Satellite Application Facility on Climate Monitoring

Scientific Report ORR V2

Validation of CM-SAF Cloud Products using MSG/SEVIRI Data

Reference Number:	**SAF/CM/DWD/SMHI/KNMI/SR/CLOUDS/2**
Issue/Revision Index:	**1.0**
Date:	**27 June 2005**

Document Signature Table

	Name	Function	Signature	Date
Author	Karl-Göran Karlsson Erwin Wolters Peter Albert Anke Tetzlaff Rob Roebeling Werner Thomas Sheldon Johnston	Project scientists		
Approval	Annegret Gratzki	Science co-ordinator		
Approval				
Release	Martin Werscheck	Project manager		
Eumetsat Approval				

Document Change Record

Issue/Revision	Date	DCN No.	Changed Pages/Paragraphs
1.0	27.06.2005	SAF/CM/DWD/SMHI/KNMI/SR/CLOUDS/2	First Version

Distribution List

Internal Distribution	
Name	**No. Copies**
DWD archive	1

External Distribution		
Company	**Name**	**No. Copies**
EUMETSAT		1

Table of Contents

1 INTRODUCTION 1

2 CLOUD OBSERVATION AT THE CABAUW, CHILBOLTON AND PALAISEU OBSERVATION SITES 2

2.1 INDIVIDUAL CLOUD OBSERVATION INSTRUMENTS 2
2.1.1 Lidar 2
2.1.2 Cloud radar 2
2.1.3 Microwave radiometer 3
2.1.4 Pyranometer 3
2.2 MEASUREMENT CAMPAIGNS 3
2.2.1 CloudNET 3
2.2.2 CESAR 3

3 VALIDATION OF THE FRACTIONAL CLOUD COVER PRODUCT (CFC) 5

3.1 VALIDATION METHOD 5
3.2 VALIDATION RESULTS FOR DAILY MEAN VALUES OVER LAND – INDIVIDUAL SYNOP STATIONS 7
3.3 VALIDATION RESULTS FOR MONTHLY MEAN VALUES OVER LAND – INDIVIDUAL SYNOP STATIONS 8
3.4 VALIDATION RESULTS FOR MONTHLY MEAN VALUES FOR LAND, SHIP AND BUOY OBSERVATIONS – GRIDDED DATA 10
3.5 VALIDATION RESULTS FOR INSTANTANEOUS OBSERVATIONS 10
3.6 SUMMARY OF CFC VALIDATION RESULTS 12

4 VALIDATION OF THE CLOUD TYPE PRODUCT (CTY) 15

4.1 DETAILS OF THE VALIDATION METHOD 15
4.2 RESULTS 17
4.3 SUMMARY OF CTY VALIDATION RESULTS 21

5 VALIDATION OF THE CLOUD TOP PRODUCT (CTH/CTP/CTT) 22

5.1 SUMMARY OF SAFNWC/METEO-FRANCE CTH VALIDATION STUDY 22
5.2 VALIDATION METHOD USED FOR THE ORR V2 VALIDATION 24
5.3 RESULTS 25
5.4 SUMMARY OF CTH/CTP/CTT VALIDATION RESULTS 26

6 VALIDATION OF THE THERMODYNAMIC CLOUD PHASE PRODUCT (CPH) 29

6.1 SCIENTIFIC APPROACH AND SPECIFICATION 29
6.2 VALIDATION METHOD 29
6.3 RESULTS 30
6.4 FURTHER QUALITY STUDIES 35
6.5 SUMMARY OF THE CPH VALIDATION RESULTS 36

7 VALIDATION OF THE CLOUD OPTICAL THICKNESS (COT) PRODUCT 37

7.1 VALIDATION METHOD 37
7.2 RESULTS 38
7.3 DISCUSSION 43
7.4 SUMMARY OF COT VALIDATION RESULTS 45

8 VALIDATION OF THE CLOUD WATER PATH PRODUCT (CWP) 46

8.1 VALIDATION METHOD 46
8.2 RESULTS 47
8.3 DISCUSSION 52

8.4 SUMMARY OF CLWP VALIDATION RESULTS 52

9 SUMMARY AND CONCLUSIONS 54

10 REFERENCES 61

List of Tables

Table 1.1 *The CM-SAF cloud products based on MSG/SEVIRI data.* 1

Table 3.1 *Bias for daily and monthly mean cloud fractions above land derived from collocated MSG and Synop observations. Fractional clouds were weighted with a factor of 0.75.* 13

Table 3.2 *Bias corrected Root Mean Square deviation for daily and monthly mean cloud fractions above land derived from collocated MSG and Synop observations. Fractional clouds were weighted with a factor of 0.75.* 13

Table 3.3 *Uncorrected Root mean square deviation for daily and monthly mean cloud fractions above land derived from collocated MSG and Synop observations. Fractional clouds were weighted with a factor of 0.75.* 14

Table 4.1 *Percentage of matches between cloud radar and MSG derived CTY's for each investigated month.* 18

Table 4.2 *Linear correlation coefficient (R^2), mean error (bias) and RMS error for daily MSG CTY cloud frequencies compared with daily cloud radar derived frequencies.* 20

Table 4.3 *Comparison of MSG- and cloud radar retrieved monthly cloud type frequencies. The notation < 12 N per day means that all data is used while >12 N per day means that only a subset of data is used requiring that more than 12 radar measurements per day are available.* 20

Table 5.1 *Statistical results from Meteo-France's CTH validation study (Daloze and Haeffelin, 2005).* 23

Table 5.2 *Summary of error statistics for monthly and daily results from Cabauw and Chilbolton. First two columns show monthly means and remaining columns error statistics based on daily averages and some basic observation statistics. The quality column represents the percentage of MSG pixels that were flagged as having good quality.* 26

Table 6.1 *Contingency table for CPH product.* 30

Table 6.2 *Global skill scores, water and ice difference scores for May, July, October and December 2004. Numbers of collocated measurements are given between parentheses.* 31

Table 6.3 *Global skill scores, water and ice difference scores for May, July, October and December 2004 for single layer (upper table) and multi layer cases (lower table). The number of collocated measurements is given between parentheses.* 32

Table 6.4 *Global skill scores for CPH obtained by ISCCP type algorithm and corresponding water and ice difference scores for May, July, October and December 2004. The numbers of collocated measurements are given in parentheses.* 34

Table 6.5 *Median and standard deviation of differences in relative water occurrence for May, July, October and December 2004. Numbers of observations are given between parentheses.* 35

Table 6.6 *Mean monthly occurrences for water and ice retrievals (in %) for MSG/SEVIRI and for lidar/radar (in parentheses).* 36

Table 7.1 *Slope, offset and linear correlation coefficients of hourly and daily mean MSG/SEVIRI and pyranometer derived Cloud Optical Thickness for May, July, October and December 2004.* 41

Table 7.2 *Monthly mean MSG/SEVIRI (MSG) and pyranometer (pyr.) derived cloud optical thicknes and median and standard deviation of the difference between MSG and pyr. retrieved daily mean COT values.* 42

Table 8.1 *Slope, offset and linear correlation coefficients of hourly and daily mean MSG/SEVIRI and microwave radiometer derived Cloud Liquid Water Path for May, July, October and December 2004.* 51

Table 8.2 *Monthly mean MSG/SEVIRI (MSG) and microwave radiometer (MW) cloud liquid water path and median and standard deviation of the difference between MSG and MW retrieved daily mean CLWP values.* 51

Table 9.1 *ORR V2 validation results for the CM-SAF cloud products in* ***May 2004****. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.* 55

Table 9.2 *ORR V2 validation results for the CM-SAF cloud products in* ***July 2004****. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.* 56

Table 9.3 *ORR V2 validation results for the CM-SAF cloud products in* ***October 2004****. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.* 57

Table 9.4 *ORR V2 validation results for the CM-SAF cloud products in* ***December 2004****. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.* 58

List of Figures

Figure 3.1 *Geographical location of Synop stations used for SEVIRI cloud mask validation. The numbers below the image show the total number of available observations for each Synop type.* 5

Figure 3.2 *Scatter plot of MSG and Synop daily mean cloud fractions (left) and frequency of occurrence of differences (right). Results are shown for July 2004; all available pixels were used; fractional clouds were weighted with a factor of 0.5.* 7

Figure 3.3 *Bias (left) and root mean square deviations (right) for daily mean cloud fractions derived from collocated MSG and Synop observations. Values are shown as a function of the fractional clouds factor; individual plots are shown for each month and for all available pixels / only the high confidence pixels, respectively.* 8

Figure 3.4 *Monthly mean cloud fraction for July 2004; all available pixels were used; fractional clouds were weighted with a factor of 0.5: Scatter plot of MSG and Synop monthly mean cloud fractions (upper left) and frequency of occurrence of differences (upper right). Spatial distribution of monthly mean cloud fraction from Synop observations (middle left) and spatial distribution of relative differences (middle right). Total number of collocated observations (lower left) and total number of daily averages used for calculating the monthly mean values (lower right).* 9

Figure 3.5 *Bias (left) and root mean square deviations (right) for monthly mean cloud fractions derived from MSG and Synop daily averages. Values are shown as a function of the fractional clouds factor; individual plots are shown for each month and for all available pixels / only the high confidence pixels, respectively.* 10

Figure 3.6 *Bias and root mean square deviation of instantaneous MSG and Synop cloud fractions for July 2004 and for* ***Midlatitude regions****.* 11

Figure 3.7 *Bias and root mean square deviation of instantaneous MSG and Synop cloud fractions for July 2004 and for* ***Nordic regions.*** 11

Figure 3.8 *Bias and root mean square deviation of instantaneous MSG and synop cloud fractions for December 2004 and for* ***Midlatitude regions****.* 12

Figure 3.9 *Bias and root mean square deviation of instantaneous MSG and Synop cloud fractions for December 2004 and for* ***Nordic regions****.* 12

Figure 3.10: *Bias for monthly mean cloud fractions derived from instantaneous MSG and Synop data. Upper left: Ship observations, Upper right: Buoy data, Lower: Land observations. Values are shown as a function of the fractional clouds factor; individual plots are shown for each month and for all available pixels / only the high confidence pixels, respectively.* 14

Figure 4.1 *Daily cloud type frequencies (%) retrieved from cloud radar data plotted against daily cloud frequencies (%) retrieved from MSG CTY estimations.* ***Upper left)*** *May 2004,* ***Upper right)*** *July 2004,* ***Lower left)*** *October 2004.* ***Lower right)*** *December 2004.* 19

Figure 5.1 *Scatter plot depicting MSG/SEVIRI-derived and radar-observed individual (instantaneous) CTH values for each studied month in 2004 (****Upper left: May, Upper right: July, Lower left: October, Lower right: December****). Also shown are the CTH values that were labelled semi-transparent and those with good quality flags. The diamond shapes depict the main mean values (in the radar case measured in a 30-minute time window) while the red squares represent the semi-transparent CTH (as given by the satellite-derived quality flag). The x-axis represents the satellite CTH in km while the y-axis represents the radar CTH.* 27

Figure 5.2 *Scatter plot depicting MSG/SEVIRI and RADAR daily averaged CTH for each month.* 28

Figure 6.1 *Frequencies of the difference between MSG/SEVIRI and ground-based water occurrence for May, July, October and December 2004.* 33

Figure 7.1 *Frequency distributions of MSG/SEVIRI (Meteosat-8) and pyranometer-derived cloud optical thickness for May, July and October 2004 for Cabauw, The Netherlands. The left panels present the frequency distributions of MSG/SEVIRI Cloud Optical Thickness (COT), the centre panels the frequency distributions of pyranometer COT and the right panels the frequency distributions of difference between MSG and pyranometer COT.* 39

Figure 7.2 *Scatterplot of hourly pyranometer COT plotted against MSG/SEVIRI retrieved COT for May, July and October 2004 for Cabauw, The Netherlands.* 40

Figure 7.3 *Scatterplot of daily mean pyranometer COT plotted against MSG/SEVIRI retrieved COT for May, July and October 2004 for Cabauw, The Netherlands. The results are plotted on a logarithmic scale.* 41

Figure 7.4. *Error in retrieved cloud optical thickness (-) and droplet effective radius assuming errors of ± 1, 2 and 3% in the 0.63 □m reflectances (left panel) and 1.61□m reflectances (right panel).* 43

Figure 7.5 *Cloud optical thickness vs. diffuse to clear-sky irradiance ratio with uncertainty range.* 44

Figure 8.1 *Frequency distributions of MSG/SEVIRI (Meteosat-8) and microwave radiometer observed Cloud Liquid Water Path (CLWP) for May and July 2004 for Chilbolton, UK. The left panel present the frequency distributions of MSG/SEVIRI CLWP, the centre panels the frequency distributions of microwave radiometer CLWP and the right panels the frequency distributions of difference between MSG/SEVIRI and microwave radiometer CLWP.* 47

Figure 8.2 *Frequency distributions of MSG/SEVIRI and microwave radiometer observed cloud liquid water path for October and, December 2004 for Chilbolton, UK. The left panels present the frequency distributions of MSG/SEVIRI CLWP, the centre panel the frequency distributions of microwave radiometer CLWP and the right panel the frequency distributions of difference between MSG/SEVIRI and microwave radiometer CLWP.* 48

Figure 8.3 *Scatterplot of hourly microwave radiometer liquid water path plotted against MSG/SEVIRI retrieved CLWP for May, July, October and December 2004 for Chilbolton, UK. The results are plotted on a logarithmic scale.* 49

Figure 8.4 *Scatterplot of daily mean microwave radiometer Liquid Water Path (LWP) plotted against MSG/SEVIRI retrieved CLWP for May, July, October and December 2004 for Chilbolton, UK.* 50

1 INTRODUCTION

This report presents the results of a validation of the MSG/SEVIRI cloud products for the CM-SAF. It was prepared for the Operational Readiness Review of the second version of the CM-SAF production system (CM-SAF ORR V2).

The CM-SAF cloud products and their main characteristics (i.e., notation, spatial and temporal resolution) are summarised in the following Table 1.1:

Table 1.1 *The CM-SAF cloud products based on MSG/SEVIRI data*

Product	Acronym	Resolution			
		Spatial	Temporal		
			Daily	Monthly	MMDC (Monthly Mean Diurnal Cycle)
Fractional cloud cover	CFC	15 km	✓	✓	✓
Cloud type	CTY	15 km	✓	✓	✓
Cloud top temperature, height, and pressure	CTT, CTH, CTP	15 km	✓	✓	✓
Cloud phase	CPH	15 km	✓	✓	✓
Cloud optical thickness	COT	15 km	✓	✓	✓
Cloud water path	CWP	15 km	✓	✓	✓

All products will be shortly introduced and described below in association with the presentation of validation results. However, for a detailed description of product content and methodology the reader is referred to the CM-SAF User Manual of Products (CM-SAF UMP, 2005).

The goal for the validation activity has been to provide results based on at least four full months of data for which all cloud products have been validated simultaneously. The chosen months were selected in order to give an idea of the typical performance of the cloud products under different seasons and they are: May 2004, July 2004, October 2004 and December 2004.

Originally, it was considered to validate results for a full year but the recommendation from the Review Board for the previous review (SIVVRR V2) to immediately start implementation of the next version of the SAFNWC MSG/SEVIRI cloud software (Version 1.2) has forced through a reduced validation ambition here. Instead, we will later repeat the described ORR V2 validation activities with the updated MSG cloud software as soon as possible after ORR V2 with the goal of upgrading the Version 2 production system with the new MSG cloud software already during autumn 2005. Consequently, the idea is to use the ORR V2 results (achieved accuracies) given in this report as a minimum requirement for allowing an upgrade of the production system and this will be tested in a repeated validation exercise after ORR V2.

On a longer time perspective, subsequent validation activities will also produce validation results with the upgraded MSG cloud software for a full year of data until the second operations review (OR 2) in the second half of 2006. All planned post-ORR V2 validation activities have been described in a separate Validation Plan document that was delivered to the ORR V2 Review Board in May 2005.

In the following, sections 3 to section 8 describe the used observations, the validation methods and the detailed validation results for each of the validated six cloud products. Finally, section 9 gives a short summary of all validation results and a discussion on the product maturity and any possible consequences concerning future development efforts.

Since the validation data sets from a few particularly chosen observation sites are fundamental for the validation of many of the products, these are first described in the section 2 prior to the presentation of the results.

2 CLOUD OBSERVATION AT THE CABAUW, CHILBOLTON AND PALAISEU OBSERVATION SITES

2.1 Individual cloud observation instruments

In this sub-section a summary is given of the instruments that were used to measure cloud physical parameters and which have been used in this study to validate a majority of the CM-SAF cloud products.

2.1.1 Lidar

The lidar is a high power laser that emits short pulses of light. The light scattered back from atmospheric particles and molecules is recorded in a time-gated fashion. The time of flight between emission of the laser pulse and reception of the echo is used to measure the range at which the backscattering has occurred.

Depending on wavelength, lidar is most sensitive to particles with radii between 1 and 3 μm, which are typically radii of cloud water particles.

2.1.2 Cloud radar

The cloud radar measures two cloud physical parameters, namely the distance between the instrument and the cloud particles and the velocity of the moving particles. The cloud radar is most sensitive to scattering of particles with high effective radii but the maximum sensitivity in this respect depends also on the operating frequency. The two radars used here (in Cabauw and Chilbolton) operate at a frequency of 35 GHz leading to an operational wavelength of 8.6 mm which makes it sensible to cloud droplets within the range of 0 to about 200 micron (Rayleigh scattering regime). Considering these radii it is not surprising that the cloud radar is especially suited to measure the particle volume and moving velocity of ice crystals (Donovan *et al.*, 1998).

The finite travel time of the signal from the radar to a target and back causes a frequency difference between the transmitted and received signal. This frequency difference can be obtained by multiplying the transmitted signal and the received signal (mixing) and combine this with low pass filtering of the mixed signal. The resulting signal is called the beat signal. The frequency of the beat signal is directly related to the distance of the target. Further, cloud radar measures the Doppler shift of moving particles. By measuring this shift, the velocity of the targets can be determined. The frequencies in the beat signal are, however, already used to determine the distance of the target. The velocities can still be obtained by measuring the phase shifts for succeeding sweeps.

2.1.3 Microwave radiometer

Microwave radiometers measure incoherent radiant electromagnetic energy. From the ground, zenith-pointing radiometers measure energy radiated (emitted) by atmospheric gases and liquid water in the form of cloud droplets and rain. This energy is dependent on the measurement frequency and is proportional to the amount of material present in the atmosphere. Radiometer measurements at selected frequencies are used to make estimates of integrated water vapor path (WVP) and liquid water path (LWP).

The technique to relate sky brightness temperature (BT) at two or more frequencies to WVP and LWP is accomplished by constructing Look Up Tables (described as a Training Set) of WVP, LWP, and corresponding BTi values. Measurements of BTi from the radiometers can then be transformed into estimates of WVP and LWP.

2.1.4 Pyranometer

The thermoelectric pyranometer is an instrument that can accurately measure broadband hemispherical irradiance in the solar spectral region. Pyranometers are mainly used at meteorological stations to measure the downwelling solar irradiance at the surface. It can either measure the global or the diffuse irradiance. The accuracy of standard pyranometers should be about 3% according WMO standards. However, mean solar irradiances may be underestimated by 3 – 8% due to pollution on the instrument dome, thermal cooling of the detectors and long maintenance intervals (Deneke, 2002).

2.2 Measurement campaigns

In this sub-section a summary is given of the measurement campaigns that were used to acquire the cloud observations for the validation of MSG/SEVIRI-derived cloud physical parameters in the CM-SAF project.

2.2.1 CloudNET

CloudNET is a EU-funded research project that aims to use data obtained quasi-continuously at three remote sensing stations for the development and implementation of cloud remote sensing synergy algorithms. The project started on 1st of April 2001 and will end on 1st April 2005. The three experimental research sites are situated in Cabauw (The Netherlands), Chilbolton (UK) and Palaiseau (France). Each site is equipped with radar, lidar and a suite of passive instrumentation. The use of active instruments (lidar and cloud radar) results in detailed vertical profiles of important cloud parameters, which cannot be derived from current satellite sensing techniques. For these already existing cloud remote sensing stations (CRS-stations) a network will be operated for at least a two year period (2003-2004) to build a co-coordinated, harmonized and joint data archive. The observations will be used to evaluate four operational numerical models and to demonstrate the role that could be played by an operational network of cloud remote sensing stations.

2.2.2 CESAR

The Cabauw Experimental Site for Atmospheric Research (CESAR) is located in The Netherlands. The site consists of a large set of instruments to study the atmosphere and its interaction with the land surface. The CESAR objectives are monitoring of long-term tendencies in atmospheric changes, studies of atmospheric and land surface processes for climate modeling, validation of space-borne observations and development and implementation of new measurement techniques. Monitoring operations started in 2000 and will continue till 2010 or later. The site is equipped with remote sensing instruments, in situ tower instruments and in situ ground instruments. Among others, instruments like lidar, radar, ceilometer and pyranometer are operated.

3 VALIDATION OF THE FRACTIONAL CLOUD COVER PRODUCT (CFC)

In the framework of the CM-SAF, the MSG/SEVIRI cloud mask developed by the Nowcasting SAF is routinely applied to SEVIRI measurements. The cloud masking algorithm applied to MSG SEVIRI data is a multi-spectral thresholding algorithm making use of dynamically adjusted thresholds. These are pre-calculated using radiative transfer calculations where input data sets such as physiography (e.g. land use, topography, etc.) and Numerical Weather Prediction (NWP) analyses are used. The methodology is described more in detail by Derrien and LeGleau (2003) and in SAFNWC SUM/1 (2004).

This validation report focuses on the performance of the MSG cloud mask with respect to daily and monthly average values. Instantaneous results will also be shortly discussed and compared to recent findings of the Nowcasting SAF. For the validation of both daily mean and monthly mean products, five months of MSG/SEVIRI data have been compared to Synop observations (April, May, July, October and December 2004). Figure 3.1 shows the geographical location of all used Synop stations, subdivided into land stations, ship and buoy measurements. Stations identified as ship measurements but clearly located on land are mobile stations.

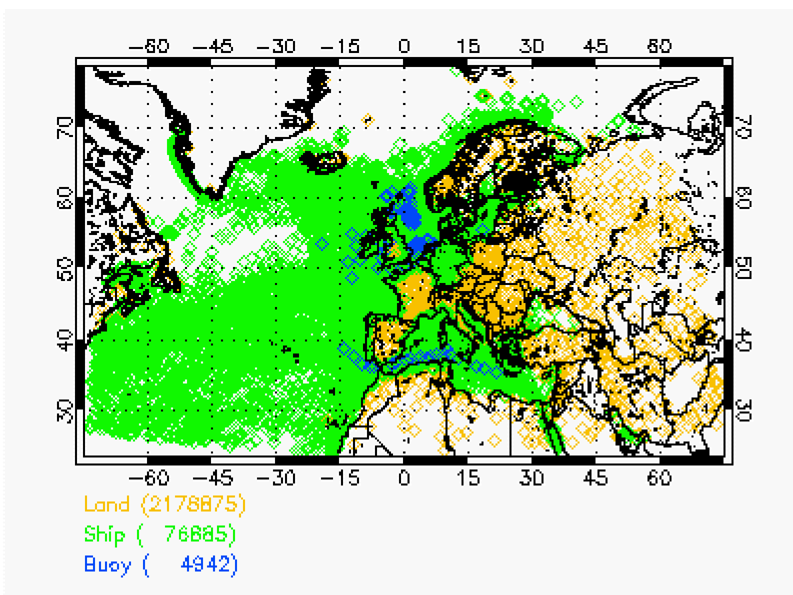

Figure 3.1 *Geographical location of Synop stations used for SEVIRI cloud mask validation. The numbers below the image show the total number of available observations for each Synop type.*

3.1 Validation method

When Synop observations are compared to satellite measurements the strongly differing viewing techniques have to be properly considered. While the satellite instrument is downward looking with, in the case of SEVIRI, a pixel size between 3 and 15 km depending on geographic latitude, the ground-based observer is upward-looking, reporting one observation representative for more or less the complete visible sky. Additionally, different quantities are reported: While the SEVIRI observation is divided in 4 classes (clear sky, cloudy, fractional clouds, snow or ice on the ground) the Synop observation provides a cloud fraction in octa. Here, 0 octa stands for no cloud visible at all, while 8 octa stands for completely overcast sky. As soon as one cloud is visible, 1 octa are reported while as soon as one hole in the cloud cover is visible, 7 octa are reported. Between 2 and 6 octa, the observation should represent the actual cloud fraction above the synop station.

Additionally, the different temporal sampling of the measurements has to be considered. Synop observations are either made hourly or in 3 or 6 hours intervals, respectively. MSG-1 measurements are available with a 15 minutes time interval. As all Synop measurements are done at the full hour, the MSG-1 slot starting 15 minutes before was chosen for comparison. Scanning starts at the South Pole thus effectively reducing the time difference to less than 10 minutes.

The following approach was chosen for the comparison of SEVIRI and Synop observations: For all Synop observations, all SEVIRI pixels from the closest time slot within a radius of 30 km around the Synop station were considered. In a first step, the quality flag provided by the MSG processing software was analysed in order to exclude pixels analysed with a low confidence from further processing. In a second step, the cloud fraction above the Synop stations was calculated from the remaining MSG pixels:

$$CFC_{MSG} = \frac{\sum cloudy + f_{frac} \sum fractional\, clouds}{\sum all\ pixels}.$$

The numerator is calculated from the sum of all pixels classified either as cloudy or partly cloudy, the latter multiplied by a factor between 0 and 1 in order to correct for the fact that fractional clouds do not completely fill the satellite's field of view. The denominator is calculated from the sum of all valid pixels, i.e. pixels which have not been classified as undecided and who don't show the low confidence flag.

The cloud fraction was converted to octa in order to allow an easy comparison with the synop observations. Additionally, in order to mimic the special treatment of octa levels 0 and 8 in the Synop observations, only MSG cloud fractions less than 0.01 were kept in the "0 octa" class and only MSG cloud fractions greater than 0.99 were kept in the "8 octa" class.

It should be stressed that we have not used the operational daily and monthly mean average CM-SAF product in this validation study. The operational product is based on hourly SEVIRI data and produced in (15 km)^2 resolution. Instead, new daily and monthly MSG averages were calculated exclusively at the times and locations where Synop data was available. This was done due to the poor spatial and temporal coverage of the Synop data compared to the satellite observations.

The comparison for land stations is based on the individual Synop stations. Daily averages were only calculated where a minimum of 4 collocated MSG and Synop observations were available during the day; monthly averages were calculated from the daily mean values where a minimum of 10 daily averages were available during the month.

The situation is different for ship and buoy observations. Especially ships travel far within one day, which is why the calculation of daily and monthly averages for fixed locations is impossible. Consequently, a simpler approach was followed here: The available instantaneous MSG and Synop collocations were gridded into a regular latitude / longitude grid with a spatial resolution of 2°. For each month, all derived cloud fractions within each grid box have been averaged to one monthly mean value. No information about daily means is therefore available. However, especially the resulting information about the bias between MSG and Synop observations can help to clarify the question whether cloud fraction above sea surfaces is generally overestimated by satellite. This feature has been obvious for CM-SAF results from polar satellites based on the NOAA AVHRR cloud masking method.

3.2 Validation results for daily mean values over land – individual Synop stations

For each day of the 5 processed months, daily average cloud fractions were calculated for all Synop stations which met the requirement mentioned above.

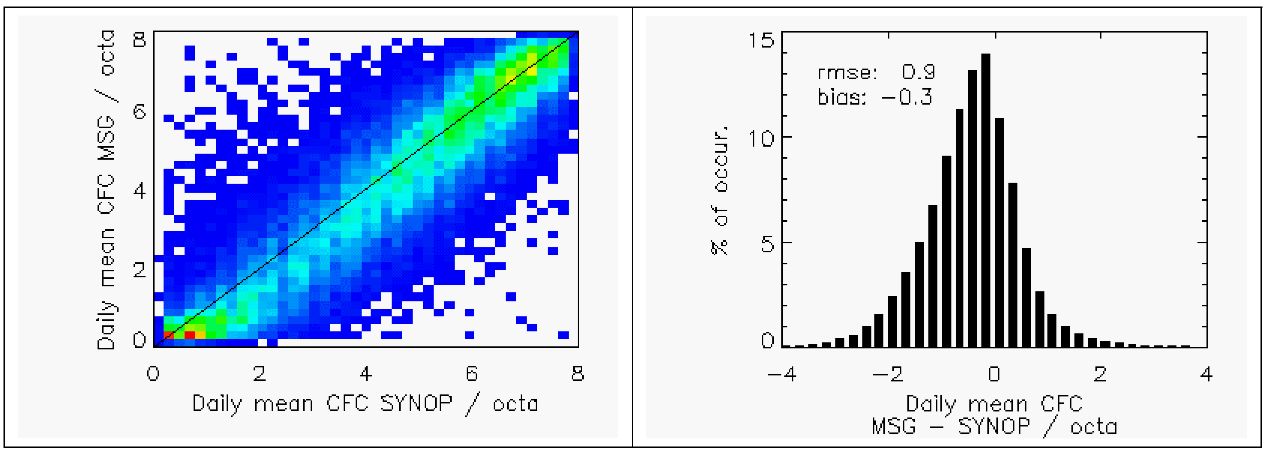

Figure 3.2 *Scatter plot of MSG and Synop daily mean cloud fractions (left) and frequency of occurrence of differences (right). Results are shown for July 2004; all available pixels were used; fractional clouds were weighted with a factor of 0.5.*

It is important to note that two parameters or options have a strong potential influence on the results: Firstly, during the calculation of the MSG cloud fraction, the quality flag provided with each MSG pixel can be used to exclude MSG results with a low confidence rating. Secondly, the factor used for weighting the pixels identified as fractional clouds also has a large potential effect on the averaged result. Both effects were investigated in this study by performing the validation once with all available pixels and once where only those pixels were used which were assigned the *high confidence* quality flag. For both cases, individual validations were performed for fractional cloud factors f_{fraq} of 0, 0.5 and 1.

Exemplarily, the results for July 2004 are shown in the following figures: The colour coded scatter plot of MSG and Synop daily mean cloud fractions is shown together with the histogram of differences in Figure 3.2. Here, all available pixels were used and fractional clouds were weighted with a factor of 0.5. The (bias corrected) root mean square deviation and bias are given in octa; the appropriate cloud cover percentages are 11.3 % and -3.8 %, respectively.

The (bias corrected) root mean square deviations and biases are shown as a function of the fractional clouds factor for all months and both quality flag options in Figure 3.3. It can be seen that the influence of the fractional clouds factor is strong compared to the option of limiting the dataset to the high confidence pixels. As it is to be expected, the mean cloud fraction increases with increasing cloud factor, therefore the bias (calculated as MSG – Synop) increases, too. Results for April and May show significantly larger bias values compared to the other months. This might be related to the fact that during those months the MSG calibration coefficients have not been constant. Later validations performed for spring 2005 will have to prove this assumption. Assuming that it is true we would propose a fractional clouds factor of around 0.75 for the operational system. Although for most months the effect of limiting the dataset to the high confidence pixels is comparably small, the results for December suggest that this is a reasonable option.

The bias corrected root mean square deviation of cloud fractions is more or less independent of both the quality flag and the fractional clouds factor. It is between 0.9 and 1.3 octa or 11 and 16 % cloud cover, respectively, for most of the cases, which is slightly above the target accuracy of 10 % for cloud cover.

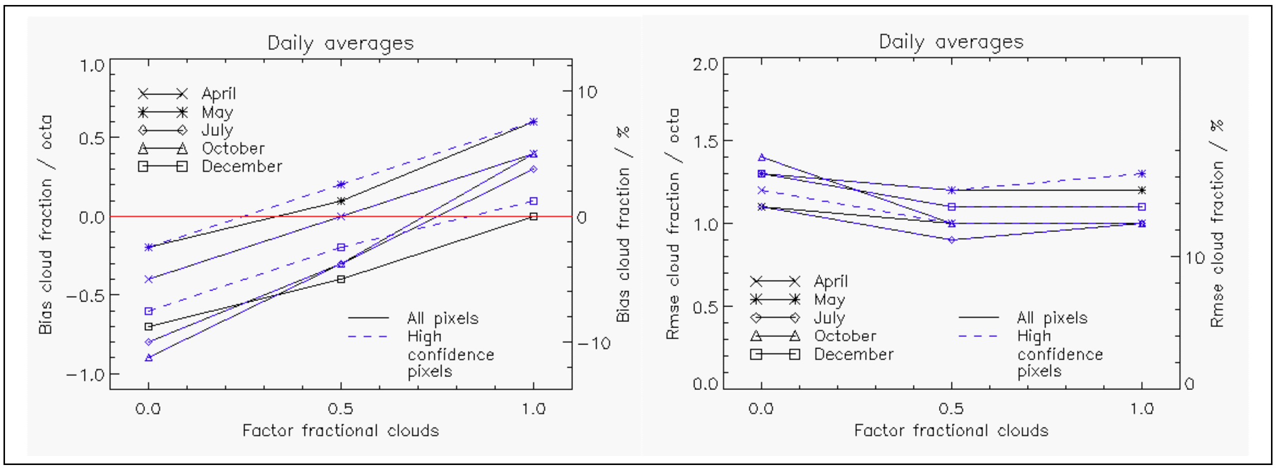

Figure 3.3 *Bias (left) and root mean square deviations (right) for daily mean cloud fractions derived from collocated MSG and Synop observations. Values are shown as a function of the fractional clouds factor; individual plots are shown for each month and for all available pixels / only the high confidence pixels, respectively.*

3.3 Validation results for monthly mean values over land – individual Synop stations

The monthly mean averages have been calculated from the daily mean values for all months and Synop stations where a minimum of 10 daily averages were available per month. The results for July 2004 are again shown exemplarily in Figure 3.4.

It can generally be expected that further averaging reduces the standard deviation of errors and does not significantly affects the biases. However, the biases for daily means and monthly means are not identical as averaging over all daily mean values at once is not mathematical equal to first calculating monthly averages and then averaging those in a second step. In the latter case the individual daily mean values have a relative weight dependant on the total number of daily averages available for each station. This expectation is well fulfilled for July, where the bias corrected root mean square deviation reduced from 0.9 octa (11 % cloud cover) to 0.5 (6 % cloud cover) octa when going from daily to monthly averages.

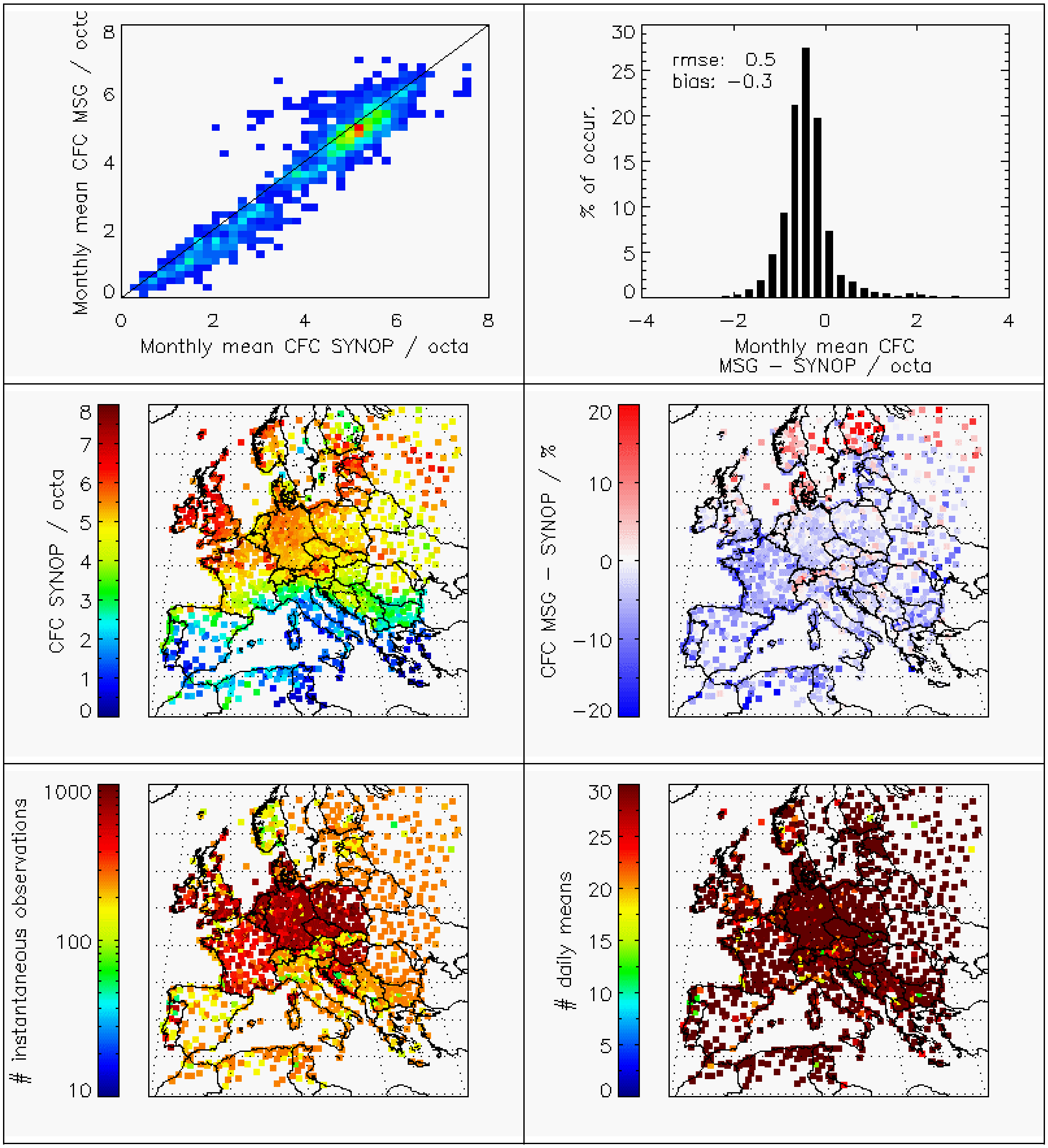

Figure 3.4 *Monthly mean cloud fraction for July 2004; all available pixels were used; fractional clouds were weighted with a factor of 0.5: Scatter plot of MSG and Synop monthly mean cloud fractions (upper left) and frequency of occurrence of differences (upper right). Spatial distribution of monthly mean cloud fraction from Synop observations (middle left) and spatial distribution of relative differences (middle right). Total number of collocated observations (lower left) and total number of daily averages used for calculating the monthly mean values (lower right).*

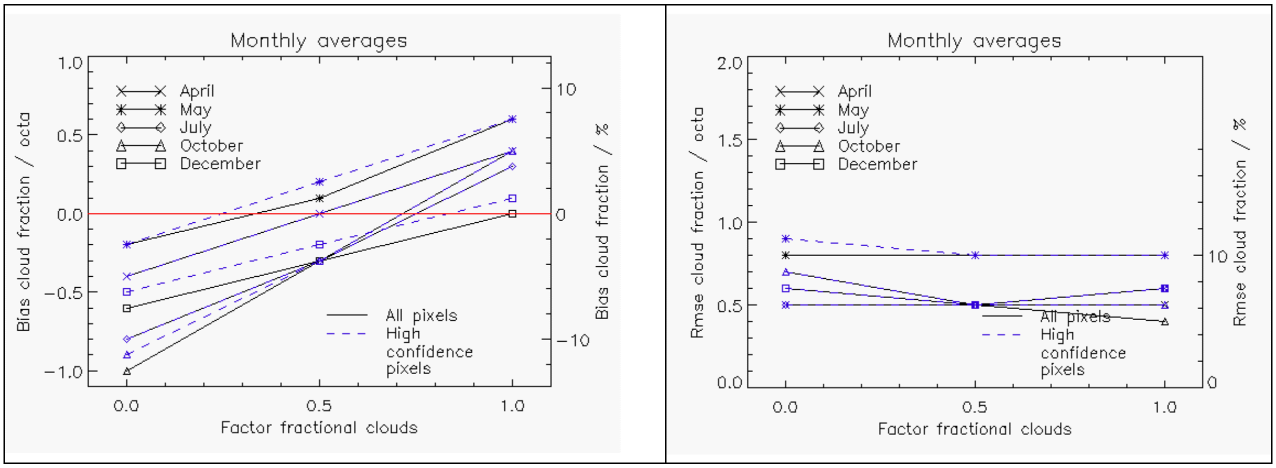

Figure 3.5 *Bias (left) and root mean square deviations (right) for monthly mean cloud fractions derived from MSG and Synop daily averages. Values are shown as a function of the fractional clouds factor; individual plots are shown for each month and for all available pixels / only the high confidence pixels, respectively.*

The combined results for all months are shown in Figure 3.5. Again, the biases did not change significantly when compared to the results from the daily mean calculations; however, it is well visible that for all cases the bias is well below the target accuracy of 10 %.

3.4 Validation results for monthly mean values for land, ship and buoy observations – gridded data

Although the gridding of the data was used in order to allow the calculation of monthly mean values especially for moving ship observations, the results for land observations are also included here. A comparison of these results with the above section can show whether the simpler approach followed here still produces useful information about the quality of the monthly mean product. Figure 3.10 shows the bias between MSG and Synop cloud fractions as a function of the fractional clouds factor for all three Synop types individually.

The results over land shown in the lower panel clearly repeat the results of the above section, indicating that the simplification did not significantly alter the results. The results for ship and buoy observations clearly indicate a strong overestimation of cloud fraction by the satellite data in the order of 10 %. Further investigations will have to show whether this overestimations are e.g. linked to very thin cirrus clouds which are difficult to detect from the ground but whose detection from space is easier over sea than over land.

3.5 Validation results for instantaneous observations

Recently, a validation of the MSG cloud mask algorithm was conducted by the Nowcasting SAF (Derrien and Le Gleau, 2005). MSG results have been compared to Synop observations over central Europe for the period November 2003 to April 2004. The approach for comparing satellite pixel data to ground based Synop observations was similar to the one presented here. However, only the 9 pixel surrounding the pixel closest to the Synop station were used there (while all pixels within 30 km from the Synop station were used here); and every pixel detected as cloud contaminated was counted as fully cloudy. For that reason, our results presented in this particular sub-section are all based on a fractional cloud factor of 1.

Following the approach of the SAFNWC, results have been stratified in terms of solar elevation and in terms of geographical latitude. Three classes of illumination and two geographic classes have been defined as follows: Day: solar elevation > 10°, Twilight, -3° < solar elevation < 10°, night: solar elevation < -3°; Midlatitude: latitude < 55°, Nordic: latitude > 55°.

The following four figures show the bias and (uncorrected) root mean square deviation between instantaneous MSG and Synop cloud fraction as a function of Synop cloud fraction. Results are presented for two months in 2004 (July and December) and for two regions (Midlatitude and Nordic).The chosen months represents the two extreme seasons summer and winter. For the two months not shown here (May and October) results are more or less found as a simple interpolation of the results for the two extreme months. Results for each of the processed months are limited to land observations between 20°W and 40°E.

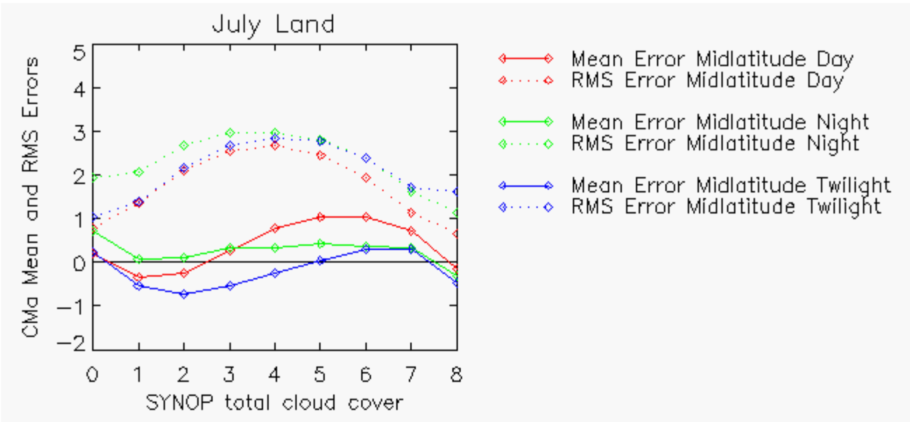

Figure 3.6 *Bias and root mean square deviation of instantaneous MSG and Synop cloud fractions for July 2004 and for* ***Midlatitude regions****.*

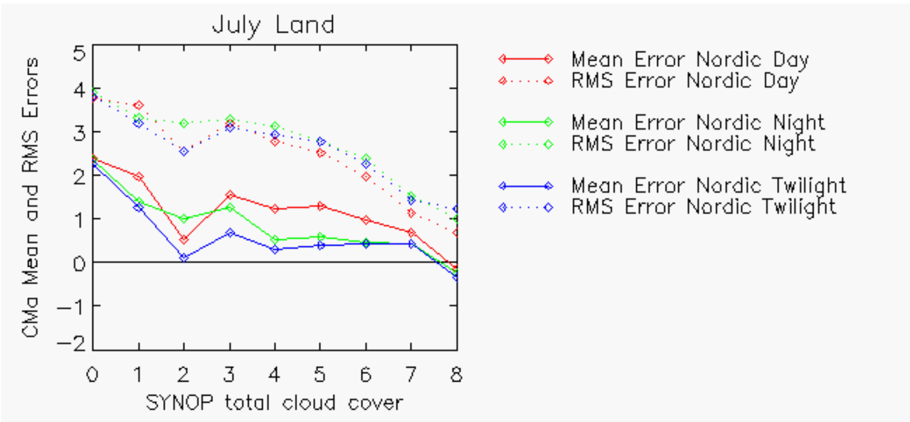

Figure 3.7 *Bias and root mean square deviation of instantaneous MSG and Synop cloud fractions for July 2004 and for* ***Nordic regions.***

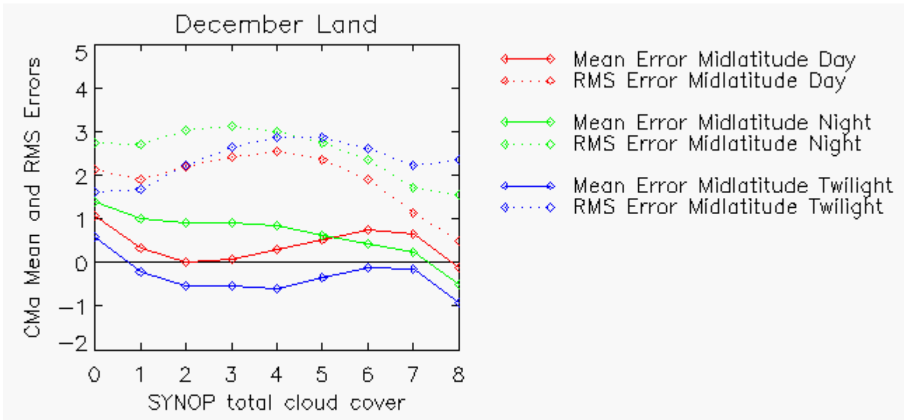

Figure 3.8 *Bias and root mean square deviation of instantaneous MSG and synop cloud fractions for December 2004 and for* ***Midlatitude regions****.*

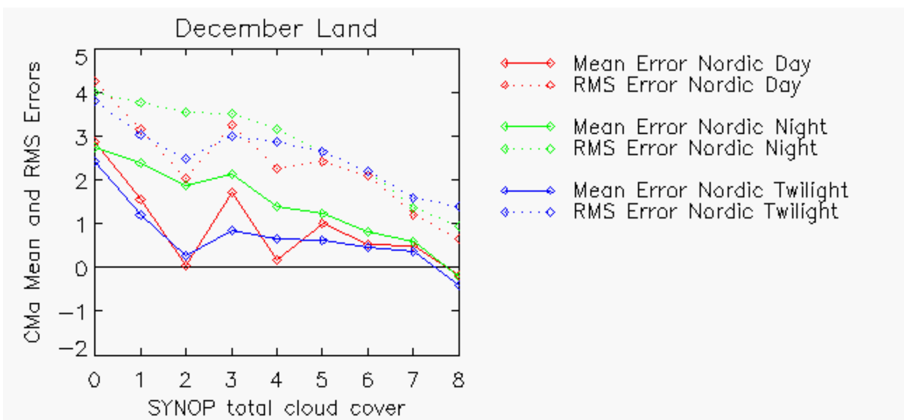

Figure 3.9 *Bias and root mean square deviation of instantaneous MSG and Synop cloud fractions for December 2004 and for* ***Nordic regions****.*

The results for all months repeat more or less the findings reported by Derrien and Le Gleau (2005). The cloud mask performance is generally better for fully cloud free or fully overcast situations. Also, the results for the Nordic regions show an overestimation of cloudiness mostly in clear situations. Cloud fractions generally seem to be overestimated during night time. During twilight conditions, the agreement between MSG and Synop seems to be even better than during daytime, however, this result has to be further investigated, especially with regard to the influence of the fractional clouds factor.

3.6 Summary of CFC validation results

A detailed validation of the NWCSAF MSG SEVIRI cloud mask algorithm was performed with Synop observations. The focus was put on daily and monthly averages. Synop observations of cloud fractions were compared to cloud fractions calculated from MSG observations within 30 km of the Synop stations.

Over land surfaces, daily averages were calculated individually for each Synop station. Subsequently, monthly averages were calculated from the daily mean results. The resulting agreement between averaged MSG and Synop data depends strongly on the factor with which fractional clouds are weighted during the calculation of the individual MSG cloud fractions. The best agreement can be achieved using a factor of 0.75. The agreement also improves slightly when all pixels without the *high confidence* quality flag are excluded from the averages. Bias and bias corrected root mean square deviations are summarised for these settings in Table 3.1 and Table 3.2. For comparison, also uncorrected root mean square deviations are shown in Table 3.3 and it reveals very small differences compared to the bias-corrected ones. Apart from April and May 2004, where the varying MSG calibration coefficients may explain the increased differences, biases are below 0.1 octa or below 1.25 %. Root mean square deviations for daily averages are slightly larger than 10 %, for monthly averages they are clearly below.

The averaged results over the ocean show a significant overestimation of satellite cloud fraction of around 1 octa.

The comparison of instantaneous results confirmed the findings presented recently by the Nowcasting SAF. Individual measurements of cloud fraction by synop and MSG show a mean difference of mostly below 1 octa, with a root mean square deviation below 3 octas.

Future validation activities will have to show whether the clear overestimation of cloud fraction above sea is possibly linked to thin cirrus clouds and whether a constant weighting factor for fractional clouds can really be applied for all seasons and all geographical areas, or whether more sophisticated techniques must be applied.

Table 3.1 *Bias for daily and monthly mean cloud fractions above land derived from collocated MSG and Synop observations. Fractional clouds were weighted with a factor of 0.75.*

	April 2004	May 2004	July 2004	October 2004	December 2004
Daily mean / octa	0.2	0.4	0.0	0.0	-0.1
Daily mean / %	2.5	5.0	0.0	0.0	-1.3
Monthly mean / octa	0.2	0.3	0.0	0.0	-0.1
Monthly mean / %	2.5	3.8	0.0	0.0	-1.3

Table 3.2 *Bias corrected Root Mean Square deviation for daily and monthly mean cloud fractions above land derived from collocated MSG and Synop observations. Fractional clouds were weighted with a factor of 0.75.*

	April 2004	May 2004	July 2004	October 2004	December 2004
Daily mean / octa	1.0	1.2	0.9	1.0	1.1
Daily mean / %	12.5	15	11.2	12.5	13.8
Monthly mean / octa	0.5	0.8	0.5	0.5	0.5
Monthly mean / %	6.3	10.0	6.3	6.3	6.3

Table 3.3 *Uncorrected Root mean square deviation for daily and monthly mean cloud fractions above land derived from collocated MSG and Synop observations. Fractional clouds were weighted with a factor of 0.75.*

	April 2004	May 2004	July 2004	October 2004	December 2004
Daily mean / octa	1.0	1.3	0.9	1.0	1.1
Daily mean / %	12.7	15.8	11.2	12.5	13.8
Monthly mean / octa	0.5	0.9	0.5	0.5	0.5
Monthly mean / %	6.7	10.7	6.3	6.3	6.4

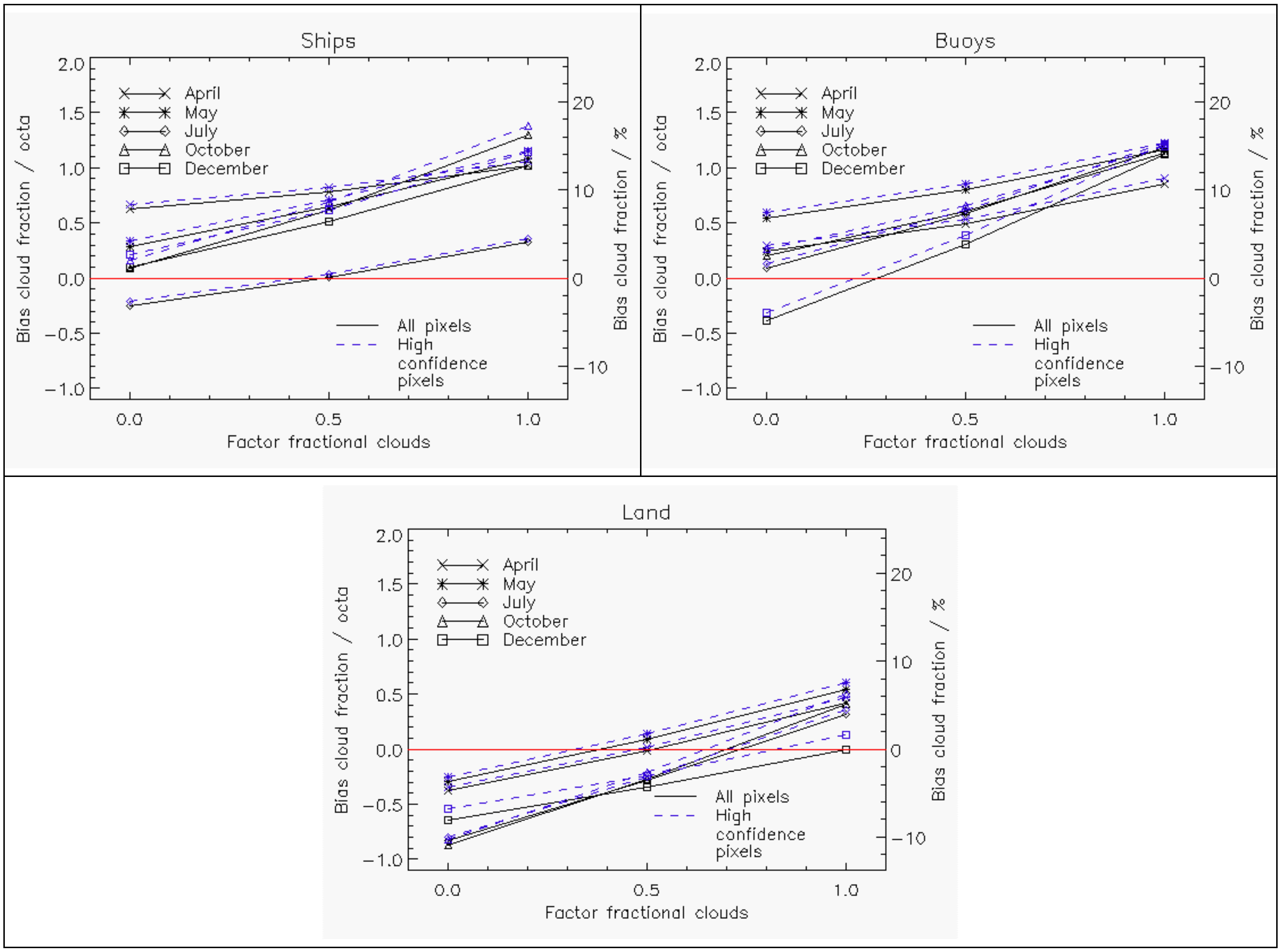

Figure 3.10: *Bias for monthly mean cloud fractions derived from instantaneous MSG and Synop data. Upper left: Ship observations, Upper right: Buoy data, Lower: Land observations. Values are shown as a function of the fractional clouds factor; individual plots are shown for each month and for all available pixels / only the high confidence pixels, respectively.*

4 VALIDATION OF THE CLOUD TYPE PRODUCT (CTY)

The CM-SAF will produce daily and monthly averaged cloud type frequencies derived from MSG data. The MSG cloud type product (CTY) is generated using a multispectral thresholding algorithm and the detailed methodology is described by Derrien and LeGleau (2003) and in SAFNWC SUM/1 (2004). The cloud type assignments of the NWC-SAF product are grouped, by the CM-SAF, into five categories (Low level clouds, Mid level clouds, High opaque clouds, High semi-transparent clouds and Fractional clouds), and for each of this categories daily and monthly cloud frequencies, are computed (further details see CM-SAF UMP, 2005).

The validation of cloud type categorizations is not a straight-forward and obvious procedure when being based on ordinary Synop observations (as a contrast to the previous case of validating the CFC product). The reason is that the different geometrical observation conditions (i.e., satellite viewing downward from top of atmosphere while the surface observer is viewing upward from the surface) make comparisons very difficult. For example, in cases of multi-layered cloudiness the surface may report exclusively low-level clouds while the satellite reports exclusively mid-level or high-level clouds. In this case it is impossible to evaluate the skill of the observation/interpretation since both the surface observation and the satellite interpretation fail in giving a correct description of true cloud conditions.

The objective of this validation report is to present a thorough validation of CM-SAF MSG CTY frequencies. In contrast to the previous CM-SAF validation studies for SIVVRR V2 (System Integration Verification and Validation Readiness Review – validation results given by SIVVRR SR 1 (2005) the number of data sets has been significantly enhanced: Four months of information provided by ground based cloud radar data from the Cabauw station combined with radiosonde measurements are compared with condensed cloud type categories of the MSG product. Similar to the initial SIVVRR V2 validation, we consider cloud radar measurements (in contrast to SYNOP observations) as a reliable source to describe the occurrence and the exact vertical location of cloud layers, although to some extent being limited by some detection problems for optically thin and high clouds. We have simulated satellite viewing conditions (i.e., viewing from space) by assigning cloud types to the cloud radar detected uppermost cloud layer using vertical temperature and pressure information from radiosonde measurements.

4.1 Details of the validation method

Cloud height data retrieved from radar profiles of the Cabauw station were compared to radiosonde measurements of De Bilt (The Netherlands; 52.10 N, 5.18 E) and the corresponding cloud pressure and cloud temperature data were retrieved. In contrast to the initial SIVVRR V2 validation of the CM-SAF cloud products (SIVVRR SR 1, 2005), radar profiles from the Chilbolton station were unfortunately not analysed. This is explained by the much smaller accessible data set from the Chilbolton site (less than 50 % of the Cabauw data coverage) and the lack of adequate radiosonde data for Chilbolton. The option to use Cabauw radiosonde information also for Chilbolton was rejected due to the relatively large distance between the Cabauw and Chilbolton sites.

The cloud radar data used for this validation study covers May, July, October and December in 2004. In July and May the Cabauw radar data set was available nearly continuously from 7:30 am to 5 pm. However, in December and October major data gaps occurred. In total, eighteen days with acceptable radar coverage were available in October and 21 days in December. The cloud radar data from the Cabauw ground-station were processed using a software routine developed by KNMI. For each 30 minute time window a mean cloud radar cloud top height (CTH) value was provided centred on the MSG observation time (slot time). To account for non-homogeneous cloud decks and multi-layer cloud situations a minimum and maximum CTH value was computed.

CTY assignments (Low-level Clouds, Mid-level clouds, High-level clouds) were provided for the corresponding mean, minimum and maximum CTH values. In contrast to normal surface observation oriented CTY definitions, cloud-top altitudes rather than cloud base altitudes are used in order to simulate satellite viewing conditions. Radar retrieved CTH values were converted to CTY assignments using corresponding radiosonde temperature [K] and pressure measurements [hPa] as:

- Low level cloud: $T > 0.8*T_{850hPa} + 0.2*T_{700hPa} - 8$
- Mid level cloud: $0.8*T_{850hPa} + 0.2*T_{700hPa} - 8 > T > 0.5*T_{500hPa} – 0.2*T_{700hPa} + 178$
- High level cloud: $T < 0.5*T_{500hPa} – 0.2*T_{700hPa} + 178$

This vertical CTY separation is in line with the CM-SAF CTY assignment implemented in the MSG algorithm. A different CTY separation approach (based on plain pressure levels) is implemented in the CM-SAF cloud algorithms to define CTY's using NOAA AVHRR data. However, the initial SIVVRR V2 validation study (SIVVRR SR 1, 2005) showed that the two different approaches produced very similar results. Consequently, only the originally used MSG CTY definition was used here.

Daily radar-derived cloud type frequencies were computed by calculating the percentage of low, mid and high level clouds from 30 minute cloud type assignments at the available MSG observation times for each measurement day. The calculation was only performed in case both MSG and AVHRR observation were available.

Single cloud radar samples with high temporal resolution (intervals of a few seconds) were not available and this limits the accuracy of the radar-retrieved CTY frequencies. In case of broken cloud situations or in the presence of non-continuous multi layer cloud decks, one CTY assignment based on a 30 minute averaged CTH does not necessarily reflect the actual cloud situation and might thus lead to erroneous CTY frequencies. In order to study the effect of CTY misclassifications caused by 30 minute averaged CTH values, we have performed a direct MSG CTY/radar CTY comparison for two different data sets: The first data set includes all MSG and radar CTY assignments of an investigated month, while the second data set only included cases with a) similar CTY assignments for maximum, minimum and mean CTH in a 30 min time window and b) radar retrieved single layer flags (here considered as single and continuous cloud layers). In this way we e.g. want to see how often an averaged 30-minute CTH value erroneously could result in a mid-level cloud type when in reality only low level and high level clouds are present simultaneously. In this way we hope to better understand how results are affected in multi-layer cloud situations.

The daily and monthly CTY frequencies were computed without any quality or layer restriction. It was found that such restrictions would significantly reduce the number of available data sets leading to unreliable daily frequencies.

A further limitation of the computed daily radar CTY frequencies is the fact that the radar data were not filtered for rain clouds. Rain droplets can significantly reduce the radar signal and hence lead to incorrect CTH measurements. Finally, the daily radar cloud type frequencies of each investigated month were averaged to compute monthly CTY frequencies in order to make comparison with the corresponding MSG-derived monthly CTY product.

MSG CTY frequencies were calculated following the same principles as used when computing the official monthly CM-SAF product but now only based on the cases with corresponding matching cloud radar observations. In a 3 km x 3 km pixel window centered on the cloud radar station, MSG CTY values were retrieved and re-classified into the categories ‘Low-level clouds’, ‘Mid-level clouds’ and ‘High-level’ clouds. In addition, quality flags are derived for the investigated pixel boxes based on the quality flag information associated with the basic cloud type algorithm. Similar to the SIVVRR V2 validation study (SIVVRR SR 1, 2005) a pixel window instead of a single MSG pixel was chosen for the radar/MSG inter-comparison to account for geometric dislocations of high cloud layers (the so-called parallax effect). However, in contrast to the SIVVRR V2 validation study, the size of the pixel window box was significantly reduced. This is done mainly to be coherent with MSG pixel window sizes applied to validate the physical cloud products (see following sections 6, 7 and 8). A detailed study has outlined that this reduced pixel window will not affect the validation results significantly. Due to the fact that the cloud radar does not allow the discrimination between semi-transparent and opaque cloud types, these two cloud groups were merged into one high-level cloud type category. Furthermore, the category Fractional clouds is not studied at all here due to problems in giving it a proper category definition in the cloud radar data set. For the entire pixel window the coverage (in %) of each cloud type category was computed. Daily averages were computed based on CTY assignments derived from hourly MSG data. Only those MSG data were taken into consideration where a corresponding radar measurement was available. Similar to the calculation of radar CTY frequencies, monthly averages were computed from daily CTY frequencies.

4.2 Results

Table 4.1 shows the summarised results for all individual comparisons. Between 59 % (July) and 73 % (December) of all MSG derived CTY assignments agree with radar cloud categories. The percentage of matches is rather high for low level and high level clouds (maximum 83 %), while it is for all month significantly reduced for mid level clouds. We suspect here (as previously mentioned) that many of the radar-retrieved mid level cloud type assignments are actually caused by temporally averaged CTH values from multiple cloud layers (e.g., high and low level cloud decks) or broken clouds and they could consequently be incorrect. This becomes obvious if we only consider single layer cases[1]: Here the number of mid level clouds is for all month significantly reduced and, in addition, the overall percentage of matches is higher for each investigated month (73 % July to 81 % December). The usage of the MSG quality flag in combination with single layer data sets does not appear to affect the results at all which should be noticed. It probably means that the deviations do not appear to be connected to well-known deficiencies of the MSG Cloud Type algorithm but rather to the deficiencies of the method of assigning the cloud type categories from the cloud radar information.

[1] Single layer: A single layer is defined here in case the CTY assignment of minimum, maximum and mean CTH value in 30 min time window is similar and the single layer flag is set for the radar CTH retrieval.

Table 4.1 *Percentage of matches between cloud radar and MSG derived CTY's for each investigated month.*

	all data		single layer[1]		good quality and single layer	
	N	*percentage matches*	*N*	*percentage matches*	*N*	*percentage matches*
2004 May						
All CTY's	153	**72 %**	94	**76 %**	80	**76 %**
…. low level clouds	*… 53*	*… 81 %*	*… 25*	*… 84 %*	*… 23*	*… 82 %*
… mid level clouds	*… 10*	*… 20 %*	-	-	-	-
… high level clouds	*… 90*	*… 74 %*	*… 69*	*… 74 %*	*… 57*	*… 74 %*
2004 July						
All CTY's	167	**59 %**	62	**73 %**	52	**73 %**
… low level clouds	*… 58*	*… 59 %*	*… 22*	*… 64 %*	*… 21*	*… 67 %*
… mid level clouds	*… 38*	*… 37 %*	*… 3*	*… 33 %*	*… 2*	*… 50 %*
… high level clouds	*… 71*	*… 70 %*	*… 37*	*… 81 %*	*… 29*	*… 79 %*
2004 Oct						
All CTY's	75	**64 %**	41	**75 %**	36	**72 %**
… low level clouds	*… 29*	*… 83 %*	*… 18*	*… 83 %*	*… 15*	*… 80 %*
… mid level clouds	*… 20*	*… 30 %*	*… 4*	*… 25 %*	*… 4*	*… 25 %*
… high level clouds	*… 26*	*… 70 %*	*… 19*	*… 80 %*	*… 17*	*… 76 %*
2004 Dec						
All CTY's	135	**73 %**	83	**81 %**	70	**80 %**
… low level clouds	*… 69*	*… 77 %*	*… 41*	*… 85 %*	*… 33*	*… 85 %*
… mid level clouds	*… 9*	*… 33 %*	-	-	-	-
… high level clouds	*… 57*	*… 74 %*	*… 42*	*… 79 %*	*… 37*	*… 78 %*

Transferring the results into daily averages yields results as visualised in Figure 4.1 and as summarized in Table 4.2.

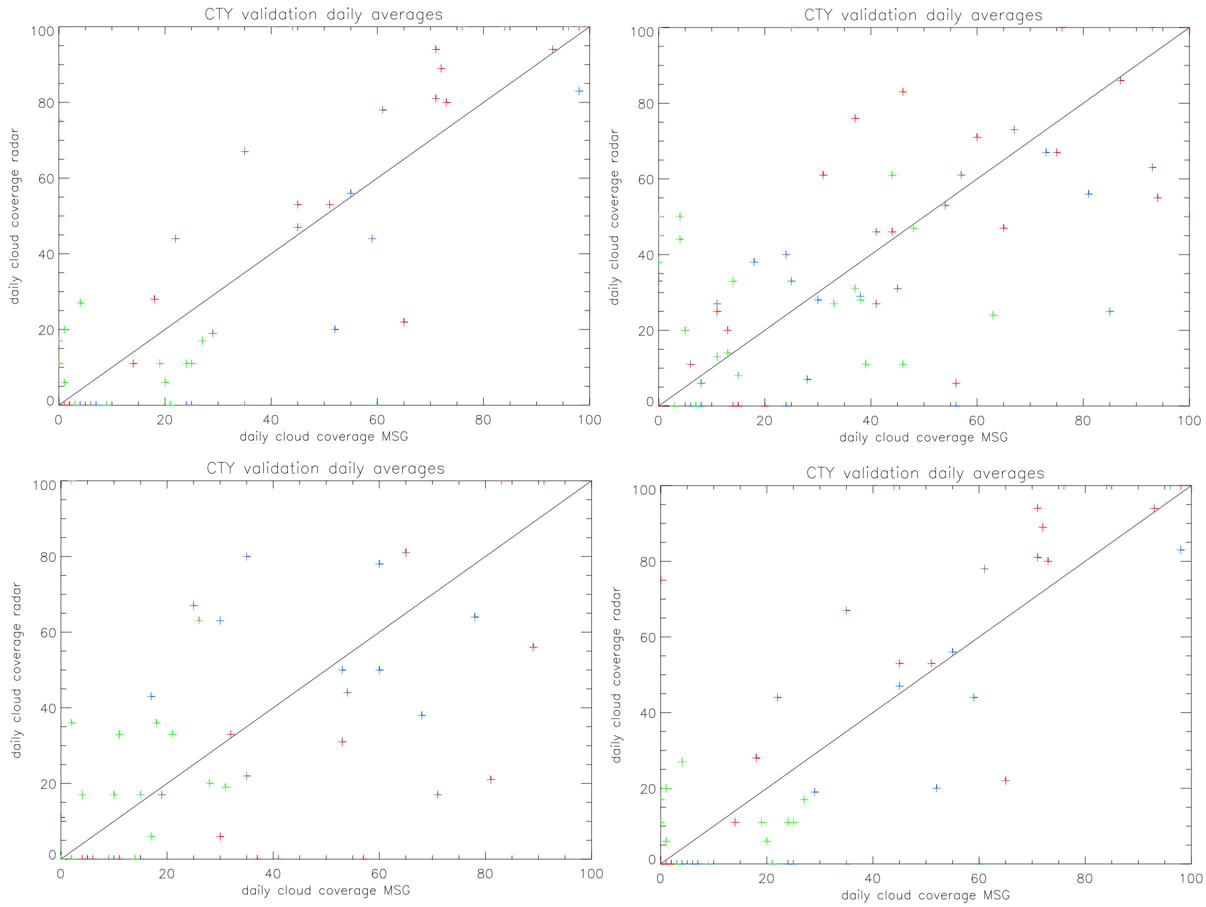

Figure 4.1 Daily cloud type frequencies (%) retrieved from cloud radar data plotted against daily cloud frequencies (%) retrieved from MSG CTY estimations. **Upper left)** May 2004, **Upper right)** July 2004, **Lower left)** October 2004. **Lower right)** December 2004.

+ low level cloud category, + mid level cloud category, + high level cloud category

The daily results reflect approximately the direct MSG/radar CTY comparison considering the fact that no layer restriction was performed here. The linear correlations are rather good for Low-level clouds and High-level-clouds, while they are poor for Mid-level clouds. However, the scatter in the results is significant even for the Low-level and High-level cloud category with RMS errors ranging from 20 % to 30 %. The relatively low linear correlation in October might be due to the limited coverage of radar data.

Finally, if looking at monthly mean values in Table 4.3 the monthly frequencies are naturally improving through the smoothing of the results through the averaging procedure.

Table 4.2 *Linear correlation coefficient (R^2), mean error (bias) and RMS error for daily MSG CTY cloud frequencies compared with daily cloud radar derived frequencies.*

	Low-level clouds Daily CTY frequencies			Mid-level clouds Daily CTY frequencies			High-level clouds Daily CTY frequencies		
	R^2	bias [%]	RMS [%]	R^2	bias [%]	RMS [%]	R^2	bias [%]	RMS [%]
May	0.9	10	23	0.3	1	10	0.8	-10	25
July	0.7	9	28	0.2	-10	32	0.9	2	20
October	0.6	2	26	0.0	-14	30	0.7	11	28
December	0.8	1	23	0.4	4	13	0.7	-5	27

Table 4.3 *Comparison of MSG- and cloud radar retrieved monthly cloud type frequencies. The notation < 12 N per day means that all data is used while >12 N per day means that only a subset of data is used requiring that more than 12 radar measurements per day are available.*

	Low level clouds Monthly cloud frequency		mid level clouds monthly cloud frequency		high level clouds monthly cloud frequency	
	cloud radar %	MSG %	cloud radar %	MSG %	cloud radar %	MSG %
May						
< 12 N per day	42.2	51.8	5.1	5.9	52.7	42.2
> 12 N per day	25.0	32.8	8.0	7.2	67.1	59.9
July						
< 12 N per day	36.6	47.1	25.2	14.5	38.2	38.6
> 12 N per day	30.7	28.9	23.3	25.8	46.1	45.4
October						
< 12 N per day	38.5	45.7	26.8	11.8	34.8	42.5
> 12 N per day	40.8	28.7	24.0	15.3	35.3	56.2
December						
< 12 N per day	56.5	56.0	7.0	15.5	36.6	28.6
> 12 N per day	43.5	49.8	8.2	12.8	48.4	37.5

A small improvement of the correlation between radar and monthly MSG CTY frequencies can be seen if we exclude days with less than 12 measurements per day. Results including the latter sub-set could here be considered as less statistically relevant. With the exception of the results for October (mainly explained by the limited coverage of cloud radar data), the monthly CTY frequencies agree within +/- 10 %. This agrees well with the initial results reported in SIVVRR SR 1 (2005) which is encouraging when considering that we have now included data sets also valid for the more problematic autumn and winter seasons. For the later, there is a less obvious use and availability of visible satellite information and also increased problems for infrared-based retrievals due to similarities of surface temperatures and cloud top temperatures.

Despite the limitations of the applied validation method (no rain filter, daily CTY frequencies are computed from 30 min averaged radar CTH values etc.) monthly MSG and radar retrieved CTY frequencies show good correlations and we believe that this should give good confidence in the quality of the CM-SAF MSG derived CTY frequencies.

4.3 Summary of CTY validation results

The MSG CTY product has been validated using detailed information about cloud layer occurrence and cloud layer altitudes provided by cloud radar information for four months in 2004 (May, July, October and December). A transfer of initial cloud height information from cloud radars into corresponding cloud top temperature and cloud pressure values has enabled the creation of a set of cloud type categories that could be compared to the MSG/SEVIRI CTY product. In general the results confirm previous SIVVRR V2 validation results (SIVVRR SR 1, 2005). If looking closely to case-to-case results, Low-level and High-level cloud assignments appear to work well but problems are evident for Mid-level clouds. However, this problem is shown to be largely explained by limitations of the validation method rather than by incorrect MSG CTY assignments. Daily CTY frequencies show good linear correlations (R^2 between 0.6 and 0.9) although with a considerable scatter of results. However, for the monthly average the difference in the respective occurrences of the cloud categories show small values (less than 10 %).

5 VALIDATION OF THE CLOUD TOP PRODUCT (CTH/CTP/CTT)

The CM-SAF MSG/SEVIRI Cloud Top retrievals are generated using the algorithms developed by the EUMETSAT Nowcasting Satellite Application Facility (SAFNWC) project. The product and the used algorithm(s) are briefly described in CM-SAF UMP (2005) and more details are provided by the corresponding SAFNWC Scientific User Manual (SAFNWC SUM/1, 2004).

In the previous SIVVRR V2 validation study (see SIVVRR SR 1, 2005) cloud top results from May 2004 were evaluated. In this study, the dataset was increased to cover the four seasons of the year represented by the following four months in 2004: May, July, October, and December. As for the SIVVRR V2 study we are only treating the CTH product in this study. The other two versions of the cloud top product (Cloud Top Temperature – CTT – and Cloud Top Pressure – CTP) are just other representations of the results from the same algorithm and should then yield consistent results with the CTH results.

The CTH product was processed using the SAFNWC/MSG software package version 1.1. The MSG derived CTH results were validated against CTH retrievals derived using cloud radar measurements from two observation sites: Chilbolton in the U.K. and Cabauw in the Netherlands. The data set was, however, not complete for neither of these stations. Cabauw was the only station with data for each of the four months and even then, some days were missing. Hourly cloud top estimations as well as daily and monthly CTH averages were compared with the corresponding radar-derived CTH values.

A similar validation study—focusing only on the instantaneous values—has recently been conducted by Meteo-France in Visiting Scientist study of the SAFNWC project (Daloze and Haeffelin, 2005). Due to its high relevance also for the CM-SAF work, it is first summarized below. Following this summary, the methods and results of the CM-SAF study will be presented and further discussed.

5.1 Summary of SAFNWC/Meteo-France CTH Validation study

Meteo France conducted a very thorough validation study for the CTH cloud product, including the use of both cloud radar and LIDAR instruments from the observation site in Palaiseau/Paris. This enabled Meteo-France to validate further aspects of the MSG/SEVIRI cloud products. For example, the CTH retrieval method for semi-transparent clouds was also verified as the lidar instrument was able to provide a ground-based semi-transparent CTH estimation. However, exactly how this was made is not reported by Daloze and Haeffelin (2005).

Meteo France used a combination of cloud radar and lidar information to verify MSG/SEVIRI CTH products. The software version used to derive the CTH product was the same as being used here for the CM-SAF ORR V2 review. The MSG data collected for the study spanned a full thirteen months from 01/09/2003 to 31/10/2004 and was collected between 06h00 and 18h00 UTC. A region of 33 x 33 pixels, centered on the observation site called SIRTA, was extracted and used for the study. A total of 16987 SEVIRI scenes were processed to compare the cloud products while a 5 x 3 pixel-window, centered on SIRTA, was used to calculate the cloud parameters.

Meteo France used a combination of radar and lidar measurements to measure cloud properties. Data retrieved from the synergy of the radar and lidar information are called the RALI data in the study. In this context, it should be known that lidar data alone can not provide reliable CTH estimates for the following situations:

- Optically thick water clouds
- Ice clouds overlying a solid continuous layer of optically thick water clouds

Also, in multi-layer situations when a solid continuous layer of optically thick water clouds is underlying a layer of optically thin ice clouds at a high altitude, the RALI combination did not give reliable CTH estimates. RALI data was obtained while operating in one of the following modes:

- Mode 0: measurement during radar data acquisition time
- Mode 1: measurement during simultaneous radar and lidar data acquisition time
- Mode 2: measurement during lidar data acquisition time
- Mode 3: measurement during radar and/or lidar data acquisition time

For all CTH validation studies, mode 1 gave the most reliable data, but the number of combined lidar/radar observations is smaller than lidar alone (mode 2) or radar alone (mode 0). Therefore they validated CTH values for

- Opaque clouds, using the mode 0 dataset
- Semi-transparent clouds: using the mode 2 dataset

The data in the study was averaged temporally over a 30 minute period. Also, when multiple cloud layers were detected the MSG CTH value was compared to the closest layer to the ground.

A total of 1080 pairs of observations were found suitable for this validation. The frequency distribution of the RALI CTH, averaged over 30 minutes, was at maximum for heights ranging between 8 and 9 km while the SEVIRI CTH, averaged over 15 pixels, showed a frequency maximum between 7 and 8 km. Thus, the SAFNWC derived CTH product showed some underestimation of (mainly) opaque cloud top heights. The statistical results of this comparison gave a negative bias of 420 meters.

Despite the inherent difficulties encountered when comparing the measurements from the satellite's space view with the RALI's point source measurements, the results showed good agreement between the instruments for both opaque and semi-transparent clouds. The results are summarised in Table 5.1 below (notice that two methods – Intercept vs radiance ratioing – are used for the semi-transparent retrieval).

Table 5.1 *Statistical results from Meteo-France's CTH validation study (Daloze and Haeffelin, 2005).*

Comparison	*Bias (m)*	*STD (m)*	*Number of observations*
OPAQUE SEVIRI CTH versus RALI CTH	- 420	1180	1077
SEVIRI SEM-Tr. CTH (intercept) versus LNA CTH	- 920	1110	317
SEVIRI SEM-Tr. CTH (radiance ratioing) versus RALI CTH	60	990	214

5.2 Validation method used for the ORR V2 validation

This section briefly describes the methods used to compare CM-SAF MSG/SEVIRI CTH retrievals with that of the cloud radar.

The cloud radar data were collected at Cabauw (51.971 North 4.927 East) and Chilbolton (51.1445 North 1.4370 West) during the CloudNet campaign. A detailed description of the CloudNet campaign can be found at the following site:

http://www.met.reading.ac.uk/radar/cloudnet/index.html

The cloud radars in Cabauw and Chilbolton were operated during the CloudNet campaign usually from 7:30 am to 5:00 pm. The dataset had days with only a few hours of measurements and not all months had the full amount of corresponding days. Moreover, only data from Cabauw was available for all four months studied. For example, one month had only eighteen days worth of data and some of these days had only one or two hours of data. Therefore, the cloud data used was limited and the total observations per month varied from 95 to 195. This averages out to about 3-6 observations per day and per month. The cloud radar data were processed and the cloud boundaries were computed using a software routine developed at KNMI. For more details about the radar cloud retrieval, see SIVVRR SR 1 (2005).

Generally, clouds and clear sky can be unambiguously detected by the cloud radar through the appearance of sharp boundaries. However, very high and optically thin cloud layers might not be detected by the cloud radar and hence CTH obtained by the cloud radar does not always reflect the height of the uppermost cloud layer. In order to compare the radar time series with the nearly instantaneous MSG/SEVIRI measurements, the radar CTH retrievals were sampled over 30 minutes centred on quarter to and quarter past the hour (MSG/SEVIRI observations are always made a quarter to the hour).

The CTH retrievals were averaged over a 30-minute time window to give the mean cloud radar CTH value. Because cloud decks are not necessarily continuous in a 30-minute time window, additional information about the CTH variation in the time window is provided by means of minimum and maximum CTH. The method used here for calculating the mean value is slightly different to the method used for SIVVRR V2. Here, a more efficient and accurate method was applied. Daily and monthly CTH averages were produced by calculating averages from continuous 30-minute CTH cloud radar data. In addition, a multiple layer flag was also compiled and used in this study. It is important to point out that the methods used for identification of multiple layers by CM-SAF and by Meteo-France in the previously described study are different.

For the MSG/SEVIRI and cloud radar CTH inter-comparison, results from hourly MSG/SEVIRI scenes were used since this is the chosen temporal resolution of the basic MSG/SEVIRI data set in the CM-SAF. In total, 834 MSG/SEVIRI cloud top estimations and corresponding cloud radar data sets (Cabauw and Chilbolton) were available for this validation study. MSG/SEVIRI CTH pixels were selected in 3 x 3 km pixel boxes centred on the position of the radar station to account for broken cloud situations and geometric dislocations (discussed in SIVVRR SR 1, 2005). Beside the average, maximum, and minimum CTH value, results from semi-transparent or opaque cloud categories were compiled using the CTH product flag. In addition, the quality flags were also extracted for this study to investigate if maybe this additional information could be utilised to improve the correlation between MSG/SEVIRI and radar CTH estimations.

5.3 Results

The results from the comparison between daily and monthly MSG/SEVIRI CTH means and the cloud radar daily and monthly CTH means are given in Table 5.2. They show good agreement between the radar observation and the MSG/SEVIRI estimation on a monthly mean basis. However, for individual daily means variations were quite large. From the six cases, i.e., a combination of the different sites and months, MSG/SEVIRI results showed higher CTH values in two out of three cases or in 66.7% of the time.

The magnitude varied from month to month but for the most part, radar mean observations were about half a kilometre (or less) lower than the monthly-mean CTH for MSG/SEVIRI. This difference was, however, sometimes greater than one kilometre (for Cabauw in May) or as small as less than 100 metres (for Chilbolton in May and for Cabauw in July).

The correlation between MSG and radar daily CTH values varied from about 0.72 to 0.92 for the studied months but these variations appeared to have no significant connection to any of the other quality parameters extracted from the dataset. For example, it was hoped that the quality flag would show a connection to a change in the correlation factor by showing that the outliers are indeed MSG CTHs of poor quality, but this proved not to be the case. The outliers remained a mixture of both good and poor quality MSG CTHs. The amount of good/bad quality flagged (**?**) pixels would sometimes be less than 50% with a correlation of 0.88 or be around 50% with a lower correlation. A check was also done against the number of semi-transparent pixels, but this also proved inconclusive.

The radar dataset offered, as well, another possible means of “corralling” the outliers under one significant cause. Multiple cloud layers could have been the cause of a few outliers as was shown in the report by Daloze and Haeffelin (2005), but this also proved to be a dead end. The radar dataset used in the study had very few multiple layer flags set—even when the max and min layers were over 5 km apart. This casts some doubt on the effic7iency of the used “flagging” method. When there was a multi-layer flag in the radar observation, the MSG CTH was compared to the radar minimum CTH. While this was in keeping with Meteo-France’s methods, only in a very few cases it lead to some improvement. But at the same time other comparisons were actually worsened by this. Nevertheless, Figure 5.1 and Figure 5.2 show that, despite the outliers, the agreement is quite good for the daily means but there is a large variation from month to month. As expected, the calculation of the daily averaged mean CTHs (shown in Figure 5.2) smooth out the influence of these outliers.

Table 5.2 *Summary of error statistics for monthly and daily results from Cabauw and Chilbolton. First two columns show monthly means and remaining columns error statistics based on daily averages and some basic observation statistics. The quality column represents the percentage of MSG pixels that were flagged as having good quality.*

Site	Monthly Mean MSG CTH [m]	Monthly Mean RADAR CTH [m]	Daily Mean RMS [m]	Daily Mean Error [m]	Total Obs	Quality [%]	Non Opaque [%]	Multiple Layers [%]	Correlation
May 2004									
Chilbolton	**3285**	**3218**	1478	149	109	48	85	9	0.76
Cabauw	**3916**	**5041**	2196	-1132	153	50	66	21	0.72
July 2004									
Chilbolton	**4491**	**3849**	2000	643	195	49	75	10	0.73
Cabauw	**5215**	**5170**	1075	45	167	53	71	18	0.92
October 2004									
Cabauw	**4912**	**5589**	1977	-677	75	57	77	16	0.73
December 2004									
Cabauw	**4307**	**3971**	1236	336	135	38	77	0	0.88

5.4 Summary of CTH/CTP/CTT validation results

The CTH validation showed that the MSG/SEVIRI-derived CTH and the cloud radar derived CTH are comparable after taking the monthly mean. While this is encouraging for the CM-SAF purposes, it is also clear that there is still room for improvement. The outliers in both Figure 5.1 and Figure 5.2 show that there are still many individual cases showing CTHs with large differences between the compared observations/estimations. This is also evident from the large RMS errors found for the daily averages. The option to use quality information from both the MSG/SEVIRI product's quality flag and the radar's multiple layer flag did unfortunately not improve the results in this respect. Finally, it must be stated that it is not fully clear to what degree these deviations are really explained by true product deficiencies or whether there still are substantial error sources involved due to deficiencies of the validation method (i.e., problems in comparing the two kinds of observations/measurements).

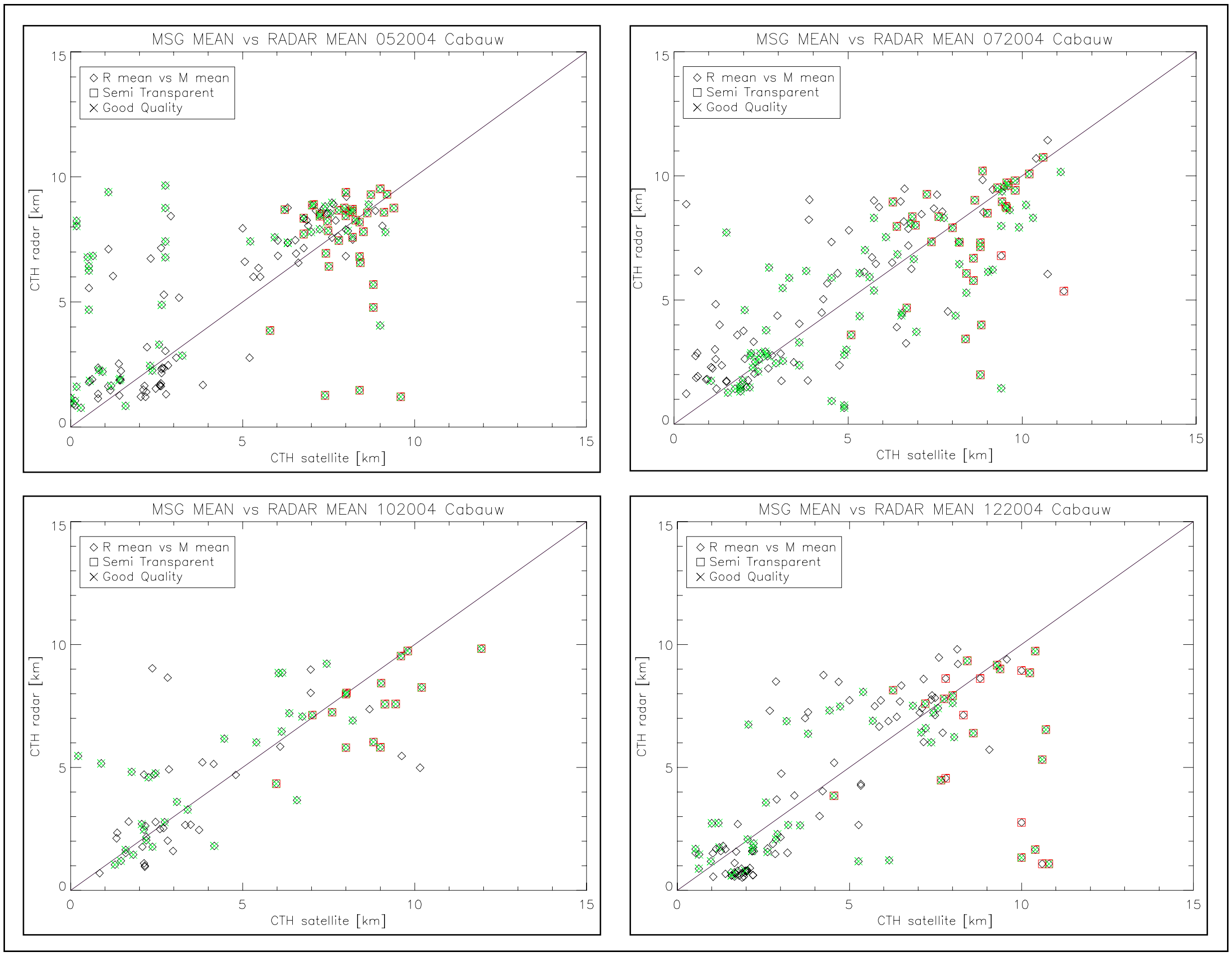

Figure 5.1 *Scatter plot depicting MSG/SEVIRI-derived and radar-observed individual (instantaneous) CTH values for each studied month in 2004 (**Upper left: May, Upper right: July, Lower left: October, Lower right: December**). Also shown are the CTH values that were labelled semi-transparent and those with good quality flags. The diamond shapes depict the main mean values (in the radar case measured in a 30-minute time window) while the red squares represent the semi-transparent CTH (as given by the satellite-derived quality flag). The x-axis represents the satellite CTH in km while the y-axis represents the radar CTH.*

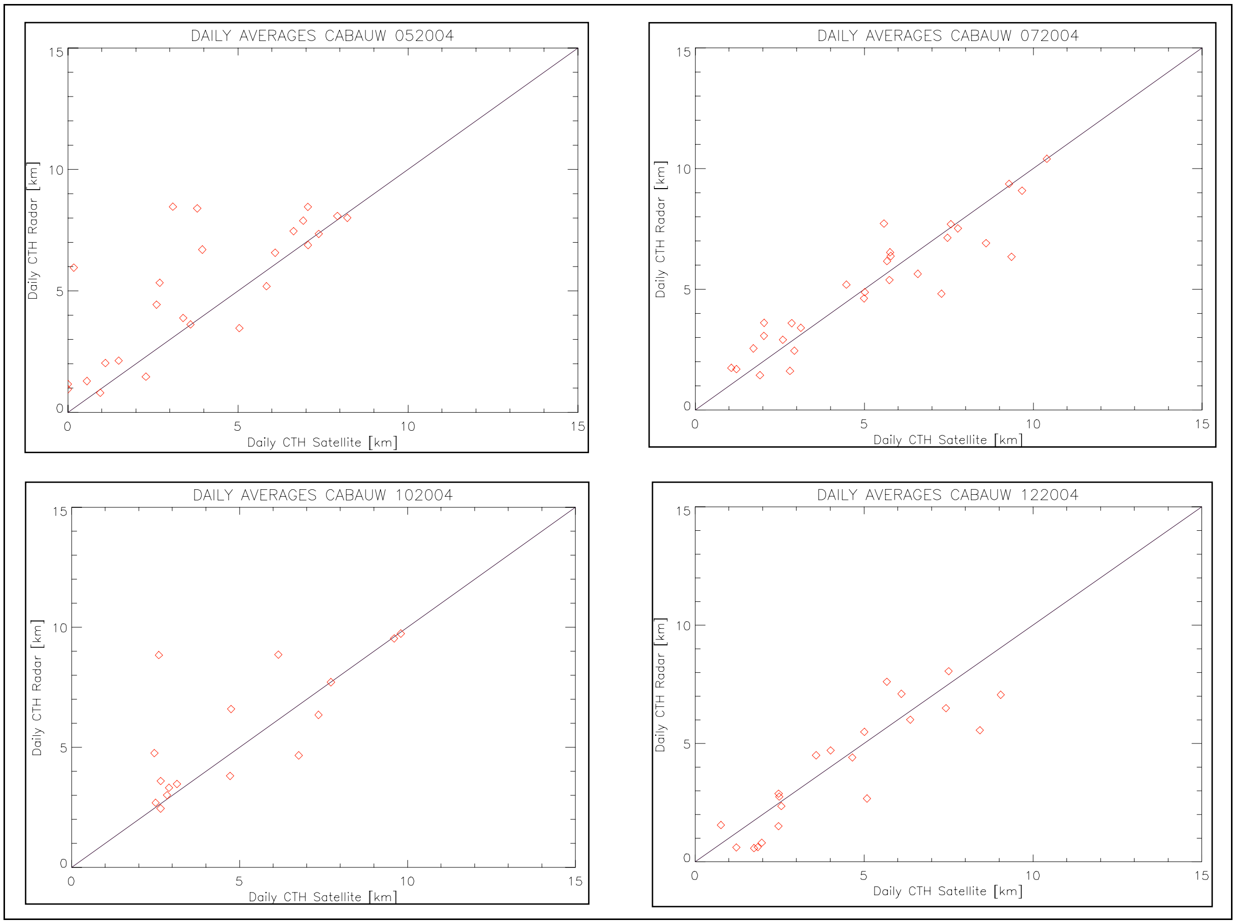

Figure 5.2 *Scatter plot depicting MSG/SEVIRI and RADAR daily averaged CTH for each month.*

6 VALIDATION OF THE THERMODYNAMIC CLOUD PHASE PRODUCT (CPH)

This product gives information on the cloud thermodynamic phase. The product is obtained with use of reflectances at the 0.6 and 1.6 μm channels. Further, retrieval of the ice phase is given an additional evaluation by means of a cloud top temperature check.

6.1 Scientific Approach and Specification

Retrieval of cloud thermodynamic phase is based upon the fact that liquid and solid cloud particles behave differently in terms of reflectance at the 1.6 μm channel. At 0.6 μm, both liquid water and ice have very small values of the imaginary index of refraction, which determines the amount of absorbed solar radiation. At 1.6 μm, the imaginary index of refraction is higher for ice particles than for liquid particles, thus resulting in a smaller reflection from ice clouds. As a result of this, the 1.6 μm channel can be useful to distinguish water from ice clouds (see Baum *et al.*, 2000b). More details of the CPH product is given in CM-SAF UMP (2005).

For the ice phase retrieval the satellite data is also checked for Cloud Top Temperatures < 265 K in addition to the evaluation of the 0.6 and 1.6 μm reflectances. If this condition is true, the pixel is assigned according to the 0.6/1.6 micron retrieval, otherwise the liquid phase is chosen.

6.2 Validation method

Satellite method

The MSG/SEVIRI cloud thermodynamic phase (CPH) is retrieved for a satellite pixel area of 3x3 pixels centred on the observation stations of Cabauw and Chilbolton. Having a single pixel area of ~18 km^2 at 50^o latitude, the validation area represents a total area of ~140 km^2. Hourly MSG/SEVIRI images were collected from the Initial Operational database at DWD. The product output values are water or ice categories for each pixel. Only satellite cases with a cloud cover greater than 78% are included, i.e. at least 7 cloud flagged pixels out of 9. The comparison of the CPH product with ground-based measurements can thus be considered valid for more or less overcast cloud systems. To avoid mixed phase effects, the satellite dataset is further restricted to slots in which at least 80% of the cloud flagged pixels is dominated by one of the two phases. Finally, these conditions result in one cloud phase for the stations of Cabauw and Chilbolton for each satellite time slot.

Ground-based method

Satellite retrieved CPH values are compared to CPH values derived from a cloud radar and lidar combination. This combination was described in detail by Hogan *et al*. (2003). For the ground-based retrieval, every 15 (Cabauw) or 30 seconds (Chilbolton) information on the cloud phase (water or ice) is deduced for up to a maximum of four layers. For the comparison with MSG/SEVIRI retrievals, a time window with duration of 30 minutes centred on the MSG/SEVIRI slot time was chosen to be more or less representative for the horizontal extent of the satellite pixel validation area. Only cloud phase of the topmost cloud layer is considered in this validation. Horizontal resolution of the lidar/radar combination method is a few tens of meters and is thus much finer than the satellite pixel resolution. As a result, in the ground-based retrieval small-scale effects like e.g. patches of cirrus over a lower cloud can play a role. Therefore, a cloud phase is only assigned to the time window period if at least 80% of this period is dominated by one of the phases. Cases with less than 80% of one phase within the time period are rejected from the dataset.

It should be noted that the lidar/radar method is taken as reference; however the lidar/radar method is not always correct. A possible situation is the following: an ice cloud with very small particles (~5 μm) is overlying an optically thick water cloud. As a result of the large droplets in the water cloud, the lidar beam will saturate. In addition, since the ice particles are too small the cloud radar beam will not be able to detect the ice cloud layer aloft. As a consequence, the above described situation will result in a retrieved water phase, whereas the ice phase should have been obtained.

Analysis method

From the dataset of collocated MSG/SEVIRI and ground-based cloud phases for each month contingency tables are made, see Table 6.1.

Table 6.1 *Contingency table for CPH product.*

	Satellite water	***Satellite ice***
Surface water	n_a	n_b
Surface ice	n_c	n_d

The elements in this table represent the following categories: satellite water – surface water (n_a), satellite water – surface ice (n_c), satellite ice - surface water (n_b), satellite ice – surface ice (n_d). From these categories we obtain a global skill score for the CPH retrieval algorithm by:

$$Skill_{weighted} = (n_a + n_d)/(n_a + n_b + n_c + n_d)$$

Difference scores for the water phase (d_w) and for the ice phase (d_i) are calculated to indicate the amount of MSG/SEVIRI CPH retrievals being different from the ground-based retrievals:

$$d_w = n_b/(n_a + n_b)$$

$$d_i = n_c/(n_c + n_d)$$

6.3 Results

Results: all cases

Table 6.2 presents the results of the comparison of MSG/SEVIRI CPH retrievals to collocated results from the lidar/radar method. As can be seen, the total number of observations is strongly biased towards the spring and summer season, having 84% of the total observations. This is not only a result of the decreasing availability of visible channel information during autumn and winter, but also due to the fact that for the Chilbolton station no ground-based cloud phase information for the months of October and December 2004 is available yet. The CPH algorithm shows a rather good performance with global skill scores of 0.80 and 0.70 for the months of May and July 2004 respectively. For the entire dataset, a global skill score of 0.69 is obtained. Results are less favourable for the months of October and December 2004, which might be a result of high solar zenith angles, causing uncertainties in the 0.6 and 1.6 μm reflectances. However, since for these months the number of collocated observations are small, more observations should be included to draw more firm conclusions on this.

Table 6.2 *Global skill scores, water and ice difference scores for May, July, October and December 2004. Numbers of collocated measurements are given between parentheses.*

	May 2004	July 2004	Oct 2004	Dec 2004	Total
Global skill score	0.80 (144)	0.70 (140)	0.57 (23)	0.58 (32)	0.69 (339)
Water difference	0.03 (73)	0.12 (58)	0.00 (10)	0.00 (17)	0.06 (158)
Ice difference	0.38 (71)	0.61 (82)	0.77 (13)	0.67 (15)	0.54 (181)

The difference scores for the ice phase are much higher compared to the difference scores for the water phase, with 54 % of the investigated ice phase retrievals being different from the ground-based retrieval. A possible explanation for these differences is that thin ice clouds are not always detected, despite for the increased ability to detect these type of clouds as a result of the slanted view of MSG/SEVIRI. When an ice cloud is overlying a water cloud, the probability for detecting the ice cloud depends upon the optical thickness of both the water and the ice cloud. An optically thick water cloud decreases the possibility for the overlying ice cloud to be detected, due to the higher reflectance of water clouds at 1.6 μm. As a contrast, an optically thick ice cloud over an optically thin water cloud will likely have the cloud set to the ice phase. For optically thin water clouds the MSG/SEVIRI retrieval can also be different from the ground-based retrieval result. That is, due to semi-transparency of the cloud, surface reflectance at 1.6 μm is contributing to the total 1.6 μm reflectance, which results in a lower 1.6 μm reflectance signal. Since ice clouds have a lower 1.6 μm reflectance than water clouds, the low 1.6 μm reflectance signal may result in the ice phase.

Another possible error source is the collocation problem. Clouds can be assigned too far north due to the large viewing angle at mid-latitudes. This effect is also dependent on cloud height.

Results: single-layer and multi-layer cases

The cloud radar dataset from Cabauw and Chilbolton gives information on cloud properties for four cloud layers at maximum. Therefore, it is possible to distinguish single-layer from multi-layer clouds. Each sample in the 30-minute time window period is evaluated for presence of at least a second cloud layer. Hence, if the number of cloud radar samples with more than one cloud layer exceeds 50% of the cloudy samples, multi-layered clouds are assumed to be present in the surroundings of the observation stations. As a result of this approach, the ground-based dataset is divided into 263 single-layer and 76 multi-layer cases. Global skill score and difference scores are calculated in the same way as for the original dataset. Results are presented in Table 6.3.

Table 6.3 *Global skill scores, water and ice difference scores for May, July, October and December 2004 for single layer (upper table) and multi layer cases (lower table). The number of collocated measurements is given between parentheses.*

Single layer cases

	May 2004	July 2004	Oct 2004	Dec 2004	Total
Global skill score	0.87 (112)	0.67 (104)	0.58 (19)	0.51 (28)	0.75 (263)
Water difference	0.03 (67)	0.12 (58)	0.00 (10)	0.00 (17)	0.06 (152)
Ice difference	0.29 (45)	0.59 (46)	0.89 (9)	0.73 (11)	0.50 (111)

Multi layer cases

	May 2004	July 2004	Oct 2004	Dec 2004	Total
Global skill score	0.56 (32)	0.36 (36)	0.50 (4)	0.50 (4)	0.47 (76)
Water difference	0.00 (6)	- (0)	- (0)	- (0)	0.00 (6)
Ice difference	0.54 (26)	0.64 (36)	0.50 (4)	0.50 (4)	0.59 (70)

From the results it can be seen that global skill scores for single-layer cases are higher than for multi-layer cases. Furthermore, focusing on the months of May and July 2004, the ice difference is lower for the single-layer cases. Since most of the multi-layer cases contain thin cirrus clouds over water clouds and knowing that thin cirrus clouds are difficult to detect by MSG/SEVIRI, it is not very surprising that cases with thin ice over water clouds will raise the ice difference score.

Relative occurrence of the water phase

Another way to assess differences between MSG/SEVIRI CPH and cloud phase values derived from ground-based measurements is done by calculation of the differences in the relative occurrence of the water phase between MSG/SEVIRI and the surface. All MSG/SEVIRI cases with a cloud cover greater than the threshold of 78% are included and relative occurrence of the water phase within the 3x3 satellite validation pixel area is calculated. The relative water occurrences are compared to the same quantity obtained from the collocated 30 minutes lidar/radar measurements and differences are calculated. Subsequently, frequency distributions are made from the resulting dataset of differences.

The frequency distributions of the differences are shown in Figure 6.1. Positive differences indicate a higher relative water occurrence for MSG/SEVIRI than for the ground-based measurements, negative values show lower relative water occurrences for MSG/SEVIRI compared to the surface. From Figure 6.1 it follows that for the months of May and December 2004 differences are small for 50-60% of the observations. Median differences range from 0% in December to +7% in July. However, a significant amount (15-25%) of the cases shows differences of +90-100%. For these cases all MSG/SEVIRI CPH retrievals result in water, whereas the lidar/radar method will detect ice for almost the entire time window period. For the months of October and December 2004 the number of large positive differences is higher than for May and July 2004.

An explanation for this might be the increased path through the atmosphere as a result of the high solar zenith angles in the autumn and winter season. This may result in increased reflectances at both the 0.6 and 1.6 μm channel. This raises the probability for a cloud flagged pixel to be detected as water cloud.

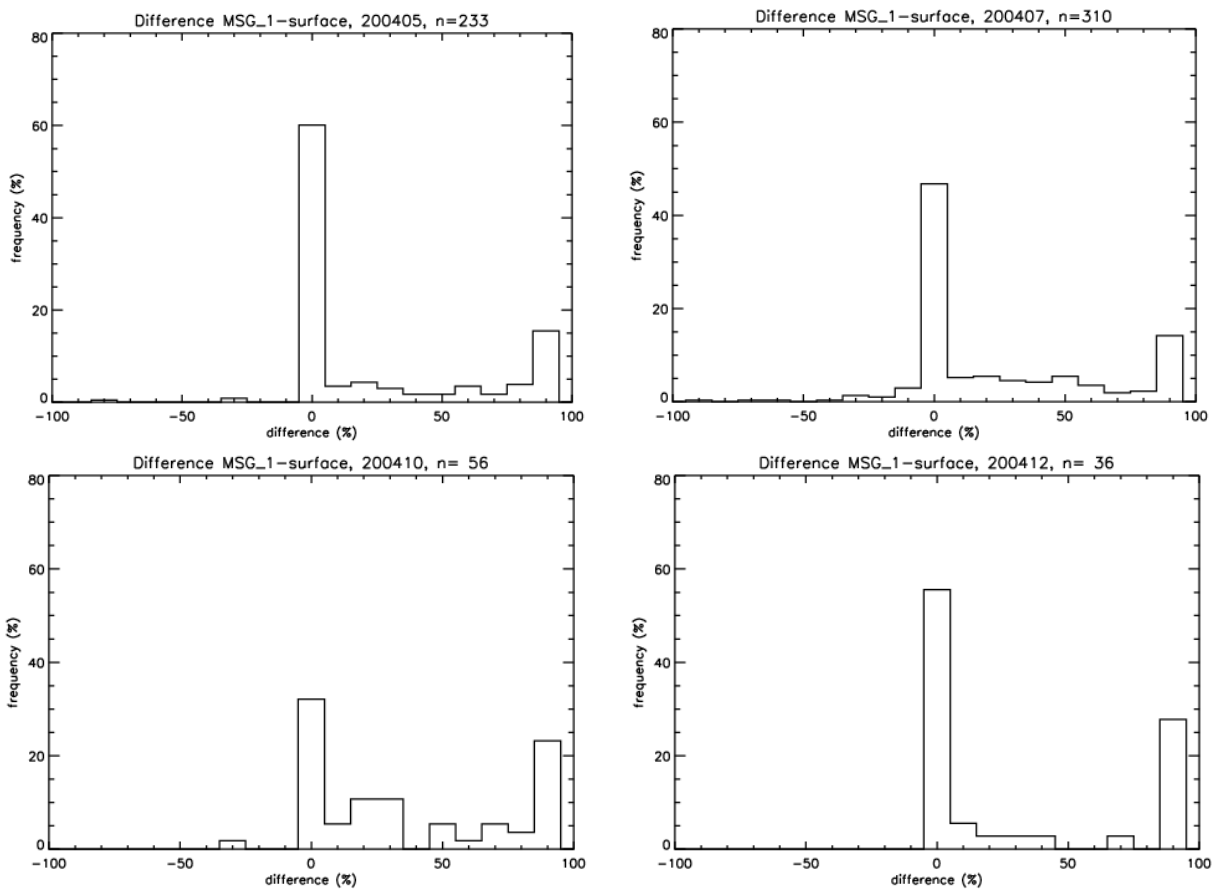

Figure 6.1 *Frequencies of the difference between MSG/SEVIRI and ground-based water occurrence for May, July, October and December 2004.*

Comparing the CM-SAF CPH algorithm to the ISCCP-type of cloud phase algorithm

To assess the quality of the current CPH algorithm more thoroughly, results of the original collocated dataset are compared to cloud thermodynamic phase retrievals done with the algorithm used in the CM-SAF software version for polar satellite data (NOAA AVHRR). This algorithm is based on the approach used in the ISCCP project (Rossow and Schiffer, 1999), which applies threshold values to Cloud Top Temperatures (CTT) to assign cloud thermodynamic phase to a cloud flagged pixel. When the CTT of a cloud flagged pixel is lower than 233 K, the cloud is assumed to consist of ice, whereas a CTT higher than 260 K will result in the liquid phase. Cloud flagged pixels with intermediate CTT values are assumed to consist of both phases. In contrast to the current algorithm, the ISCCP type approach has the advantage that it is not dependent upon the availability of the visible channel, which makes it also available during nighttime.

Cloud Top Temperatures are taken from the current CM-SAF CTT product. For each cloud flagged pixel, CTT is evaluated with the above mentioned temperature thresholds. Hence, similar to the validation of the current CPH algorithm, a satellite cloud cover threshold of 78% (7 out of 9 pixels) is used as well as the 80% threshold for the dominant phase. The result is a single cloud phase for each time slot which is collocated with the cloud phase retrieved from lidar/radar. To obtain a clean comparison, only slots during daytime are investigated. Calculation of skill and difference scores is done in an identical way as for the current CPH retrieval algorithm.

Results of the comparison are shown in Table 6.4. When compared to the current CPH algorithm results, one can see that the total number of collocated observations has decreased by 45%. This is due to the fact that for each pixel the ISCCP type algorithm has three possible output values (water, ice and mixed phase), whereas the current CPH algorithm only obtains water or ice. As a result, the probability that more than 80% of the cloud flagged pixels in the validation area is dominated by either water or ice will decrease.

Comparing results of Table 6.4 to results of Table 6.2 shows that for the months of May, July and October 2004 global skill scores are lower for the ISCCP type algorithm. The total global skill score of the ISCCP type algorithm for this dataset is 0.54, which is 0.15 lower than the current algorithm. Water and ice difference scores are higher for the ISCCP type algorithm, indicating that the retrieved cloud thermodynamic phase by the algorithm differs more often from the ground-based values than the implemented CPH algorithm.

For the dataset investigated it is indicated that the newly developed CPH retrieval algorithm has a better performance than the ISCCP type algorithm used in the previous version. However, more detailed research is needed before drawing firm conclusions. Moreover, it must be emphasized that the current algorithm depends on reflectances at 0.6 μm, which restricts the retrievals to daytime cloud scenes only, whereas a temperature threshold method will also be available at nighttime.

Table 6.4 *Global skill scores for CPH obtained by ISCCP type algorithm and corresponding water and ice difference scores for May, July, October and December 2004. The numbers of collocated measurements are given in parentheses.*

	May 2004	July 2004	Oct 2004	Dec 2004	Total
Global skill score	0.55 (80)	0.52 (84)	0.50 (14)	0.63 (8)	0.54 (186)
Water difference	0.17 (42)	0.38 (32)	0.14 (7)	0.20 (5)	0.25 (86)
Ice difference	0.76 (38)	0.54 (52)	0.86 (7)	0.67 (3)	0.65 (100)

6.4 Further quality studies

Daily and monthly differences in the relative water occurrence between satellite retrieved CPH and cloud phase by the lidar/radar method were also calculated.

To be consistent with the validation of the instantaneous results, the initial 3x3 pixel area was kept. The daily mean relative water occurrence of MSG/SEVIRI is calculated by dividing all cloud flagged pixels labelled as water in the validation area by the total number of cloud flagged pixels for each day. This is done for the Cabauw and Chilbolton station for the months of May and July 2004 and for the Cabauw station for October and December 2004.

Daily mean water occurrence measured from the surface was calculated by dividing the number of samples of the water phase by the number of samples flagged as cloudy. The daily mean difference of relative water occurrence was calculated as MSG/SEVIRI – lidar/radar method. For each month the median and standard deviation of differences were calculated from the daily mean differences. It should be stressed that all cloudy pixels observed in the 3x3 pixel area were included. This is in contrast with the cloud cover selection criterion used for the instantaneous validation, where only cases with at least 7 cloudy pixels out of 9 were included.

Table 6.5 shows the median and standard deviation of differences in the relative water occurrence between MSG/SEVIRI and the lidar/radar method for each month of data. As for the validation of the instantaneous measurements, the number of collocated observations is much higher in the spring and summer season.

Table 6.5 *Median and standard deviation of differences in relative water occurrence for May, July, October and December 2004. Numbers of observations are given between parentheses.*

	May 2004	July 2004	October 2004	December 2004
$\text{Median}_{\text{difference}}$ (%)	25.9 (543)	19.1 (645)	54.5 (154)	55.6 (125)
$\sigma_{\text{difference}}$ (%)	17.5	18.6	16.6	34.3

All months show a positive median difference, which indicates an overestimation of MSG/SEVIRI water phase clouds compared to the lidar/radar observations. This is in line with the results of the instantaneous validation. In addition, the median difference of the monthly results is higher than for the instantaneous results, but this could be an effect of difference in the cloud cover selection criterion.

The standard deviation of difference is about 17%, except for December 2004.

Further, the median difference for the months of October and December is much higher than for May and July. The higher median and standard deviation of difference might be explained by the low number of observations for October and December (only 20% of the total observations) and by less favourable bi-directional reflectance conditions due to high solar zenith angles giving higher reflectances at 0.6 and 1.6 μm.

Table 6.6 *Mean monthly occurrences for water and ice retrievals (in %) for MSG/SEVIRI and for lidar/radar (in parentheses).*

	May 2004	July 2004	Oct 2004	Dec 2004
Water occurrence	80.9 (48.3)	77.1 (53.9)	88.9 (38.5)	83.1 (37.2)
Ice occurrence	19.1 (51.7)	22.9 (46.1)	11.1 (61.5)	16.9 (62.8)

6.5 Summary of the CPH validation results

Four months of CPH algorithm retrievals have been validated with ground-based cloud phase retrievals from a lidar and radar combination method at the Cabauw and Chilbolton observation sites. Global skill scores are best for the months of May and July 2004 (0.80 and 0.70). The entire dataset shows a global skill score of 0.69. Water and ice difference scores, which indicate the amount of MSG/SEVIRI CPH retrievals being different from the ground-based retrievals are 0.06 and 0.54 respectively. The high ice difference score can to a large extent be attributed to semi-transparent and multi-layer cases. Results are also presented in Table 6.6 in terms of the mean monthly occurrences of the ice and water phases which is compliant with the CM-SAF monthly CPH product.

For single-layer and multi-layer cases we see that the global skill score is higher for single-layer than for multi-layer cases. Further, multi-layer cases have a higher ice difference score, indicating that for these cases a higher amount of MSG/SEVIRI CPH retrievals is different from the ground-based retrievals.

An additional quality assessment is done by comparison of results of the currently implemented CPH retrieval algorithm to results of the algorithm used in the previous version, which is based on CTT thresholds. It is shown that for the dataset subject to validation the global skill score of the current CPH algorithm is higher by 0.15.

Results of the current CPH algorithm can be improved by including infrared information, like e.g. brightness temperature difference between 8.5 and 11.0 μm (see Baum *et al.*, 2000). An additional advantage of adding infrared channels is that the algorithm will also be applicable at night time, whereas it is restricted to daytime scenes in its current state. A development of a night time CPH scheme will be considered for future CM-SAF development activities.

7 VALIDATION OF THE CLOUD OPTICAL THICKNESS (COT) PRODUCT

This product provides information on the Cloud Optical Thickness (COT) for pixels that are flagged cloudy by the cloud detection test. The retrieval algorithm is based on 0.6 and 1.6 μm channel data of MSG/SEVIRI. As an intermediate product the cloud droplet effective radius is retrieved as well.

The method to retrieve the cloud optical thickness utilizes solar reflected measurements at non-absorbing and absorbing wavelengths. The underlying principle of the method is that cloud reflectance at a non-absorbing channel in the visible wavelength region is primarily a function of the optical thickness, whereas the cloud reflectance at a water or ice absorbing channel in the near infrared is primarily a function of cloud particle size (Nakajima and King, 1990). The cloud optical thickness and the droplet effective radius are retrieved by combining the information of non-absorbing and absorbing channels. Nakajima and Nakajima (1995) developed a method to retrieve cloud optical thickness and effective particle radius based on satellite channel radiances at 0.6, 3.7 and 10.8 μm. The method used in the CM-SAF is based on the same principle, but utilizes the 1.6 μm near infrared channel instead of the 3.7 μm channel (Watts, 1998, Jolivet, 2000, Roebeling *et al.*, 2003).

7.1 Validation method

The validation of MSG/SEVIRI cloud optical thickness retrievals is done with ground-based pyranometer measurements. Pyranometers measure broadband hemispherical irradiance in the solar spectrum. Methods have been developed to relate pyranometer observations to narrowband cloud optical thicknesses. The principle of these methods is the unique relationship of transmitted solar radiation to cloud optical thickness. A radiative transfer model is used to relate broadband irradiance observations to narrowband cloud optical thickness (Deneke, 2002). The cloud optical thickness is calculated for an imaginary plane-parallel homogeneous cloud layer that produces the same radiation field at the surface as the actual cloud layer. For the cloud microphysical properties a representative water (or ice) cloud has to be assumed with a given droplet size distribution, effective radius and effective variance.

MSG/SEVIRI and pyranometer-derived cloud optical thickness values were compared for three months in 2004: May, July and October. Hourly MSG/SEVIRI images were collected for the CM-SAF baseline area that covers Europe and Northern Africa, from the initial operational database at the DWD. The validation was restricted to the measurement site of Cabauw, The Netherlands (51.97°N, 4.93°E). For the validation in May and July daylight hour observations were used between 8:00 and 18:00 UTC and in October between 9:00 and 16:00 UTC. The intention was also to use observations in December between 11:00 and 13:00 UTC. However, it turned out that the validation had to be cancelled for December 2004 due to lack of useful pyranometer data for Cabauw.

The pyranometer cloud optical thickness was derived with the approach proposed by Barnard and Long (2004), who give a simple empirical equation to approximate cloud optical thickness as function of the solar broadband irradiances. Because this approach is most closely satisfied for fully overcast skies, cases of broken cloudiness were excluded. Since there is no direct way to determine if pyranometer observations represent cases of fully overcast skies it was assumed that skies were overcast when all MSG retrievals within a 3x3 pixels subset were flagged cloudy and the residual between direct and diffuse pyranometer radiances was close to zero. The difference between the transmission through the cloud and the transmission through the atmospheric layers above and below the cloud is ignored due to lack of detailed information on vertical distribution of the atmosphere's optical properties.

The comparison was done for hourly, daily and monthly cloud optical thickness values. The daily and monthly mean cloud optical thickness values (COT_avg) were calculated with the logarithm of the cloud optical thickness to take account of the quasi-logarithmic dependence of cloud albedo on cloud optical thickness, using the following equation:

$$COT\_avg = \exp\left(\frac{\sum_{1}^{n} \log(COT\_i)}{n} \right)$$

where COT_i is the instantaneous cloud optical thickness value and n the number of observations during the averaging period.

The validation of MSG/SEVIRI COT retrievals was done for both water and ice clouds, which is consistent with the pyranometer observations that cannot be sorted on cloud thermodynamic phase. Because the pyranometer COT retrievals are less reliable for large optical thicknesses all retrievals with an optical thickness >250 were excluded. The MSG cloud optical thickness was extracted for the satellite pixel that collocated best with the Cabauw observation site. The pyranometer cloud optical thickness values were calculated for the observations closest to the time that MSG scans the Cabauw observation site, about 13 minutes after MSG starts scanning the southeast side of the Earth disk.

7.2 Results

Frequency distributions of pyranometer and MSG-retrieved optical thickness values are analysed to make a statistical comparison of ground-based and satellite derived cloud optical thicknesses. Figure 7.1 presents for May, July and October 2004 the frequency distributions of MSG/SEVIRI and pyranometer retrieved COT and the frequency distributions of differences between MSG/SEVIRI and pyranometer COT. The number the valid observations per month are restricted due to excluding pyranometer observations of broken cloud fields and with a COT > 250. The sizes of the datasets vary between 39 observations in October and 70 observations in July 2004, which is on the low side for a statistical analysis. For both instruments the frequencies of cloud optical thickness are log-normally distributed. However, the optical thicknesses retrieved from pyranometer are significantly higher than the optical thicknesses retrieved from MSG/SEVIRI. The median pyranometer cloud optical thickness varies between 20 and 30, while the median MSG/SEVIRI cloud optical thickness varies between 5 and 15. Moreover, the minimum pyranometer optical thickness is about 6 whereas the minimum MSG/SEVIRI cloud optical thickness is about zero. This high minimum optical thickness results from selecting only completely overcast sky cases for the retrieval of pyranometer optical thickness. This selection criterion causes an overestimation of pyranometer optical thickness, whereas the MSG optical thickness are not necessarily overestimated since low MSG/SEVIRI COT values can be found due to collocation errors and differences in the sampling area. Both pyranometer and MSG/SEVIRI have a significantly lower median optical thickness for May than for July 2004, with a MSG optical thickness of 5.6 in May and 9.3 in July 2004 and a pyranometer optical thickness

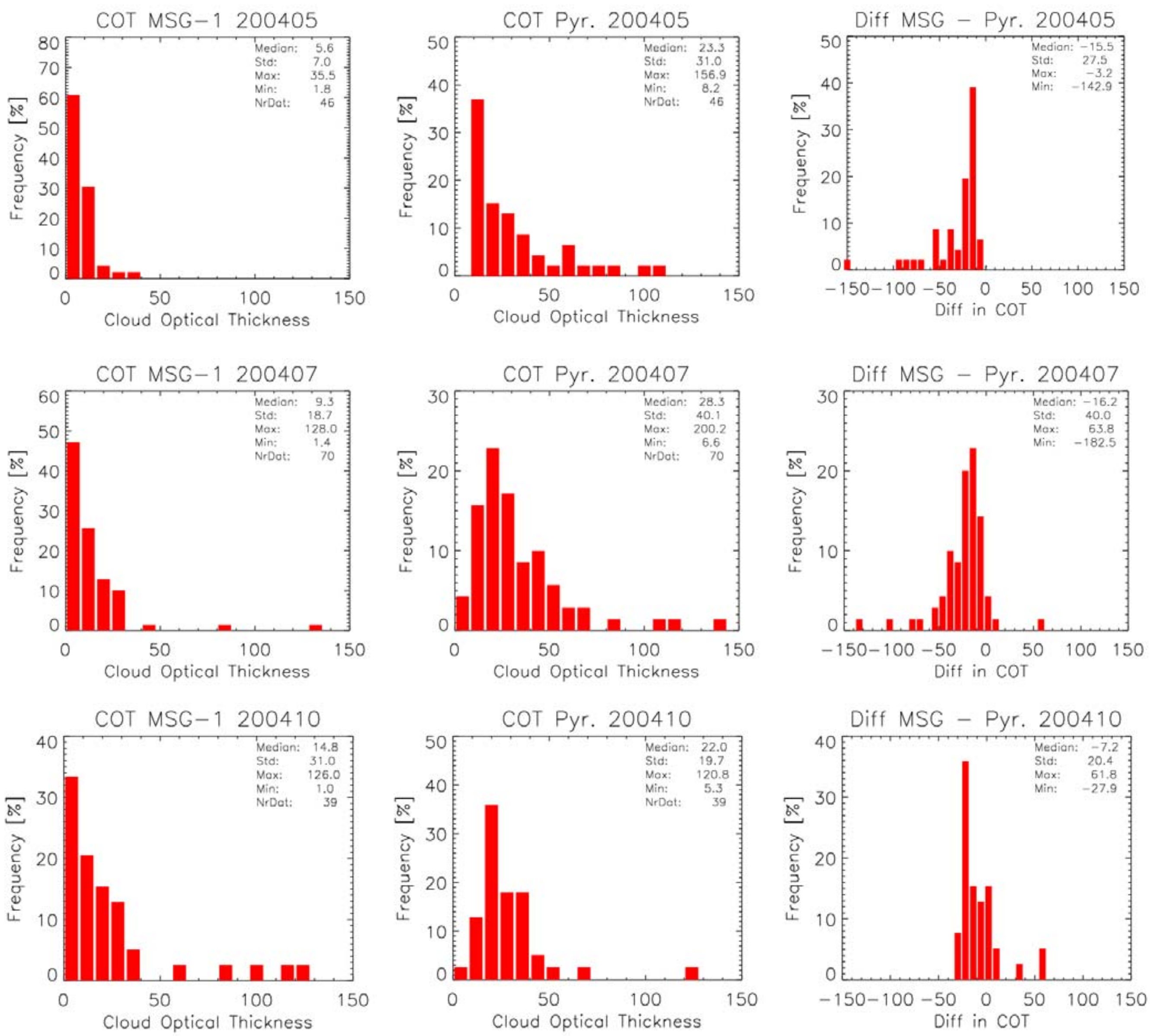

Figure 7.1 *Frequency distributions of MSG/SEVIRI (Meteosat-8) and pyranometer-derived cloud optical thickness for May, July and October 2004 for Cabauw, The Netherlands. The left panels present the frequency distributions of MSG/SEVIRI Cloud Optical Thickness (COT), the centre panels the frequency distributions of pyranometer COT and the right panels the frequency distributions of difference between MSG and pyranometer COT.*

of 23.3 in May and 28.3 in July 2004. The figures of the differences between MSG/SEVIRI and pyranometer cloud optical thickness show for almost all observations that MSG/SEVIRI optical thicknesses are lower than pyranometer optical thicknesses.

Figure 7.2 presents the scatter plots of the logarithm of hourly pyranometer and MSG optical thicknesses. For the three observation months there is a rather weak but positive the relationship between pyranometer and MSG COT. The graphs clearly reveal the underestimation of MSG retrievals. For May and July the slope of the linear regression equation is lower than 0.2, which indicates about 5 times lower MSG optical thicknesses than pyranometer optical thicknesses. The results for October 2004 correlate best (Corr ~ 0.76). In addition, the slope of linear regression equation is higher than 1 (slope ~1.2). The large scatter in the plots most likely originates from spatial mismatch between ground-based and satellite observations. The spatial resolution of MSG images over The Netherlands is about 18 km^2. The representativeness of the pyranometer observations depends on the altitude of the cloud base and the distribution of clouds around the ground-based station. This means that the satellite and the ground-based cloud optical thickness values are likely to represent somewhat different cloud scenes. However, their statistical distributions are expected to be more consistent.

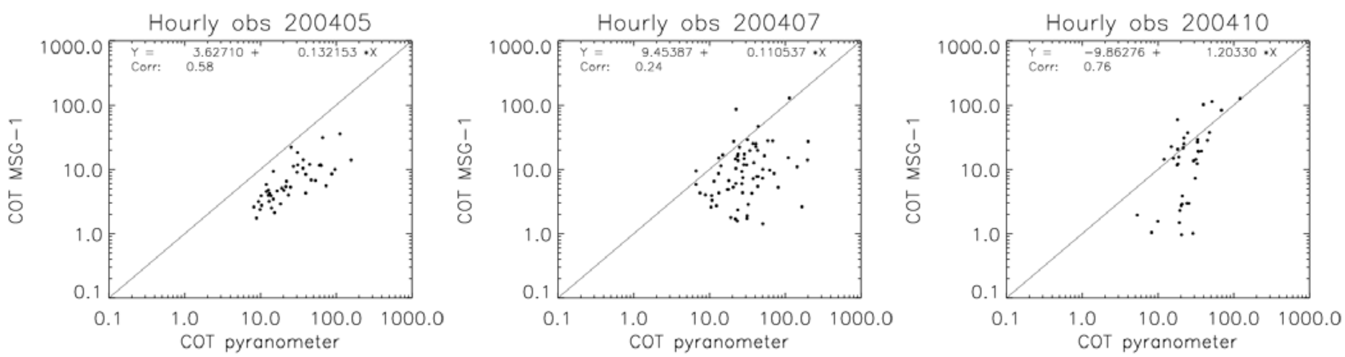

Figure 7.2 *Scatterplot of hourly pyranometer COT plotted against MSG/SEVIRI retrieved COT for May, July and October 2004 for Cabauw, The Netherlands.*

The observed differences cannot be explained by collocation or sampling errors. Part of the observed differences can be dedicated to the sensitivity of the pyranometer cloud optical thickness retrievals to fluctuations in cloud and atmospheric properties. The empirical relationship of Barnard and Long (2004) does not consider variations in solar zenith angle, aerosol optical thickness, cloud particle effective radius and water vapour and ozone concentration. Moreover, the empirical relationship is based on observations of 3 ARM sites (Alaska, Southern Great Planes and Papua New Guinea) and was never tested for Cabauw. The non-linear behaviour of cloud optical thickness as a function of radiances causes an underestimation of cloud optical thicknesses derived from satellite due to sub-pixel in-homogeneities of about 30% for GCM grid boxes (Cahalan et al. 1994). There are two sources of error that are not included in the retrieval because they are difficult to measure from the ground i.e., the fractions of ice and water in the cloud and the in-homogeneities within the cloud and in cloud cover.

To reduce the effect of spatial mismatching daily mean optical thicknesses were calculated from MSG and pyranometer hourly COT values. The number of observations that were used to

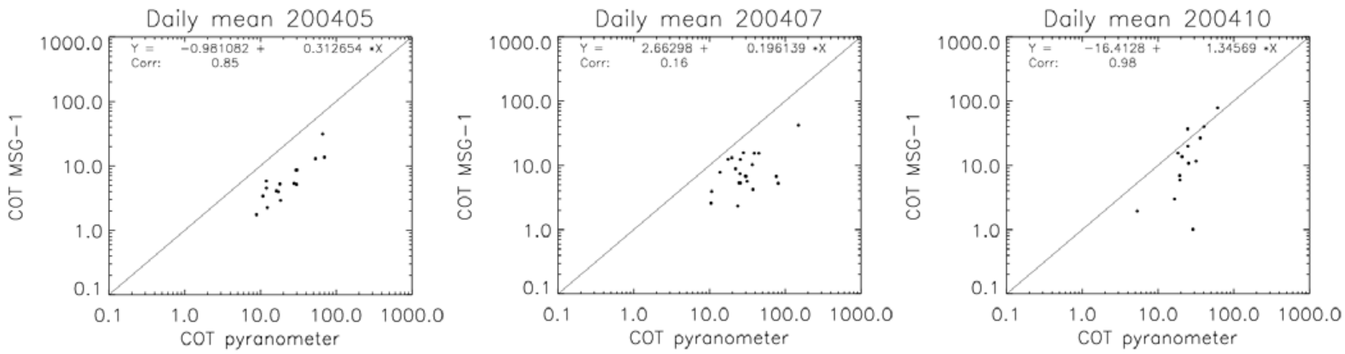

Figure 7.3 *Scatterplot of daily mean pyranometer COT plotted against MSG/SEVIRI retrieved COT for May, July and October 2004 for Cabauw, The Netherlands. The results are plotted on a logarithmic scale.*

calculate the daily mean varies between 2 and 9 hourly observations per day. Figure 7.3 presents the scatterplots of daily means of pyranometer and MSG/SEVIRI COT for May, July and October. The results suggest that the linear correlation between daily MSG/SEVIRI and pyranometer COT improved compared to the correlations of the hourly observations. The highest correlation is observed for October 2004 (Corr. = 0.98). However, for almost all data pairs the MSG/SEVIRI cloud optical thicknesses remain significantly lower than pyranometer optical thicknesses.

Table 7.1 lists the slope, offset and correlation coefficient of the linear regression between hourly and monthly mean pyranometer and MSG/SEVIRI cloud optical thicknesses. For May and October 2004 the correlations of daily mean optical thicknesses from MSG/SEVIRI and pyranometer improved compared to the hourly correlations. For July 2004 the comparison of daily means does not differ much with the comparison of hourly optical thicknesses.

Table 7.1 *Slope, offset and linear correlation coefficients of hourly and daily mean MSG/SEVIRI and pyranometer derived Cloud Optical Thickness for May, July, October and December 2004.*

	Hourly products				**Daily mean products**			
	nr data	**slope**	**offset**	**corr.**	**nr days**	**slope**	**offset**	**corr.**
May 2004	**46**	**0.13**	**3.63**	**0.58**	**16**	**0.31**	**-0.98**	**0.85**
July 2004	**70**	**0.11**	**9.45**	**0.24**	**22**	**0.20**	**2.66**	**0.16**
October 2004	**39**	**1.20**	**-9.86**	**0.76**	**14**	**1.35**	**-16.41**	**0.98**
December 2004	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

Table 7.2 presents the monthly statistics of the comparison of pyranometer and MSG/SEVIRI cloud optical thickness. The monthly statistics were calculated from the daily mean optical thicknesses. The logarithm of daily cloud optical thickness values was used to take account of the quasi-logarithmic dependence of cloud albedo on cloud optical thickness. The number of days differs from month to month due to differences in cloudiness, broken cloud cases and occurrence of clouds with COT > 250. Both MSG/SEVIRI and pyranometer have a higher mean COT for July than for May 2004. However, during the three observation months MSG/SEVIRI observes the highest monthly mean COT for October 2004, while the pyranometer retrieves for this month the lowest monthly mean COT. The 3 months data set is too short for a statistically significant validation of monthly mean MSG/SEVIRI COT retrievals. The absolute differences between the median pyranometer and MSG/SEVIRI derived cloud optical thickness are large, between -20 for July and -7 for October 2004.

Table 7.2 Monthly mean MSG/SEVIRI (MSG) and pyranometer (pyr.) derived cloud optical thicknes and median and standard deviation of the difference between MSG and pyr. retrieved daily mean COT values.

		COT MSG	COT pyr.	MSG – pyr.	MSG – pyr.
	Days	**Mean**	**Mean**	**Median of diff.**	**Std. of diff.**
May 2004	**16**	**5.6**	**21.9**	**-13.2**	**13.9**
July 2004	**22**	**7.8**	**29.4**	**-20.2**	**25.4**
October 2004	**14**	**11.0**	**23.5**	**-7.1**	**11.8**
December 2004	**-**	**-**	**-**	**-**	**-**

The MSG/SEVIRI radiances were calibrated with the pre-launch calibration coefficients provided by EUMETSAT. Govaerts and Clerici, 2004 demonstrated that the vicarious calibration of the Meteosat-8 channels is stable and shows minor drift. However, the MSG/SEVIRI radiances have not been inter-compared with other imaging satellites. For a few test cases Roebeling et al. 2004 found considerable differences between calibrated MSG-1 and NOAA-17 reflectances, ranging between ~ 4% for the 0.6 micron channel and ~20% for the 1.6 micron channel. Moreover, variations in atmospheric water vapour and satellite viewing geometry affect the level of absorption over the channel, which may result in lower top of atmosphere reflectances (Podlasly and Wollenweber, 2002). Figure 7.4 shows the errors in cloud optical thickness and droplet effective radius due to errors in 0.63 μm and 1.61 μm simulated reflectances. For clouds with τ > 60 the retrieval of optical thickness is very sensitive to errors in 0.63 μm reflectances, where reflectances errors of ±3 % in can propagate to optical thickness errors of ±30 %. The errors in effective radius are about 0.7 for errors of ±3 % in 0.63 μm reflectances. The retrieval of effective radius is more sensitive to errors in 1.61 μm reflectances, with effective radius errors between 0.8 and 1.5 μm for errors of ±3 %.

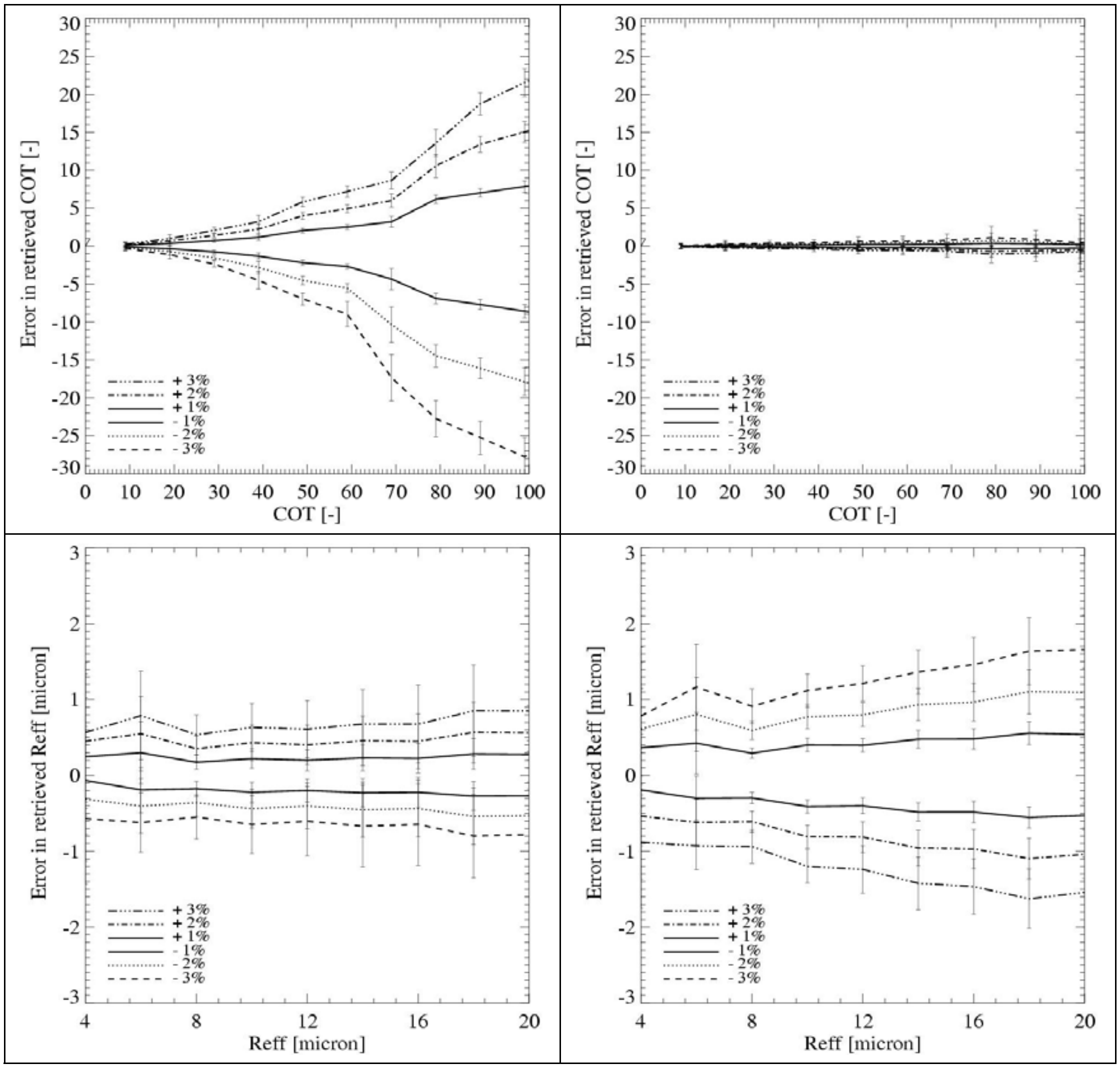

Figure 7.4. Error in retrieved cloud optical thickness (-) and droplet effective radius assuming errors of ± 1, 2 and 3% in the 0.63 µm reflectances (left panel) and 1.61µm reflectances (right panel).

7.3 Discussion

In its current state it is not possible to confirm that the MSG/SEVIRI COT product is compliant with the user's requirements. The observed difference between the median of pyranometer and MSG/SEVIRI derived COT of about 15 are rather large. The MSG/SEVIRI COT values are about 80% lower than the pyranometer COT values, which is demonstrated by the ~0.3 slope of the linear regression equation between daily means of pyranometer and MSG/SEVIRI derived COT. It needs to be mentioned that we found 15% lower optical thicknesses from the DWD initial operational dataset of ORR_V2 than the optical thicknesses from the local KNMI dataset that were used for SIVVRR_V2. Further research is needed to find the reason for these unforeseen differences.

As indicated in SIVVRR SR 1 (2005) we used the empirical relationship of Barnard and Long (2004) to relate irradiance measurements to visible cloud optical thickness. This empirical approach provides an estimate of optical thickness as a function of the ratio between diffuse and clear-sky irradiance. Variables such as solar zenith angle, aerosol optical thickness, cloud particle effective radius and water vapour and ozone are not taken into account. Large differences in the median values of the pyranometer COT values are found if we calculate pyranometer cloud optical thicknesses using the uncertainty range of COT values of the simulations by Barnard and Long. Figure 7.5 illustrates that the median of pyranometer COT could change with about +/-7 units due to these uncertainties. We suggest determining a more accurate relationship between COT and solar irradiance by replacing the empirical approach used by Barnard and Long (2004) with Look Up Tables of COT and solar irradiance calculated with a radiative transfer model.

To determine the pyranometer COT 1-minute mean diffuse radiances were used. The influence of using longer sampling times for the pyranometer derived COT was tested and resulted in a drop of the median COT value of about 2-3 COT values. To increase the sampling period an independent method is needed to determine whether the pyranometer observations represent overcast cases. There are two sources of error that are not included in the retrieval because they are difficult to measure from the ground i.e.: the fractions of ice and water in the cloud and the in homogeneities within the cloud and in cloud cover. Finally, it needs to be mentioned that the median values of MSG/SEVIRI COT are close to the median COT values of other satellite derived COT products, for example ISCCP COT over Northern Europe gives a mean COT of about 5 over the period 1983-2000.

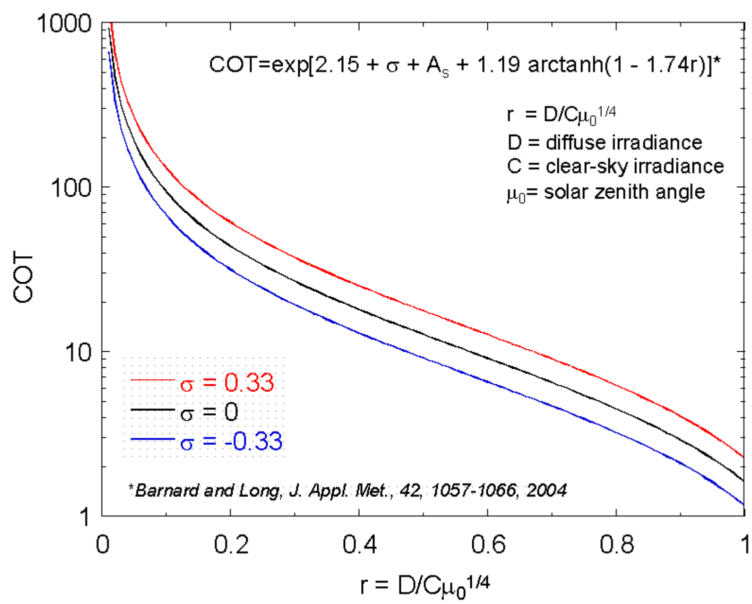

Figure 7.5 *Cloud optical thickness vs. diffuse to clear-sky irradiance ratio with uncertainty range.*

7.4 Summary of COT validation results

Results of the COT retrieval algorithm indicate that the MSG/SEVIRI cloud optical thickness values have smaller median and minimum values than the pyranometer retrieved values. The maximum frequency of differences between pyranometer and MSG/SEVIRI retrieved optical thickness is about 15, however differences up to 200 are observed. Part of the differences is caused by the selection criterion for the ground-based retrieval, which only allows completely overcast skies for the retrieval of pyranometer optical thickness. This may result in an overestimation of the pyranometer optical thickness. Other sources of error are mismatches in collocation and the sensitivity of the pyranometer retrieval algorithm to in homogeneities in cloud and atmospheric properties. To minimize collocation errors, daily averages of MSG/SEVIRI and pyranometer retrieved COT values were compared and this gave higher correlations, with a maximum of 0.98 for October 2004. This is significantly better than the correlation of the hourly observations of about 0.76 in October 2004. The current validation activity was done for completely overcast skies only; future validation studies should also include broken cloud fields.

8 VALIDATION OF THE CLOUD WATER PATH PRODUCT (CWP)

This product provides information on the Cloud Liquid Water Path (CLWP) given in gm^{-2}. The CLWP is calculated as a function of the Cloud Optical Thickness and the droplet effective radius estimate ($r_{e(1.6\ \mu m)}$). Notice, however, that in most presentations of CM-SAF products this product is abbreviated as CWP (as in the section heading) for consistency with the naming convention for other products (three letter acronyms). However, in this section we will make use of the CLWP notation to clearly indicate that it is not treating the ice water path with the same accuracy (as commented further below).

The Cloud Liquid Water Path is derived from the cloud optical thickness (τ_{vis}) and the droplet effective radius estimates (r_e). The effective radius is an intermediate product that is retrieved together with the cloud optical thickness from 0.6 and 1.6 μm cloud radiances. The cloud liquid water path is estimated with the equation of Stephens (1978):

$$CLWP = \frac{2}{3}.\tau_{vis}.r_e.\rho_l$$

where ρ_l is the density of liquid water and r_e is the droplet effective radius of water particles.

The equation given above is currently applied for both water and ice clouds. The water path estimates for ice clouds will be less reliable because the equation was originally defined for water clouds. KNMI and IFM in Kiel are currently working on an improved approach for the retrieval of ice cloud properties within a CM-SAF Visiting Scientist activity entitled “Improved retrieval of ice cloud properties”.

8.1 Validation method

The MSG/SEVIRI and microwave radiometer cloud liquid water path retrievals were compared for Chilbolton (UK) for May, July, October and December 2004. Hourly MSG/SEVIRI (Meteosat-8) images for the CM-SAF baseline area were collected from the DWD initial operational database. The MSG/SEVIRI cloud liquid water path was extracted for the satellite pixel that collocated best with the Chilbolton observation site (51.14°N, -1.44°W). The microwave radiometer liquid water path observations were averaged to 20 minutes means to obtain better correlations between ground and satellite derived liquid water path. We did not use cloud radar and lidar observations to separate ice from water clouds. Note that the microwave radiometer is insensitive to ice clouds and will not measure any liquid water path for these clouds. Although the validation is limited to satellite pixels with cloud thermodynamic phase being water, the satellite algorithm may retrieve a liquid (or ice) water path for ice clouds. Because microwave radiometer observations for thick clouds are known to be unreliable it was decided to exclude Cloud Liquid Water Path values > 750 gm^{-2} from both microwave observations and satellite retrievals. For May and July 2004 observations with rain were excluded from the validation dataset, because the microwave radiometer observations tend to become unrealistic for rain clouds. It was not possible to exclude observations with rain for October and December 2004 due to lack of rainfall data in the CLOUDNET LWP database.

8.2 Results

A statistical analysis of frequency distributions of microwave radiometer and MSG/SEVIRI cloud liquid water path is performed to determine the accuracy and precision of MSG/SEVIRI CLWP retrievals. The sizes of the datasets range between 37 observations in December and 125 observations in July 2004 which is on the low side for a statistical analysis but sufficient to calculate a monthly mean CLWP. Figure 8.1 and Figure 8.2 present the frequency distributions of MSG/SEVIRI and microwave radiometer cloud liquid water path and the frequency distributions of difference between microwave radiometer and MSG/SEVIRI CLWP. These distributions are presented for May and July 2004 (Figure 8.1) and October and December 2004 (Figure 8.2). Although the frequency distributions of microwave and MSG/SEVIRI liquid water path are not identical there are many similarities. The mean values of MSG/SEVIRI and microwave radiometer retrieved liquid water path vary between 50 and 100 gm^{-2}. From the frequency distributions it can be seen that the microwave radiometer liquid water path is often higher than the MSG/SEVIRI liquid water path. The differences are largest for May and July 2004, where the mean CLWP from MSG/SEVIRI is about 30-40% lower than microwave radiometer observed LWP. For October and December 2004 the mean CLWP from MSG/SEVIRI is respectively ~5% higher and ~10% lower than microwave radiometer LWP. The frequency distributions of differences between microwave

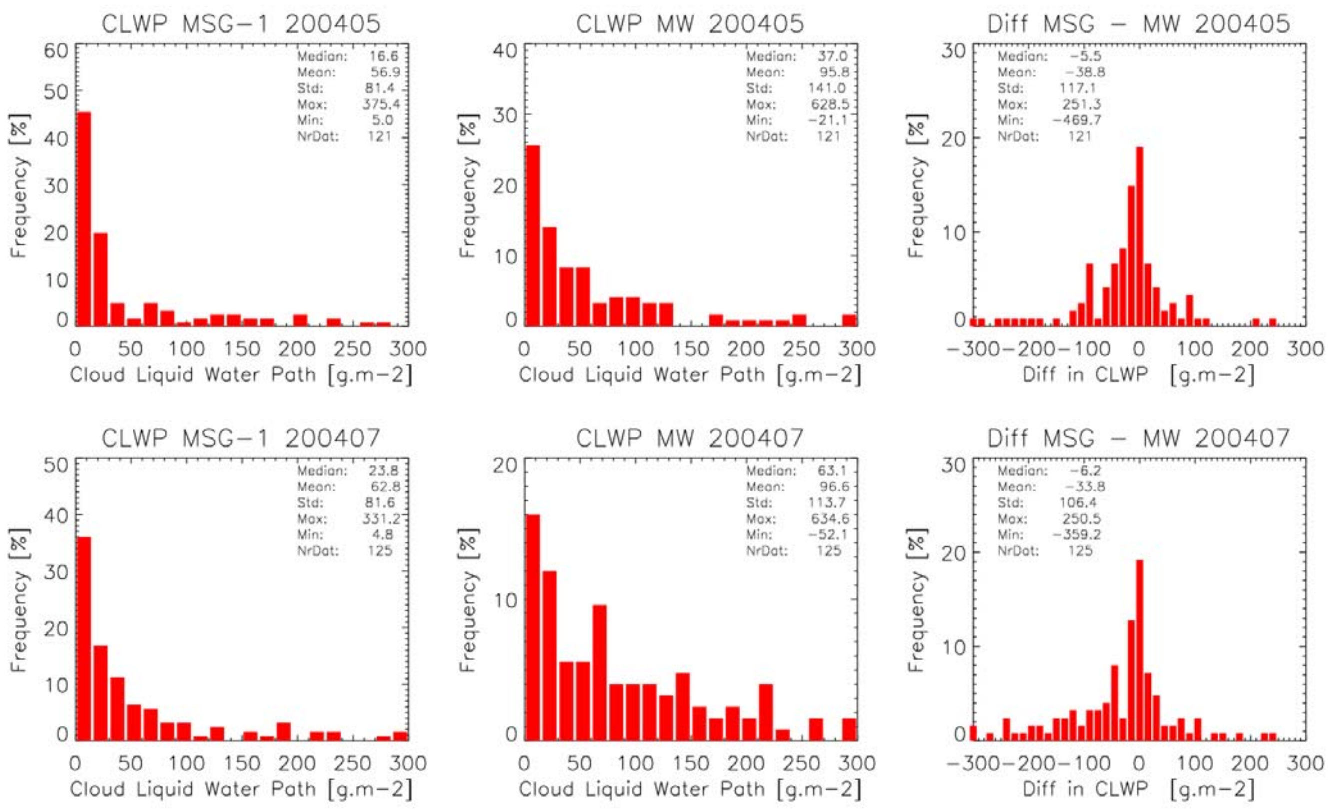

Figure 8.1 *Frequency distributions of MSG/SEVIRI (Meteosat-8) and microwave radiometer observed Cloud Liquid Water Path (CLWP) for May and July 2004 for Chilbolton, UK. The left panel present the frequency distributions of MSG/SEVIRI CLWP, the centre panels the frequency distributions of microwave radiometer CLWP and the right panels the frequency distributions of difference between MSG/SEVIRI and microwave radiometer CLWP.*

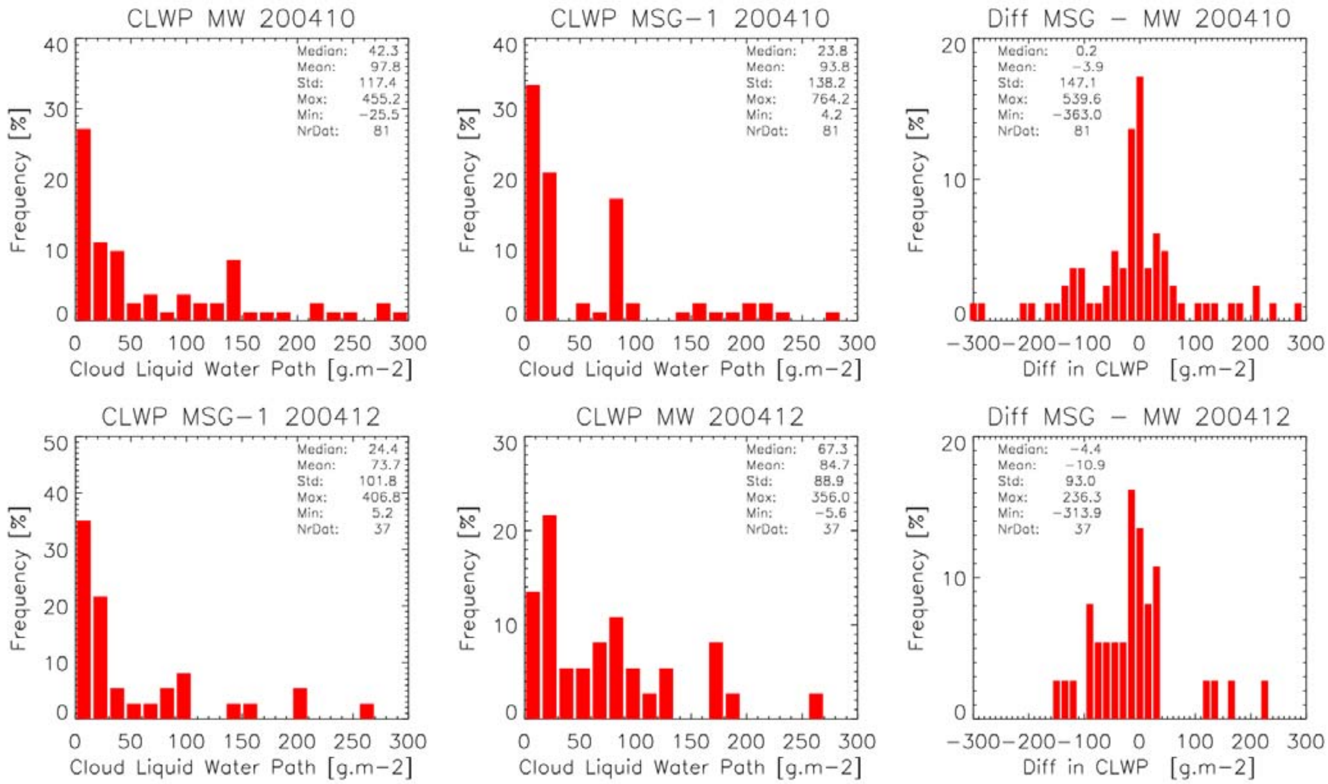

Figure 8.2 *Frequency distributions of MSG/SEVIRI and microwave radiometer observed cloud liquid water path for October and, December 2004 for Chilbolton, UK. The left panels present the frequency distributions of MSG/SEVIRI CLWP, the centre panel the frequency distributions of microwave radiometer CLWP and the right panel the frequency distributions of difference between MSG/SEVIRI and microwave radiometer CLWP.*

radiometer and MSG/SEVIRI CLWP are normally distributed and have a strong peak for differences between 0 and –10 gm^{-2}. For about 60% of the observations the differences range between –20 and 20 gm^{-2}.

Figure 8.3 presents the scatter plots of microwave radiometer liquid water path versus MSG/SEVIRI CLWP for the hourly observations. The low slopes of the linear regression equation indicate that microwave radiometer LWP is underestimated 1.5 to 3 times. However, the graphs clearly reveal that the data-pairs centre well around the 1:1 line. The hourly microwave radiometer LWP observations are weakly but positively related to MSG/SEVIRI CLWP, with correlations between 0.34 in October and 0.57 in May 2004. The large scatter of microwave radiometer LWP versus MSG/SEVIRI CLWP most likely originates from spatial mismatch between ground-based and satellite observations. The 20-minutes mean microwave radiometer LWP may not well represent the MSG/SEVIRI CLWP. The microwave radiometer observation represents a relatively small area overhead the instrument whereas the MSG/SEVIRI observation represents an area of ~18 km^2. This means that the satellite and the ground-based CLWP values are likely to represent somewhat different cloud scenes. However, their statistical distributions are expected to be more consistent.

Figure 8.4 shows that the effect of spatial mismatching is reduced when daily means of microwave radiometer and MSG/SEVIRI CLWP are compared instead of hourly CLWP values. The daily mean liquid water path was calculated from MSG/SEVIRI and microwave radiometer hourly CLWP values. The number of observations that was used to calculate the daily means varied between 2 and 9 hourly observations per day. The results suggest that the daily mean CLWP values from microwave radiometer and MSG/SEVIRI are reasonable and positively correlated for most months, which is an improvement compared to the correlation between the hourly CLWP retrievals. The best result is observed for July 2004 with a correlation of 0.82.

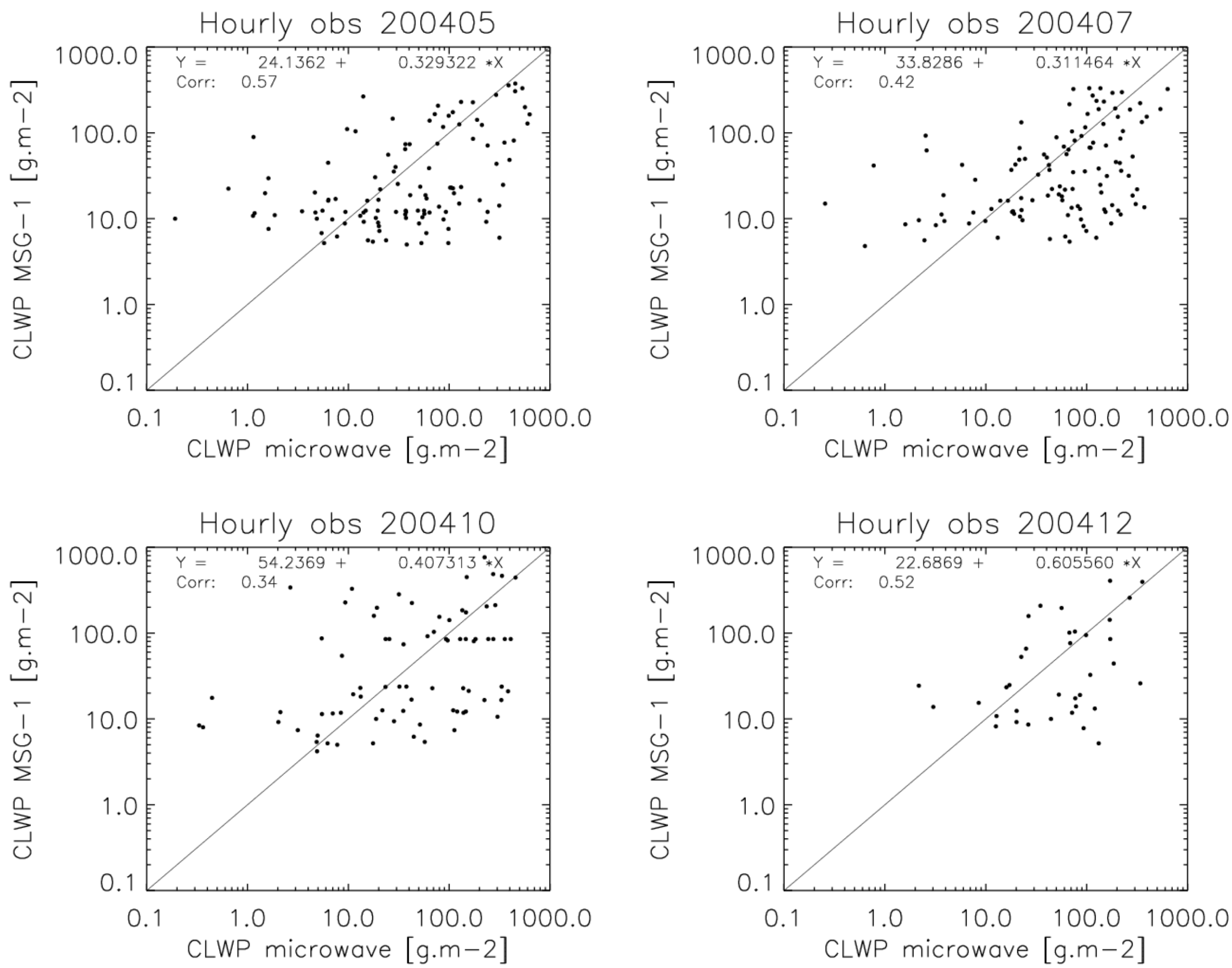

Figure 8.3 Scatterplot of hourly microwave radiometer liquid water path plotted against MSG/SEVIRI retrieved CLWP for May, July, October and December 2004 for Chilbolton, UK. The results are plotted on a logarithmic scale.

summarizes the slope, offset and correlation coefficient of the linear regression between hourly and monthly mean microwave radiometer and MSG/SEVIRI CLWP. For May, July and December 2004 the correlations between daily mean microwave radiometer and MSG/SEVIRI CLWP improved compared to the correlations between hourly CLWP values. For October 2004 the correlation of the daily means (0.04) is lower than the correlation of the hourly observations (0.34). For all months the slope of the linear regression equation of the comparison of daily means is closer to 1 than the slope of the comparison of the hourly observations. Averaged over the day MSG/SEVIRI underestimates the microwave radiometer LWP about 1.5 to 2 times. The standard deviation of the differences can be as large as 117 $g.m^{-2}$ and the correlation coefficients can vary between 0.04 and 0.82. These results reveal that further study is necessary to explain the various instantaneous differences.

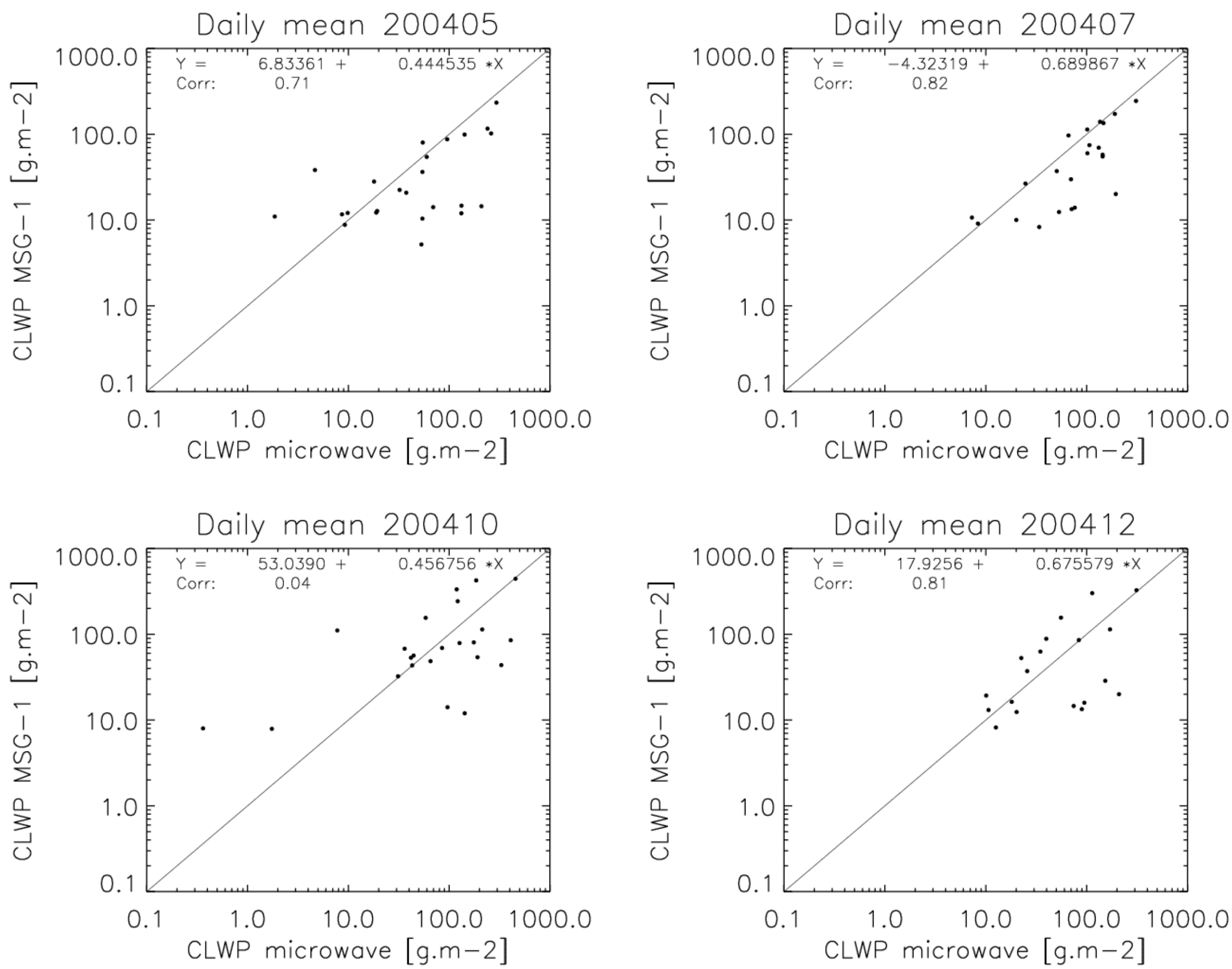

Figure 8.4 *Scatterplot of daily mean microwave radiometer Liquid Water Path (LWP) plotted against MSG/SEVIRI retrieved CLWP for May, July, October and December 2004 for Chilbolton, UK.*

Table 8.1 *Slope, offset and linear correlation coefficients of hourly and daily mean MSG/SEVIRI and microwave radiometer derived Cloud Liquid Water Path for May, July, October and December 2004.*

	Hourly products			Daily mean products		
	nr data	**offset**	**corr**	**nr days**	**offset**	**corr**
May 2004	**121**	**24.14**	**0.57**	**121**	**6.83**	**0.71**
July 2004	**125**	**33.83**	**0.42**	**125**	**-4.32**	**0.82**
October 2004	**81**	**34.24**	**0.34**	**81**	**53.04**	**0.04**
December 2004	**37**	**22.68**	**0.52**	**37**	**17.93**	**0.81**

Table 8.2 lists the statistics of the monthly comparison between microwave radiometer and MSG/SEVIRI retrieved liquid water path. The monthly statistics were calculated from the daily mean cloud liquid water path values. The number of days that was used for the calculation of the monthly statistics is indicated in the table. The monthly mean CLWP from microwave radiometer varies during the validation months between ~80 gm^{-2} in December and ~130 gm^{-2} in October 2004. Whereas the monthly mean CLWP from MSG/SEVIRI varies between mean values of ~44 gm^{-2} in May and ~112 gm^{-2} in October 2004. Both microwave radiometer and MSG/SEVIRI observed the highest CLWP in October 2004. The median of differences between daily MSG/SEVIRI and microwave radiometer CLWP values varies between ~2 gm^{-2} and ~ -25 gm^{-2}. This is a very acceptable result considering the dynamic variation in CLWP values and the uncertainties in the ground-based microwave radiometer retrievals of about 20 gm^{-2} (Dong et al. 2000 and Crewell and Löhnert 2003). The latter errors will likely contribute to some of the discrepancies between MSG/SEVIRI and microwave radiometer CLWP retrievals. The four month dataset is too limited to validate the monthly MSG/SEVIRI CLWP retrievals. During the four months there is no correlation between MSG/SEVIRI and microwave radiometer derived CLWP. To perform a statistically significant validation on monthly CLWP retrievals requires at least one year of monthly products and microwave radiometer monthly CLWP observations.

Table 8.2 *Monthly mean MSG/SEVIRI (MSG) and microwave radiometer (MW) cloud liquid water path and median and standard deviation of the difference between MSG and MW retrieved daily mean CLWP values.*

		CLWP MSG	**CLWP MW**	**MSG – MW**	**MSG – MW**
	nr days	**Mean**	**Mean**	**Median of diff**	**Std. of diff.**
May 2004	**24**	**44.1**	**83.9**	**-9.8**	**60.6**
July 2004	**22**	**64.2**	**99.3**	**-25.5**	**44.7**
October 2004	**23**	**112.2**	**129.5**	**0.38**	**131.9**
December 2004	**19**	**73.0**	**81.5**	**2.0**	**81.2**

8.3 Discussion

The frequency distributions of the differences clearly show that there is almost no bias between microwave and MSG/SEVIRI CLWP. The slope of the regression equation between daily means of microwave and MSG/SEVIRI CLWP is about 0.7, which indicates a 30% underestimation of CLWP by MSG/SEVIRI. However, the standard deviations of the difference are higher than 100 gm^{-2}. Therefore it is too early to conclude if the MSG/SEVIRI CLWP product is compliant with the user requirements. Part of the observed differences originate from a difference of about -15% between the cloud liquid water path values from the DWD initial operational dataset (used for ORR_V2) and from the local KNMI dataset (used for SIVVRR_V2). Further research is needed to find the reason for these unforeseen differences.

A positive bias can occur because for very thick clouds the MSG/SEVIRI CLWP retrieval is limited by a maximum value of 2/3*128*24=2050 gm^{-2}, whereas the microwave radiometer often observes much higher CLWP values. Because microwave radiometer observations for thick clouds are known to be unreliable it was decided to exclude both from microwave observations and satellite retrievals CLWP values > 750 gm^{-2}. The higher percentage of low CLWP values from MSG/SEVIRI most probably results from differences due to the size of the sampling area. It should be realized that the variations that we find in the microwave radiometer data often occur due to sub-pixel variations. In order to reduce the influence of rapid CLWP variations we sampled the CLWP values over a 20 minutes period, aiming to represent more or less the field of view of the satellite. However, the 20 minutes means of a series of point measurements has a different statistical meaning than an 18 km^2 CLWP value derived from MSG/SEVIRI. The frequency of low CLWP values from microwave can easily be manipulated by increasing the sampling period, which will result in an increased percentage of low CLWP values.

To improve the validation results a number of steps can be taken. Firstly, the validation period needs to be extended to at least 6 months and 3 CLOUDNET observation sites to get more reliable statistics. Preferably a selection should be made of cases with constant microwave CLWP values over the day and month in order to exclude sampling mismatches. These cases would provide the most reliable data for validation of MSG/SEVIRI CLWP. Since the initial operational database at DWD comprises hourly MSG cloud products, and not the 15 minutes products, the number observations that meet the selection criteria during a day or month is occasionally on the low side at the validation sites. Secondly, to find a representative sampling period requires a thorough statistical study on the relationship between ground-based point measurements and satellite based pixel observations. The proposed way forward is to simulate CLWP images and ground-observations with a 3-dimensional cloud simulator and determine a representative sampling period for the microwave radiometer CLWP observations.

8.4 Summary of CLWP validation results

Comparison of half-hourly MSG/SEVIRI CLWP values with hourly results from the microwave radiometer shows that for about 50% of the observations the differences between microwave and MSG/SEVIRI CLWP range between –20 and +20 gm^{-2}. However, microwave retrieved CLWP is often higher, as is indicated by the mean difference between microwave and MSG/SEVIRI CLWP between about –5 gm^{-2} in October and 40 gm^{-2} in May 2004. The magnitude of the mean differences is strongly influenced by the CLWP values for very thick clouds that are significantly higher for microwave radiometer than for MSG/SEVIRI.

For thick clouds the microwave radiometer observations are known to be unreliable or erratic due to precipitation. Thick clouds influence the monthly mean CLWP value significantly, even after excluding CLWP values > 750 gm^{-2} from both microwave observations and satellite retrievals and microwave radiometer observations with precipitation (only for May and July). However, thick clouds rarely occur and have little influence on the median of differences between microwave and MSG/SEVIRI CLWP that are between about 0 gm^{-2} in October and –7 gm^{-2} in July 2004. The high standard deviation of differences may be explained by collocation errors and the sensitivity of the ground-based method to variations in cloud properties. In order to minimize collocation errors, comparison of daily averages of MSG/SEVIRI and microwave radiometer retrieved CLWP values show a correlation about 0.7, which is significantly better than the correlation about 0.5 for the hourly observations.

9 SUMMARY AND CONCLUSIONS

The results of the ORR V2 validation activities are summarised for all six CM-SAF cloud products in the following four tables where each table shows results for each of the four studied months (May, July, October and December 2004).

Notice that for the CFC product results are presented for a fractional sub-pixel cloudiness of 0.75 which was shown to give the best agreement with observed cloud amounts. Furthermore, for this parameter we also present Bias and RMS errors of the monthly mean but it should be noticed that this is based on statistics for all the indivudual Synop stations used. For all other cloud parameters, we compare to single observation sites (e.g. Chilbolton and Cabauw) which means that it is impossible to estimate bias errors or RMS errors of the monthly mean (we have here only four monthly values to compare with). Consequently, we can only provide a rough estimate of the bias error in this case which is rather a plain difference (equal to the Bias error of the daily mean in most cases) between the two monthly estimates from satellite and surface observations for all the other cloud parameters.

For the CTY product, it needs to be stressed that we summarise the monthly results calculated only on days with more than 12 useful radar measurements (30 minute periods) in order to make the estimation more confident. We also have to emphasise again that the bias of the monthly means is just a simple difference between the monthly mean from the MSG/SEVIRI algorithm and the corresponding mean for the cloud radar observation. The same is true also for the remaining cloud products.

For the CTH product, results from the two stations Cabauw and Chilbolton have been summed up as weighted averages in the tables according to the number of available observations.

For the CPH product we have not calculated daily averages, mainly due to the low number of useful matches. Nevertheless, a rough estimate of the quality of the monthly means is calculated which we believe indicates the overall performance of the product.

Table 9.1 *ORR V2 validation results for the CM-SAF cloud products in **May 2004**. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.*

Product **Results for MAY 2004**	**Acronym**	**Accuracy according to SIVVRR V2 validation (quantity in parenthesis)**		
		Instantaneous	**Daily**	**Monthly**
Fractional cloud cover	CFC	(Not calculated)	Land:+5.0/15.8 % (Bias/RMS)	Land:+3.8/10.7% Sea: +12/- % (Bias/RMS)
Cloud type	CTY	72 % (correctly radar matched Low, Mid and High cloud categories)	Low: 0.9 Mid: 0.3 High: 0.8 (correlation) Low: +10/23 % Mid: +1/10 % High: -10/25 % (Bias/RMS)	Low: +7.8 % Mid: -0.8 % High: -7.2 % (Monthly mean difference)
Cloud top temperature, height, and pressure	CTT, CTH, CTP	0.67-0.69 (correlation CTH)	-629/1897 m (Bias/RMS CTH) 0.72-0.76 (correlation CTH)	-629 m (Monthly mean difference CTH)
Cloud phase	CPH	80 % (Percentage correct overall matches)	(Not calculated)	Water: +32.6 % Ice: -32.6 % (Monthly mean difference)
Cloud optical thickness	COT	0.58 (Correlation)	0.85 (Correlation) -13.2 (Median difference)	-16.3 (Montly mean difference)
Cloud water path	CWP	0.57 (Correlation)	0.71 (Correlation) -9.8 gm^{-2}) (Median difference	-39.8 gm^{-2} (Monthly mean difference)

Table 9.2 *ORR V2 validation results for the CM-SAF cloud products in* ***July 2004****. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.*

Product **Results for July 2004**	**Acronym**	**Accuracy according to SIVVRR V2 validation (quantity in parenthesis)**		
		Instantaneous	**Daily**	**Monthly**
Fractional cloud cover	CFC	(Not calculated)	Land:0.0/11.2 % (Bias/RMS)	Land:0.0/6.3% Sea: +3/- % (Bias/RMS)
Cloud type	CTY	59 % (correctly radar matched Low, Mid and High cloud categories)	Low: 0.7 Mid: 0.2 High: 0.9 (correlation) Low: +9/28 % Mid: -10/32 % High:+2/20 % (Bias/RMS)	Low: -1.8 % Mid: +2.5 % High: -0.7 % (Monthly mean difference)
Cloud top temperature, height, and pressure	CTT, CTH, CTP	0.66-0.73 (correlation CTH)	+368/1573 m (Bias/RMS CTH) 0.73-0.92 (Correlation)	+ 368 m (Monthly mean difference CTH)
Cloud phase	CPH	70 % (Percentage correct overall matches)	(Not calculated)	Water: +23.2 % Ice: -23.2 % (Monthly mean difference)
Cloud optical thickness	COT	0.24 (Correlation)	0.16 (Correlation) -20.2 (Median difference)	-21.6 (Monthly mean difference)
Cloud water path	CWP	0.42 (Correlation)	0.82 (Correlation) -25.5 gm^{-2} (Median difference)	-35.1 gm^{-2} (Monthly mean difference)

Table 9.3 *ORR V2 validation results for the CM-SAF cloud products in **October 2004**. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.*

Product Results for October 2004	Acronym	Accuracy according to SIVVRR V2 validation (quantity in parenthesis)		
		Instantaneous	**Daily**	**Monthly**
Fractional cloud cover	CFC	(Not calculated)	Land:0.0/12.5 % (Bias/RMS)	Land:0.0/6.3% Sea: +13/- % (Bias/RMS)
Cloud type	CTY	64 % (correctly radar matched Low, Mid and High cloud categories)	Low: 0.6 Mid: 0.0 High: 0.7 (correlation) Low: +2/26 % Mid: -14/30 % High:+11/28 % (Bias/RMS)	Low: -12.1 % Mid: -8.7 % High: +20.9 % (Monthly mean difference)
Cloud top temperature, height, and pressure	CTT, CTH, CTP	0.75 (correlation CTH)	-677/1977 m (Bias/RMS CTH) 0.73 (Correlation CTH)	-677 m (Monthly mean difference CTH)
Cloud phase	CPH	57 % (Percentage correct overall matches)	(Not calculated)	Water: +50.4 % Ice: -50.4 % (Monthly mean difference)
Cloud optical thickness	COT	0.76 (Correlation)	0.98 (Correlation) -7.1 (Median difference)	-12.5 (Monthly mean difference)
Cloud water path	CWP	0.34 (Correlation)	0.04 (Correlation) +0.38 gm^{-2} (Median difference)	-17.3 gm^{-2} (Monthly mean difference)

Table 9.4 *ORR V2 validation results for the CM-SAF cloud products in* ***December 2004****. Results are given for individual (instantaneous – if calculated), daily and monthly estimations.*

Product **Results for December 2004**	**Acronym**	**Accuracy according to SIVVRR V2 validation (quantity in parenthesis)**		
		Instantaneous	**Daily**	**Monthly**
Fractional cloud cover	CFC	(Not calculated)	Land:-1.3/13.8 % (Bias/RMS)	Land:-1.3/6.3% Sea: +10/- % (Bias/RMS)
Cloud type	CTY	73 % (correctly radar matched Low, Mid and High cloud categories)	Low: 0.8 Mid: 0.4 High: 0.7 (correlation) Low: +1/23 % Mid: +4/13 % High:-5/27 % (Bias/RMS)	Low: +6.3 % Mid: +4.6 % High: -10.9 % (Monthly mean difference)
Cloud top temperature, height, and pressure	CTT, CTH, CTP	0.69 (correlation CTH)	+336/1236 m (Bias/RMS CTH) 0.88 (correlation CTH)	+336 m (Monthly mean difference CTH)
Cloud phase	CPH	58 % (Percentage correct overall matches)	(Not calculated)	Water: +45.9 % Ice: -45.9 % (Monthly mean difference)
Cloud optical thickness	COT	(Not calculated – missing data!)	(Not calculated – missing data!)	(Not calculated – missing data!)
Cloud water path	CWP	0.52 (Correlation)	0.81 (Correlation) +2.0 gm^{-2} (Median difference)	-8.5 gm^{-2} (Monthly mean difference)

The following concluding remarks can be given with reference to the results shown in previous tables and in previous sub-sections with detailed descriptions for each product:

CFC:

This product appears to be the most well-tuned and reliable of the CM-SAF cloud products, probably as a result of many years of product development and validation within the SAFNWC project. Monthly mean values agree very well with corresponding values based on surface observations, especially for the studied summer and autumn months. However, some differences in performance are seen over land and ocean surfaces and also over the seasons. This leads to quite large quality differences over land and ocean for the winter season when cloud amounts are overestimated by approximately 10 % over oceanic areas while there is a small but noticeable underestimation of cloud cover over land. This feature is common to what has previously been found for the NOAA AVHRR product (see CM-SAF UMP, 2005). Further development has to try to mitigate this problem since a complete solution is difficult to achieve. The lost contrast in satellite imagery between cloudy and cloud-free areas in the winter season is the main cause for the problem over land. Better prospects are found for the oceanic overestimation problem where various methods for sub-pixel fractional cloudiness estimation can be tried.

CTY:

This product is very difficult to objectively validate, partly because of a lack of a real firmly established definition of how to vertically separate various cloud types as a function of cloud top temperatures/pressures/heights. The approach used here is utilising assumptions made by the particular product retrieval method and may not be fully representative of other ways of defining the various cloud types. With respect to this problem, results have to be used with caution.

Based on cloud radar retrievals of dominating cloud tops, results show a reasonable agreement with fairly low deviations concerning the monthly average values. However, it is clear that the used radar data set has not offered dense enough measurements in time to really estimate the quality of a true monthly average. A large number of days have had too few measurements to allow a proper definition of the daily average and this has consequently limited the results. This is obvious if e.g. comparing the results of the daily averages with the monthly averages. However, despite the big difficulties we believe that the performance is at least reasonable. The confidence in the product is also strengthened further if studying the results for cases with only single cloud layers. Here, we find most often improvements of the fraction of correctly matched categories with 10-20 % from the figures given in the previous tables.

We are aware that the current definition of this product is limited since it basically only takes into account the vertical position of the cloud top. A complete redefinition of this cloud product which also takes into account other cloud parameters such as CPH, COT and CWP will be considered for the CDOP phase.

CTT/CTH/CTP:

The cloud top product has been validated by using cloud radar estimations of cloud top heights and results are only shown for the cloud top height product (CTH) assuming that results should be similar for the other two product representations.

Mainly as an effect of the many kind of problems of getting a representative way of comparing measurements and satellite observations, results show a considerable scatter when looking at individual cases. This is clear from the quite high RMS errors. However, it is encouraging with respect to the climate monitoring task to find a quite low bias for the monthly mean value (varying between -677 m and +368 m). It also agrees quite well with the referenced deeper SAFNWC Visiting Scientist study suggesting an overall bias of cloud top heights of between -920 m and + 60 m.

A serious remaining problem here is the treatment of the thinnest and highest cloud layers, especially in the cases when thin Cirrus layers are superposed over optically thick water clouds. We suspect that these cases explain the overall tendency for a slight underestimation of cloud top heights. Continued development has to focus on getting a better treatment of multi-layer cloud situations and to investigate improved methods for thin Cirrus detection.

CPH:

The cloud phase product has been validated using a combined cloud radar/lidar method. Results show encouraging skills for water clouds but a significantly lower skill for ice clouds. It shows up in the monthly mean product as a general overestimation of clouds with the liquid water phase (about 20-30 % in spring and summer and up to 50 % for other seasons). This particular ice cloud problem is with a high probability explained by similar principles/conditions as the previously mentioned problem of a correct cloud top determination in case of thin Cirrus layers superposed over optically thick water clouds. The problem seems to be highly depending on the seasons (sun elevation) since the error tends to become larger in the winter season. However, the low number of useful observations during these months makes the results uncertain.

It is clear that a method for handling the cloud water phase problem also during night conditions would be highly desirable. Such methods have been suggested based on a multi-spectral use of infrared channels and they will be considered in the future CM-SAF work.

COT:

The optical thickness product has been validated using an empirical relationship between pyranometer broadband irradiance measurements and cloud optical depth. Results show a significant negative bias of about 10-20 COT units (i.e., average ground based values are above 20 while satellite estimations are below 10). It is, however, premature to conclude whether this is a truly significant feature or whether considerable deficiencies exist, for example in the used calibration method for the SEVIRI 1.6 micron channel. Other satellite-based methods show the same feature (i.e., much lower values than for ground based estimations) and there is an ongoing discussion about the causes of the differences.

CWP:

The cloud (liquid) water path product has been validated using microwave radiometer measurements. Just as for the COT product, results indicate a significant underestimation of the satellite-based monthly average (-10 to -40 gm^{-2}). However, it is clear that there is a large influence on the average results from few large measured CWP values since median difference values show not always negative biases but slightly positive values for some months. Also, results are partly compensated by the influence of the effective radius parameter which means that the negative biases are not as significant as for COT.

10 REFERENCES

Barnard, J.C. and C.N. Long, 2004: A simple empirical equation to calculate cloud optical thickness using shortwave broadband measurements, *J. Appl. Meteor.,* **43***,* 1057-1066.

Baum, B.A., P.F. Soulen, K.I. Strabala, M.D. King, S.A. Ackerman, W.P. Menzel and P. Yang, 2000: Remote sensing of cloud properties using MODIS airborne simulator imagery during SUCCESS: 2. Cloud thermodynamic phase, *J. Geophys. Res.,* **105***,* 11781-11792.

Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider, 1994: The albedo of fractal stratocumulus clouds. J. Atmos. Sci., 51, 2434-2455.

CM-SAF UMP, 2005: User Manual of Products of the Climate Monitoring SAF (included in the ORR V2 documentation).

Crewell S., and U. Löhnert, 2003: Accuracy of cloud liquid water path from ground-based microwave radiometry 2. Sensor accuracy and synergy, Radio Sci., 38 (3), 8042.

Daloze, J.-F. and M. Haeffelin, 2005: Validation of SAFNWC/MSG Cloud Top Height using ground-based lidar and radar measurements, SAFNWC Visiting Scientist Report.

Deneke, 2002: Influence of clouds on the solar radiation budget, *KNMI Scientific Report, WR-2002-09*.

Derrien, M. and H. LeGleau, 2003: SAFNWC/MSG SEVIRI cloud products, *Proc. 2003 EUMETSAT Meterological Satellite Conference, Weimar, Germany, 29 Sep – 3 Oct. 2003,* EUM P 39, 191-198.

Derrien, M. and Le Gleau, H., 2005: MSG/SEVIRI cloud mask and type from SAFNWC; accepted by Int. J. Remote Sens.

Dong X., P. Minnis, T. P. Ackerman, E. E. Clothiaux, G. G. Mace, C. N. Long, and J. C. Liljegren, 2000: A 25-month Database of Stratus Cloud Properties Generated from Ground-based Measurements at the ARM SGP Site. J. Geophys. Res. 105, 4529-4538.

Donovan, D.P., H. Bloemink, and A.C.A.P. van Lammeren, 1998: Analysis of ERM synergy by use of CLARA observations, *Final Report, ESA Contract No. 12953/98/NL/6D*.

Govaerts Y. and Clerici M., 2004, MSG-1/SEVERI Solar Channels Calibration Commisioning Activity Report, EUM/MSG/TEN/04/0024

Hogan, R. J., A. J. Illingworth, J. P. V. Poiares Baptista and E. J. O'Connor, 2003: Characteristics of mixed-phase clouds, Part II: A climatology from ground-based lidar. Quart. J. Roy. Meteorol. Soc., 129, 2117-2134.

Jolivet D., A.J. Feijt and P. Watts, 2000: Requirements for synergetic use of the ERM imager, *ESA Contract RFQ/3-9439/98/NL/GD*.

Nakajima T. and M.D. King, 1990: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements, part 1: Theory, *J. Atm. Sci.,* **47***,* 1878-1893.

Nakajima T. Y. and T. Nakajima, 1995: Determination of cloud microphysical properties from NOAA AVHRR measurements for FIRE and ASTEX regions, *J. Atm. Sci,* **52***,* 4043 – 4059.

Podlasly,C. and Wollenweber G., 2002: Correction of the surface reflectance with BDRF to improve cloud detection; Proceedings EUMETSAT satellite data user's conference, Dublin, 2002.

Roebeling R.A. and A.J. Feijt, 2004, Comparison of meteosat-8 and NOAA-17 based cloud microphysical properties, Proceedings EUMETSAT data users conference, Prague, Czech, 31 May – 4 June 2004

Roebeling R.A., H. Hauschildt, D. Jolivet, E.A. van Meijgaard and A.J. Feijt, 2003: Retrieval and validation of METEOSAT-8 and AVHRR based cloud physical parameters in the CM-SAF. *Proceedings EUMETSAT Data Users Conference, 29 September – 3 October, 2003, Weimar, Germany*.

Rossow, W.B. and R.A. Schiffer, 1999: Advances in Understanding Clouds from ISCCP, Bull. Amer. Meteor. Soc, 80, 2261-2287.

SAFNWC SUM/1, 2004: Scientific User Manual for the PGE01-02-03 of the SAFNWC/MSG/SEVIRI

SIVVRR SR 1, 2005: Scientific validation report for SIVVRR V2 for the CM-SAF cloud products: Initial validation of CM-SAF cloud products using MSG/SEVIRI data, SAF/CM/DWD/SMHI/KNMI/SR/CLOUDS/1.

Stephens G. L., G. W. Paltridge and C. M. R. Platt, 1978: Radiation profiles in extended water clouds. III. Observations. J. Atm. Sci, 35, 2133-2141.

Watts, P. D., C. T. Mutlow, A. J. Baran, and A. M. Zavody, 1998: Study on Cloud Properties derived from Meteosat Second Generation Observations. Final Report, EUMETSAT ITT no. 97/181.