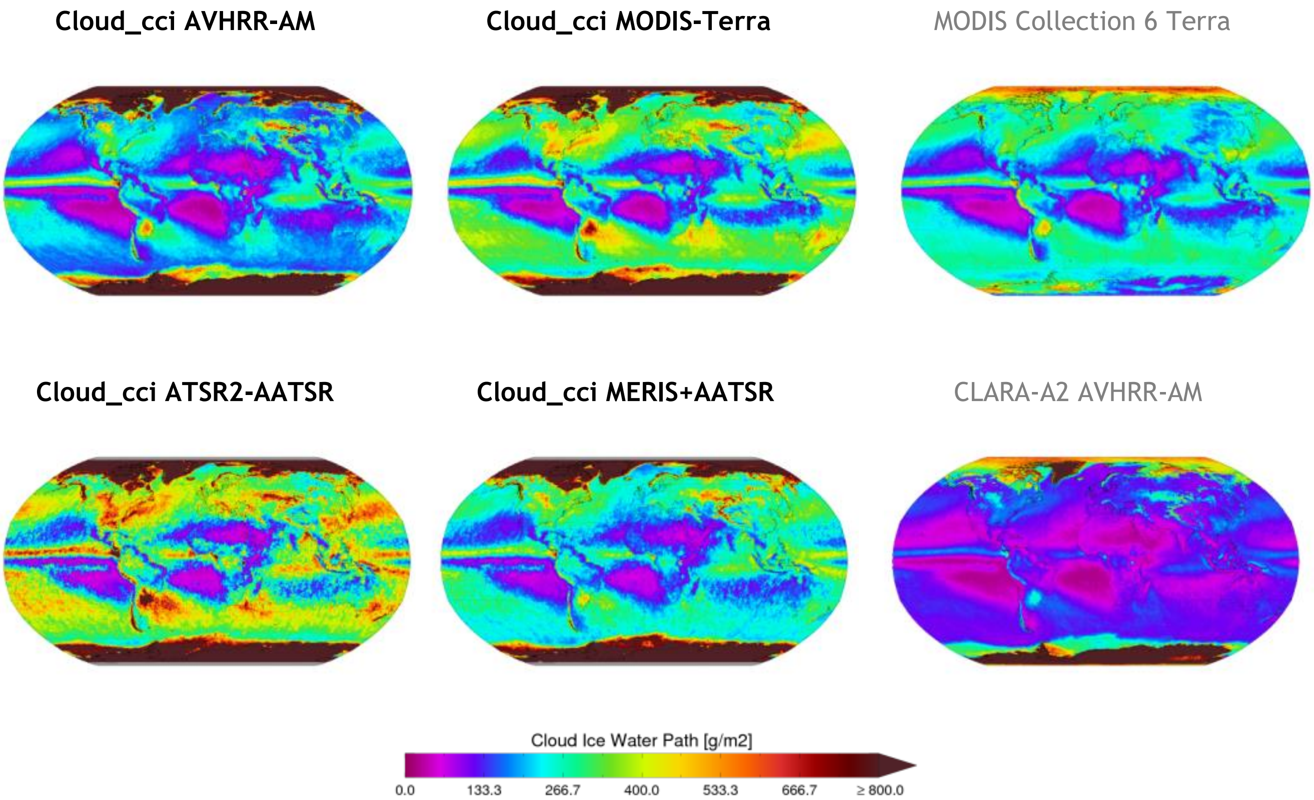
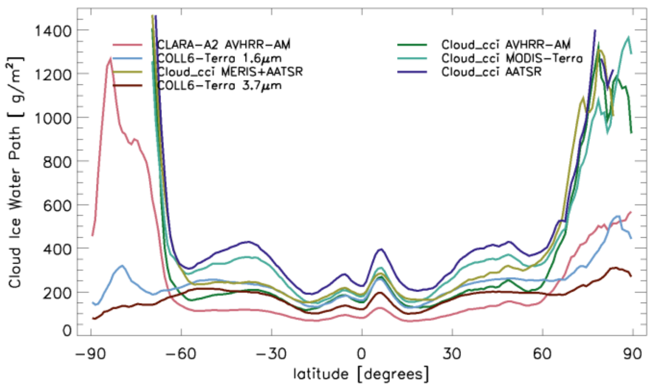


Figure 4-30 *Globally gridded means of cloud ice water path for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*



Latitude weighted Means of Cloud Ice Water Path [g/m2]

Dataset-name	Global	± 60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	130.31	103.83	504.51	116.10	88.44	132.00	317.41	173.64	113.21
Cloud_cci AVHRR-AM	248.50	182.19	1221.78	192.52	166.21	214.95	616.99	347.72	209.30
COLL6-TERRA	215.75	207.96	236.95	255.04	173.23	253.03	324.61	214.91	216.10
Cloud_cci MODIS-TERRA	320.17	260.05	1243.55	332.38	207.22	332.35	607.01	398.10	289.83
Cloud_cci MERIS+AATSR	286.15	229.81	1143.87	242.28	205.05	280.13	708.32	341.24	262.63
Cloud_cci AATSR	394.61	324.25	1590.84	385.73	273.99	389.37	704.77	464.16	366.21

Figure 4-31 *Cloud Ice Water Path from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Terra the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Terra the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

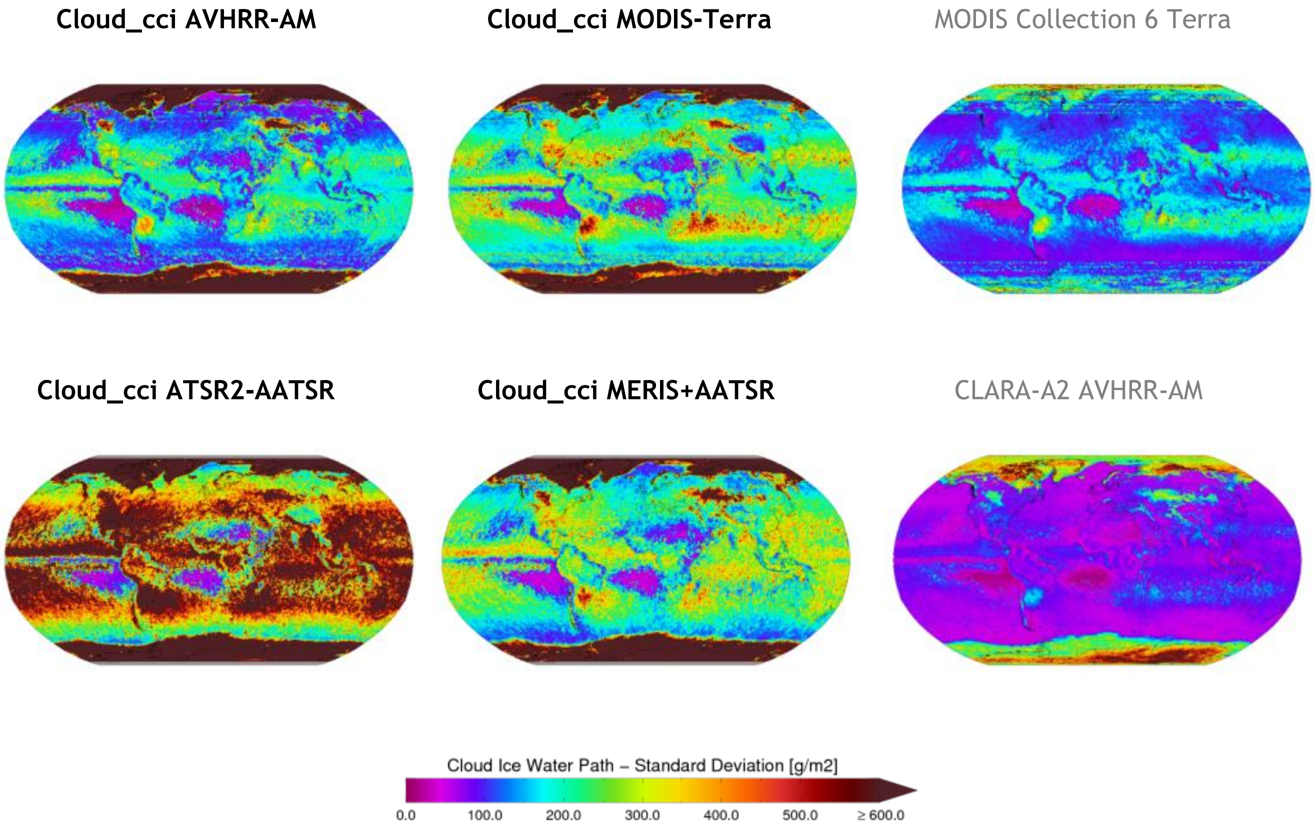


Figure 4-32 *Globally gridded means of cloud ice water path standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Terra the 2.1µm product is plotted.*

General findings

- Large deviations are found for the different datasets, although the global patterns are similar, which show low mean ice water path values for the subtropics, in particular in the stratocumulus regions, and higher values in the inner tropic and mid-latitudes. Extremely high values are found again for the Polar Regions in all datasets except in MODIS C6 Terra. The high mean IWP values in the Polar Regions are partly one order of magnitude higher than in the rest of the globe.
- Similar to LWP, the largest temporal variability for IWP occurs in regions with highest IWP and vice versa. General highest variability is seen in Cloud_cci ATSR2-AATSR, while lowest variability is seen for CLARA-A2.
- All datasets are stable in time beyond 2003 in terms of global (60S-60N) mean IWP. Highest values here are approx. 340g/m² for Cloud_cci ATSR2-AATSR, lowest values are approx. 100g/m² for CLARA-A2. Inhomogeneities are again seen for Cloud_cci AVHRR-AM for the transitions of NOAA-12 to NOAA-15 and of NOAA-15 to NOAA-17. Values for NOAA-12 and NOAA15 are very high (a factor of 3) compared to NOAA-17 and Metop-A. A small jump in IWP is also found for the transition from ATSR2 to AATSR in the end of 2002.

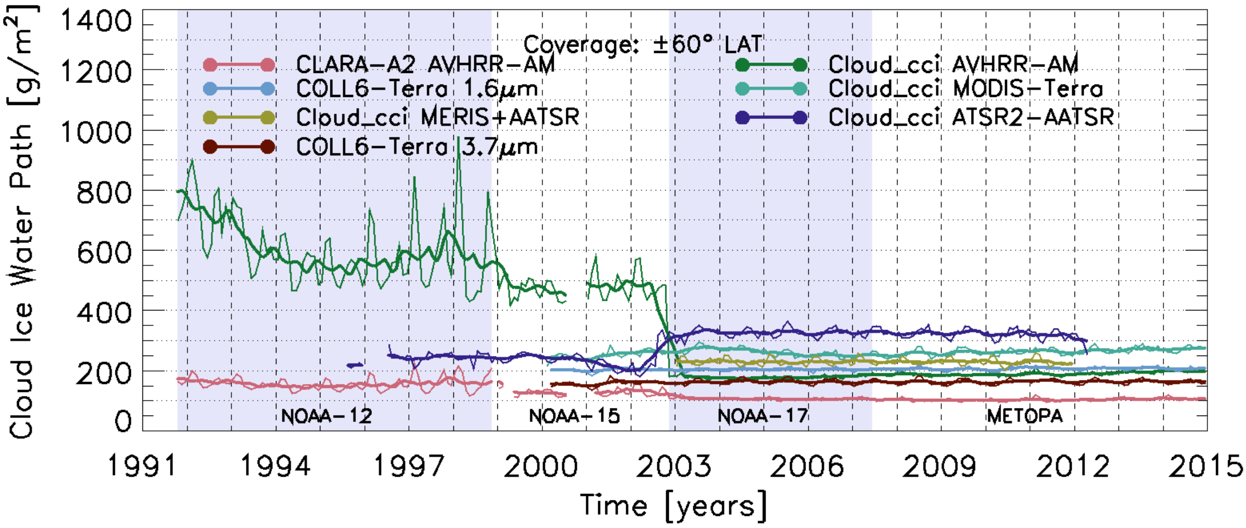


Figure 4-33 *Time series of "in-cloud" cloud ice water path for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6μm and 3.7μm products are shown.*

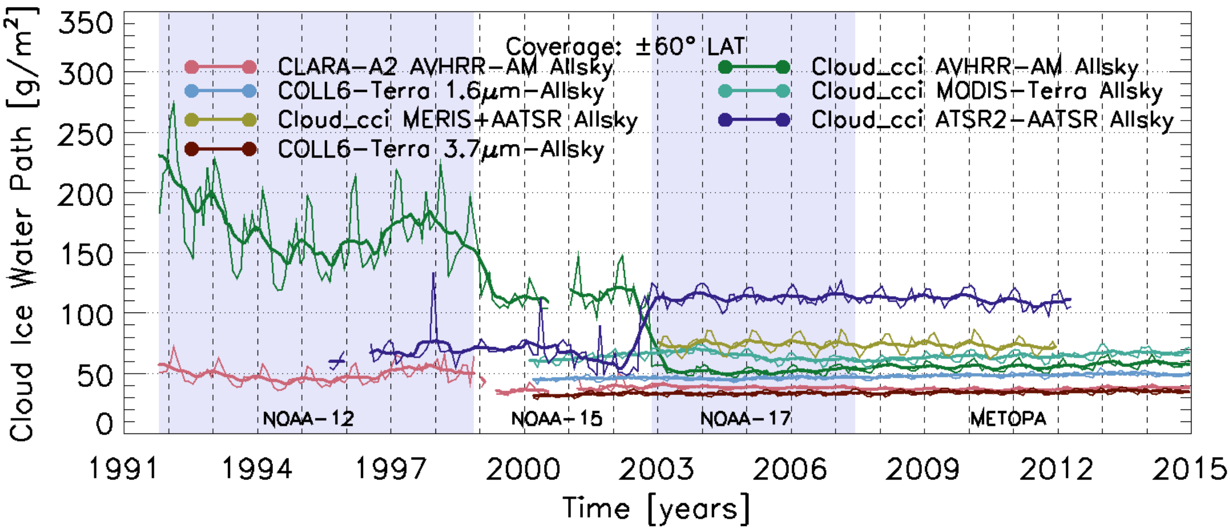


Figure 4-34 *Time series of "all-sky" cloud ice water path for all morning satellite retrievals. For MODIS Collection 6 Terra the 1.6μm and 3.7μm products are shown.*

Discussion

As mentioned in the general findings, large deviations are found among all datasets including the Cloud_cci datasets, although the general meridional structure as present in all datasets. Cloud_cci ATSR2-AATSR, and to some extend also Cloud_cci MODIS-Terra, has much higher IWP compared to the references datasets. Although, Cloud_cci IWP doesn't seem to have such high IWP, it is certainly also affected by the erroneous ice LUTs (as mentioned above for CER_{ice}) which certainly explains the high values for Cloud_cci MODIS-Terra an Cloud_cci ATSR2-AATSR. For Cloud_cci AVHRR-AM some compensating errors are likely to be present. All Cloud_cci datasets exhibit extremely high, and certainly unrealistic, values for the polar regions, basically above/below 60N/60S. This is likely to be fixed by the ice LUT bug fix and the BRDF bug fixes mentioned in the discussion for CER_{ice}.

The time series of IWP and allsky IWP reveal only little year-to-year variability for the 60S-60N mean IWP, this is common among all datasets. Cloud_cci ATSR2-AATSR shows again problems at the satellite transition; the same is seen for Cloud_cci AVHRR-AM for NOAA12 to NOAA15 and NOAA15 to NOAA17 transitions. For NOAA15 and NOAA17, Cloud_cci AVHRR-AM values are very high and unstable due to the bad illumination condition in their near-twilight orbit which is in addition drifting with time.

Table 4-10 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. The biases for the 1.6µm data (Cloud_cci AVHRR-AM after 2003 and MERIS+AATSR) are fairly small. For Cloud_cci MODIS-Terra and ATSR2-AATSR (both using the 3.7µm channel and are compared against MODIS-Terra C6 3.7µm) have very large positive biases, which is mainly due to the 3.7µm LUT bug in CC4CL as mentioned before. The bc-RMSE is between 130 and 380 g/m^2 with lowest values for Cloud_cci AVHRR-AM and Cloud_cci MODIS-Terra, and largest values for Cloud_cci ATSR2-AATSR. The strong negative trend of AVHRR-AM in the common period (2003-2011) is again caused by some NOAA-16 data, for which the 3.7µm channel was used, and for which the IWP values are significantly larger than for NOAA-17 and Metop-A, entered the calculation of the trends. Otherwise, small negative trends are found for Cloud ATSR2-AATSR and MERIS-AATSR, while a positive trends in seen in MODIS-Terra C6 1.6µm data.

Table 4-10: *Evaluation scores for Cloud_cci Level-3C Ice Water Path (morning satellites) based on comparison to MODIS Terra C6 from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	$AVHRR\text{–}AM_{1.6}$	$MODIS\text{–}Terra_{3.7}$	$ATSR2\text{–}AATSR_{3.7}$	$MERIS\text{+}AATSR_{1.6}$
Bias [g/m^2]	−22.41	+100.36	+151.44	+14.22
bc−RMSD [g/m^2]	+132.09	+136.76	+376.34	+217.38
Cloud_cci trend [g/m^2 / decade]	+14.85	−6.85	−3.72	−3.25
C6−Terra trend [g/m^2 / decade]	+5.10	+0.97	+0.90	+5.76
Δ trend [g/m^2 / decade]	+9.75	−7.82	−4.62	−9.01

4.1.9 Liquid Cloud Fraction

In this subsection the liquid cloud fraction of Cloud_cci AVHRR-AM, Cloud_cci ATSR2-AATSR, Cloud_cci MODIS-Terra, Cloud_cci MERIS+AATSR, MODIS Terra C6 and CLARA-A2 (morning satellites only) are compared by means of multi-annual mean (Figure 4-35), zonal mean (Figure 4-36) and standard deviation (Figure 4-37), all for a common time period, and time series plots (Figure 4-38). It should be noted that for MODIS-Terra C6 the “Cloud_Phase_Infrared” phase product is chosen for the comparison in this section.

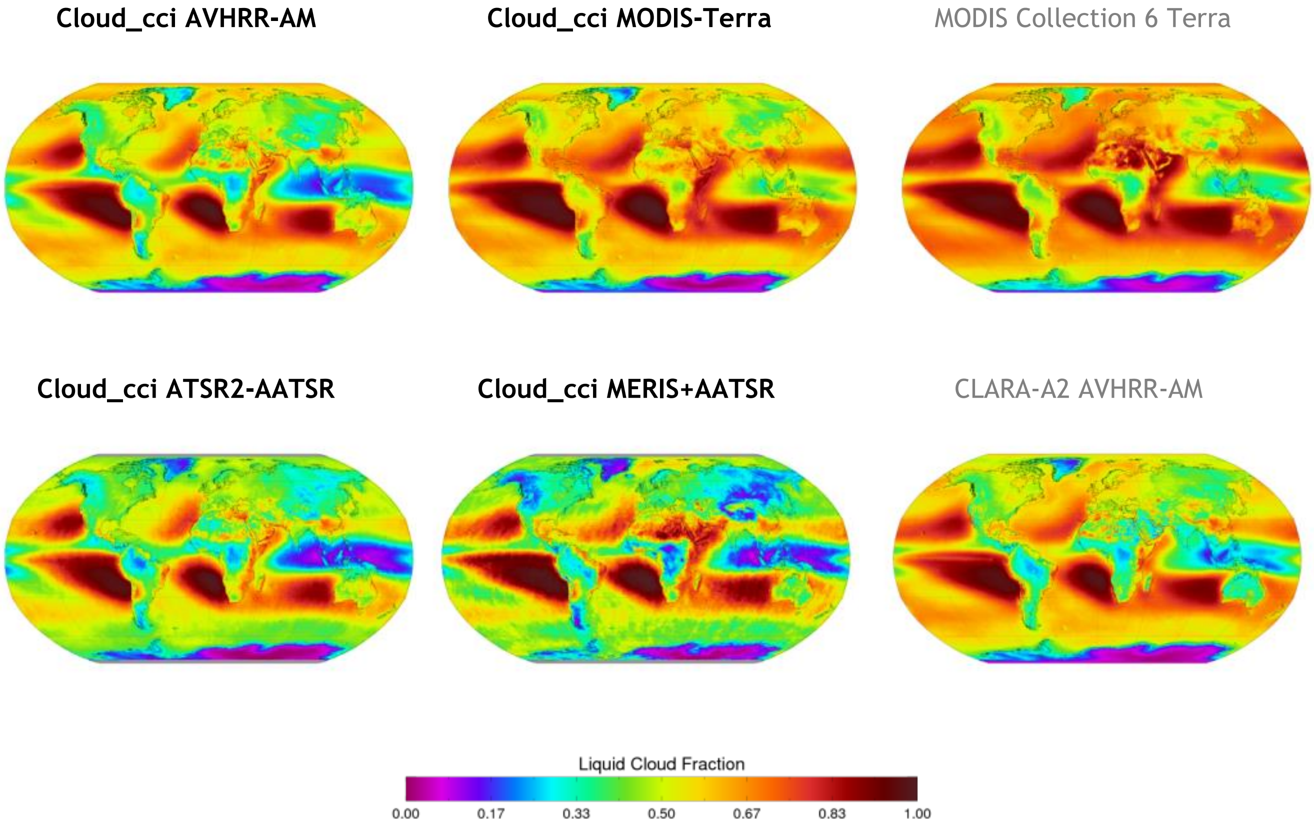
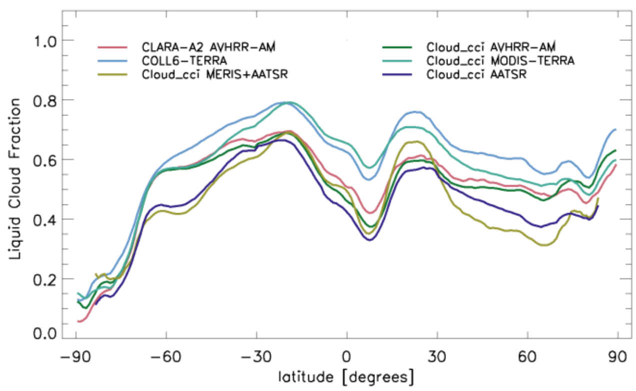


Figure 4-35 *Globally gridded means of liquid cloud fraction for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*



Latitude weighted Means of Liquid Cloud Fraction [%]

Dataset-name	Global	± 60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-AM	56.05	58.03	37.55	62.47	58.13	53.32	48.83	42.21	61.70
Cloud_cci AVHRR-AM	54.01	55.56	38.07	60.81	55.21	51.25	49.94	43.60	58.24
COLL6-TERRA	65.24	67.74	40.98	68.66	69.17	62.92	57.17	54.85	69.47
Cloud_cci MODIS-TERRA	62.93	65.69	38.19	63.57	69.41	57.65	52.00	52.71	67.08
Cloud_cci MERIS+AATSR	51.49	52.84	34.71	53.09	56.61	41.64	35.23	40.80	55.59
Cloud_cci AATSR	49.34	51.37	31.85	54.27	51.96	46.86	39.48	39.68	53.21

Figure 4-36 *Liquid Cloud Fraction from 2003 to 2011. Left: Zonal means. Right: Latitude weighted means for different regions. Highest and lowest values of each region are highlighted.*

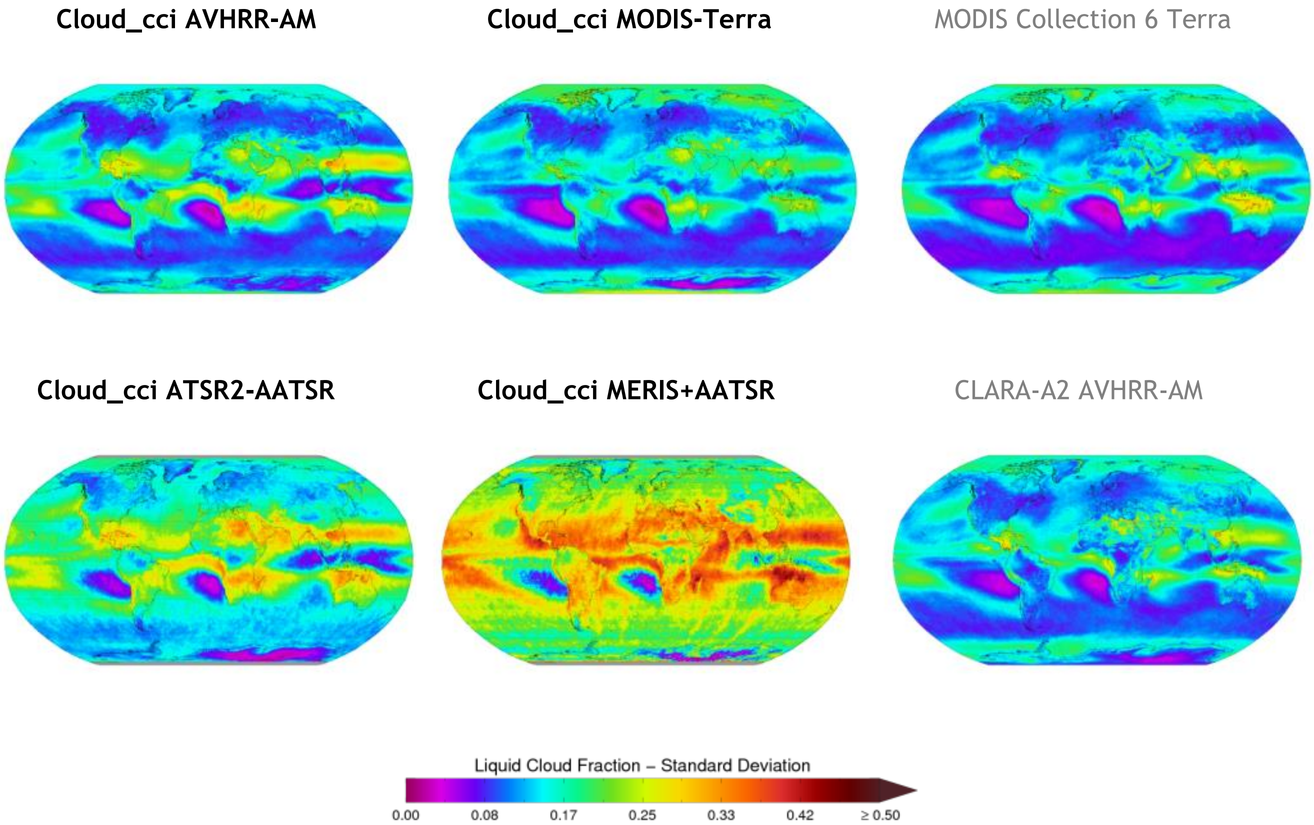


Figure 4-37 *Globally gridded means of liquid cloud fraction standard deviation for all morning satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*

General findings

- All datasets agree on the main global patterns of liquid cloud fraction (LCF): high LCF (close to 1) in the stratocumulus regions, low values in the inner tropics with a distinct local minimum in the pacific warm pool region. Another local minimum is found for the Antarctica region in all datasets with nearly no liquid cloud tops in some part. Cloud_cci MODIS-Terra and MODIS Terra C6 have more liquid clouds then the other datasets.
- The spread among the datasets is up to 20% in some latitude band, but it does not increase towards the Polar Regions as it was seen for most other cloud variables in the subsections above.
- The temporal variability of LCF is relative similar for Cloud_cci AVHRR-AM, Cloud_cci MODIS-Terra, MODIS Terra C6 and CLARA-A2. It's a bit increased for Cloud_cci ATSR2-AATSR and largest for Cloud_cci MERIS+AATSR. However, for all datasets the largest temporal variability is found in the subtropical regions and lowest for southern hemispheric stratocumulus regions.

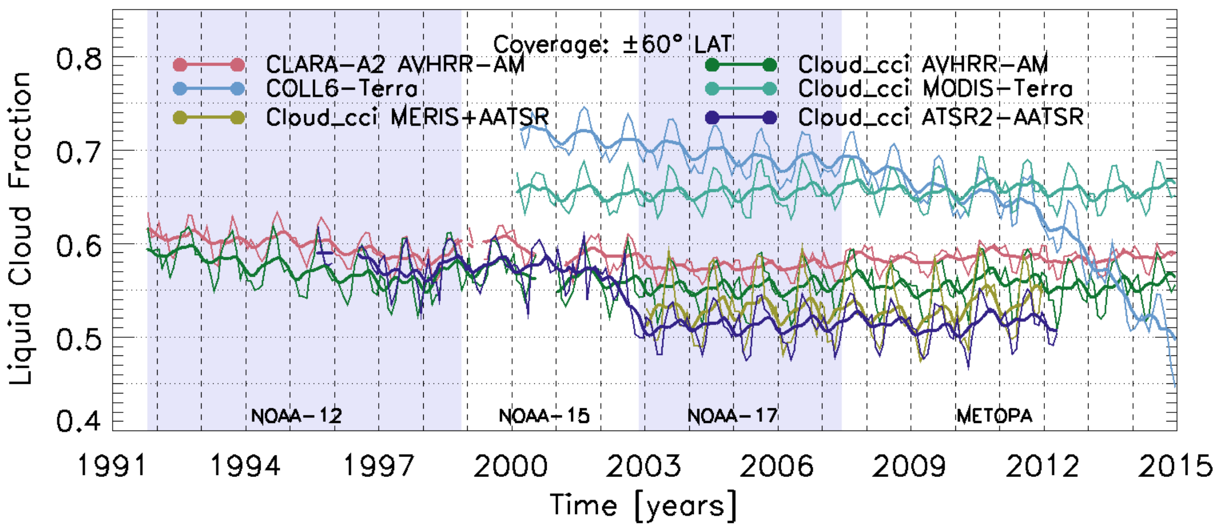


Figure 4-38 *Time series of liquid cloud fraction for all morning satellite retrievals.*

Discussion

The global patterns of LCF of the Cloud_cci datasets are very similar to the reference data. Regions with dominant liquid cloud occurrence are identified as well as regions with high occurrences of ice clouds, e.g. topical regions over the eastern Indian Ocean and western Pacific Ocean. There is a LCF of about 50% of liquid clouds over the artic regions while the liquid clouds fraction in Antarctica is very low. In particular the agreement in the latter is very convincing. Apart from the general agreement, the Cloud_cci datasets show systematic deviations among them and in comparisons to the reference datasets. Cloud_cci MODIS-Terra has relative high LCF similar to MODIS-Terra C6, which could indicate that small, low cumulus clouds might be identified with higher spatial resolution of the MODIS sensor. However, AATSR-based datasets do not show these even though being on the same spatial resolution. It is rather the opposite; Cloud_cci ATSR2-AATSR and MERIS-AATSR have lower LCF than all other datasets for most parts of the globe. A possible explanation for this is that the Pavolonis et al. (2005) scheme used for cloud typing is fine-tuned for AVHRR, which can lead to shortcoming if the channels deviate slightly. In contrast to the cloud detection, this has not been accounted for in the CC4CL and FAME-C cloud typing procedure. All Cloud_cci LCF data are relative stable in time, except a small decreasing trend seen for Cloud_cci AVHRR-AM during NOAA-12's lifetime and a jump found for Cloud_cci ATSR2-AATSR at the sensor transition in 2002. The seasonal cycle for the 60S-60N mean values agrees well among the Cloud_cci datasets and also with respect to the reference data. It should be mentioned that for MODIS Terra C6 a strong decrease in LCF with time is found while the slope of the decrease gets also larger with time. This is due to an increasing degredation of the MODIS-Terra 8.7µm channel. Cloud_cci MODIS-Terra does not use this channel, thus is not affected by the 8.7µm channel degradation, and remains relative constant with time.

Table 4-10 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Terra C6. Since MODIS-Aqua C6 has exceptionally high LCF, all Cloud_cci datasets appear to have negative biases when compared against C6. The biases are somewhat larger for Cloud_cci AVHRR-AM and ATSR2-AATSR, while being a bit smaller for MERIS-AATSR and smallest for Cloud_cci MODIS-Terra. The bc-RMSE values range from 8 to 25 %. The trends of the Cloud_cci datasets are consistently small and of positive sign, while MODIS-Terra C6 shows an enormous negative trend of 7% / decade, which is also visible in Figure 4-38 and due to the degradation for MODIS-Terra's 8.7 µm channel with time.

Table 4-11: *Evaluation scores for Cloud_cci Level-3C Liquid Cloud Fraction (morning satellites) based on comparison to C6-Terra from 2003-2011, in case of MERIS+AATSR the C6-Terra Liquid Cloud Fraction is daytime only. The scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset			
	AVHRR−AM	MODIS−Terra	ATSR2−AATSR	MERIS+AATSR
Bias [%]	−12.18	−2.05	−16.32	−8.54
bc−RMSD [%]	+10.54	+8.21	+14.64	+24.58
Cloud_cci trend [% / decade]	+1.16	+0.95	+0.77	+0.98
C6−Terra trend [% / decade]	−7.24	−7.24	−7.25	−0.70
Δ trend [% / decade]	+8.41	+8.19	+8.01	+1.68

4.2 Afternoon satellites

In this subsection the afternoon satellite datasets are compared, these are:

- Cloud_cci AVHRR-PM (see Section 2.2 for details)
- Cloud_cci MODIS-Aqua (see Section 2.2 for details)
- CLARA-A2 (only subset of afternoon satellites used; see Section A.6 for details)
- MODIS Aqua Collection 6 (see Section A.4 for details)
- PATMOS-x (only subset of afternoon satellites used; see Section A.3 for details)
- ISCCP (0300AMPM data used, see Section A.7 for details)

Figure 2-3 shows the local solar observation times of all satellites mentioned. For interpreting the following comparisons of cloud optical thickness, cloud effective radius, liquid water path and ice water path correctly it is necessary to revisit the channel settings of all sensors/datasets:

For Cloud_cci AVHRR-PM, CLARA-A2 and PATMOS-x the 3.7µm channel was used for the products compared, except for a 2-year period (approx. 2000-2002) for which the 1.6µm channel was switched on on NOAA-16/AVHRR. MODIS-Aqua has three NIR channels and MODIS Collection 6 data usually has optical property products for each of them. In our comparisons, the maps always show the 3.7µm product, while in the time series and zonal mean plots 1.6µm and 3.7µm MODIS collection 6 products are visualized if both exist. For Cloud_cci MODIS-Aqua only the 3.7µm channel was used.

Table 4-12 *Near-infrared channels utilized for afternoon satellite datasets compared in this section. This information is needed for correctly interpreting the comparison results shown for cloud optical thickness, cloud effective radius, liquid water path and cloud water path. ISCCP does not use any Near-infrared channel. Time periods given are approximated.*

Channel used	Cloud_cci AVHRR-PM	Cloud_cci MODIS-Aqua	CLARA-A2	PATMOS-x	MODIS Aqua Collection 6	ISCCP
1.6µm	2000-2002	-	2000-2002	2000-2002	(2002-2014)^{T}	-
2.1µm	-	-	-	-	(2002-2014)^{M}	-
3.7µm	1982-2000 2003-2014	2002-2014	1982-2000 2003-2014	1982-2000 2003-2014	2002-2014^{T}	-

4.2.1 Cloud Fraction

In this subsection the cloud fraction of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-39), zonal mean (Figure 4-40) and standard deviation (Figure 4-41), all for a common time period, and time series plots (Figure 4-42).

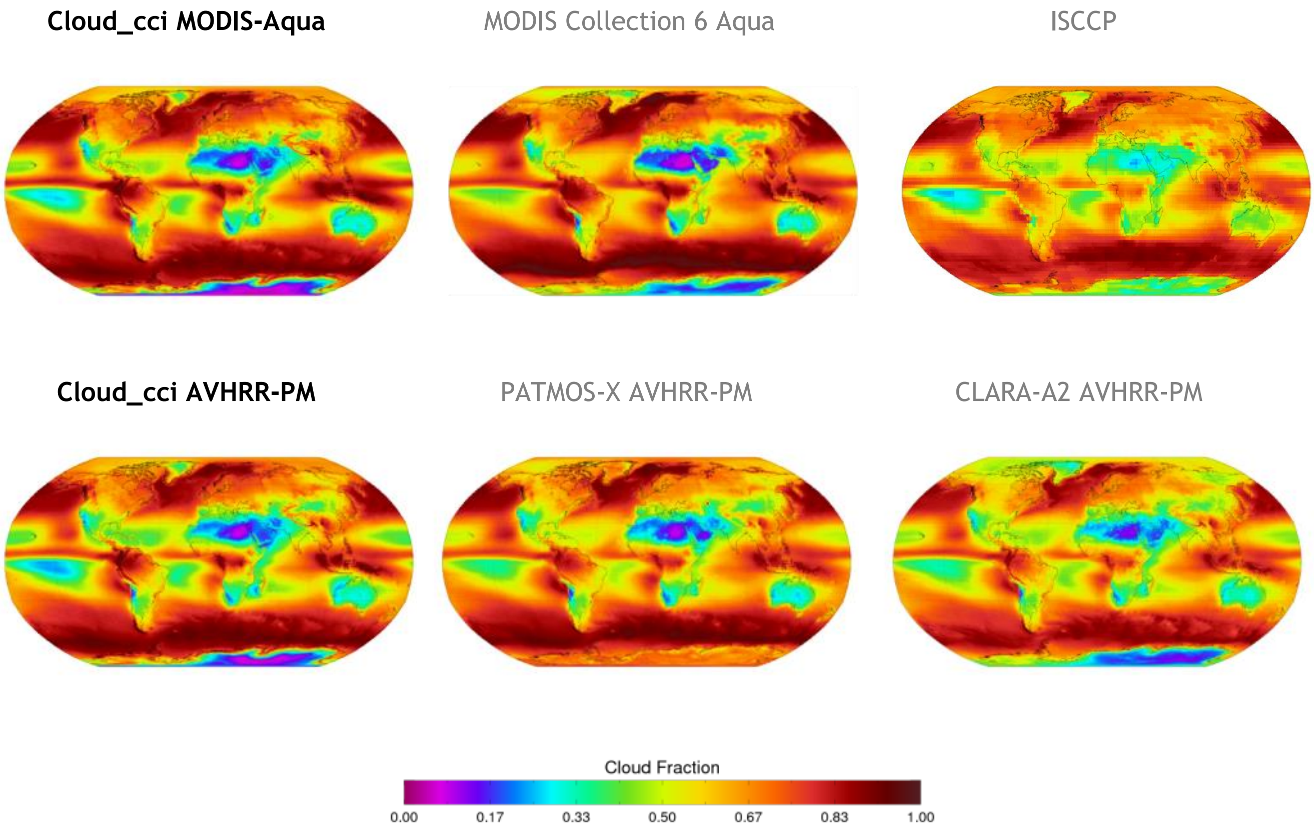
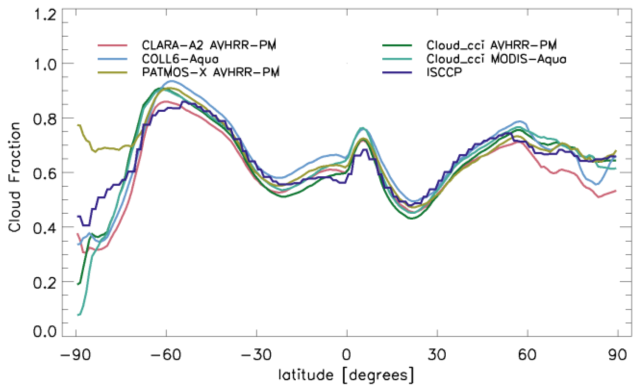


Figure 4-39 *Globally gridded means of cloud fraction for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*



Latitude weighted Means of Cloud Fraction [%]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-PM	61.90	61.98	62.07	74.42	56.71	63.93	60.63	51.22	66.26
Cloud_cci AVHRR-PM	63.04	61.90	70.94	76.78	55.08	65.66	69.80	53.74	66.81
COLL6-Aqua	67.22	67.25	66.12	81.13	61.61	68.79	67.89	55.04	72.18
Cloud_cci MODIS-Aqua	64.99	64.22	69.40	77.69	58.14	67.35	70.61	54.43	69.28
PATMOS-X AVHRR-PM	65.16	63.97	77.17	78.47	58.58	64.23	68.46	54.11	69.66
ISCCP	64.13	63.62	67.86	77.63	56.91	67.97	66.93	56.24	67.34

Figure 4-40 *Cloud Fraction from 2003 to 2011. Left: Zonal means. Right: Latitude weighted means for different regions. Highest and lowest values of each region are highlighted.*

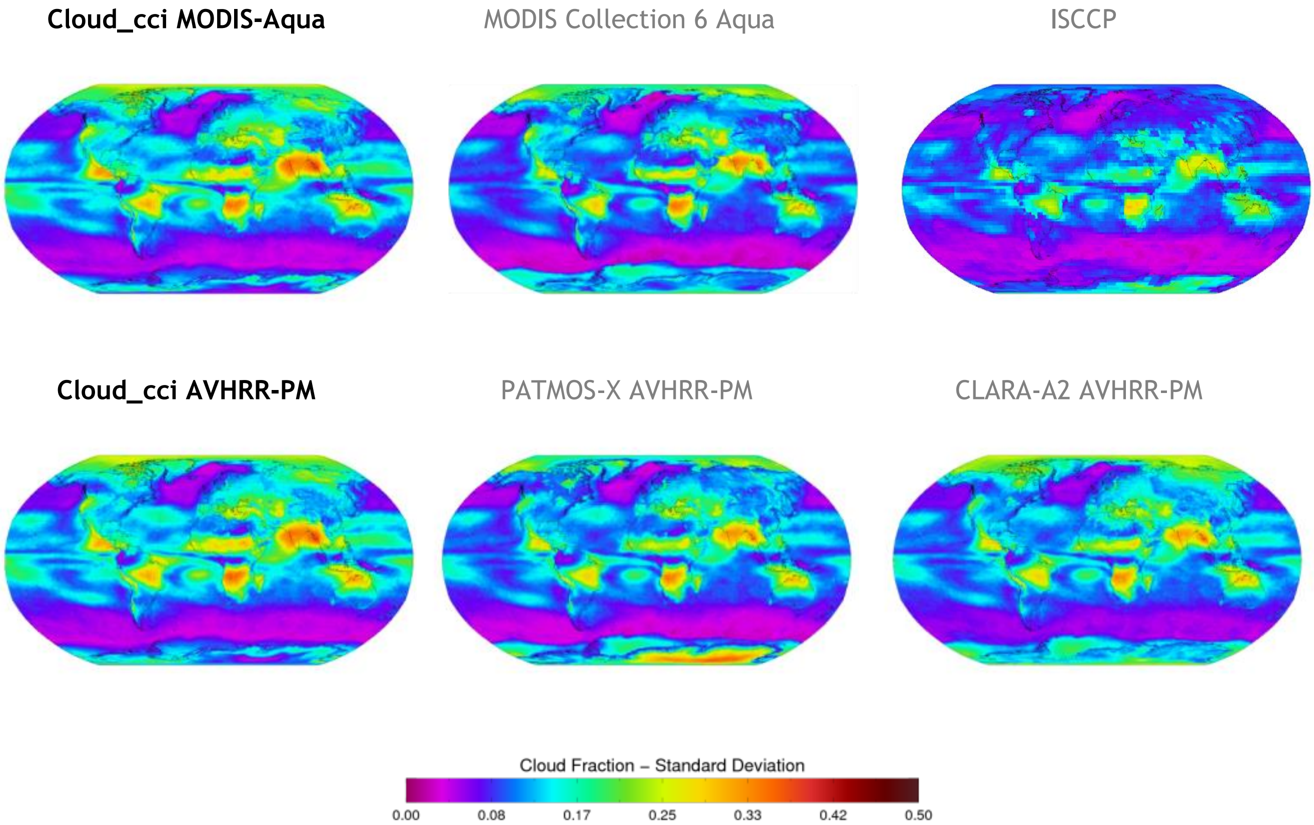


Figure 4-41 *Globally gridded means of cloud fraction standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*

General findings

- The mean CFC is very similar for all datasets with very similar spatial patterns, e.g. (1) high CFC in the tropics, oceanic stratocumulus regions and mid-latitude storm track regions, (2) low CFC in subtropical subsidence regions and Antarctica. Largest spread among the datasets is found for Polar Regions is up to 10% (Arctic) and 70% (Antarctic). On average, MODIS-Aqua C6 has highest cloud fraction and Cloud_cci AVHRR-PM lowest.
- The temporal variability of CFC is also very similar between all datasets considered, with the exception of ISCC which has lower variability than the other Highest variability is found in the outer tropics, in particular over land. Lowest variability at the inner tropic and the oceanic storm track regions of the mid-latitudes.
- While after 2003 all datasets seem relative stable, relative strong inhomogeneities are visible for PATMOS-x, CLARA-A2 and Cloud_cci AVHRR-PM at the satellite transitions and during the lifetimes of the early NOAA satellites NOAA-7 through NOAA-14. Similar issues are found for ISCCP which is additionally affected by transitions between GEO satellites used.

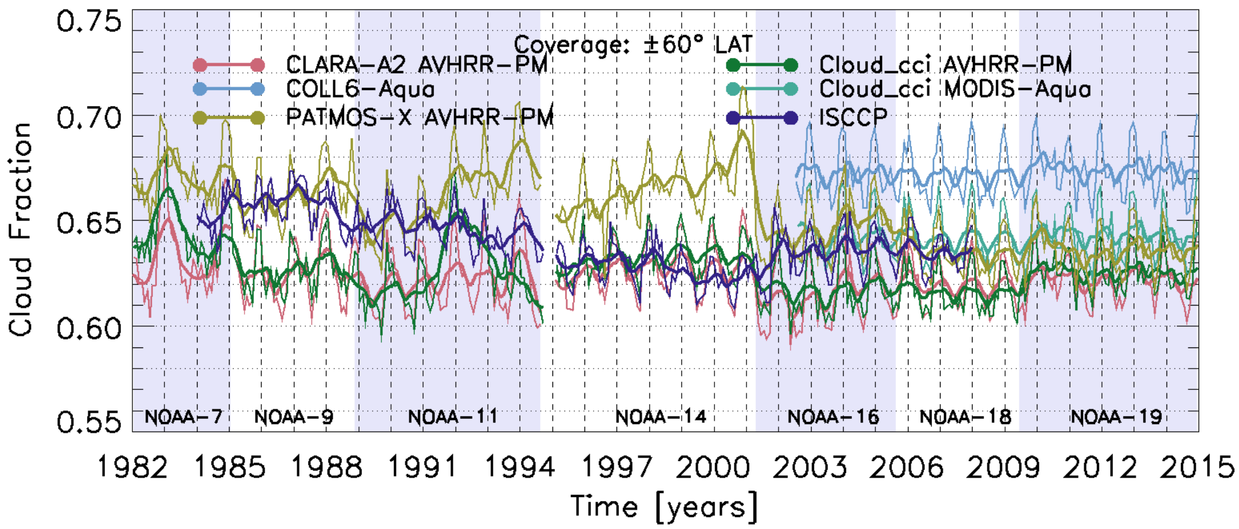


Figure 4-42 *Time series of cloud fraction for all afternoon satellite retrievals.*

Discussion

As for the morning satellite based Cloud_cci dataset, also Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua CFC agrees well with the reference datasets within a 10% range for all parts of the globe within the common period of 2003 to 2011. Only in the Polar Regions one reference datasets deviates more that the given range from the others. Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua are also relative stable after 2001, with only a small jumps of about 1% at the transition from NOAA18 to NOAA19. Before 2001 more inhomogeneities are found for Cloud_cci AVHRR-PM. These are caused by satellite drifts of the early NOAA satellite, i.e. NOAA-07, -09, -11 and -14, possibly overlaid by some error introduced by aging of the sensors. Two significant positive anomalies are found for Cloud_cci AVHRR-PM, each in a two to three year period after the severe volcano eruptions of El Chichón (1982) and Pinatubo (1991). These anomalies are not found for CLARA-A2. ISCCP also shows significant inhomogeneities at AVHRR-carrying satellite transitions, probably overlaid by changes due to transitions between the GEO satellites used.

Table 4-13 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. Cloud_cci biases are below 6% (general underestimation) and bc-RMSE values between 6 and 8 %. The trends are relative low, but for Cloud_cci AVHRR-PM still above 1%/decade mainly due to a jump from NOAA18 to NOAA19.

Table 4-13: *Evaluation scores for Cloud_cci Level-3C cloud fraction (afternoon satellites) based on comparison to C6-Aqua from 2003-2011. The scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR–PM	MODIS–Aqua
Bias [%]	–5.36	–3.04
bc–RMSD [%]	+7.24	+6.77
Cloud_cci trend [% / decade]	+1.29	+0.19
C6–Aqua trend [% / decade]	+0.34	+0.34
Δ trend [% / decade]	+0.95	–0.16

4.2.2 Cloud Top Pressure

In this subsection the cloud top pressure of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-43), zonal mean (Figure 4-44) and standard deviation (Figure 4-45), all for a common time period, and time series plots (Figure 4-46).

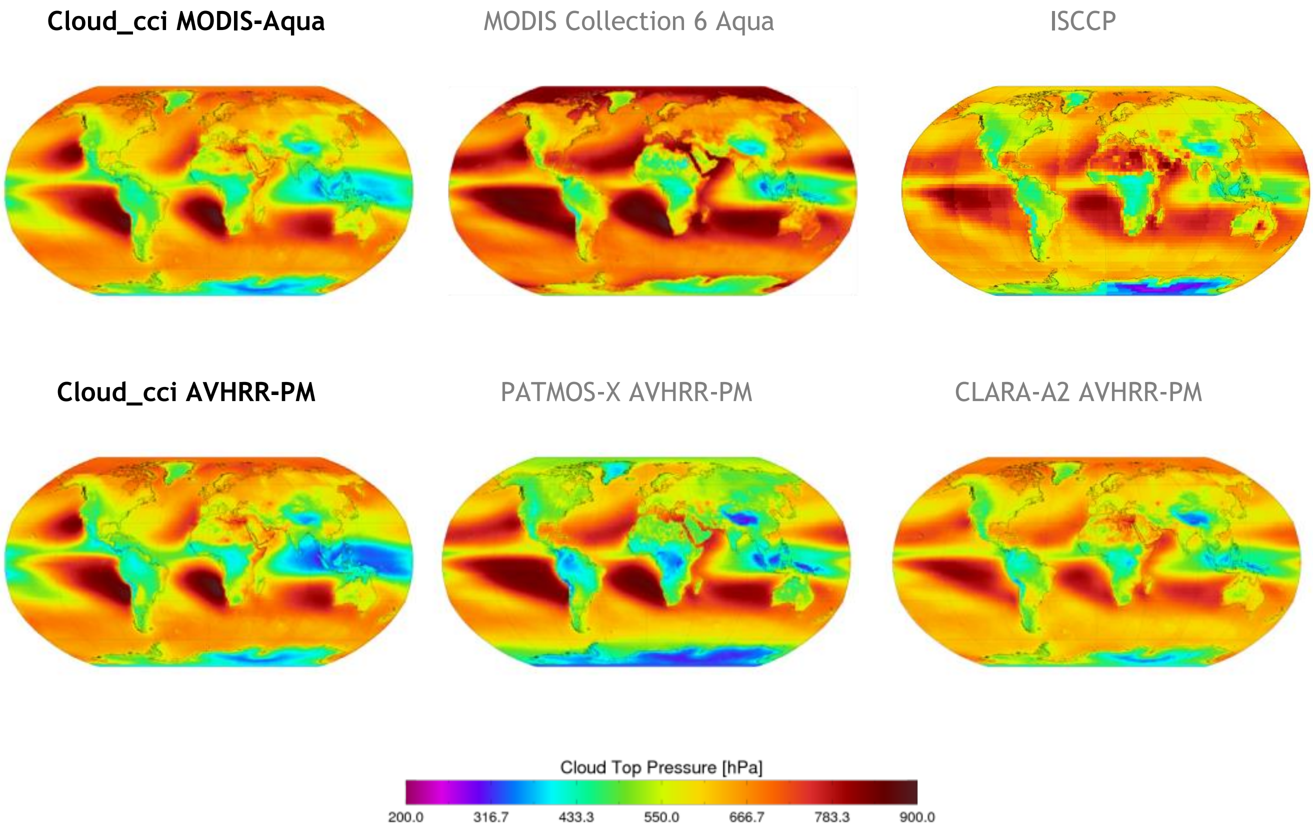
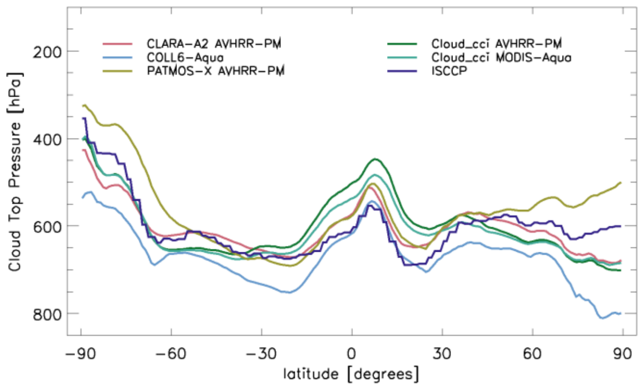


Figure 4-43 *Globally gridded means of cloud top pressure for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*



Latitude weighted Means of Cloud Top Pressure [hPa]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-PM	611.66	612.49	572.62	630.38	615.55	586.39	638.96	546.62	638.13
Cloud_cci AVHRR-PM	593.45	590.09	573.91	655.98	562.40	599.90	656.29	539.72	615.29
COLL6-Aqua	668.77	668.43	631.20	693.54	665.57	651.21	711.24	585.05	702.93
Cloud_cci MODIS-Aqua	608.97	608.06	573.41	664.92	585.97	611.58	656.17	548.34	633.62
PATMOS-X AVHRR-PM	598.66	613.33	465.64	642.16	617.25	573.87	542.35	482.66	645.95
ISCCP	624.91	631.66	557.49	641.99	641.59	594.25	605.57	548.72	656.01

Figure 4-44 *Cloud Top Pressure from 2003 to 2011. Left: Zonal means. Right: Latitude weighted means for different regions. Highest and lowest values of each region are highlighted.*

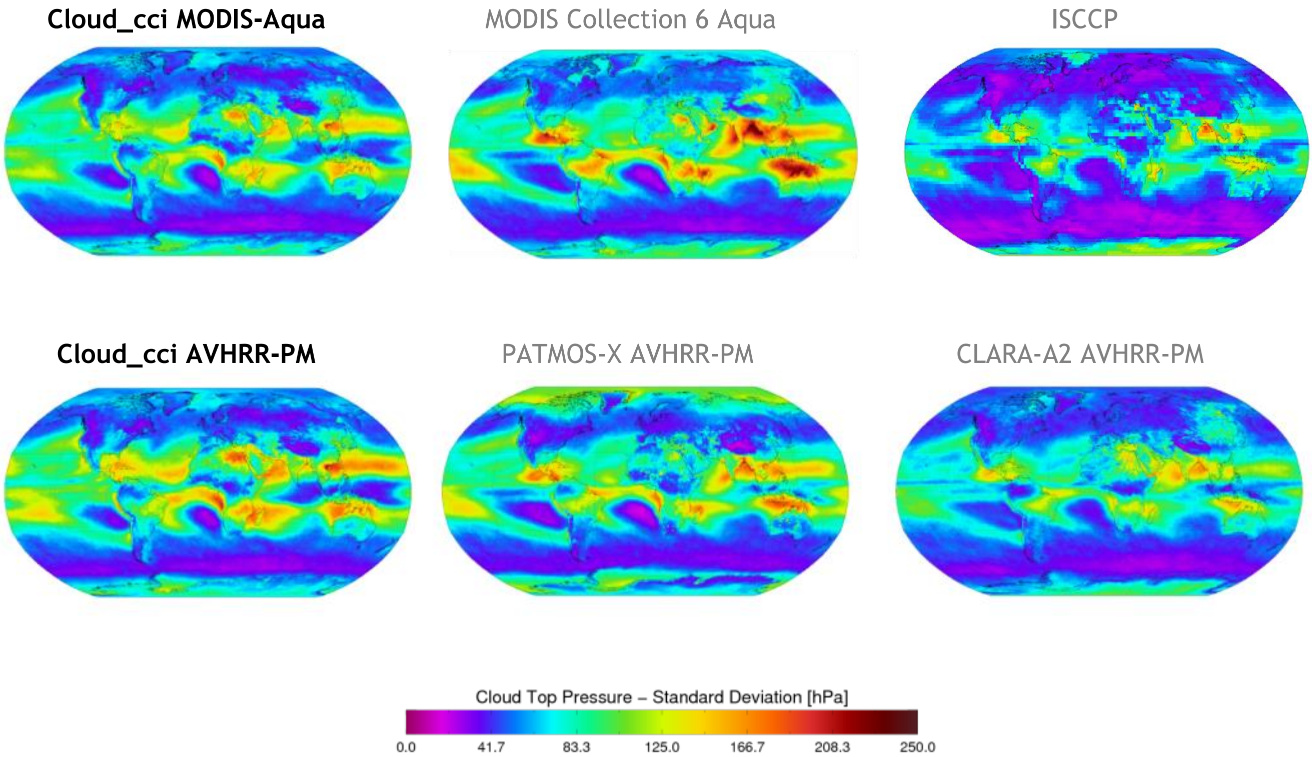


Figure 4-45 *Globally gridded means of cloud top pressure standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*

General findings

- All datasets show the same main features: (1) high mean CTP for the stratocumulus regions, (2) low mean CTP in the tropics and Antarctica, (3) mean CTP for mid-latitudes are between 600 and 750 hPa. MODIS Aqua C6 show highest CTP nearly everywhere. Spread among the datasets increases towards the Polar Regions. Mean maps for ISCCP reveal some stripes which most likely corresponds to the edges of the GEO satellites used.
- Similar temporal variability given in all datasets (MODIS C6 highest and ISCCP lowest values), with highest variability found in outer tropics and lowest variability in the mid-latitudes.
- Time series plots of global mean CTP (60S-60N) show a relative stable behaviour of all datasets with small trends found for the early NOAA satellites in PATMOS-x, CLARA-A2 and Cloud_cci AVHRR-PM, leading to jumps at the satellite transitions. In addition ISCCP exhibits a positive trend.

Table 4-14: *Relative fraction of low, mid-level, and high clouds calculated from 1-dimensional histograms for a common time period of 2003 to 2011 (excluding the Polar Regions higher/lower than +/-60°latitude). Highest values are in red, while lowest values are shown blue. * ISCCP has slightly different definitions of low/mid/high 500/700hPa instead of 440/680hPa and does not provide data for the whole period (2003-2007 only).*

Dataset	Low clouds	Mid-level clouds	High clouds
CLARA-A2 AVHRR-PM	33.54%	24.87%	41.60%
Cloud_cci AVHRR-PM	45.25%	20.40%	34.34%
Coll6 MODIS-Aqua	55.66%	10.66%	33.68%
Cloud_cci MODIS-AQUA	47.54%	21.25%	31.22%
PATMOS-x AVHRR-PM	46.57%	22.59%	30.84%
ISCCP*	50.56%	22.11%	27.33%

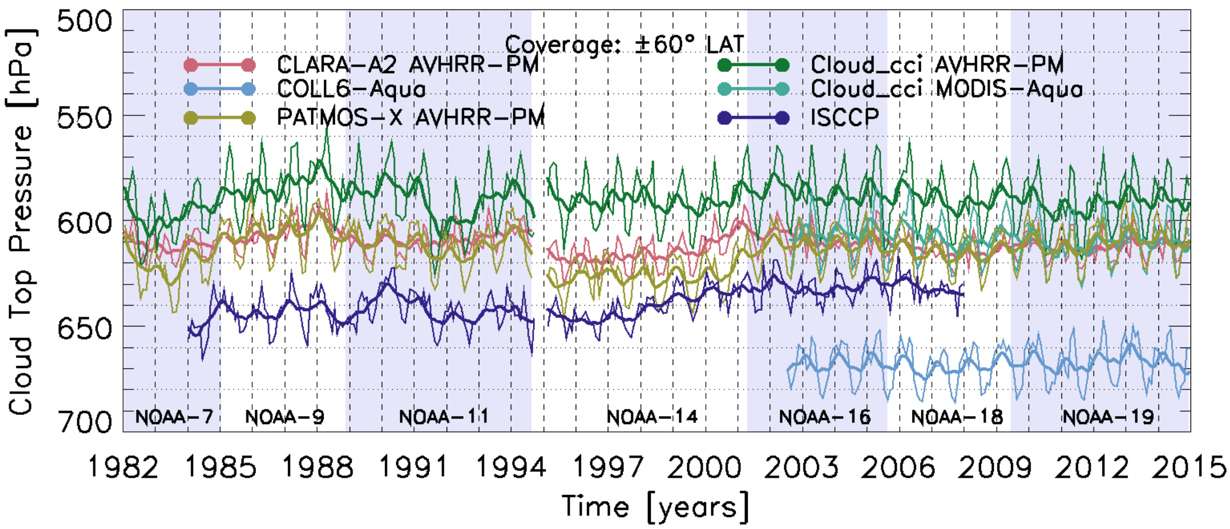


Figure 4-46 *Time series of cloud top pressure for all afternoon satellite retrievals.*

Discussion

The Cloud_cci AVHRR-PM and MODIS-Aqua CTP data reflects all major global features when compared to the reference data. In the mid latitudes and polar regions, they are more or less in the middle of the spread of all dataset. Only in the tropic the Cloud_cci dataset depict extreme values, both having the lowest CTP there on average. As mean values of CTP give only limited information, due to the bimodal underlying distribution, these comparisons have to be interpreted with caution.

Table 4-14 gives the relative fraction of low, mid- and high-level clouds, which is made by combining CTP and CFC. The relative fraction of the three layers is very similar for Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua and PATMOS-x, while CLARA-A2 has more mid-level and high clouds and MODIS-Aqua much more low-level clouds. Both Cloud_cci datasets are stable in time with two exceptions for AVHRR-PM. After the eruptions of El Chichón (1982) and Pinatubo (1991) positive CTP anomalies are found both lasting for a few years. This is most likely an artefact caused by the positive anomalies in cloud fraction found in these periods, as described in the previous subsection. Wrongly detected cloudy pixels are likely to cause high CTP retrievals, which will bias the CTP mean values.

Table **4-15** presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. There is a general underestimation of CTP in Cloud_cci dataset compared to MODIS-Aqua C6 (biases between -80 and -60hPa). The bc-RMSE values are between 77 and 91 hPa. Both Cloud_cci datasets show a clear positive bias of about +9 hPa/decade while MODIS-Aqua C6 trend is clearly smaller (+1hPa/decade) although of the same sign.

Table 4-15: *Evaluation scores for Cloud_cci Level-3C Cloud Top Pressure (afternoon satellites) based on comparison to C6- Aqua from 2003-2011, in case of MERIS+AATSR the C6-Aqua Cloud Top Pressure is daytime only. The scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR–PM	MODIS–Aqua
Bias [hPa]	–78.34	–60.38
bc–RMSD [hPa]	+90.31	+77.91
Cloud_cci trend [hPa / decade]	+9.10	+9.49
C6–Aqua trend [hPa / decade]	+1.11	+1.11
Δ trend [hPa / decade]	+7.99	+8.38

4.2.3 Cloud Optical Thickness – Liquid clouds

In this subsection the cloud optical thickness of liquid clouds (COT_{liq}) of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-47), zonal mean (Figure 4-48) and standard deviation (Figure 4-49), all for a common time period, and time series plots (Figure 4-50). The COTs of the products refer to slightly different wavelengths[1].

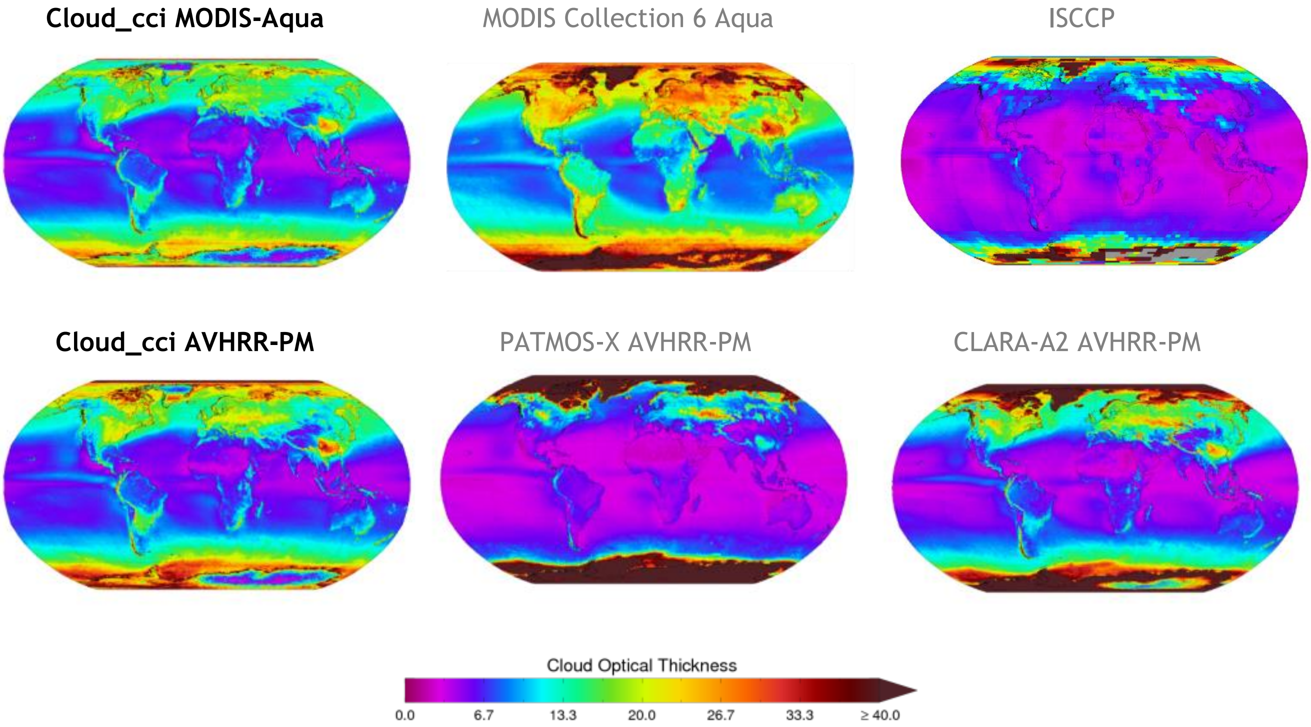
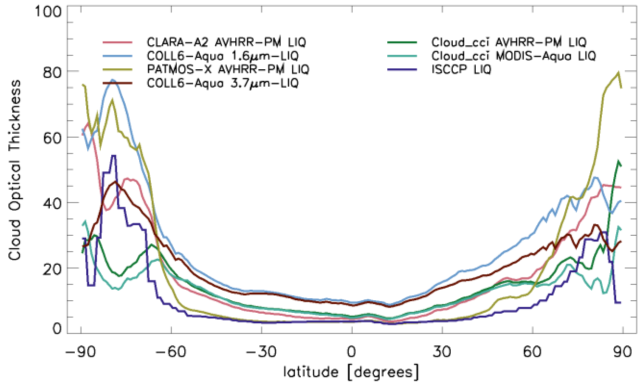


Figure 4-47 *Globally gridded means of liquid cloud optical thickness for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*



Latitude weighted Means of Liquid Cloud Optical Thickness

Dataset–name	Global	±60° Lat.	Ant–arctica	Mid–Lat. South	Tropics	Mid–Lat. North	Arctic	Land	Sea
CLARA–A2 AVHRR–PM	9.33	7.44	34.42	9.70	4.97	12.25	26.96	12.86	7.96
Cloud_cci AVHRR–PM	9.76	8.64	24.16	11.26	6.26	12.70	19.56	11.39	9.13
COLL6–Aqua	14.97	13.14	33.45	15.55	10.15	18.89	28.77	20.52	12.79
Cloud_cci MODIS–Aqua	9.22	8.37	19.66	11.25	6.00	12.11	16.40	10.86	8.58
PATMOS–X AVHRR–PM	7.14	4.71	39.44	5.53	3.51	7.30	30.84	9.70	6.15
ISCCP	4.88	4.09	16.66	4.63	3.48	5.33	14.88	5.62	4.60

Figure 4-48 *Cloud optical thickness - Liquid from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Aqua the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

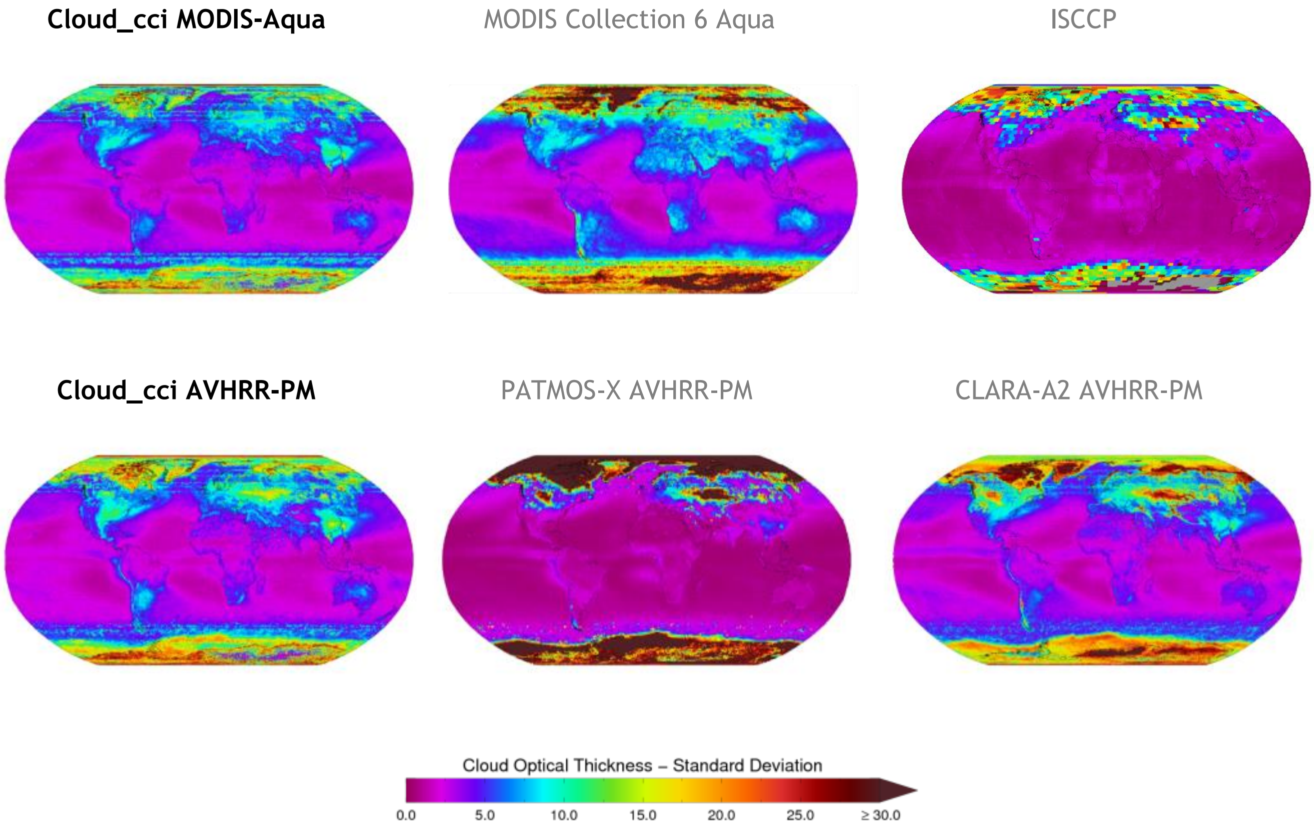


Figure 4-49 *Globally gridded means of liquid cloud optical thickness standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*

General findings

- While the main global patterns of COT_{liq} seem similar among all datasets (lower values in the tropics and subtropics, higher values in the higher latitudes), the absolute values are relatively divers. MODIS Aqua C6 COT_{liq} is highest for most parts of the globe except the polar regions in which PATMOS-x and CLARA-A2 provide highest values Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua are in between and comparable to each other. Apart from the polar regions, PATMOS-x and ISCCP show much lower mean COT_{liq} than the other datasets.
- The spread among the datasets is largest in the high latitudes. The temporal variability of COT_{liq} has similar spatial patterns as the mean values with highest variability in regions of higher mean COT_{liq}. Also, the differences in temporal variability between the datasets are similar to the difference in the mean values.
- The time series of the MODIS datasets show a very stable behaviour, however, revealing a small negative trend between 2005 and 2012. The AVHRR-based datasets (PATMOS-x, CLARA-A2 and Cloud_cci AVHRR-PM) again suffer from orbital drift, creating a drift in the mean COT_{liq} for the early satellites and jumps at the satellite transitions, which seem most pronounced for PATMOS-x.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

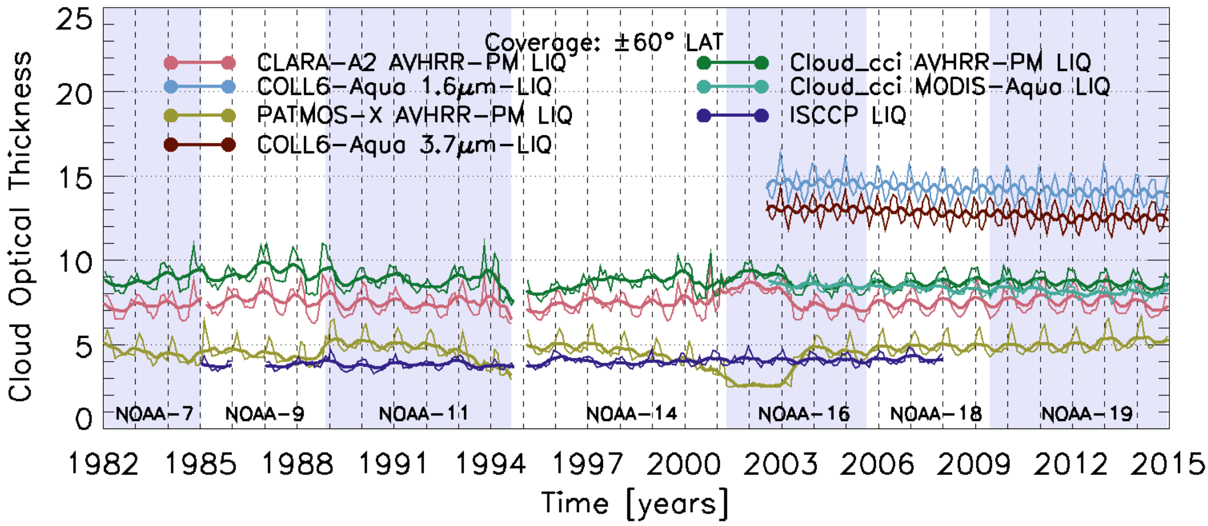


Figure 4-50 *Time series of liquid cloud optical thickness for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

Discussion

MODIS C6 COT_{liq} data (1.6 and 3.7µm) present relative high values, while ISCCP and PATMOS-X is at the lower end. The latter is mainly due to the logarithmic averaging done for PATMOS-x. When averaging logarithmically, lower values get higher weights compared to arithmetical averaging, thus the logarithmic average is significantly lower than the arithmetic average. Cloud_cci datasets AVHRR-PM and MODIS-Aqua lie in the middle and agree in most spatial features found on the globe. As the spread among all datasets increases a lot towards the poles, no judgement of the Cloud_cci quality can be made, based on these comparisons in polar regions. One can only suspect whether higher COT values for liquid clouds in polar regions are realistic or not compared to other regions. As polar regions are often characterized by snow and ice condition at the surface one cannot rule out an impact of those on COT retrievals of semi-transparent clouds if the surface conditions and/or their effects are not well captured.

The Cloud_cci AVHRR-PM time series of mean COT_{liq} in 60S-60N is relative constant over the whole time period. A slightly increase in 2001 marks the 1.6µm channel era of NOAA-16. Cloud_cci MODIS-Aqua is very close to Cloud_cci AVHRR-PM, however there is a negative trend MODIS which is also seen in all the Coll6 MODIS products. If this might be related to a sensor specific problem, e.g. visible channels degradation, or else needs further research.

Table 4-16 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua have small negative biases compared to C6. The bc-RMSE values are slightly smaller than for the morning datasets. Cloud_cci MODIS-Aqua and MODIS-Aqua C6 show a persistent, small, negative trend. Which we assume is due to the degradation of one MODIS VIS channel. The Cloud_cci AVHRR-PM datasets looks relative stable with a small positive trend.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

Table 4-16: *Evaluation scores for Cloud_cci Level-3C Optical Thickness - Liquid (afternoon satellites) based on comparison to C6-Aqua from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR−PM$_{3.7}$	MODIS−Aqua$_{3.7}$
Bias	−3.90	−4.20
bc−RMSD	+4.59	+4.33
Cloud_cci trend [/ decade]	+0.37	−0.37
C6−Aqua trend [/ decade]	−0.54	−0.63
Δ trend [/ decade]	+0.91	+0.26

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

4.2.4 Cloud Optical Thickness – Ice clouds

In this subsection the cloud optical thickness of ice clouds (COT_{ice}) of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-51), zonal mean (Figure 4-52) and standard deviation (Figure 4-53), all for a common time period, and time series plots (Figure 4-54). The COTs of the products refer to slightly different wavelengths[1].

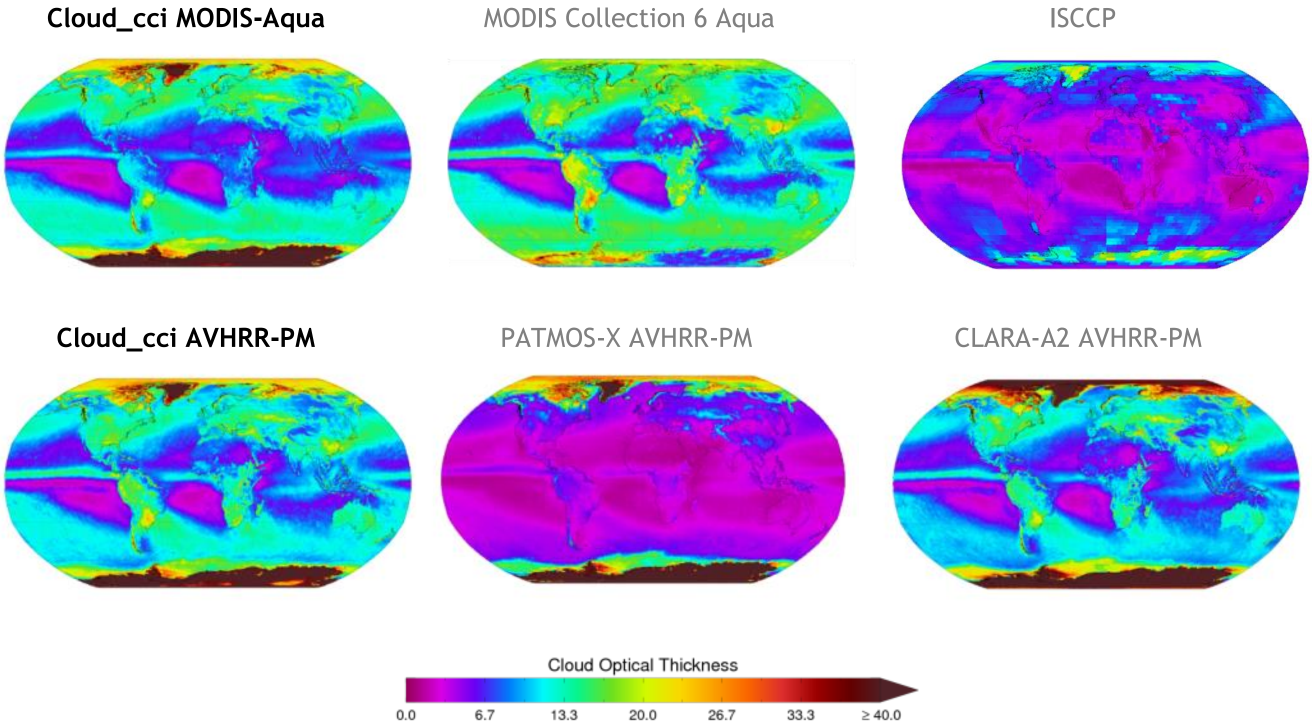
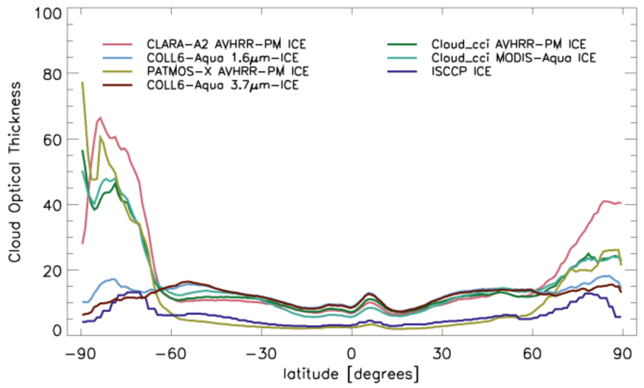


Figure 4-51 *Globally gridded means of ice cloud optical thickness for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*



Latitude weighted Means of Ice Cloud Optical Thickness

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-PM	11.04	9.45	33.28	10.55	8.18	11.63	21.63	14.78	9.54
Cloud_cci AVHRR-PM	11.00	9.99	25.33	11.49	8.67	12.09	16.90	14.02	9.80
COLL6-Aqua	11.67	11.33	14.67	13.92	9.50	13.61	14.49	13.37	11.00
Cloud_cci MODIS-Aqua	10.59	9.48	26.24	12.67	7.13	12.76	17.45	12.77	9.74
PATMOS-X AVHRR-PM	4.47	3.08	24.77	3.62	2.43	4.35	14.46	6.71	3.58
ISCCP	4.62	4.21	8.97	5.50	3.35	5.14	8.17	4.77	4.57

Figure 4-52 *Cloud optical thickness - Ice from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Aqua the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

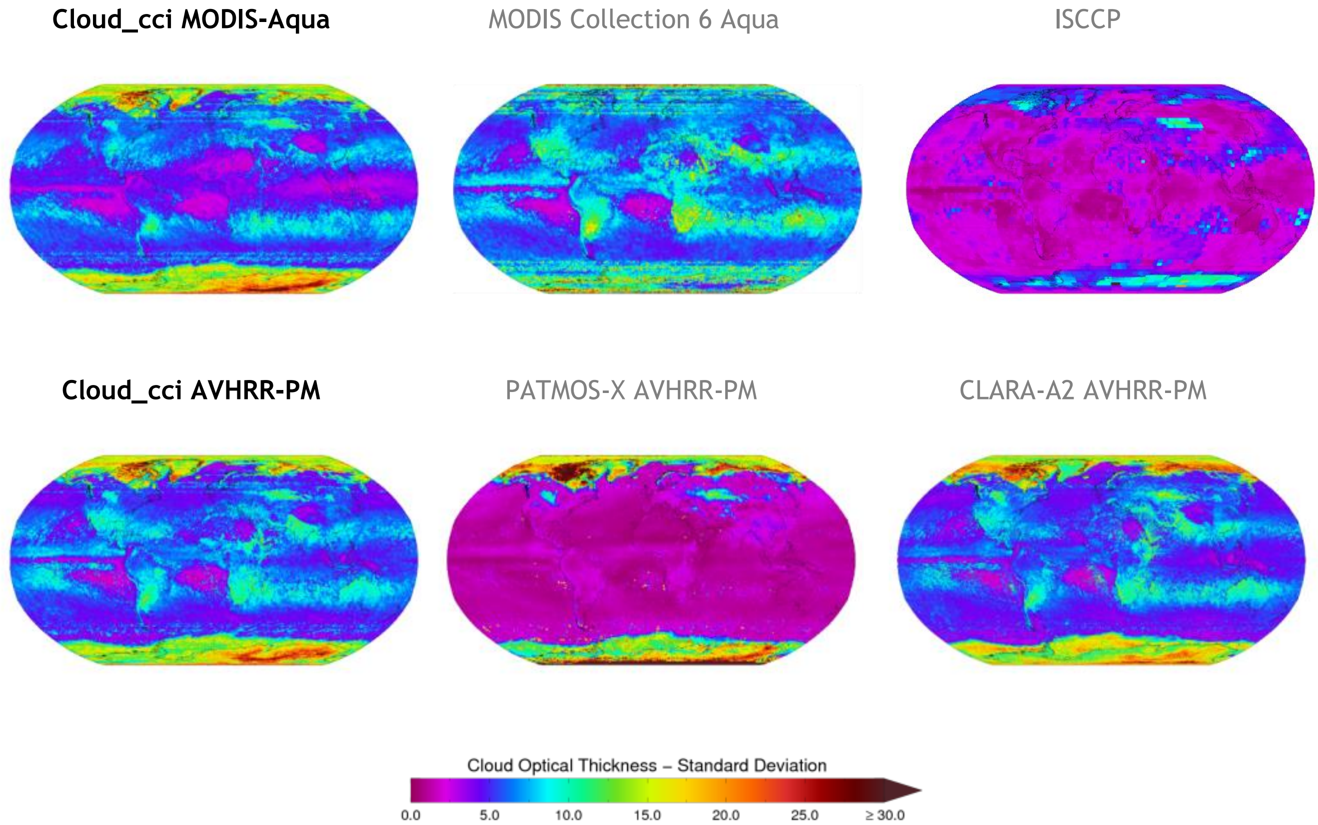


Figure 4-53 *Globally gridded means of ice cloud optical thickness standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*

General findings

- In contrast to COT_{liq}, the COT_{ice} is more similar in the datasets in consideration, except in ISCCP and PATMOS-x which has much lower mean values for most parts of the globe. In all datasets, lower values are found in the subtropics, medium values in large parts of the mid- and higher latitude, extremely high values over the snow and ice covered surfaces in the Polar Regions, where also the spread among the datasets is highest.
- Except PATMOS-x, all datasets agree on: Large temporal variability in COT_{ice} is found in the subtropics outside the stratocumulus regions in which the variability is lowest. The variability is also relative low in the mid-latitudes. In ISCCP and PATMOS-x the temporal variability is very low compared to the other datasets. Polar Regions are characterized by high variability in all datasets.
- As for COT_{liq}, the stability of the COT_{ice} time series of the AVHRR-based datasets suffer from the satellite drift of the early NOAA satellites, creating drifts and jumps for PATMOS-x, CLARA-A2 and Cloud_cci AVHRR-PM. The Cloud_cci MODIS-Aqua time series is characterized by a small decreasing trend between 2005 and 2012.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

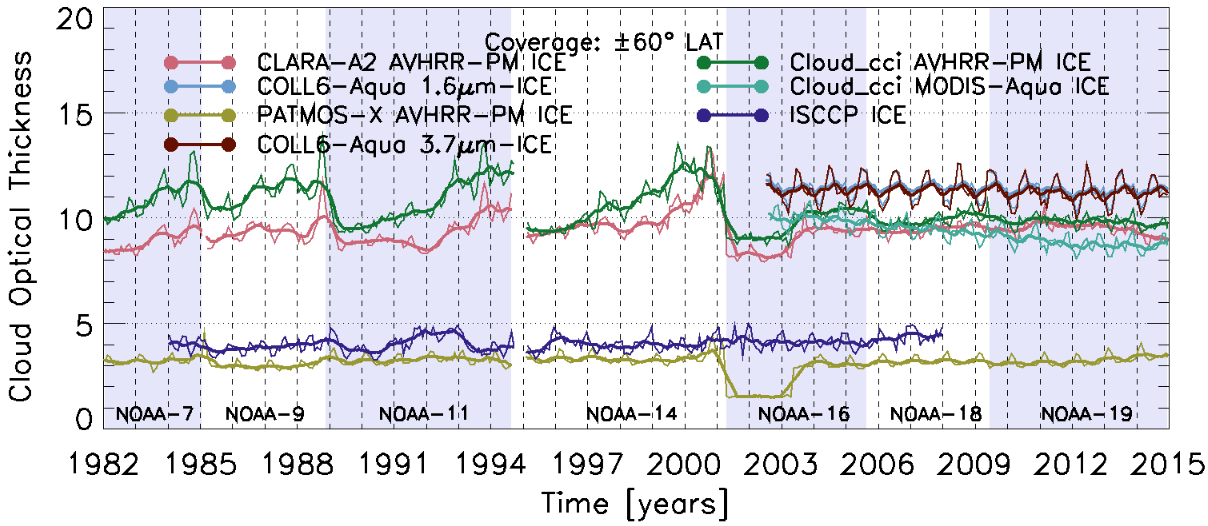


Figure 4-54 *Time series of ice cloud optical thickness for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

<u>Discussion</u>

The cloud optical thickness of ice clouds COT_{ice} is very similar for the Cloud_cci dataset AVHRR-PM and MODIS-Aqua compared to the reference data between 60S and 60N. The spread towards the polar regions is high among all datasets, with the Cloud_cci datasets being somewhere in the middle, while CLARA-shows highest and MODIS-Aqua C6 lowest COT_{ice} mean values there. Considering the average for 60S-60N, the Cloud_cci datasets are 1 to 2 optical thicknesses lower then MODIS-Aqua C6. The time series plots exhibit a strong response of the Cloud_cci AVHRR-PM mean COT_{ice} to the satellite drift of NOAA-07, -09, -11 and -14, which also leads to strong jumps in the time series at the ties of satellites transitions. Visible is also the 1.6µm period in the beginning of NOAA-16. These inhomogeneities due to the satellite drifts appear a bit smaller for PATMOS-x compared to Cloud_cci-PM and CLARA-A2, even though this dataset is also AVHRR-based. The reason for this is not entirely known yet, however, the generally smaller absolute values might attenuate the absolute values of the inhomogeneities. Cloud_cci MODIS-Aqua has a small but constant decreasing trend until 2012, which can most likely be attributed to a degradation of the used MODIS visible channel.

Table 4-17 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. Biases for Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua wrt. to MODIS-Aqua C6 are very similar (slightly negative). The bc-RMSE values are also nearly identical (values around 5 optical thicknesses). While Cloud_cci AVHRR-PM has almost no decadal trend, same for MODIS-Aqua C6, there is a clear negative trend in Cloud_cci AVHRR-PM of -1.3 / decade, which is due to a MODIS channel degradation as mentioned above.

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

Table 4-17: *Evaluation scores for Cloud_cci Level-3C Optical Thickness - Ice (afternoon satellites) based on comparison to C6- Aqua from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR−PM$_{3.7}$	MODIS−Aqua$_{3.7}$
Bias	−1.24	−1.64
bc−RMSD	+4.84	+5.54
Cloud_cci trend [/ decade]	−0.09	−1.31
C6−Aqua trend [/ decade]	−0.03	−0.07
Δ trend [/ decade]	−0.06	−1.24

[1]Wavelengths the COT of the individual datasets refer to: Cloud_cci: 0.55µm, CLARA-A2: 0.6µm, MODIS C6: 0.65µm (assumed to be the same as for C5), ISCCP: unknown. The impact of the representative wave length within 0.5-0.7µm on COT is assumed to be very minor.

4.2.5 Cloud Effective Radius – Liquid clouds

In this subsection the cloud effective radius of liquid clouds (CER_{liq}) of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-55), zonal mean (Figure 4-56) and standard deviation (Figure 4-57), all for a common time period, and time series plots (Figure 4-58). ISCCP is only included for the time series plots as no ISCCP CER_{liq} data is available after the year 2000.

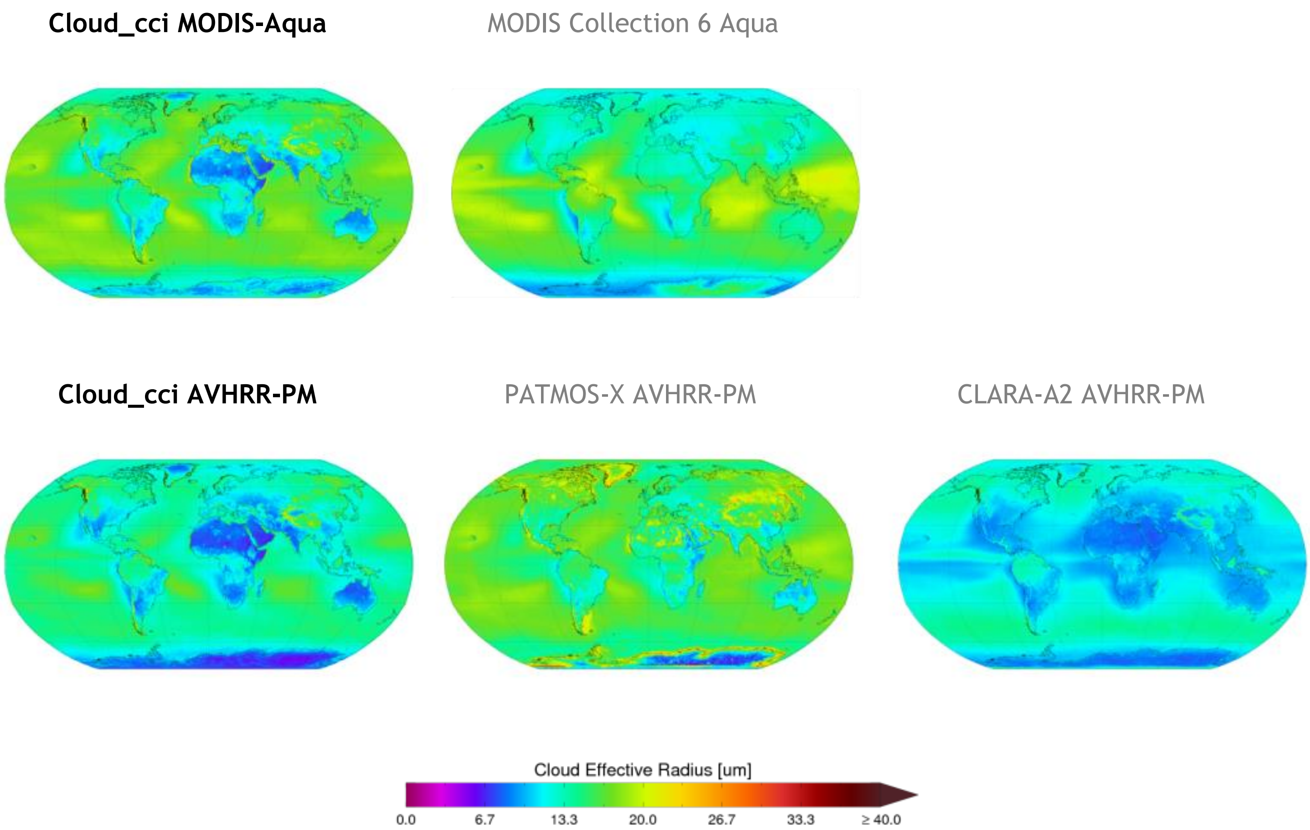
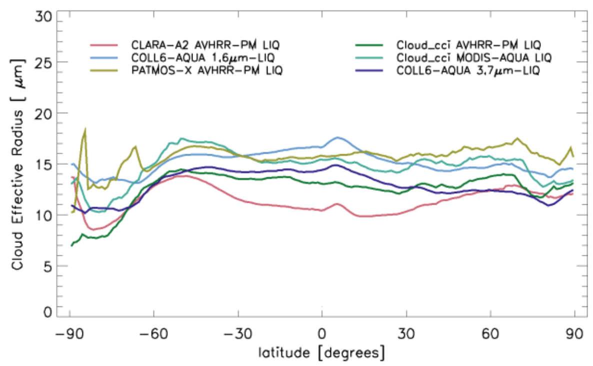


Figure 4-55 *Globally gridded means of liquid cloud effective radius for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*



Latitude weighted Means of Liquid Cloud Effective Radius [micron]

Dataset−name	Global	± 60° Lat.	Ant−arctica	Mid−Lat. South	Tropics	Mid−Lat. North	Arctic	Land	Sea
CLARA−A2 AVHRR−PM	11.27	11.22	11.01	12.97	10.56	11.37	12.49	10.55	11.55
Cloud_cci AVHRR−PM	13.16	13.24	11.15	14.01	13.07	12.96	13.28	11.22	13.91
COLL6−AQUA	15.25	15.60	11.51	15.37	16.38	13.70	12.86	13.61	15.90
Cloud_cci MODIS−AQUA	15.31	15.44	13.02	16.80	15.05	15.21	14.87	12.43	16.44
PATMOS−X AVHRR−PM	15.90	15.89	15.12	16.28	15.69	16.09	16.62	14.64	16.39

Figure 4-56 *Cloud effective radius - Liquid from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Aqua the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

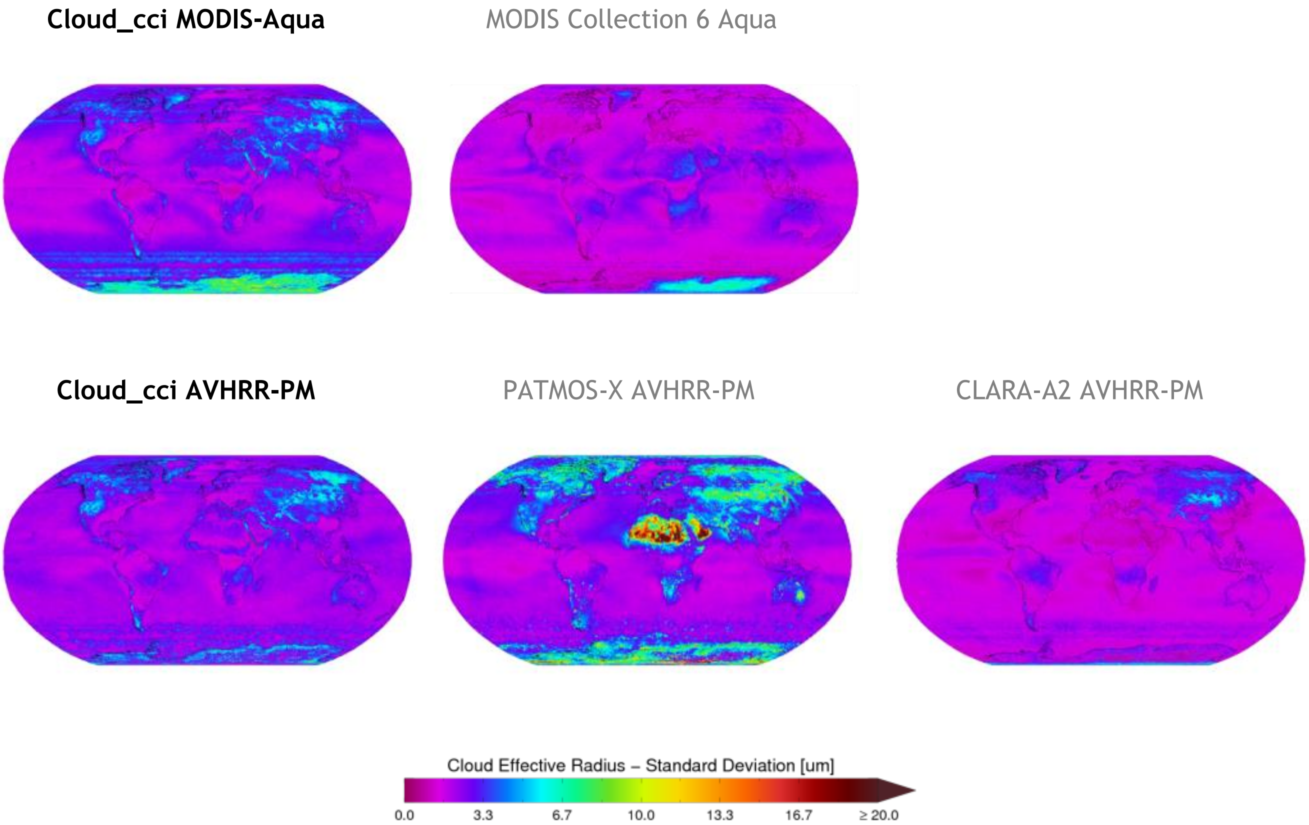


Figure 4-57 *Globally gridded means of liquid cloud effective radius standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*

General findings

- The mean cloud effective radius of liquid clouds varies a bit between all datasets with CLARA-A2 CER_{liq} being at the lower end and highest values found for PATMOS-x and Cloud_cci MODIS-Aqua for most parts of the globe. The spread among the datasets in the Polar Regions is for CER_{liq} comparable to the rest of the globe.
- The temporal variability is general very low for all datasets, except over Antarctica and some land regions of which the Sahara regions clearly sticks out for PATMOS-x CER_{liq}.
- The time series of the AVHRR-based datasets (and ISCCP) reveal again trends in the course of the lifetime of the early NOAA satellites as well as in homogeneities at satellite transitions resulting from these.

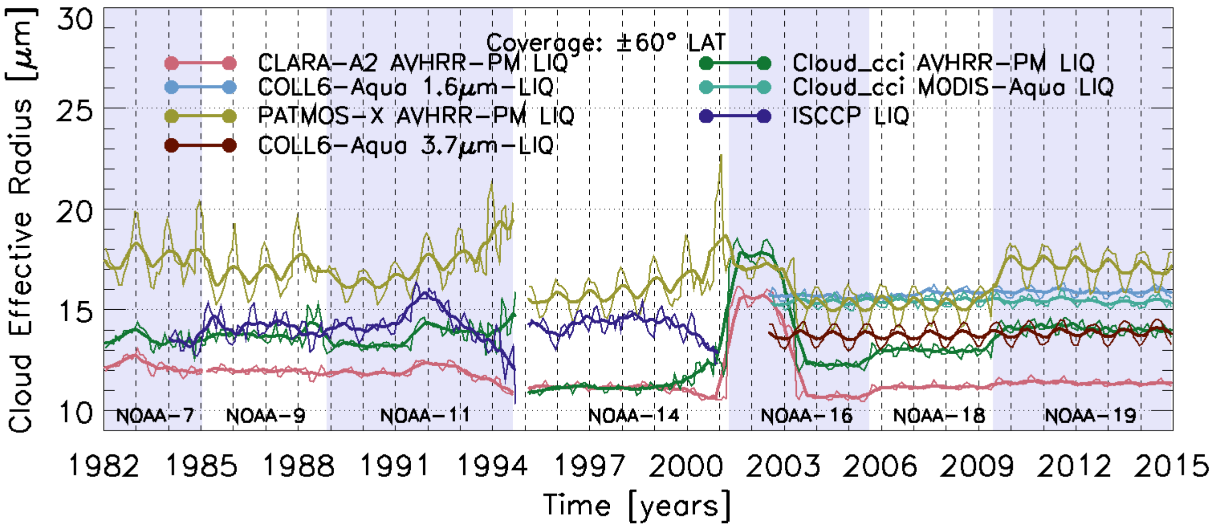


Figure 4-58 *Time series of liquid cloud effective radius for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6μm and 3.7μm products are shown.*

Discussion

The spread in CER_{liq} among all datasets is relative large (about 50% of the mean of all datasets) throughout all latitude bands. Thus, the Cloud_cci datasets are difficult to judge based on these comparisons. One can at least conclude that the relative latitudinal dependence of the mean CER_{liq} in Cloud_cci AVHRR-PM and MODIS-Aqua are similar to the reference data. For another feature one can also find an agreement between all datasets: the lower CER_{liq} in the stratocumulus regions, especially near the coasts, compared to the outer regions of the stratocumulus further westwards and partly already trade cumulus regions. The Cloud_cci data show a rather strong land-sea contrast in mean CER_{liq} (lower values over land compared to sea) which cannot be found in the reference datasets.

The time series of Cloud_cci AVHRR-PM shows some peculiarities: relative constant values for certain periods, but significant jumps at the times of satellite transitions. The low values for NOAA-14 are very suspicious. The higher values at the beginning for NOAA-16 are most likely due to the 1.6μm channel being available and used, compared to the 3.7μm channel available and used at all other times.

Table 4-18 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. Compared to the 3.7μm CER_{liq} of MODIS-Aqua C6, Cloud_cci AVHRR-PM has a small negative bias and Cloud_cci MODIS-Aqua a small positive bias. The bc-RMSE values are also low with values of 2.2 and 2.3μm, respectively. Cloud_cci AVHRR-PM does show a clear positive trend of 1.7μm / decade which is not confirmed by both MODIS-Aqua datasets. The time series plots indicate that sudden increases (jumps) from NOAA-16 to NOAA-18 and from NOAA-18 to NOAA-19 are responsible for the trend found.

Table 4-18: *Evaluation scores for Cloud_cci Level-3C Effective Radius - Liquid (afternoon satellites) based on comparison to C6- Aqua from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR–PM$_{3.7}$	MODIS–Aqua$_{3.7}$
Bias [µm]	–0.50	+1.71
bc–RMSD [µm]	+2.34	+2.21
Cloud_cci trend [µm / decade]	+1.71	+0.13
C6–Aqua trend [µm / decade]	+0.18	+0.18
Δ trend [µm / decade]	+1.53	–0.04

4.2.6 Cloud Effective Radius – Ice clouds

In this subsection the cloud effective radius of ice clouds (CER_{ice}) of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-59), zonal mean (Figure 4-60) and standard deviation (Figure 4-61), all for a common time period, and time series plots (Figure 4-62). ISCCP is only included for the time series plots as no ISCCP CER_{ice} data is available after the year 2000.

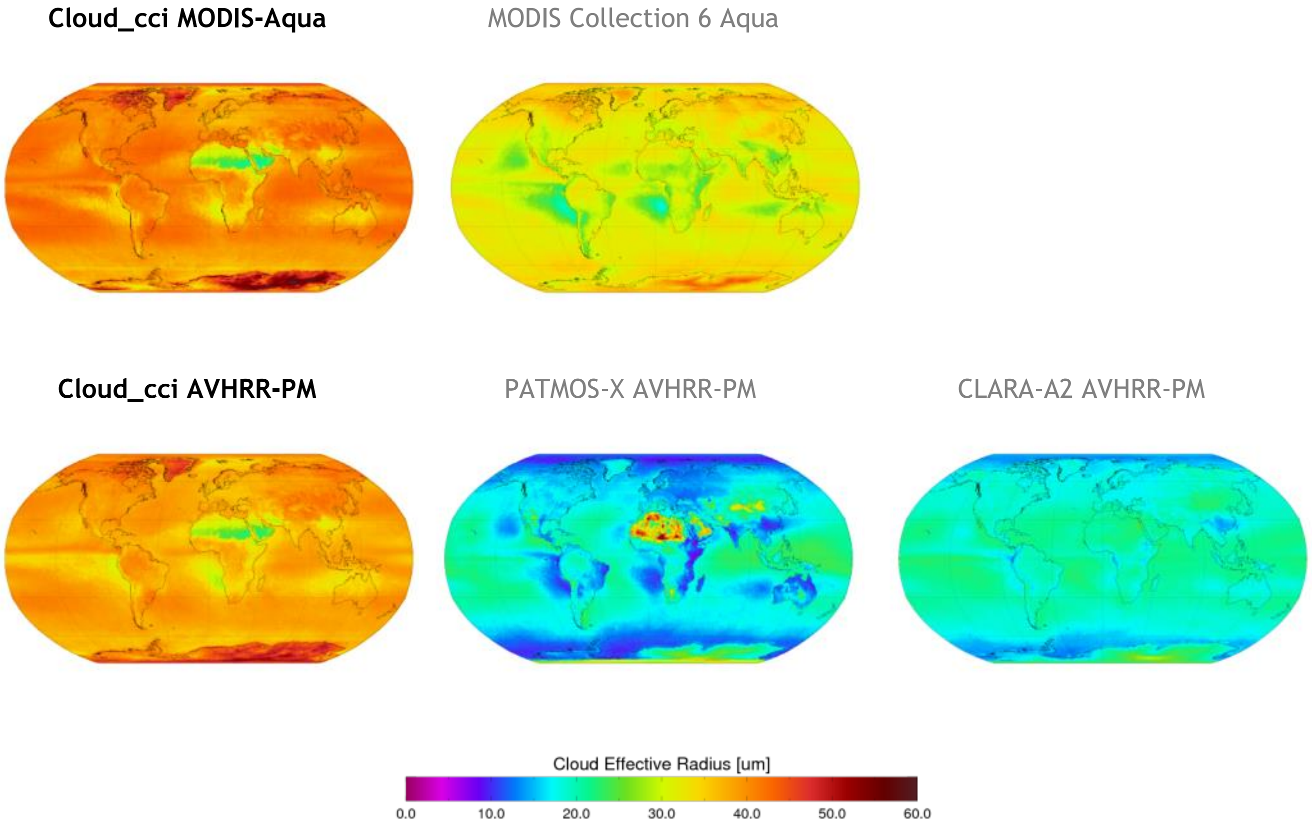
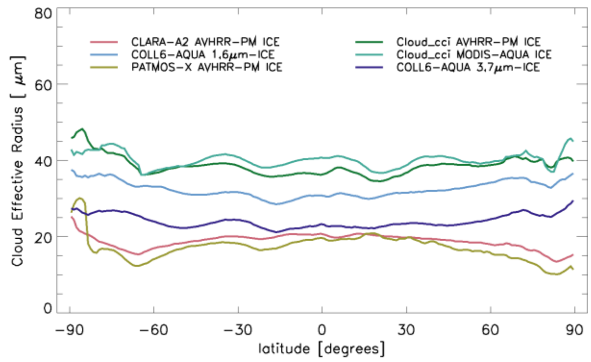


Figure 4-59 *Globally gridded means of ice cloud effective radius for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*



Latitude weighted Means of Ice Cloud Effective Radius [micron]

Dataset–name	Global	±60° Lat.	Ant–arctica	Mid–Lat. South	Tropics	Mid–Lat. North	Arctic	Land	Sea
CLARA–A2 AVHRR–PM	19.45	19.69	16.78	19.05	20.17	19.10	17.20	19.05	19.61
Cloud_cci AVHRR–PM	37.58	37.38	39.20	38.58	36.48	38.62	39.86	37.24	37.71
COLL6–AQUA	31.51	31.15	34.29	32.13	30.37	32.25	34.91	31.36	31.58
Cloud_cci MODIS–AQUA	39.69	39.63	39.98	39.91	39.28	40.29	40.56	38.23	40.26
PATMOS–X AVHRR–PM	17.96	18.32	14.49	17.50	18.80	17.74	14.05	17.93	17.98

Figure 4-60 *Cloud effective radius - Ice from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Aqua the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

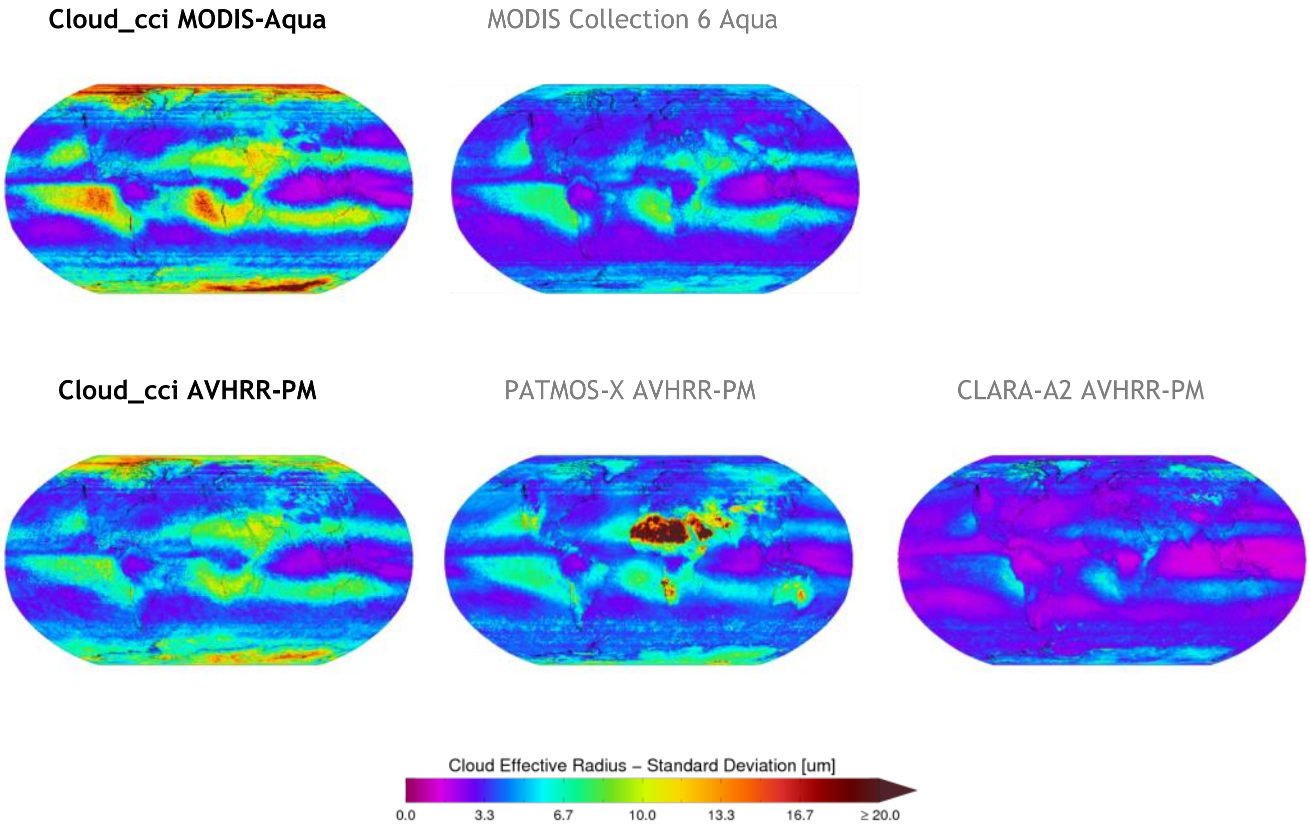


Figure 4-61 *Globally gridded means of ice cloud effective radius standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*

General findings

- Very high mean CER_{ice} are found for the Cloud_cci datasets MODIS-Aqua and AVHRR-PM, while comparable low values are found for PATMOS-x and CLARA-A2. No strong latitudinal dependence is found for mean CER_{ice}.
- Highest temporal variabilities of CER_{ice} are found in the subtropics, especially in the stratocumulus regions, which are characterizes by very low ice clouds frequencies of occurrences. The temporal variability is, as the mean values, highest for Cloud_cci MODIS-Aqua and Cloud_cci AVHRR-PM.
- The MODIS-based dataset are stable in time, although the large deviation among them remains throughout the MODIS era. AVHRR-based datasets show again jumps at the satellite transitions of the early, significantly drifting NOAA satellite. Another feature is found in the period of 2000 to 2003 for which the 1.6µm channel was switched on and used from NOAA-16, while the rest of the time period always the 3.7µm was switched on and thus used.

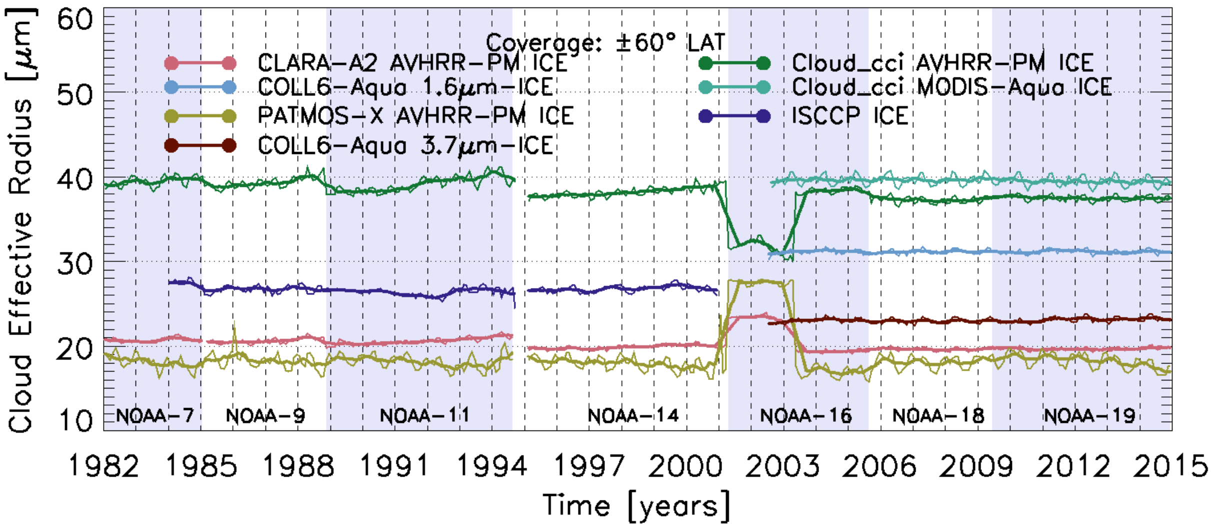


Figure 4-62 *Time series of ice cloud effective radius for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

Discussion

The spread in CER_{ice} among the datasets is even larger than for CER_{liq} which prevents drawing too many conclusions from this intercomparison. On the other hand, as already mentioned when comparing the morning satellite datasets, the CC4CL scheme used wrong LUTs for ice cloud properties, which led to a very strong overestimation of CER_{ice}. This is also visible in this section for Cloud_cci AVHRR-PM and MODIS-Aqua showing mean CER_{ice} values of about 40 µm. Keeping this in mind, there is only little confidence in the quality of CER_{ice} as currently included. The corresponding bug has been fixed in the meantime which will lead to CER_{ice} of much better quality in Cloud_cci AVHRR-PM and MODIS-Aqua dataset version v3.0

Table 4-19 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. For both Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua large biases are found when compared against MODIS-Aqua C6 reflecting the 3.7µm LUT bug issue mentioned above. The bc-RMSE values are moderate with values around 5µm. Cloud_cci AVHRR-PM shows a relative large trend of 0.86 µm / decade which is mainly caused by a few months of 1.6µm CER_{ice} retrievals being included in the beginning of the common 2003- 2011 time period (see Figure 4-62). Apart from that, all time series look very stable with only small jumps at satellite transitions for Cloud_cci AVHRR-PM.

Table 4-19: *Evaluation scores for Cloud_cci Level-3C Effective Radius - Ice (afternoon satellites) based on comparison to C6- Aqua from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR−PM$_{3.7}$	MODIS−Aqua$_{3.7}$
Bias [µm]	+14.44	+16.74
bc−RMSD [µm]	+5.15	+4.92
Cloud_cci trend [µm / decade]	+0.86	−0.15
C6−Aqua trend [µm / decade]	−0.06	−0.04
Δ trend [µm / decade]	+0.91	−0.12

4.2.7 Liquid Water Path

In this subsection the cloud liquid water path (LWP) of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-63), zonal mean (Figure 4-64) and standard deviation (Figure 4-65), all for a common time period, and time series plots (Figure 4-66 and Figure 4-67).

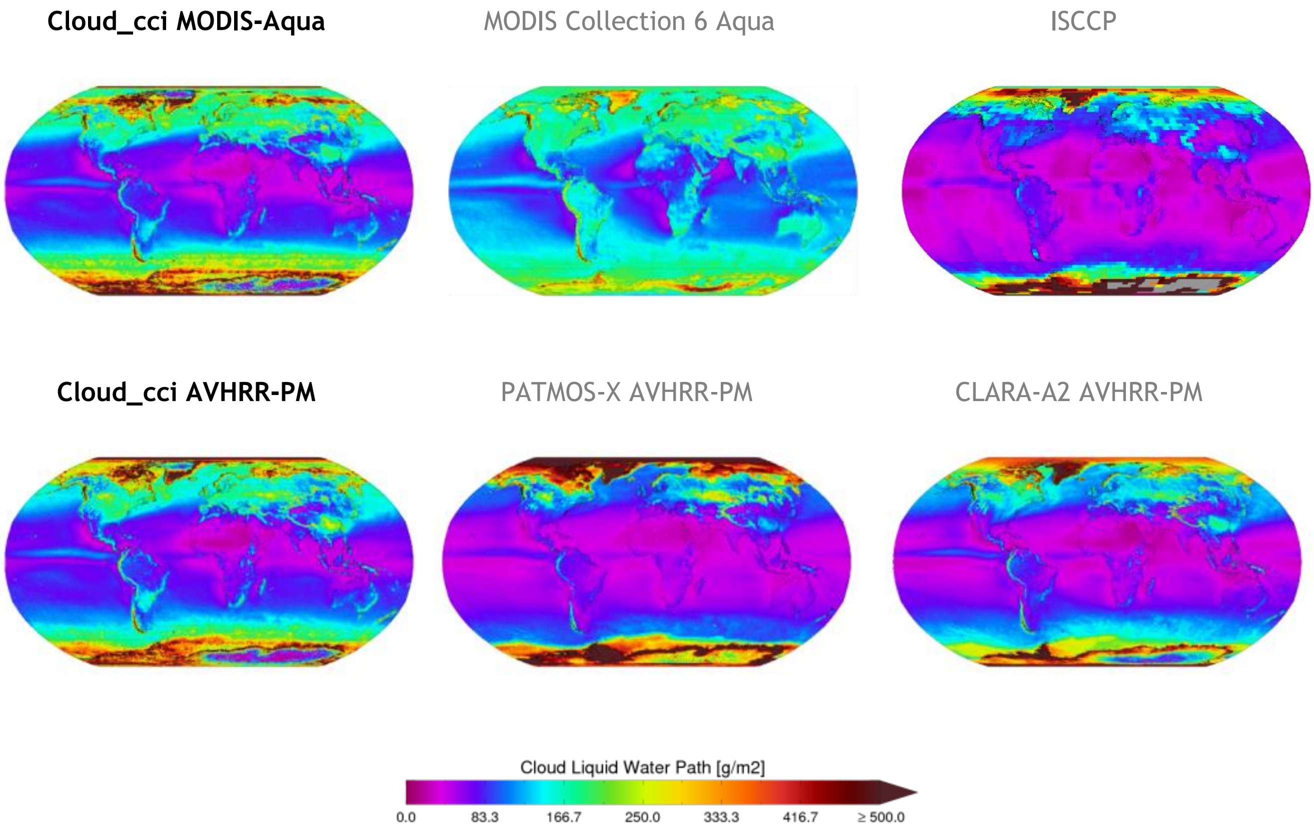
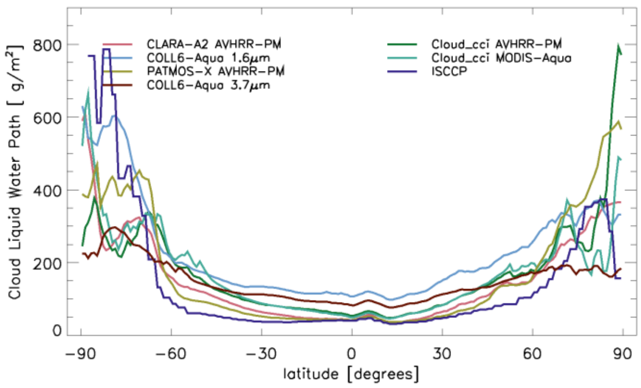


Figure 4-63 *Globally gridded means of cloud liquid water path for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*



Latitude weighted Means of Cloud Liquid Water Path [g/m2]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-PM	84.62	70.60	255.42	98.46	48.05	107.28	228.53	107.07	75.89
Cloud_cci AVHRR-PM	107.71	92.61	292.34	127.45	66.27	131.99	246.99	119.73	103.07
COLL6-Aqua	131.87	123.89	206.91	143.13	106.50	152.15	197.00	160.54	120.63
Cloud_cci MODIS-Aqua	111.60	96.03	309.62	144.58	63.82	138.05	237.48	122.87	107.22
PATMOS-X AVHRR-PM	82.18	62.29	328.74	78.38	44.56	96.51	290.58	104.85	73.40
ISCCP	58.39	48.25	205.16	52.37	39.02	70.85	191.69	74.85	52.31

Figure 4-64 *Cloud liquid water path from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Aqua the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

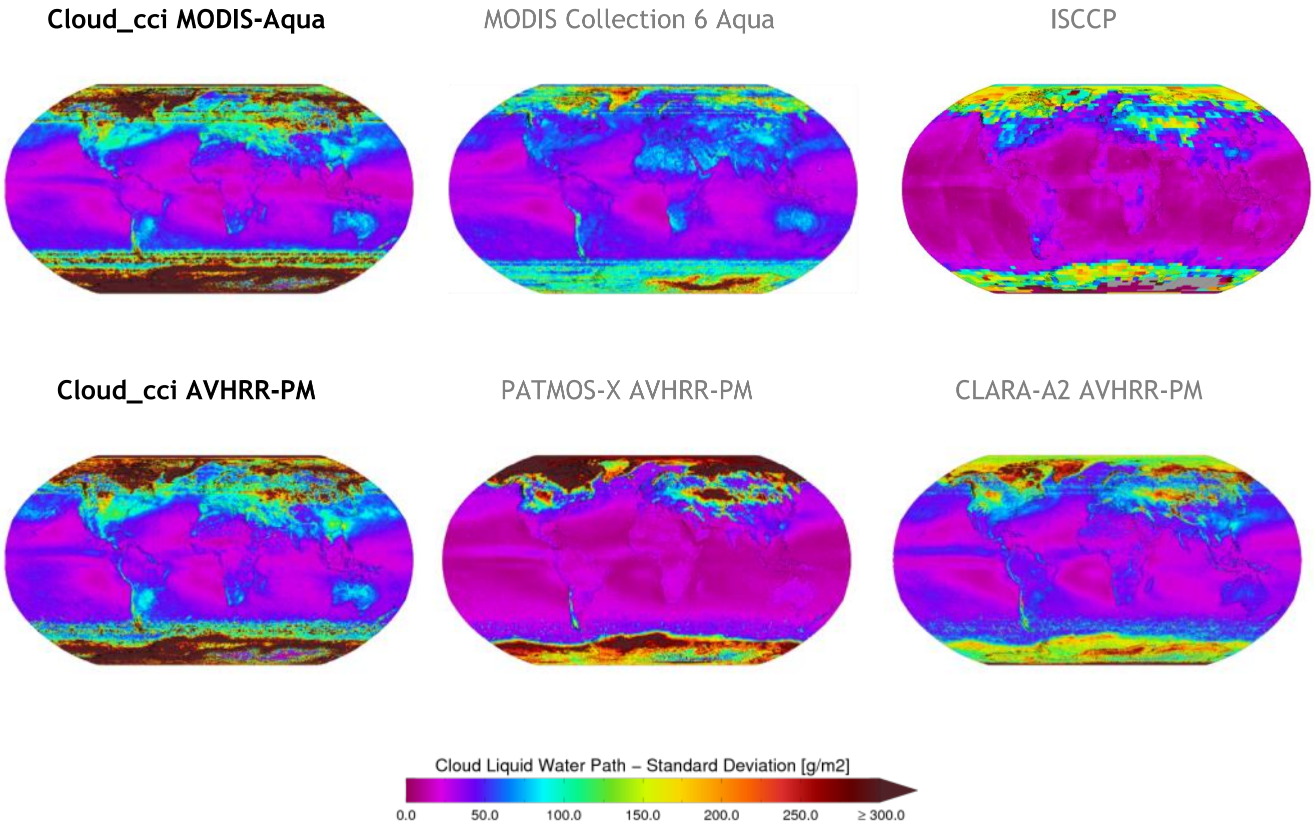


Figure 4-65 *Globally gridded means of cloud liquid water path standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*

General findings

- The mean LWP in all datasets shows similar spatial patterns with generally lower values in the tropics (a small local maximum near the ITCZ), increasing towards the mid- and high latitudes and (partly) reaching very high values in the Polar Regions, where also the spread among the datasets is highest. Apart from the polar regions, MODIS C6 exhibits highest and ISCCP lowest mean values among all datasets.
- The temporal variability of LWP shows similar pattern: low in the tropics and subtropics and high in the high latitudes.
- Time series of mean LWP show a significant seasonal cycle. The stability of the LWP time series of Cloud_cci AVHRR-PM, CLARA-A2 and PATMOS-x are again affected changing local observations time among the NOAA satellites and the by satellite drift of the early NOAA satellites that does enlarge this effect. For the later AVHRR-carrying satellites and MODIS, the time series seem relative stable, but again a small decreasing trend for LWP if found for all MODIS datasets between 2005 and 2012. In the all-sky time series plots the spread among the datasets is reduced.

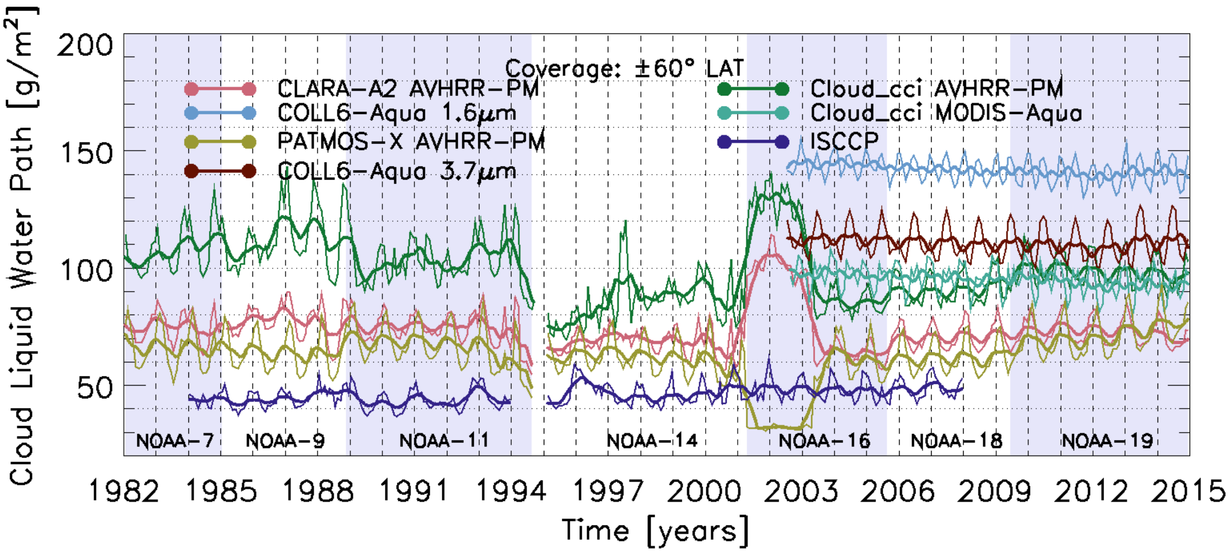


Figure 4-66 *Time series of "in-cloud" cloud liquid water path for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

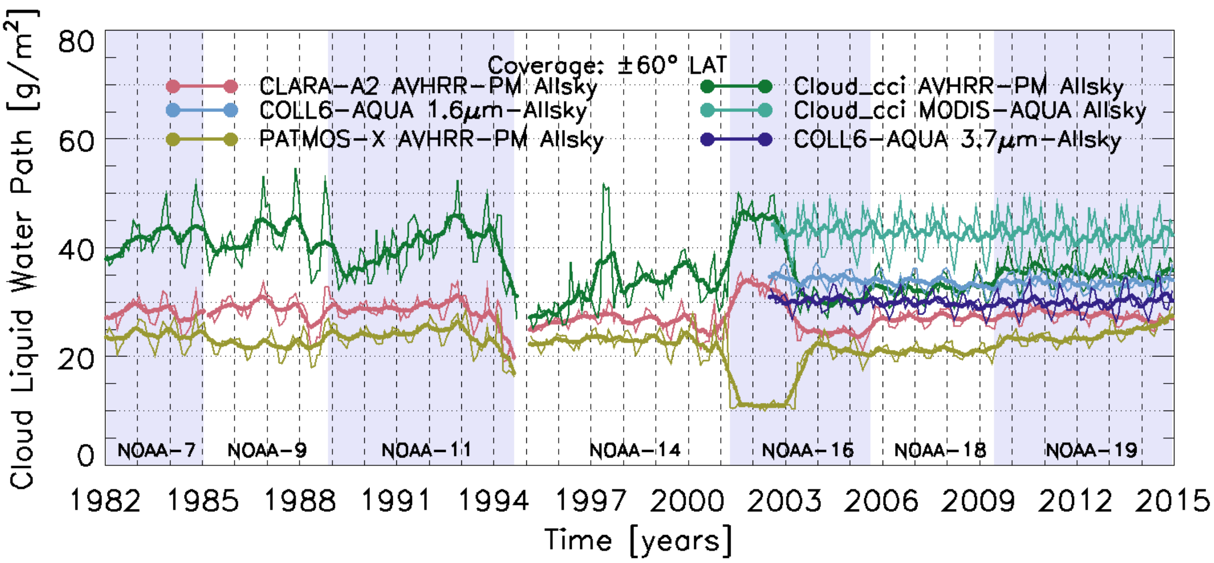


Figure 4-67 *Time series of "all-sky" cloud liquid water path for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

Discussion

For a large portion of the globe (60S-60N) the Cloud_cci mean LWP lies in the middle between MODIS-Aqua C6 data (highest) and ISCCP and PATMOS-x (lowest). Cloud_cci AVHRR-PM and MODIS-Aqua agree with the other datasets on increasing in-cloud LWP from tropical regions towards the mid-latitudes. They also agree on smaller LWP values in the eastern parts of the stratocumulus regions, compared to the western parts and compared to trade cumulus regions. Another regional feature, the LWP 'hotspot' in South-east Asia is well represented in the Cloud_cci datasets.

The time series plots again present the difficulties that the drifting, early satellites cause in Cloud_cci AVHRR-PM. LWP values increase with leave time and 'fall back' at the transition to the next sensor. However, it should be possible to increase the stability by applying a proper drift correction. Also found is the 2 year period in which the NOAA-16 1.6µm channel is switched on and used, causing significantly higher mean LWP values in this period. In the all-sky time series plot, all datasets seem to agree a bit better, which indicates that difference in CFC play a significant role in the in-cloud comparisons.

Table 4-20 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. The biases wrt. MODIS-Aqua C6 for both Cloud_cci AVHRR-PM and Cloud_cci MODIS-Terra are negative and relative small. The bc-RMSE values are similar around 70g/m². While both MODIS-Aqua datasets show a slight negative trend, Cloud_cci AVHRR-PM has a high positive trend, which is cause by significant jump for LWP between NOAA-16 and NOAA-18 and between NOAA-18 and NOAA-19.

Table 4-20: *Evaluation scores for Cloud_cci Level-3C Liquid Water Path (afternoon satellites) based on comparison to C6- Aqua from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR–PM$_{3.7}$	MODIS–Aqua$_{3.7}$
Bias [g/m^2]	–15.94	–12.60
bc–RMSD [g/m^2]	+68.62	+76.32
Cloud_cci trend [g/m^2 / decade]	+15.39	–2.91
C6–Aqua trend [g/m^2 / decade]	–4.00	–4.58
Δ trend [g/m^2 / decade]	+19.39	+1.67

4.2.8 Ice Water Path

In this subsection the cloud ice water path (IWP) of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-68), zonal mean (Figure 4-69) and standard deviation (Figure 4-70), all for a common time period, and time series plots (Figure 4-71 and Figure 4-72).

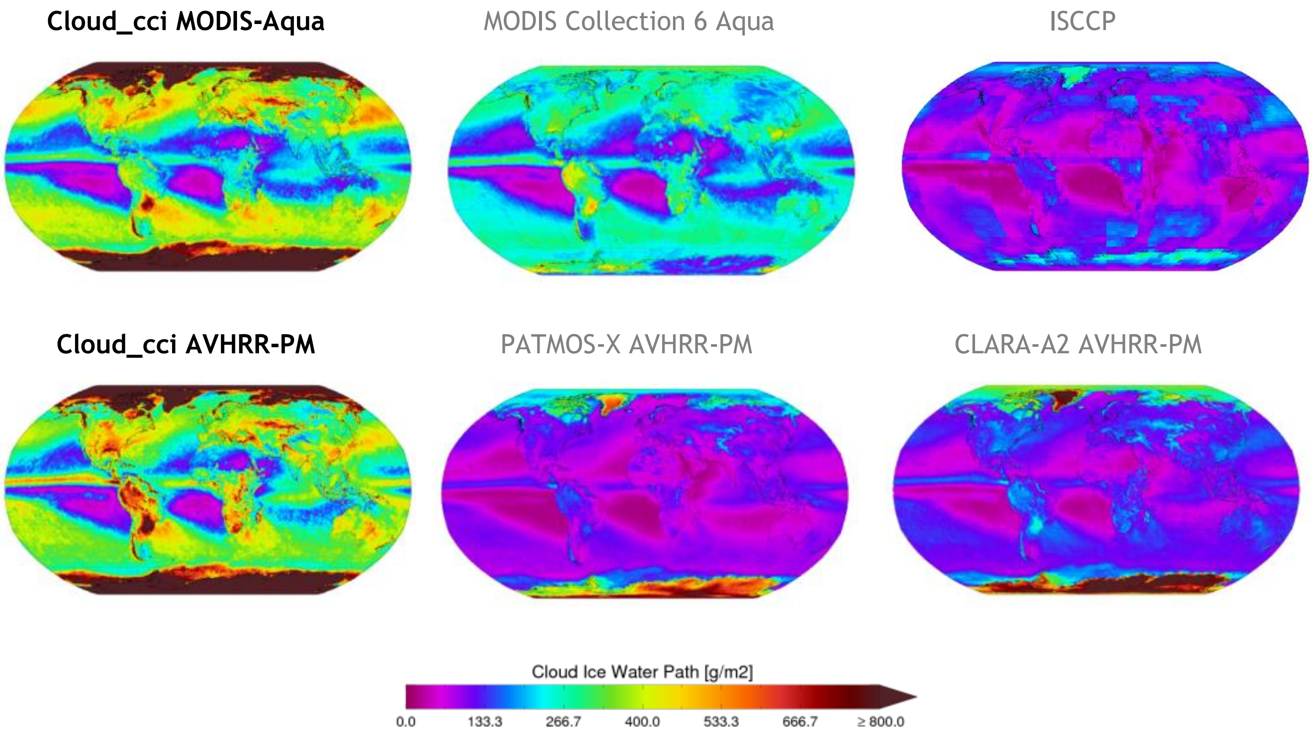
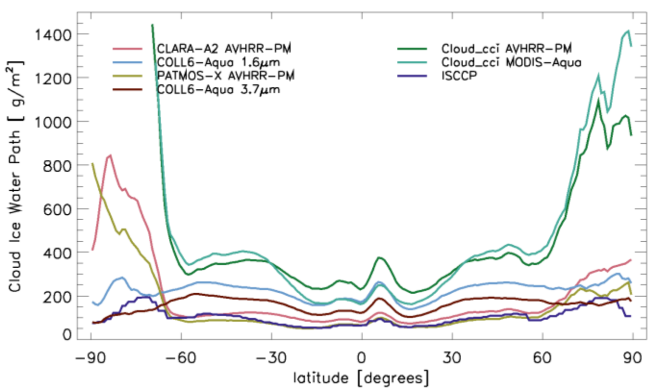


Figure 4-68 *Globally gridded means of cloud ice water path for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*



Latitude weighted Means of Cloud Ice Water Path [g/m2]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-PM	122.42	106.85	353.76	117.04	95.63	125.80	215.41	163.05	106.14
Cloud_cci AVHRR-PM	362.48	308.84	1165.46	347.15	274.62	363.33	639.88	493.74	310.64
COLL6-Aqua	211.18	207.89	229.80	249.30	178.31	245.18	248.18	235.70	201.46
Cloud_cci MODIS-Aqua	351.41	282.64	1372.37	383.63	208.12	387.71	722.93	467.00	306.22
PATMOS-X AVHRR-PM	87.37	75.30	261.32	84.36	65.67	93.41	175.94	114.73	76.54
ISCCP	87.85	82.79	137.50	102.54	67.63	102.51	134.01	93.49	85.59

Figure 4-69 *Cloud Ice water path from 2003 to 2011. Left: Zonal means. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown. Right: Latitude weighted means for different regions. For MODIS Collection 6 Aqua the 2.1µm product is used. Highest and lowest values of each region are highlighted.*

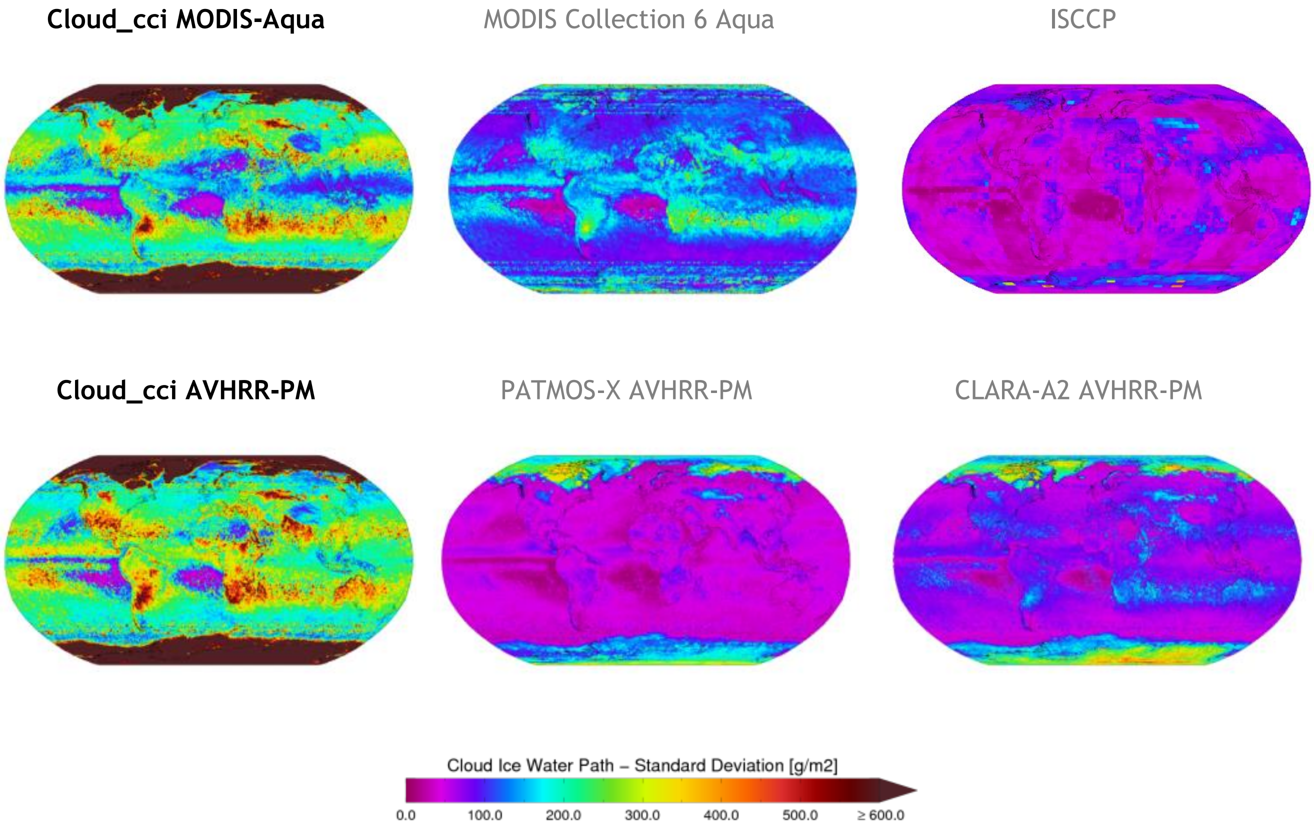


Figure 4-70 *Globally gridded means of cloud ice water path standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey. For MODIS Collection 6 Aqua the 2.1µm product is plotted.*

General findings

- The mean IWP is high for in the ITCZ for all datasets, decreasing in the subtropics and increasing again towards the mid-latitudes with higher values over land. Polar regions show extremely high values, in particular in the Cloud_cci datasets which also have highest values among the datasets for the rest of the globe.
- Apart from the polar regions, highest temporal variability for LWP is found tropical and subtropical land regions.
- In the time series plots, Cloud_cci AVHRR-PM exhibits a very high sensitivity to the local observation time creating trends and jumps in IWP during the life time of the early NOAA satellites and at their transitions. This effect is smaller (possibly due to the significantly lower absolute values for IWP) for CLARA-A2, ISCCP and PATMOS-x. A significant jump is found in 2000 to 2003 for all AVHRR datasets because of the NIR channel switch of NOAA16 (1.6µm instead of 3.7µm).

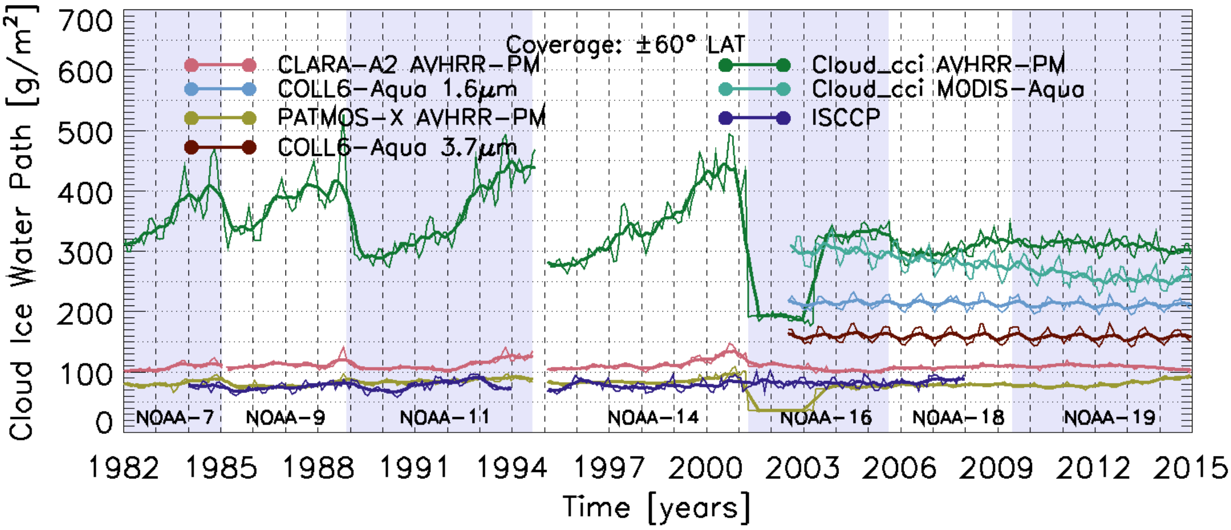


Figure 4-71 *Time series of "in-cloud" cloud ice water path for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

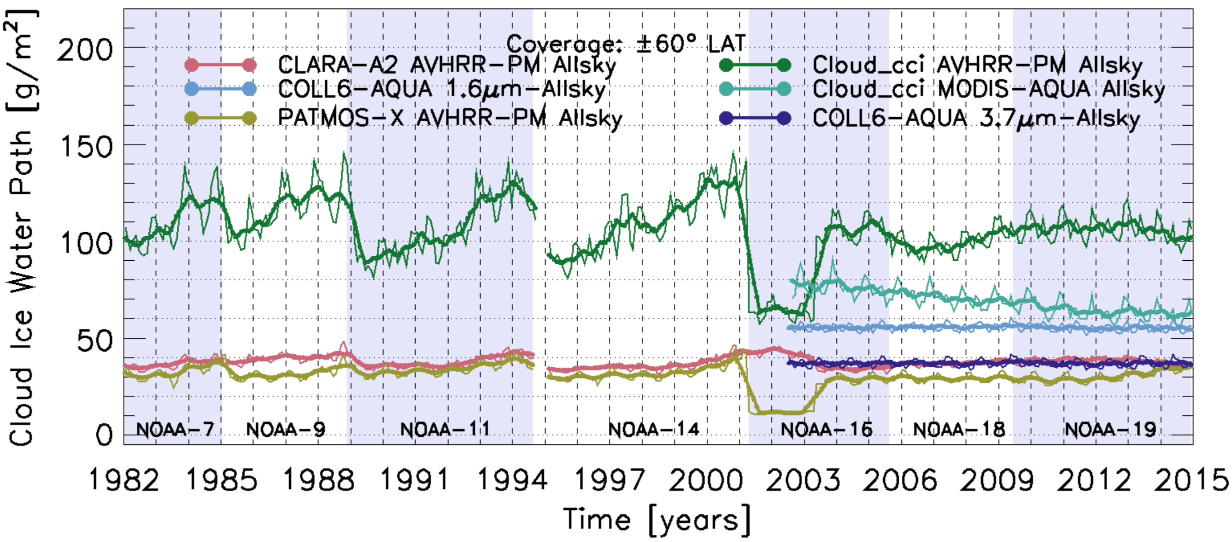


Figure 4-72 *Time series of "all-sky" cloud ice water path for all afternoon satellite retrievals. For MODIS Collection 6 Aqua the 1.6µm and 3.7µm products are shown.*

Discussion

Although the Cloud_cci dataset reflect the spatial pattern of mean IWP, they present values that are about 2 times higher than the reference datasets. Since is mainly due to the LUT bug described before, that caused in particular too high CER_{ice}. As IWP is diagnosed using COT_{ice} and CER_{ice} this overestimation has a direct effect on the IWP. Thus, the absolute IWP values of Cloud_cci AVHRR-PM and MODIS-Aqua are not trustable. However, relative information, e.g. that one region has higher IWP the another, should be inferable from the data.

The drift of the early NOAA satellites has a significant impact on the stability of the Cloud_cc AVHRR-PM dataset. For Cloud_cci MODIS-Aqua, the degradation of one of the MODIS visible channels becomes clear.

Table 4-21 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. For both Cloud_cci MODIS-Aqua and AVHRR-PM (both using the 3.7µm channel and are compared against MODIS-Terra C6 3.7µm) very large positive biases are found, which is mainly due to the 3.7µm LUT bug in CC4CL as mentioned before. The bc-RMSE values lie between 154 and 185 g/m². While MODIS-Aqua C6 3.7µm data has nearly no trend in the considered period, Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua shoe relative large trends, but of opposite sign. For Cloud_cci AVHRR-PM it is due to some NOAA-16 data (which are of much lower IWP due to the use of the 1.6µm channel, are included. The Cloud_cci MODIS-Aqua IWP stability is strongly affected by a growing degradation of one of the MODIS visible channels, which MODIS-Aqua C6 does not use.

Table 4-21: *Evaluation scores for Cloud_cci Level-3C Ice Water Path (afternoon satellites) based on comparison to C6- Aqua from 2003-2011. As different shortwave IR channels were used for the Cloud_cci datasets, the MODIS reference product based on the same shortwave IR channel is used for comparison. The used channel od both the Cloud_cci and MODIS data is indicated by a subscript at the Cloud_cci dataset name (1.6: 1.6µm channel, 3.7: 3.7µm channel). In addition, the scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	$AVHRR\text{–}PM_{3.7}$	$MODIS\text{–}Aqua_{3.7}$
Bias [g/m^2]	+149.79	+126.76
bc–RMSD [g/m^2]	+183.58	+176.96
Cloud_cci trend [g/m^2 / decade]	+20.76	–49.04
C6–Aqua trend [g/m^2 / decade]	–1.73	–1.90
Δ trend [g/m^2 / decade]	+22.49	–47.15

4.2.9 Liquid Cloud Fraction

In this subsection the liquid cloud fraction LCF of Cloud_cci AVHRR-PM, Cloud_cci MODIS-Aqua, CLARA-A2 (afternoon satellites only), MODIS Aqua C6, ISCCP and PATMOS-x (afternoon satellites only) are compared by means of multi-annual mean (Figure 4-73), zonal mean (Figure 4-74) and standard deviation (Figure 4-75), all for a common time period, and time series plots (Figure 4-76).

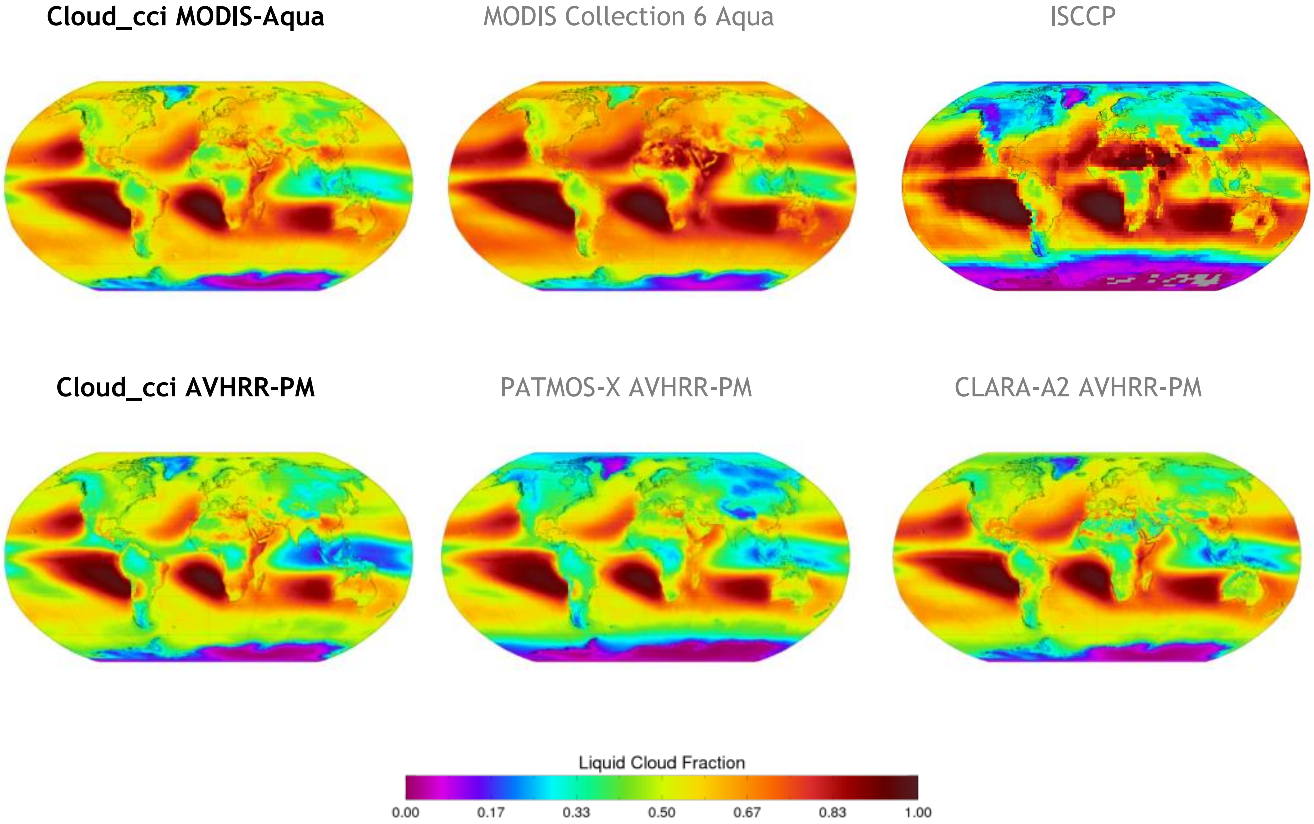
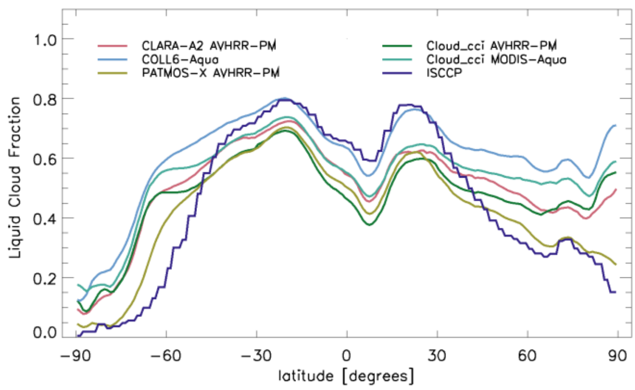


Figure 4-73 *Globally gridded means of liquid cloud fraction for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*



Latitude weighted Means of Liquid Cloud Fraction [%]

Dataset-name	Global	±60° Lat.	Ant-arctica	Mid-Lat. South	Tropics	Mid-Lat. North	Arctic	Land	Sea
CLARA-A2 AVHRR-PM	56.34	59.20	32.58	60.17	61.01	53.30	43.01	44.47	61.19
Cloud_cci AVHRR-PM	51.98	54.09	33.02	55.85	55.33	48.93	43.70	44.18	55.15
COLL6-Aqua	65.84	68.39	41.47	69.00	69.98	63.44	57.37	56.32	69.73
Cloud_cci MODIS-Aqua	58.64	60.79	37.66	62.09	61.90	56.46	51.80	50.44	61.97
PATMOS-X AVHRR-PM	50.26	54.15	18.55	54.54	57.50	44.62	31.70	37.08	55.63
ISCCP	60.09	63.46	12.72	55.97	71.59	48.20	29.95	50.02	63.78

Figure 4-74 *Liquid cloud Fraction from 2003 to 2011. Left: Zonal means. Right: Latitude weighted means for different regions. Highest and lowest values of each region are highlighted.*

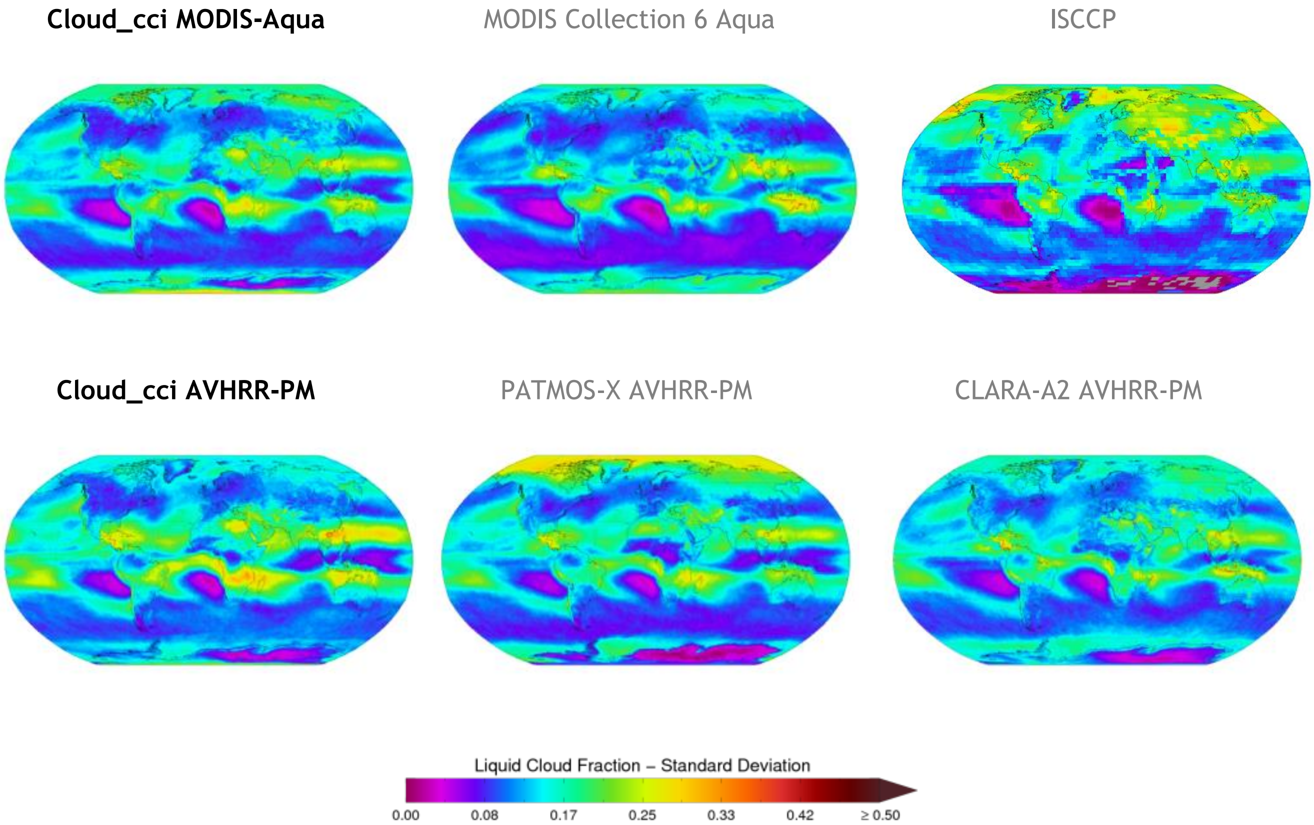


Figure 4-75 *Globally gridded means of liquid cloud fraction standard deviation for all afternoon satellite retrievals averaged over the years 2003 to 2011. Reference data named in grey.*

General findings

- The LCF of all datasets show the characteristic patterns with lower LCF in the inner topics and very high LCF for the stratocumulus regions. The LCF is decreasing again towards the higher latitudes. Extremely low values are found for Antarctica. Highest LCF is found for MODIS Aqua C6, lowest for PATMOS-x and ISCCP outside the tropics and for Cloud_cci AVHRR-PM inside the tropics.
- The stratocumulus regions are also the regions with the lowest temporal variability in LCF. Relative large variability is found in the outer tropics / inner subtropics over land and over oceans without persistent stratocumulus regimes. The temporal variability in the mid-latitudes is relative low. All datasets seem to agree more or less on these findings.
- The time series plots of LCF show significant seasonal cycles in the global (60S-60N) mean in all datasets. Cloud_cci AVHRR-PM shows largest discontinuities among the AVHRR-based datasets for the early NOAA satellites.

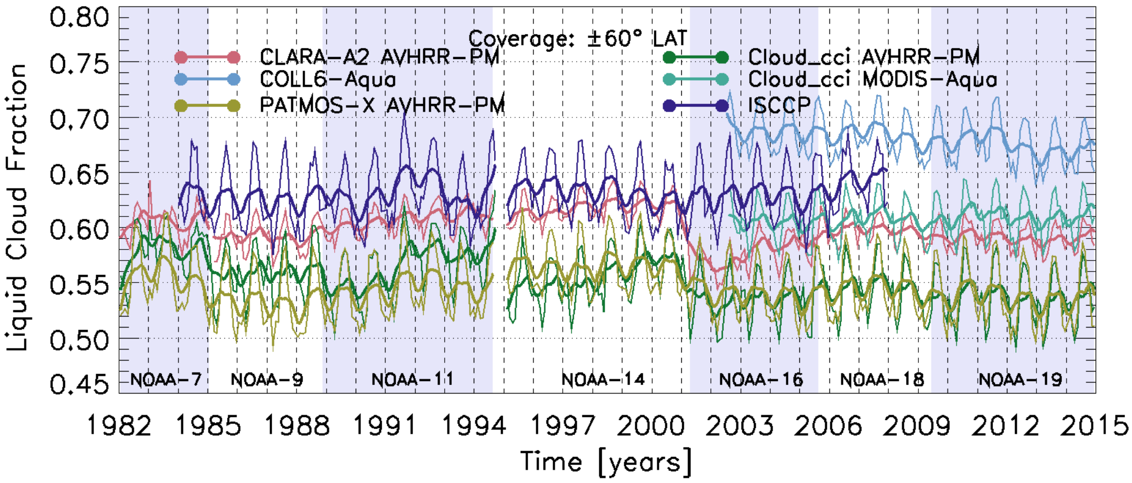


Figure 4-76 *Time series of liquid cloud fraction for all afternoon satellite retrievals.*

Discussion

The LCF of Cloud_cci AVHRR-PM and MODIS-Aqua show, similar to the morning satellite datasets, a high level of agreement with the reference data, in particular in relative terms, e.g. global spatial patterns. In terms of absolute LCF values, MODIS has about 10% higher LCF nearly everywhere except in the southern high latitudes; ISCCP has lowest LCF anywhere outside the tropical region values. In terms of long-term stability, Cloud_cci AVHRR-PM is always within 5%, however with slight, short-term trends for the early NOAA satellites and two positive anomalies related to the volcano eruptions of El Chichón (1982) and Pinatubo (1991), for which also positive CFC and CTP anomalies were found. Falsely detected low clouds would cause positive LCF anomalies which indicates that these two anomalies are artefacts. A Pinatubo eruption related anomaly can also be found for ISCCP.

Table 4-22 presents evaluation scores inferred from comparisons of Cloud_cci datasets with MODIS-Aqua C6. As seen for MODIS-Terra C6, MODIS-Aqua C6 has a very high LCF, significantly higher than the Cloud_cci datasets. Biases are between -7 and -15 %. The bc-RMSE values amount to 7.8 and 9.8, respectively. Trends for Cloud_cci AVHRR-PM and Cloud_cci MODIS-Aqua are small and positive, the trend for MODIS-Aqua C6 small and negative. All in all, all three datasets are relative stable and comparable.

Table 4-22: *Evaluation scores for Cloud_cci Level-3C Liquid Cloud Fraction (afternoon satellites) based on comparison to C6- Aqua from 2003-2011, in case of MERIS+AATSR the C6- Aqua Liquid Cloud Fraction is daytime only. The scores were calculated separately for each Cloud_cci dataset by including all valid data points pairwise in the MODIS and the Cloud_cci dataset. This can introduce some small variability in the trend values for MODIS.*

Measure	Cloud_cci dataset	
	AVHRR–PM	MODIS–Aqua
Bias [%]	–14.30	–7.60
bc–RMSD [%]	+9.80	+7.85
Cloud_cci trend [% / decade]	+0.60	+1.12
C6–Aqua trend [% / decade]	–0.43	–0.43
Δ trend [% / decade]	+1.02	+1.54

5. Homogeneity analysis of cloud fraction (CFC) using SYNOP

To be suitable for climate analysis, satellite-derived cloud datasets have to meet the challenging requirements including those for accuracy, precision and decadal stability (URDv2). In this section the Cloud_cci datasets are examined if they comply with the requirement for decadal stability defined as the change of the CFC accuracy in time (i.e. per decade).

The analysis was performed using a subset of 874 SYNOP sites derived from the ECMWF archive (Figure 5-1). From the initial over 6000 available sites we selected only those where observations were continuously performed in 1982–2014 (i.e. to cover the longest Cloud_cci dataset) at least every 6 hours. For each site we used cloud amount observed with the highest temporal frequency (up to 1 hour) that was reported for the whole 33-year period. Thus the frequency of the observations could vary between sites, but remained stable in time for each site. Next, the instantaneous octa estimates were transferred to continuous cloud fractional cover (0-100%) by 0-8 octa equal 0, 10, 25, 40, 50, 60, 75, 90, 100%, respectively. Situations when 9 octa was reported (i.e. fog) were excluded from the analysis. We then aggregated instantaneous observations to monthly means. Finally, we excluded sites for which the Standard Normal Homogeneity Test (SNHT, see Appendix B for details) detected any inhomogeneity in the time series of CFC monthly anomalies at 99% confidence level.

Corresponding satellite-derived time series were extracted from individual 0.5x0.5 degree cells within which the sites were located. The accuracy was revealed by a mean bias error calculated against synoptic observations at a monthly time step. In addition, the bc-RMSE was calculated to check whether precision of CFC is stable over time. Performance statistics were computed for each SYNOP site and further aggregated to a global scale, land, ocean and climate zones (Figure 5-1). It has to be noted that the performance statistics were calculated only for SYNOP sites, which coverage differs among climate zones (e.g. limited in tropical and polar zones). Therefore, the analyses were also carried out based on monthly anomalies calculated separately for each grid and then aggregated to climate zones.

The homogeneity assessment was conducted by means of the SNHT. A break in a time series was identified if the statistic T(k) of the SNHT reached the critical value at the 95% confidence level (Appendix B). Following the guidelines of Aguilar et al. (2003) and Toreti et al. (2011) we first carried out the relative SNHT based on the de-trended mean monthly cloud fraction difference between satellite-derived CFC and SYNOP. This difference was aggregated over all grids where the SYNOP sites were located. Assuming homogeneity of the reference (SYNOP) data, a break detected by the relative test should reveal an inhomogeneity in the Cloud_cci series caused by non-climatic factors such as changes in satellites.

For a trend analysis we used linear trends derived using Theil-Sen estimates (Theil, 1950) and their significance was estimated with the Mann-Kendall test (Kendall, 1938; Mann, 1945). For multiple comparisons of the statistical significance of each grid, we applied the adjustment of p-value using the method of Benjamini and Hochberg (1995).

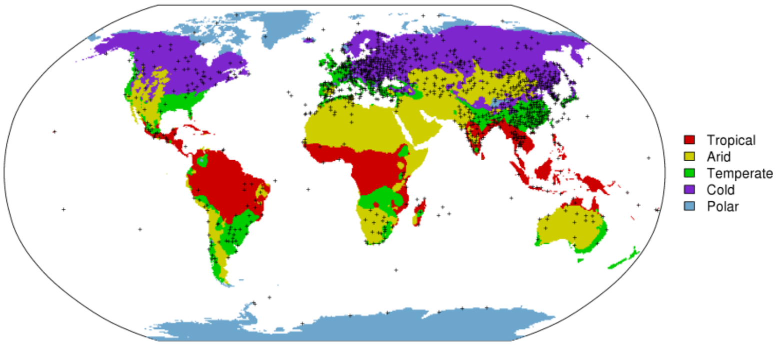


Figure 5-1 *The SYNOP sites (marked as +) overlaid on the Köppen-Geiger map of climate zones (Peel et al., 2007).*

In the following, the accuracy, precision and decadal stability are presented separately for each sensor and its morning and afternoon orbit. Exhaustive results are presented for AVHRR-AM, AVHRR-PM and (A)ATSR, which are affected by changes in consecutively operating sensors. These can lead to immediate alternation in accuracy and precision of cloud retrievals among sensors due to inter-calibration issues, satellite orbital drift, as well as changes in equatorial crossing time. Results for MODIS-AQUA and MODIS-TERRA are shortened, as they are derived by one satellite/sensor.

5.1 Cloud_cci AVHRR-AM

Morning NOAA satellites consist of NOAA-12, 15, 17 and Metop-A. On a global scale, a mean bias for each satellite is below 1% for all satellites except NOAA-12, which reveals a bias of 8.65% (Table 5-1). The differences between accuracies among satellites (that can lead to inhomogeneities) are distinct for all zones apart from Tropics and Ocean. Time series of monthly cloud fraction anomalies confirm a break between NOAA-12 and other morning NOAA satellites (Figure 5-2). The break is also evident for the Ocean (taking into account all grids), even though diminished in the bias averaged over the limited number of SYNOP sites (Figure 5-3). Hence, only for the tropics the inhomogeneity is not apparent.

Cloud fraction of NOAA-12 and NOAA-15 are less precise (bc-RMSE >12%) than NOAA-17 and Metop-A (bc-RMSE < 10%, Table 5-2). NOAA-12 reveals constantly higher bc-RMSE, while higher bc-RMSE for NOAA-15 is caused by outlying errors in 2000-2001 (Figure 5-4) that are recommended for a further investigation.

Figure 5-5 presents results of SNHT performed based on the bias and monthly anomaly time series. The former evidently exposes the inhomogeneity at the turn of 1998 and 1999 (i.e. NOAA-12/NOAA-15) for all zones except Tropics. Moreover, the break is detected by the relative SNHT for the Ocean in 1994-1995 (i.e. within the NOAA-12 operation), which is not confirmed by a relative SNHT run over monthly anomalies. Though, it should be further investigated if the break is not caused by inaccuracies of synoptic observations at limited number of sites in the Ocean zone.

The inhomogeneity between NOAA-12 and NOAA-15 is largely responsible for a negative trend in CFC monthly anomalies calculated for each 0.5 x 0.5° grid for the whole time period (Figure 5-6 a and b). When outlying NOAA-12 is removed from the analysis, no significant trend is detected (Figure 5-6c and d).

Table 5-1 *Bias of CC4CL-AVHRR-AM mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns) and NOAA missions (rows).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
AVHRR-AM	3.12	3.18	-3.74	1.69	2.32	2.23	4.33	8.77
12	8.65	8.75	-2.53	2.93	8.32	6.29	11.88	15.11
15	0.24	0.28	-4.07	1.67	-0.34	-0.45	0.42	6.08
17	0.64	0.69	-4.55	0.27	-0.52	0.82	0.96	5.00
Metop-A	0.97	1.02	-4.18	1.45	-0.13	0.66	1.36	6.63

Table 5-2 *Bias-corrected root mean square error (bc-RMSE) of CC4CL-AVHRR-AM mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns) and NOAA missions (rows).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
AVHRR-AM	11.46	11.44	12.25	11.82	12.61	9.39	11.89	13.00
12	12.87	12.83	11.96	11.53	14.76	9.46	13.39	13.87
15	12.04	12.01	14.47	13.63	12.16	11.88	11.42	12.88
17	8.96	8.93	11.31	11.36	9.64	7.46	8.66	11.39
Metop-A	9.20	9.16	11.81	11.32	10.22	7.72	8.80	10.95

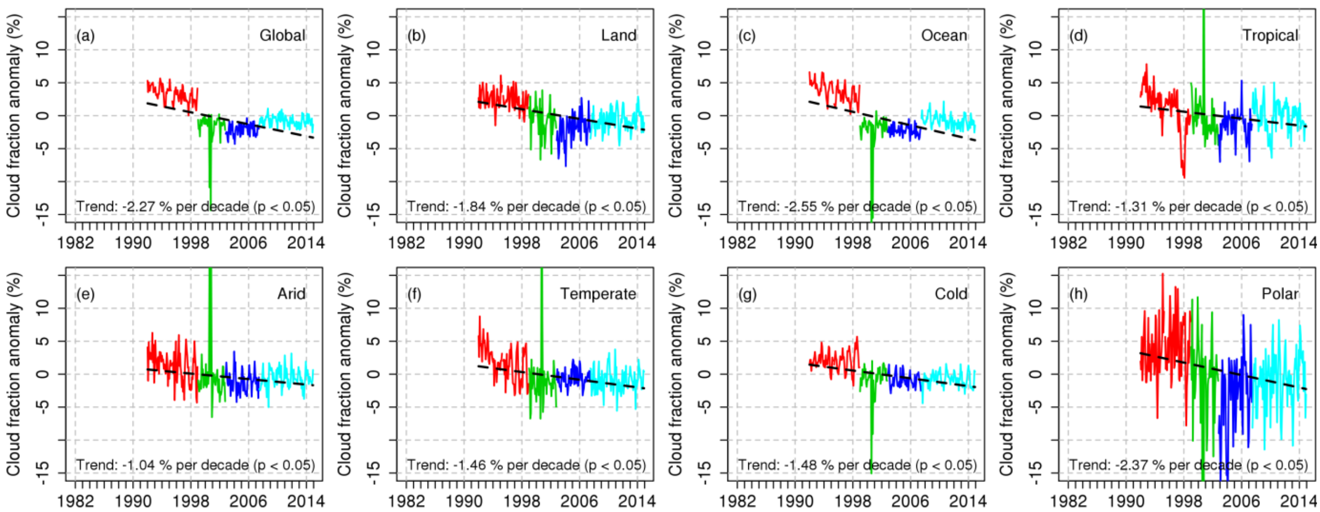


Figure 5-2 *Time series of CC4CL-AVHRR-AM monthly cloud fraction anomalies aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: NOAA-12 (red), NOAA-15 (green), NOAA-17 (blue), and Metop-A (cyan). Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

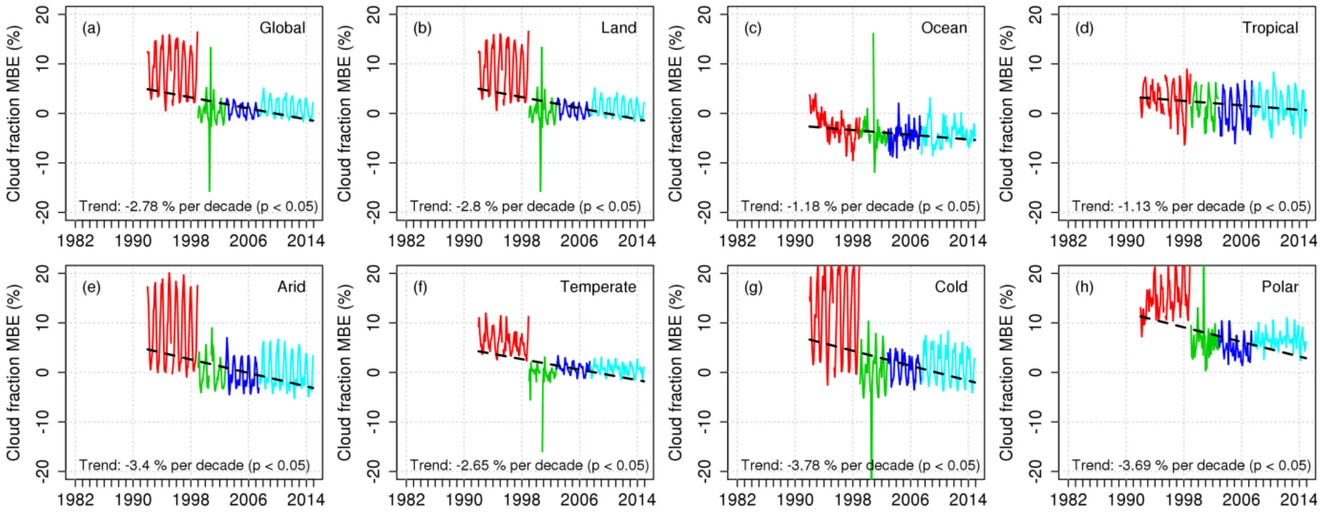


Figure 5-3 *Time series of CC4CL-AVHRR-AM bias (MBE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: NOAA-12 (red), NOAA-15 (green), NOAA-17 (blue), and Metop-A (cyan). Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

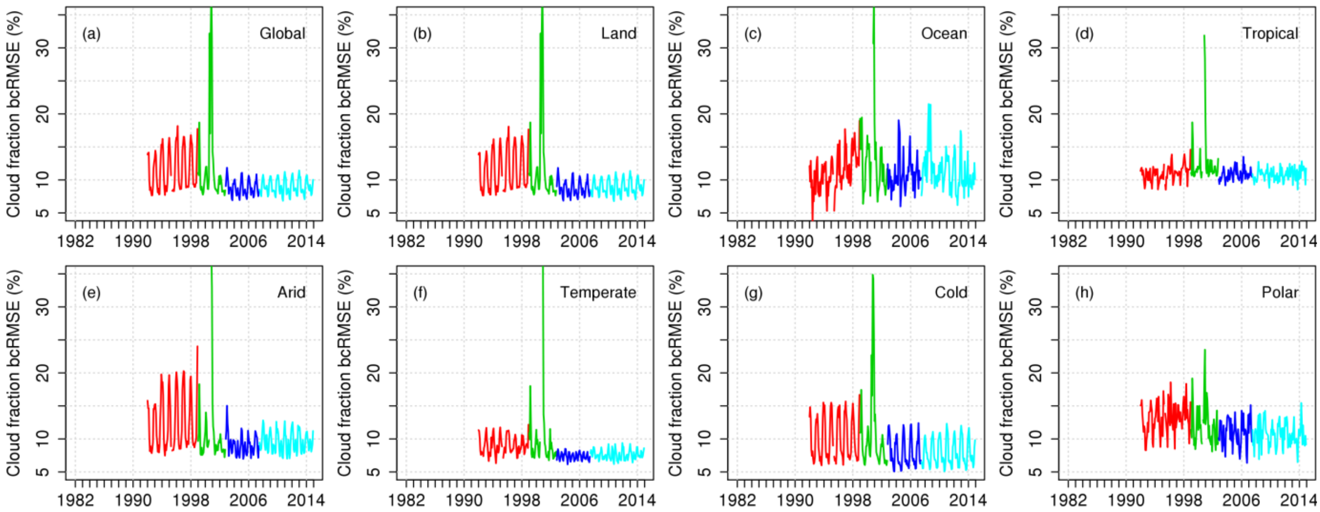


Figure 5-4 *Time series of CC4CL-AVHRR-AM bias-corrected root mean square error (bc-RMSE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: NOAA-12 (red), NOAA-15 (green), NOAA-17 (blue), and Metop-A (cyan).*

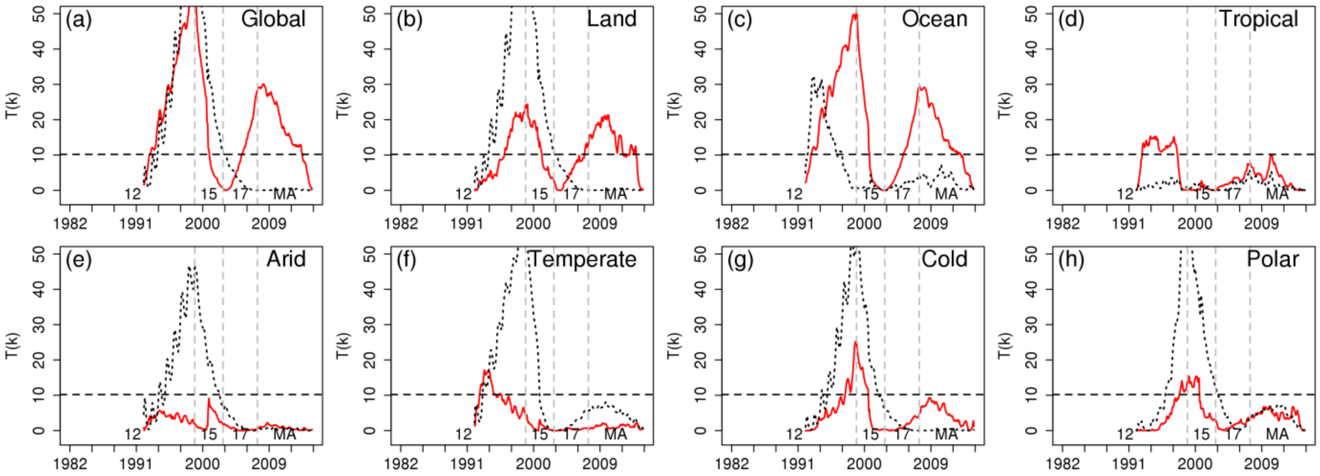
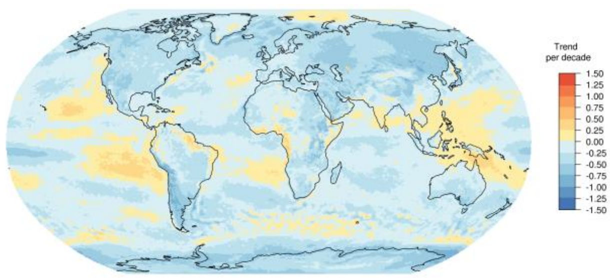
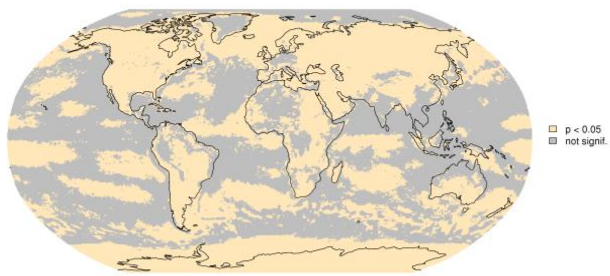


Figure 5-5 *Results of Standard Normal Homogeneity Tests (SNHT) for global (a), land (b), ocean (c) and climate zones (d-h). Relative SNHT applied to de-trended mean monthly cloud fraction difference between CC4CL-AVHRR-AM and SYNOP (black dotted line). Absolute SNHT applied to de-trended CC4CL-AVHRR-AM monthly cloud fraction anomalies for all grids (red solid line).The horizontal dashed line indicates a critical value of statistic T which signifies a break in the time series. The vertical dashed lines show changes in the satellite missions (NOAA-12, 15, 17 and Metop-A).*

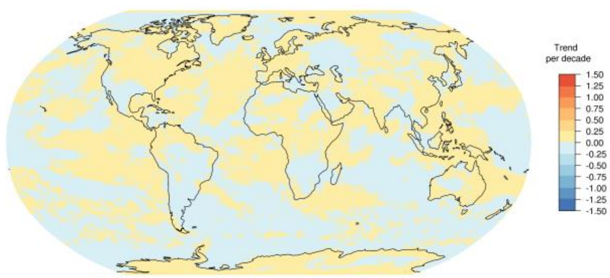
(a) Monotonic trend per decade (1992-2014)



(b) Trend statistical significance (1992-2014)



(c) Monotonic trend per decade (1999-2014)



(d) Trend statistical significance (1999-2014)

Figure 5-6 *Map of CC4CL-AVHRR-AM Theil-Sen monotonic trend (a) and its statistical significance according to the Mann-Kendall test adjusted using Benjamini-Hochberg method (b) based on the cloud fraction monthly standardized anomalies in 1992-2014. Figures c and d present trend for a limited period without NOAA-12.*

5.2 Cloud_cci AVHRR-PM

Afternoon NOAA satellites consist of NOAA-7, 9, 11, 14, 16, 18, and 19. On a global scale, a mean bias is positive of 4%, and it varies from 3.1% for NOAA-11 to 5.88% for NOAA-14 (Table 5-3). The bias differs among climatic zones, but error variability is limited within each zone. Similarly, comparable precision (bc-RMSE) is revealed for all sensors and zones (Table 5-4). A pronounced exception is the Polar zone, for which all sensors reveal lower performance.

Monthly anomalies presented in Figure 5-7 do not show evident inhomogeneities between sensors except a turn of NOAA-7 and 9 at a global scale, and for Arid and Temperate zones. The higher CFC anomalies of NOAA-7 is mostly responsible for its negative trend (significant for all but Tropics).Yet, for the Cold zone higher anomalies are revealed for both NOAA-7 and 9 (Figure 5-7g).

Time series of the bias reveal larger positive bias for NOAA-14 than for other sensors, but only for Land. It should also be noticed that this bias increases with time, most notably for the Cold zone (Figure 5-8g).

Figure 5-8d and Figure 5-9d reveal a peculiar positive trend in the bias (statistically significant) and bc-RMSE for the Tropical zone that it is recommended for an additional investigation. Besides this, bc-RMSE has a lower spread since NOAA-16 than for previous sensors (Figure 5-9a). In addition, outlying high bc-RMSE values can be seen for the first month of NOAA-9.

The relative SNHT unveils inconspicuous breaks (i.e. only slightly exceeding the critical value) between NOAA-11 and NOAA-14 (Polar, Cold and Temperate zones), as well as between NOAA-16 and NOAA-18 (Arid and Polar zones). These breaks are however not confirmed by the absolute SNHT, which concurrently detects inhomogeneity within the operation of NOAA-7.

Figure 5-11a and b present CFC trends estimated per 0.5x0.5 grid for the whole covered period. Statistically significant are mostly negative trends over the Ocean including the Antarctic Circumpolar Current (around 50°S). Positive trends are significant over Polar regions such as Greenland and Antarctica. However, NOAA-7 was removed from the analysis, due to the potential inhomogeneities detected by SNHT, the sign (positive or negative) of a per-grid trend generally remains the same, but much weaker and with decreased area of statistical significance (Figure 5-11c and d).

Table 5-3 *Bias of CC4CL-AVHRR-PM mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns) and NOAA missions (rows).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
AVHRR-PM	4.02	4.08	-2.29	3.54	2.5	3.72	4.91	8.95
7	5	5.06	-1.3	3.02	3.57	4.38	6.39	11
9	3.39	3.44	-2.01	1.69	1.95	2.58	4.88	9.45
11	3.1	3.14	-2.01	1.99	1.89	3.2	3.73	6.82
14	5.88	5.95	-1.48	4.59	4.16	5.35	7.26	10.65
16	4.14	4.21	-3.84	3.66	2.65	3.44	5.29	10.29
18	3.08	3.13	-3.07	4.13	1.37	2.97	3.58	7.15
19	3.35	3.41	-2.44	5.05	1.76	3.55	3.37	8

Table 5-4 *Bias-corrected root mean square error (bc-RMSE) of CC4CL-AVHRR-PM mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns) and NOAA missions (rows).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
AVHRR-PM	9.8	9.77	10.91	11.4	10.67	8.17	9.69	12.23
7	10.9	10.88	11.48	10.65	11.82	9.11	11.3	12.75
9	10.42	10.4	10.89	11.27	11.47	8.58	10.5	12.54
11	9.81	9.8	10.33	10.7	10.11	8.37	10.11	12.78
14	9.96	9.92	11.24	11.51	10.91	8.04	9.87	12.8
16	9.47	9.43	10.21	11.71	10.8	7.74	8.84	11.93
18	8.93	8.88	12.03	11.43	9.88	7.49	8.3	11.3
19	9	8.97	10.36	11.81	9.87	7.74	8.29	10.61

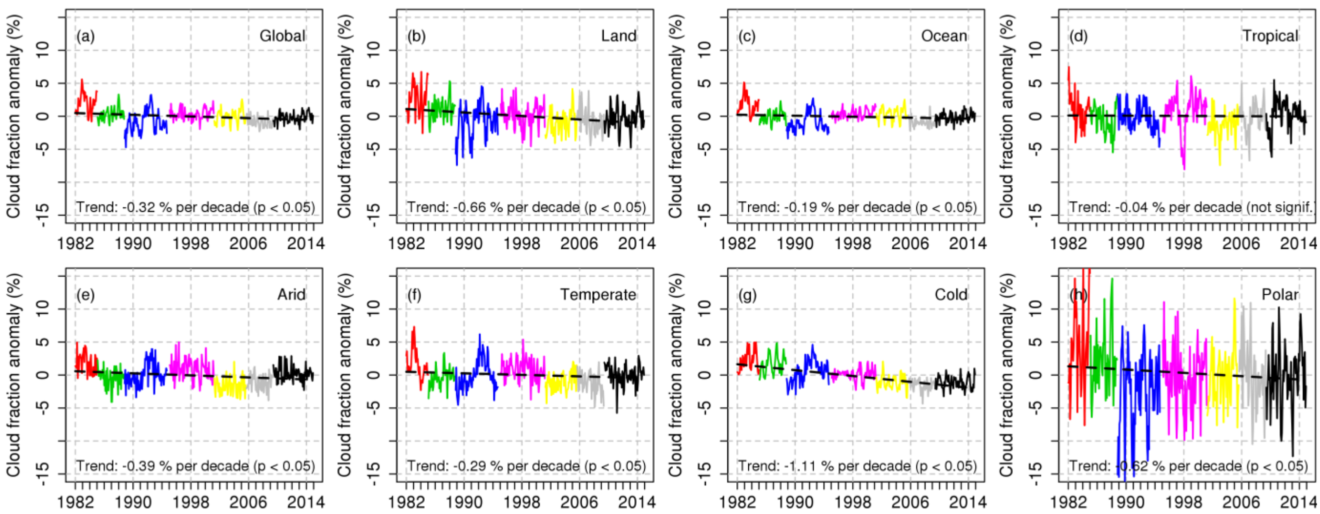


Figure 5-7 *Time series of CC4CL-AVHRR-PM monthly cloud fraction anomalies aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: NOAA-7 (red), NOAA-9 (green), NOAA-11 (blue), NOAA-14 (pink), NOAA-16 (yellow), NOAA-18 (grey) and NOAA-19 (black). Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

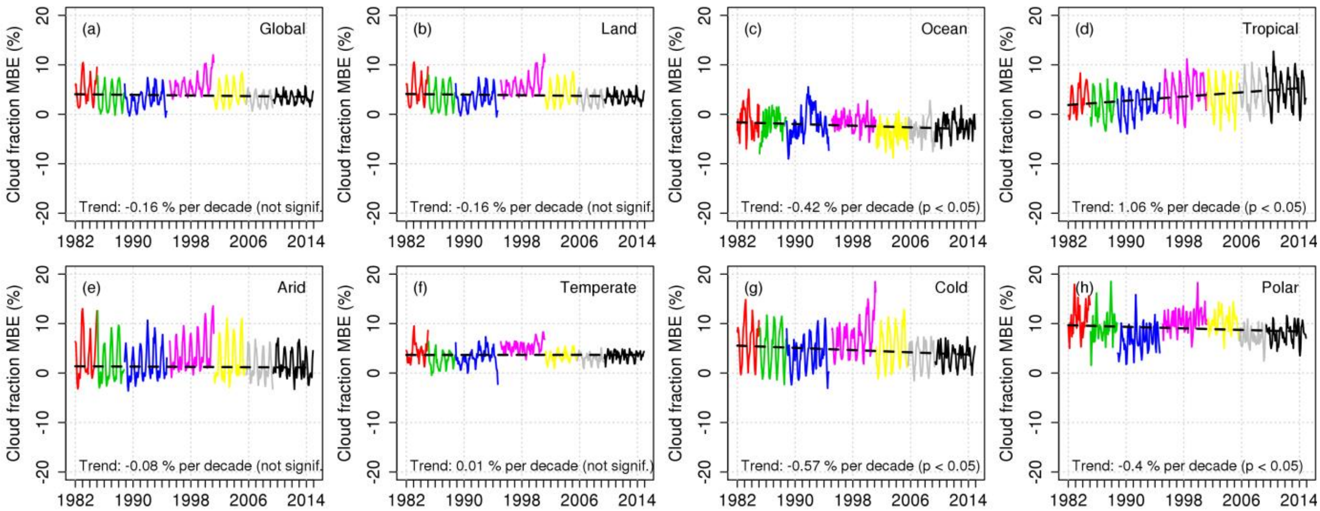


Figure 5-8 *Time series of CC4CL-AVHRR-PM bias (MBE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: NOAA-7 (red), NOAA-9 (green), NOAA-11 (blue), NOAA-14 (pink), NOAA-16 (yellow), NOAA-18 (grey) and NOAA-19 (black). Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

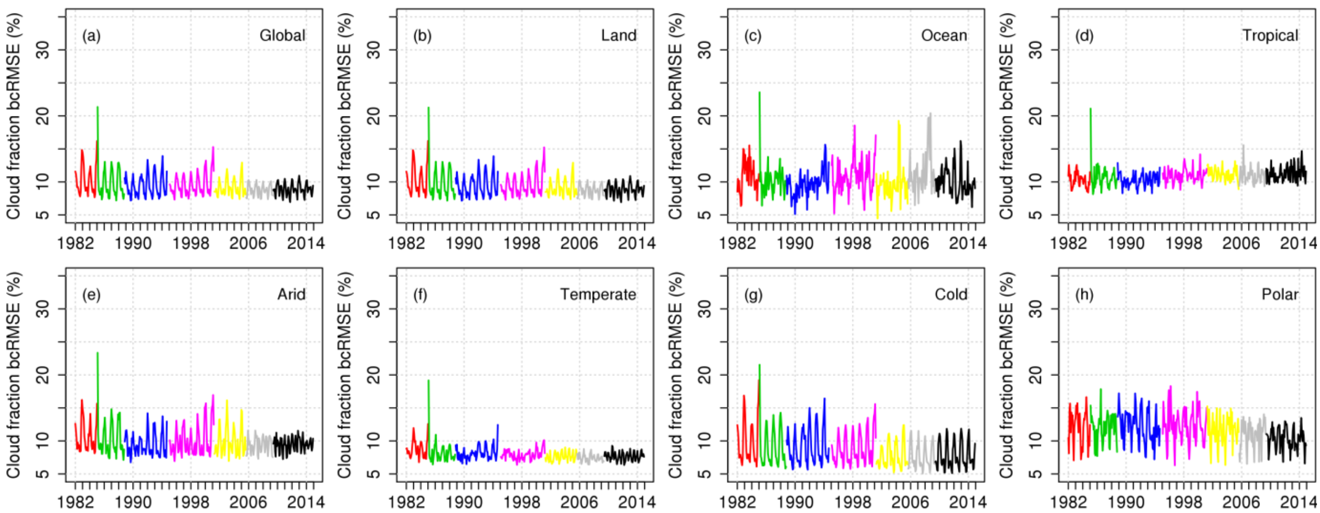


Figure 5-9 *Time series of CC4CL-AVHRR-PM bias-corrected root mean square error (bc-RMSE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: NOAA-7 (red), NOAA-9 (green), NOAA-11 (blue), NOAA-14 (pink), NOAA-16 (yellow), NOAA-18 (grey) and NOAA-19 (black).*

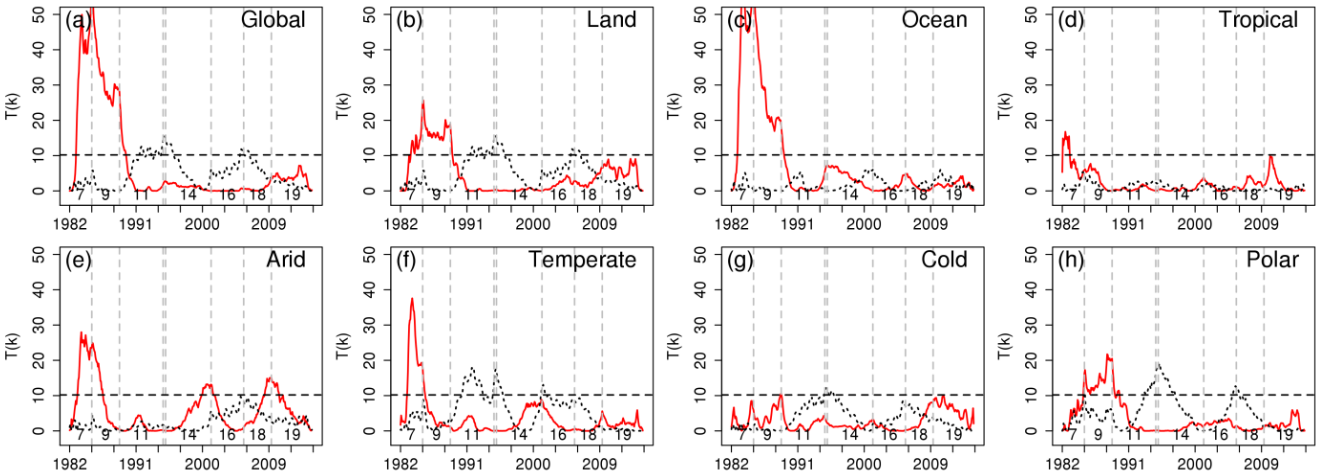


Figure 5-10 *Results of Standard Normal Homogeneity Tests (SNHT) for global (a), land (b), ocean (c) and climate zones (d-h). Relative SNHT applied to de-trended mean monthly cloud fraction difference between CC4CL-AVHRR-PM and SYNOP (black dotted line). Absolute SNHT applied to de-trended CC4CL-AVHRR-PM monthly cloud fraction anomalies for all grids (red solid line).The horizontal dashed line indicates a critical value of statistic T which signifies a break in the time series. The vertical dashed lines show changes in the satellite missions (NOAA-7, 9, 11, 14, 16, 18 and 19).*

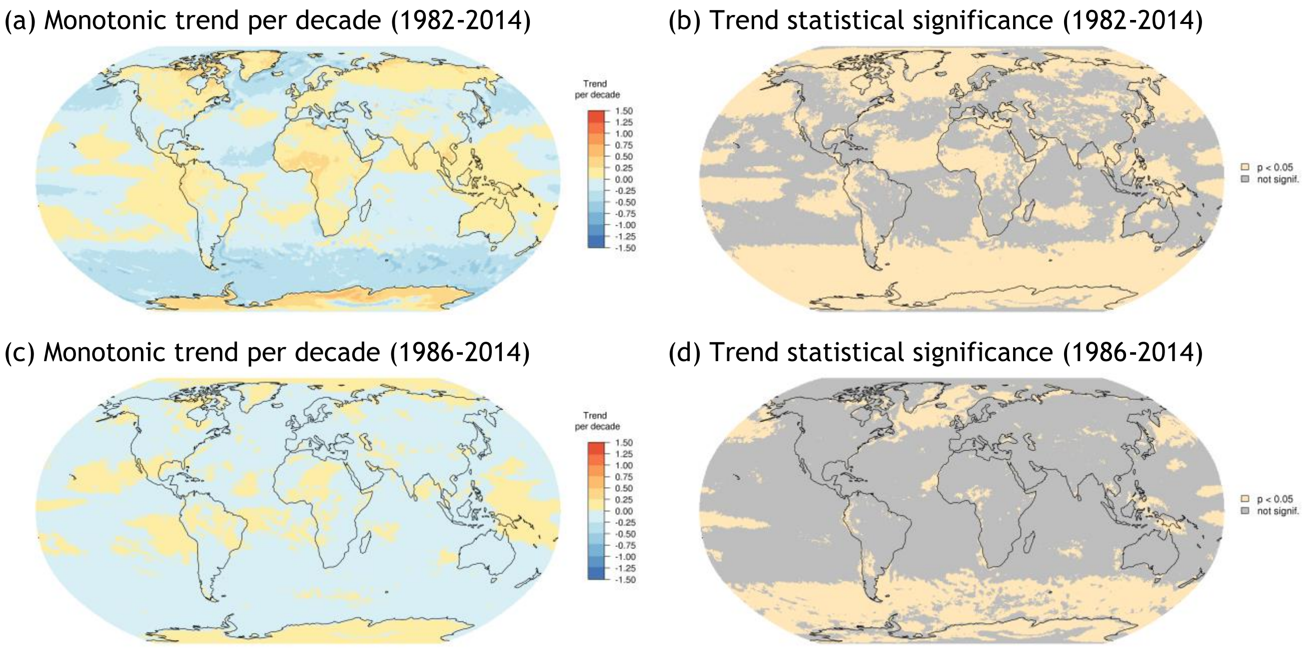


Figure 5-11 *Map of CC4CL-AVHRR-PM Theil-Sen monotonic trend (a) and its statistical significance according to the Mann-Kendall test adjusted using Benjamini-Hochberg method (b) based on the cloud fraction monthly standardized anomalies in 1982-2014. Figures c and d present trend for a limited period without NOAA-7.*

5.3 Cloud_cci ATSR2-AATSR

This dataset consists of retrievals from ATRS2-ERS2 and AATSR-ENVISAT. On a global scale, a mean bias is positive of below 1%, but the performance difference between the sensors is noticeable, especially for the Ocean zone (Table 5-5). Both sensors reveal a similar CFC precision (bc-RMSE) of 11-13% (Table 5-6).

The time series of the monthly anomalies (Figure 5-12) and bias (Figure 5-13) expose the major inhomogeneity between the sensors, however much more prominent in the anomaly time series. That is also confirmed by the SNHT, whose absolute version run over anomalies detects more evident breaks than a relative one run over the bias. This suggests that inhomogeneities are present for areas where SYNOP sites are not available. Further investigation unveiled that the inhomogeneities are more severe for daytime (Figure 5-16) than night-time (Figure 5-17) acquisitions.

The trends estimated per each grid unearthed that individual orbits might be responsible for the major inhomogeneity causing a positive trend in monthly anomalies (Figure 5-18a and b). These suspicious trend are visible for daytime acquisitions (Figure 5-18c). However, a map of trends for night-time retrievals (Figure 5-18e) also brings out some unexpected pattern seen as a sharp edge along the equator. Moreover, a large positive trend at Polar zone is more prominent for night-time estimations. A trend analysis run for AATSR only (i.e. 2003-2011) does not show any significant trends in CFC (Figure 5-18g and h).

Table 5-5 *Bias (MBE) of CC4CL-(A)ATSR mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns) and satellite missions (rows).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
(A)ATSR	-0.87	-0.81	-7.25	-0.72	-2.78	-1.15	-0.03	5.11
ATSR2	-2.06	-1.98	-11.51	-3.13	-4.58	-1.36	-1.32	3.16
AATSR	-0.07	-0.04	-4.42	0.87	-1.59	-1.01	0.82	6.41

Table 5-6 *Bias-corrected root mean square error (bc-RMSE) of CC4CL-(A)ATSR mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns) and satellite missions (rows).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
(A)ATSR	12.55	12.51	14.98	15.14	13.69	11.56	11.65	13.47
ATSR2	13.40	13.34	16.12	16.12	14.74	11.97	12.64	14.18
AATSR	11.88	11.86	13.45	14.24	12.82	11.28	10.86	12.81

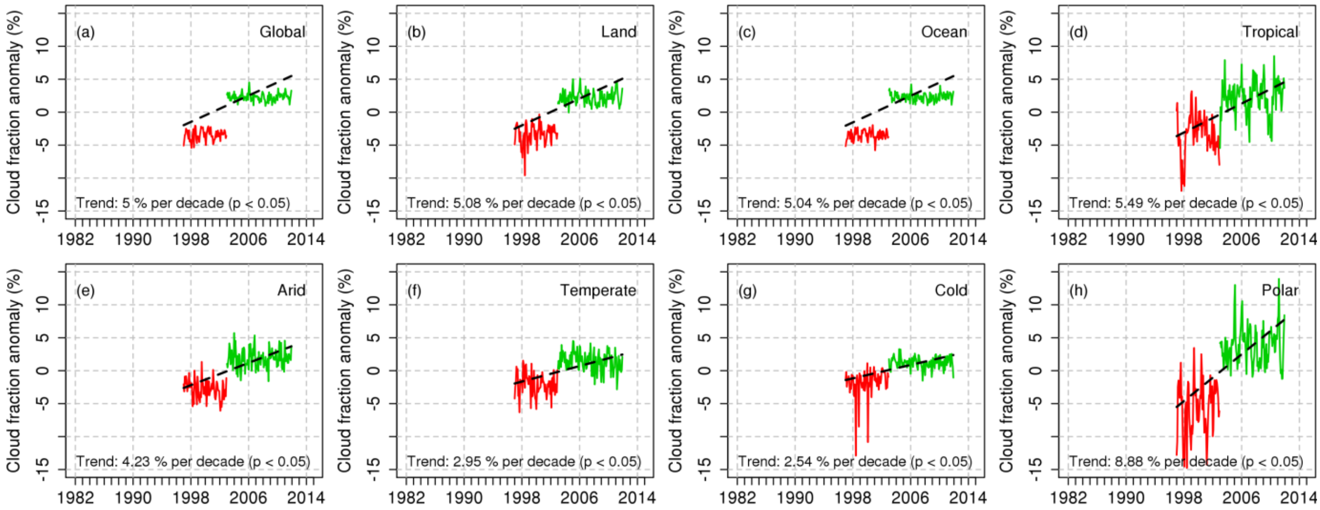


Figure 5-12 *Time series of CC4CL-(A)ATSR monthly cloud fraction anomalies aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: ATSR2-ERS2 (red) and AATSR-ENVISAT. Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

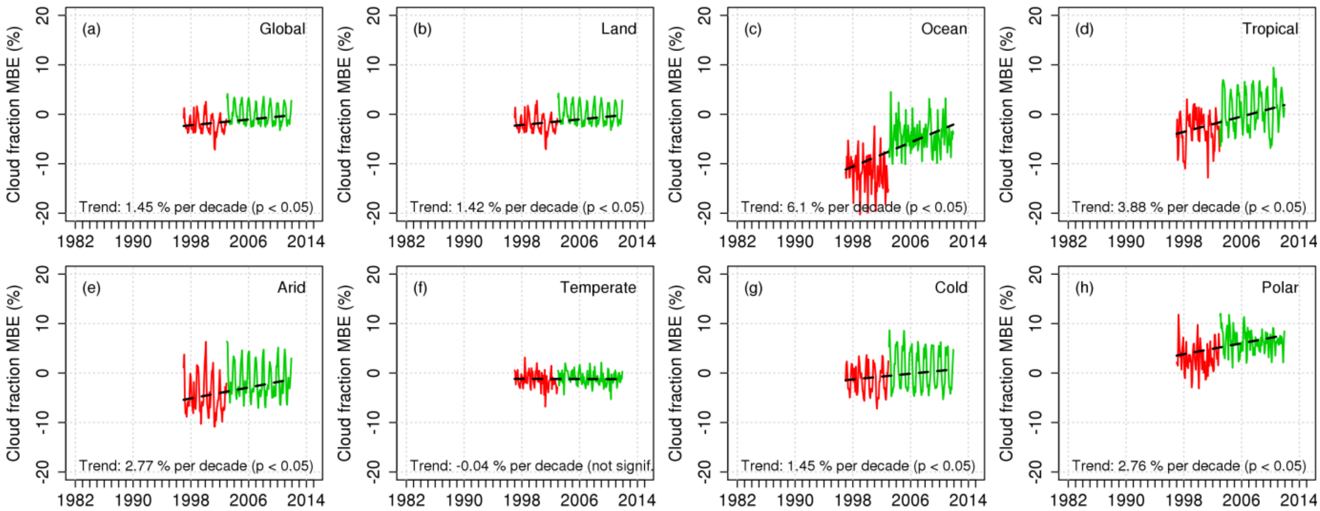


Figure 5-13 *Time series of CC4CL-(A)ATSR bias (MBE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: ATSR2-ERS2 (red) and AATSR-ENVISAT (green). Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

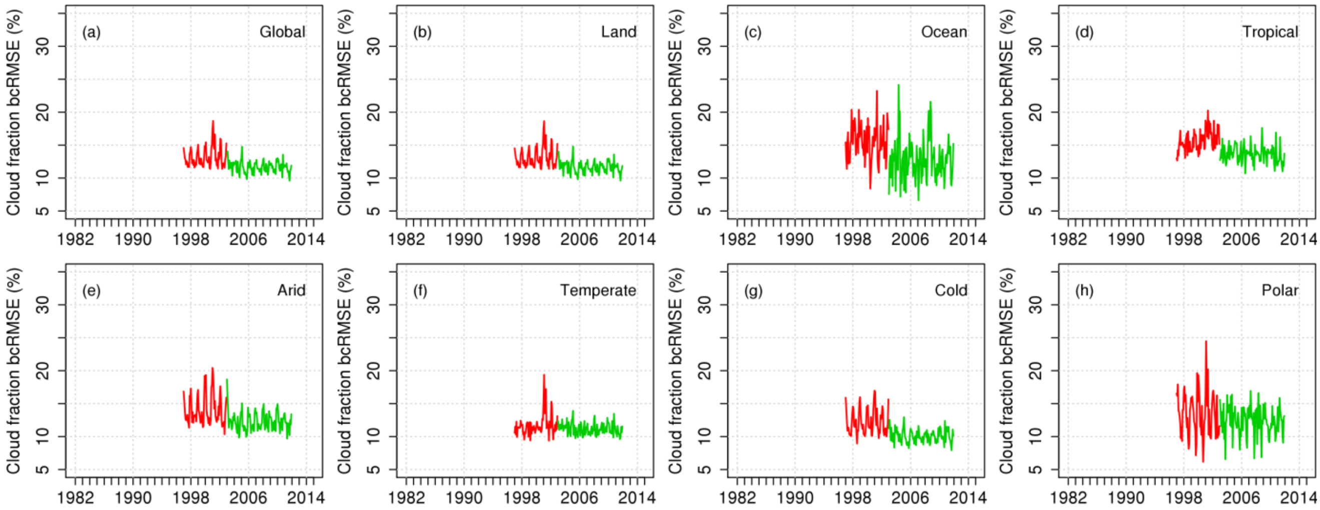


Figure 5-14 *Time series of CC4CL-(A)ATSR bias-corrected root mean square error (bc-RMSE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Colours represent consecutive satellite missions: ATSR2-ERS2 (red) and AATSR-ENVISAT (green).*

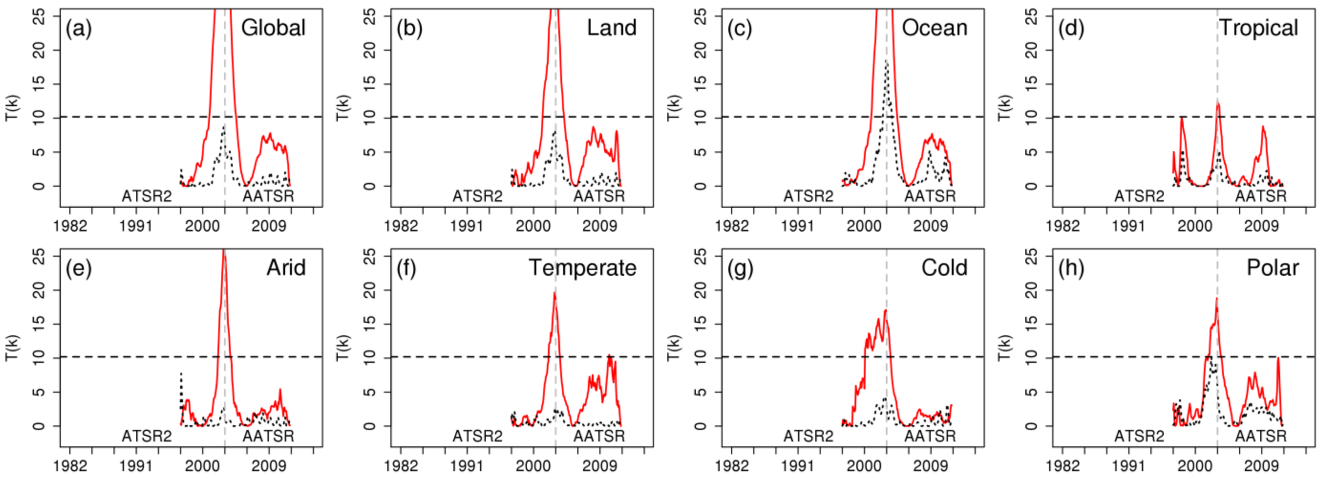


Figure 5-15 *Results of Standard Normal Homogeneity Tests (SNHT) for global (a), land (b), ocean (c) and climate zones (d-h). Relative SNHT applied to de-trended mean monthly cloud fraction difference between CC4CL-(A)ATSR and SYNOP (black dotted line). Absolute SNHT applied to de-trended CC4CL-(A)ATSR monthly cloud fraction anomalies for all grids (red solid line).The horizontal dashed line indicates a critical value of statistic T which signifies a break in the time series. The vertical dashed line shows a change in the satellite missions (ATSR2-ERS2, AATSR-ENVISAT).*

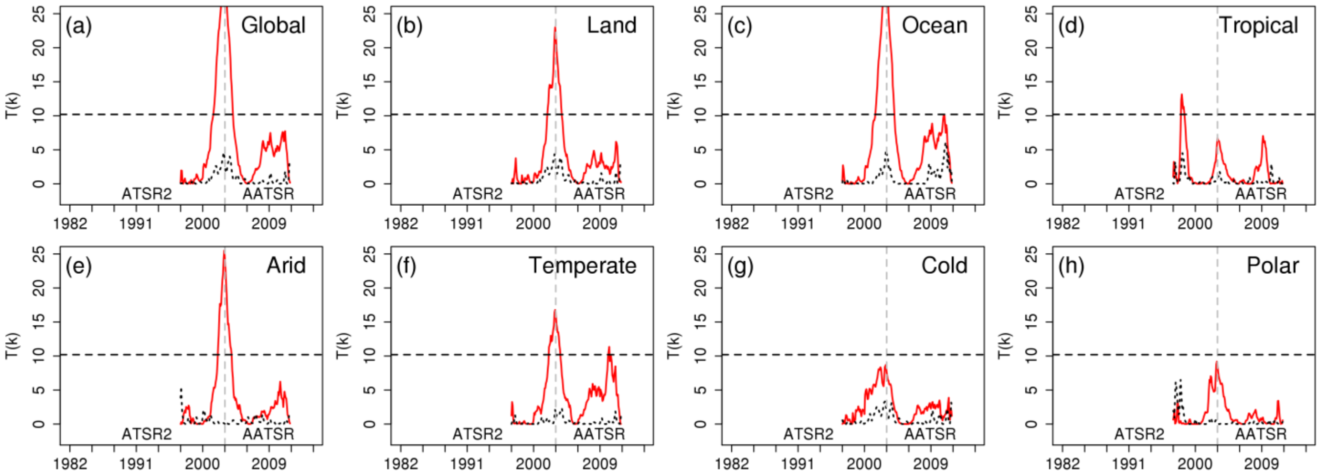


Figure 5-16 *As previous but applied to daytime (A)ATSR acquisitions only.*

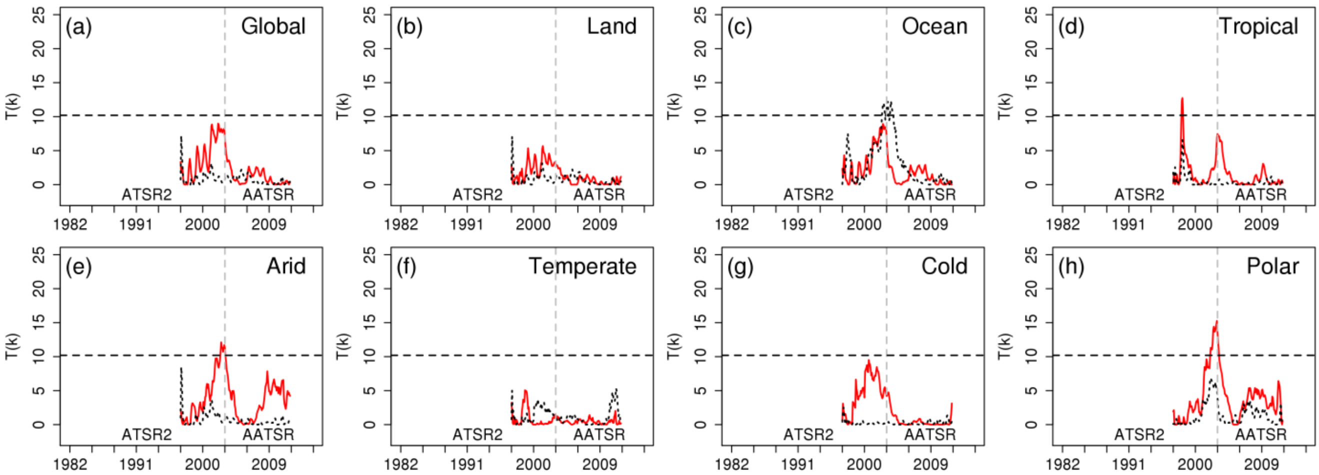


Figure 5-17 *As previous but applied to night-time (A)ATSR acquisitions only.*

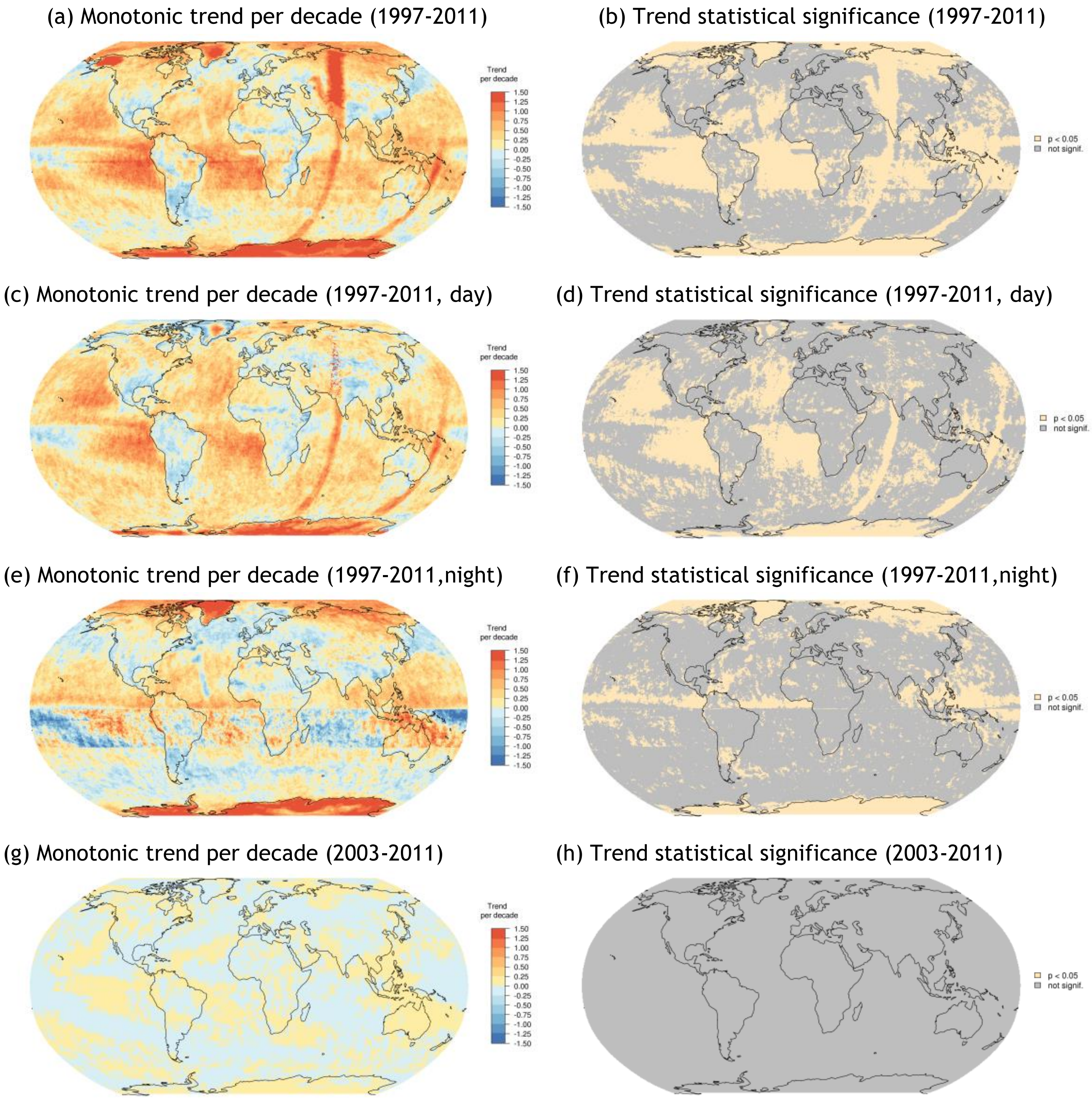


Figure 5-18 *Map of CC4CL-(A)ATSR Theil-Sen monotonic trend and its statistical significance according to the Mann-Kendall test adjusted using Benjamini-Hochberg method based on the cloud fraction monthly standardized anomalies in 1997-2011: all day (a-b), daytime (c-d), night-time (e-f), as well as all day but for AATSR only (g-h).*

5.4 Cloud_cci MODIS-Terra and MODIS-Aqua

CFC estimates from AQUA and TERRA are less vulnerable to inhomogeneities, as both sensors continuously operated during the covered years. Therefore, the results below are combined and shortened as compared to ones presented previously.

Both sensors reveal better performance at the Ocean than on Land, especially concerning bias (Table 5-7). Precision varies from below 8% for the Temperate zone to 12% in the Polar region.

The bias has a strong annual cycle, but on average remains stable over time. Only for the Temperate zone a negative trend in the bias is significant, -0.43% and -0.73% per decade for TERRA and AQUA, respectively. The bias time series are shown only for TERRA since there are very similar for AQUA (Figure 5-19).

Time series of bc-RMSE reveal some months of outlying low precision for the Ocean, e.g. in winter 2004 and 2008 for both AQUA (Figure 5-20c) and TERRA (Figure 5-21c). The latter also unveils high bc-RMSE at the beginning of the time series, both for Land and Ocean.

Homogeneity analysis does not detect any breaks in AQUA estimates, both for absolute and relative SNHT (Figure 5-23). Yet, SNHT identifies inhomogeneities in TERRA CFC for the Ocean, most probably due to errors in the beginning of the time series reflected in bc-RMSE.

Trends estimated for each grid show similar patterns for both AQUA and TERRA. For Polar and Greenland trends are significant, but only detected by TERRA. Interestingly, both sensors identify a significant positive trend (>1% per decade) over an Arctic Sea around 60˚E, as well as over the Aral Sea.

Table 5-7 *Bias and bias-corrected root mean square error (bc-RMSE) of CC4CL-TERRA and CC4CL-AQUA mean monthly cloud fraction as compared to synoptic observations aggregated for climate zones (columns).*

	Global	Land	Ocean	Tropical	Arid	Temperate	Cold	Polar
Bias (TERRA)	3.67	3.7	0.47	3.52	3.19	2.1	4.79	10.4
bc-RMSE (TERRA)	10.87	10.87	10.63	11.65	13.77	7.92	10.76	11.9
Bias (AQUA)	4.89	4.93	0.48	4.68	4.26	3.5	6.07	10.22
bc-RMSE (AQUA)	10.6	10.6	10.2	11.26	13.04	7.76	10.75	11.96

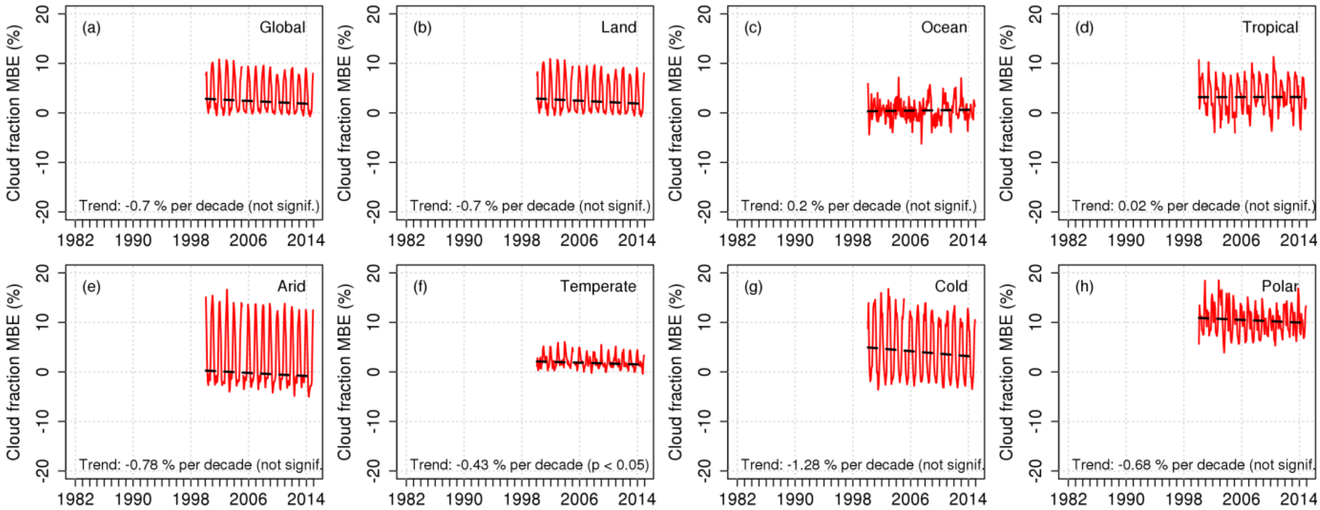


Figure 5-19 *Time series of CC4CL-TERRA bias (MBE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h). Theil-Sen linear trend (dashed line) and its Mann-Kendall statistical significance is also provided.*

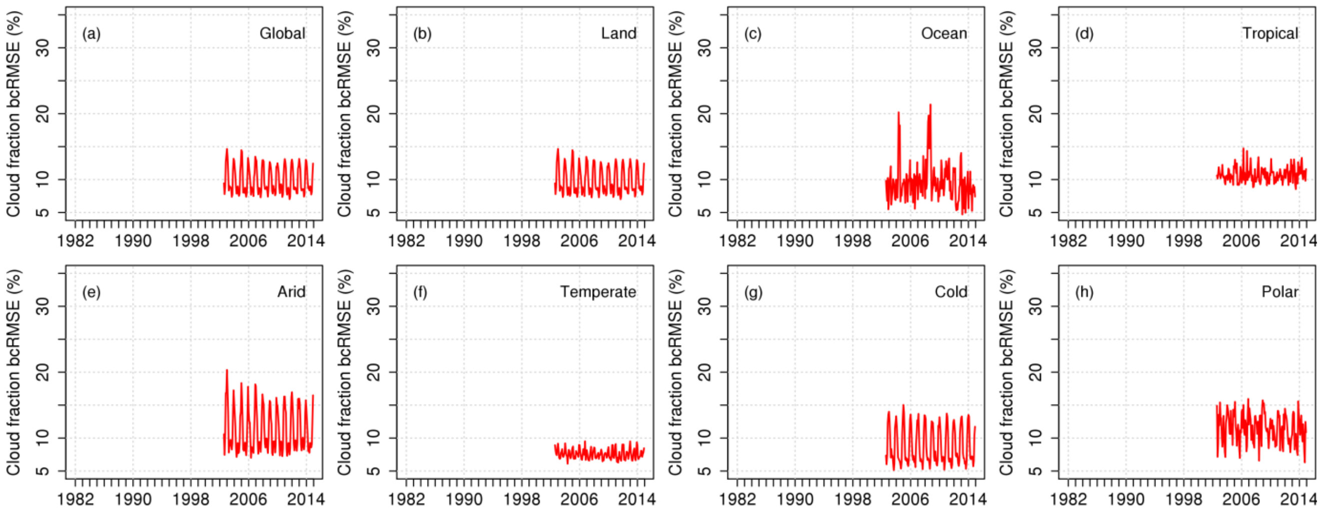


Figure 5-20 *Time series of CC4CL-AQUA bias-corrected root mean square error (bc-RMSE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h).*

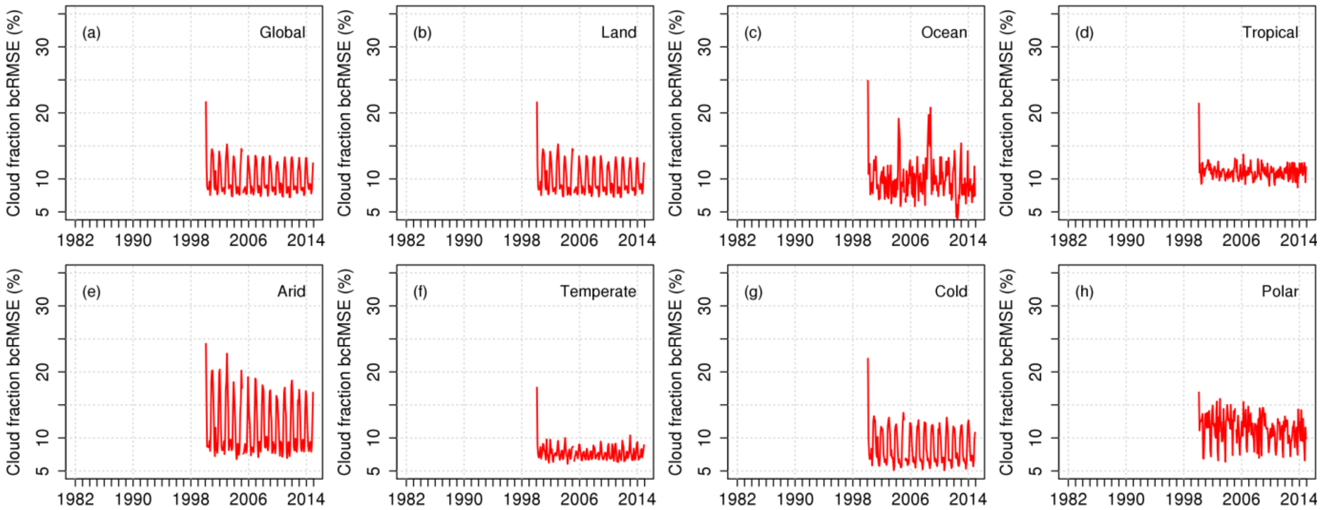


Figure 5-21 *Time series of CC4CL-TERRA bias-corrected root mean square error (bc-RMSE) aggregated for global (a), land (b), ocean (c) and climate zones (d-h).*

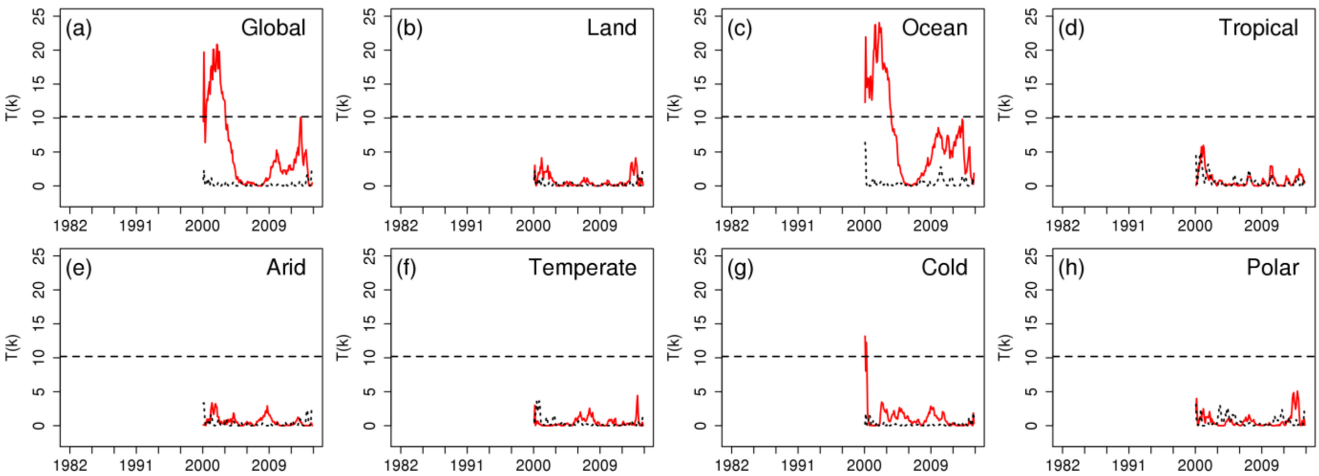


Figure 5-22 *Results of Standard Normal Homogeneity Tests (SNHT) for global (a), land (b), ocean (c) and climate zones (d-h). Relative SNHT applied to de-trended mean monthly cloud fraction difference between CC4CL-TERRA and SYNOP (black dotted line). Absolute SNHT applied to de-trended CC4CL-TERRA monthly cloud fraction anomalies for all grids (red solid line).The horizontal dashed line indicates a critical value of statistic T which signifies a break in the time series.*

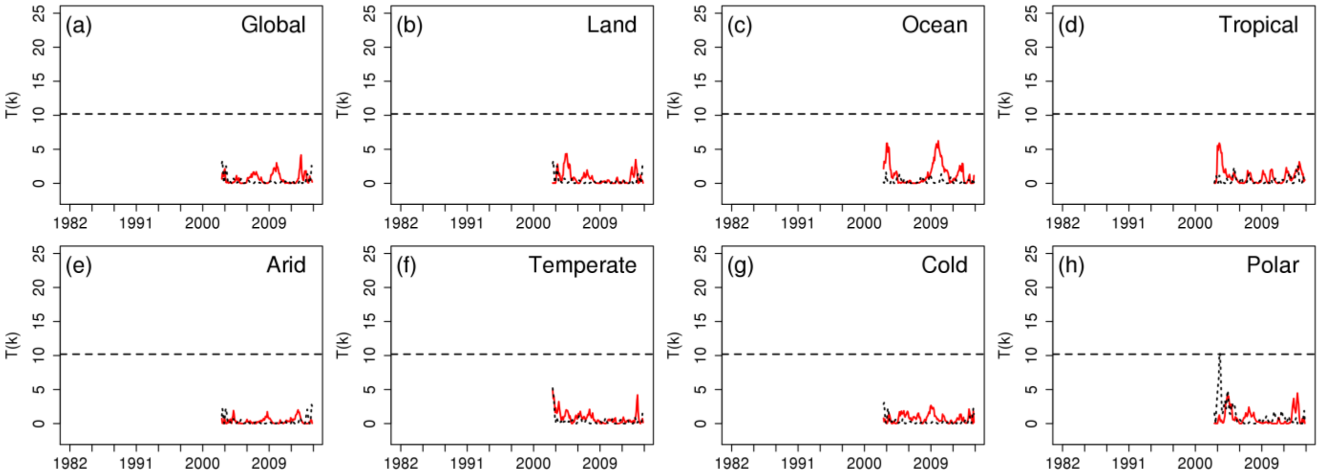
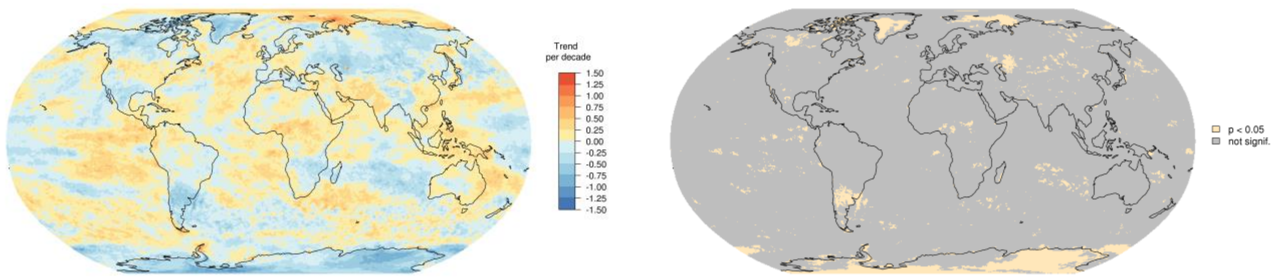


Figure 5-23 *Results of Standard Normal Homogeneity Tests (SNHT) for global (a), land (b), ocean (c) and climate zones (d-h). Relative SNHT applied to de-trended mean monthly cloud fraction difference between CC4CL-AQUA and SYNOP (black dotted line). Absolute SNHT applied to de-trended CC4CL-AQUA monthly cloud fraction anomalies for all grids (red solid line).The horizontal dashed line indicates a critical value of statistic T which signifies a break in the time series.*

(a) Monotonic trend per decade (TERRA) (b) Trend statistical significance (TERRA)



(c) Monotonic trend per decade (AQUA) (d) Trend statistical significance (AQUA)

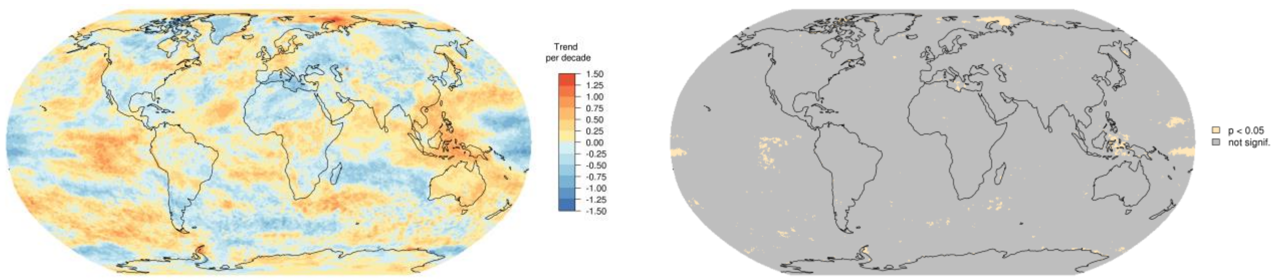


Figure 5-24 *Map of CC4CL-TERRA (2000-2014) and CC4CL-AQUA (2003-2014) Theil-Sen monotonic trend (a, c) and the statistical significance (b, d) according to the Mann-Kendall test adjusted using Benjamini-Hochberg method based on the cloud fraction monthly standardized anomalies.*

6. Uncertainty validation

In this section the pixel-based random uncertainties associated with the OE-based cloud properties CTP, LWP and IWP are validated. Mandatory for such a validation is a very accurate reference source. For CTP, the CALIOP data is used, for LWP an AMSR-E product and for IWP the DARDAR product. All Cloud_cci data used in this section are from the AVHRR-PM dataset, which is expected to perform similar to all other datasets in terms of pixel-based uncertainty characterizations. For the cloud properties CER and COT no reliable reference data source is available which would allow a similar uncertainty validation as presented for CTP, LWP and IWP. However, as LWP and IWP (and their uncertainties) are derived from CER and COT retrievals (and their uncertainties) the uncertainty validation for LWP and IWP are considered to implicitly contain a CER and COT uncertainty validation. For cloud mask CMA the best reference data (CALIOP) has been used to statistically determine the CMA uncertainty, thus an additional CMA uncertainty validation is neither possible nor needed.

The reported pixel-based uncertainties X_{unc} for a given variable X (both being a result of the optimal estimation technique) represent the 68% confidence interval that the true value is within $X \pm X_{unc}$. Given this, it can be assumed that for 68% of all pixels the truth is within $X \pm X_{unc}$. We will use this rationale in the following to analyse a large set of collocated Cloud_cci and reference observations for the validation of the uncertainties. As the reference data are also not perfect, the reference data uncertainty is also considered in this exercise. In Section 6.1 the CTP is considered (separately for liquid and ice clouds: Section 6.1.1 and 6.1.2, respectively) and Sections 6.2 and 6.3 give the results for LWP and IWP, respectively.

It needs to be noted that the presented framework considers random uncertainties only. For this reason, occurring systematic biases between Cloud_cci and the reference data are removed if significant.

In all equations of this Section i is used as pixel index, d for the pixel-based difference between Cloud_cci and the reference, u for the pixel-based total uncertainty (consisting of Cloud_cci uncertainty and reference data uncertainty) and β as pixel-based ratio between the absolute value of d and u.

6.1 Cloud top pressure

As mentioned above CALIOP data is used as best estimate of the truth for CTP. Unfortunately, no uncertainty measure exists for the CALOP CTP. For this reason we set all CALIOP CTP uncertainties to 10hPa arbitrarily.

Equations 6-1 to 6-3 list the calculations of d, u and β as defined above. The CTP uncertainty (CTP_unc) validation is separated into liquid and ice clouds, for which both phase retrievals (Cloud_cci and CALIOP) had to agree.

$$d_i = CTP_i^{cci} - CTP_i^{caliop}$$ **Equation 6-1**

$$u_i = \sqrt{\left(CTP\_unc_i^{cci}\right)^2 + \left(CTP\_unc_i^{caliop}\right)^2}$$ **Equation 6-2**

$$\beta_i = \frac{|d_i|}{u_i}$$ **Equation 6-3**

6.1.1 Liquid clouds

Figure 6-1 shows boxplots for CTP_unc^{cci} and CTP_unc^{caliop} as function of CTP. The majority of the CTP_unc^{cci} does now show a dependence on the CTP^{cci} retrieval itself and lies between 10 and 50 hPa. Exceptions are the low-level clouds below 850hPa for which the median of CTP_unc^{cci} becomes smaller, along with an increase in spread. Few outliers exist in each bin with very small or very high uncertainties. As mentioned above no CTP_unc^{caliop} information are provided along with CTP^{caliop}, which were in turn set to 10 hPa in this exercise (see right hand side of Figure 6-1).

A small systematic bias is found between CTP^{cci} and CTP^{caliop} for liquid clouds. After analysing this, 12 hPa is subtracted from CTP^{cci} data to centre the frequency distribution of d around 0 (not shown). From here onwards, we consider all remaining deviations to be of random nature only.

Figure 6-2 shows a 2-dimensional histogram over all CTP^{cci} , CTP^{caliop} pairs for all liquid cloud pixels used. In addition a histogram over β is given. Ideally, the portion of all cases that have a β value of 1 or below (β1-portion hereafter), should amount to 68%. Thus 68% should fall into the first bin. However, the histogram only reports 42% with $\beta \leq 1$. In the second bin ($1 < \beta \leq 2$) about 23% are found. The numbers decrease further with increasing bin. This investigation suggests too small values for CTP_unc^{cci} given that the assumed CTP_unc^{caliop} is not unrealistically small. In the following it is investigated if the appropriateness of CTP_unc^{cci} is a function of CTP itself. For this, the portion of cases with $\beta \leq 1$ with respect to all cases within certain CTP bins is investigated (Figure 6-3 to Figure **6-5**).

Two features are prominent in Figure 6-3: (1) the increase of the β1-portion for high clouds reaching values near 80% (although there only few liquid clouds at these altitudes), and (2) the decrease for low clouds, reaching values below 40% for the lowest clouds (high CTPs). The second feature is significant in a sense that it represents a large set of pixels (see right hand side of Figure 6-3). Looking at the results stratified for CTP^{caliop} bins (Figure 6-4), the β1-portion is nearly 40% for almost all clouds. Exception here is that for the (very few) high level liquid clouds, the β1-portion goes down. Figure 6-5 only reveals that the β1-portion is large enough when the differences between CTP^{cci} and CTP^{caliop} are small anyway.

Table 6-1 reports correlation coefficients of d with (a) CTP_unc^{cci}, (b) with the combined uncertainty u and (c) with β. The correlation coefficients for (a) and (b) are very close to zero. Correlation (c) is 0.83.

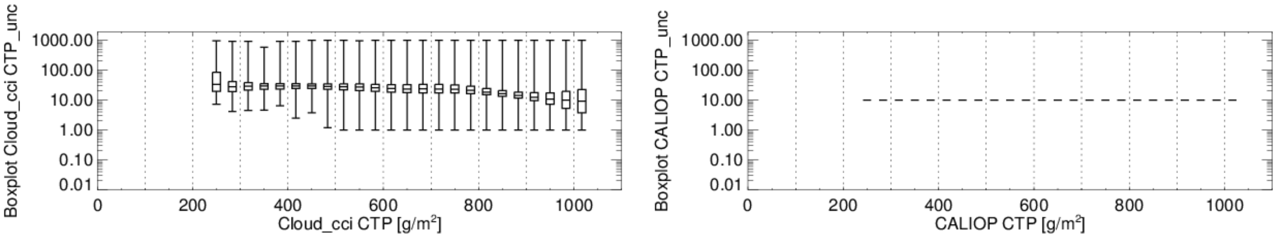


Figure 6-1 *Boxplot of CTP uncertainties (CTP_unc) shown as function of CTP for Cloud_cci (left) and CALIOP (right) for* ***liquid clouds****. Boxes reflect minimum and maximum as well as 25%, 50% and 75% percentiles of CTP_unc distribution in each CTP bin. As no CTP uncertainty is available to CALIOP, all values have been set to 10hPa. CTP uncertainties below 1hPa or above 1000hPa where omitted.*

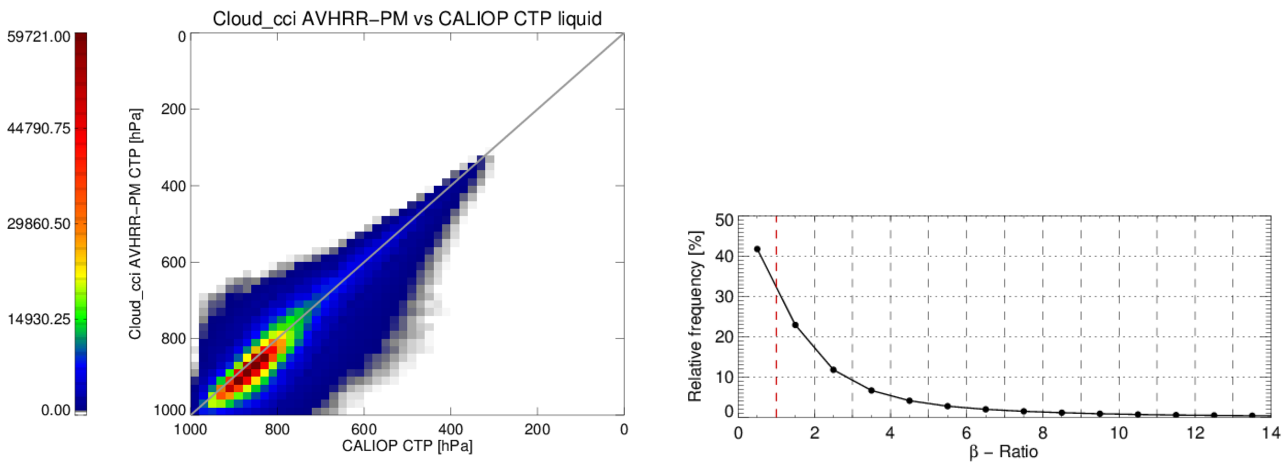


Figure 6-2 *Left: 2-dimensional frequency histogram of collocated Cloud_cci and CALIOP CTP pairs for* ***liquid clouds*** *before applying bias correction. Right: Histogram of β-ratio (as defined in Equation 6-3) over all collocations. Vertical dashed lines reflect bin borders (bin width is 1).*

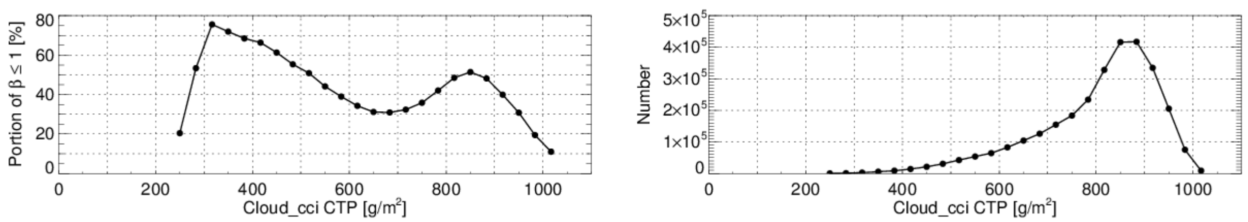


Figure 6-3 *Left: Portion of all cases for which $\beta \leq 1$ (β1-portion) shown as function of Cloud_cci CTP for* ***liquid clouds****. Right: Corresponding number of pixels per Cloud_cci CTP bin.*

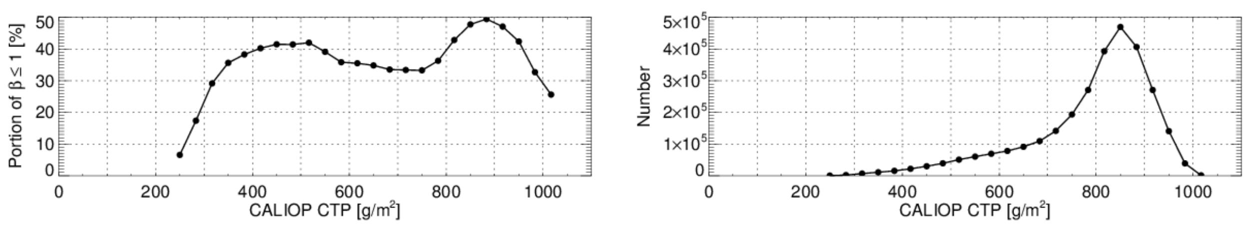


Figure 6-4 *As Figure 6-3 but as function of CALIOP CTP.*

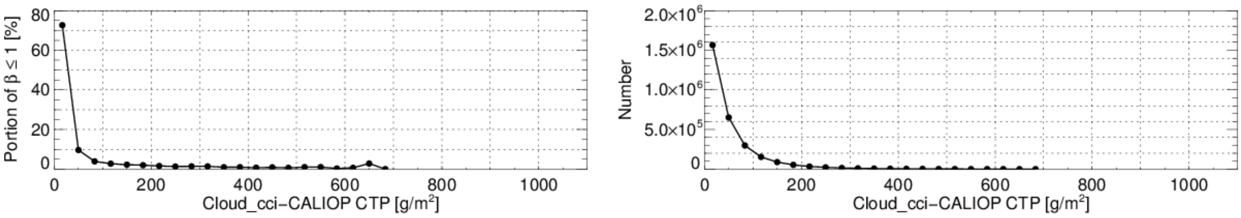


Figure 6-5 *As Figure 6-3 but as function of the absolute difference Cloud-cci CTP – CALIOP CTP.*

Table 6-1 *Correlation coefficients for liquid clouds. See text for details.*

(a)	Correlation ($\lvert d\rvert$, CTP_unc^{cci})	0.074
(b)	Correlation ($\lvert d\rvert$, u)	0.075
(c)	Correlation ($\lvert d\rvert$, β)	0.833

6.1.2 Ice clouds

Figure 6-6 shows boxplots for CTP_unc^{cci} and CTP_unc^{caliop} as function of CTP. The majority of CTP_unc^{cci} does now show a strong dependence on the CTP^{cci} retrieval itself and lies between 10 and 50 hPa. Exceptions are firstly the low-level ice clouds below 750hPa for which the spread becomes larger. Secondly, for high level ice clouds with CTPs below 200 hPa the CTP_unc^{cci} becomes larger. Few outliers exist with very small or very high uncertainties. As mentioned above for liquid clouds, no CTP_unc^{caliop} are available, which were also set to 10hpa in this exercise for ice clouds.

A moderate systematic bias is found between CTP^{cci} and CTP^{caliop} for ice clouds. After analysing the frequency distribution of d, 60 hPa is subtracted from CTP^{cci} data to centre the distribution around 0 (not shown). From here onwards we consider all remaining deviations to be of random nature only.

Figure 6-7 shows a 2-dimensional histogram over all CTP^{cci} , CTP^{caliop} pairs for all ice cloud pixels used. In addition a histogram over β is given. The histogram only reports 34% with $\beta \leq 1$. In the second bin ($1 < \beta \leq 2$) about 17% are found. These values are even lower than for liquid clouds. This investigation suggests too small values for CTP_unc^{cci} also for ice clouds given that the assumed CTP_unc^{caliop} is not unrealistically small. In the following it is investigated if the appropriateness of CTP_unc^{cci} is a function of CTP^{cci} itself. For this, the portion of cases with $\beta \leq 1$ with respect to all cases within certain CTP bins is investigated (Figure 6-8 to Figure **6-10**).

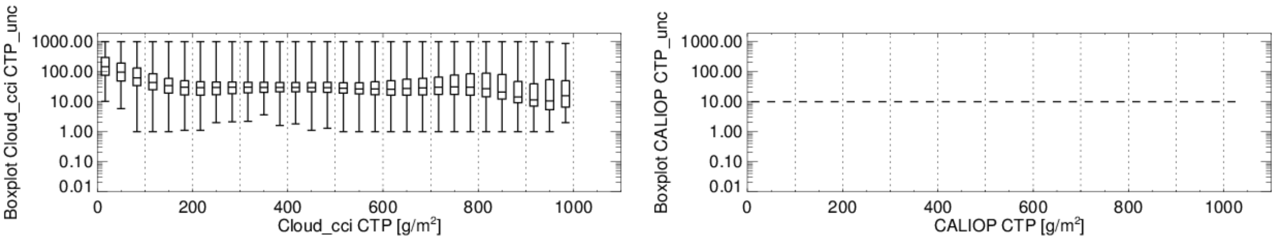


Figure 6-6 *Boxplot of CTP uncertainties (CTP_unc) shown as function of CTP for Cloud_cci (left) and CALIOP (right) for* ***ice clouds****. Boxes reflect minimum and maximum as well as 25%, 50% and 75% percentiles of CTP_unc distribution in each CTP bin. As no CTP uncertainty is available for CALIOP, all values have been set to 10hPa.*

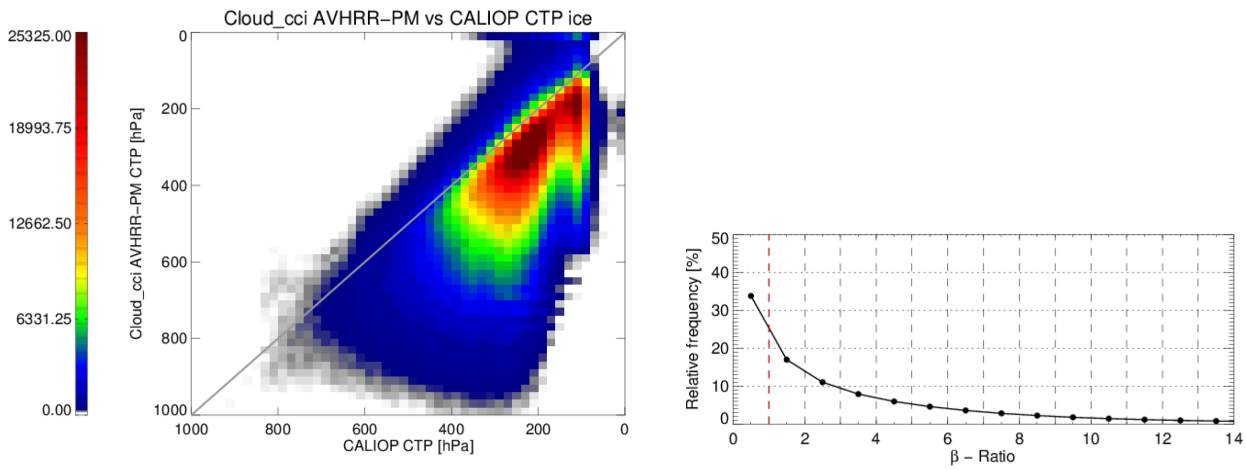


Figure 6-7 *Left: 2-dimensional frequency histogram of collocated Cloud_cci and CALIOP CTP pairs for* ***ice clouds****. Right: Histogram of Β-ratio (as defined in Equation 6-3) over all collocations. Vertical dashed lines reflect bin borders (bin width is 1).*

In Figure 6-8, except from the very high ice clouds, for which the β1-portion is 60% or above, for most parts of the retrieved CTP spectrum the CTP_unc are too low.

Looking at the results stratified for CALIOP CTP bins (Figure 6-9), the β1-portion is around 30% for almost all clouds. As for liquid cloud the β1-portion is large enough when the differences between Cloud_cci and CALIOP small (Figure 6-10).

Table 6-2 reports correlation coefficients of d with (a) CTP_unc^{cci}, (b) the combined uncertainty u and (c) β. As for liquid clouds, also for ice clouds the correlation coefficients for (a) and (b) are very close to zero. Correlation (c) is 0.69.

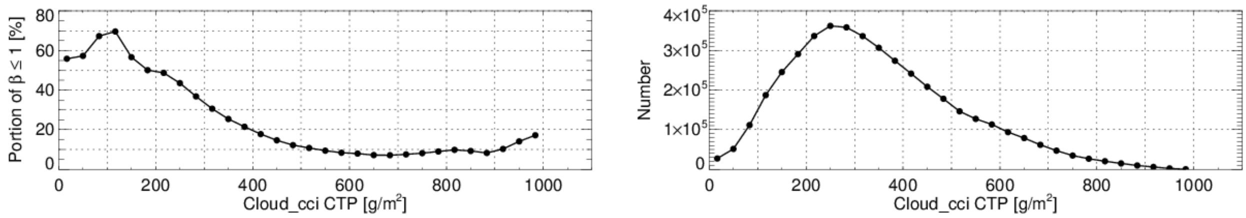


Figure 6-8 *Left: Portion of all cases for which $\beta \leq 1$ (β1-portion) shown as function of Cloud_cci CTP for* ***ice clouds****. Right: Corresponding number of pixels per Cloud_cci CTP bin.*

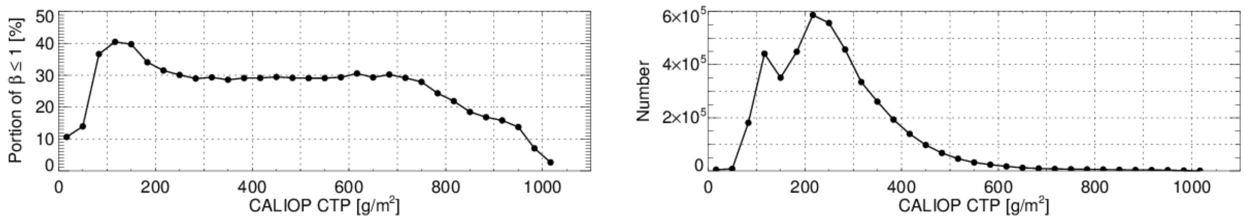


Figure 6-9 *As Figure 6-8 but as function of CALIOP CTP.*

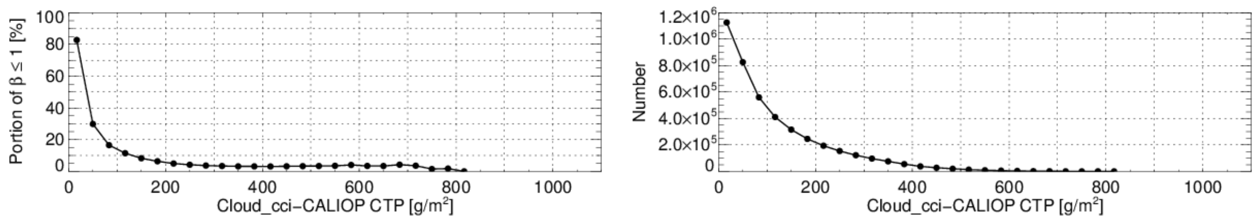


Figure 6-10 *As Figure 6-8 but as function of the absolute difference Cloud-cci CTP – CALIOP CTP.*

Table 6-2 *Correlation coefficients for ice clouds. See text for details.*

(a)	Correlation ($\|d\|$, CTP_unc^{cci})	-0.065
(b)	Correlation ($\|d\|$, u)	0.15
(c)	Correlation ($\|d\|$, β)	0.696

6.2 Liquid water path

As mentioned in the beginning of Section 6, AMSR-E LWP data (Wentz and Meissner, 2004) is used as best estimate of the LWP truth. There is no pixel-based uncertainty reported for AMSR-E LWP. Instead we use 0.017mm, which is reported as the overall root mean square error in Table 9 of Wentz and Meissner (2002), as pixel-based uncertainty measure. In the discussion below we will tackle what the impact of this compared to no uncertainty or compared to an assumed relative uncertainty is.

Only a very small systematic bias is found between LWP^{cci} and LWP^{amsre} for liquid clouds. As the frequency distribution of d was already centred around 0 (not shown), no further bias correction was applied and all remaining deviations to be of random nature only.

Equations 6-4 to 6-6 list the calculations of d, u and β as defined above. For LWP uncertainty (LWP_unc) validation all cases for which either Cloud_cci or AMSR-E reported an ice cloud top were excluded.

$$d_i = LWP_i^{cci} - LWP_i^{amsre}$$ **Equation 6-4**

$$u_i = \sqrt{\left(LWP\_unc_i^{cci}\right)^2 + \left(LWP\_unc_i^{amsre}\right)^2}$$ **Equation 6-5**

$$\beta_i = \frac{|d_i|}{u_i}$$ **Equation 6-6**

Figure 6-11 shows boxplots for LWP_unc^{cci} and LWP_unc^{amsre} as function of LWP. For LWP larger than 30g/m² the LWP_unc^{cci} grows exponentially with LWP^{cci} (near-linear relationship in logarithmic plot). As we have set a fixed uncertainty (see above) for LWP_unc^{amsre}, no sensitivity to the AMSR-E LWP exists (right hand side of Figure 6-11).

Figure 6-12 shows a 2-dimensional histogram over all LWP^{cci} , LWP^{amsre} pairs for all liquid cloud pixels available in the collocation set. In addition a histogram over β is given. As mentioned above, the β1-portion should ideally amount to 68%. Thus 68% should fall into the first bin. However, the histogram only reports about 52% with $\beta \leq 1$. However, assigning no error to AMSR-E gives 21%, assigning relative errors of 15% to each AMSR-E footprint gives 25% (not shown). In the second bin ($1 < \beta \leq 2$) about 30% are found. The number decrease further with increasing bin. This investigation suggests too small values for LWP_unc^{cci} given that the LWP_unc^{amsre} correctly presents the 68% confidence level itself. In the following it is investigated if the appropriateness of LWP_unc^{cci} is a function of LWP^{cci} itself. For this, the potion of cases with $\beta \leq 1$ with respect to all cases within certain LWP bins (Figure 6-13 to Figure **6-15**).

Figure 6-13 shows that for very large LWP^{cci} the LWP_unc^{cci} seems to have the right amplitude as the β1-portion increases to values even above 80%. However, for LWP^{cci} values below 1000g/m³ (which are the vast majority of all cases) the LWP_unc^{cci} seems to be too small. These results however become a bit inconclusive because Figure 6-14 indicates that for small LWP^{amsre} the LWP_unc^{cci} seems appropriate. Figure 6-15 reveals that the β1-portion is large enough when the differences between Cloud_cci and AMSR-E are small (below 20g/m²).

Table 6-3 reports correlation coefficients of the LWP differences (Cloud_cci minus AMSR-E) with (a) the LWP_unc^{cci}, (b) with the combined uncertainty (Cloud_cci and AMSR-E) and (c) with β. The correlation coefficients for (a) and (b) are around 0.88. Correlation (c) is lower (0.34).

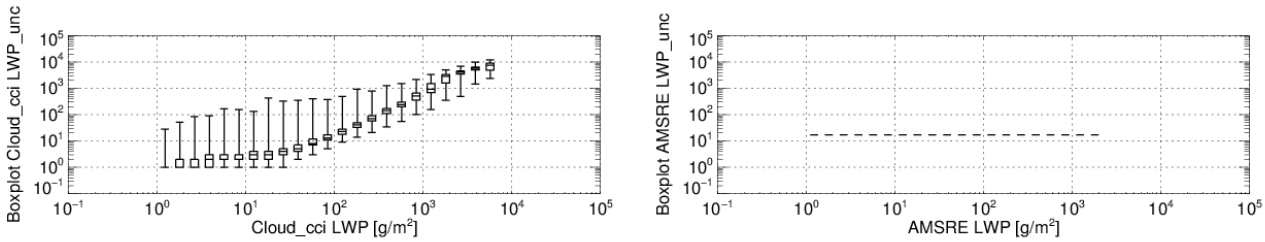


Figure 6-11 *Boxplot of IWP uncertainties (LWP_unc) shown as function of IWP for Cloud_cci (left) and AMSR-E (right). Boxes reflect minimum and maximum as well as 25%, 50% and 75% percentiles of IWP_unc distribution in each IWP bin.*

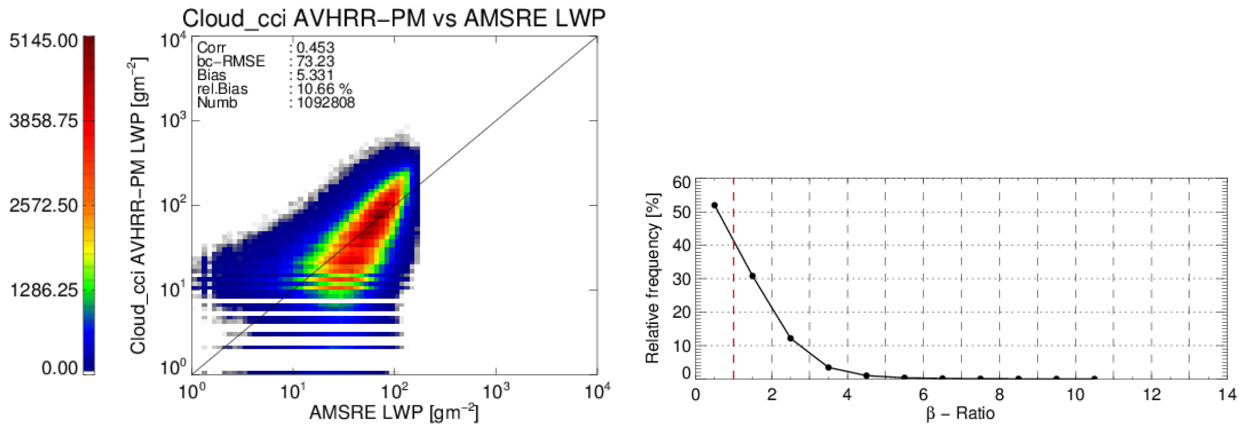


Figure 6-12 *Left: 2-dimensional frequency histogram of LWP^{cci}, LWP^{amsre} pairs (about 1 million collocated pixels). Right: Histogram of β-ratio (as defined in Equation 6-6) over all collocations. Vertical dashed lines reflect bin borders (bin width is 1).*

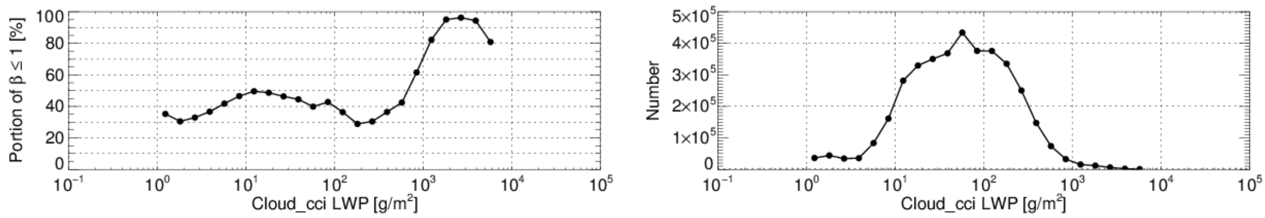


Figure 6-13 *Left: Portion of all cases for which* $\beta \leq 1$ (β1-portion) *shown as function of Cloud_cci LWP. Right: Corresponding number of pixels per Cloud_cci LWP bin.*

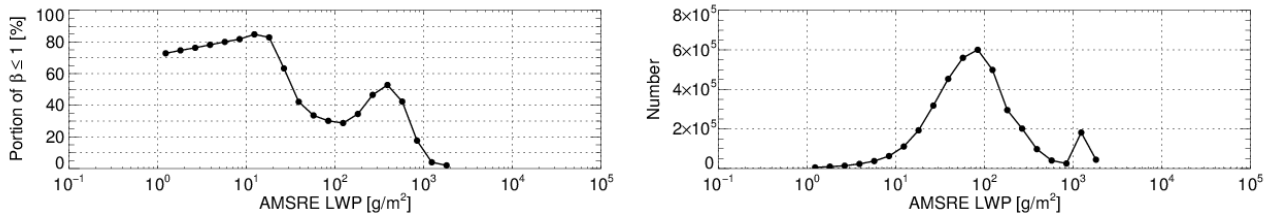


Figure 6-14 *As Figure 6-18 but as function of AMSR-E LWP.*

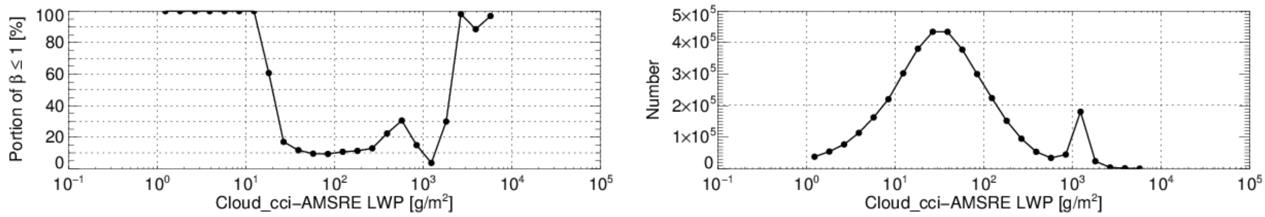


Figure 6-15 *As Figure 6-18 but as function of the difference Cloud-cci LWP – AMSR-E LWP.*

Table 6-3 *Correlation coefficients. See text for details.*

(a)	Correlation ($\|d\|$, LWP_unc^{cci})	0.886
(b)	Correlation ($\|d\|$, u)	0.879
(c)	Correlation ($\|d\|$, β)	0.336

6.3 Ice water path

As mentioned in the beginning of Section 6, DARDAR IWP data (IWP^{dardar}) is used as best estimate of the truth. IWP^{dardar} values are results of an variational scheme (Delanoë and Hogan, 2008,2010) for retrieving ice cloud properties from combined radar CloudSat, lidar (CALIOP), and infrared radiometer (e.g. MODIS) and come along with a pixel based uncertainty measure similar to Cloud_cci.

Only a very small systematic bias is found between IWP^{cci} and IWP^{dardar}. As the frequency distribution of d $w$as already centred around 0 (not shown), no further bias correction was applied and all remaining deviations to be of random nature only.

Equations 6-7 to 6-9 list the calculations of d, u and β as defined above. For IWP uncertainty (IWP_unc) validation all cases for which either Cloud_cci or DARDAR reported a liquid cloud top were excluded.

$$d_i = IWP_i^{cci} - IWP_i^{dardar}$$ **Equation 6-7**

$$u_i = \sqrt{\left(IWP\_unc_i^{cci}\right)^2 + \left(IWP\_unc_i^{dardar}\right)^2} \quad \textbf{Equation 6-8}$$

$$\beta_i = \frac{|d_i|}{u_i} \quad \textbf{Equation 6-9}$$

Figure 6-16 shows boxplots for IWP_unc^{cci} and IWP_unc^{dardar} uncertainty as function of IWP. For IWP larger than 30g/m² the IWP_unc^{cci} grows exponentially with IWP^{cci} (near-linear relationship in logarithmic plot). Right hand side of Figure 6-16 shows the same for IWP_unc^{dardar} and IWP^{dardar}, indicating similarities to Cloud_cci, with the spread however being lower than for Cloud_cci. It is also important to note that IWP_unc^{dardar} are about one order of magnitude smaller than the retrieved IWP^{dardar}, whereas for Cloud_cci it seems that the IWP_unc^{cci} is nearly as large as the retrieved IWP^{cci}.

Figure 6-17 shows a 2-dimensional histogram over all IWP^{cci} , IWP^{dardar} pairs for all ice cloud pixels used. In addition a histogram over β is given. As mentioned above, the β1-portion should ideally amount to 68%. Thus 68% should fall into the first bin. However, the histogram only reports 36% with $\beta \leq 1$. In the second bin ($1 < \beta \leq 2$) about 24% are found. The number decrease further with increasing bin. This investigation suggests too small values for IWP_unc^{cci} given that the IWP_unc^{dardar} correctly presents the 68% confidence level itself. In the following it is investigated if the appropriateness of IWP_unc^{cci} is a function of IWP itself. For this, the potion of cases with $\beta \leq 1$ with respect to all cases within certain IWP bins (Figure 6-18 to Figure **6-20**).

Figure 6-18 shows that for large IWP^{cci} values the IWP_unc^{cci} seems to have the right amplitude as the β1-portion increases to values above 60%. However, for IWP values below 1000g/m³ the IWP_unc^{cci} seems to be too small. Figure 6-19 shows on the one hand a similar peak for very high IWP^{dardar}, on the other hand however another peak for β1-portion for IWP^{dardar} between 100 and 1000g/m². Figure 6-20 reveals that the β1-portion is large enough when the differences between Cloud_cci and CALIOP are small. However, and this is again in contrast to CTP, also for large IWP differences a significant portion of the data is correctly characterized by the IWP_unc , although by far not reaching the envisaged values of 68% or above.

Table 6-4 reports correlation coefficients of the IWP differences (IWP^{cci} - IWP^{dardar}) with (a) the IWP_unc^{cci} (b) with the combined uncertainty (IWP_unc^{cci} and IWP_unc^{dardar}) and (c) with β. The correlation coefficients for (a) and (b) are around 0.5. Correlation (c) is low (0.07).

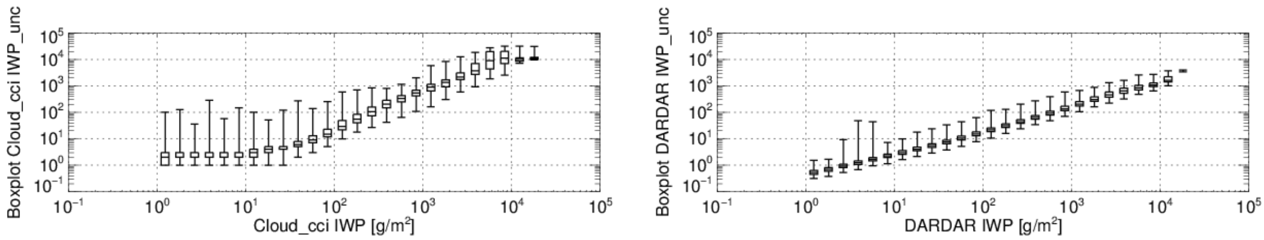


Figure 6-16 *Boxplot of IWP uncertainties (IWP_unc) shown as function of IWP for Cloud_cci (left) and DARDAR (right). Boxes reflect minimum and maximum as well as 25%, 50% and 75% percentiles of IWP_unc distribution in each IWP bin.*

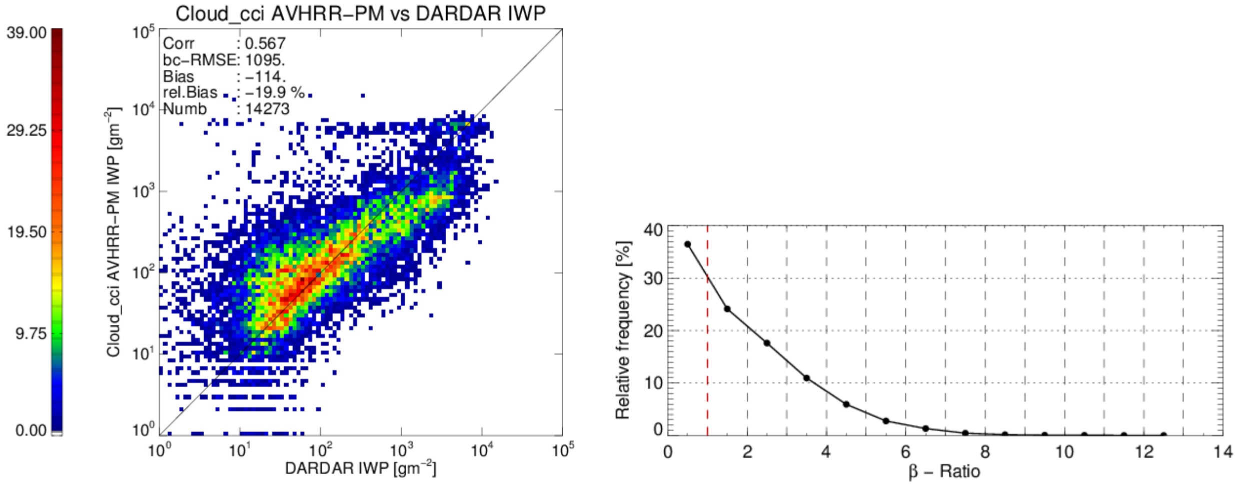


Figure 6-17 *Left: 2-dimensional frequency histogram of Cloud_cci and DARDAR IWP pairs for about 14000 collocated pixels. Right: Histogram of β-ratio (as defined in Equation 6-9) over all collocations. Vertical dashed lines reflect bin borders (bin width is 1).*

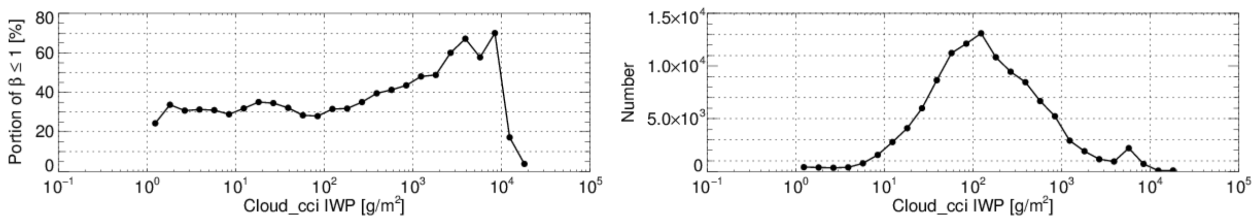


Figure 6-18 *Left: Portion of all cases for which* $\beta \leq 1$ *(*β*1-portion) shown as function of Cloud_cci IWP. Right: Corresponding number of pixels per Cloud_cci IWP bin.*

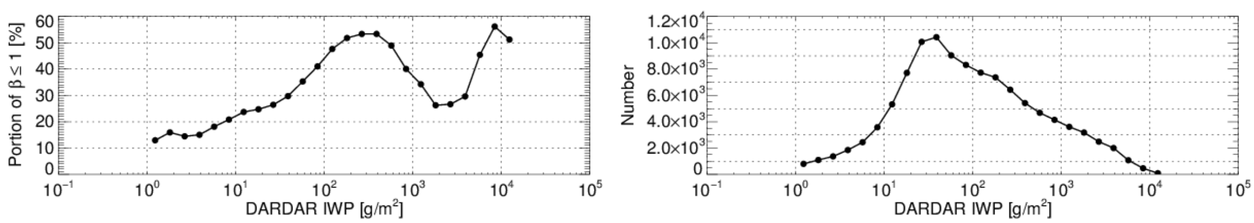


Figure 6-19 *As Figure 6-18 but as function of DARDAR IWP.*

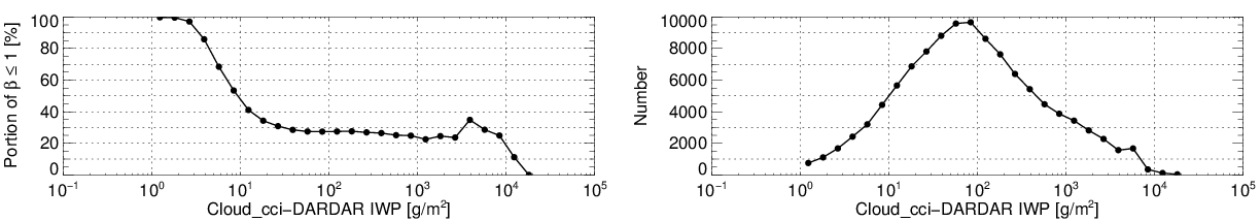


Figure 6-20 *As Figure 6-18 but as function of the absolute difference Cloud-cci IWP - DARDAR IWP.*

Table 6-4 *Correlation coefficients. See text for details.*

(a)	Correlation ($\|d\|$, IWP_unc^{cci})	0.476
(b)	Correlation ($\|d\|$, u)	0.501
(c)	Correlation ($\|d\|$, β)	0.073

6.4 Summary of the uncertainty validation

In the previous three subsections, the Cloud_cci CTP, LWP and IWP uncertainties were validated using very accurate reference observations of CTP, LWP and IWP and, if existent, their uncertainties. As the OE uncertainty describes the interval around the retrieved value in which the truth can be found with a confidence of 68%, we determined the number or cases for which the uncertainty covered the difference to the reference data as portion from all cases.

For CTP, LWP and IWP the found portion was significantly too low (between 32 and 52%) highlighting the Cloud_cci uncertainties being generally too low. This probably means that not all uncertainty sources are (correctly) included/modelled yet. One example is the AVHRR measurement uncertainty, which is likely to be larger than currently specified.

Further studies revealed that for CTP nearly no correlation between the uncertainties and the actually difference to the estimated truth could be found, which indicates the CTP uncertainty being not a good measure of the actual uncertainty of the retrieved CTP value. In contrast, the results for LWP and IWP uncertainties indicate their utility, as the correlation between LWP/IWP uncertainty and the LWP/IWP difference to the truth was around 0.88/0.5, although the absolute values of the LWP and IWP uncertainties were too low as well. Conducting some further tests, we found that multiplying the Cloud_cci IWP uncertainty by a factor of 3 raises the β1-portion to about 66%, with correlations (a), (b) and (c) (see above for definitions) nearly remaining constant.

It needs to be noted that the uncertainties of the reference data also have an important impact on the results inferred. Having these specified too low (by us or the providers of the reference data), could also explain a significant portion of the results.

7. Summary

In this document, the evaluation results for Cloud_cci v2.n0 datasets are presented. In Section 2 a general overview of Cloud_cci datasets were given and the derived cloud properties were introduced together with a discussion on the available processing levels.

In Section 3 the results of validating Cloud_cci datasets against benchmark reference data from CALIOP (a space-based Lidar), SYNOP (human cloud cover observations) and passive microwave observations of LWP were shown and discussed.

Cloud mask (CMA) validation against CALIOP proves the good cloud detection performance of CC4CL and FAME-C with hitrates generally higher than 75%, for most datasets comparisons even higher than 80%. For nearly all comparisons a slight underestimation of cloud occurrences is found in the Cloud_cci data compared to CALIOP, which is primarily due to a lack of sensitivity of passive imager data with respect to optically very thin clouds. Removing these clouds from the statistic, which is done by using CALIOP cloud optical thickness (COT), shows a clear improvement in the found detection scores. Major weaknesses were found for example for MODIS-Terra (large overestimation of cloud occurrence in the high latitudes and CC4CL-based twilight cloud detection.

Cloud phase (CPH) validation against CALIOP shows hitrates of 70% and higher when using the CALIOP phase of the uppermost detected cloud layer. Neglecting the very thin cloud layers (with COT lower than 0.15) improves the scores significantly, now being around 80%, with the exception of AVHRR-AM and MERIS-AATSR for which the scores increase only to approximately 75%. All datasets have a bias towards liquid clouds, when compared against to uppermost layer. The bias reduced for the scenario of neglecting the very thin cloud layer. Exceptions here are ATSR2-AATSR and MERIS-AATSR which show always an artificial bias towards ice clouds.

Cloud top height retrievals were also validated against CALIOP. Common for all Cloud_cci datasets is the strong underestimation of cloud top height for high, optically thin clouds. For these clouds, biases of -4000m can occur. Removing the optically very thin cloud layers at the top of the CALIOP profiles, improves the agreement between Cloud_cci and CALIOP significantly. For AVHRR-PM for example, the bias with respect to CALIOP CTH nearly disappears when removing all cloud layers with COTs (integrated from layer bottom to top of all cloud layers) lower than 1.0. For CC4CL and the FAME-C MERIS-CTP (O_2 A-band based) show very good performance for very low clouds with biases usually smaller than 230m.

Cloud_cci monthly mean cloud fraction data (CFC) were validated against equivalent values from SYNOP observations. It could be shown, that all datasets reveal a high stability of the bias compared to SYNOP throughout the time period covered. An exception here is AVHRR-AM, which clearly reveals shortcomings for AVHRR on board NOAA12, the satellite that that was used for 1991 to 1999 for AVHRR-AM. As the SYNOP data was purposely not subsampled to the satellite overpass time, the comparisons include sampling errors. From comparing the biases for AVHRR-PM and MODIS-Aqua with AVHRR-AM, ATSR2-AATSR and MERIS-AATSR it seems that the morning satellites do provide a better estimate of the climatological mean (24-hour mean) than the afternoon satellites.

Cloud_cci monthly mean allsky liquid water path (LWP) was validated against satellite-based passive microwave data with focus on the three stratocumulus regions. The results are relatively diverse among the regions and datasets. For Cloud_cci AVHRR-PM and MODIS-Aqua the LWP is in good agreement with UWISC throughout the time periods with only slight negative biases for the early NOAA satellites. For AVHRR-AM and ATSR2-AATSR there are some problems before 2002 that can surely be related to the difficult illumination conditions of NOAA-12 and NOAA15 (early morning orbit) and problems with inconsistencies in 3.7 channels for ATSR2, compared to AATSR. Both issues are however less pronounced for the Southern Atlantic stratocumulus regions.

In Section 4 Cloud_cci L3C products were compared against each other and confronted with well-established, existing datasets ('reference datasets' hereafter) of the same kind (CLARA-A2, PATMOS-x, MODIS Collection 6). These comparisons are used for plausibility studies as none of the used external datasets can be considered as truth. The comparisons were separated into morning orbit satellites and afternoon orbit satellite, with morning and afternoon referring to the local observation time of the daytime node of the orbits.

The results show that Cloud_cci datasets are comparable to reference datasets. They reflect the same global patterns when comparing multi-annual fields. In some comparison (e.g. COT in polar regions), the spread among all datasets (including the reference datasets) is large, which prevents any sound conclusion on the quality of the Cloud_cci datasets. However, most comparisons clearly indicate consensus between Cloud_cci and the reference datasets highlighting the comparable quality of the Cloud_cci datasets for most cloud properties. Exceptions are CER_{ice} and IWP for which the Cloud_cci datasets (except MERIS+AATSR) reveal a strong overestimation. The main reason for this has already been identified (error in the LUTs for ice clouds) and fixed in CC4CL, which will leads to significantly improved CER_{ice} and IWP in upcoming dataset editions.

Furthermore, time series analyses in Section 4 together with homogeneity analyses in Section reveal some drifts and jumps in the time series of some of the properties, i.e. for those datasets being composed of more than one satellite mission. These are caused by changing observations times between subsequent satellites as well as some remaining differences in spectral sensor characteristics that have not yet been completely accounted for. These inhomogeneities need to be considered when performing long-term analyses, e.g. trend detection.

As an additional validation aspect, the OE-based uncertainties were validated. Unfortunately, the Cloud_cci uncertainties of the cloud properties (apart from cloud mask/fraction) as presented in the v2.0 data are found to be too low, thus are not representing the 68% confidence interval around the retrieved values. In addition, for the presented dataset version (v2.0) the CTP, CTT, CTH uncertainties do not seem to be a good measure of the actual uncertainty of the CTP, CTT, CTH retrieval as no correlation to the real retrieval error (deviation from the truth) could be proven, and are thus recommended not to be used. For LWP and IWP (and probably also for CER, COT, and other cloud properties derived from then) the uncertainties are useable, although found to be too low as well. For the moment it is recommended to apply a correction factor of 3 to all uncertainties, except cloud mask/fraction and CTP/CTH and CTT (these uncertainties should not be used), with which a confidence interval of nearly 66% could be reached. Further research is necessary to improve the uncertainty characterisation in the retrieval, and thus in the Level-2 data.

Finally, the validation results inferred were compared to GCOS requirements (see compliance matrix in Table 7-2). For nearly all validations for which a trustable reference data source is available, the compliance to GCOS requirements could be shown, e.g. cloud fraction (all datasets) accuracy and stability, liquid water path accuracy and ice water path accuracy. Cloud top height accuracy is close to the requirements for many situations. Generally, for effective radius and optical thickness no reliable reference data is available for accuracy compliance analysis. A general problem is the assessment of the stability. In this report most stability assessments are based on comparisons to MODIS, which in turn however, is not a entirely reliable source itself as it is sometime characterized by significant trends which may or may not be true. For this reason stability compliance against MODIS is often not met, however, it often is when considering a constant climate.

Based on the evaluation results documented in this report, the project team summarized recommendations on the usage of Level-2, Level-3U and Level-3C products of Cloud_cci datasets versions 2.0 (Table 7-1). Using Level-3S products needs careful consideration of the partly large and time-varying discrepancies between the used sensors in the light of the envisaged application. Please contact the Cloud_cci team for more information (http://www.esa-cloud-cci.org/?q=support).

Table 7-1 Recommendation on the usage of Level-2, Level-3U and Level-3C products of Cloud_cci version 2.0 datasets.

Cloud mask (CMA) & Cloud fraction (CFC)

Recommendation for use:

Can be used (See limitations below)

Known limitations:

AVHRR-PM, AVHRR-AM, MODIS-Aqua, MODIS-Terra, ATSR2-AATSR (CC4CL) datasets:

- Discrimination of heavy aerosol and cloudy is not optimal, thus aerosol is sometimes flagged as clouds in such conditions. It is advised to be careful in the interpretation cloudiness in periods with dust / volcanic ash outbreaks. Cloudiness is overestimated in these conditions.
- Cloud detection during polar night over snow and ice is generally difficult. Comparisons to other AVHRR datasets reveal that the CC4CL cloud detection has also some shortcoming during polar day although it performs better than during polar night.
- Cloud detection in twilight conditions is of rather poor quality due to the small number of channels used. Due to the orbit constellation, this issue is impacting most significantly the AVHRR-AM set before 1999.
- AVHRR-PM/AM time series of CFC should not be used for trend analysis without performing a proper satellite drift correction first.
- The ATSR2-AATSR cloud detection consistency between ATSR2 and AATSR is currently poor because of a missing spectral shift adjustment for ATSR2
- Due to a limitation of passive imagers, the cloud fraction is usually biased toward lower values, compared to example CALIOP.

MERIS+AATSR (FAME-C):

- Overestimation of cloudiness in desert region, especially in Africa and Middle-East region.
- Underestimation of cloudiness above ocean related to inhomogeneous cloud fields with small clouds.
- Some problems occur in the discrimination of heavy aerosol loading with cloud, as mentioned above.
- Underestimation of cloudiness for optically thin, high cirrus clouds in tropical regions
- Overestimation of cloudiness above snow and ice surfaces.

Cloud phase & Liquid cloud fraction (CPH)

Recommendation for use:

Can be used (See limitations below)

Known limitations:

- MODIS-Aqua phase data biased towards liquid phase and should be used carefully.
- AVHRR-PM/AM time series of CFC should not be used for trend analysis without performing a proper satellite drift correction first.
- ATSR-2/AATSR is biased towards ice phase and should be used carefully

Cloud top pressure (CTP)

Recommendation for use:

Can be used (See limitations below)

Known limitations:

All datasets:

- In semi-transparent (ice) cloud conditions, the cloud top will be assigned too low.
- Multi-layer clouds are not modelled hence the CTH for cases of an upper layer of thin cirrus will effectively retrieve a radiative height (approx. 1 optical depth into the cloud).

AVHRR-PM, AVHRR-AM, MODIS-Aqua, MODIS-Terra, ATSR2-AATSR (CC4CL) datasets:

- The corrected cloud top pressure/height/temperature retrievals are not radiatively consistent.

MERIS+AATSR (FAME-C) dataset:

- For a significant part of cloudy pixels the OE does not reach a successful CTT retrieval (no convergence reached). This is in cases where the forward model did not lead to successful fitting of simulated radiances to observed radiances. This happens for example in cases of optically thin clouds or multi-layer cloud situations. These pixels are set to the undefined value.
- In case of temperature inversions the wrong CTH and CTP might be assigned due to ambiguous values or due to wrong model profiles.
- In case of very low clouds sometimes no successful retrieval can be performed due to observed radiance ratios which do not occur in the simulated radiances. This might be related to stray light correction and/or slightly inaccurate simulations of absorption in the Oxygen-A band. These pixels are set to the undefined value.
- Due to the sensitivity of the CTP retrieval to the cloud vertical extinction profile, which is unknown, a large overestimation or underestimation of CTP might occur.

Cloud optical thickness - liquid and ice clouds (COT_{liq}, COT_{ice})

Recommendation for use:

Can be used (See limitations below)

Known limitations:

- COT is a daytime product only
- In cases of wrong phase assigned, the optical thickness is likely to have significant errors.
- In the case of incorrectly assigned surface BRDF the optical depth is likely to be biased. Too high BRDF the COT will be biased low. Too low BRDF the COT will be biased high.
- In case of sub-pixel clouds or cloud borders the COT is likely to have significant errors.
- In case of optically thin clouds above (especially poorly) defined highly reflecting surface, the COT retrieval might be problematic
- For very optically thick clouds, the measurements go into saturation and thus the sensitivity of the measurement to the COT is small. Those values should be accompanied by large uncertainty values.

Cloud effective radius - liquid clouds (CER_{liq})

Recommendation for use:

Can be used (See limitations below)

Known limitations:

All datasets:

- CER is a daytime product only
- In cases of wrong phase assigned, the effective radius is likely to have significant errors.
- In case of sub-pixel clouds or cloud boarders, the effective radius is likely to have significant errors.

AVHRR-PM, AVHRR-AM, MODIS-Aqua, MODIS-Terra, ATSR2-AATSR (CC4CL) datasets:

- Due to a bug in the ice LUTs for 3.7μm, the ice affective radii are systematically too high. This affected nearly all datasets, with the exception of the MERIS+AATSR dataset and the AVHRR-AM dataset after mid of 2002.

MERIS+AATSR (FAME-C) dataset:

- In FAME-C the near-infrared channel at 1.6 micron is used. These measurements are affected by 3-d cloud structures, which are not accounted for in the physical cloud model used in the retrieval, and have a penetration depth into the cloud.

Cloud effective radius - ice clouds (CER_{ice})

Recommendation for use:

Should be used with caution (See limitations below)

Known limitations:

All datasets:

- CER is a daytime product only
- In cases of wrong phase assigned, the effective radius is likely to have significant errors.
- In case of sub-pixel clouds or cloud boarders, the effective radius is likely to have significant errors.

AVHRR-PM, AVHRR-AM, MODIS-Aqua, MODIS-Terra, ATSR2-AATSR (CC4CL) datasets:

- Due to a bug in the ice LUTs for 3.7μm, the ice affective radii are systematically too high. This affects all AVHRR-PM, MODIS-Aqua, MODIS-Terra and ATSR2-AATSR data, and data of AVHRR-AM before 2002. **The absolute values of CER_{ice} in these datasets are not reliable.**

MERIS+AATSR (FAME-C) dataset:

- In FAME-C the near-infrared channel at 1.6 micron is used. These measurements are affected by 3-d cloud structures, which are not accounted for in the physical cloud model used in the retrieval, and have a penetration depth into the cloud.

Liquid water content (LWP)

Recommendation for use:

Can be used (See limitations below)

Known limitations:

All datasets:

- LWP is a daytime product only
- Since LWP is computed from retrieved COT and CER, same limitations as for COT and CER apply for LWP.
- The method used assumes vertically homogeneous clouds, which might deviate from truth. In case of vertically inhomogeneous cloud layers, e.g. multi-layer clouds, the LWP retrieval is likely to show large errors, since the CER is retrieved from the most upper cloud layers and may not be representative for the entire vertical column.
- In cases of wrongly assigned cloud phase, i.e. ice cloud is treated as liquid cloud, the retrieved LWP will show large errors.

ATSR-2-AATSR (FAME-C) dataset:

- The LWP is currently biased too high as the phase selection for this dataset is biased towards ice.

Ice water content (IWP)

Recommendation for use:

Should be used with caution (See limitations below)

Known limitations:

All datasets:

- IWP is a daytime product only
- Similar limitations as mentioned for Cloud liquid water path (see Section **Fehler! Verweisquelle konnte nicht gefunden werden.**)

AVHRR-PM, AVHRR-AM, MODIS-Aqua, MODIS-Terra, ATSR2-AATSR (CC4CL) datasets:

- Due to a bug in the ice LUTs for 3.7μm, the IWP data are systematically too high. This affects all AVHRR-PM, MODIS-Aqua, MODIS-Terra and ATSR2-AATSR data, and data of AVHRR-AM before 2002. **The absolute values of IWP in these datasets are not reliable.**

MERIS+AATSR (FAME-C) dataset:

- Coefficients used in Heymsfield (2003) are based on observations for mid-latitude cirrus clouds. It is not exactly known how large the introduced error for deviating cloud conditions is.

Table 7-2 Compliance matrix confronting validation results of Cloud_cci data with GCOS requirements. The latter are taken from Section 2.2.1 of URDv3 (2018). Green shaded cell indicate compliance with the requirements, yellow cells nearly compliance and red cell no compliances. However, only for validation scores reported in black bold font, the reference data is trusted to provide an accurate reference for compliance analysis. For some cloud properties and datasets no validation ('n/v') was possible. Last column provides tractability of the used validation scores reported earlier in this document.

		GCOS target requirements	Achieved Cloud_cci accuracy						Comment
			AVHRR-AM	AVHRR-PM	MODIS-Terra	MODIS-Aqua	ATSR2-AATSR	MERIS+AATSR	
Cloud cover	**accuracy**	5%	**3.5 %[2]**	**-0.8 %[2]**	**4.7 %[2]**	**2.3 %[2]**	**3.4 %[12]**	**-4.3 %**	Level-2 validation against CALIOP
	stability (per decade)	3%	**0.34%**	**0.95%**	**-0.19%**	**-0.16%**	**-0.56%**	**-1.78%**	Values[4] taken from Table 4-2 and 4-13 (L3C comparisons to MODIS C6)
Cloud top height/ pressure	**accuracy[5]** (low/mid/high)	0.5/0.7/ 1.6km	n/v	**0.1/-0.8/ 0.05km[2]**	**0.07/-1.7/ -1.9km[2]**	**0.7/-1.3/ -1.4km[2]**	**0.3/-0.8/ -2.7km**	**-0.06/-1.0/ -2.6km**	Level-2 validation against CALIOP
	stability (per decade)	15hPa	16.2hPa	8.0hPa	16.7hPa	8.4hPa	14.8hPa	12.9hPa	Values[4] taken from Table 4-4 and 4-15 (L3C comparisons to MODIS C6)
Liquid cloud optical depth	**accuracy**	10%	n/v	n/v	n/v	n/v	n/v	n/v	No validation possible due to a lack of reliable reference data. through LWP and IWP validation
	stability (per decade)	2%	-1.3%	10%	3%	2.9%	-0.6%	-4.8%	Values[4] taken from Table 4-5 and 4-16 divided by mean MODIS-Terra C6 COT_{liq} (9) (L3C comparisons to MODIS C6)
Liquid cloud optical depth	**accuracy**	10%	n/v	n/v	n/v	n/v	n/v	n/v	No validation possible due to a lack of reliable reference data. through LWP and IWP validation
	stability (per decade)	2%	3.8%	-0.6%	-4.6%	-12.4%	-3.8%	-5%	Values[4] taken from Table 4-6 and 4-17 divided by mean MODIS-Terra C6 COT_{ice} (10) (L3C comparisons to MODIS C6)
Liquid water path	**accuracy**	25%	n/v	**10.7%[3]**	n/v	n/v	n/v	n/v	Level-2 validation against AMSR-E (Figure 6-12)
	stability (per decade)	5%	-13%	24%	3.5%	2%	-0.4%	-8.2%	Values[4] taken from Table 4-9 and 4-20 divided by mean MODIS C6 LWP (80g/m²) (L3C comparisons to MODIS C6)

Ice water path	**accuracy**	25%	n/v	**-19.9%**	n/v	n/v	n/v	n/v	Level-2 validation against DARDAR (Figure 6-17
	stability (per decade)	5%	3.4%	8%	-2.8%	-17%	-1.7%	-3.2	Values[4] taken from Table 4-10 and 4-12 divided by mean MODIS C6 IWP (280g/m²) (L3C comparisons to MODIS C6)
Cloud liquid effective particle radius	**accuracy**	10%	n/v	n/v	n/v	n/v	n/v	n/v	No validation possible due to a lack of reliable reference data. through LWP and IWP validation
	stability (per decade)	1µm	-1.0µm	1.53 µm	0.04µm	-0.04 µm	0.23µm	-0.7µm	Values[4] taken from Table 4-7 and 4-18 (L3C comparisons to MODIS C6)
Cloud ice effective particle radius	**accuracy**	10%	n/v	n/v	n/v	n/v	n/v	n/v	No validation possible due to a lack of reliable reference data. through LWP and IWP validation
	stability (per decade)	1µm	-0.2µm	0.91 µm	0.69µm	-0.12 µm	1.12µm	-0.12µm	Values[4] taken from Table 4-8 and 4-19 (L3C comparisons to MODIS C6)

[1] weighted mean of row 3 (Bias) of Table 3-7
[2] after applying COD threshold
[3] global ocean
[4] derived with respect to MODIS C6 which itself reveals a significant (negative) trend for some cloud properties
[5] As most of the validation results (discussed in the following) are given in terms of CTH, which is also the direct measurement of CALIOP data used as validation source, we defined CTH requirements by converting GCOS CTP accuracy requirements CTH via the US standard atmosphere (1976) profiles, obtained from https://www.digitaldutch.com/atmoscalc/, using the ΔH/ΔP derivative at three representative levels for low, mid-level and high clouds. For mid-level clouds the middle pressure level between the 680hPa and 440hPa borders was chosen: 560hPa. For low and high clouds distinct peaks exist in frequency distributions of global CALIOP CTP measures around 850hPa and 200hPa. At these two levels the conversion from GCOS CTP requirements for low and high clouds to CTH requirements were used.

8. Glossary

ACDD	Attribute Convention for Dataset Discovery
AATSR	Advanced Along Track Scanning Radiometer
ATBD	Algorithm Theoretical Baseline Document
ATSR2	Along-Track Scanning Radiometer 2
AVHRR	Advanced Very High Resolution Radiometer
BRDF	Bidirectional Reflectance Distribution Function
Bc-RMSE	Bias-corrected (bc) Root Mean Square Error
C6	(MODIS) Collection 6
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CC4CL	Community Cloud retrieval for Climate
CCI	Climate Change Initiative
CER	Cloud Effective Radius
CF	Climate and Forecast
CFC	Cloud Fractional Coverage
CFMIP	Cloud Feedback Model Intercomparison Project
CLA	Spectral Cloud Albedo
CLARA-A2	CM SAF cLouds, Albedo and Radiation dataset from AVHRR data - Edition 2
CMA	Cloud Mask
COSP	CFMIP Observation Simulator Package
CPH	Cloud Phase
COSP	CFMIP Observation Simulator Package
COT	Cloud Optical Thickness
CTH	Cloud Top Height

CTP	Cloud Top Pressure
CTT	Cloud Top Temperature
CWP	Cloud Water Path
DCHP-A	Daytime Cloud Height Property AATSR
DCHP-M	Daytime Cloud Height Property MERIS
DOI	Digital Object Identifier
DSRDP	Data Standards Requirements for CCI Data
DWD	Deutscher Wetterdiest
ECV	Essential Climate Variable
ENVISAT	Environmental Satellite
ESA	European Space Agency
FAME-C	FUB AATSR MERIS Cloud retrieval algorithm
FAR	False alarm rate
FUB	Freie Universität Berlin
FCDR	Fundamental Climate Data Record
GAC	Global Area Coverage – globally available AVHRR dataset with reduced resolution (4 km).
JCH	Joint Cloud property Histogram
ISCCP	International Satellite Cloud Climatology Project
IWP	Ice Water Path
LUT	Look-up Table
LWP	Liquid Water Path
MERIS	Medium Resolution Imaging Spectrometer
Metop	Meteorological Operational Satellite

MODIS	Moderate Resolution Imaging Spectroradiometer
MW	Microwave
NOAA	National Oceanic & Atmospheric Administration
OE	Optimal Estimation
PATMOS-x	AVHRR Pathfinder Atmospheres–Extended dataset
POD	Probability of detection
PUG	Product User Guide
PVIR	Product Validation and Intercomparison Report
RAL	Rutherford Appleton Laboratory
SYNOP	Surface synoptic observations
UWISC	University of Wisconsin

9. References

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, **2003**. Guidance on metadata and homogenization, WMO TD N. 1186 (WCDMP N. 53), 51pp.

Alexandersson, H., **1986**. A homogeneity test applied to precipitation data. J. Climatol. 6, 661–675. doi:10.1002/joc.3370060607

ATBDv5, Algorithm Theoretical Baseline Document (ATBD) – ESA Cloud_cci, Issue 5, Revision: 0, planned date of Issue: 06/2017, Available at: http://www.esa-cloud-cci.org/?q=documentation

ATBD-FAME-Cv5, Algorithm Theoretical Baseline Document (ATBD) FAME-C – ESA Cloud_cci, Issue 5, Revision: 0, planned date of Issue: 06/2017, Available at: http://www.esa-cloud-cci.org/?q=documentation

ATBD-CC4CLv5, Algorithm Theoretical Baseline Document (ATBD) CC4CL – ESA Cloud_cci, Issue 5, Revision: 0, planned date of Issue: 06/2017, Available at: http://www.esa-cloud-cci.org/?q=documentation

Baum, B. A., Menzel, W. P., Frey, R. A., Tobin, D. C., Holz, R. E., Ackerman, S. A., Heidinger, A. K., and Yang, P.: MODIS cloud-top property refinements for collection 6, Journal of applied meteorology and climatology, 51, 1145–1163, **2012**.

Benjamini, Y., and Hochberg, Y. (**1995**). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57, 289–300.

Bodas-Salcedo, A., Webb, M.J., Bony, S., Chepfer, H., Dufresne, J.L., Klein, S.A., Zhang, Y., Marchand, R., Haynes, J.M., Pincus, R. and John, V.O., **2011**. COSP: Satellite simulation software for model assessment. Bulletin of the American Meteorological Society, 92(8), p.1023.

Carbajal Henken, C.K., Lindstrot, R., Preusker, R. and Fischer, J.: FAME-C: cloud property retrieval using synergistic AATSR and MERIS observations. Atmos. Meas. Tech., 7, 3873–3890, doi:10.5194/amt-7-3873-2014, **2014**

CARv3, Climate Assessment Report (CAR) – ESA Cloud_cci, Issue 3, Revision: 0, planned date of Issue: 06/20017, Available at: http://www.esa-cloud-cci.org/?q=documentation

CECRv3, Comprehensive Error Characterization Report (CECR) – ESA Cloud_cci, Issue 3, Revision: 1, Date of Issue: 07/03/2017 Available at: http://www.esa-cloud-cci.org/?q=documentation

Delanoë, J., and R. J. **Hogan**, 2008: A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer, J. Geophys. Res., 113, D07204, doi:10.1029/2007JD009000.

Delanoë, J., and R. J. **Hogan**, 2010: Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds. J. Geophys. Res., 115, D00H29, doi:10.1029/2009JD012346.

González A., **2009**: Measurement of Areas on a Sphere Using Fibonacci and Latitude--Longitude Lattices. Mathematical Geosciences. 42 (1), 49-64. doi:10.1007/s11004-009-9257-x

Feofilov, A. G., C. J. Stubenrauch, S. Protopapadaki, and R. Armante, **2017**: Diurnal variation of high-level clouds from a synergy of the space-borne infrared sounders AIRS and IASI: detection and radiative effects. In preparation for submission to ACPD.

Heidinger, A.K., M.D. Goldberg, A. Jelenak and M.J. Pavolonis, **2005**: A new AVHRR cloud climatology, Proc. SPIE 5658, 197, doi: 10.1117/12.579047.

Heidinger, A.K., M.J. **Pavolonis**, **2009**: Gazing at Cirrus Clouds for 25 Years through a Split Window. Part I: Methodology. J. Appl. Meteor. Climatol., 48, 1100–1116. doi: 10.1175/2008JAMC1882.1

Heidinger, A. K., Evan, A. T., Foster, M., and Walther, A., **2012**: A Naïve Bayesian cloud detection scheme derived from CALIPSO and applied within PATMOS-x. Journal of Applied Meteorology and Climatology, 51, 1129–1144.

Heidinger, A. K., Foster, M. J., Walther, A., and Zhao, X.: The pathfinder atmospheres-extended AVHRR climate dataset, Bulletin of the American Meteorological Society, 95, 909–922, **2014**.

Hollmann, R., Merchant, C.J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R. and Holzer-Popp, T., **2013**. The ESA climate change initiative: Satellite data records for essential climate variables. Bulletin of the American Meteorological Society, 94(10), pp.1541-1552.

Karlsson, K. -G., & Johansson, E. (**2013**). On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO–CALIOP data: example investigating the CM SAF CLARA-A1 dataset. Atm. Meas. Tech., 6, 1271–1286, http://dx.doi.org/10.5194/amt-6-1271-2013

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Meirink, J. F., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R.: CLARA-A2: The second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmospheric Chemistry and Physics Discussions, **2016**, 1–41, doi:10.5194/acp-2016-935, http://www.atmos-chem-phys-discuss.net/acp-2016-935/, 2016.

Kendall, M.G., **1938**. A New Measure of Rank Correlation. Biometrika 30, 81–93.

Khaliq, M.N., Ouarda, T.B.M.J., **2007**. On the critical values of the standard normal homogeneity test (SNHT). Int. J. Climatol. 27, 681–687. doi:10.1002/joc.1438

Mann, H.B., **1945**. Nonparametric Tests Against Trend. Econometrica 13, 245–259.

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

McGarragh, G. R., Poulsen, C. A., Thomas, G. E., Povey, A. C., Sus, O., Stapelberg, S., Schlundt, C., Proud, S., Christensen, M. W., Stengel, M., Hollmann, R., and Grainger, R. G.: The Community Cloud retrieval for CLimate (CC4CL). Part II: The optimal estimation approach, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-333, in review, **2017**.

O'Dell, C.W., Wentz, F.J. and Bennartz, R., **2008**. Cloud liquid water path from satellite-based passive microwave observations: A new climatology over the global oceans. Journal of Climate, 21(8), pp.1721-1739.

Pavolonis, M.J., A.K. Heidinger, T.Uttal, **2005**: Daytime Global Cloud Typing from AVHRR and VIIRS: Algorithm Description, Validation, and Comparisons. J. Appl. Meteor., 44, 804–826. doi: 10.1175/JAM2236.1

Peel, M.C., Finlayson, B.L., McMahon, T.A., **2007**. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. doi:10.5194/hess-11-1633-2007

Platnick, S., King, M. D., Meyer, K. G., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Ridgway, B., et al.: MODIS cloud optical properties: User Guide for the Collection 6 Level-2 MOD06/MYD06 product and associated Level-3 datasets, Version 0.9 (beta), 17, **2014**.

Platnick, S., Meyer, K.G., King, M.D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G.T., Zhang, Z., Hubanks, P.A., Holz, R.E. and Yang, P., **2017**. The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), pp.502-525.

PUGv3.1, Product User Guide (PUG) – ESA Cloud_cci, Issue 3, Revision: 1, Date of Issue: 18/04/2017 Available at: http://www.esa-cloud-cci.org/?q=documentation

Rossow, W.B. and Schiffer, R.A., **1991**. ISCCP cloud data products. Bulletin of the American Meteorological Society, 72(1), pp.2-20.

Rossow, W.B., Walker, A.W., Beuschel, D.E. and Roiter, M.D., **1996**. International Satellite Cloud Climatology Project (ISCCP) documentation of new cloud datasets. WMO/TD, 737, p.115.

Rossow, William B., and Robert A. Schiffer. "Advances in understanding clouds from ISCCP." Bulletin of the American Meteorological Society 80, no. 11 (**1999**): 2261.

RAFCDRv1.0, Technical Report on AVHRR GAC FCDR generation – ESA Cloud_cci, Issue 1, Revision: 0, planned date of Issue: 06/2017. Available at: http://www.esa-cloud-cci.org/?q=documentation

RODCv1.0, Report on Orbital Drift Correction for AVHRR – ESA Cloud_cci, Issue 1, Revision: 0, planned date of Issue: 06/2017. Available at: http://www.esa-cloud-cci.org/?q=documentation

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**.

Stubenrauch, C.J., W.B. Rossow, S. Kinne, S. Ackerman, Cesana G., H. Chepfer, et al. Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX Radiation Panel, Bulletin of the American Meteorological Society (**2013**) http://dx.doi.org/10.1175/BAMS-D-12-00117

Stubenrauch, C. J., A. G. Feofilov, S. E. Protopapadaki, R. Armante, 2017: Cloud climatologies from the InfraRed Sounders AIRS and IASI: Strengths, Weaknesses and Applications In preparation for submission to ACPD

Sus, O., Stengel, M., Stapelberg, S., McGarragh, G., Poulsen, C., Povey, A. C., Schlundt, C., Thomas, G., Christensen, M., Proud, S., Jerg, M., Grainger, R., and Hollmann, R.: The Community Cloud retrieval for Climate (CC4CL). Part I: A framework applied to multiple satellite imaging sensors, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-334, in review, **2017**.

Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. Parts 1–3. Ned Akad Wetensch 53, 386–392, 521-525, 1397–1412.

Thomas, S.M., A.K. Heidinger, and M.J. Pavolonis, 2004: Comparison of NOAA's Operational AVHRR-Derived Cloud Amount to Other Satellite-Derived Cloud Climatologies. J. Climate, 17, 4805–4822. doi: 10.1175/JCLI-3242.1

Toreti, A., Kuglitsch, F.G., Xoplaki, E., Della-Marta, P.M., Aguilar, E., Prohom, M., Luterbacher, J., 2011. A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. Int. J. Climatol. 31, 630–632. doi:10.1002/joc.2088

URDv3, User Requirements Document (URD) – ESA Cloud_cci, Issue 3, Revision: 0, Date of Issue: 12/02/2018 Available at: http://www.esa-cloud-cci.org/?q=documentation

Vaughan, M., Powell, K., Kuehn, Rl.,Young, S., Winker, D., Hostetler, C., Hunt, W., Liu, Z., McGill, M., and Getzewich, B., 2009: Fully Automated Detection of Cloud and Aerosol Layers in the CALIPSO Lidar Measurements, J. Atmos. Oceanic Technol., 26, 2034–2050, doi: 10.1175/2009JTECHA1228.1.

Wang, T., Fetzer, E.J., Wong, S., Kahn, B.H. and Yue, Q., **2016**. Validation of MODIS cloud mask and multilayer flag using CloudSat-CALIPSO cloud profiles and a cross-reference of their cloud classifications. *Journal of Geophysical Research: Atmospheres*, *121*(19).

Wentz, F. and T. Meissner. **2000**. AMSR Ocean Algorithm. Algorithm Theoretical Basis Document, Version 2. Santa Rosa, California USA: Remote Sensing Systems. Available at: https://eospso.gsfc.nasa.gov/sites/default/files/atbd/atbd-amsr-ocean.pdf

Wentz, F. J. and T. Meissner. **2004**. *AMSR-E/Aqua L2B Global Swath Ocean Products derived from Wentz Algorithm, Version 2*. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: http://dx.doi.org/10.5067/AMSR-E/AE_OCEAN.002. [Date Accessed: 07 Sept. 2017].

Winker, D. M., Vaughan, M.A., Omar, A., Hu, Y, Powell, K.A., Liu, Z., Hunt, W.H., and Young, S.A., **2009**: Overview of the CALIPSO mission and CALIOP data processing algorithms, J. Atmos. Oceanic. Technol., 26, 2310-2323, doi:10.1175/2009JTECHA1281.1.

Annex A – Datasets for comparisons with Cloud_cci products

A.1 SYNOP: manual cloud observations from surface stations

Observations of total cloud cover made at meteorological surface stations (i.e. synoptic observations – hereafter called SYNOP) constitute one of the data records used to evaluate the cloud fractional coverage estimates. The SYNOP data used is from the local DWD archive of collected global SYNOP reports following the guidance of the *Guide to Meteorological Instruments and Methods of Observations* (WMO, 2008)

At manned stations the total cloud cover is visually estimated by human observers, at automated stations in contrast ceilometers are used for that purpose. For data quality reasons, only those SYNOP reports provided by manned airport stations were taken into account (~1800 stations globally).

SYNOP total cloud cover observations are used for the evaluation of level-3 cloud cover estimates.

Manual cloud observations are affected by many sources of error. We list some of the most important in the following:

- The observation is subjective in nature, i.e., despite clear instructions on how to make an observation, differences will appear because of different interpretations from person to person. This introduces a random noise in global cloud amount observations but may also lead to geographical biases (reflecting some systematic behaviour related to the way people have been educated/trained).
- The human eye has a detection limit for when a cloud can be clearly discernible against a cloud-free sky. This limit is somewhere in the cloud optical thickness range of 0.5-1.0 (with some dependence on solar zenith angle and on which viewing angles clouds are observed and the degree of aerosol load or haze in the troposphere). Thus, many satellite sensors have a higher sensitivity to e.g. cirrus detection than SYNOP observations.
- At night, the random error in the observations increases, naturally since the observer does not have a clear sky background against which a cloud can be observed (i.e., clouds are as dark as the cloud-free sky). However, accuracies improve in the presence of moonlight. Nevertheless, the overall effect is normally a negative bias (underestimated cloud amounts) since the observer is tempted to report cloud free conditions as soon as stars becomes visible, thus neglecting that large fractions of thin cirrus and other cloud types may still be present.
- A well-known deficiency of SYNOP observations is the scenery effect, i.e. overestimation of convective cloud towers at a slanted view (Karlsson, 2003). This effect is thus most pronounced in the summer season and for low to moderate cloud amounts when the overestimation easily can reach values of 20-30 % (1-2 octas).
- It is important to consider that most SYNOP stations are located at land stations and with higher density in developed countries. Thus, global averages tend to be biased towards land conditions in densely populated countries.

Since no rigorous study has been able to cover all those aspects in a quantitative manner (mainly because of lack of an absolute truth as reference) we can only make a very general qualitative statement about the overall quality. We would suggest that the accuracy of SYNOP observations vary between approximately +10 % (some overestimation) at daytime conditions changing to -10 % or worse (some underestimation) at night time. However, the variability (precision) probably reaches higher absolute values and it is largest during night conditions. This may lead to a strong seasonal variation with the worst accuracy and precision features during the winter season (at least at middle and high latitudes including the Polar Regions).

It is worth noting that the increasing trend to replace manual cloud observations with automatic observations from ceilometers will change the accuracy and precision of cloud observations in several ways.

Despite their subjective character and varying quality, SYNOP observations still provide a useful reference data set suitable for monitoring and validating space-based estimations of cloud coverage, especially due to their long-term availability.

A.2 CALIPSO-CALIOP

Measurements from space-born active instruments (radar + lidar) provide probably the most accurate information we can get about cloud presence in the atmosphere. The reason is the fact that the measured reflected radiation comes almost exclusively from cloud and precipitation particles and is therefore not “contaminated” by radiation from other surfaces or atmospheric constituents as is the case for measurements from most passive radiometers. In this validation study we have decided to utilise measurements from the CALIOP lidar instrument carried by the CALIPSO satellite (included in the A-Train series of satellites - Figure A-1).

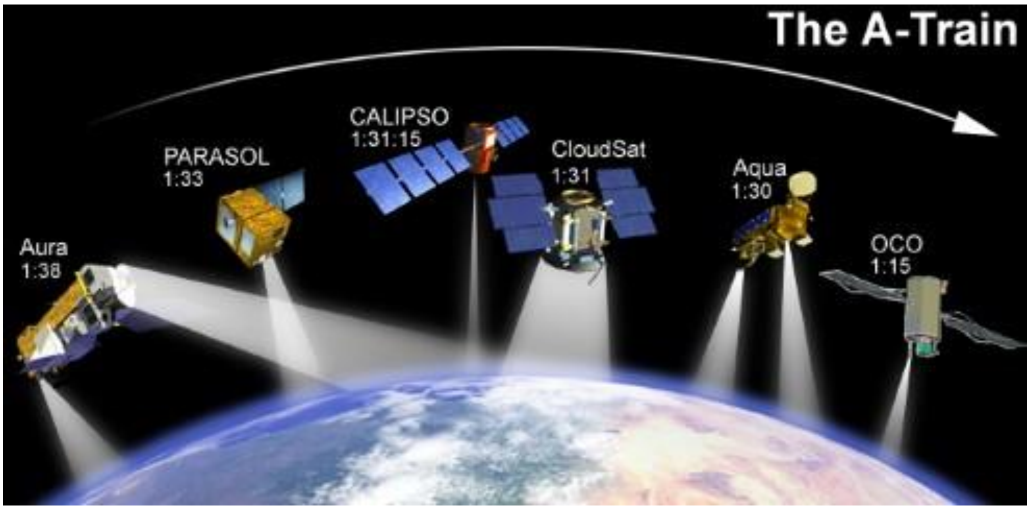


Figure A-0-1 *The Aqua-Train satellites. (Image credit: NASA)*

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite was launched in April 2006 together with CloudSat. The satellite carries the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) and the first data became available in August 2006 (Winker et al., 2009). CALIOP provides detailed profile information about cloud and aerosol particles and corresponding physical parameters (Vaughan et al., 2009).

CALIOP measures the backscatter intensity at 1064 nm while two other channels measure the orthogonally polarized components of the backscattered signal at 532 nm. The CALIOP cloud product we have used report observed cloud layers i.e., all layers observed until signal becomes too attenuated. In practice the instrument can only probe the full geometrical depth of a cloud if the total optical thickness is not larger than a certain threshold (somewhere in the range 3-5). For optically thicker clouds only the upper portion of the cloud will be sensed. The horizontal resolution of each single FOV is 333 m and the vertical resolution is 30-60 m.

The CALIOP products are available in five different versions with respect to the along-track resolution ranging from 333 m (individual footprint resolution), 1 km, 5 km, 20 km and 80 km. The four latter resolutions are consequently constructed from several original footprints/FOVs. This allows a higher confidence in the correct detection and identification of cloud and aerosol layers compared to when using the original high resolution profiles. For example, the identification of very thin Cirrus clouds is more reliable in the 5 km data record than in the 1 km data record since signal-to-noise levels can be raised by using a combined data record of several original profiles.

We used the CALIOP level-2 1 km and 5 km cloud layer data record versions 3-01, 3-02 and 3-30 (CALIPSO Science Team, 2015) for the validation purpose. The 5 km resolution data record is closest to the nominal AVHRR GAC resolution but according to Karlsson and Johansson (2013) there are some inconsistencies between results for the two resolutions which means that the total cloud amounts from the 5 km is often slightly underestimated. It means that some of the thick (opaque) boundary layer clouds that are reported in fine resolution (333 m and 1 km) data records are not reported in the higher resolution (5 km or higher) data records. This has to do with the methodology to do averaging at the longer scales (5 km or higher) where contributions from strongly reflecting clouds are removed from the original signal to facilitate detection of very thin cloud layers and aerosols. Thus, we use here the method proposed by Karlsson and Johansson (2013) combining the two CALIPSO data records (i.e., adding missed clouds at 5 km resolution which are detected at 1 km resolution). This normally gives almost 5 % higher global cloud amounts compared to if just relying on 5 km data.

The CALIOP cloud layer product reports up to 10 cloud layers per column and provides information about cloud phase and cloud type of each layer as well as the pressure, height and temperature at each layer's top.

The CALIOP data record classifies cloud layers into cloud types according to Table A-1. To be noticed here is that the ISCCP cloud type method has been used in the sense that the vertical separation of Low (categories 0-3), Medium (categories 4-5) and High (categories 6-7) clouds is defined by use of vertical pressure levels of 680 hPa and 440 hPa. However, the separation of thin and thick clouds is made using the information on whether the surface or lower layers below the current layer can be seen by CALIOP.

Table A-1 *Cloud type categories according to the CALIOP Vertical Feature Mask product*

Category 0	Low, overcast, thin (transparent St, StCu, and fog)
Category 1	Low, overcast, thick (opaque St, StCu, and fog)
Category 2	Transition stratocumulus
Category 3	Low, broken (trade Cu and shallow Cu)
Category 4	Altocumulus (transparent)
Category 5	Altostratus (opaque, As, Ns, Ac)
Category 6	Cirrus (transparent)
Category 7	Deep convective (opaque As, Cb, Ns)

We only give a quite general description of the CALIPSO data records in this section. The details concerning the actual use of the data records are elaborated further in the following sections 6.1.1.2 and 6.1.2.1.

It should be emphasized that the CALIOP measurement is probing the atmosphere very efficiently in the along-track direction since it is a nadir pointing instrument. Here, cloud dimensions down to the original FOV resolution (333 m) will be detected. However, it should be made clear that the across-track extension of the observation is still limited to 333 m. Thus, to compare CALIOP-derived results with the results of 4 km GAC AVHRR pixel data is not entirely consistent (i.e., CALIOP is only capable of covering the GAC pixel properly in one direction and not in the perpendicular direction). However, we believe that this deficiency is of marginal importance. Most cloud systems on the GAC scale will be detected, e.g., it is very unlikely to imagine elongated clouds with size and shapes below 0.3x4 km that might risk remaining undetected within a GAC pixel that coincides with a CALIOP measurement. Most clouds will have aspect ratios for the two horizontal directions that guarantee detection by CALIOP. However, it is also clear that in situations with scattered (sub-pixel) cloudiness within the GAC FOV some optically thick clouds may be detected by AVHRR cloud schemes while not being covered at all by CALIOP FOVs. Thus, some small bias between AVHRR and CALIOP observations due to this effect appears unavoidable.

It is important to consider that the CALIOP lidar instrument is much more sensitive to cloud particles than the measurement from a passively imaging instrument. It means that a significant fraction of all CALIOP-detected clouds will not be detected from imagers. This sensitivity difference also propagates into CPH and CTH, which will typically be sensed at a lower cloud layer by passive instruments compared to CALIOP (see e.g., Hamann et al., 2014). Thus, to get reasonable and justified results one should theoretically consider filtering out the contributions from the very thinnest clouds. We have applied this approach in this validation study, both in the study of cloud amounts (CFC) and cloud top heights (CTO).

The cloud detection efficiency with CALIOP is slightly different day and night because of the additional noise from reflected solar radiation at daytime that can contaminate lidar backscatter measurements. However, Chepfer et al. (2010) reports that this can introduce an artificial difference of not more than 1 % when comparing night time and daytime data.

In conclusion: Despite the fact that the CALIPSO cloud observations most likely are the best available cloud reference data record being released so far, we might still see a negative bias of a few percentage points in cloud cover when using exclusively the 5 km data record. However, in this validation effort we have tried to compensate for this effect by combining the 1 km and 5 km data records following Karlsson and Johansson (2013). Other errors, e.g. due to mis-interpretation of heavy aerosol loads as clouds, are in this respect of minor importance when judging the effect on full global orbits.

A.3 PATMOS-x

The most appropriate satellite-derived climatology to compare AVHRR-based Cloud_cci datasets with is (besides CLARA-A2) the PATMOS-x data record. The acronym stands for "AVHRR Pathfinder Atmospheres – Extended" and the corresponding cloud products have been derived using the CLAVR-x method (Clouds from AVHRR – Extended, see Heidinger et al, 2005, Pavolonis et al., 2005, Thomas et al., 2004 and Heidinger and Pavolonis, 2009). AVHRR radiances in all available spectral channels have been used to derive global cloud and radiation products over the entire lifetime of the AVHRR sensor. Some basic information about the used methodology for the derivation of various parameters is given in Table A.. To notice is that the cloud screening methodology of CLAVR-x has undergone a substantial revision lately compared to the method described by the cited references. The previous multispectral threshold approach has been replaced by a probabilistic methodology (naïve Bayesian classifier – see Heidinger et al., 2012). We have compared Cloud_cci results against the results produced by this new method. This means we have compared to PATMOS-x version v05r03. The most up-to-date publication describing the PATMOS-x data record is provided by Heidinger et al. (2014).

Table A.1 *Some basic characteristics of the PATMOS-x retrieval methods*

Product	Methodology
Cloud amount	Computed from results of a statistical naïve Bayesian cloud mask trained from CALIPSO-CALIOP cloud information
Cloud top level	Optimum Estimation (OE) retrieval
Cloud phase	Multi-channel test scheme
Cloud optical thickness	OE retrieval (with look-up tables as CM SAF but with different radiative transfer models and ice particle definitions)
Cloud effective radius	OE retrieval (with look-up tables as CM SAF but with different radiative transfer models and ice particle definitions)
Cloud liquid water path	Calculated from optical thickness and effective radius (Stephens' parameterization – same as CM SAF)
Cloud ice water path	Calculated from optical thickness and effective radius (Stephens' parameterisation – same as CM SAF)

The PATMOS-x data record is prepared exclusively as so-called level-2b products. This means that, for each satellite, data from all orbits during one day have been sub-sampled to produce only two global products per day valid for the nominal local solar time for both the descending (southbound) and ascending (northbound) observation nodes.

A.4 MODIS Collection 6

MODIS (or Moderate Resolution Imaging Spectroradiometer) is an advanced imaging instrument onboard the Terra (EOS AM) and Aqua (EOS PM) polar satellites (see http://modis-atmos.gsfc.nasa.gov/index.html).

Both Terra and Aqua orbits around the Earth are sun synchronous. Terra passes from north to south across the equator in the morning (local solar time 10:30), while Aqua passes south to north over the equator in the afternoon (local solar time 13:30). Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands or groups of wavelengths.

Since the Terra and Aqua satellites passes in very similar orbits (at least the afternoon orbit of Aqua) as the NOAA and Metop-A satellites and since MODIS observes with as much as 36 spectral channels (including all the AVHRR-like channels), corresponding cloud products from MODIS should serve as a top quality reference for corresponding cloud products retrieved from AVHRR data. MODIS uncertainties are indeed expected to be somewhat smaller than what can be obtained with AVHRR retrievals. For example: multiple CO_2 channels allow a more accurate cloud-top height determination, additional shortwave channels allow better discrimination of (thin) cirrus and a more reliable retrieval of cloud optical properties over very bright surfaces. Otherwise, uncertainties should lie in the same ballpark as for Cloud_cci, CLARA-A2 and PATMOS-x. The main limitation of MODIS is the relatively short duration of the observation period, starting in 2000. We have used the level-3 MODIS gridded atmosphere monthly global products - MOD08_M3 (Terra) and MYD08_M3 (Aqua). They contain monthly 1° × 1° degree grid average values of atmospheric parameters related to atmospheric aerosol particle properties, total ozone burden, atmospheric water vapor, cloud optical and physical properties, and atmospheric stability indices. Statistics are sorted into 1° × 1° degree cells on an equal-angle grid that spans a (calendar) monthly interval and then summarized over the globe.

For this particular study we have used data from Terra & Aqua MODIS Collection 6 (Platnick et al., 2014; Baum et al., 2012, Platnick et al., 2017). Validation results for MODIS C6 cloud properties can for example be found in Baum et al., (2012), Marchant et al. (2016) and Wang et al., (2016).

A.5 UWisc: liquid water path observations from microwave imagers

UWisc dataset

Passive microwave imagers, such as the Special Sensor Microwave/Imager (SSM/I) series, can be used to retrieve column-integrated liquid water along with water vapour and surface wind speed over ocean. Because the microwave (MW) channels fully penetrate clouds, they provide a direct measurement of the total liquid (but not solid) cloud condensate amount. For precipitating clouds an estimate of the rain water path has to be made and subtracted from the total liquid water path to retrieve the cloud liquid water path.

For the CC4CL LWP evaluation the the University of Wisconsin (UWisc) MW-based LWP climatology (O'Dell et al., 2008) was chosen as an independent reference data record. The LWP climatology is based on retrievals from various microwave radiometer instruments, including the SSM/I series, the Tropical Rainfall Measurement Mission Microwave Imager (TMI), and the Advanced Microwave Scanning Radiometer for EOS (AMSR-E). The UWisc v3 dataset spans the years 1988 – 2008.

UWisc Liquid water path estimates are reported to have a potential systematic error of 15% to 30% (O'Dell et al., 2008). Major contributors to the error are the clear-sky/beam-filling effect (leading to a systematic positive bias) and cloud-rain-partition (sign unclear). It seems that for the latter the Remote Sensing Systems (RSS) algorithm is bias high for LWP than alternative cloud-rain distribution assumptions.

As we are looking at stratocumulus region exclusively in this PVIR, we assume the latter source to be of only small significance. Thus, we assume a positive bias of UWISC LWP of about 15% compared to the truth.

Setup of evaluation

The UWisc LWP data record comprises monthly mean all-sky LWP in 1° × 1° grid boxes that is based on all available data for a specific month. In addition, for each month and each grid box over the 1988-2008 period the mean diurnal cycle of LWP is available. In order to obtain the monthly mean all-sky LWP from UWisc closest to the overpass times of the CC4CL satellites, the mean diurnal cycle parameters, available in the data record, were used to adjust the monthly mean grid box values, based on the equation:

$$\langle \mathrm{LWP}(Y,t)\rangle=\langle \mathrm{LWP}(Y)\rangle \; (A_0 + A_1 \cos(\omega(t-T_1)) + A_2 \cos(2\omega(t-T_2))) \; / \; A_0$$

where $\langle \mathrm{LWP}(Y)\rangle$ represents the uncorrected monthly mean LWP for year Y, t the local time (h), ω the radial frequency that corresponds to a 24-hour period, A0 the multi-year monthly mean, and A1 (T1) and A2 (T2) are the amplitudes (phases) of the first and second harmonics of the diurnal cycle, respectively (see also O'Dell et al. 2008). For the evaluation we distinguished the morning and afternoon satellites, using typical local overpass times of 9:30 and 14:00, respectively. To assess the impact of deviating overpass times, LWP was also calculated for 1 hour earlier and 1 hour later.

Because microwave instruments are able to penetrate through deep convective clouds or ice over water clouds and measure the LWP at lower altitudes, which is not possible for passive imagers, the present evaluation was restricted to regions with very few (<5%) ice clouds. Therefore, three well-known areas dominated by stratocumulus clouds were selected: the oceanic area west of Africa at 10°-20°S, 0°-10°E (SAF), the area west of South America at 16°-26°S, 76°-86°W (SAM), and the area west of California at 20°-30°N, 120°-130°W (NAM). See also Figure X1 for their locations

A.6 The CLARA-A2 dataset

The CM SAF cloud, albedo and radiation (CLARA-A2, Karlsson et al., 2016) dataset provides a long term (1982 – 2015) data record based on homogenized AVHRR (Very High Resolution Radiometer) measurements. The AVHRR Instrument is stationed on board the polar orbiting NOAA satellites and the EUMETSAT METOP satellites measuring in five spectral bands in the beginning of the dataset (1982 AVHRR/2) and was extended to a sixth channel in 1998 (AVHRR/3), although only accessible if switched with the previous third channel at 3.7 micron. The horizontal resolution is close to 1 km at nadir but only reduced resolutions of approximately 4 km are permanently archived and available with global coverage. The CLARA-A2 dataset includes measurements from NOAA- 7 up to 19 and the METOP-A and METOP-B missions.

The products, including macro physical and microphysical cloud properties, are provided on a L2 and L3 (daily and monthly mean) basis and are summarized as follows:

- Fractional Cloud Cover – CFC
- Cloud Top Pressure – CTP
- Cloud Top Temperature – CTT
- Cloud Top Height – CTH
- Cloud Optical Thickness – COT
- Cloud Phase – CPH
- Liquid Water Path – LWP
- Ice Water Path – IWP
- Cloud effective radius - REF
- Joint Cloud property Histogram – JCH

A.7 ISCCP

The International Satellite Cloud Climatology Project (ISCCP) provides cloud properties over a period of more than 35 years (Rossow and Schiffer, 1991; Rossow et al., 1996; Rossow and Schiffer, 1999). This project was established in 1982 as part of WCRP to collect weather satellite radiance measurements (from geostationary and polar orbiting satellites) and to analyze them to infer the global distribution of clouds, their properties, and their diurnal, seasonal and inter-annual variations. The resulting data records and analysis products are being used to study the role of clouds in climate, both their effects on radiative energy exchanges and their role in the global water cycle. This project and its results are considered to be the state of the art today on what can be derived from routine weather satellite data. ISCCP is the only other existing TCDR for cloud physical property products (here we mean products CPH, LWP and IWP). However, it has the disadvantage that it is based on different satellite types – polar and geostationary – of which most of the latter do not contain the necessary narrow-band channels for accurate retrieval of LWP and IWP.

The production of ISCCP has recently been transferred to the National Centers for Environmental Information (NCEI) and a new high-resolution version of the data record (to be denoted ISCCP-H) is under production (see https://www.ncdc.noaa.gov/isccp). Unfortunately, the new data record is still not released which means that we have compared to the previous ISCCP-D2 version covering the period 1983-2008 as prepared in the Global Energy and Water cycle Experiment (GEWEX) database (Stubenrauch et al, 2013). I.e. we have used the 0300AMPM data products which contain daytime and night-time data valid at 03:00 AM and 03:00PM local time, which is close to the observation times of the prime AVHRR afternoon satellites and Aqua.

Annex B - The Standard Normal Homogeneity Test

The Standard Normal Homogeneity Test (SNHT) seeks for inhomogeneities in a time series. It derives a statistic *T(k)* where changes in the standardized mean before and after each step (i.e. months in this study) are calculated. Large difference between the mean values before and after a time step indicates a possible break in a time series. Following Alexandersson (1986), statistic *T(k)* for time step *k* is defined as:

$$T(k) = k\bar{z}_1^2 + (n-k)\bar{z}_2^2\,;\; k \in 1, 2, \ldots, n. \qquad (A.1)$$

The standardized means $\bar{z}_1$ and $\bar{z}_2$ are calculated as:

$$\bar{z}_1 = \frac{1}{k}\sum_{i=1}^{k}\frac{(Y_i - \bar{Y})}{\sigma} \qquad (A.2)$$

$$\bar{z}_2 = \frac{1}{n-k}\sum_{i=k+1}^{n}\frac{(Y_i - \bar{Y})}{\sigma} \qquad (A.3)$$

where Y_i stands for the value at a time step i, $\bar{Y}$ for the mean, and σ for the standard deviation of the whole time series.

A large difference between the mean value before ($\bar{z}_1$) and after ($\bar{z}_2$) the time step *k* leads to high values of $T(k)$. Khaliq and Ouarda (2007) provided critical values of $T(k)$ depending on *n* which signifies a break in a time series at several confidence levels.