

# EUMETSAT Satellite Application Facility on Climate Monitoring



## Validation Report Meteosat Cloud Fractional Cover (COMET) Edition 2

DOI: [10.5676/EUM\\_SAF\\_CM/CFC\\_METEOSAT/V002](https://doi.org/10.5676/EUM_SAF_CM/CFC_METEOSAT/V002)

Cloud Fractional Cover

CM-23012

Doc.No.:


SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0

Issue:

2.0

Date:

April 11, 2023

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

## Document Signature Table

	Name	Function	Signature	Date
Authors	Reto Stöckli Anke Tetzlaff Quentin Bourgeois	CM-SAF Scientists		06.04.2023
Editor	Marc Schröder	Science Coordinator		11.04.2023
Approval	Steering Group			
Release	Rainer Hollmann	Project Manager		

## Distribution List

Internal Distribution	
Name	No. Copies
DWD Archive	1
CM SAF Team	1

External Distribution		
Company	Name	No. Copies
Public		1

## Document Change Record

Issue Revision	Date	DCN no.	Changed pages / Paragraphs
1.0	06.04.2023	SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0	Final version for DRR3.9

## Applicable Documents

Reference	Title	Code
AD 1	CM SAF Product Requirements Document	SAF/CM/DWD/PRD/4.1

## Reference Documents

Reference	Title	Code
RD 1	Algorithm Theoretical Basis Document Meteosat Cloud Fractional Cover Edition 2	SAF/CM/MeteoSwiss/ATBD/MET/CFC/2.0
RD 2	Product User Manual Meteosat Cloud Fractional Cover Edition 2	SAF/CM/MeteoSwiss/PUM/MET/CFC/2.0

## Contents

<b>1</b>	<b>Executive summary</b>	<b>11</b>
<b>2</b>	<b>The EUMETSAT Satellite Application Facility on Climate Monitoring</b>	<b>14</b>
<b>3</b>	<b>Introduction to the Meteosat measurements</b>	<b>15</b>
<b>4</b>	<b>Validation strategy</b>	<b>16</b>
<b>5</b>	<b>Data records for comparison with Meteosat CFC</b>	<b>18</b>
5.1	Manual cloud observations from surface stations (SYNOP) . . . . .	18
5.1.1	Reference SYNOP (1983–2020) . . . . .	19
5.1.2	Gridded SYNOP (2005) . . . . .	21
5.2	A-Train (CALIPSO-CALIOP) . . . . .	22
5.3	MODIS . . . . .	25
5.4	PATMOS-x . . . . .	25
5.5	CLARA-A2 . . . . .	27
5.6	CLAAS-A2 . . . . .	27
5.7	CC4CL AVHRR (ESA-Cloud-CCI) . . . . .	28
5.8	ISCCP H . . . . .	29
<b>6</b>	<b>Validation of Meteosat CFC</b>	<b>29</b>
6.1	Validation against synoptic observations . . . . .	29
6.1.1	Instantaneous (level 2) . . . . .	30
6.1.2	Hourly, Daily and monthly means . . . . .	34
6.1.3	Decadal stability of monthly means . . . . .	37
6.2	Validation against A-Train (CALIPSO-CALIOP) . . . . .	38

<b>7</b>	<b>Comparison of Meteosat First and Second Generation</b>	<b>44</b>
<b>8</b>	<b>Inter-comparison with other satellite-derived data records</b>	<b>47</b>
8.1	Inter-comparison with MODIS . . . . .	47
8.2	Inter-comparison with PATMOS-x . . . . .	50
8.3	Inter-comparison with CLARA-A2 . . . . .	53
8.4	Inter-comparison with CLAAS-A2 . . . . .	56
8.5	Inter-comparison with CC4CL-AVHRR . . . . .	59
8.6	Inter-comparison with ISCCP-H . . . . .	62
<b>9</b>	<b>Concluding remarks</b>	<b>65</b>
<b>A</b>	<b>Performance and homogeneity measures</b>	<b>68</b>
A.1	Performance statistics . . . . .	68
A.2	The Standard Normal Homogeneity Test . . . . .	68

## List of Figures

1	Mean bias error and bias-corrected root mean square error (in red) of Meteosat CFC of Edition 1 (1991–2015, top plots) and Edition 2 (1983–2020, bottom plots) as compared to synoptic observations. Thick lines with filled circles give yearly mean statistics. The yellow line indicates the Standard Normal Homogeneity Test (SNHT). . . . .	12
2	Overview of the Meteosat record used as input for the generation of Meteosat CFC. . .	16
3	SYNOP cloud observation statistics of Payerne Switzerland (WMO Site No. 06610). From top to bottom: mean cloud fraction (CFC), diurnal CFC amplitude, diurnal CFC phase (time of maximum CFC), number of monthly CFC observations. . . . .	19
4	SYNOP cloud observation statistics of Strassbourg France (WMO Site No. 07190). From top to bottom: mean cloud fraction (CFC), diurnal CFC amplitude, diurnal CFC phase (time of maximum CFC), number of monthly CFC observations. . . . .	20
5	SYNOP sites used as a reference to validate Meteosat CFC for 1983–2020, and to inter-compare Meteosat CFC monthly means (2005) with other satellite-based data records at a $1^\circ \times 1^\circ$ resolution. . . . .	21
6	The Aqua-Train satellites . . . . .	23
7	Overview of AVHRR measurements used to calculate the PATMOS-x CFC monthly means. . . . .	26
8	Skill scores and performance statistics of the instantaneous Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites. . . . .	32
9	Performance statistics of the level-2 Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites in relation to sun zenith angle (left) and satellite view zenith angle (right). . . . .	33
10	Skill scores of the level-2 Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites in relation to sun zenith angle (left) and satellite view zenith angle (right). . . . .	34
11	Performance statistics of the Meteosat CFC hourly, daily and monthly means during 1983–2020 as compared to synoptic observations at 290 sites. . . . .	36
12	Time series of mean bias error and bias-corrected root mean square error of Meteosat CFC as compared to synoptic observations at 290 sites in 1983–2020. . . . .	37
13	Exemplary matchup of the CALIOP track and the SEVIRI scan. . . . .	38
14	5-day moving average of Meteosat CFC and CALIPSO cloud fractional cover. . . . .	39

15	Meteosat CFC cloud scores as a function of the COT threshold used to discriminate clear and cloudy CALIOP observations. . . . .	40
16	Probability of detection for Meteosat CFC cloud mask resolved by cloud type of the uppermost CALIOP cloud layer. . . . .	41
17	Mean annual Meteosat CFC and reference CALIOP, and skill scores derived for COT>0 and COT>0.2 . . . . .	42
18	Meteosat CFC skill scores scores in 4-degree latitude bands. . . . .	43
19	Comparison of Meteosat CFC derived from Meteosat First and Second Generation and gridded SYNOP monthly means of 2005 at 1°×1°spatial resolution. . . . .	45
20	Monthly, diurnal and per-okta occurrence statistics of Meteosat CFC derived from Meteosat First and Second Generation and gridded SYNOP monthly means of 2004 and 2005 at 1°×1°spatial resolution. The sawtooth pattern of the diurnal analysis is caused by the different observation statistics of 3-hourly vs. 6-hourly SYNOP observations across SYNOP sites. . . . .	46
21	Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and MODIS as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics. . . . .	48
22	Inter-comparison of Meteosat CFC, MODIS and gridded SYNOP monthly means of 2005 at 1° × 1° spatial resolution. . . . .	49
23	Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and PATMOS-x as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics. . . . .	51
24	Inter-comparison of Meteosat CFC, PATMOS-x and gridded SYNOP monthly means of 2005 at 1° × 1° spatial resolution. . . . .	52
25	Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and CLARA-A2 as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics. . . . .	54
26	Inter-comparison of Meteosat CFC, CLARA-A2 and gridded SYNOP monthly means of 2005 at 1° × 1° spatial resolution. . . . .	55
27	Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and CLAAS-A2 as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics. . . . .	57
28	Inter-comparison of Meteosat CFC, CLAAS-A2 and gridded SYNOP monthly means of 2005 at 1° × 1° spatial resolution. . . . .	58

29	Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and CC4CL-AVHRR as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics. . . .	60
30	Inter-comparison of Meteosat CFC, CC4CL-AVHRR and gridded SYNOP monthly means of 2005 at $1^\circ \times 1^\circ$ spatial resolution. . . . .	61
31	Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and ISCCP-H as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics. . . . .	62
32	Inter-comparison of Meteosat CFC, ISCCP-H and gridded SYNOP monthly means of 2005 at $1^\circ \times 1^\circ$ spatial resolution. . . . .	63



## List of Tables

1	A summary of validation results with respect to the target bias accuracy requirement of 3% for the monthly Meteosat CFC. . . . .	13
2	A summary of validation results with respect to the bias-corrected RSME target precision requirement of 15% for the monthly Meteosat CFC. . . . .	13
3	Accuracy and precision requirements for the CM-SAF Meteosat CFC daily and monthly means (% refer to absolute CFC values). . . . .	17
4	Cloud type categories according to the CALIOP Vertical Feature Mask product. . . .	24
5	Performance statistics of level-2 Meteosat CFC as compared with synoptic observations at 290 sites. . . . .	31
6	Skill scores of the instantaneous Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites. . . . .	35
7	Performance statistics of Meteosat CFC hourly, daily and monthly means (1983–2020) as compared to synoptic observations at 290 sites. . . . .	35
8	Summary of validation results for the Meteosat CFC-based cloud mask against CALIOP.	40
9	Comparison statistics of gridded CFC derived by Meteosat First Generation, Meteosat Second Generation and SYNOP for 2004 and 2005. . . . .	44
10	Inter-comparison statistics of gridded Meteosat CFC, MODIS and SYNOP for 2005. . .	48
11	Inter-comparison statistics of gridded Meteosat CFC, PATMOS-x and SYNOP for 2005.	50
12	Inter-comparison statistics of gridded Meteosat CFC, CLARA-A2 and SYNOP for 2005.	53
13	Inter-comparison statistics of gridded Meteosat CFC, CLAAS-A2 and SYNOP for 2005.	56
14	Inter-comparison statistics of gridded Meteosat CFC, CC4CL-AVHRR and SYNOP for 2005. . . . .	59
15	Inter-comparison statistics of gridded Meteosat CFC, ISCCP-H and SYNOP for 2005.	64
16	Contingency matrix for the Meteosat-based cloud mask and reference synoptic or CALIPSO observations. . . . .	68

## List of Acronyms

AD	Applicable Documents
AVHRR	Advanced Very High Resolution Radiometer
bcRMSE	Bias-corrected Root Mean Square Error
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CCI	Climate Change Initiative
CC4CL	The ESA CCI Community Cloud Retrieval for Climate
CDR	Climate Data Record
CFC	Cloud Fractional Cover
CLARA-A2	(CM SAF's) Cloud, Albedo and Radiation dataset (version 2)
CLAAS-A2	CM SAF CLOUD property dAtAset using SEVIRI
CM SAF	(EUMETSAT's) Satellite Application Facility on Climate Monitoring
COT	Cloud Optical Thickness
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Record
GCOS	Global Climate Observing System
HIRS	High Resolution Infrared Radiation Sounder
IR	Infrared channel
ISCCP	International Satellite Cloud Climatology Project
MBE	Mean Bias Error
MFG	Meteosat First Generation
MODIS	Moderate Resolution Imaging Spectroradiometer
MVIRI	Meteosat Visible and Infra-red Imager
MSG	Meteosat Second Generation
NOAA	National Oceanic and Atmospheric Administration
PATMOS-x	Pathfinder Atmospheres Extended

RD	Reference Documents
SAF	EUMETSAT's Satellite Application Facility
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SNHT	Standard Normal Homogeneity Test
VIS	Visible channel
WMO	World Meteorological Organization
WV	Water vapour channel

## 1 Executive summary

This CM SAF report provides the validation and the inter-comparison of Meteosat Cloud Fractional Cover (CFC) Edition 2 calculated with the processing framework GeoSatClim. The Meteosat CFC data record is derived from measurements taken by the Meteosat Visible and Infrared Imager (MVISIRI) and the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) onboard the EUMETSAT Meteosat First (MFG) and Second (MSG) Generation satellites. The covered time period ranges from 1983 to 2020, thus includes MFG 2–7 and MSG 1–4 (Meteosat 3–11, respectively).

This report presents an evaluation of the following CM SAF data record:

**Cloud Fractional Cover                      CM-23012                      (CFC)**

In Edition 2 the following shortcomings of Edition 1 were addressed, and are well summarized in Figure 1. Please note that the accuracy requirements, drawn in green shaded colors, were tightened from Edition 1 to Edition 2.

1. Extension back to 1983 using new Meteosat 2–7 infrared calibration
2. Prolongation forward up to 2020 using Meteosat 10 and 11
3. Removal of inhomogeneity during 1994–1996
4. Decreasing negative biases during winter months
5. Tightening accuracy and precision requirements
6. Improving on decadal stability

**Averaged over all reference sites, Meteosat CFC complies with monthly optimal requirements for accuracy and precision as compared to SYNOP. Taking accuracy and precision requirements simultaneously into account – monthly optimal, target and threshold requirements are met by 20%, 61% and 80% of sites, respectively. These values are lower than in Edition 1 since accuracy and precision requirements were tightened for Edition 2. For instance the monthly threshold accuracy requirement was lowered from 10% to 5% and the monthly threshold precision requirement was lowered from 20% to 15%. So in absolute terms both accuracy and precision increased in Edition 2, while they decreased with respect to tightened requirements. Our analysis also demonstrates an alarming decrease in SYNOP observation quality, substantially hampering the ability to perform stability analysis and satellite algorithm development. Two examples in this report demonstrate that multi-decadal SYNOP CFC (part of GCOS ECVs) time series cannot be used properly to validate satellite-based CDRs over Switzerland and France.**

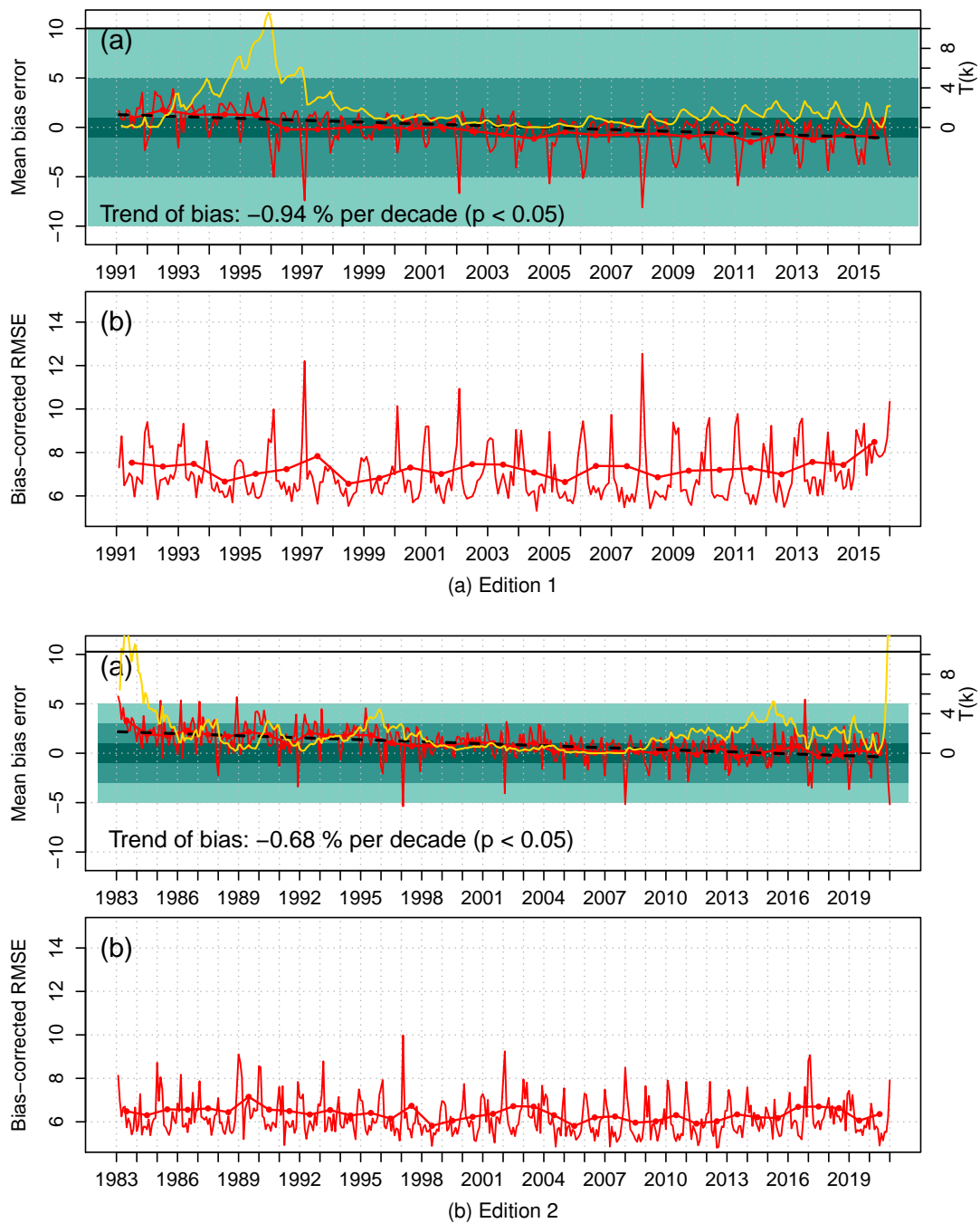


Figure 1: Mean bias error and bias-corrected root mean square error (in red) of Meteosat CFC of Edition 1 (1991–2015, top plots) and Edition 2 (1983–2020, bottom plots) as compared to synoptic observations. Thick lines with filled circles give yearly mean statistics. The yellow line indicates the Standard Normal Homogeneity Test (SNHT).

The two reference data records used for validation are ground-based observation sources (SYNOP) and retrievals from an active satellite instrument (CALIPSO/CALIOP). Inter-comparisons to other passive satellite imager derived CFC data records such as MODIS, PATMOS-x, CLARA-A2, CLAAS-A2, CC4CL-AVHRR and ISCCP-H are provided as well.

The evaluation includes both Level-2 and Level-3 comparisons. The Level-3 comparison demonstrates the applicability of the aggregated data record available to the user and is needed to check the agreement with product accuracy requirements. Level-2 comparisons allow a more in-depth analysis of potential issues of the Meteosat CFC provided on the native temporal resolution. All evaluations presented here were done in the light of the product accuracy requirements. The evaluation scores and their compliance with the target requirements of accuracy and precisions are given in Table 1 and Table 2, and can be briefly summarized as follows:


Table 1: A summary of validation results with respect to the target bias accuracy requirements of 3% for the monthly Meteosat CFC. Results from inter-comparison to other satellite-derived data records are marked in [blue](#). For all inter-comparisons the reference is SYNOP.

<b>Dataset</b>	<b>Achieved accuracy</b>
Meteosat CFC	0.9%
MODIS	<a href="#">4.9%</a>
PATMOS-x	<a href="#">-0.1%</a>
CLARA-A2	<a href="#">3.1%</a>
CLAAS-A2	<a href="#">2.7%</a>
CC4CL-AVHRR	<a href="#">5.1%</a>
ISCCP-H	<a href="#">6.1%</a>

Table 2: A summary of validation results with respect to the bias-corrected RSME target precision requirement of 15% for the monthly Meteosat CFC. Results from inter-comparison to other satellite-derived data records are marked in [blue](#). For all inter-comparisons the reference is SYNOP.

<b>Dataset</b>	<b>Achieved precision</b>
Meteosat CFC	6.2%
MODIS	<a href="#">7.8%</a>
PATMOS-x	<a href="#">8.5%</a>
CLARA-A2	<a href="#">6.4%</a>
CLAAS-A2	<a href="#">6.3%</a>
CC4CL-AVHRR	<a href="#">6.3%</a>
ISCCP-H	<a href="#">9.4%</a>

The Meteosat CFC accuracy and precision results together with the reported decadal stability <1%, fulfilling the target decadal stability requirements, demonstrate that the Edition 2 of the Meteosat CFC CFC TCDR is suitable for climate monitoring, climate model evaluation and regional climate analysis applications. It is thus useful to extend in both space and time century-long ground-based climate observations. This second version of the data record builds on a newly developed state-of-the-art retrieval system for only two heritage channels. It employs a cyclic generation of clear sky background fields, on continuous cloud scores and on a Bayesian cloud fraction estimation using concurrent information on cloud state and variability. The algorithm depends on well characterized IR radiances and VIS reflectances from the Meteosat Fundamental Climate Data Record (FCDR) provided by EUMETSAT. Further guidance on how to use the products is given in the product user manual (RD 2). The mentioned accuracy requirements are reported in detail in the product requirements document (AD 1). The algorithm theoretical basis document describes the full algorithm (RD 1).

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/GFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

## 2 The EUMETSAT Satellite Application Facility on Climate Monitoring

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


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

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

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

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

- Application of highest standards and guidelines as lined out by GCOS for the satellite data processing,
- Processing of satellite data within a true international collaboration benefiting from developments at international level and pollinating the partnership with own ideas and standards,
- Intensive validation and improvement of the CM SAF climate data records,
- Taking a major role in data set assessments performed by research organisations such as

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

WCRP (World Climate Research Program). This role provides the CM SAF with deep contacts to research organizations that form a substantial user group for the CM SAF CDRs,

- Maintaining and providing an operational and sustained infrastructure that can serve the community within the transition of mature CDR products from the research community into operational environments.

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

### 3 Introduction to the Meteosat measurements

The Meteosat CFC is based on 38 years (1983–2020) of data from EUMETSAT’s geostationary Meteosat satellites of the First and Second Generation. The satellites are located at longitude of 0° (except Meteosat-8 at 3.7°W) directly above the equator at an altitude of about 36 000 km. Both satellite generations have a field of view that extends to around 80°N/S and 80°W/E. However Meteosat CFC data record covers area up to 65°N/S and 65°W/E due to a reported decrease of accuracy at higher satellite view angles.

The Meteosat First Generation (MFG) satellites carried the Meteosat Visible and Infrared Imager (MVIRI), a radiometer that measures the earth’s disk every 30 minutes in 3 spectral bands (so-called Meteosat heritage channels) covering visible and infrared wavelengths: the broadband visible channel (VIS, 500–900 nm), the water-vapor absorption channel (WV, 570–710 nm), and the thermal infrared channel (IR, 1050–1250 nm).

The Meteosat Second Generation (MSG) satellites have been equipped with the Spinning Enhanced Visible and Infrared Imager (SEVIRI) scanning the earth surface every 15 minutes in 12 spectral channels ranging from the visible to the thermal infrared. The Meteosat CFC retrieval algorithm employs two visible channels (560–710 nm and 740–880 nm) to simulate MVIRI’s broadband channel and the thermal infrared channel (980–1180 nm).

The periods covered by consecutive Meteosat missions used for Meteosat CFC data record are shown in Figure 2. The data record is based on 22 years (1983–2004) of MFG MVIRI and 16 years (2005–2020) of MSG SEVIRI data, respectively. However, Meteosat-7 (MFG-7) and Meteosat-8 (MSG-1) operated simultaneously in 2004 and 2005. A comparison of CFC between MFG and MSG has been carried out for 2005. The results of this comparison are shown in Section 7.

The radiances of MFG MVIRI visible channel were calibrated using factors published by EUMETSAT (2010). The MFG MVIRI WV and IR channels inter-calibration factors were provided by EUMETSAT (V. John, R. Roebeling and J. Schulz, personal communication). These inter-calibration factors are based on a daily inter-calibration of MFG MVIRI with the High Resolution Infrared Radiation Sounder (HIRS) instrument on board the National Oceanic and Atmospheric Administration (NOAA) polar orbiting platforms (John et al., 2019). For MSG SEVIRI, the original radiance calibration factors were used as provided as part of the Level 1.5 radiance data by EUMETSAT (EUMETSAT, 2017). For a more detailed instrument specification, description of the calibration and pre-processing, the reader is referred to the Algorithm Theoretical Basis Document (RD 3).



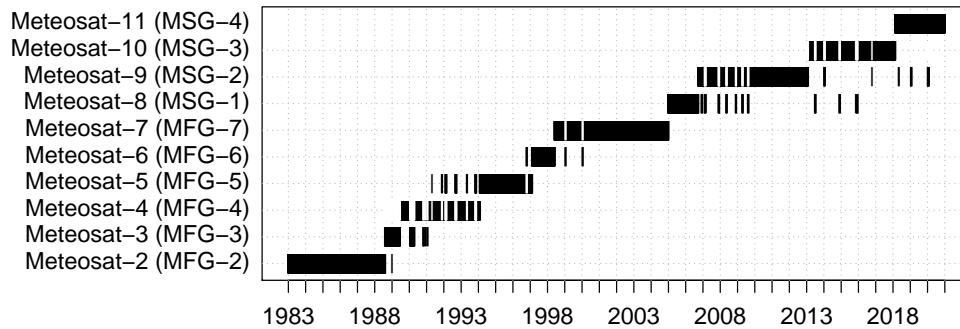


Figure 2: Overview of the Meteosat record used as input for the generation of Meteosat CFC.

## 4 Validation strategy

The purpose of the validation effort is to characterize the cloud product in terms of its accuracy and precision, thus to give a guidance for the product applicability. Furthermore, the data record is confronted with the product requirements stated in AD 1 and their compliance is reported. The requirements are determined for three categories “Threshold”, “Target” and “Optimal”. The definition of the categories follows the definitions from the WMO Observing Systems Capability Analysis and Review Tool (OSCAR, <http://www.wmo-sat.info/oscar/observingrequirements>):

- **Threshold:** The threshold is the minimum requirement to be met to ensure that data are useful and that the data will be provided to CM SAF user.
- **Target:** The target is an intermediate level between threshold and optimum which, if achieved, would result in a significant improvement for the targeted application. It is a realistic requirement which is expected to be reached by CM SAF taking into account the challenges and available sensors.
- **Optimum:** The optimum is an ideal requirement above which further improvements are not necessary. It is the best possible requirement under optimal conditions of input and auxiliary data, available sensors and retrievals. Ideally the optimum requirement is bound to the GCOS requirements for using the ECV in the field of climate monitoring and climate change detection.

The requirements for these three categories are summarized for the CM SAF Meteosat CFC data record in Table 3.

Requirements on decadal stability are 3%, 2% and 0.3% for threshold, target and optimal classes. These requirements and the ones listed in Table 3 are defined after taking into account requirements from different users and user groups. The most well-established reference here are the recommendations issued by GCOS (2006). However, values are also influenced by requirements from users working with regional climate monitoring and regional climate modeling applications.

The CM SAF Meteosat CFC data record consists of level-3 hourly, daily and monthly means for the period 1983–2020. The validation task comprises evaluation of the daily and monthly CFC to check if they comply with CM SAF requirements. However, to evaluate Meteosat CFC against reference data records (i.e. SYNOP and CALIPSO/CALIOP), the collocations are performed for

Table 3: Accuracy and precision requirements for the CM-SAF Meteosat CFC daily and monthly means (% refer to absolute CFC values).

	Accuracy requirement (Mean bias error)	Precision requirement (Bias-corrected root mean square error)
<i>Threshold</i>		
Hourly means	5%	35%
Daily means	5%	25%
Monthly means	5%	15%
<i>Target</i>		
Hourly means	3%	30%
Daily means	3%	20%
Monthly means	3%	10%
<i>Optimal</i>		
Hourly means	1%	25%
Daily means	1%	15%
Monthly means	1%	5%

instantaneous level-2 data. The accuracy and precision are then provided on the instantaneous time scale (for which no requirement is formulated) and for the collocations aggregated to daily and monthly time scale. Inter-comparisons of Meteosat CFC with other satellite-based data records are performed based on monthly means, except PATMOS-x for which collocations are first rendered at instantaneous level, and then aggregated to monthly. Details on the validation (or comparison) exercises are given in each section of results dedicated to a reference (or inter-compared) data record.


The chosen validation references may be subdivided into two groups:

Group 1. Independent cloud observations:

- Cloud amount observations from surface stations (SYNOP) (time period 1983–2020)
- Cloud amount from the CALIPSO cloud lidar (CALIOP) (selected year 2010)

Group 2. Other satellite-derived data records:

- Cloud amount from Moderate Resolution Imaging Spectroradiometer (MODIS) (time period 2004–2020)
- Cloud amount from PATMOS-x (time period 1983–2020)
- Cloud amount from CLARA-A2 (time period 1983–2020)
- Cloud amount from CLAAS-A2 (time period 2004–2020)
- Cloud amount from ESA-Cloud-CCI project CC4CL AVHRR (time period 1983–2016)
- Cloud amount from ISCCP-H project (time period 1983–2017)

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

The first group of observations is the most important one since it fulfills the condition that the observation reference must be independent. Thus, results achieved from comparisons with this group of observations will be related to the expected accuracy and precision requirements.

To assess the accuracy of Meteosat CFC at both level-2 and level-3 we use the mean bias error, i.e. mean difference between Meteosat CFC and reference data. To express the precision of Meteosat CFC compared to a reference data record, the bias-corrected root mean squared error (bcRMSE) is used. In addition, to compare the discrete level-2-based binary cloud condition (clear or cloudy derived from CFC thresholds), we use the skill scores such as probability of detection (POD), false alarm ratio (FAR) and the Hanssen-Kuiper's discriminant (KSS). These scores can be viewed as a measure of precision. Although requirements for skill scores are not stated in AD 1 for the product under validation (CM-23012), in the following we refer to target requirements formulated for the CM SAF SEVIRI-based data record (CM-21011), i.e. 90% for POD and 15% for FAR. See Appendix A.1 for details on performance statistics and skill scores used herein. For a trend analysis we use linear trends derived using Theil-Sen estimates (Theil, 1950) and their significance is estimated with the Mann-Kendall test (Kendall, 1938; Mann, 1945).

In the following, the used reference data records are introduced in Section 5. Section 6 presents validation results performed against reference data records of Group 1, including an analysis of decadal stability (Section 6.1.3). Further, Section 7 gives an assessment of differences in the performance of Meteosat CFC derived from MFG and MSG. Then, results of the inter-comparisons with data records of Group 2 are presented in Section 8, followed by the final conclusions (Section 9).

## 5 Data records for comparison with Meteosat CFC

### 5.1 Manual cloud observations from surface stations (SYNOP)

Observations of cloud amount made at meteorological surface stations (i.e. synoptic observations hereafter called SYNOP) constitute one of the data records used to evaluate Meteosat CFC. All observations from national meteorological services or weather stations that were recorded, encoded and transmitted to ECMWF following the standard of the World Meteorological Organization were considered. At manned stations the total cloud cover (also known as cloud amount) is visually estimated by human observers. In contrast, ceilometers are used for that purpose at automatic stations. However, for data quality and consistency reasons, only those SYNOP reports provided by manned stations were taken into account.

Two selections of SYNOP sites were used in this report, both taken from the ECMWF archive. First, a reference SYNOP data record covering the full time range is used to assess if Meteosat CFC fulfills accuracy and precision requirements (Table 3). It consists of quality-controlled synoptic observations (1983–2020) at 290 sites. Second, a selection of a greater number of SYNOP sites provided a source for a grid-based comparison of Meteosat CFC with other satellite-based data records for the year 2005. In both cases synoptic observations used for training the Bayesian classifier of Meteosat CFC within the GeoSatClim framework (RD 1) were excluded. The division of synoptic observations into the two data records is caused by a different number of sites available for 1 and 38 years, as well as different quality control procedures applied to 1-year and multi-annual time series. For both selections, the original okta scale (0–8) of observations was transformed to cloud fractional cover as

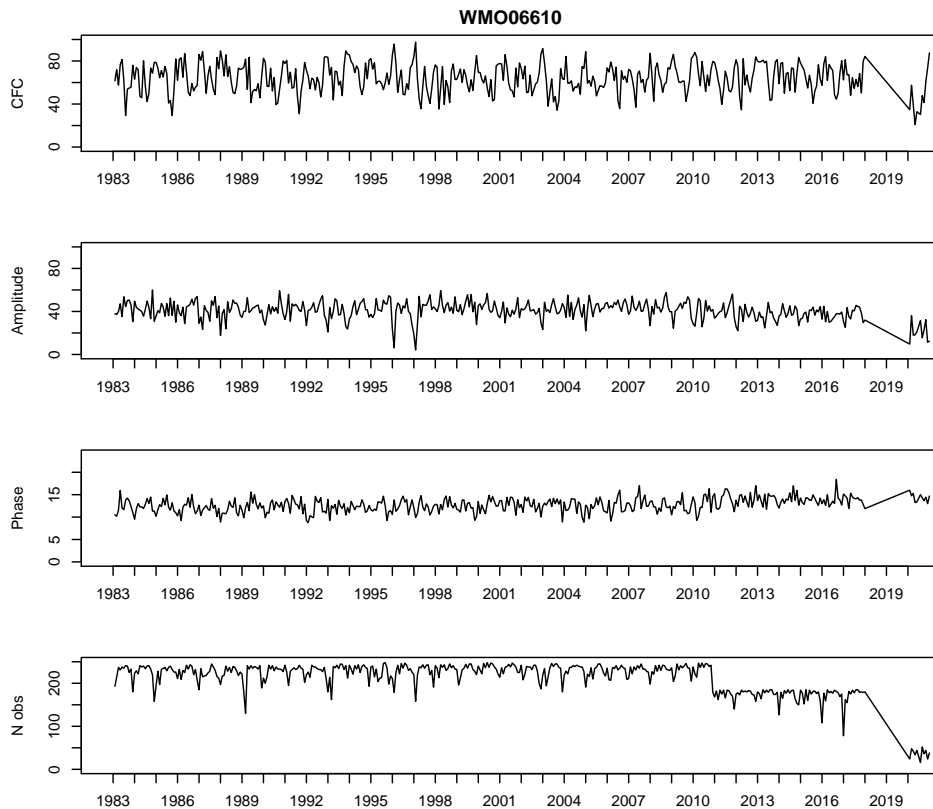


Figure 3: SYNOP cloud observation statistics of Payerne Switzerland (WMO Site No. 06610). From top to bottom: mean cloud fraction (CFC), diurnal CFC amplitude, diurnal CFC phase (time of maximum CFC), number of monthly CFC observations.

0, 10, 25, 40, 50, 60, 75, 90, 100% (WMO standard).

### 5.1.1 Reference SYNOP (1983–2020)

The initial ECMWF archive contains data for over 6000 globally distributed sites. From these we selected sites for a geographic range of 60°N to 60°S and 60°W to 60°E, and for which the satellite viewing angle is below 70°. In order to ensure the collection of long-term homogenous data series, we selected stations where observations were continuously performed in 1983–2020, at least every 6 hours with a maximum break of 30 days. For each site we used cloud amount observed with the highest temporal frequency (up to 1 hour) that was reported for the whole 38-year period. Thus the frequency of observations could vary between sites, but remained stable in time for each site. Further, we excluded sites for which the Standard Normal Homogeneity Test (SNHT, see Section A.2 for details) detected any inhomogeneity in a time series of cloud amount monthly anomalies.

The amount of SYNOP sites with enough stability and without gaps usable for validating 38 years has slightly increased from 237 to 290. This was possible by accessing a more consistent SYNOP archive at ECMWF where all WMO SYNOP messages are stored. However, much of the southern hemisphere are not covered and we have further made alarming observations regarding the state of some SYNOP observations in Europe:

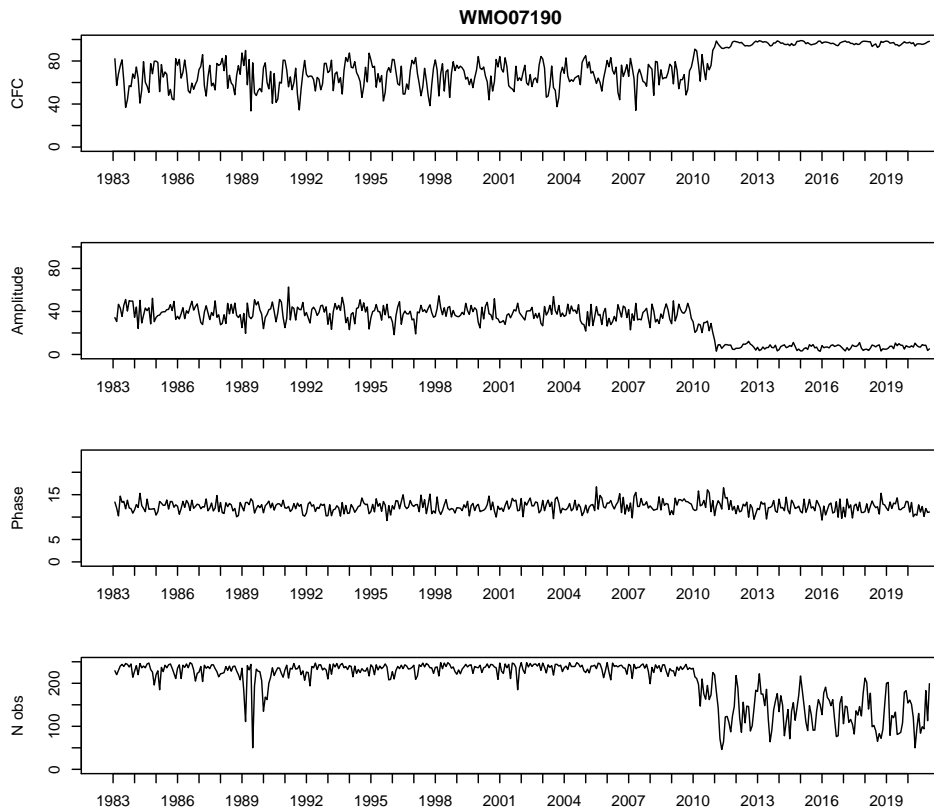


Figure 4: SYNOP cloud observation statistics of Strassbourg France (WMO Site No. 07190). From top to bottom: mean cloud fraction (CFC), diurnal CFC amplitude, diurnal CFC phase (time of maximum CFC), number of monthly CFC observations.

- High quality sites like Payerne in Switzerland start to have gaps which disallow trend analysis and inter-comparison to satellite observations. Payerne reduced the number of observations in 2011, did not report any SYNOP cloud amount during 2018 and 2019 and resumed in 2020 with around 10-20% of its previous observation frequency (Fig. 3). While there are still enough available SYNOP observations in central Europe, Payerne is of high relevance because it hosts an array of automatic ground and ground-based remote sensing instruments together with aerosol, water vapor and upper atmospheric profile measurements. Such a set of instrumentation cannot be often found. But it is required to develop satellite-based algorithms and to quality check / de-bias human-based cloud observations.
- Whole countries like France start to degrade their SYNOP observations. France reported the full 0–8 okta range until 2010 and replaced its observations with a method where they only report CFC above 6 okta. It decreases the dynamic range of cloud observations to 6–8 okta, rising mean cloud amount by 20% and decreasing diurnal amplitude (Fig. 4). At the same time the number of observations were reduced to 50–70% of pre-2011 conditions. As shown in Fig. 5 a) a large gap now results in western Europe covering all of France.

Since SYNOP stations are unevenly distributed in geographic space, a regular grid spacing was created. For each  $2^\circ \times 2^\circ$  geographic grid cell the SYNOP site with the most valid observations was used. This yielded a total of 290 SYNOP sites distributed over the full Meteosat disc (Fig. 5a).

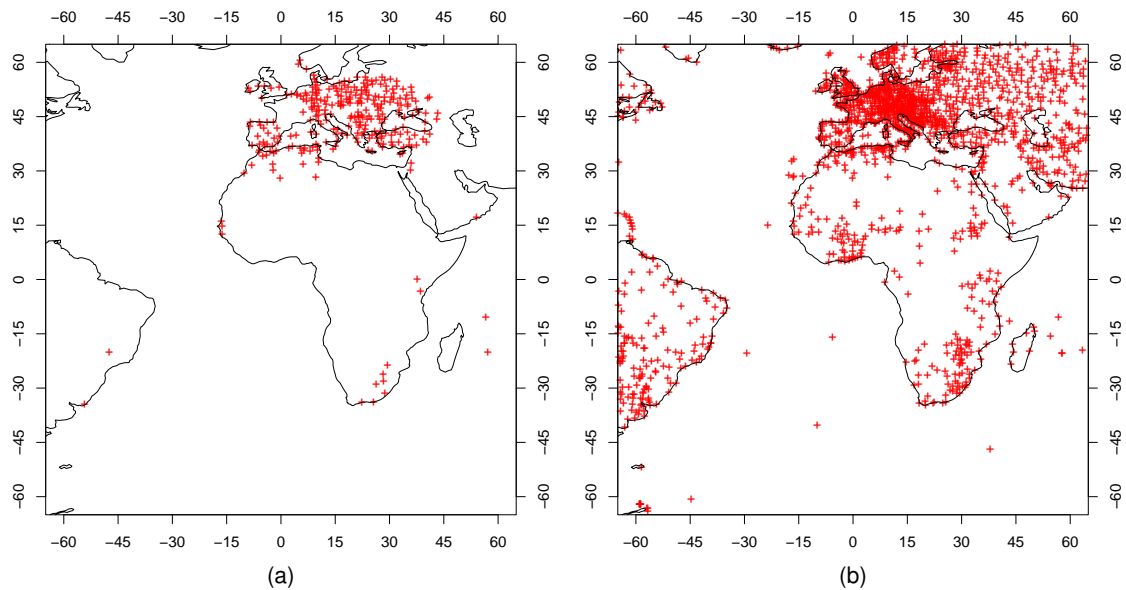


Figure 5: SYNOP sites used as a reference to validate Meteosat CFC for 1983–2020 (a), and to inter-compare Meteosat CFC monthly means (2005) with other satellite-based data records at a  $1^\circ \times 1^\circ$  resolution (b). Note that synoptic observations shown in (b) were spatially aggregated.


### 5.1.2 Gridded SYNOP (2005)

Grid-based comparisons of Meteosat CFC and other satellite-based data records (listed in Section 4) in relation to SYNOP were performed based on CFC monthly means of 2005. For this purpose the gridded monthly means of SYNOP CFC were created. The selection of reference SYNOP sites (described in a previous section) was driven by a need for accuracy and homogeneity (even with a cost of a limited number of sites). In order to create gridded SYNOP monthly means we selected 2326 sites providing observations in 2005 seamlessly within  $65^\circ\text{N}$  to  $65^\circ\text{S}$  and  $65^\circ\text{W}$  to  $65^\circ\text{E}$  window (Fig. 5b). Further we aggregated instantaneous observations to monthly means, and then performed a spatial aggregation from point data into a  $1^\circ \times 1^\circ$  grid by calculating the mean of all sites located within a grid cell. This resulted in 1143 grid cells with monthly mean SYNOP CFC.

### Uncertainty and error sources

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 behavior related to the way people have been 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

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/GFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

some dependence on solar zenith angle, 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. This is natural 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 become 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 okta).
- It is important to consider that most SYNOP stations are located on land 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 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 nighttime. However, the variability (precision) probably reaches higher absolute values and it is largest during night conditions. This may lead to a strong seasonal variation in quality 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. This may possibly lead to improved accuracies at night time but there is also a considerable risk that the precision figures degrades, mainly as an effect of that ceilometers only observing a very small fraction of the sky.

Despite their subjective character and varying quality, SYNOP observations still provide the most useful reference data set suitable for monitoring and validating space-based estimations of cloud coverage, particularly due to their long-term availability. They remain the only reference for long term studies on climatological cloud cover state, variability and trends.

## 5.2 A-Train (CALIPSO-CALIOP)

Measurements from space-borne active instruments (radar + lidar) provide the most accurate information on cloud presence in the atmosphere. Measured reflected radiation comes almost exclusively from cloud and precipitation particles and is therefore not “contaminated” by radiation from the ground surface or atmospheric constituents as is the case with most passive radiometers. In this validation study we have decided to utilize measurements from the CALIOP lidar instrument carried by the CALIPSO satellite, included in the A-Train series of satellites (Fig. 6). The Cloud-Aerosol

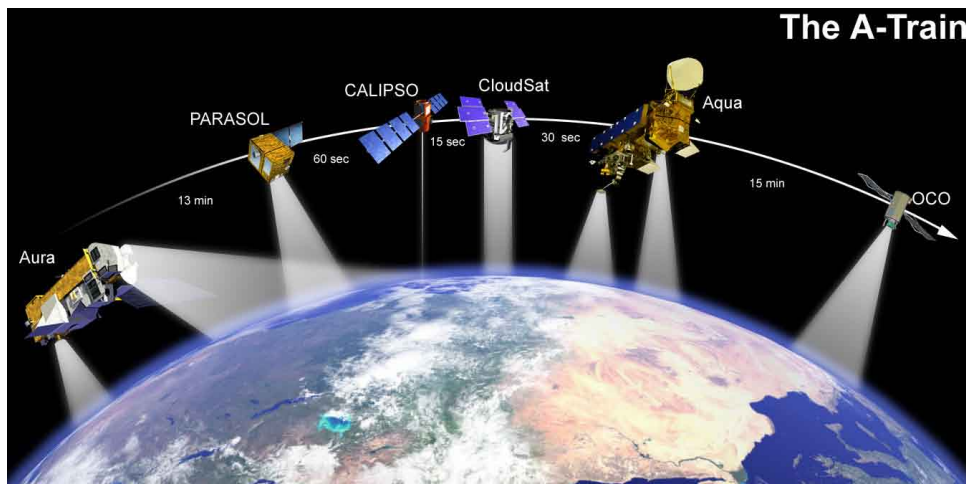


Figure 6: The Aqua-Train satellites. (Image credit: NASA)

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. CALIOP provides detailed profile information about cloud and aerosol particles and corresponding physical parameters.

CALIOP measures the backscatter intensity at 1064 nm while two other channels measure the orthogonally polarized components of the backscattered signal at 532 nm. 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 measurement 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 (resolution based on the spacing between consecutive footprints of 70 m), 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 detection and identification of cloud and aerosol layers compared to the original footprint resolution. 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 5-km cloud layer data record versions 3-01 (CALIPSO Science Team, 2015) for validation purpose since this resolution is closest to the nominal SEVIRI resolution and because of the advantages described above. It 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 4. 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



Table 4: 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:	Alto cumulus (transparent)
Category 5:	Altostratus (opaque, As, Ns, Ac)
Category 6:	Cirrus (transparent)
Category 7:	Deep convective (opaque As, Cb, Ns)

seen by CALIOP.


### Uncertainty and error sources

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 70 m, the individual footprint of the lidar beam. Thus, to compare CALIOP-derived results with the results of 3–5-km MVIRI/SEVIRI pixel data is not entirely consistent (i.e., CALIOP is only capable of covering the MVIRI/SEVIRI 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 MVIRI/SEVIRI scale will be detected, e.g., it is very unlikely to imagine elongated clouds with size and shapes below  $0.3 \times 4$  km that might risk remaining undetected within a MVIRI/SEVIRI pixel that coincides with a CALIOP measurement. Most clouds will have aspect ratios for the two horizontal directions that guarantee detection by CALIOP.

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 by 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 (i.e., saying something on the performance of the applied cloud detection method for clouds that should be theoretically detectable) one should consider filtering out the contributions from the very thinnest clouds. We have employed this approach in this validation study.

The cloud detection efficiency with CALIOP is slightly different during 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 the Meteosat CFC validation context we have used the 5 km CALIOP data record since this resolution is closest to the nominal SEVIRI resolution. However, the results in different data records from CALIPSO, related to different horizontal resolutions (with the five options 333 m, 1 km, 5 km, 20 km and 80 km) are unfortunately not entirely consistent (see e.g., Karlsson and Johansson, 2013). It means that some of the thick (opaque) boundary layer clouds that are reported in fine

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

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 boundary layer clouds are removed from the original signal to facilitate detection of very thin cloud layers and aerosols.

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 percent in the CALIOP-derived cloud cover when using the 5 km data record. 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 accumulated results based on a large number of full global orbits. This also concerns problems with reduced signal-to-noise ratios due to solar contamination during daytime.

Note that the CALIPSO validation was not updated since Edition 1 since the chosen year 2010 was hardly modified by the performed algorithm changes and prolongation of the time series.

### 5.3 MODIS

MODIS (Moderate Resolution Imaging Spectroradiometer) is an advanced imaging instrument on-board the Terra (EOS AM) and Aqua (EOS PM) polar satellites (see <http://modis-atmos.gsfc.nasa.gov/index.html>). Terra's orbit around the Earth is sun synchronous and timed so that it 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.

We have used the level-3 Combined Terra and Aqua COSP files derived from MODIS Standard L2 Cloud Fraction Product derived from the MODIS Cloud Mask ([http://dx.doi.org/10.5067/MODIS/MCD06COSP\\_M3\\_MODIS.061](http://dx.doi.org/10.5067/MODIS/MCD06COSP_M3_MODIS.061)). The data are monthly  $1^\circ \times 1^\circ$  grid average values of the selected atmospheric parameter. Statistics are sorted into  $1^\circ \times 1^\circ$  cells on an equal-angle grid that spans a (calendar) monthly interval and then summarized over the globe. We used CFC combined from Terra and Aqua Collection 6.1, thus starting from 2002 when Aqua was launched.

#### Uncertainty and error sources

Stubenrauch et al. (2012) reports, mainly based on comparisons with CALIPSO-CALIOP, a probability of detection (POD) ranging from 70–75% over Polar Regions in the Polar night to roughly 90% over all other surfaces (including Polar Regions in the Polar summer).

### 5.4 PATMOS-x

PATMOS-x stands for "Pathfinder Atmospheres Extended", a suite of cloud products developed by NOAA. The CFC global data record was generated from the AVHRR sensors onboard NOAA satellites. 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)

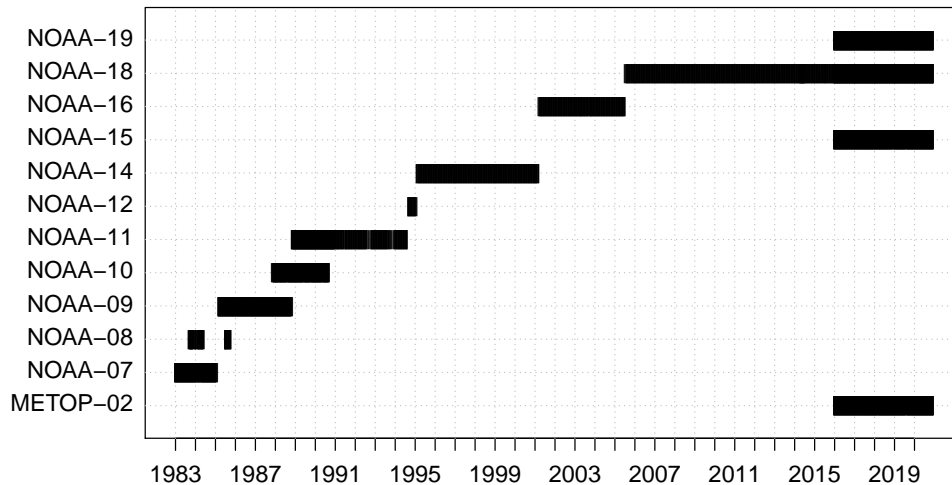



Figure 7: Overview of AVHRR measurements used to calculate the PATMOS-x CFC monthly means.

employing AVHRR radiances in all available spectral channels. Cloud fraction was computed from results of a statistical naïve Bayesian cloud mask trained on CALIPSO-CALIOP cloud information (Heidinger et al., 2012).

The pixel-level product used here is based on measurements from global area coverage (GAC) data. Each GAC pixel is the average of four 1.1-km AVHRR pixels. The geographic area represented by one GAC pixel is approximately 3 km by 5 km. Specifically we use the level-2b product (<https://www.ncdc.noaa.gov/cdr/atmospheric/avhrr-cloud-properties-patmos-x>), which is the level-2 pixel-level data subsampled to a 0.1° equal-angle global grid. Since PATMOS-x's CFC monthly means are not provided by the data producer, the collocations Meteosat CFC–PATMOS-x were performed at level-2 with a maximum time difference of 10 minutes (see Bojanowski et al., 2014), and then aggregated to monthly means. To limit the amount of data to process, we used the ascending node of one satellite per month (Fig. 7). Only simultaneous Meteosat CFC and PATMOS-x observations were used to calculate the monthly means. Similarly, when comparing with SYNOP, CFC from all three sources had to be present with a maximum time difference of 10 minutes. The level-2-based collocation was chosen to avoid the influence of temporal aggregation, which is sensitive to, e.g., choice of sensor(s) in case of satellite mission overlap (e.g., NOAA-18, METOP-A), missing data and aggregation method.

### Uncertainty and error sources

Stubenrauch et al. (2012) reports, mainly based on comparisons with CALIPSO-CALIOP, a probability of detection (POD) ranging from 70–75% over Polar Regions in the Polar night to roughly 90% over all other surfaces (including Polar Regions in the Polar summer). Deviations with respect to ISCCP and MODIS Level 3 data records are reported to be less than 5% over most regions.

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

## 5.5 CLARA-A2

The CLARA-A2 data record is a second version of the CM SAF's global data record of cloud, surface albedo and surface radiation products derived from homogenized measurements of the Advanced Very High Resolution Radiometer (AVHRR) onboard the polar orbiting NOAA (7–18) and Metop (A) satellites. The horizontal resolution is close to 1 km at nadir but only reduced resolution of approximately 4 km are permanently archived and available with a global coverage. The products, including macrophysical and microphysical cloud properties, are provided on a L2 and L3 (daily and monthly mean) basis. Herein we used CFC monthly means compiled from all AVHRR sensors over a time period 1982–2019 and provided by CM SAF on a global regular latitude-longitude grid with 0.25 degree resolution.

### Uncertainty and error sources

The CLARA-A2 monthly means are not corrected for orbital drift and for sampling issues. The respective overpass times of polar sensors change during the lifetime of satellites. Prior to 1991 there was only a single satellite with an afternoon orbit. From 1991 to 2002 both a morning and an afternoon orbit were covered. After the launch of METOP-A in 2002 four satellites can cover a large part of the diurnal cycle. Users of CLARA-A2 monthly mean data need therefore to be careful with interpreting trends as they may be due to sampling issues of the diurnal cycle of cloud cover (K.-G. Karlsson, personal communication).


According to the CLARA-A2 validation report (CM SAF, 2016a), CLARA-A2 CFC is on average about 3% lower than SYNOP and CALIOP. Precision of CLARA-A2 CFC monthly means is of 7% as compared to SYNOP.

## 5.6 CLAAS-A2

The CLAAS-2 data record is based on 12 years (2004–2017) of Level 1.5 SEVIRI data provided by EUMETSAT. The data comprise images that have already undergone certain modifications by EUMETSAT: they have been corrected for all unwanted radiometric and geometric effects, geolocated using a standardized projection and calibrated. The time series of SEVIRI reflectances was carefully calibrated against the Moderate Resolution Imaging Spectroradiometer (MODIS), which improved the retrieval of microphysical parameters. The calibration method is outlined in Meirink et al. (2013). With the calibration also homogenization was performed because the same MODIS instrument was used for all MSG satellites. The IR radiances of SEVIRI were used as provided by EUMETSAT, relying on the on-board black-body calibration.

Cloud fraction and cloud top properties were derived using following Derrien and Le Gléau (2005), and the CPP (Cloud Physical Properties) algorithm (Roebeling et al., 2006), which retrieves cloud thermodynamic phase, cloud optical thickness, cloud particle effective radius, and liquid/ice water path. All parameters can be accessed on Level 2 basis, which means in original SEVIRI pixel size (3×3 km at sub-satellite point) and at 15-minute resolution.

The level-2 (instantaneous pixel) cloud mask comprises 6 categories: cloud free, cloud contam-

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/CFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

inated, cloud filled, undefined, snow/ice at the surface (under cloud-free conditions), and non-processed. In this report we employed CFC product available as monthly composites on a regular latitude/longitude grid with a spatial resolution of  $0.05^\circ \times 0.05^\circ$ , which is defined as the fraction of cloudy pixels per grid box compared to the total number of analysed pixels in the grid box. Pixels are counted as cloudy if they belong to the classes cloud filled or cloud contaminated. Fractional cloud cover is expressed in percent.

## Uncertainty and error sources

According to the CLAAS-2 validation report (CM SAF, 2016b), CLAAS-2 CFC is on average about 4% and 3% higher than SYNOP and CALIOP, respectively. Precision of CLAAS-2 CFC monthly means is of 10% as compared to SYNOP.

## 5.7 CC4CL AVHRR (ESA-Cloud-CCI)

The Community Optimal estimation Cloud retrieval for Climate (CC4CL) AVHRR data record is the cloud physical properties ECV's produced by the ESA-Cloud-CCI project. The project initiated by European Space Agency (ESA) has focused on cloud studies in the frame of its Climate Change Initiative (CCI) running over the time period 2010–2016. The ESA-Cloud-CCI has aimed at adapting and developing the state-of-the-art cloud retrieval schemes to be applied to the longest existing time series of the cloud observations available from polar orbiting satellites with AVHRR and AVHRR-like sensors (Stengel et al., 2015). The cloud properties (i.e. cloud cover, cloud top height and temperature, cloud optical thickness, cloud effective radius, and liquid and ice water paths) are derived by means of an optimal-estimation-based retrieval framework for: (1) the AVHRR heritage product (1982–2016) comprising (Advanced) Along Track Scanning Radiometer (A)ATSR, AVHRR and Moderate-Resolution Imaging Spectroradiometer (MODIS), and (2) the (A)ATSR-Medium Resolution Imaging Spectrometer (MERIS) product (2002–2012).

The CC4CL is an optimal estimation retrieval that can be used to determine cloud properties from visible/infrared satellite radiometers. CC4CL is based on the ORAC retrieval (Oxford RAL retrieval of Aerosol and Cloud) algorithm (Poulsen et al., 2012), with further developments done in this project. The algorithm fits radiances computed from LUTs created from DIScrete Ordinate Radiative Transfer (DISORT) (Stamnes et al., 2000) to the TOA signal measured by the satellite by varying the cloud optical depth, effective radius, cloud top pressure, phase and surface temperature.

The AVHRR-derived products are pixel-based level 2 and global level 3 data projected on an equal-angle grid. The level-3 products are provided as daily near-nadir with 0.05 degree horizontal resolution. In this report we employed monthly averages (1991–2014) aggregated on a 0.5 degree latitude-longitude grid for each individual instrument.

## Uncertainty and error sources

ESA-Cloud-CCI project is still ongoing and final data records are not yet provided to the public. Therefore the number of independent evaluations are limited. Referring to the product validation

report, the global CFC monthly means derived from AVHRR sensors reveal bias from -10 to 5%, and a RMSE of 10-20% as compared to synoptic observations.

## 5.8 ISCCP H

The International Satellite Cloud Climatology Project (ISCCP) provides data on various aspects of cloud variables like cloud amount, optical thickness, and cloud top temperature. The data, which have been collected since 1983, are available at 3-hour and monthly intervals with global coverage. The product has more than 200 variables for physical and chemical characteristics, identification methods, and ancillary data (e.g., snow and ice cover, atmospheric temperature and humidity, etc.). ISCCP-H uses input data from full-resolution Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) and 10 km geostationary imagery – an increase from 30 km in the ISCCP-D series. The H series CDR period of record extends from July 1983 through June 2017 (Young et al., 2019).

## 6 Validation of Meteosat CFC

### 6.1 Validation against synoptic observations

Synoptic total cloud cover observations in 1983–2020 from 290 sites (see Section 5.1.1) were used for the evaluation of instantaneous (level-2) Meteosat CFC. Performance statistics for daily and monthly means are also provided, however these were calculated from match-ups rendered at level-2. This means that exclusively Meteosat CFC estimates which had corresponding synoptic observations (undertaken at the same UTC time and within a range of a given pixel) were collocated and further aggregated to daily/monthly means. This evaluation reveals the accuracy and precision of instantaneous Meteosat CFC, aggregated to daily and monthly means. Evaluation of level-3 monthly means is shown in Section 8, where Meteosat CFC is inter-compared with gridded SYNOP and other satellite-derived data records.

In order to evaluate the performance of Meteosat CFC both mean bias error (MBE, accuracy parameter) and bias-corrected root mean square error (bcRMSE, precision parameter) were calculated and then compared to the target requirements specified in Table 3 (page 17). Values for bcRMSE are expected to be larger for instantaneous estimates than for daily or monthly means. For MBE, requirements are the same for instantaneous and aggregated data. In addition, we assessed the performance of the Meteosat CFC-based binary cloud mask by means of skill scores. The binary cloud mask was defined as: clear (0–1 okta,  $CFC \leq 10\%$ ) and cloudy (7–8 okta,  $CFC \geq 90\%$ ). The skill scores are widely used by the climate modeling and cloud remote sensing community. We provide them for comparability reasons despite no requirements on binary cloud detection are defined for the continuous Meteosat CFC as part of the CM SAF Product Requirements Document (AD 1).

### 6.1.1 Instantaneous (level 2)

Table 5 reveals detailed statistics, both averaged over all match-ups as well as for different climate zones, illumination and viewing angles, seasons and the different satellites. The overall difference between Meteosat CFC and SYNOP is -0.7% with a bias corrected RMSE of approximately 29%.

Averaged over 290 sites, MBE of Meteosat CFC is close to 0% except for Arid and Ocean zones, for which the bias is negative of approximately -6 to -8% (Table 5). However, bias at individual sites (Fig. 8e) does not reveal distinct spatial patterns, except an overestimation of 5–15% in Anatolia. A decrease in precision of Meteosat CFC estimates can be seen in Europe from SW to NE (Fig. 8f) due to a lower probability of detection (Fig. 8b).

Meteosat CFC slightly overestimates SYNOP (MBE=3.6%) and is less precise (bcRMSE=34%) during nighttime. It underestimates SYNOP during daytime (MBE=-1.4%) but with higher precision (bcRMSE=24%). It is endorsed by a more detailed performance assessment of Meteosat CFC in dependence on sun zenith angle (Fig. 9a, 9c, and 9e). Figure 10a reveals almost linear but very minor decrease of POD with an increase of sun zenith angle, while FAR (Fig. 10c) has a distinct jump from 8 in daytime to 1 during night-time acquisitions (Table 6).

On average, at low viewing zenith angles ( $VZA < 30^\circ$ ) Meteosat CFC underestimates SYNOP by approximately 10% (Table 8). This dependency is due to the very few long-term SYNOP sites in the arid zone of Africa located at a low view angle and is discussed further down in this section. A decrease of Meteosat CFC precision with increasing VZA is explicit (Fig. 9d, 9f and 10f).

A seasonal cycle of Meteosat CFC performance is revealed by a negative bias and larger bcRMSE during winter months (DJF), and a positive bias and lower bcRMSE in summer months (JJA) (Table 5). However the impact of seasons is diminished, as statistics are averaged over the northern and southern hemispheres. The performance cycle can be seen more distinctly in the time series of monthly MBE and bcRMSE shown on page 37.

Table 5: Performance statistics of level-2 Meteosat CFC as compared with synoptic observations at 290 sites.

	N	mean Meteosat	mean SYNOP	MBE	bcRMSE
overall	40913095	58.51	57.82	0.68	28.54
Cold	19277694	65.06	64.22	0.84	29.70
Temperate	17454963	55.37	54.59	0.78	27.52
Tropical	3240251	36.46	37.23	-0.77	26.23
Arid	233024	50.27	56.16	-5.89	31.88
Ocean	83214	50.99	58.86	-7.87	27.06
day	19750494	58.15	59.56	-1.41	22.93
night	15287019	58.48	54.86	3.62	33.91
twilight	5875582	59.77	59.69	0.08	29.60
DJF	10055486	67.80	68.33	-0.53	32.17
JJA	10555889	47.89	46.91	0.98	25.41
MAM	10263138	59.30	57.85	1.44	27.36
SON	10038582	59.55	58.75	0.80	28.94
MFG	20486319	59.12	58.10	1.01	29.20
MSG	20426776	57.89	57.54	0.35	27.87
VZA < 30	209231	38.29	47.95	-9.65	32.09
VZA 30-50	6565693	39.34	40.71	-1.37	26.23
VZA > 50	34138171	62.32	61.18	1.14	28.92



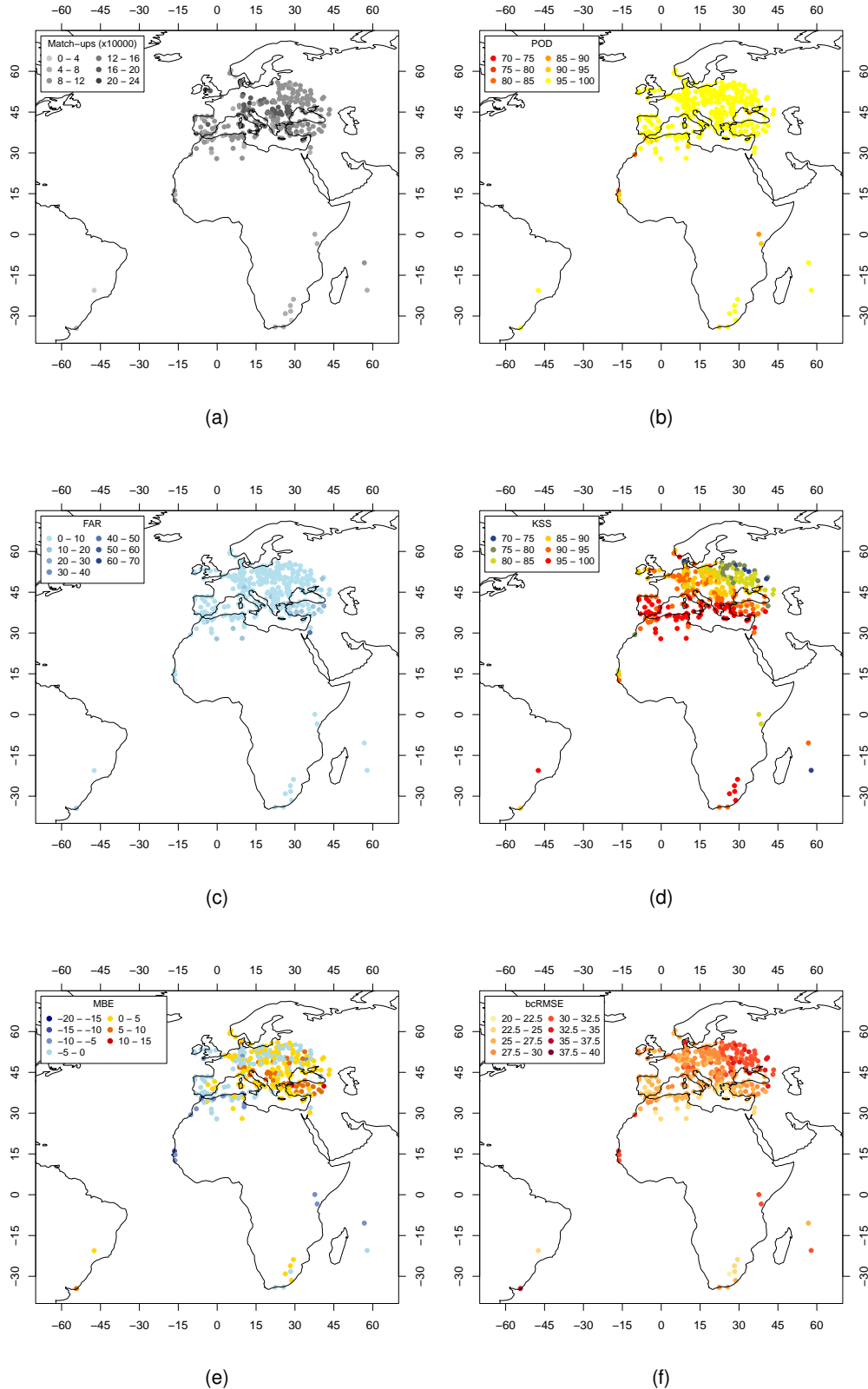


Figure 8: Skill scores and performance statistics of the instantaneous Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites: (a) number of Meteosat CFC–SYNOP match-ups, (b) probability of detection, (c) false alarm ratio, (d) Hanssen-Kuiper’s discriminant, (e) mean bias error, (f) bias-corrected root mean square error. Note that (a) presents a total number of match-ups (used for calculation of MBE and bcRMSE), which is lower for a binary mask.

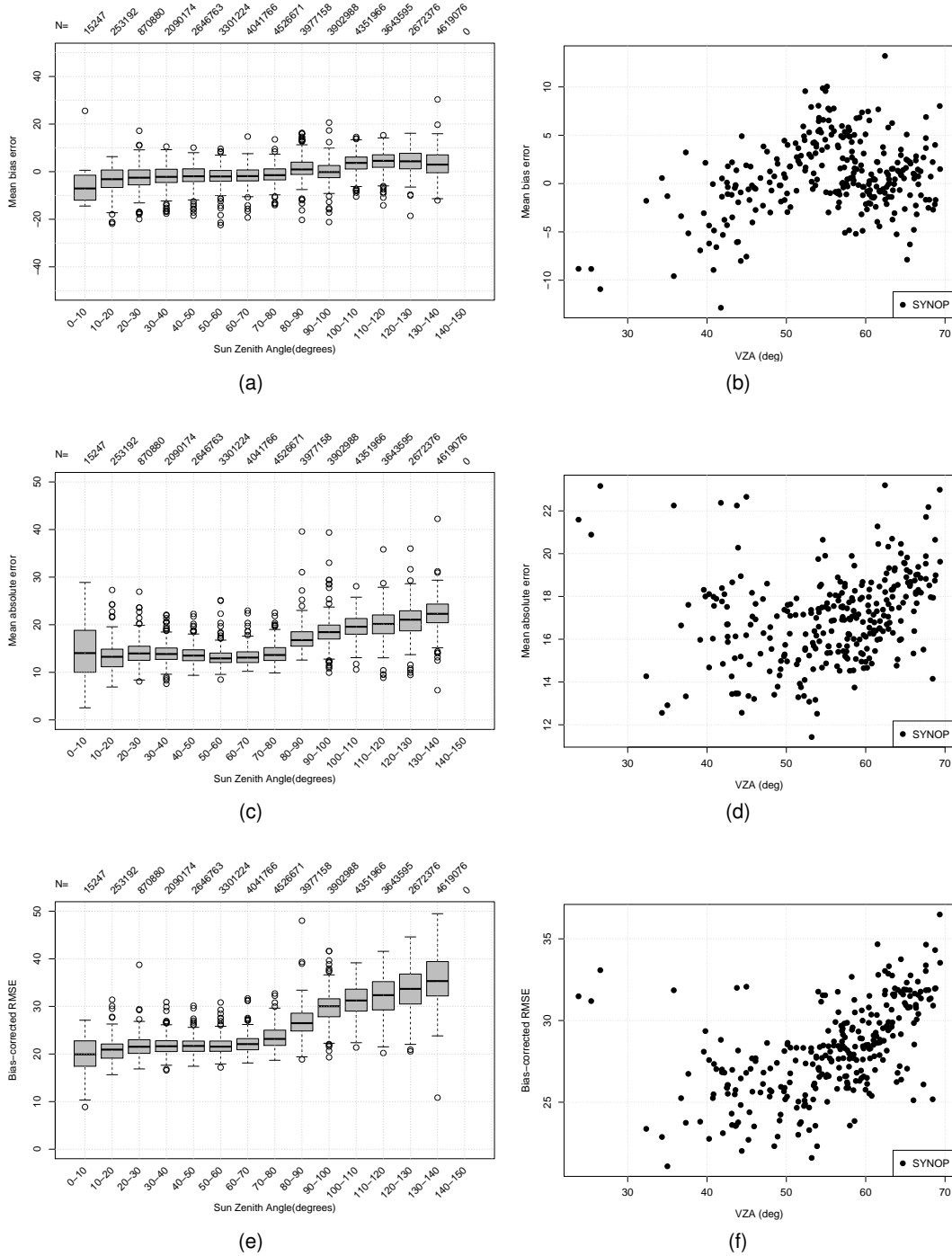


Figure 9: Performance statistics of the level-2 Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites in relation to sun zenith angle (left) and satellite view zenith angle (right).

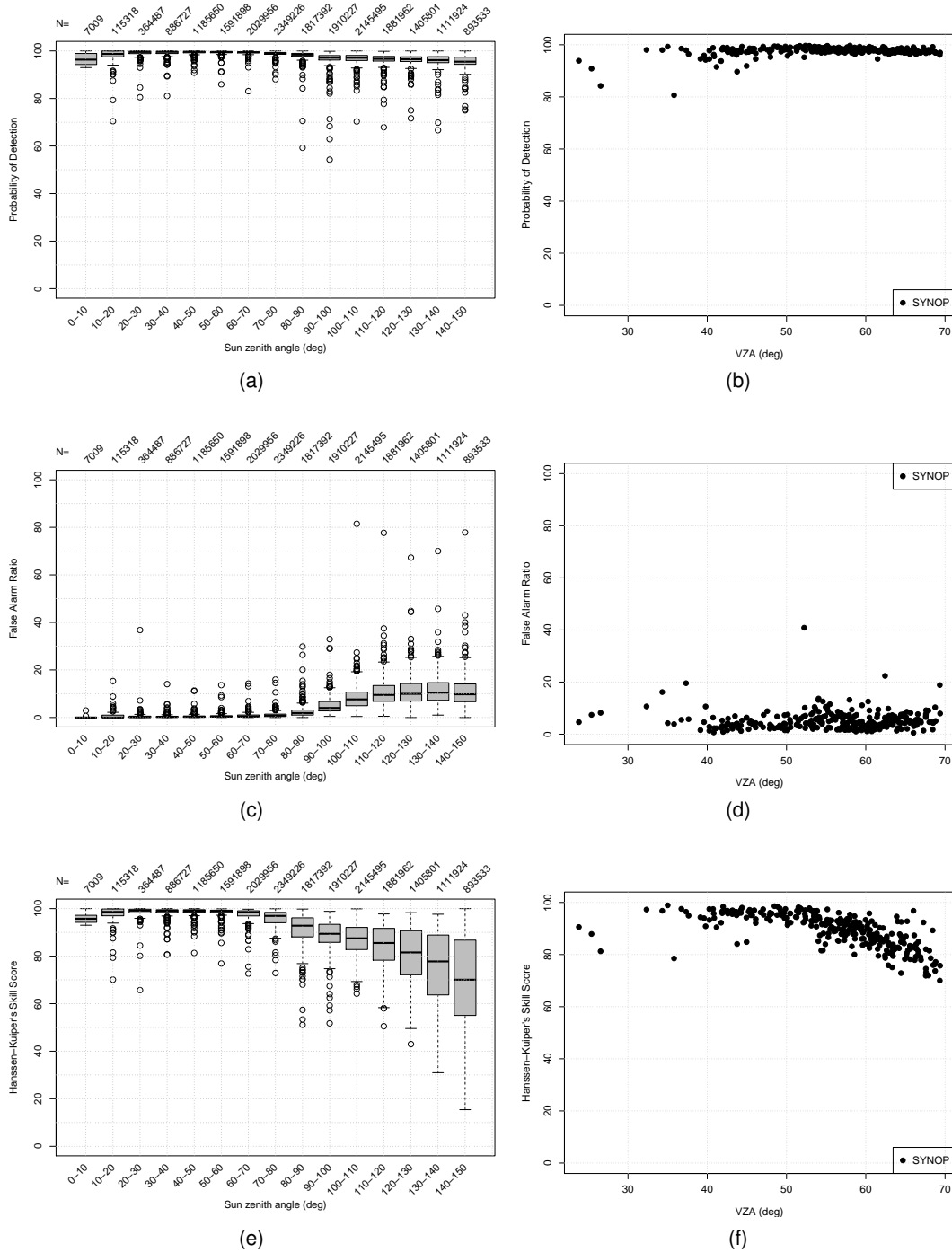


Figure 10: Skill scores of the level-2 Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites in relation to sun zenith angle (left) and satellite view zenith angle (right).

### 6.1.2 Hourly, Daily and monthly means

Averaged over all 290 reference SYNOP sites, Meteosat CFC fulfills the optimal accuracy and simultaneously the target precision requirements (Table 3) for hourly, daily and monthly means. The optimal accuracy requirements are: 1% for MBE met by hourly, daily and monthly Meteosat CFC

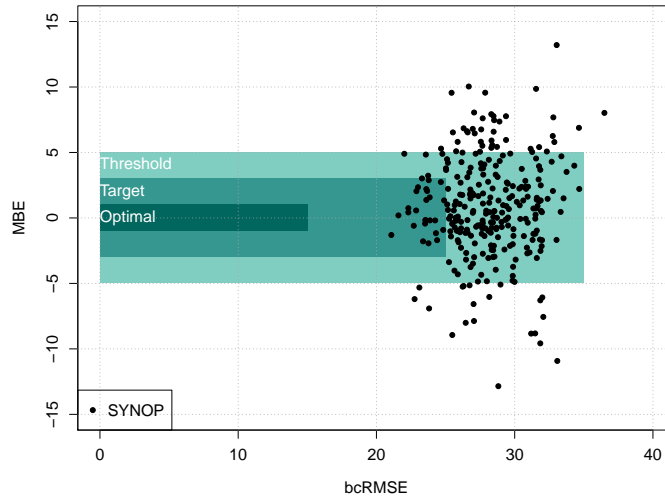
Table 6: Skill scores of the instantaneous Meteosat CFC during 1983–2020 as compared to synoptic observations at 290 sites.

	N	mean Meteosat	mean SYNOP	POD	FAR	KSS
overall	20115720	63.90	62.60	97.84	4.15	90.76
Cold	9919015	74.10	72.50	97.74	4.44	85.78
Temperate	8328425	58.20	57.20	98.12	3.56	93.27
Tropical	1461727	26.20	25.70	96.73	5.03	94.95
Arid	73999	54.90	54.90	93.96	6.09	86.53
Ocean	21909	74.70	74.10	98.73	2.07	92.78
day	9450439	64.40	64.50	99.20	0.73	97.88
night	7857823	62.20	58.90	96.32	8.67	83.19
twilight	2807458	67.00	66.50	97.21	3.46	90.29
DJF	5394662	78.40	76.80	96.89	5.14	79.52
JJA	4855853	44.50	43.90	98.55	2.85	96.29
MAM	4879036	65.50	64.10	98.63	3.44	92.35
SON	4986169	65.60	64.00	97.84	4.41	89.81
MFG	10036370	64.70	63.40	97.69	4.24	90.21
MSG	10079350	63.20	61.80	98.00	4.05	91.29
VZA < 30	85103	32.60	33.90	90.18	6.28	87.09
VZA 30-50	2824071	31.40	31.00	97.33	4.03	95.49
VZA > 50	17206546	69.40	67.90	97.90	4.15	88.92

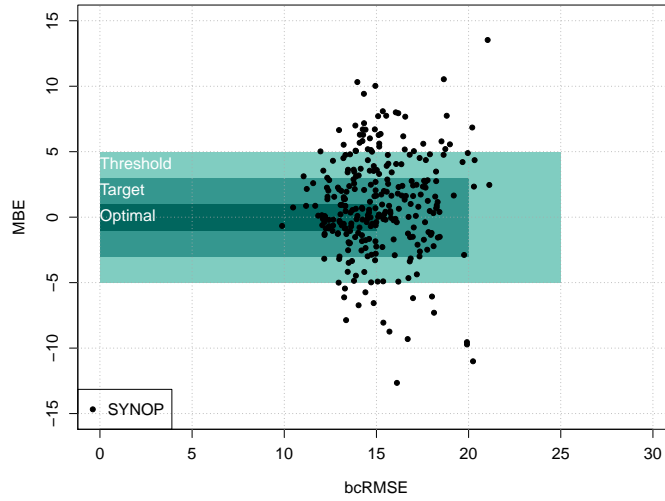
means. Similarly, bcRMSE of 20.2%, 12.4% and 6.1% comply with target precision requirements of 30%, 20% and 10% for hourly, daily and monthly means, respectively (Table 7). Taking accuracy and precision requirements simultaneously into account — optimal, target and threshold requirements are met by 20%, 61% and 80% of sites, respectively (Figure 11).

Table 7: Performance statistics of Meteosat CFC hourly, daily and monthly means (1983–2020) as compared to synoptic observations at 290 sites.

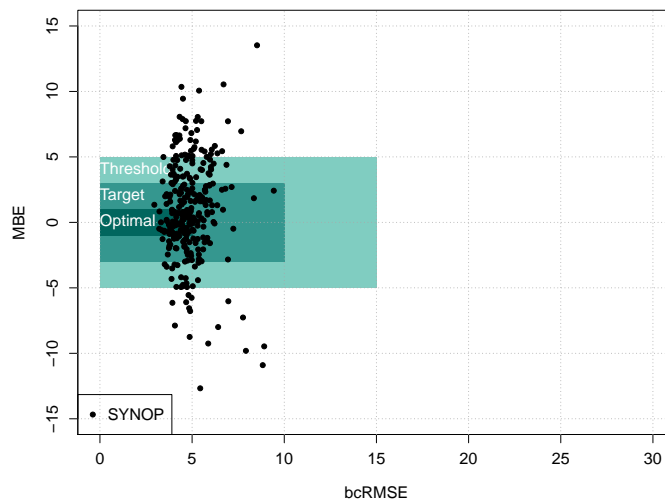
temporal aggregation	N	mean Meteosat	mean SYNOP	MBE	bcRMSE
hourly	3173729	49.6	49.1	0.5	20.2
daily	272838	49.8	49.3	0.5	12.4
monthly	8987	49.9	49.3	0.5	6.1



(a)



(b)



(c)

Figure 11: Performance statistics of the Meteosat CFC hourly (a), daily (b) and monthly (c) means during 1983–2020 as compared to synoptic observations at 290 sites.

### 6.1.3 Decadal stability of monthly means

The AD 1 specifies the requirements for the decadal stability of Meteosat CFC, which are of 0.3%, 2%, and 5% for optimal, target and threshold, respectively. Decadal stability reveals the change of CFC accuracy in time (i.e. per decade). To assess the stability we computed MBE and bcRMSE for each month in 1983–2020 over 290 SYNOP sites. As shown by a dashed line in Figure 12a, the trend in bias is of -0.68% per decade, thus within the target requirements.

The breaks in stability were tested by the relative SNHT test which, following the guidelines of Aguilar et al. (2003) and Toreti et al. (2011), was carried out based on the de-trended mean monthly cloud fraction difference between Meteosat CFC and SYNOP. At the beginning and at the end of the time series  $T(k)$  of the relative test exceeds the critical value, which is equal 10.02 at the 95% confidence level (see Appendix A.2 for SNHT definition and details). The explanation for this homogeneity, very likely caused by non-climatic factors, is not trivial, as there was no change of satellites in those periods (see Figure 2). We further see that other long and independent satellite series like CLARA-A2, PATMOS-x or ISCCP-H (see further down) exercise a positive bias towards SYNOP, which could lead to a hypothesis on changed SYNOP observation practices in the early 1980s. Such a hypothesis would need to be further examined.

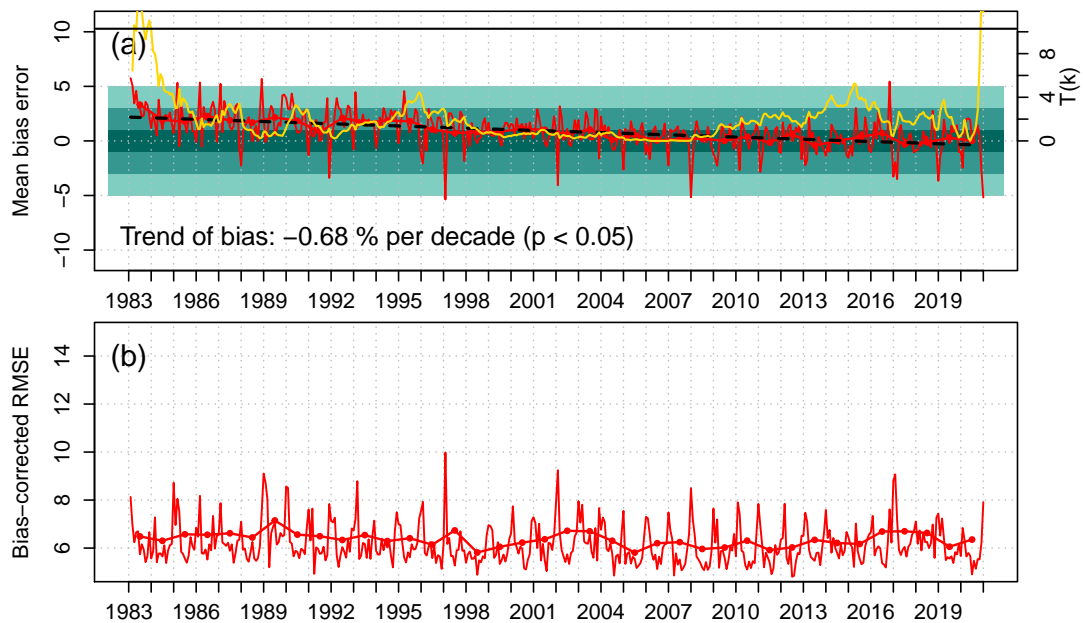


Figure 12: Time series of mean bias error (a) and bias-corrected root mean square error (b) of Meteosat CFC as compared to synoptic observations at 290 sites in 1983–2020. Thick lines with filled circles give yearly mean statistics. The black dashed line represents a Theil-Sen linear trend provided with its Mann-Kendall statistical significance. Colored rectangles reveal the accuracy requirements. A yellow solid line reveals the  $T(k)$  statistic from the Standard Normal Homogeneity Test.

## 6.2 Validation against A-Train (CALIPSO-CALIOP)

The validation against CALIOP on CALIPSO requires the spatial and temporal matching of observations from SEVIRI (used in Meteosat CFC for 2010) and CALIOP. Collocations were computed using spatial nearest neighbor search and scan line-based time matching. For each CALIOP measurement, the nearest neighbor in the SEVIRI grid is determined yielding its line number in the image. The matching SEVIRI scan is given by the scan where the acquisition time of this particular scan line is closest to the timestamp of the CALIOP measurement. An example matchup is shown in Figure 13. Maximum collocation distances are 5 km and 7.5 minutes in space and time. Note that the CALIPSO validation was not updated since Edition 1 since the chosen year 2010 was hardly modified by the performed algorithm changes and prolongation of the time series.

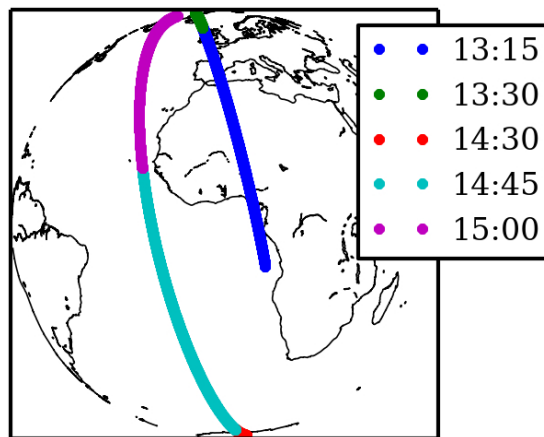


Figure 13: Exemplary matchup of the CALIOP track and the SEVIRI scan. The labels display the nominal timestamp of the SEVIRI scan that was matched to the identically colored ground track segment.

Due to the advanced lidar technique, CALIOP is much more sensitive to high and optically thin clouds than SEVIRI. Therefore, we do not only compare Meteosat CFC against the uppermost cloud layer detected by CALIOP, but also compare against CALIOP data which was filtered by means of the cloud optical thickness (COT). The latter can tell us more about how accurate Meteosat CFC is relative to the potential of the SEVIRI sensor.

We derived a binary CALIOP cloud mask by interpreting all CALIOP measurements with total column  $COT > 0$  (and  $COT > 0.2$ ) as cloudy. The Meteosat CFC was converted to a binary cloud mask by setting  $CFC > 50\%$  to cloudy and  $CFC \leq 50\%$  to clear sky. This was required to allow for a comparison against the CALIOP binary cloud mask.

Figure 14 shows a time series of cloud fraction from Meteosat CFC and CALIOP at all collocations in 2010. Meteosat underestimates CFC as compared to CALIOP ( $COT > 0.0$ ) and overestimate as compared to CALIOP ( $COT > 0.2$ ). The underestimation is expected as a passive sensor reveals lower sensitivity to thin clouds. The fluctuation of MBE revealed at the bottom panel of Fig. 14 need to be further investigated. Table 8 summarizes the performance of Meteosat CFC over all collocations (denoted as *All*), as well as for different subsets related to light conditions and land/ocean. Averaged over all months and collocations, Meteosat CFC bias is of -9.99% which is within the threshold

requirements. Probability of cloud detection is stable above 70% for all subsets. However, Meteosat CFC has a higher probability to incorrectly detect clear sky during night-time ( $FAR_{clr}=34.13\%$ ) compared to daytime ( $FAR_{clr}=30.24\%$ ). Yet, MBE (in absolute terms) is lower for night-time observations. Cloud detection is very similar in terms of accuracy and precision over land and sea.

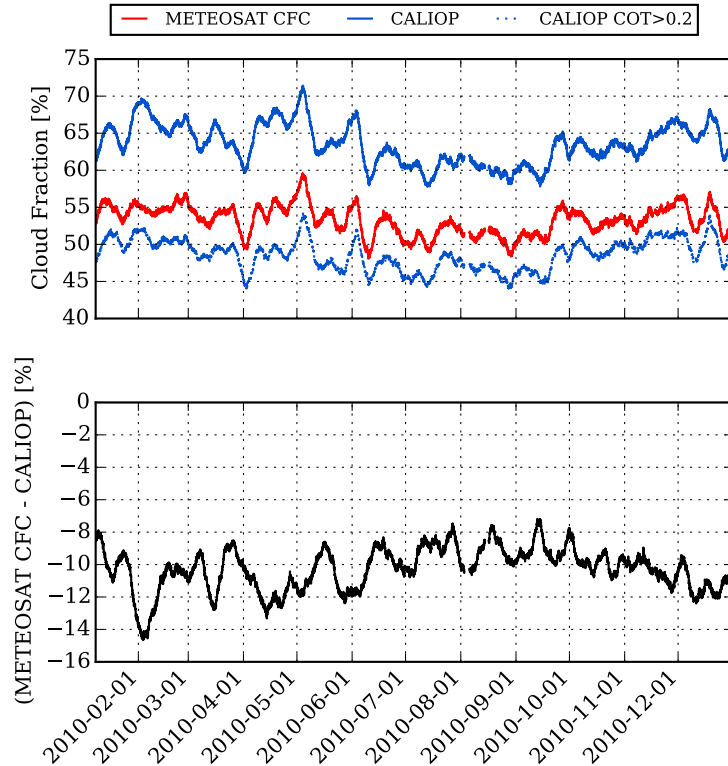


Figure 14: 5-day moving average of Meteosat CFC and CALIPSO cloud mask (yielding a cloud fraction, top) and a difference between both (bottom).

We also analyzed the impact of removing CALIOP measurements with total column COT smaller than a certain threshold. We found that SEVIRI is indeed less sensitive to optically thin clouds. In Figure 15 the probability of detection increases with the COT threshold used to distinguish clear and cloudy CALIOP measurements. However, it does not imply that optically clouds thinner than, say  $COT=0.1$ , can generally not be detected by SEVIRI, because the false alarm ratio also increases with the COT threshold. It is more likely to miss a cloud with SEVIRI, if it is optically thin. Thus there are two effects happening simultaneously when increasing the CALIOP COT threshold (Figure 15):

- Optically thin CALIOP clouds not detected by SEVIRI are reset to cloud-free, hence the cloud POD increases.
- Optically thin CALIOP clouds detected by SEVIRI are reset to cloud-free, leading to an increased False Alarm Ratio.

The coupling of these effects causes the Hitrate and KSS to peak at  $COT=0.15$ .

The contribution of different CALIOP cloud types to the probabilities of cloud detection are shown in Figure 16. Cloud analyses with respect to the CALIOP cloud type have a very strong ice cloud bias.



Table 8: Summary of validation results for the Meteosat CFC-based cloud mask against CALIOP. 'Clr' stands for clear, and 'cld' for cloudy. All values in percent (0–100).

	CALIOP COT>0.0					CALIOP COT>0.2				
	All	Day	Night	Sea	Land	All	Day	Night	Sea	Land
POD clr	86.38	88.20	84.37	85.64	88.12	80.18	81.46	78.76	79.06	82.29
FAR clr	32.10	30.24	34.13	31.78	32.55	11.88	9.39	14.61	12.01	11.59
POD cld	76.31	75.55	77.01	77.57	73.75	88.60	89.75	87.63	89.10	87.55
FAR cld	9.38	9.08	9.64	9.55	9.05	19.08	20.07	18.20	19.19	18.93
Hit rate	80.01	80.48	79.55	80.41	79.23	84.28	85.20	83.38	84.06	84.73
KSS	62.70	63.75	61.38	63.03	61.87	68.79	71.21	66.39	68.16	69.84
MBE	-9.99	-10.31	-9.68	-9.11	-11.69	4.62	5.54	3.72	5.10	3.71
bcRMSE	43.58	42.96	44.18	43.31	44.04	39.38	38.07	40.59	39.60	38.90

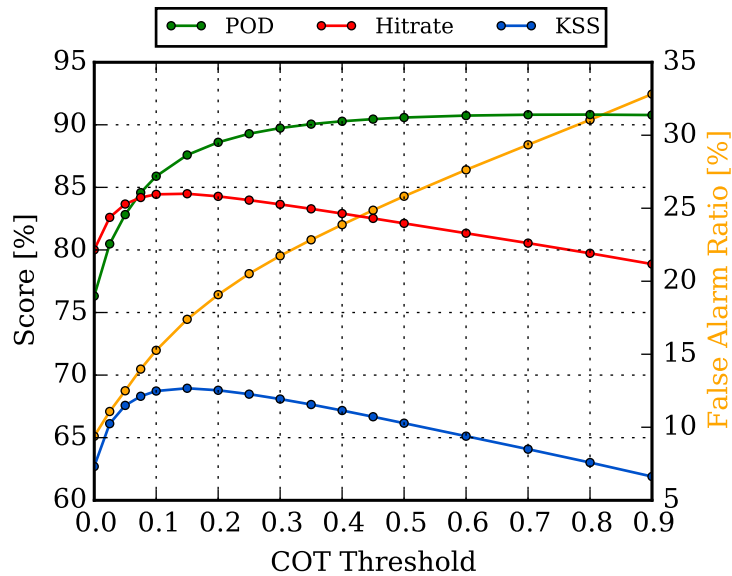


Figure 15: Meteosat CFC cloud scores as a function of the COT threshold used to discriminate clear and cloudy CALIOP observations. KSS denotes the Hanssen-Kuiper's Discriminant.

It is found that CALIOP provides a cloud type classification for 98% of the ice clouds, but only for 30% of the liquid clouds.

Altostratus and deep convective clouds are detected with almost 100% probability by Meteosat CFC, and Altocumulus almost reaches precision requirements of 90%. Cirrus clouds closely miss the threshold requirements. However, when ignoring CALIOP measurements with total column COT < 0.2, Cirrus meet target requirements. Transition Stratocumulus is detected with a probability of 60% for COT > 0.2. Concurrently, only 20% of low, broken Cumulus clouds are detected regardless a threshold for COT. However, the occurrence probability for this cloud type is two orders of magnitude lower than for example Cirrus clouds.

Figure 17 presents Meteosat CFC performance statistics over Meteosat disc remapped to a regular 1.5°×1.5°grid. Mean annual Meteosat CFC and CALIOP CFC (with COT > 0) reveal similar spatial patterns, however with Meteosat CFC underestimating CALIOP over the Atlantic Ocean within 0–10°N as well as for the tropics over Africa. This can be explained by a significant contribution

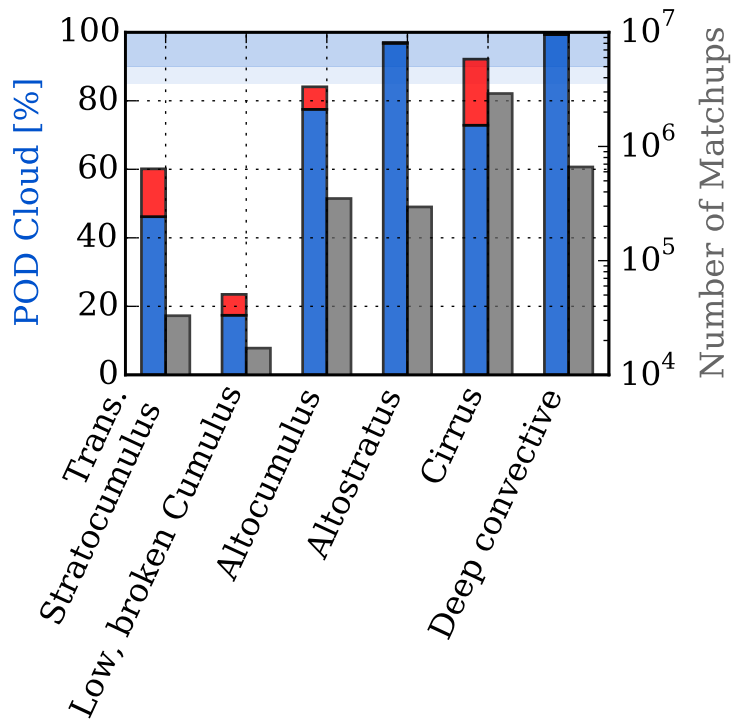


Figure 16: Probability of detection for Meteosat CFC cloud mask resolved by cloud type of the uppermost CALIOP cloud layer. Blue bars were computed by interpreting all CALIOP measurements with total column COT > 0 as cloudy. Red bars were derived using COT > 0.2 as cloud criterion. Gray bars give the number of matchups by cloud type.

of Cirrus clouds in this region, which can be missed by Meteosat CFC. This underestimation is indeed no longer visible when CALIOP CFC is estimated using COT > 0.2, thus excluding optically thin clouds. Probability of detection exceeds 90% over large areas i.e. Northern and Southern Atlantic Ocean, South America, Europe and South Africa. Two main spots of lower POD are located over Atlantic Ocean around 10°S–0°, Western Indian Ocean, as well as over desert (Sahara and Arabian Peninsula). Averaged over latitudinal bands (Fig. 18), POD does not comply with threshold requirement (of 85%) in a zone between 30°S and 50°N. However when excluding thin clouds by COT > 0.2, POD is above 85% for all latitudes. Concurrently, FAR complies with target requirements of 15% (Fig. 17).

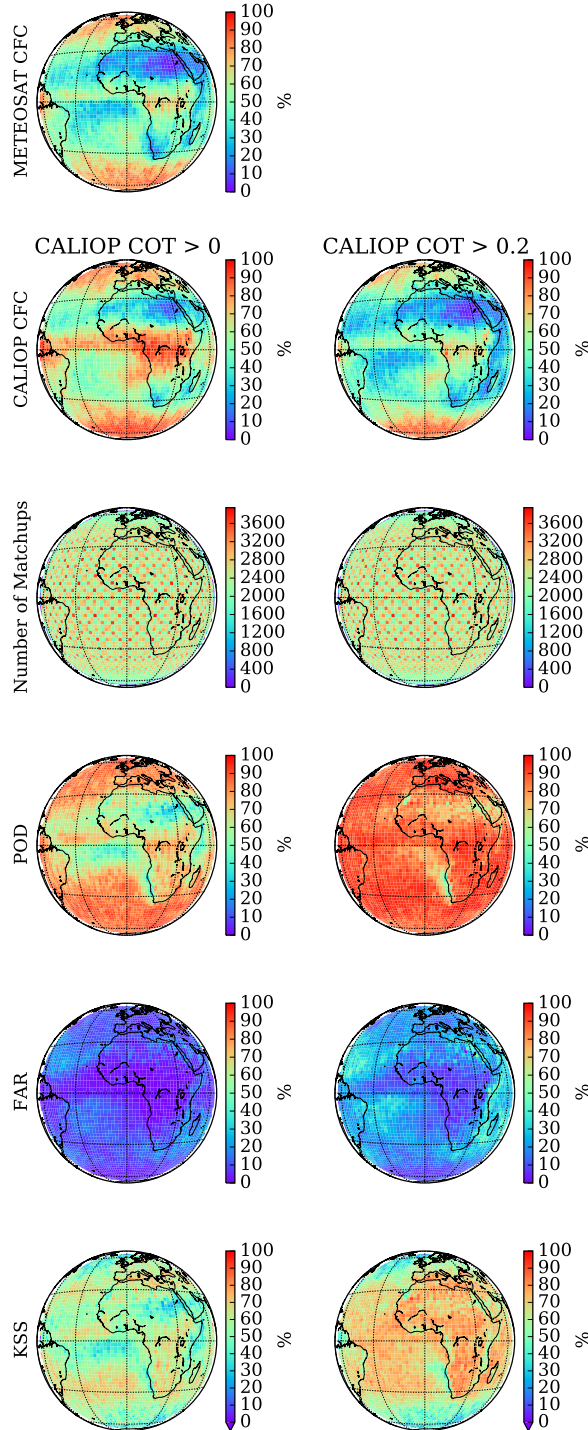


Figure 17: Mean annual CFC of Meteosat and reference CALIOP, and skill scores derived for COT > 0 (left column) and COT > 0.2 (right columns)

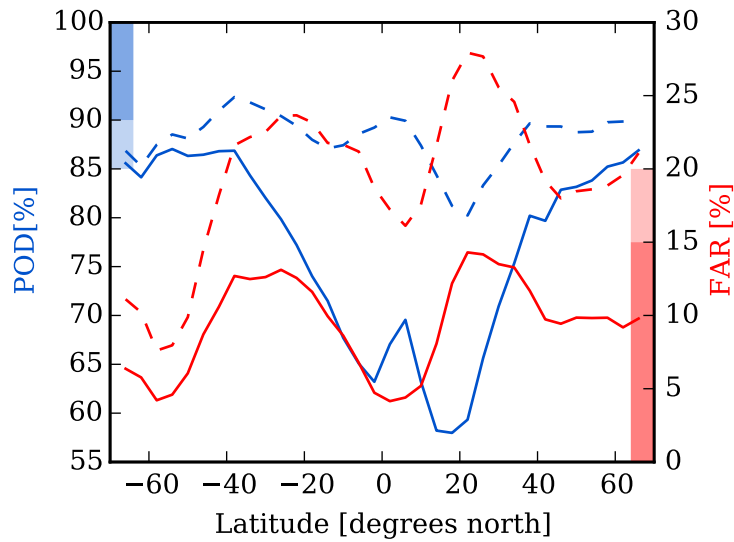


Figure 18: Meteosat CFC skill scores in 4-degree latitude bands. Darker and lighter blue/red shadings indicate the target and threshold POD/FAR requirements, respectively. Solid and dashed lines reveal analysis using total column COT > 0 and COT > 0.2 as CALIOP cloud criterion.

## 7 Comparison of Meteosat First and Second Generation

The homogeneity analysis of the Meteosat CFC bias against SYNOP does not reveal a break during the transfer from MFG/MVIRI to MSG/SEVIRI (2004 to 2005, see Section 6.1.3). This analysis was performed using reference CFC time series at 290 sites (Section 5.1.1). Differences in Meteosat CFC employing MVIRI and SEVIRI might still be present at locations where synoptic observations are not available (e.g. ocean regions). Therefore a more extensive comparison of monthly mean MVIRI- and SEVIRI-based CFC is carried out during 2004 and 2005 at the  $1^\circ \times 1^\circ$  resolution. MFG CFC and MSG CFC are compared at all  $1^\circ \times 1^\circ$  grids within  $60^\circ\text{N}/\text{S}$ – $60^\circ\text{W}/\text{E}$ , as well as against gridded SYNOP CFC monthly means (Section 5.1.2).

MSG-based CFC is only slightly more precise (bcRMSE=5.97%) than MFG-based CFC (bcRMSE=6.03%) as shown in Table 9. Spatial distribution of MBE among grids do not reveal any distinct differences between both satellites (Fig. 19e–f), with both mainly underestimating CFC in Africa at  $0^\circ$ – $2^\circ\text{N}$ . Compared to the long-term analysis presented before, SYNOP sites used here are not quality screened since the goal was to create a short term but spatially complete grid (see Section 5.1.1). Several grid cells especially in tropical Africa reveal possible issues with the site-based data, which is supported by the inter-comparisons to other satellite data in the next section.

Over all grids, CFC from MFG is on average 0.31% lower than CFC from MSG (Table 9). A similar difference is revealed for January, whereas for June the mean difference is close to 1%. Yet, the differences are not equally distributed in space (Fig. 19c–d). Concurrently, close to the African coast around  $10^\circ\text{S}$ – $20^\circ\text{S}$  (a region of tropical marine stratocumulus) MFG CFC has up to 5% larger values in January than MSG. Unfortunately these differences in ocean regions can neither be evaluated against synoptic observations, as these are not available, nor against CALIPSO/CALIOP launched only in 2006. Therefore, Meteosat CFC must be used with caution when analyzing cloud cover in tropical marine stratocumulus area where climate models have been shown to have difficulties in simulating the magnitude as well as the variability of albedo.

Table 9: Comparison statistics of gridded CFC derived by Meteosat First Generation, Meteosat Second Generation and SYNOP for 2004 and 2005.

	N	MBE	bcRMSE
MFG – SYNOP (all months)	6936	0.16	6.03
MFG – SYNOP (January)	578	1.11	6.54
MFG – SYNOP (June)	578	-0.10	5.83
MSG – SYNOP (all months)	6576	0.57	5.97
MSG – SYNOP (January)	548	0.88	7.24
MSG – SYNOP (June)	548	0.96	5.57

Overall Fig. 20 demonstrates that the seasonal cycle and diurnal cycle of MFG and MSG are highly comparable. The difference between the two satellite sensors are smaller than the difference of each sensor to the reference SYNOP. Also the occurrence statistics of CFC classes shown in the lower left sub-figure shows how precise the bayesian estimation of those classes operates for both sensors.

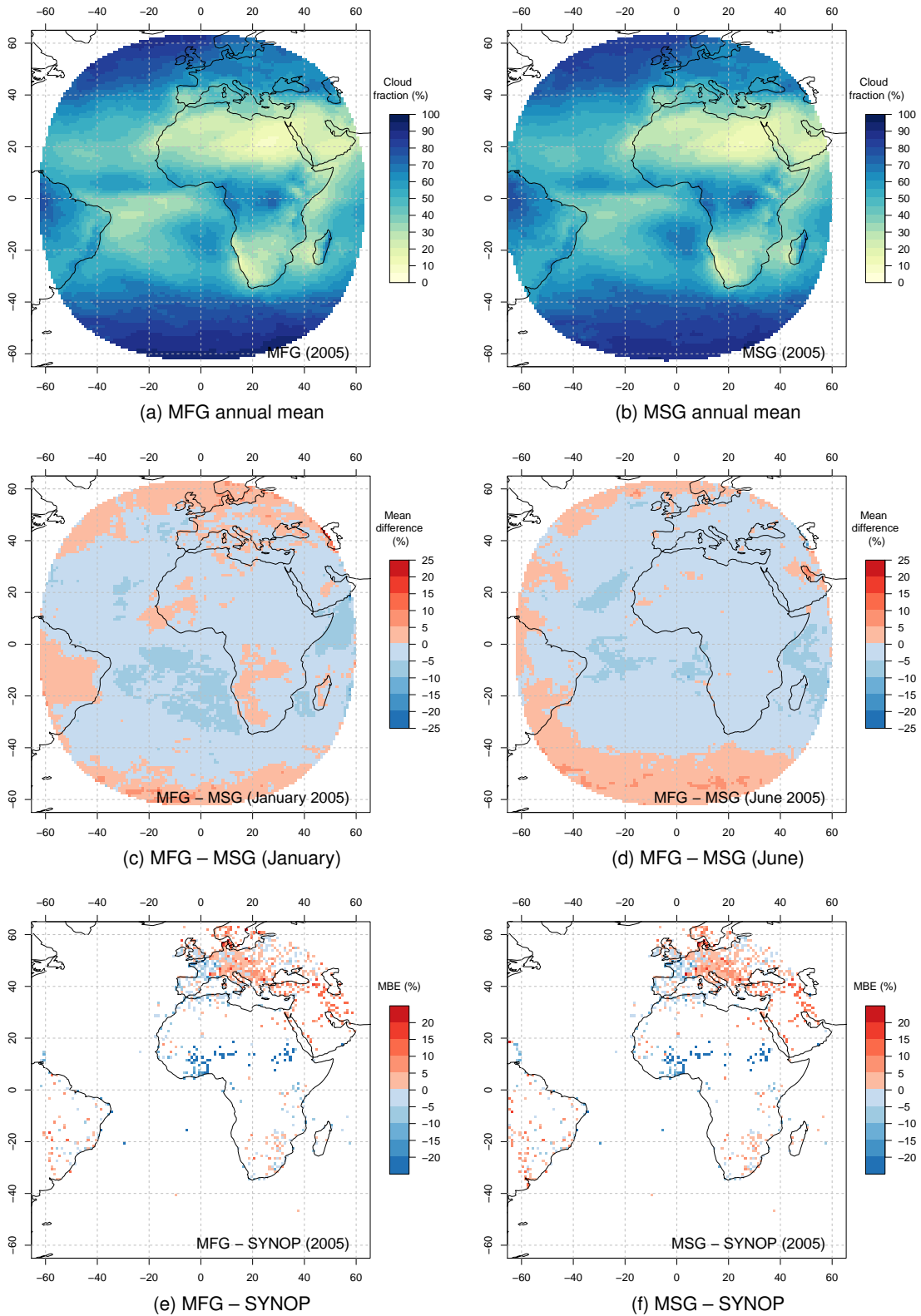


Figure 19: Comparison of Meteosat CFC derived from Meteosat First and Second Generation and gridded SYNOP monthly means of 2005 at  $1^\circ \times 1^\circ$  spatial resolution.

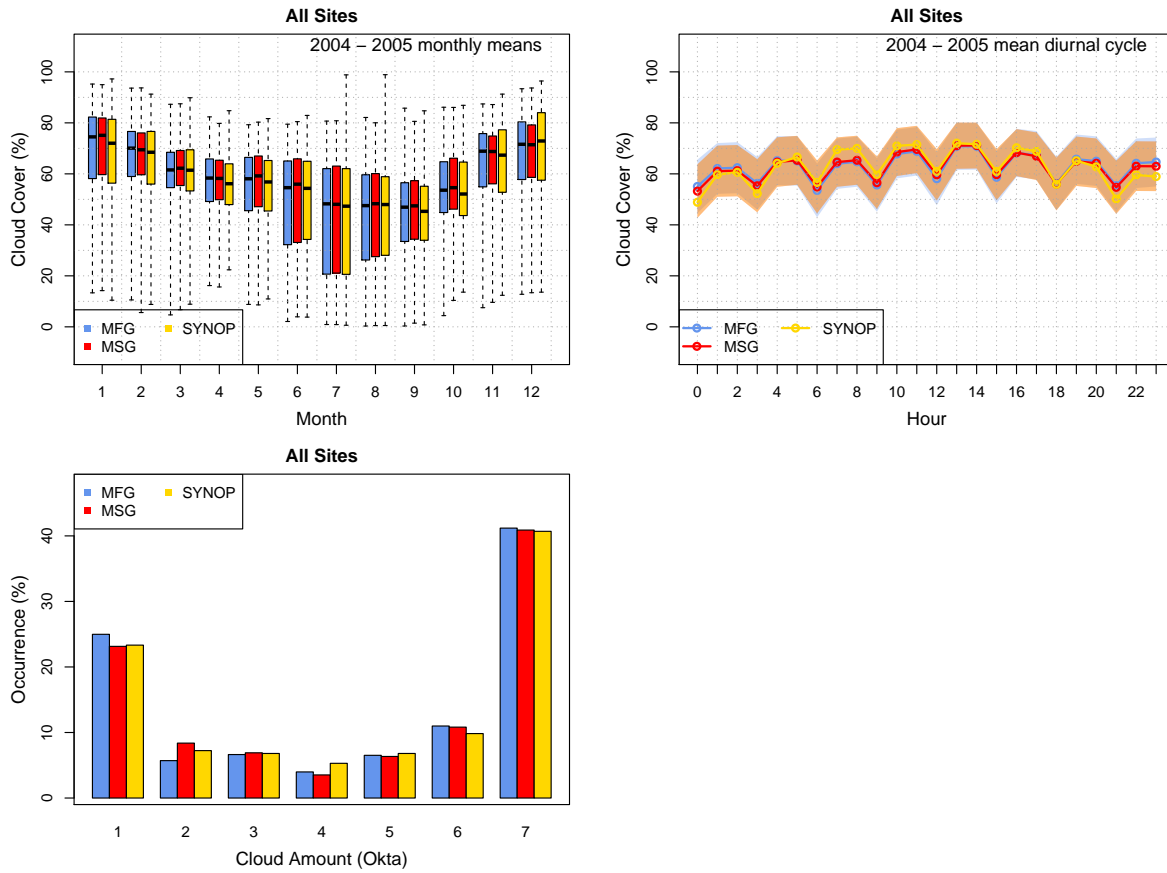


Figure 20: Monthly, diurnal and per-okta occurrence statistics of Meteosat CFC derived from Meteosat First and Second Generation and gridded SYNOP monthly means of 2004 and 2005 at  $1^\circ \times 1^\circ$  spatial resolution. The sawtooth pattern of the diurnal analysis is caused by the different observation statistics of 3-hourly vs. 6-hourly SYNOP observations across SYNOP sites.

## 8 Inter-comparison with other satellite-derived data records

Comparisons of Meteosat CFC with other satellite-based data record and synoptic observations were carried out based on level 3 monthly means. Temporal aggregation of each data record was independent. For Meteosat CFC, monthly means were derived from instantaneous estimates in 3 steps: (i) hourly means derived from instantaneous, (ii) daily means derived from at least 4 hourly values, and (iii) monthly means derived from at least 20 daily values. All inter-compared satellite-derived data records were obtained in level-3. The only exception was PATMOS-x, not available in the original level-3 monthly means, for which collocations with Meteosat CFC were made for level-2 data and further CFC's were aggregated to monthly.

To investigate differences between Meteosat CFC and each inter-compared data record we carried out three analyses:

1. Comparison of time series of CFC monthly means averaged over 290 reference SYNOP sites (see Section 5.1.1) for the whole overlapping period;
2. Comparison of CFC monthly means aggregated to  $1^\circ \times 1^\circ$  for grids where synoptical observations are available (see Section 5.1.2) for 2005;
3. Comparison of Meteosat CFC and inter-comparison data record at  $1^\circ \times 1^\circ$  for the whole Meteosat disc for 2005.

### 8.1 Inter-comparison with MODIS

Figure 21 presents the Meteosat CFC and MODIS CFC as compared to synoptic observations aggregated over 290 sites. Both satellite-based data records agree with SYNOP during summer months when CFC has its yearly minimum. Yet, the differences are evident for winter months (CFC seasonal maximum) when MODIS tends to overestimate and Meteosat CFC to underestimate SYNOP CFC. The underestimation of Meteosat CFC during winter months may be related to a deficient snow/cloud separation when using only two heritage channels of Meteosat. The MBE time series for both satellites reveal an opposite annual cycle (Fig. 21b), which is a basis for further investigation in order to prepare a next version of the Meteosat CFC. MODIS is less precise than Meteosat CFC according to bcRMSE shown in Figure 21c. Note that we have used the MODIS CFC derived from the MODIS cloud mask. The MODIS CFC derived from cloud physical properties tended to show the opposite behavior: underestimation compared to SYNOP.

These findings are confirmed by a comparison carried out at a grid level for 2005. For grids where SYNOP CFC is available, Meteosat CFC overestimates synoptic observations by approximately 1% while MODIS overestimates by more than 4%, and bcRMSE are 6.2% and 7.8%, respectively (Table 10). Yet, when aggregated over the whole Meteosat disc, Meteosat CFC is 0–10% lower than MODIS CFC. The difference between both data records is indeed mostly visible for the ocean (where SYNOP is not present), for which MODIS CFC is 5–35% higher than Meteosat CFC (Figure 22c). Larger difference are also seen for African desert areas, where MODIS is higher than Meteosat CFC.



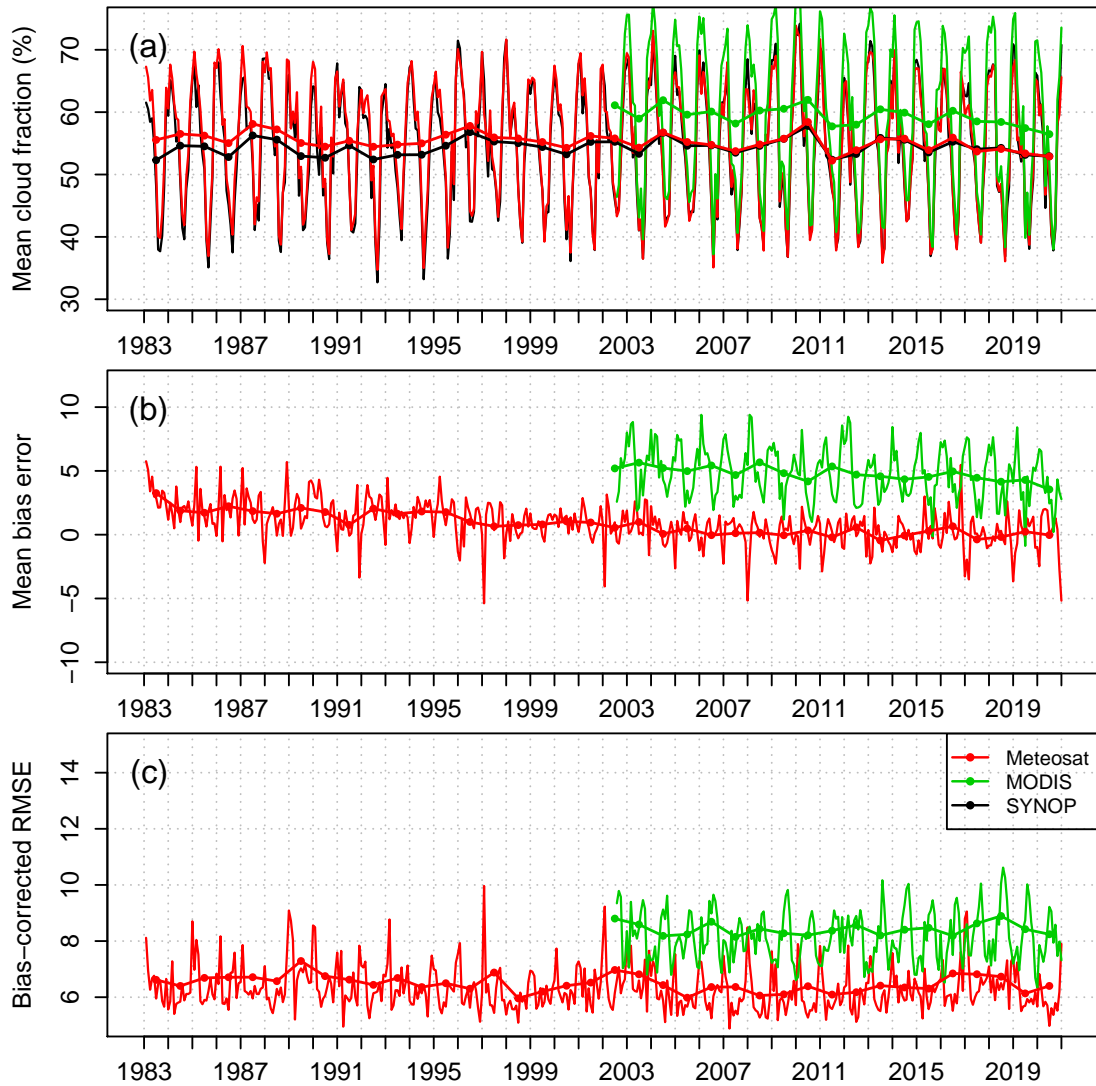


Figure 21: Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and MODIS as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics.

Table 10: Inter-comparison statistics of gridded Meteosat CFC, MODIS and SYNOP for 2005.

	MBE	bcRMSE
Meteosat CFC – SYNOP	0.92	6.16
MODIS – SYNOP	4.88	7.84

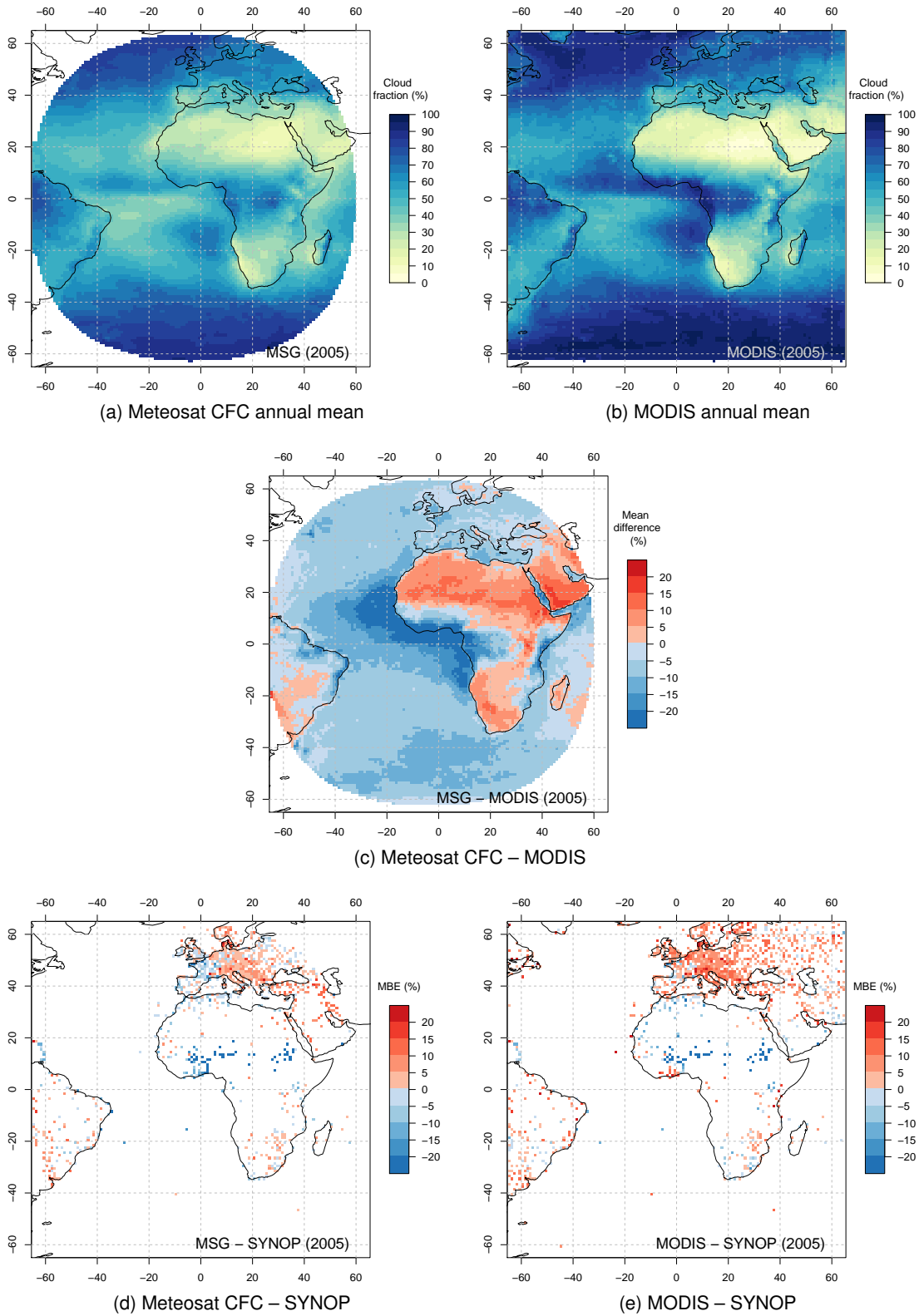


Figure 22: Inter-comparison of Meteosat CFC, MODIS and gridded SYNOP monthly means of 2005 at  $1^\circ \times 1^\circ$  spatial resolution.

## 8.2 Inter-comparison with PATMOS-x

Figure 23 presents the Meteosat CFC and PATMOS-x CFC as compared to synoptic observations aggregated over 290 sites. Meteosat CFC agrees with PATMOS-x after 2006, however both slightly underestimate SYNOP. Before 2000 PATMOS-x has overall higher CFC than Meteosat, but the differences between both data records are more complex. The main feature is an increase of PATMOS-x bias (Fig. 23b) and bcRMSE (Fig. 23c) with a lifetime of NOAA-11, NOAA-14 and NOAA-16. Since Meteosat CFC and PATMOS-x collocations were performed at level-2, the variation of performance statistics is not related to orbital drift. It can be however related to sensor degradation and/or inter-calibration issues, and it is recommended for a further investigation.

Grid-based comparison carried out at whole Meteosat disc for 2005 reveals PATMOS-x CFC 1% lower than Meteosat CFC. The difference between both data records is mostly visible for the ocean within 20°S–20°N (Figure 24c). For grids where SYNOP CFC is available, Meteosat CFC overestimates synoptic observations by approximately 1% while PATMOS-x is just on target (Table 11).

Table 11: Inter-comparison statistics of gridded Meteosat CFC, PATMOS-x and SYNOP for 2005.

	MBE	bcRMSE
Meteosat CFC – SYNOP	0.92	6.16
PATMOS-x – SYNOP	-0.08	8.59

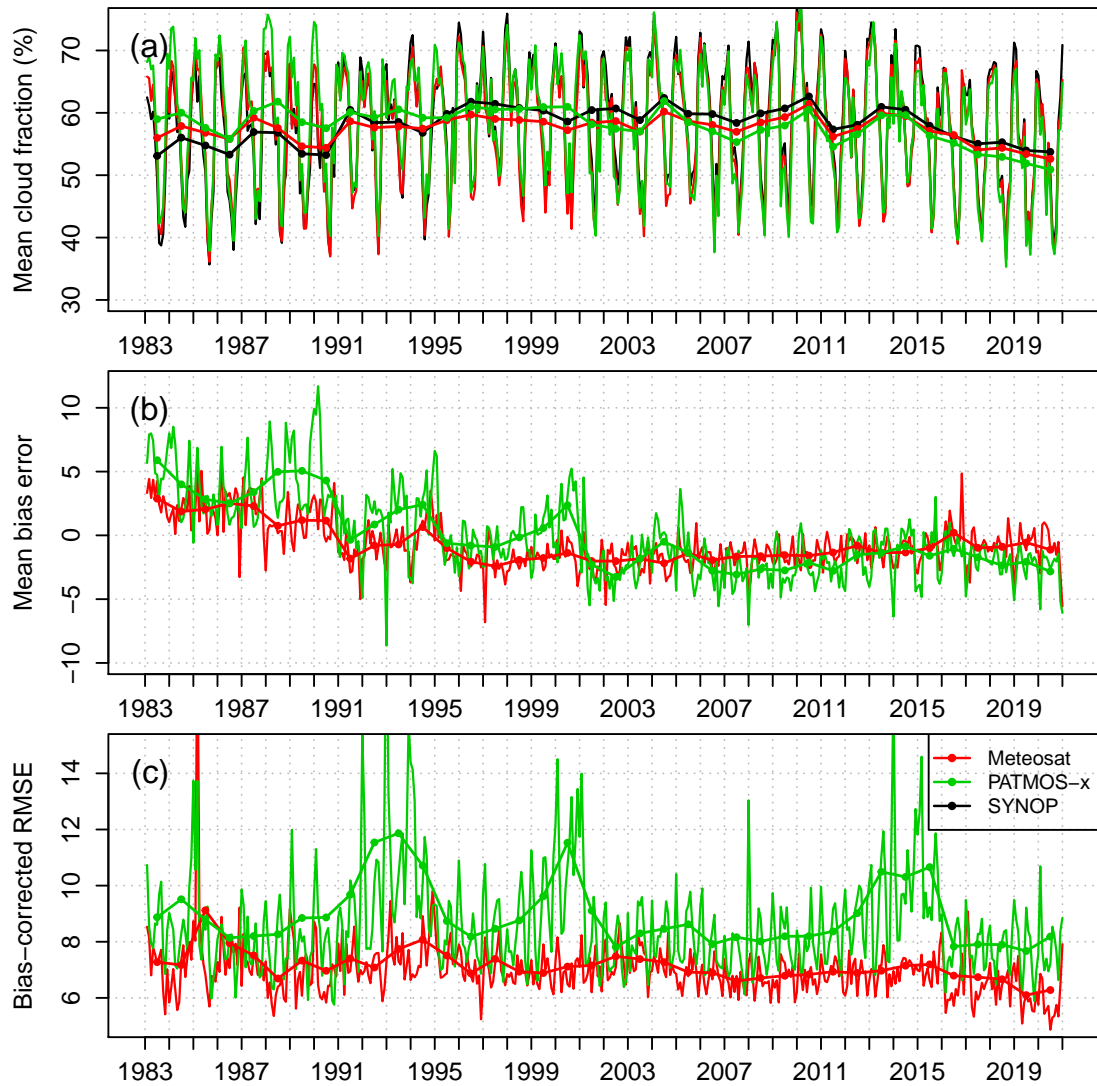


Figure 23: Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and PATMOS-x as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics.

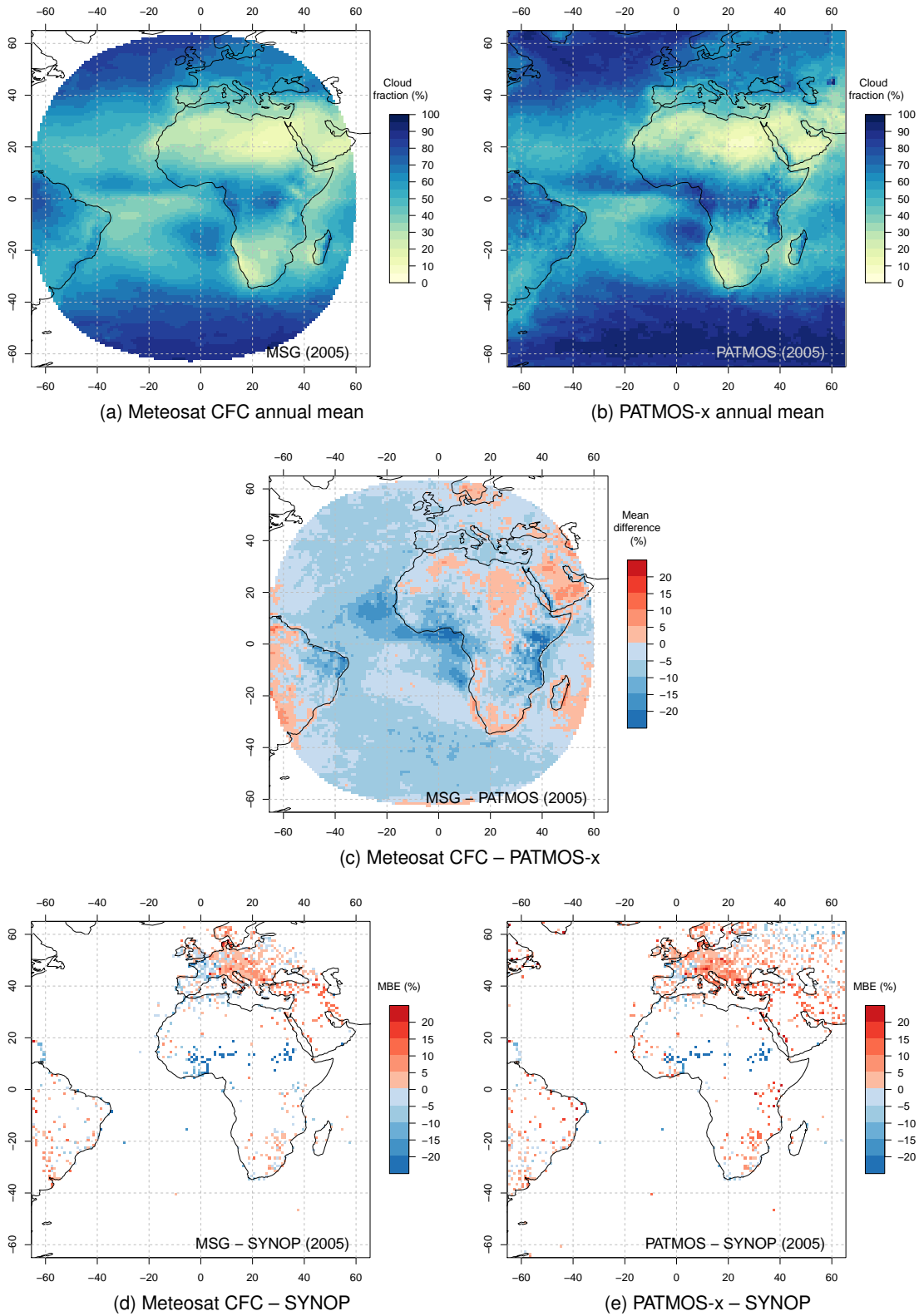


Figure 24: Inter-comparison of Meteosat CFC, PATMOS-x and gridded SYNOP monthly means of 2005 at 1° × 1° spatial resolution.

### 8.3 Inter-comparison with CLARA-A2

Figure 25 presents the Meteosat CFC and CLARA-A2 CFC as compared to synoptic observations aggregated over 290 sites. Meteosat CFC and CLARA-A2 agree with SYNOP during summer months when CFC has its yearly minimum. Yet, the differences are evident for winter months (CFC seasonal maximum) when CLARA-A2 tends to overestimate SYNOP CFC. This overestimation is larger prior to 2003 when no METOP satellite was present (Fig. 25b). Overall precision of both data records is similar, but bCRMSE of Meteosat CFC reaches higher values in winter months.

These findings are confirmed by a comparison carried out at a grid level for 2005. For grids where SYNOP CFC is available, Meteosat CFC overestimates synoptic observations by approximately 1% while CLARA-A2 overestimates by 3%, and bCRMSE are 6.2% and 6.5%, respectively (Table 12). When aggregated over the whole Meteosat disc, Meteosat CFC is 6% lower than CLARA-A2 CFC. CLARA-A2 CFC is greater for most of the disc with maxima close to the African coast around 10°S-20°S (a region of tropical marine stratocumulus), as well as for a few spots in Sahara. Concurrently, CLARA-A2 CFC is lower than Meteosat CFC over South America (Figure 26c). The performance of both data records as compared to gridded SYNOP CFC does not reveal which data record is more accurate for these areas.

Table 12: Inter-comparison statistics of gridded Meteosat CFC, CLARA-A2 and SYNOP for 2005.

	MBE	bCRMSE
Meteosat CFC – SYNOP	0.92	6.16
CLARA-A2 – SYNOP	3.08	6.47

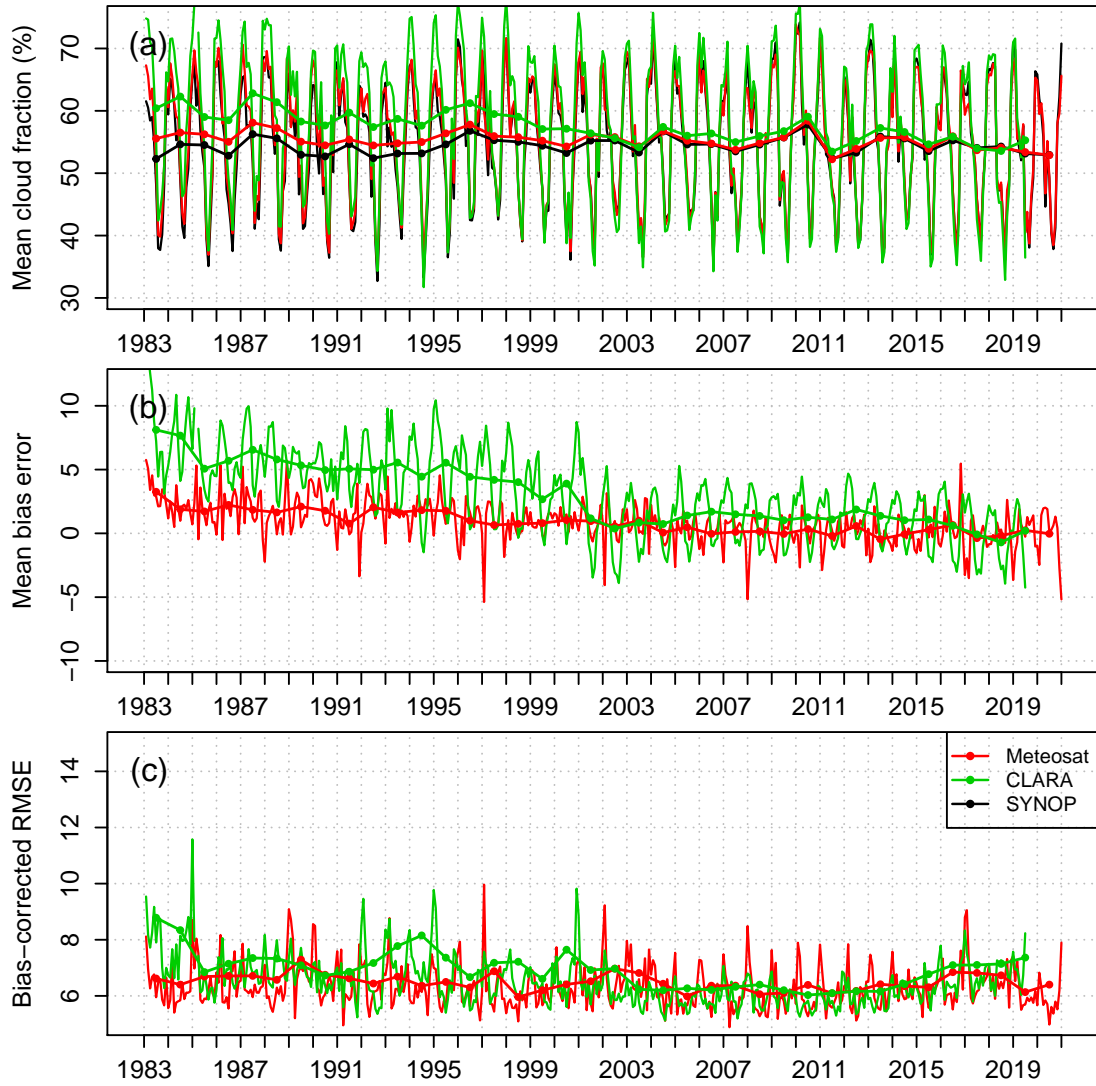


Figure 25: Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and CLARA-A2 as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics.

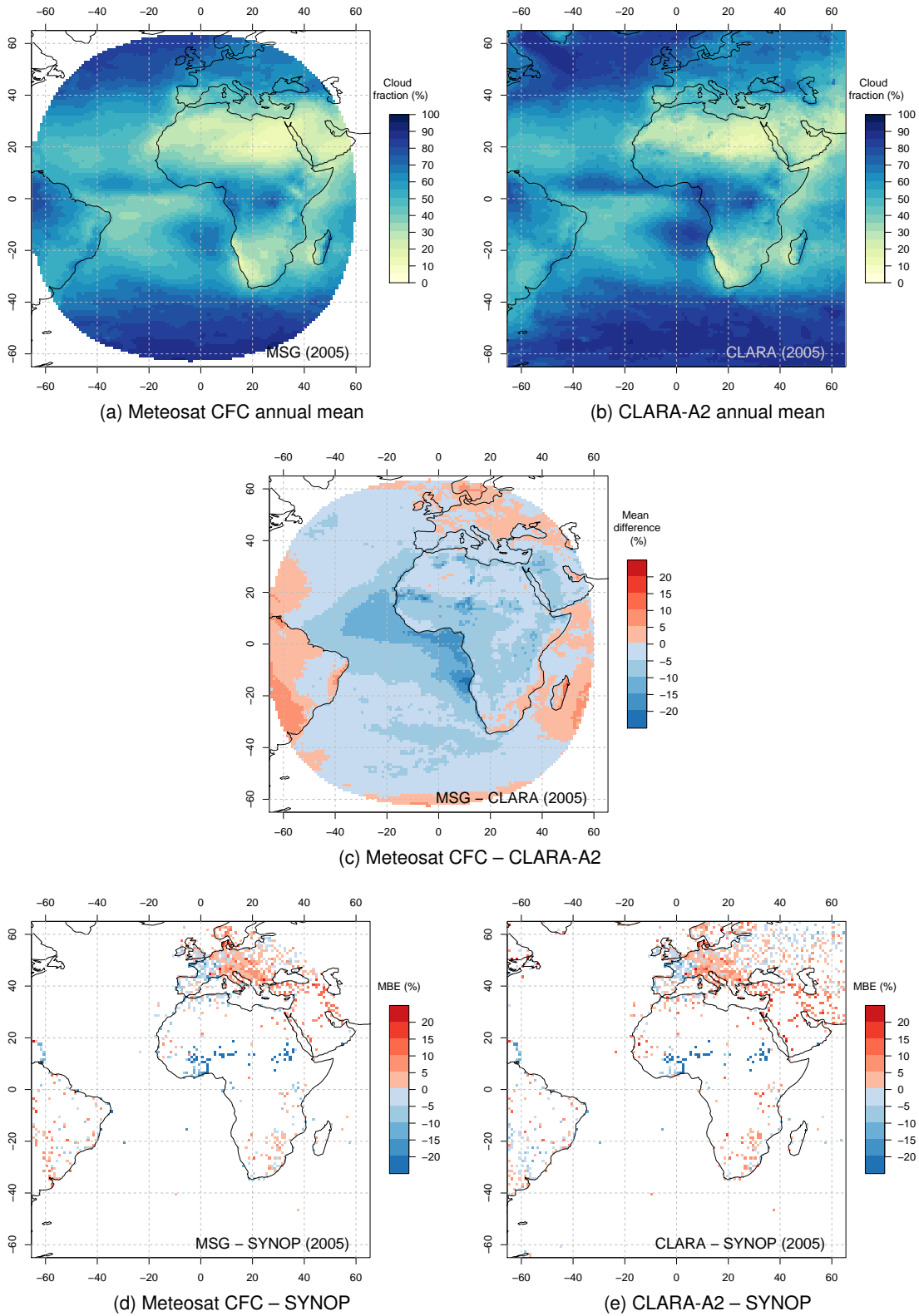


Figure 26: Inter-comparison of Meteosat CFC, CLARA-A2 and gridded SYNOP monthly means of 2005 at  $1^\circ \times 1^\circ$  spatial resolution.



## 8.4 Inter-comparison with CLAAS-A2

Figure 27 presents the Meteosat CFC and CLAAS-A2 CFC as compared to synoptic observations aggregated over 290 sites. Time series of Meteosat CFC and CLAAS-A2 agree well, but with a latter shifted toward higher CFC values. The average better correspondence of Meteosat CFC with SYNOP is most probably due to training of GeoSatClim versus SYNOP. Time series of MBE endorse that CLAAS-A2 CFC slightly overestimates SYNOP CFC, however it has to be noted that MBE of CLAAS-A2 has a lower spread, i.e. Meteosat CFC reveals greater (negative) bias for winter months, which is not observed in CLAAS-A2 (Fig. 27b). Moreover, Meteosat CFC is more precise than CLAAS-A2 in summer, but it is just the opposite for winter months (Fig. 27c). Table 11 and Figure 28 summarize results for a grid-based comparison for 2005. For grids where SYNOP CFC is available, the main difference in performance of Meteosat CFC and CLAAS-A2 is in Europe, where CLAAS-A2 reveals too high CFC values (Fig. 28d–e). Overall Meteosat CFC overestimates by 1% and CLAAS-A2 overestimates by 2.7% SYNOP CFC. A direct comparison shown in Figure 28c reveals greater CLAAS-A2 CFC values over ocean as compared to Meteosat CFC.

Table 13: Inter-comparison statistics of gridded Meteosat CFC, CLAAS-A2 and SYNOP for 2005.

	MBE	bcRMSE
Meteosat CFC – SYNOP	0.92	6.16
CLAAS-A2 – SYNOP	2.73	6.32

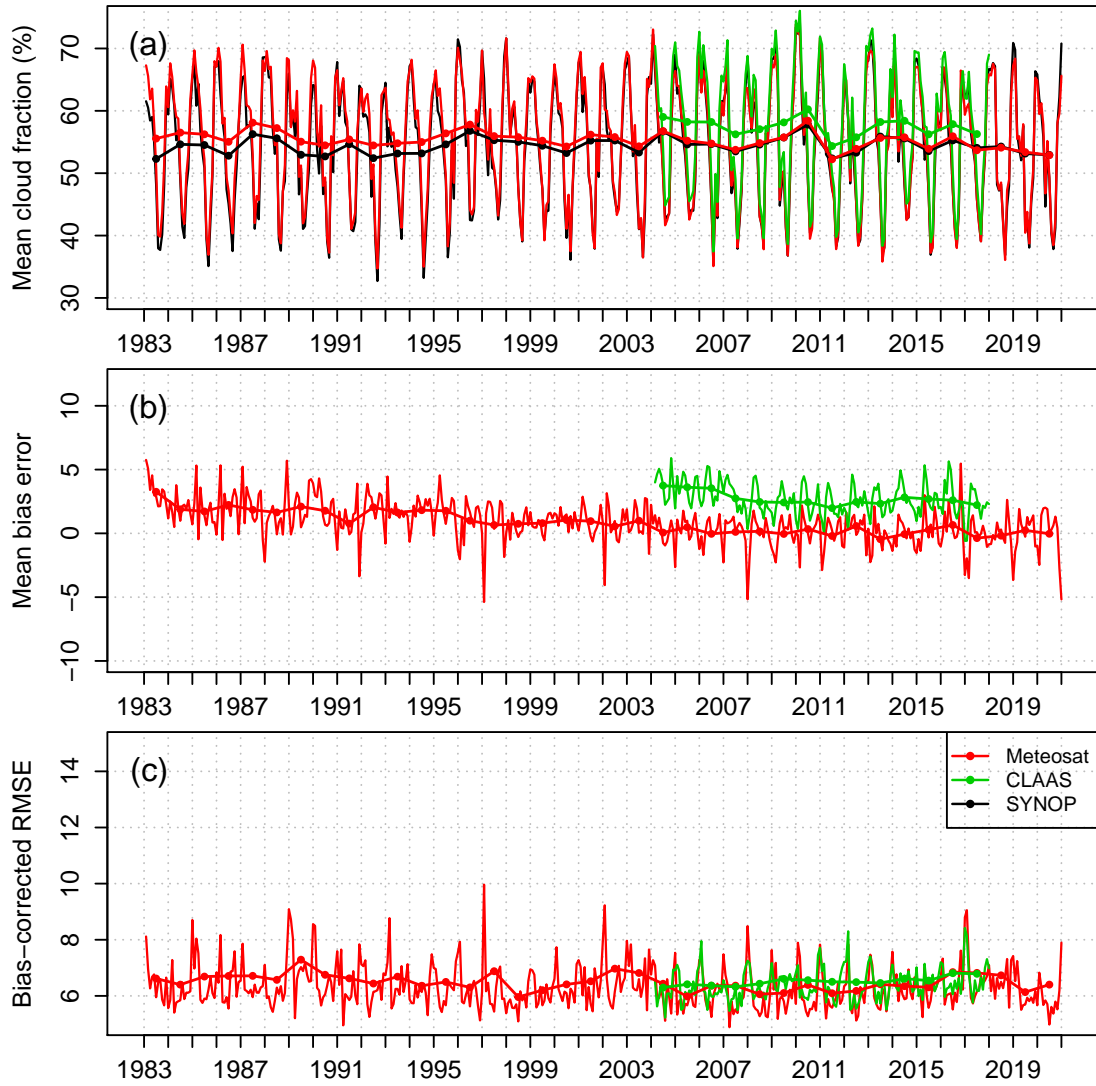


Figure 27: Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and CLAAS-A2 as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics.

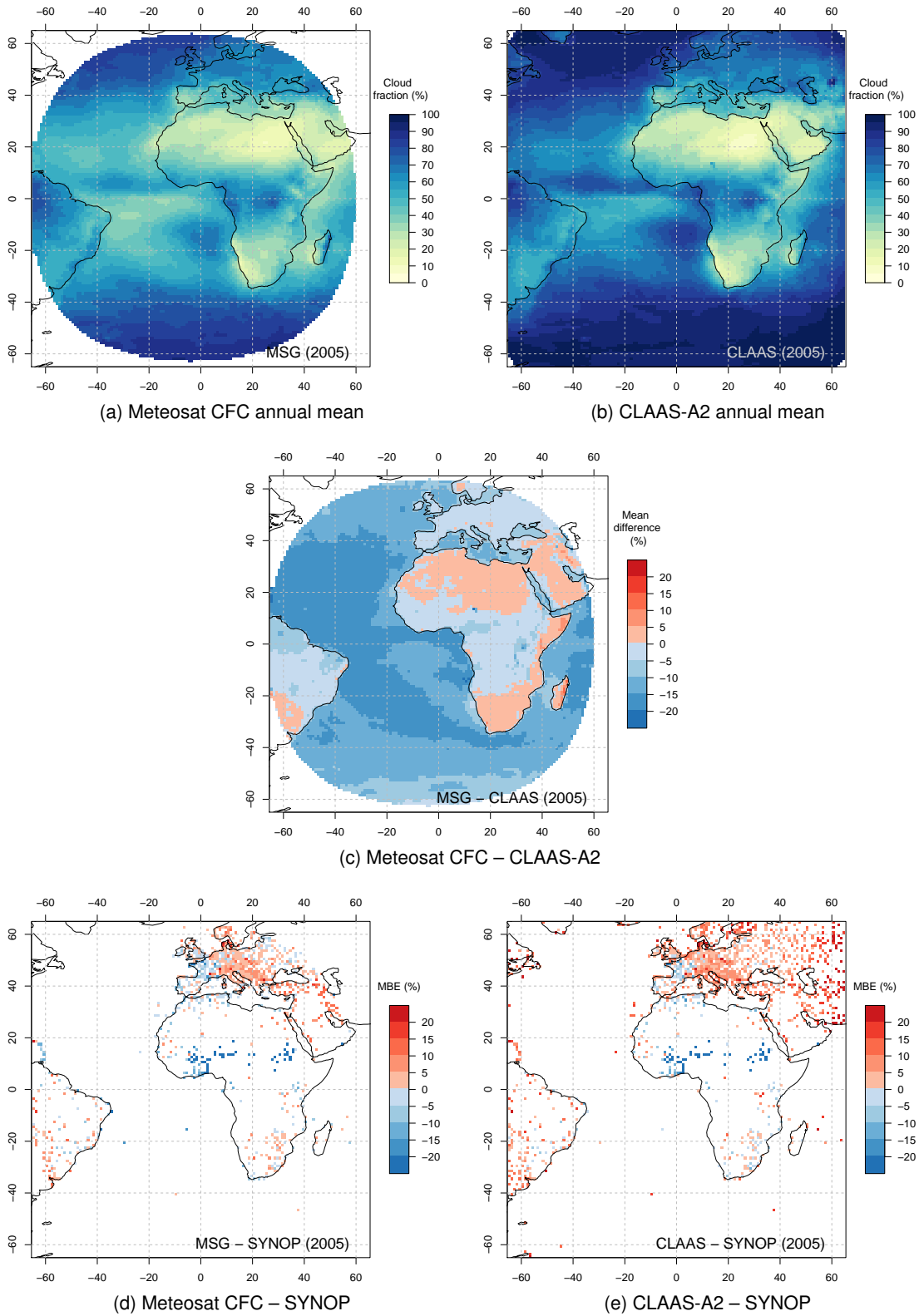


Figure 28: Inter-comparison of Meteosat CFC, CLAAS-A2 and gridded SYNOP monthly means of 2005 at 1° x 1° spatial resolution.

## 8.5 Inter-comparison with CC4CL-AVHRR

Figure 29 presents the Meteosat CFC and CC4CL-AVHRR CFC as compared to synoptic observations aggregated over 290 sites. Meteosat CFC and CC4CL-AVHRR agree with SYNOP during summer months when CFC has its yearly minimum. Yet, the differences are evident for winter months (CFC seasonal maximum) when CC4CL-AVHRR tends to overestimate SYNOP CFC. This overestimation is slightly lower after 2002, i.e. for NOAA-16, NOAA-18 and NOAA-19 (Fig. 29b). Overall precision of both data records is similar, but bCRMSE of Meteosat CFC reaches higher values in winter months.

These findings are confirmed by a comparison carried out at a grid level for 2005. For grids where SYNOP CFC is available, Meteosat CFC overestimates synoptic observations by approximately 1% while CC4CL-AVHRR overestimates by 5.1%, and bCRMSE are 6.2% and 6.4%, respectively (Table 14). When aggregated over the whole Meteosat disc, Meteosat CFC is almost 5% lower than CC4CL-AVHRR CFC. CC4CL-AVHRR CFC is greater for most of the disc with maxima close to the African coast around 10°S–20°S (a region of tropical marine stratocumulus). Concurrently, CC4CL-AVHRR CFC is lower than Meteosat CFC over South America (Figure 30c). The performance of both data records as compared to gridded SYNOP CFC does not reveal which data record is more accurate for these areas.

Table 14: Inter-comparison statistics of gridded Meteosat CFC, CC4CL-AVHRR and SYNOP for 2005.

	MBE	bCRMSE
Meteosat CFC – SYNOP	0.92	6.16
CC4CL-AVHRR – SYNOP	5.13	6.35

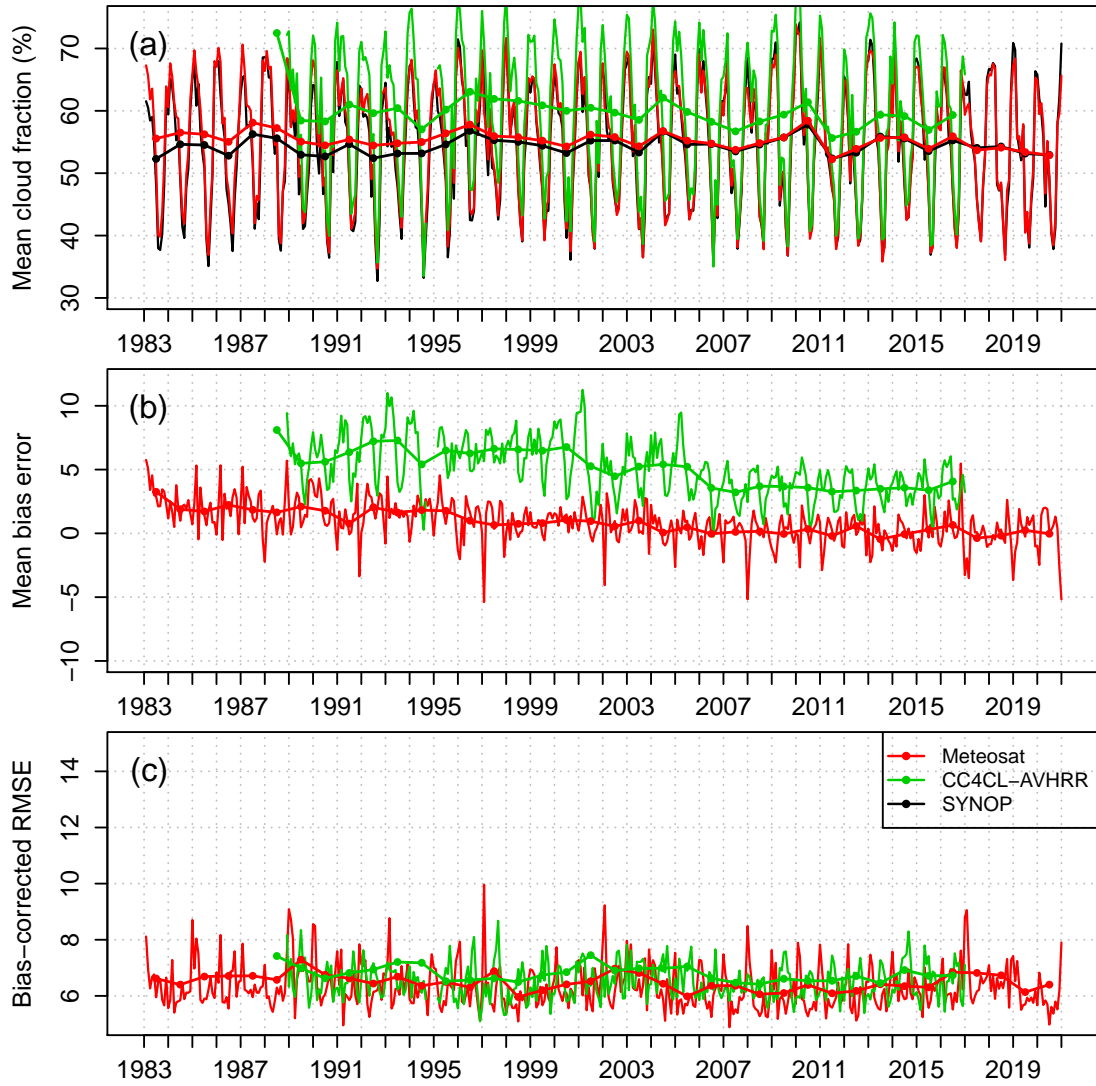


Figure 29: Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and CC4CL-AVHRR as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics.

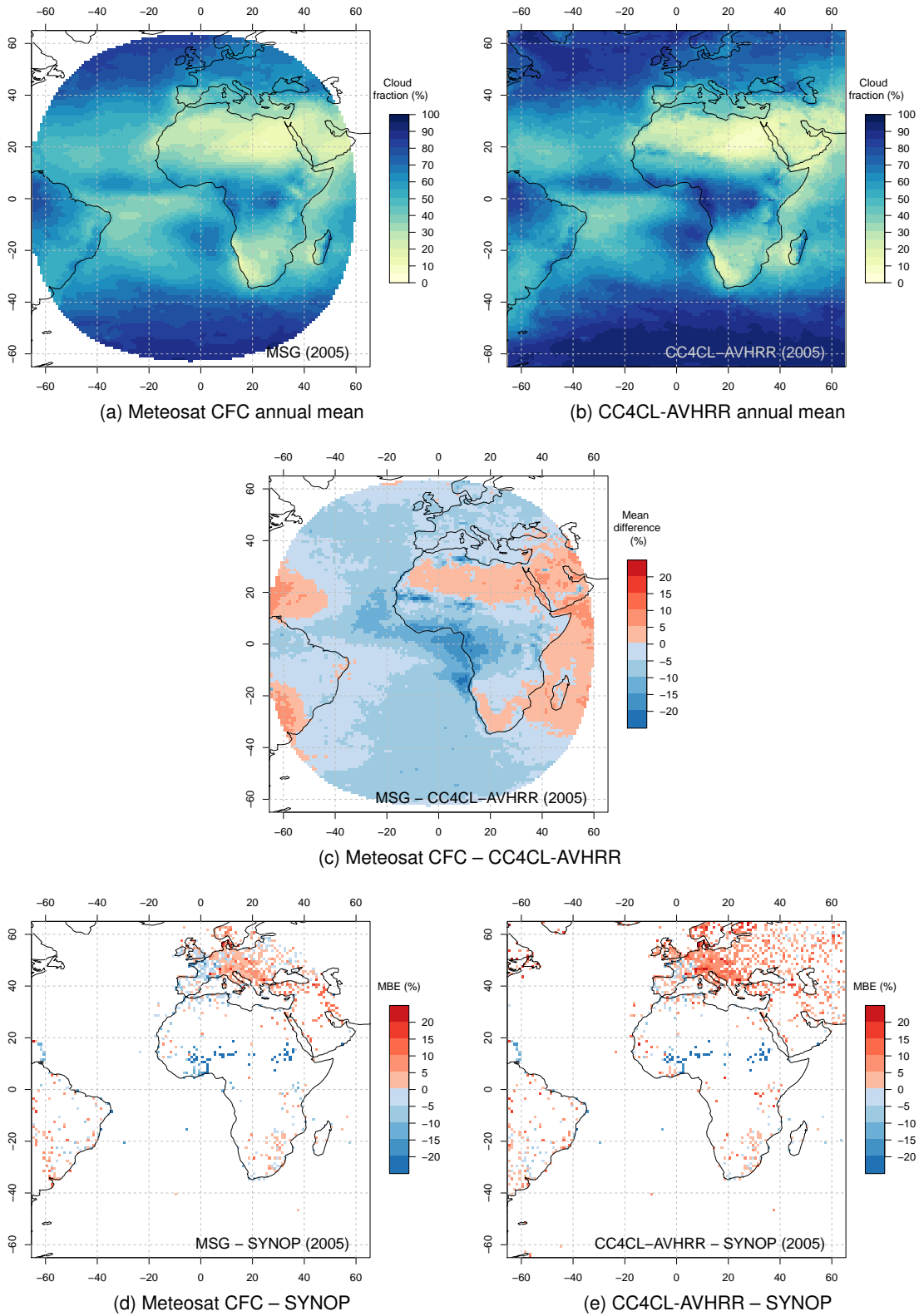


Figure 30: Inter-comparison of Meteosat CFC, CC4CL-AVHRR and gridded SYNOP monthly means of 2005 at 1° x 1° spatial resolution.

## 8.6 Inter-comparison with ISCCP-H

Figure 31 presents the Meteosat CFC and ISCCP-H as compared to synoptic observations aggregated over 290 sites. ISCCP-H shows a dampened seasonal cycle, reaching its maximum during winter months comparable to Meteosat CFC, but showing too high CFC values during summer.

These findings are confirmed by a comparison carried out at a grid level for 2005. For grids where SYNOP CFC is available, Meteosat CFC overestimates synoptic observations by approximately 1% while ISCCP-H overestimates by 6.1%, and bcRMSE are 6.2% and 9.4%, respectively (Table 15). When aggregated over the whole Meteosat disc, Meteosat CFC is almost 5–10% lower than ISCCP-H CFC. ISCCP-H CFC is greater for most of the disc. The spatial distribution of differences to SYNOP do not reveal systematic biases.

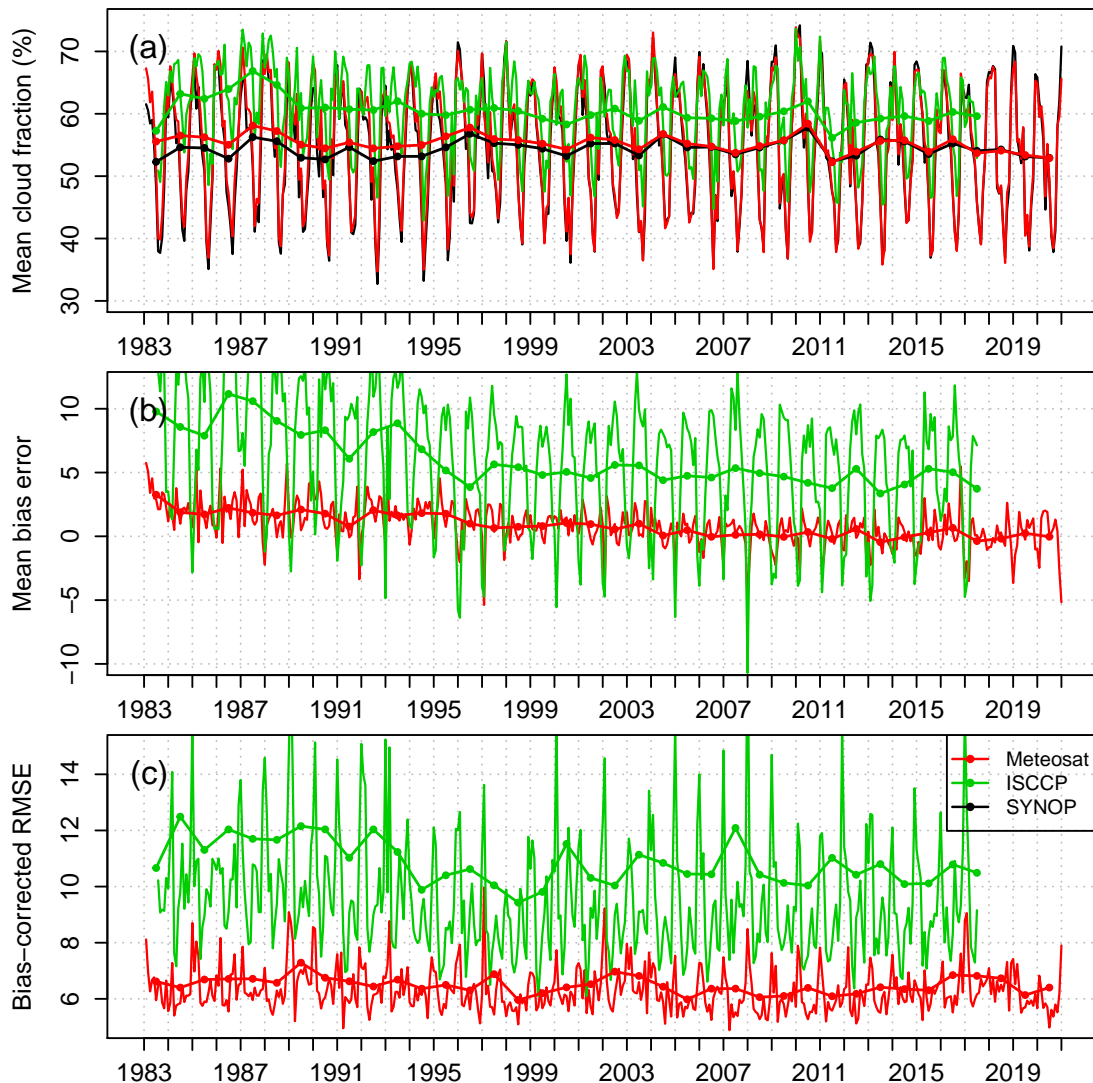


Figure 31: Time series of monthly means (a), mean bias error (b) and bias-corrected root mean square error (c) of Meteosat CFC and ISCCP-H as compared to synoptic observations at 290 sites. Thick lines with filled circles give yearly mean statistics.

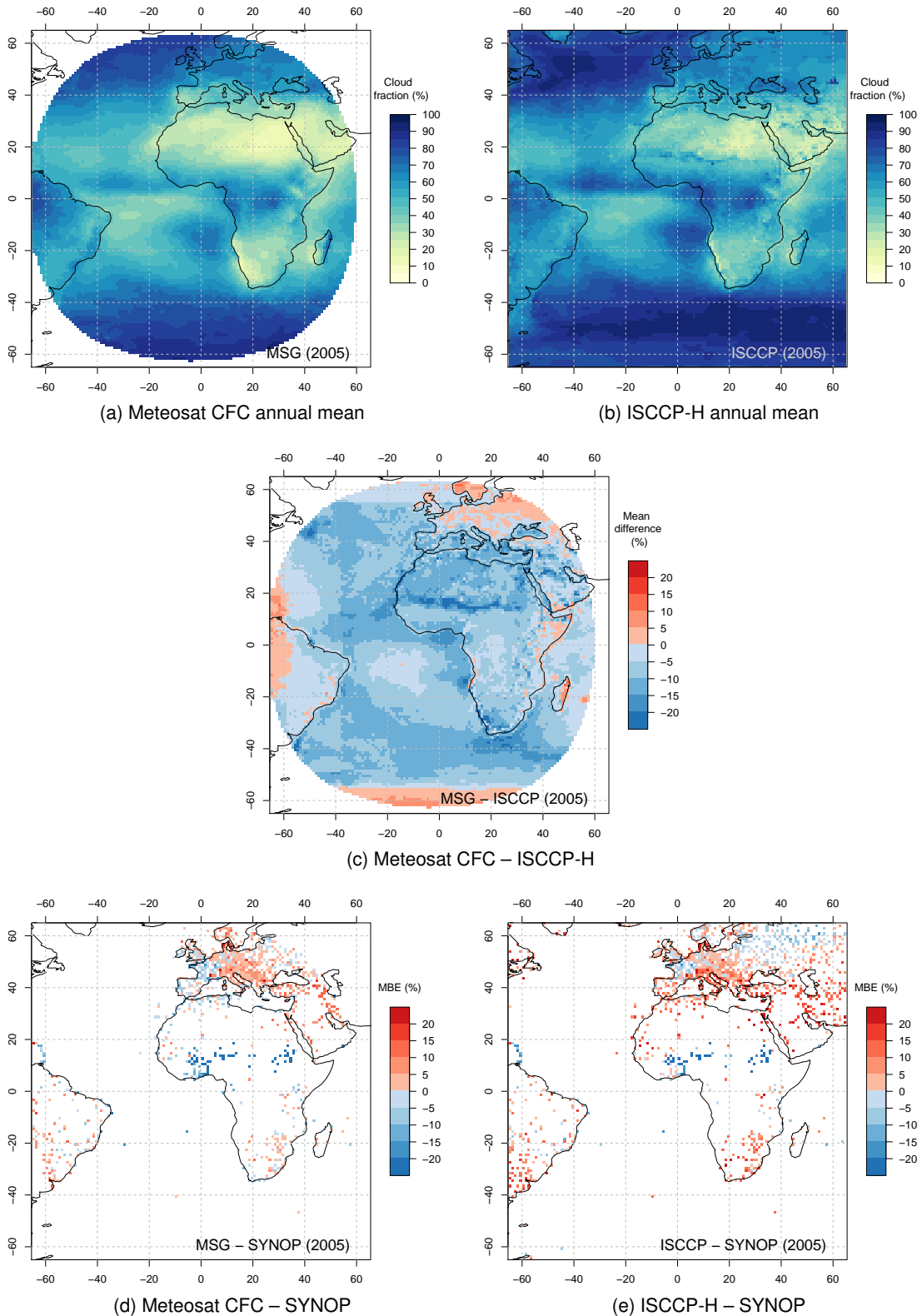


Figure 32: Inter-comparison of Meteosat CFC, ISCCP-H and gridded SYNOP monthly means of 2005 at  $1^\circ \times 1^\circ$  spatial resolution.



Table 15: Inter-comparison statistics of gridded Meteosat CFC, ISCCP-H and SYNOP for 2005.

	MBE	bcRMSE
Meteosat CFC – SYNOP	0.92	6.16
ISCCP-H – SYNOP	6.08	9.44

## 9 Concluding remarks


This report documents the validation of Edition 2 of the Meteosat CFC data record. For the study we used reference data records from independent observation sources (SYNOP and CALIPSO/CALIOP) with different measurement strategies. We also inter-compared Meteosat CFC to concurrent state-of-the-art satellite-based data records derived from geostationary and polar orbiting passive visible and infrared imaging sensors (MODIS, PATMOS-x, CLARA-A2, CLAAS-A2, ISSCP-H and CC4CL-AVHRR). This broad spectrum of reference and inter-compared measurements supported the required independence and variety in the evaluation process.

The validation was based on data records at different processing levels: Level 2 (instantaneous data) and Level 3 (daily and monthly means). Despite missing requirements on the instantaneous values (Level 2) they were included in the evaluation process. This brings further insights into the precision of the derived cloud data records on high temporal resolution and simplifies the detective work needed to improve Meteosat CFC for a next release.

The following issues will need further attention:

- Comparison to SYNOP reveals a positive bias in the early 1980's. Since several other inter-comparisons (to CLARA-A2, PATMOS-x and ISSCP-H) using different sensors or algorithms show a similar positive bias, the problem might be related in the SYNOP observations or a combination of SYNOP observation changes with changes in e.g. atmospheric forcing like aerosol or water vapor.
- The decrease of SYNOP coverage and/or usable data in France and Switzerland strongly hampering the use of satellite data activities. The potential automatization of SYNOP cloud observations should lead to an increase of SYNOP cloud observations and it should even enhance the quality of SYNOP cloud observations since human observer biases will decrease.
- We observe a VZA dependence of Meteosat CFC using inter-comparisons. Meteosat CFC is underestimated at low VZA and overestimated at high VZA. A VZA dependence might be a result of a lower dynamic range between cloudy and clear sky at high VZA, but it could also be a result of the view geometry effect: at high VZA we look through cloud decks from the side, potentially increasing the CFC amount compared to a nadir CFC estimation. A VZA correction of the CFC estimation might be needed.
- The negative CFC bias during winter has been reduced in Edition 2, but time series analysis suggests that this effect is still visible. The bayesian classifier might need to be adjusted for the absolute clear sky brightness temperature and reflectance.

With this report we tried to describe the positive features as well as possible sources of error as accurately as possible. We emphasize that Meteosat CFC is the longest high temporal (sub-daily) and high spatial ( $0.05^\circ \times 0.05^\circ$ ) resolution CFC climatology with 38 years length. Users are encouraged to report back findings on applicability, stability and potential deficiencies to CM SAF (email: [cmsaf.contact@dwd.de](mailto:cmsaf.contact@dwd.de)).

	<b>Validation Report</b> <b>Meteosat Cloud Fractional Cover</b> <b>Edition 2</b>	Doc: SAF/CM/MeteoSwiss/VAL/MET/GFC/2.0 Issue: 2.0 Date: April 11, 2023
--	--	--

## References

- Aguilar, E., Auer, I., Brunet, M., Peterson, T., and Wieringa, J. (2003). Guidance on metadata and homogenization. Technical report, WMO TD N. 1186 (WCDMP N. 53).
- Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *J. Climatol.*, 6(6):661–675.
- Bojanowski, J., Stöckli, R., Tetzlaff, A., and Kunz, H. (2014). The impact of time difference between satellite overpass and ground observation on cloud cover performance statistics. *Remote Sens.*, 6:12866–12884.
- Chepfer, H., Bony, S., Winker, D. M., Cesana, G., Dufresne, J., Minnis, P., Stubenrauch, C., and Zeng, S. (2010). The GCM Oriented CALIPSO Cloud Product (CALIPSO- GOCCP). *J. Geophys. Res.*, 105:D00H16.
- CM SAF (2016a). Validation Report of CM SAF Cloud, Albedo, Radiation data record, AVHRR-based, Edition 2 (CLARA-A2). Validation Report SAF/CM/SMHI/VAL/GAC/CLD/2, Issue 2.0, EUMETSAT Satellite Application Facility on Climate Monitoring.
- CM SAF (2016b). Validation Report of CM SAF SEVIRI cloud products Edition 2 (CLAAS-2). Validation Report SAF/CM/KNMI/VAL/SEV/CLD/2, Issue 2.0, EUMETSAT Satellite Application Facility on Climate Monitoring.
- Derrien, M. and Le Gléau, H. (2005). MSG/SEVIRI cloud mask and type from SAFNWC,. *International Journal of Remote Sensing*, 26:4707–4732.
- EUMETSAT (2010). Meteosat VIS channel calibration information. PDF-TEN-VIS-CHANNEL-CALIB, EUMETSAT, EUMETSAT, Am Kavalleriesand 31, D-64295 Darmstadt, Germany.
- EUMETSAT (2017). MSG Level 1.5 Image Data Format Description. Technical Report EUM/MSG/ICD/105, EUMETSAT Am Eumetsat-Allee 1, D-64295 Darmstadt, Germany.
- GCOS (2006). Systematic observation requirements for satellite-based products for climate. Technical Report GCOS-107, WMO-TD No. 1338, Global Climate Observing System, CH-1212 Geneva 2, Switzerland.
- Hamann, U., Walther, A., Baum, B., Bennartz, R., Bugliaro, L., Derrien, M., Francis, P. N., Heidinger, A., Joro, S., Kniffka, A., Le Gléau, H., Lockhoff, M., Lutz, H.-J., Meirink, J. F., Minnis, P., Palikonda, R., Roebeling, R., Thoss, A., Platnick, S., Watts, P., and Wind, G. (2014). Remote sensing of cloud top pressure/height from SEVIRI: analysis of ten current retrieval algorithms. *Atmospheric Measurement Techniques*, 7:2839–2867.
- Heidinger, A., Goldberg, M., Jelenak, A., and Pavolonis, M. (2005). A new AVHRR cloud climatology. *Proc. SPIE 5658*, 197.
- Heidinger, A. K., Evan, A. T., Foster, M. J., and Walther, A. (2012). A Naive Bayesian Cloud-Detection Scheme Derived from CALIPSO and Applied within PATMOS-x. *Journal of Applied Meteorology and Climatology*, 51:1129–1144.
- John, V. O., Tabata, T., Ruethrich, F., Roebeling, R. A., Hewison, T., and Schulz, J. (2019). On the recalibration of infrared (ir) and water vapour (wv) radiance time-series from european and japanese geostationary satellites. *Submitted to Remote Sensing*.

- Karlsson, K.-G. (2003). A ten-year cloud climatology over Scandinavia derived from NOAA AVHRR imagery. *Int. J. Climatol.*, 23:1023–1044.
- Karlsson, K.-G. and 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. *Atmos. Meas. Tech. Discuss.*, 6:1093–1141.
- Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30:81–93.
- Khaliq, M. and Ouarda, T. (2007). On the critical values of the standard normal homogeneity test (SNHT). *Int. J. Climatol.*, 27:681–687.
- Mann, H. (1945). Nonparametric tests against trend. *Econometrica*, 13:245–259.
- Meirink, J., Roebeling, R., and Stammes, P. (2013). Inter-calibration of polar imager solar channels using SEVIRI. *Atm. Meas. Tech.*, 6:2495–2508.
- Pavolonis, M., Heidinger, A., and Uttal, T. (2005). Daytime Global Cloud Typing from AVHRR and VIIRS: Algorithm Description, Validation, and Comparisons. *J. Appl. Meteor.*, 44:804–826.
- Poulsen, C. A., Siddans, R., Thomas, G. E., Sayer, A. M., Grainger, R. G., Campmany, E., Dean, S. M., Arnold, C., and Watts, P. D. (2012). Cloud retrievals from satellite data using optimal estimation: evaluation and application to ATSR. *Atmos. Meas. Tech.*, 5:1889–1910.
- Roebeling, R. A., Feijt, A. J., and Stammes, P. (2006). Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *Journal of Geophysical Research-Atmospheres*, 111(D20).
- Stammes, K., Tsay, S.-C., Wiscombe, W., and Lazlo, I. (2000). Disort, a general-purpose fortran program for discrete-ordinate-method radiative transfer in scattering and emitting layered media: Documentation of methodology. Technical Report DISORT Report v1.1, Dept. of Physics and Engineering Physics Stevens Institute of Technology Hoboken, NJ 07030.
- Stengel, M., Mieruch, S., Jerg, M., Karlsson, K.-G., Scheirer, R., Maddux, B., Meirink, J., Poulsen, C., Siddans, R., Walther, A., and Hollmann, R. (2015). The Clouds Climate Change Initiative: Assessment of state-of-the-art cloud property retrieval schemes applied to AVHRR heritage measurements. *Remote Sensing of Environment*, 162:363–379.
- 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., Heidinger, A., and Pavolonis, M. (2004). Comparison of NOAA’s operational AVHRR-derived cloud amount to other satellite-derived cloud climatologies. *J. Climate*, 17:4805–4822.
- Toreti, A., Kuglitsch, F., Xoplaki, E., Della-Marta, P., Aguilar, E., Prohom, M., and 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.
- Wilks, D. (2011). *Statistical Methods in the Atmospheric Sciences*. International Geophysics. Academic Press, 704 pp.
- Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., and Rossow, W. B. (2019). On the recalibration of infrared (ir) and water vapour (wv) radiance time-series from european and japanese geostationary satellites. *Earth Syst. Sci. Data*, 10:583–593.

## A Performance and homogeneity measures

### A.1 Performance statistics

Given that  $E_k$  is the Meteosat CFC and  $M_k$  is the CFC from reference SYNOP or inter-compared satellite-derived data record, for the time step  $k$ , and  $n$  is the length of the time series, the performance statistics are defined as:

Mean bias error: 
$$\text{MBE} = \frac{1}{n} \sum_{k=1}^n (E_k - M_k)$$

Bias-corrected root mean square error: 
$$\text{bcRMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^n (E_k - M_k - \text{MBE})^2}$$

The skill scores are defined for the contingency matrix (Table 16) following Wilks (2011):

Probability of (cloud) detection: 
$$\text{POD}_{\text{cld}} = 100 \cdot \frac{a}{a+c}$$

Probability of (clear-sky) detection: 
$$\text{POD}_{\text{clr}} = 100 \cdot \frac{d}{b+d}$$

False alarm (of cloud) ratio: 
$$\text{FAR}_{\text{cld}} = 100 \cdot \frac{b}{b+d}$$

False alarm (of clear-sky) ratio: 
$$\text{FAR}_{\text{clr}} = 100 \cdot \frac{c}{c+d}$$

Hit rate: 
$$\text{HitRate} = 100 \cdot \frac{a+d}{N}$$

Hanssen-Kuipers discriminant: 
$$\text{KSS} = 100 \cdot \frac{ad-bc}{(a+c)(b+d)}$$

If not specified cloudy or clear, POD and FAR refer to probability of (cloud) detection and false alarm (of cloud) rate, respectively.

Table 16: Contingency matrix for the Meteosat-based cloud mask and reference synoptic or CALIPSO observations.

		Reference cloud mask		
		Cloudy	Cloud-free	
Meteosat cloud mask	Cloudy	$a$	$b$	$a + b$
	Cloud-free	$c$	$d$	$c + d$
Total		$a + c$	$b + d$	$N$

### A.2 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:

$$T(k) = k\bar{z}_1^2 + (n - k)\bar{z}_2^2; k \in 1, 2, 3, \dots, n. \quad (1)$$

The standardized means  $\bar{z}_1^2$  and  $\bar{z}_2^2$  are calculated as:

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

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

where  $Y_i$  stands for the value at a time step  $i$ ,  $\bar{Y}$  for the mean, and  $\sigma$  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. In this report for  $n = 300$  (25 years  $\times$  12 months) we employed the critical value of 10.02 for 95% of confidence level.