Workshop Handouts

Cloud Retrieval Evaluation Workshop-4 (CREW-4)

Image: Courtesy and Copyright by Alexander Riehn, Hotel am Badersee, Grainau

4-7 March 2014, Grainau, Germany, Europe

Organized by Deutscher Wetterdienst. Financially supported by EUMETSAT, the European Space Agency through the ESA Cloud CCI project, the CM SAF project, and Deutscher Wetterdienst

Program Committee

Bryan Baum (co-chair), Rob Roebeling (co-chair), Ralf Bennartz, Ulrich Hamann, Andrew Heidinger, Jan Fokke Meirink, Martin Stengel, Andi Walther, Phil Watts, and Anke Thoss

1   Table of Contents

2   Introduction

Clouds strongly modulate the energy balance of Earth and its atmosphere through their interaction with solar and thermal radiation (Cess et al. 1989). However, because cloud properties vary on time scales of seconds to days, and also spatially on scales from meters to thousands of kilometers, clouds have to be represented in a simplified way in climate and weather forecast models. The formation of clouds as well as their direct and indirect feedbacks to the climate system contributes largely to the uncertainty in climate predictions (Solomon et at. 2007). Measurements of the global distributions of these properties and their diurnal, seasonal, and inter annual variations are critical for improving our understanding of the role of clouds in the weather and climate systems.

The overarching objectives of CREW are to bring together scientists working on cloud retrievals so as to identify and address research questions related to cloud parameter retrievals; to enhance communication; to develop international partnerships; to provide a comparison and validation platform; and finally to provide retrieval verification and validation statistics. The investigations of CREW focus primarily on Level 2 products and are therefore a complementary effort to GSICS (Global Space-Based Inter-Calibration System) with the focus on Level 1 data and the GEWEX Cloud Assessment (Stubenrauch et al., 2013) comparing average Level 3 cloud properties on climatological scales. In the GEWEX it found that differences of Level 3 products could be primarily attributed to different sensor sensitivities. But it found that deviations of the retrieval datasets may also be caused by differences in Level 2 retrieval methods and Level 2 to Level 3 aggregation methods. Hence, one recommendation of the GEWEX Cloud Assessment is to investigate these issues in more detail being the primary goal of the CREW project.

A common Level 2 retrieval database was built to determine the strengths and weaknesses of the state-of-art cloud retrieval algorithms. The CREW database consists of cloud property retrievals from different algorithms for passive imagers (SEVIRI, MODIS, AVHRR, POLDER, and/or AIRS), complemented with cloud measurements that serve as a reference (CALIOP, CPR, AMSU, MISR) for a number of "golden days." Sixteen scientific institutes from Europe and the USA contributed data among others the EUMETSAT central facilities, the Nowcasting SAF and the Climate Monitoring SAF. Thus, CREW is an inter-comparison with the most participating institutions since the pre-ISCCP algorithm inter-comparisons (Rossow et al., 1985). The datasets were used to evaluate the retrieval quality of cloud optical, micro- and macro physical properties for different cloud and observation conditions. The success of this project can be attributed to the inter-comparison strategy of the algorithms. The idea is to have an independent, objective and consistent inter-comparison of a large number of state of the art retrieval algorithms. The direct comparison with independent validation datasets help to understand the potentials and limitations of the cloud retrievals with passive imagers, allows quantifying the accuracy of the retrieval products and in this way provides a path toward optimizing these retrievals for both climate monitoring research and climate and weather model analysis. Besides focusing on instantaneous cloud parameter retrievals, CREW also seeks to observe and understand methods that are used to prepare daily and monthly cloud parameter climatologies. Finally, particular attention is given to increase the traceability and uniformity of different long-term and homogeneous records of cloud parameters.

Another important component of the workshop is the discussion on the results of the algorithm and sensor comparisons and validation studies. CREW provides a forum for international satellite-based cloud retrieval teams to share their experience with state-of-the-art cloud parameter retrievals from passive imaging satellite observations. Initially the collaboration was established at the EUMETSAT funded Cloud Workshops held in Norrköping, Sweden. 25 scientists participated in this workshop. The second workshop was held in Locarno, Switzerland, in 2009, and had about 45 participants. The third CREW in Madison, Wisconsin, USA had about 70 participants from universities, research institutes, and satellite agencies in Europe and the United States. In March 2014 the 4th Cloud Retrieval Evaluation Workshop will be held in Grainau, Germany. The topics of the workshop are retrieval and evaluation methods as well as application of cloud retrievals in climate, weather and nowcasting applications.

3   Sensor Calibration

Courtesy: Jan Fokke Meirink (KNMI), Phil Watts (EUMETSAT)

List of open questions for the calibration working group (extendable)

- What calibration adjustments have been made for the provided CREW datasets?
  - SEVIRI? AVHRR? VIIRS? MODIS?
- Is the nominal/operational/pre-launch/onboard calibration accuracy OK for real-time/near-realtime applications (like Nowcasting)?
- How do we define a recalibrated time-series (FCDR)?
  - At what point are calibration corrections sufficient for permitting climate studies?
  - Does it include other requirements than just radiance adjustments (e.g. navigation, sampling, orbital drift, …..,etc)?
  - Can we ever achieve accuracies permitting climate trend detection (e.g. according to requirements as described in BAMS papers by Ohring et al (2007) and Wielicki et al. 2013)?
- What calibration accuracy do we require for the VIS, NIR, SWIR and LWIR channels?
- What calibration approach meets the requirements of the cloud community best?
- What role can GSICS play (i.e., is there a need for additional or complementary calibration monitoring efforts by GSICS or others)?
- What are the main improvements concerning calibration in MODIS Collection 6 datasets?
- Are there calibration issues regarding data from CloudSat and CALIPSO that we need to consider when using them as a reference?
- What are the plans for updating the calibration of historic AVHRR data?
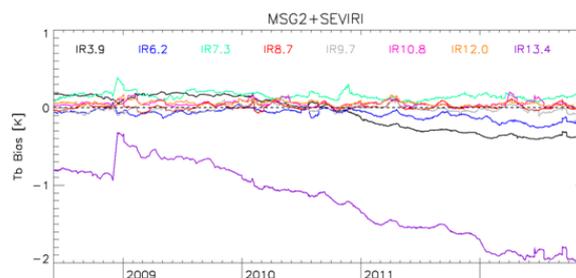 (Inform about SCOPE-CM AVHRR FCDR project!)

For a proper Level 2 retrieval it is essential to pay careful attention to the correct calibration of the measured radiances (Level 1). Very comprehensive advice for calibrations are given by GSICS being an international collaborative effort initiated in 2005 by WMO and the CGMS to monitor, improve and harmonize the quality of observations from operational weather and environmental satellites of the Global Observing System (GOS). GSICS aims at ensuring consistent accuracy among space-based observations worldwide for climate monitoring, weather forecasting, and environmental applications.

Figure 1: Example for GSCIS inter-calibration of MSG2-SEVIRI with GEO-LEO IR inter-calibration. Pay special attention to the 13.4 μm channel.

Thermal sensors on SEVIRI are calibrated onboard while this is not done for the solar ones. Some groups made suggestions to increase the radiance in the VIS08 channel by 8% and to decrease the radiance of the NIR 1.6 by 2% compared to the operational EUMETSAT calibration. Figure 2 shows results of an inter-calibration study with MODIS suggesting such calibration offsets.
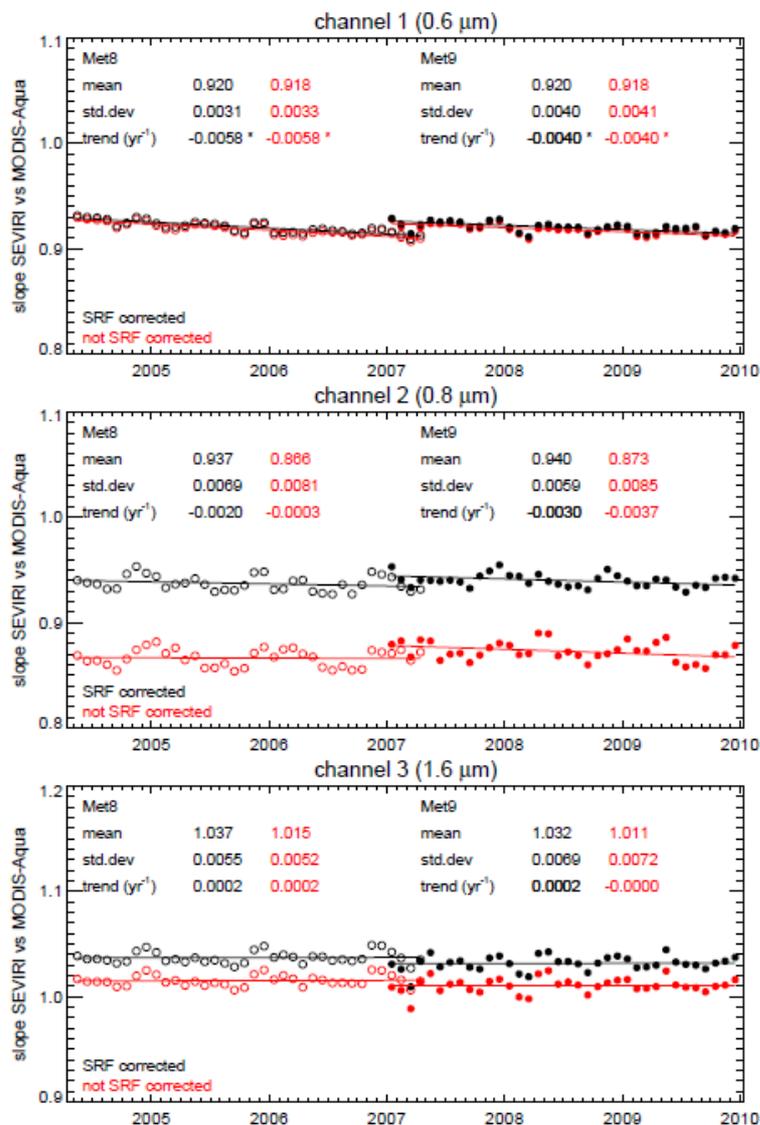
Figure 2: Time series of inter-calibration slopes for SEVIRI-Meteosat against MODIS-Aqua for three solar channels. The open circles are the monthly slopes for Meteosat-8, while the filled circles are for Meteosat-9. The solid lines are linear fits through those monthly slopes. Black symbols and lines correspond to data that have been corrected for SRF differences, while red symbols and lines correspond to data that have not been corrected for SRF differences. Mean, standard deviation (relative to the trend line), and trend of the inter-calibration slopes are indicated in the plots. Trends that are significant at the 95% level are marked by an asterisk. Figure taken from Meirink et al. (2013).

Figure 3 shows the sensitivity of derived cloud properties, namely the cloud optical thickness (COT), effective radius (Reff), cloud top pressure (Pc) and the liquid water path to this calibration change (LWP) to the above mentioned modification of the MSG calibration.

Figure 3: The effect of the a change of 8% of the VIS08 and -2% of the NIR16 channel of the MSG/SEVIRI measurements on Cloud Optical Thickness (COT), 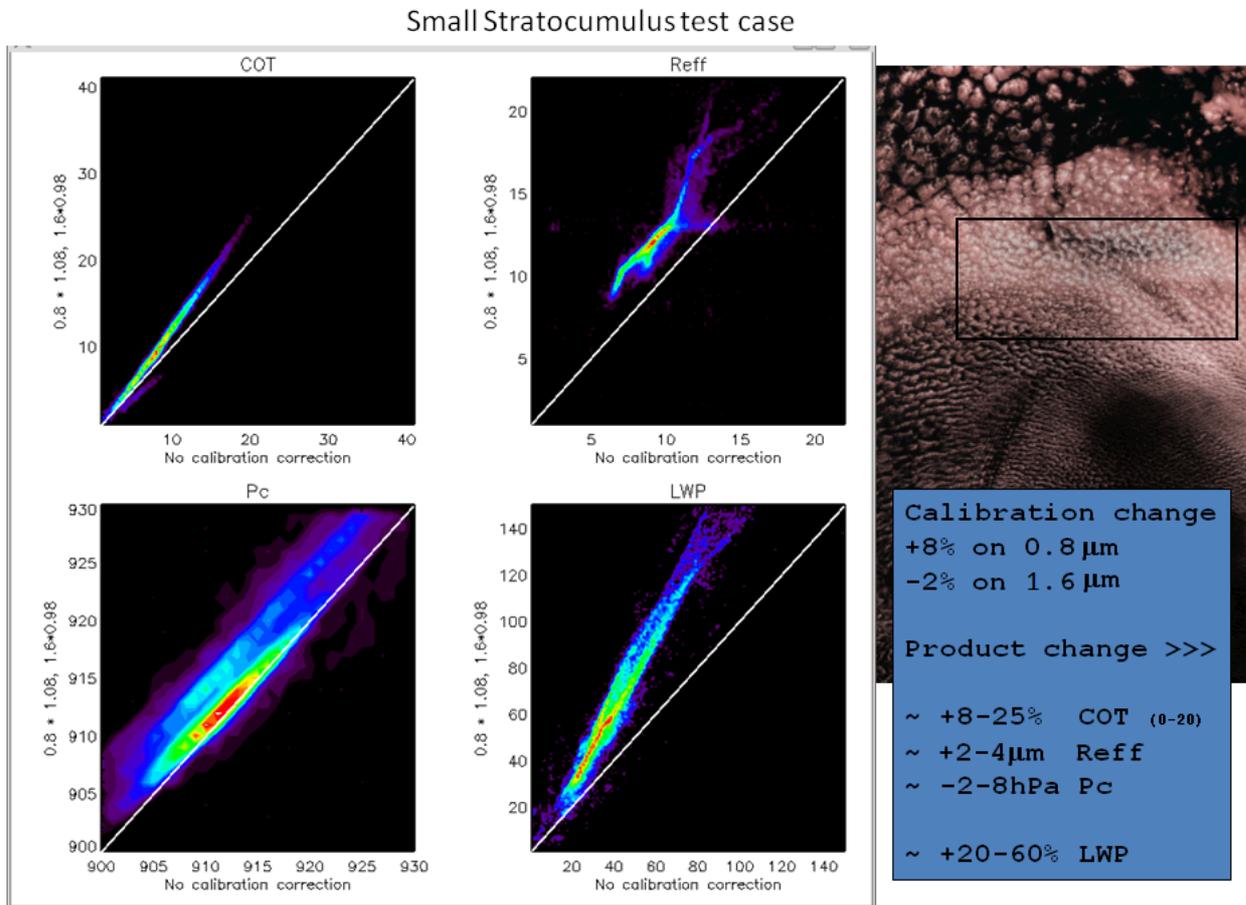effective radius (Reff), cloud top pressure (Pc), and liquid water path (LWP). The sensitivity was calculated for a marine stratocumulus area, shown on the left side.

4 Level-2: Assessment of Cloud Parameter Retrievals

4.1 Inter-comparison of SEVIRI products

Courtesy: Ulrich Hamann (KNMI, MeteoSwiss), Andi Walther (SSEC), and the CREW team

List of open questions for the cloud retrieval working groups (extendable)
- What ice crystal models are best suited?
- Do multi layer needs to be modelled separately, and if so how modelled?
- How to deal with sub-pixel cloudiness?
- Do we need vertically inhomogeneous clouds?
- How to deal with three-dimensional cloud radiative effects?
- Use of prior information in retrievals (e.g. optimal estimation)?
- How sensitive is the method to the ancillary data used?
- Are more channels always better than fewer channels?
- Should the infrared and optical cloud properties be consistent?
- Is the method robust to multiple solutions?
- Is the method able to use the full spectral information?
- Does the method use temporal and spatial information?
- Is the method able to use more complex cloud models?
- Does the method have the ability to estimate the retrieval uncertainty?
- Does the method have a model versus reality consistency check?
- How to detect cloud over snow and ice (day/night)?
- How to derive cloud physical and optical properties over bright surfaces?
- How can the cloud retrieval community go towards standard definitions (e.g. for cloud mask, cloud top height, cloud phase)?

4.1.1 Cloud mask

Twelve groups provided cloud mask data for the SEVIRI disk. The data was transformed into a binary cloud mask with values 0 for no observation or space, 1 for cloud and 2 for cloud free. Figure 4 shows cloud masks of the 12 groups for the noon scene of 13th of June 2008. Cloudy pixels are indicated as bright areas, cloud-free areas are blue for ocean and green for land surfaces. The OCA algorithm does not retrieve an own product, but applies the cloud mask obtained by the MPEF algorithm. All algorithms catch the same distribution of cloud on the MSG disk with most cloud in the tropics and in the west wind regions. There is a distinct difference in the total cloud amount for this scene ranging from 41 percent (FUB) to 61 percent (MFR).
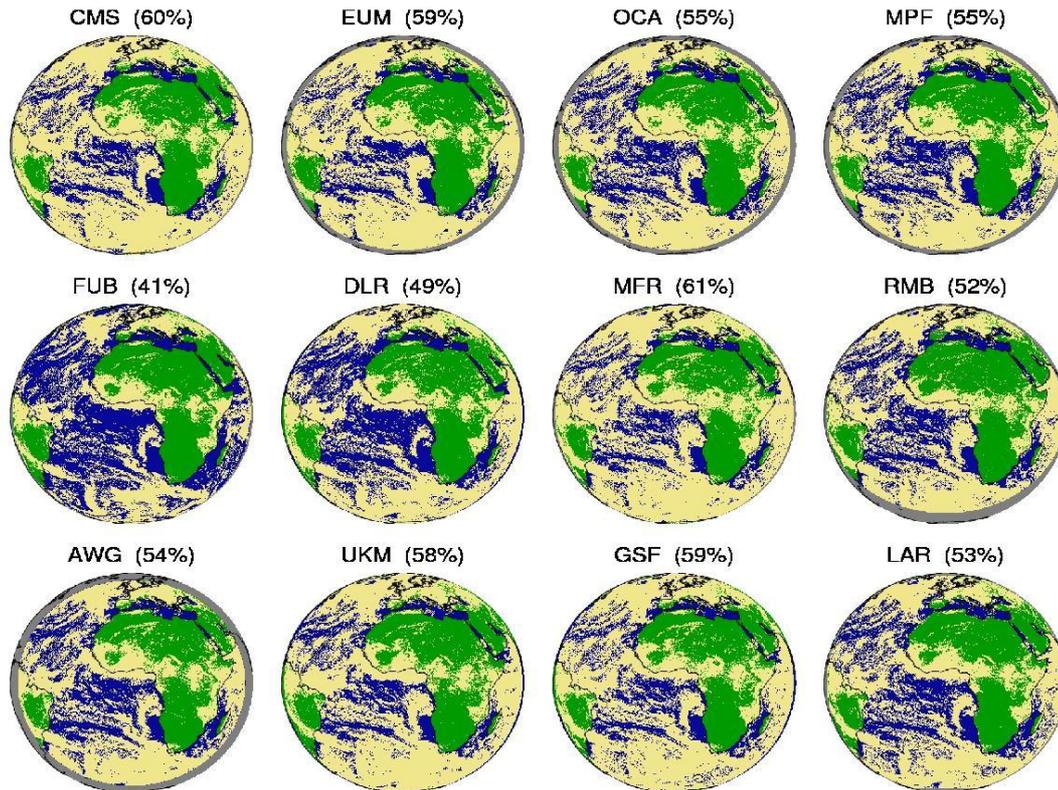

Figure 4: Cloud masks of the twelve CREW algorithms for 13-06-2008 at 12:00 UTC.

Figure 5 shows the 'number of disagreements', i.e. the number of algorithms for which the cloud mask differs from the majority of algorithms. Areas with a high number of disagreements need to be further investigated. The deviations over Northern Africa are probably caused by different detection thresholds for thin cirrus clouds. There might also be biomass burning aerosol that is classified as clouds by some algorithms. A second area of disagreements - the southern part of the Arabian Peninsula and the adjacent sea - is marked with a blue ellipse. MODIS measurements show there is dust storm is this area. Some algorithms classify the dust storm as clouds, in particular over the sea. A third area in the west of Angola is marked by a red rectangle. The reason for this disagreement is likely the different interpretation of partly cloudy pixels by the different algorithms.
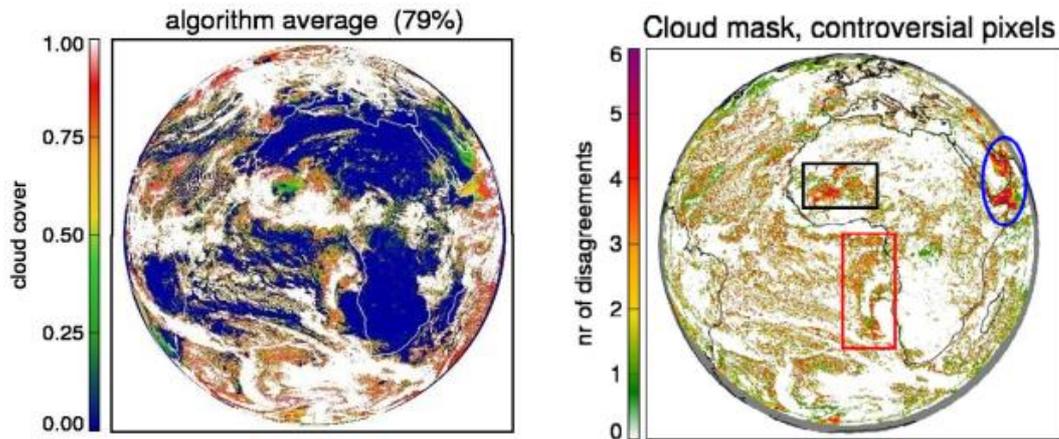

Figure 5: The left figure shows the multi algorithm average of all cloud masks for 13-06-2008 at 12:00UTC. The right figure shows the number of disagreements of the cloud detection. With 12 algorithms participating in the inter-comparison, the maximum number of disagreements is 6. The marked areas show specific problems of cloud detection like thin cirrus clouds over land (Sahara), misclassification of dust as cloud (Arabian Peninsula), and different classification of partly covered cloud pixels (Southern Atlantic).

Figure 6 shows the latitudinal mean of the binary cloud mask. The rough distribution of clouds is captured by all algorithms. During this day of the north hemispheric summer the Inter-tropical convergence zone is shifted northwards. The cloud amount has a local minimum in the subtropical regions at 20º S and 20º N. In the mid latitudes the average cloud amount is higher again. The agreement of the different SEVIRI algorithms is good, except the FUB algorithm derives a lower cloud amount compared to the other algorithms. The deviations are somewhat larger in the Southern hemisphere. For this specific day in the Northern Hemispheric summer, the solar zenith angle is larger for the Southern hemisphere, making the cloud detection more challenging.
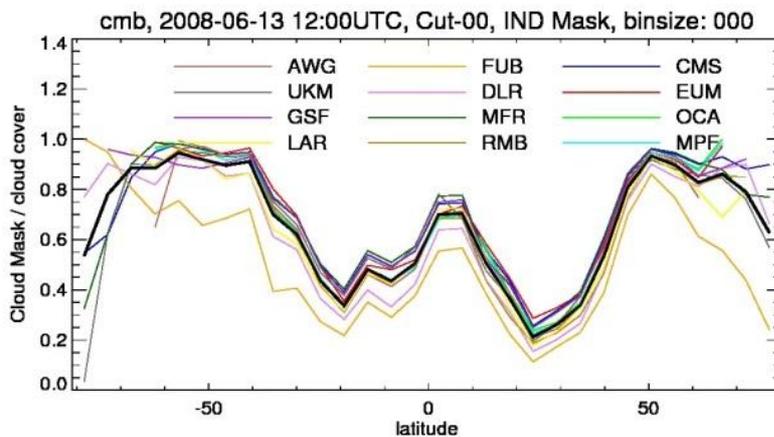

Figure 6: The latitudinal mean of the binary cloud mask for 13-06-2008 at 12:00UTC.

4.1.2 Cloud phase

The determination of the cloud phase is an important step in every cloud retrieval algorithm. As cloud ice crystals and water droplets have different spectral characteristics and their phase functions deviate drastically, the retrieval of the cloud optical depth and the effective radius depends critically on the cloud phase. The retrieved cloud phases of the participating algorithms are shown in Figure 7. Areas with ice tops are shown in red, water cloud in blue. The EUM, MPF, MFR and AWG algorithm indentify some pixels either as mixed or as unknown phase shown in green. The MPF algorithm shows these cases in the tropical Southern Atlantic, the MFR algorithm retrieve mixed phased primarily at the outer boundary of cloudy areas and the AWG algorithm retrieves mixed phase clouds at the most Northern and Southern part of the SEVIRI disk. The ratio satellite pixels identified as water clouds compared to the total number of cloudy pixels varies between 0.47 (LAR) to 0.82 (RMB).
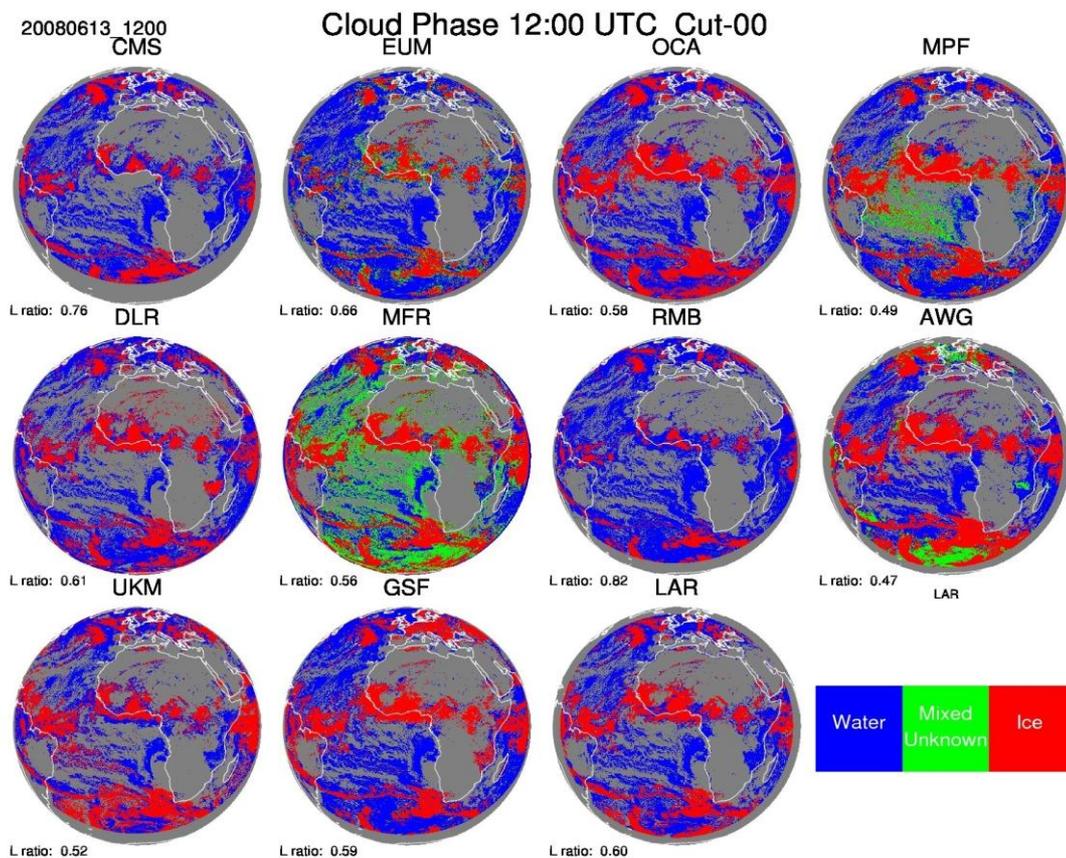
Figure 7: Cloud phase of the twelve CREW algorithms for 13-06-2008 at 12:00 UTC. Red areas shows cloud with ice top, water clouds are shown in blue and in case of unknown or mixed cloud phases in green. The ratio of water to the total number of cloudy pixels is given as L ration at the lower left of the diagram.

On the left hand side of Figure 8 the multi algorithm average of the cloud phase is shown, on the right hand side shows the latitudinal mean of the cloud phase. In the inner tropical convergence zone deep convective systems can be identified. The central areas of convective systems are identified as ice by all algorithms. On the other hand the phase identification for the outer part of the convective systems is not consistent. The inner parts of the deep convective systems usually have a high optical depth, whereas the outer parts might be thin cirrus layers or multi layer systems where a cloud phase determination is more challenging. Due to

this, the mean cloud phase differs among the algorithms substantially from 30% (RMB) to 85% (MPF, MFR, AWG). In the trade wind regions most of the clouds are classified as water clouds. Exceptions are the MPF algorithm, that identifies many clouds in the tropical South Atlantic as mixed phase, and the MFR algorithm, that identifies the borders of many clouds as mixed. Thus the latitudinal average of MPF and MFR is higher 15-20% higher than the multi algorithm average in those regions. In the mid latitude low pressure systems with the associated frontal systems can be identified. The RMB algorithm tend to detect more water cloud than the multi algorithm average, while for the AWG algorithm the ice coverage tend to show larger values than the average, especially for high latitudes.
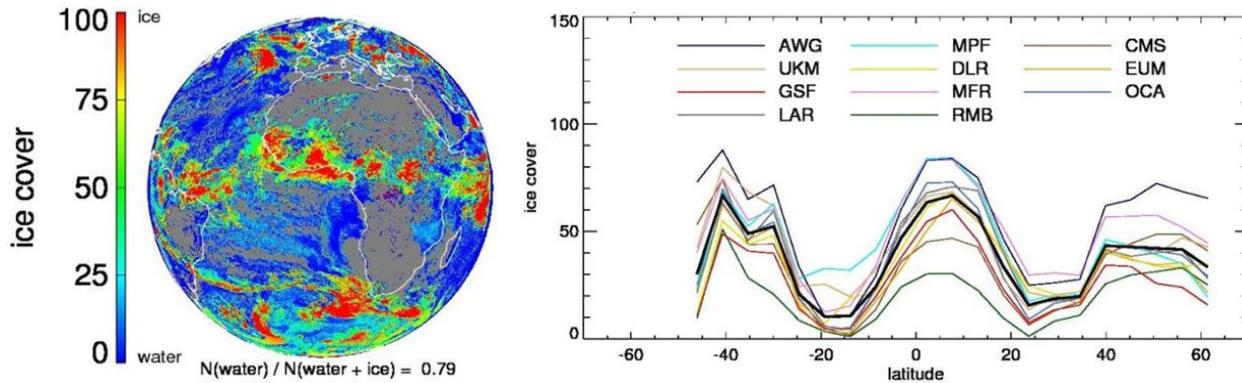
Figure 8: The left hand side shows the multi algorithm average of the cloud phase (water=0, unknown=50, ice=100). Light blue, green and yellow colors mark the areas were the multi algorithm ensemble cannot certainly determine one cloud phase. The right hand side shows the zonal average for the individual algorithms, only pixels were every algorithm detected a cloud is taken into account (common cloud mask), both figures for 13-06-2008 12 UTC.

4.1.3 Cloud top pressure

For the inter-comparison we look at the CTP, as this property is directly provided by ten algorithms, whereas CTH is provided by five algorithms only. Figure 9 shows the CTP derived by the algorithms for 13 June 2008, 13:45UTC. The zonal distribution of the CTP is comparable for all datasets. High clouds are present in the inter-tropical convergence zone (ITCZ). Adjacent to them, low clouds are most common in the marine stratocumulus region between 30º S and 30º N. In the mid latitudes synoptic systems with their frontal structures can be identified. The derived CTP means range from 577 hPa to 424 hPa. The smallest mean CTPs (the highest clouds) are retrieved by MFR (424 hPa), CMS (432 hPa) and AWG (439 hPa), the algorithms showing the largest mean CTPs (the lowest clouds) are EUM (558 hPa) and MPF (577 hPa). Averaging is performed with the logarithm of CTP and afterwards converted into a pressure again. In this way, the mean CTP is more comparable to the mean CTH. Note that the cloud masks differ between the algorithms, whereby the mean CTPs are influenced. Some algorithms also limit the domain for retrievals due to high viewing or solar zenith angles and/or sun glint.
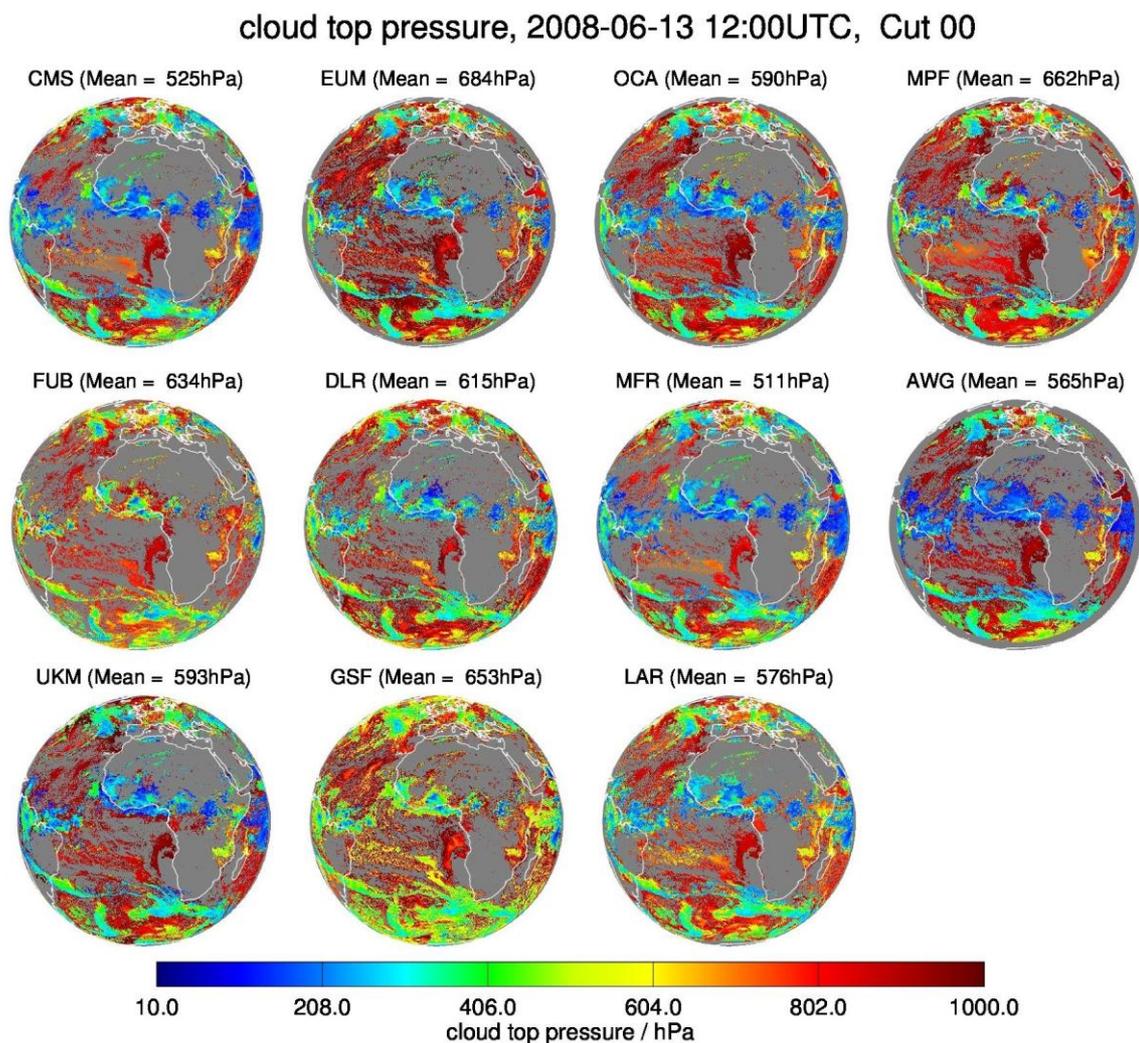

Figure 9: Cloud top pressure of the twelve CREW algorithms for 13-06-2008 at 12:00 UTC.

In Figure 10 some basic statistics of the multi algorithm ensemble are presented. In Figure 10a we show the number of algorithms that detect a cloud and provide a CTP value for the observed satellite pixel. In general, the agreement of the cloud detection among the algorithms is good, in particular for the central parts of the cloud systems. However, at the edges of the cloud systems the cloud detection results differ. The ability to detect a cloud decreases when the sub-pixel cloud fraction and the optical thickness decreases, see also Figure 10d showing the multi algorithm ensemble average of the cloud optical depth (COD). There might also be overestimations of the cloud cover by some algorithms due to misinterpretation of aerosols or cloud free scenes. In particular, false cloud detection may occur in case of large uncertainties in the surface albedo, emission and temperature. The multi algorithm average of the CTP is shown in Figure 10b. The area displayed is limited to regions where all SEVIRI retrievals detect clouds. Figure 10c shows the multi algorithm ensemble standard deviation of the logarithm of CTP. To eliminate the influence of different cloud masks, only pixels are shown for which all algorithms provide retrieved CTP values.
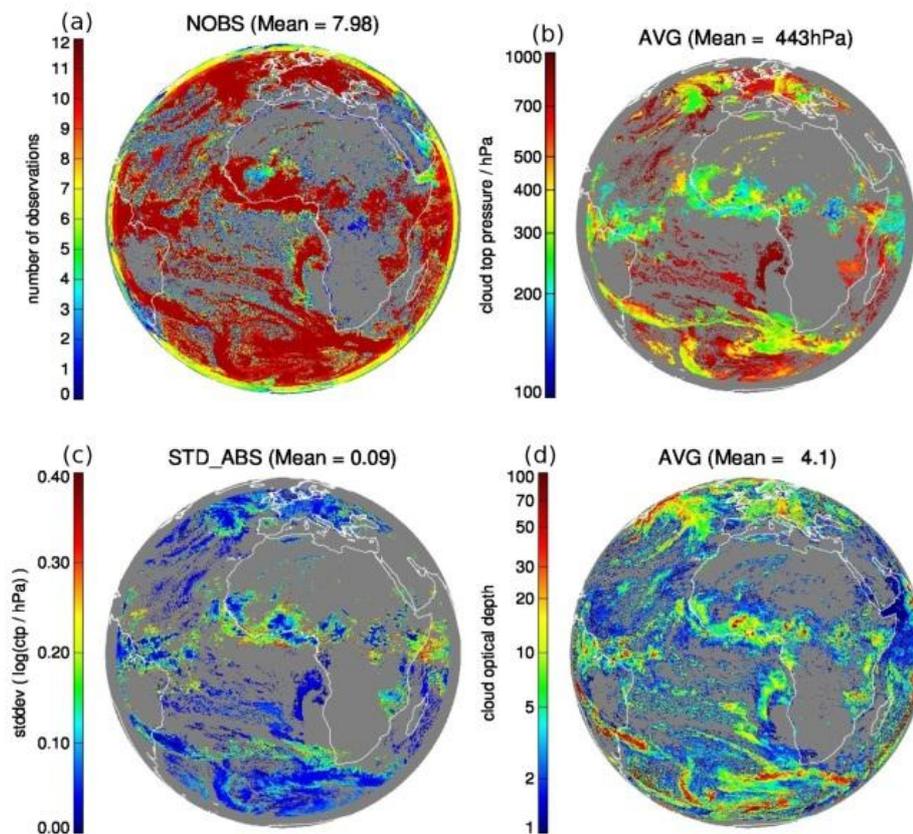
Figure 10: Multi algorithm ensemble statistics. Panel (a) displays the number of algorithms that provide a CTP value. Panel (b) shows the multi algorithm average of the cloud top pressure (CTP). Panel (c) shows the standard deviation of the log10(CTP). In Panel (b) and (c) values are shown for areas only, where all retrievals detect clouds (common mask) to eliminate effects of different sample sizes. Panel (d) shows the multi algorithm average for the cloud optical depth (not limited to the common mask). All images are for 13 June 2008, 12:00UTC.

Figure 11 shows the latitudinal mean of the cloud top pressure. The uppermost panel shows the datasets reduced to the common cloud mask. The agreement here is better compared to the second panel, where the datasets with the individual cloud masks are shown. The relative standard deviation of the cloud top pressure is about 15-20% in the extra-tropics and up to 55% in the tropics while using individual cloud masks. Reducing the data sets to the common cloud pixels, the agreement improves by some 5% in the extra-tropics and up to 15% in the tropics.

Figure 11: Latitudinal mean of the Cloud Top Pressure of eleven algorithms for 13 June 2008at 12:00 UTC. In the upper panel only satellite pixels were used, for which all eleven retrievals derive a result for the CTP, whereas the plot in the middle panel is based on all available cloudy pixels. The black line shows the average of all SEVIRI algorithms. In the lower panel, the relative standard deviation of the algorithm ensemble is shown.

Figure 12 shows the cloud top pressure averaged by the viewing zenith angle. Small CTP values in the tropics are associated with the ITCZ. For viewing zenith angles above 60 degrees we observe a slight decrease of the CTP. As the line of sight is more tilted in this region, the slanted cloud optical depth increase. Therefore clouds with thin cloud optical depth are detected more easy and the CTP is detected closer to the physical cloud top.
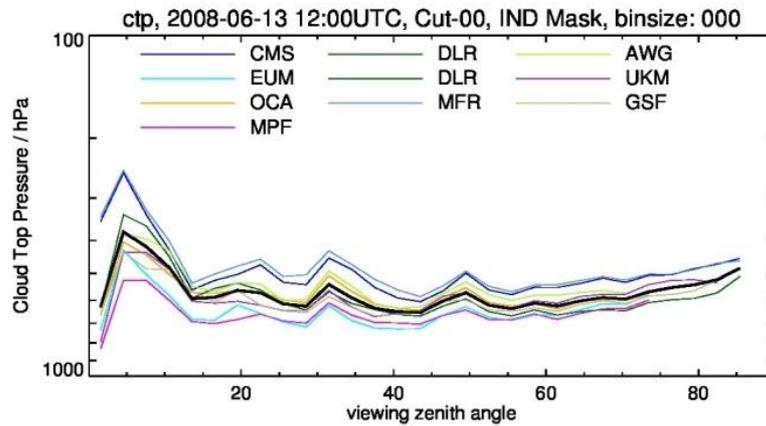
Figure 12: The dependence of the cloud top pressure on the viewing zenith angle for 13-06-2008 at 12:00UTC. The upper panel shows the diagram for the original datasets. The lower panel shows the mean for the common cloud mask.

Figure 13 shows the histograms of the cloud top pressure. On the left side datasets with their individual cloud masks are shown. We observe different cloud occurrence frequency for the boundary layer clouds. The right hand side shows the histograms for the datasets reduced to the common cloud mask. By this procedure the cloud occurrence frequency is reduced by a factor of 2 to 3. Most histograms show a distribution with two cloud occurrence maxima, one around 250hPa and a second one below 800hPa.
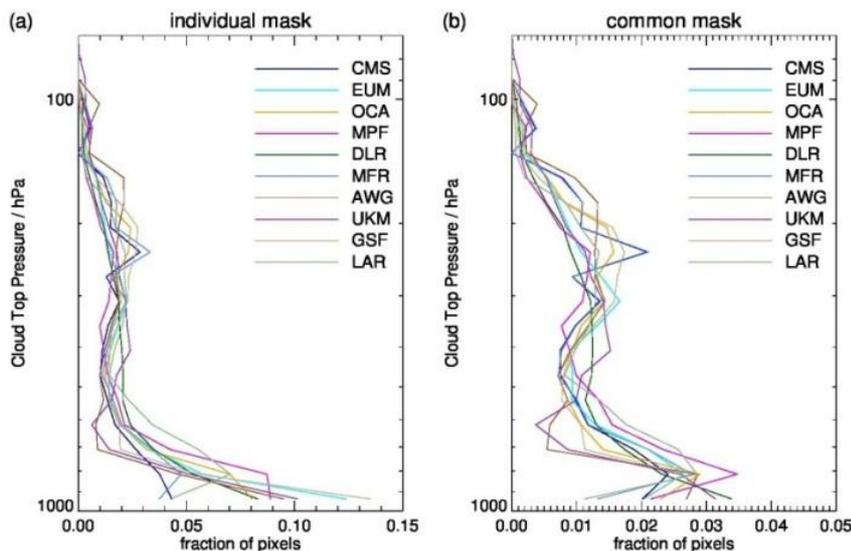
Figure 13: Histograms of the cloud top pressure for 13-06-2008 at 12:00UTC. The left panel shows the diagram for the original datasets. The right panel shows the mean for the common cloud mask.

The inter-comparison of the SEVIRI cloud top pressure algorithms and the comparison to CALIOP and CPR measurements has been submitted to AMTD (Hamann et al, 2014). At the moment the paper is in the discussion status.

4.1.4 Cloud top height
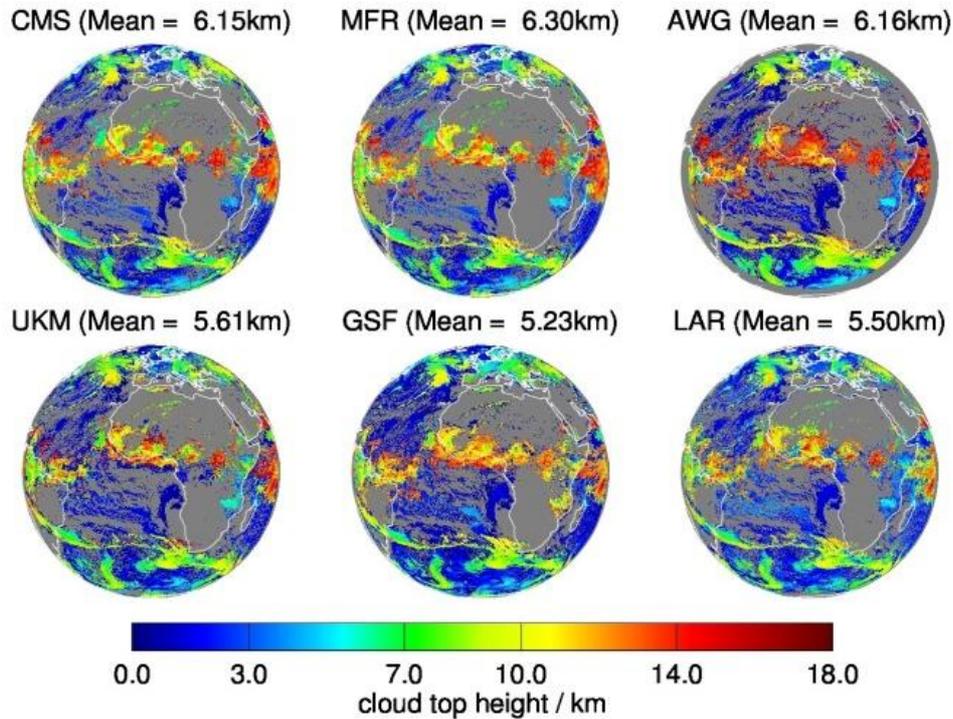
Figure 14: Cloud top height of the six SEVIRI algorithms for 13-06-2008 at 12:00 UTC.

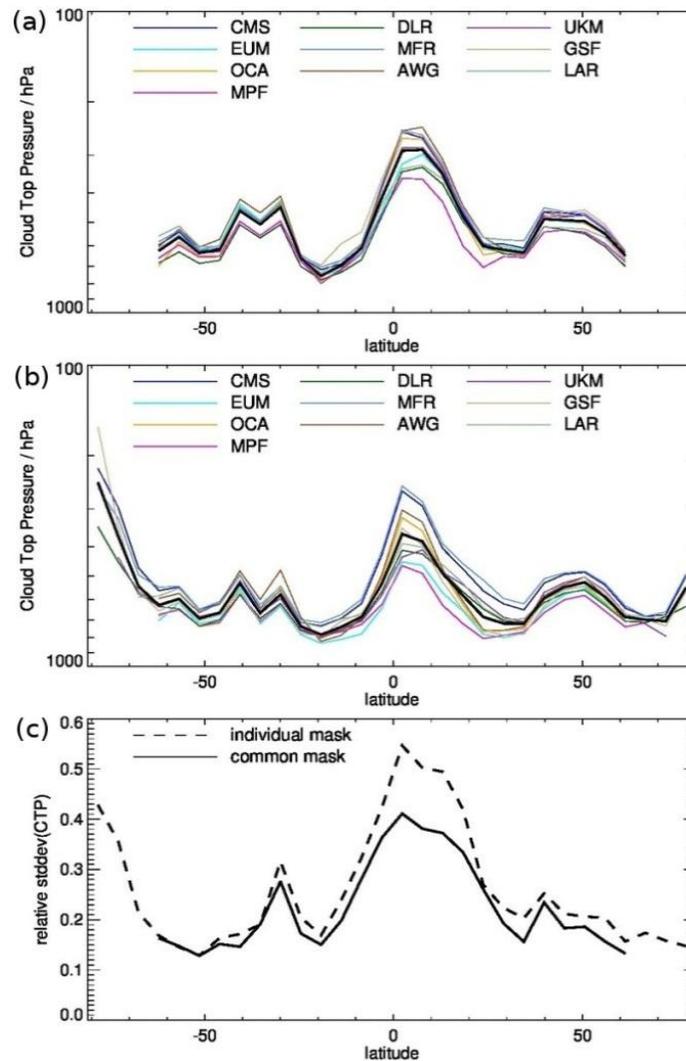**Figure 15: Latitudinal mean of the Cloud Top Height of six algorithms for 13 June 2008at 12:00 UTC. In the upper panel only satellite pixels were used, for which all eleven retrievals derive a result for the CTH, whereas the plot in the middle panel is based on all available cloudy pixels. The black line shows the average of all SEVIRI algorithms.**
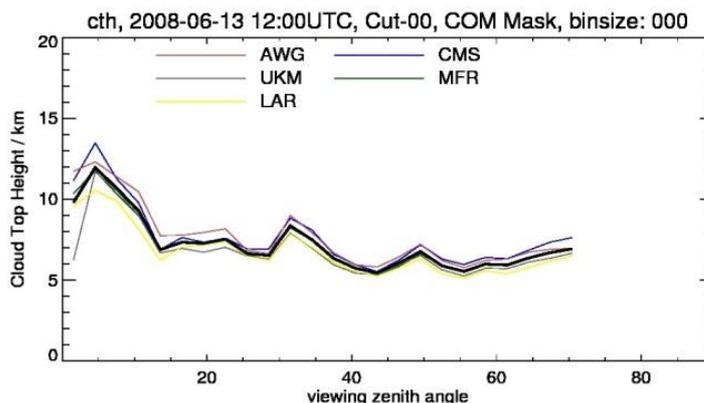
Figure 16: The dependence of the cloud top height on the viewing zenith angle for 13-06-2008 at 12:00UTC. The Figure shows the values for the common cloud mask.

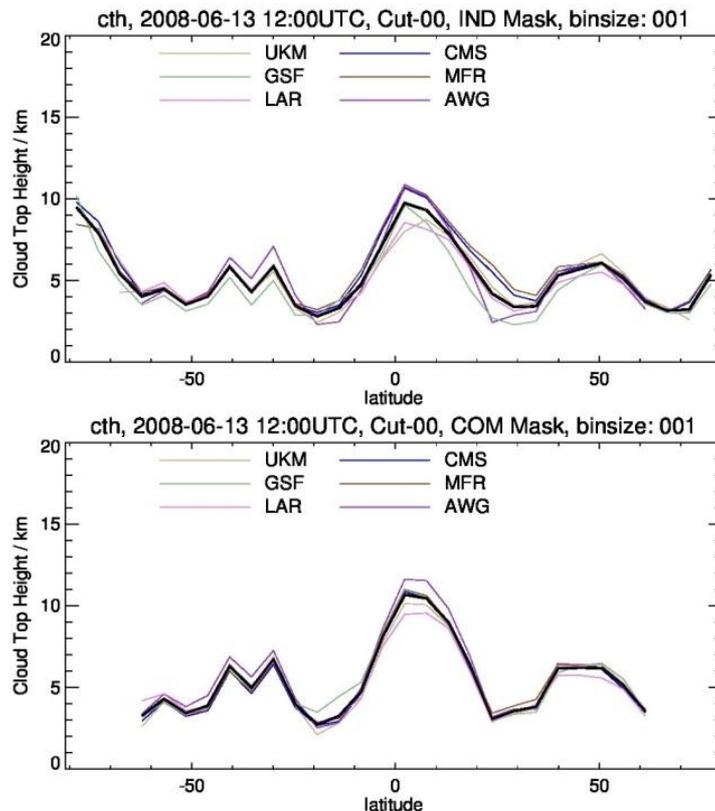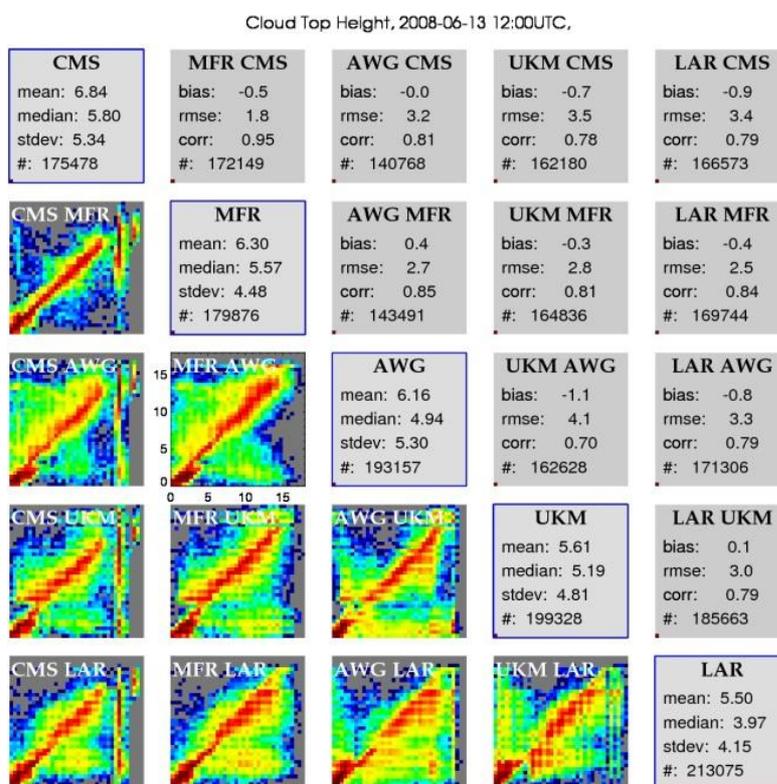**Figure 17: This image type shows joint histogram scatter plot matrix and statistics for all algorithm pairs for the Cloud Top Height 13-06-2008 at 12:00UTC. The small thumbnail images below the main diagonal show scatter plots as a distribution of frequency in a log space. The matrix diagram shall give a rough overview of the results. More detailed information can be obtained through individual files in the single subdirectory. At one box in the upper left edge of the image the range of the axes is printed, those are valid for all thumbnails. Statistical parameters mean value; median value, standard deviation and the number of valid observations, for each algorithm are written in the boxes in the main diagonal from upper left to lower right. The boxes above this main diagonal specifies statistical parameters bias, root mean square error, correlation coefficient and the number of common valid pixels of the comparison between two algorithms.**

Courtesy: Sauli Joro (EUMETSAT)

A common problem of cloud top height retrieval is the boundary layer, where often temperature inversions cause ambiguities in the CTT to CTH conversion. Furthermore, the temperature profiles used as auxiliary data in the retrieval often have uncertainties. Therefore, most groups developed an adjustment for CTH over sea, if a temperature inversion is present. Most often the CTH is moved to a position close to the temperature inversion. EUMETSAT investigated the position of the cloud top height detected by CALIOP with respect to the position of the temperature inversion to estimate the uncertainty of this assumption.
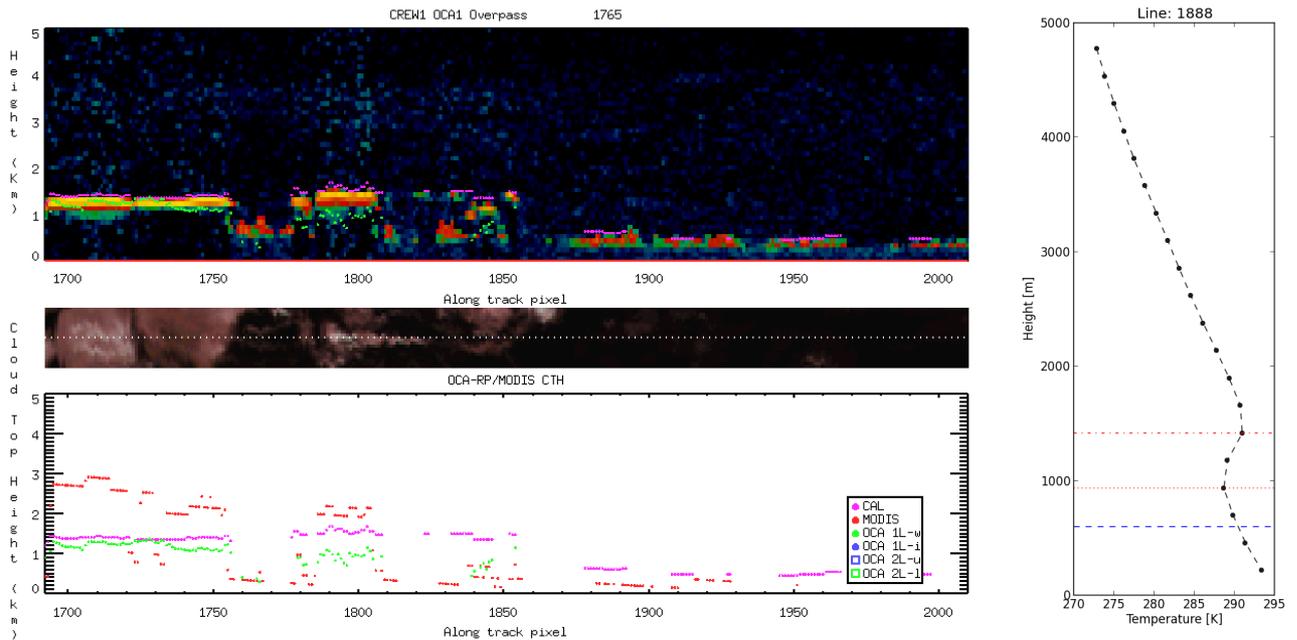
Figure 18: Left side … Right side…

4.1.5   Cloud optical depth
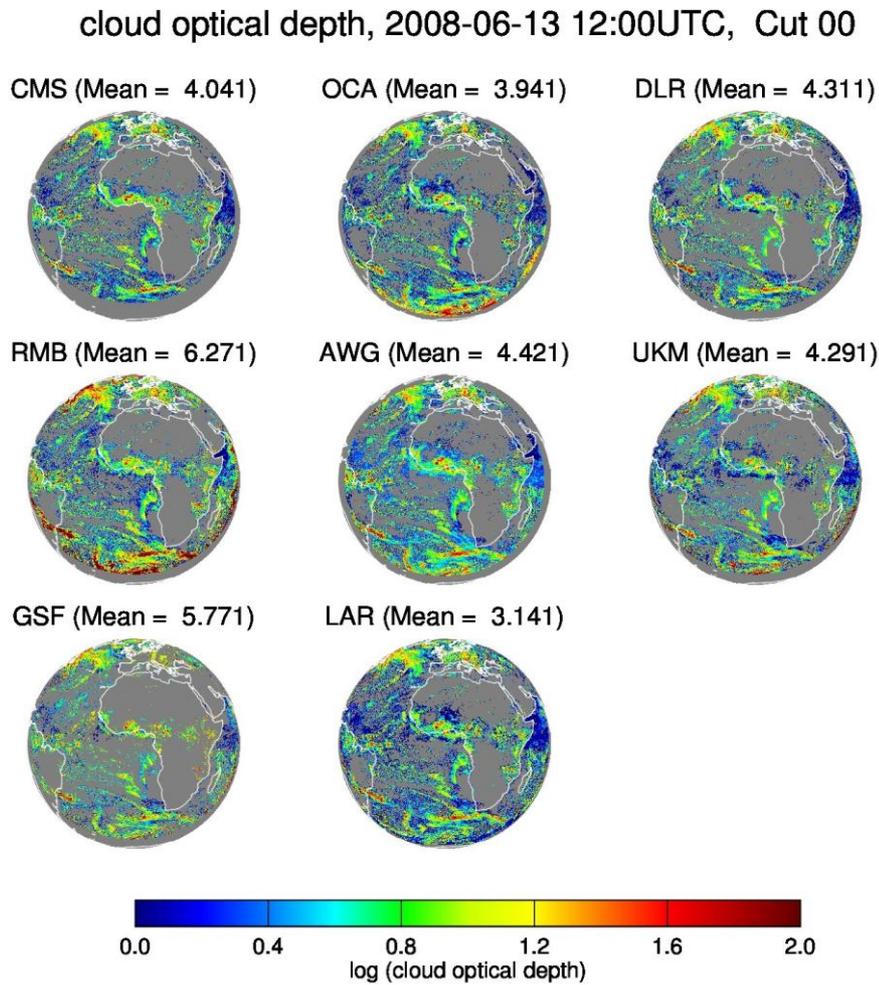
Figure 19: Cloud optical depth for eight SEVIRI algorithms and the COX algorithm for 13-06-2008 at 12:00 UTC.
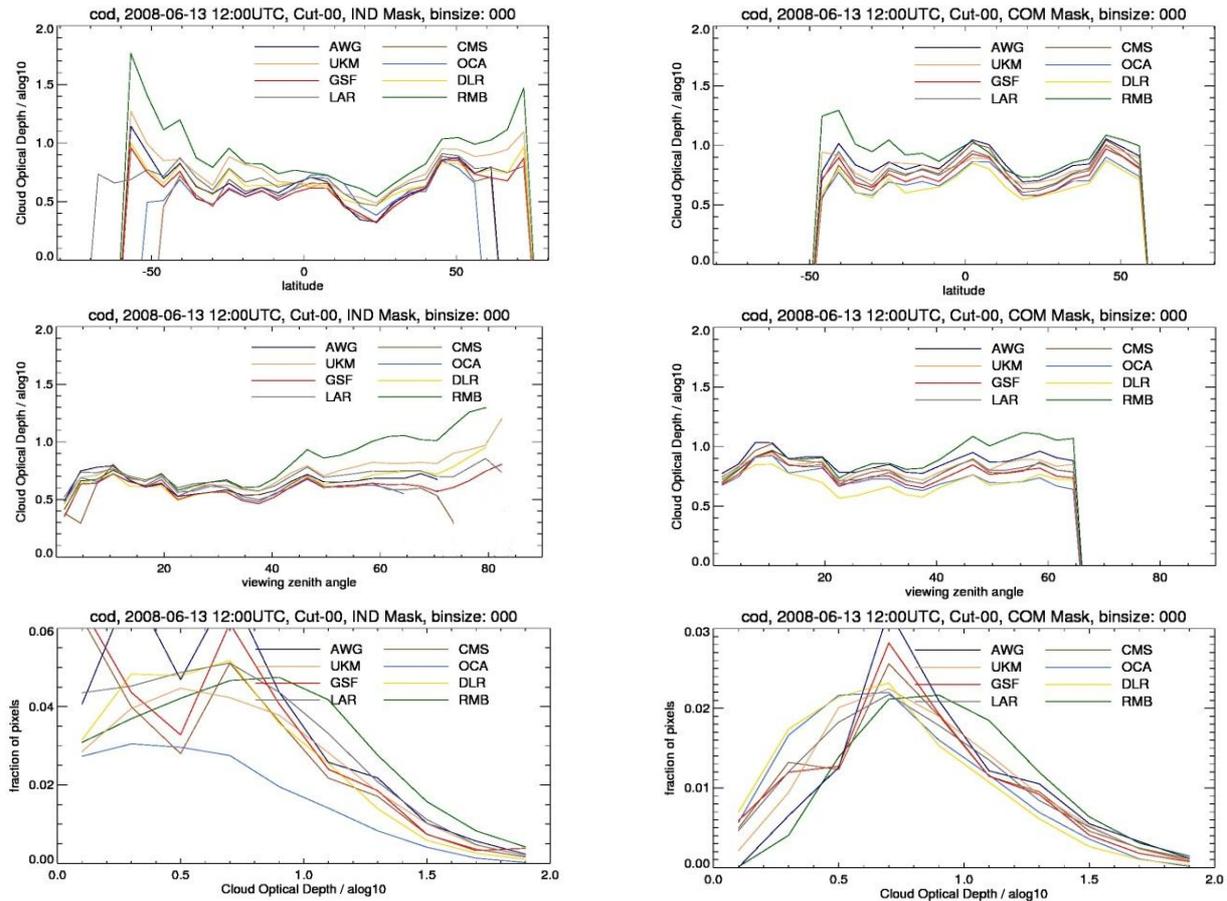
Figure 20: All diagrams show statistics of the cloud optical depth. The first row shows the dependence on the latitude, the second row the dependence on the viewing zenith angle and the third row the histograms. The left hand side shows the cloud datasets with individual cloud masks, the right hand side the datasets filtered by the common mask.

Figure 21: This image type shows joint histogram scatter plot matrix and statistics for all algorithm pairs for the Cloud Optical Depth for 13-06-2008 at 12:00UTC. The small thumbnail images below the main diagonal show scatter plots as a distribution of frequency in a log space. The matrix diagram shall give a rough overview of the results. More detailed information can be obtained through individual files in the single subdirectory. At one box in the upper left edge of the image the range of the axes is printed, those are valid for all thumbnails. Statistical parameters mean value; median value, standard deviation and the number of valid observations, for each algorithm are written in the boxes in the main diagonal from upper left to lower right. The boxes above this main diagonal specifies statistical parameters bias, root mean square error, correlation coefficient and the number of common valid pixels of the comparison between two algorithms.

4.1.6   Effective radius

effective radius, 2008-06-13 12:00UTC, Cut 00

CMS (Mean = 14.1micro m)   OCA (Mean = 17.2micro m)   DLR (Mean = 14.2micro m)

AWG (Mean = 21.4micro m)   UKM (Mean = 15.8micro m)   GSF (Mean = 18.9micro m)

LAR (Mean = 19.7micro m)   RGB picture

0.0   12.0   24.0   36.0   48.0   60.0
ref / \μm

**Figure 22: Cloud effective radius for seven SEVIRI algorithms for 13-06-2008 at 12:00 UTC.**
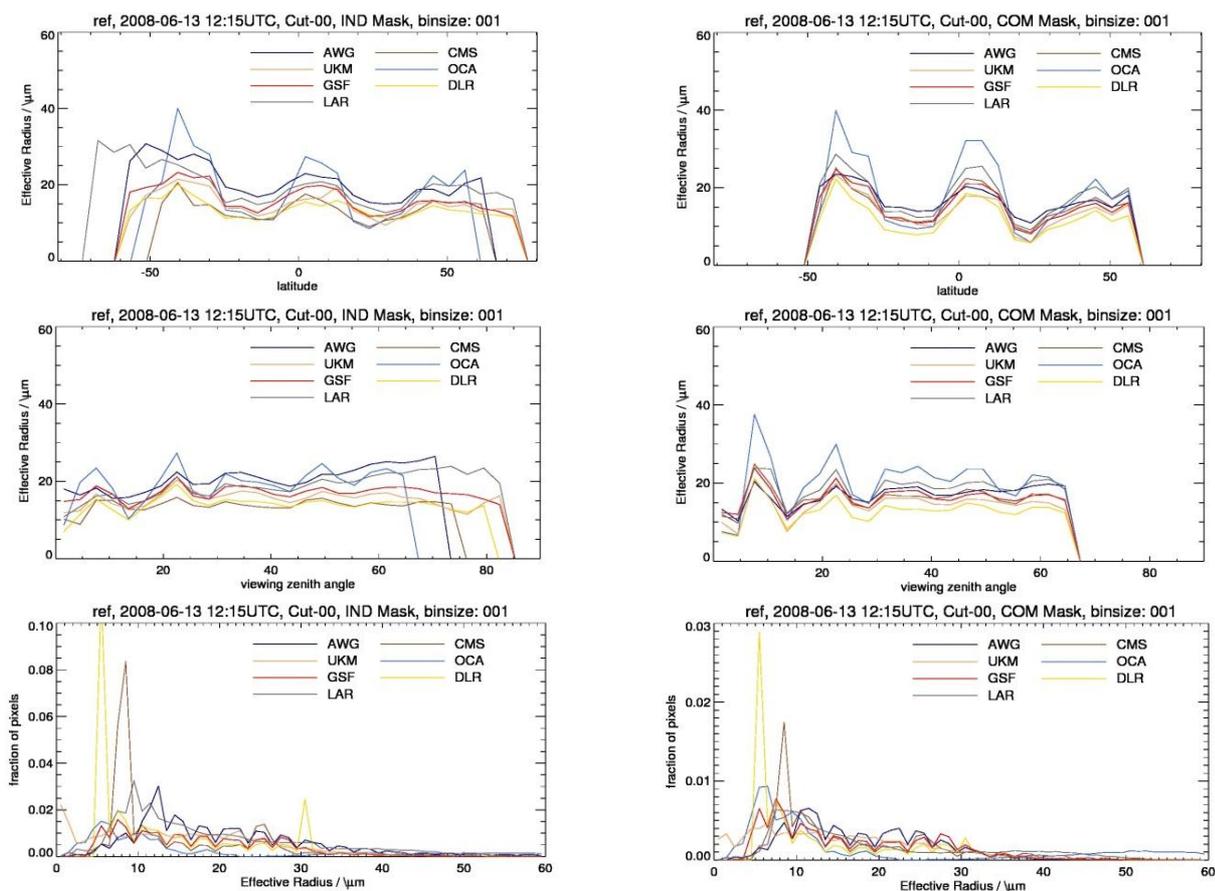
Figure 23: All diagrams show the effective radius. The first row shows the dependence on the latitude, the second row the dependence on the viewing zenith angle and the third row the histograms. The left hand side shows the cloud datasets with individual cloud masks, the right hand side the datasets filtered by the common cloud mask.
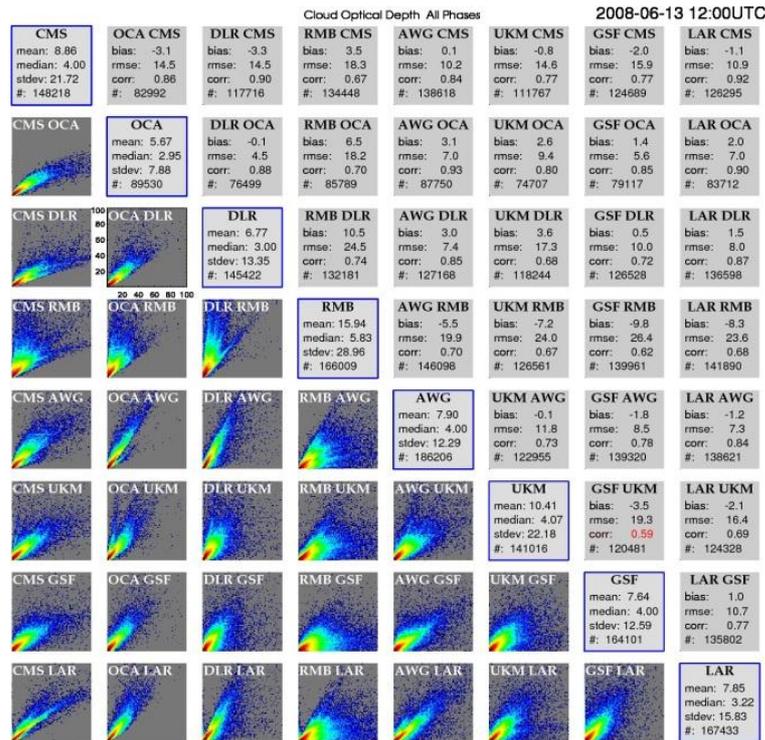
Effective Radius All Phases   20080613_1215
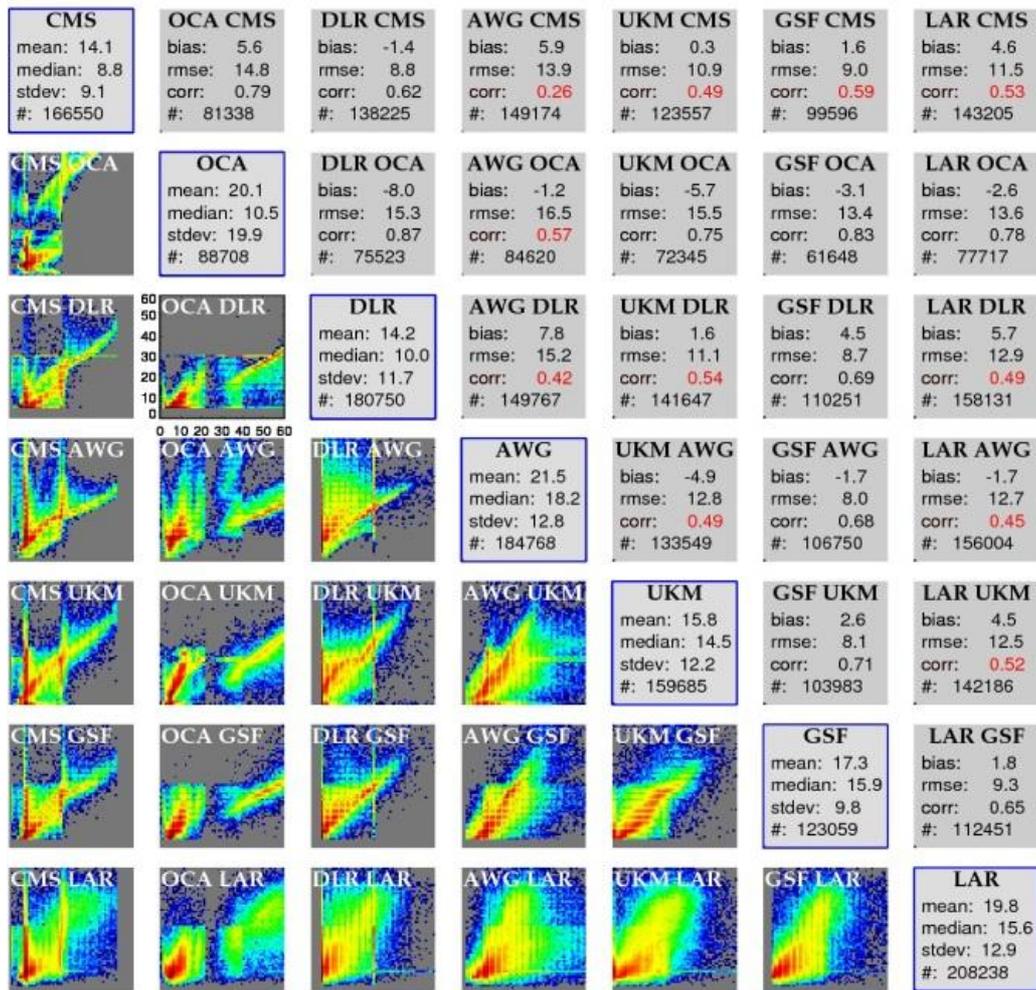
Figure 24: This image type shows joint histogram scatter plot matrix and statistics for all algorithm pairs for the Effective Radius for 13-06-2008 at 12:00UTC. The small thumbnail images below the main diagonal show scatter plots as a distribution of frequency in a log space. The matrix diagram shall give a rough overview of the results. More detailed information can be obtained through individual files in the single subdirectory. At one box in the upper left edge of the image the range of the axes is printed, those are valid for all thumbnails. Statistical parameters mean value; median value, standard deviation and the number of valid observations, for each algorithm are written in the boxes in the main diagonal from upper left to lower right. The boxes above this main diagonal specifies statistical parameters bias, root mean square error, correlation coefficient and the number of common valid pixels of the comparison between two algorithms.

4.1.7   Liquid Water Path
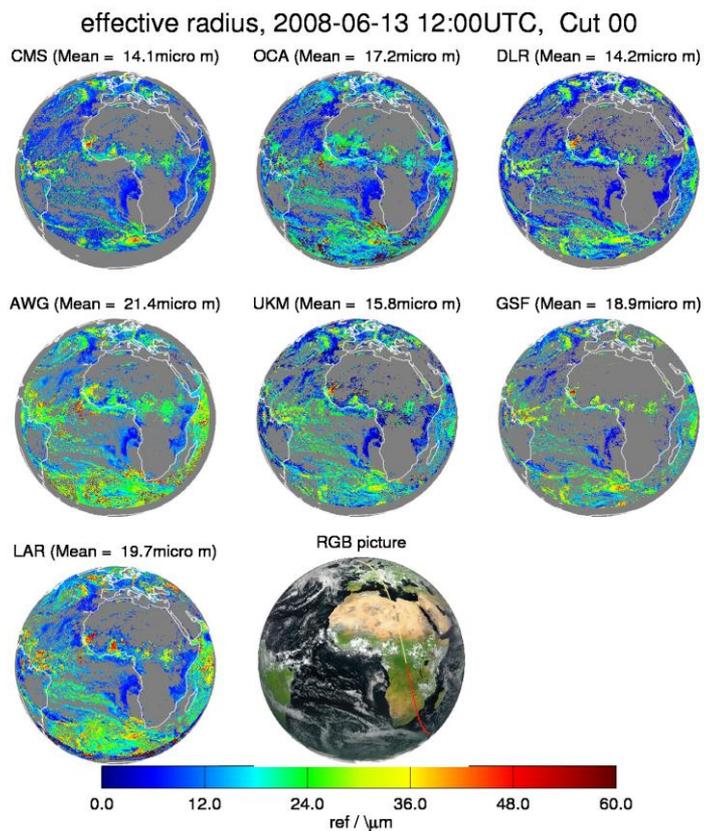
CMS (Mean =   87.g/m^2)   AWG (Mean =   58.g/m^2)

GSF (Mean = 116.g/m^2)   LAR (Mean =   93.g/m^2)

0.0   40.0   80.0   120.0   160.0   200.0
liquid water path / g/m^2

Figure 25: Liquid water path for four SEVIRI algorithms for 13-06-2008 at 12:00 UTC.
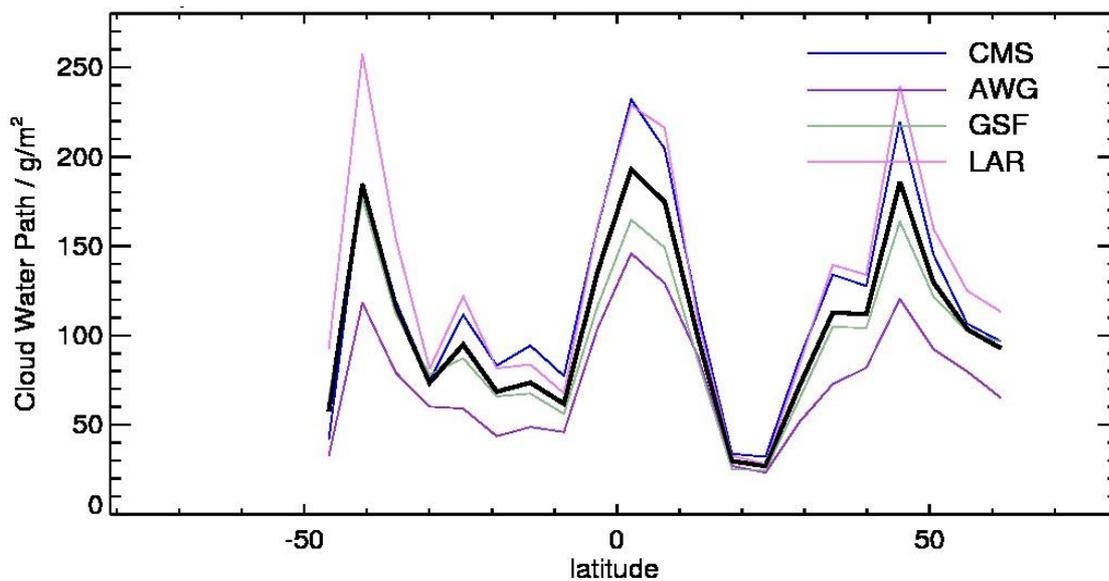
Figure 26: Cloud water path for four SEVIRI algorithms average along a constant latitude for 13-06-2008 at 12:00 UTC.
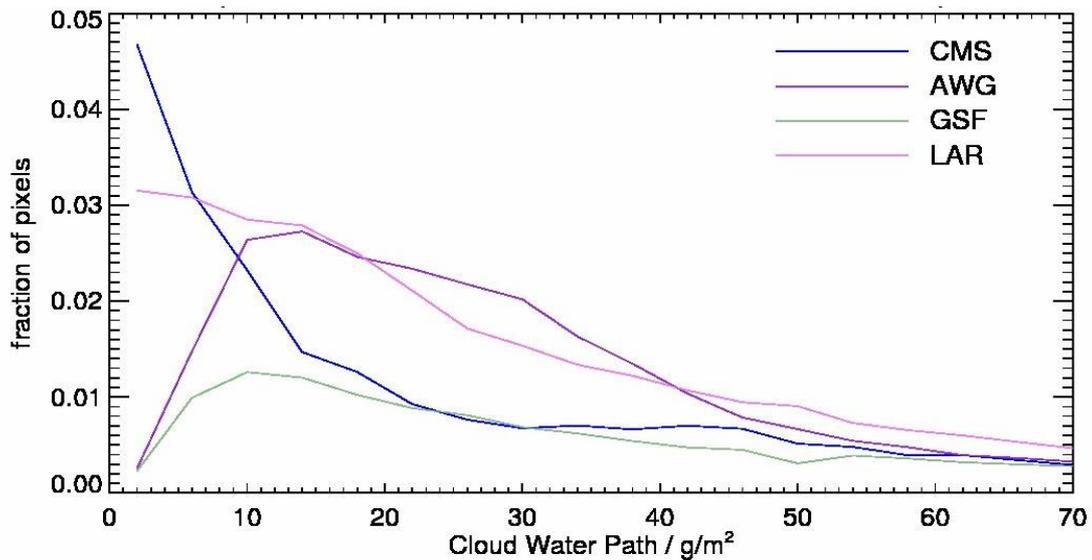
Figure 27: Histogram of the cloud water path (water and ice clouds) for the individual mask, 13-06-2008 at 12:00UTC.
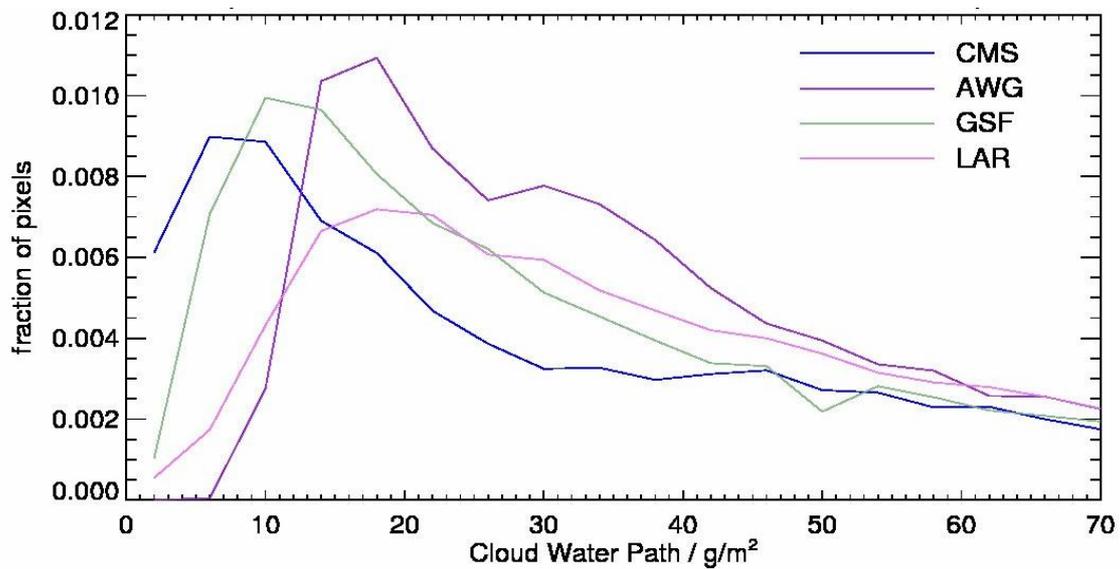
Figure 28: Histogram of the cloud water path (water and ice clouds) for the individual mask, 13-06-2008 at 12:00UTC.

Cloud Water Path All Phases   20080613_1200

| CMS	AWG CMS	GSF CMS	LAR CMS
mean: 87.	bias: -28.7	bias: -18.2	bias: 8.2
median: 27.	rmse: 107.5	rmse: 98.7	rmse: 102.4
stdev: 165.	corr: 0.81	corr: 0.85	corr: 0.88
#: 145501	#: 125616	#: 89159	#: 135874
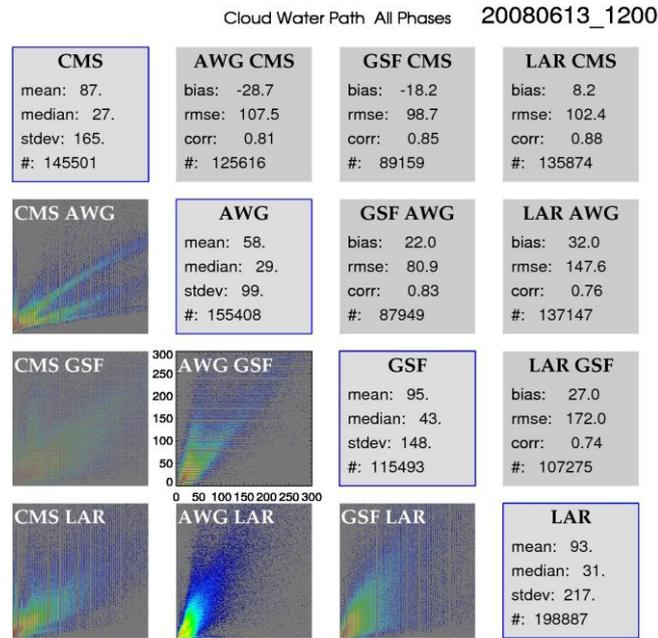
Figure 29: This image type shows joint histogram scatter plot matrix and statistics for all algorithm pairs for the Cloud Water Path for 13-06-2008 at 12:00UTC. The small thumbnail images below the main diagonal show scatter plots as a distribution of frequency in a log space. The matrix diagram shall give a rough overview of the results. More detailed information can be obtained through individual files in the single subdirectory. At one box in the upper left edge of the image the range of the axes is printed, those are valid for all thumbnails. Statistical parameters mean value; median value, standard deviation and the number of valid observations, for each algorithm are written in the boxes in the main diagonal from upper left to lower right. The boxes above this main diagonal specifies statistical parameters bias, root mean square error, correlation coefficient and the number of common valid pixels of the comparison between two algorithms.
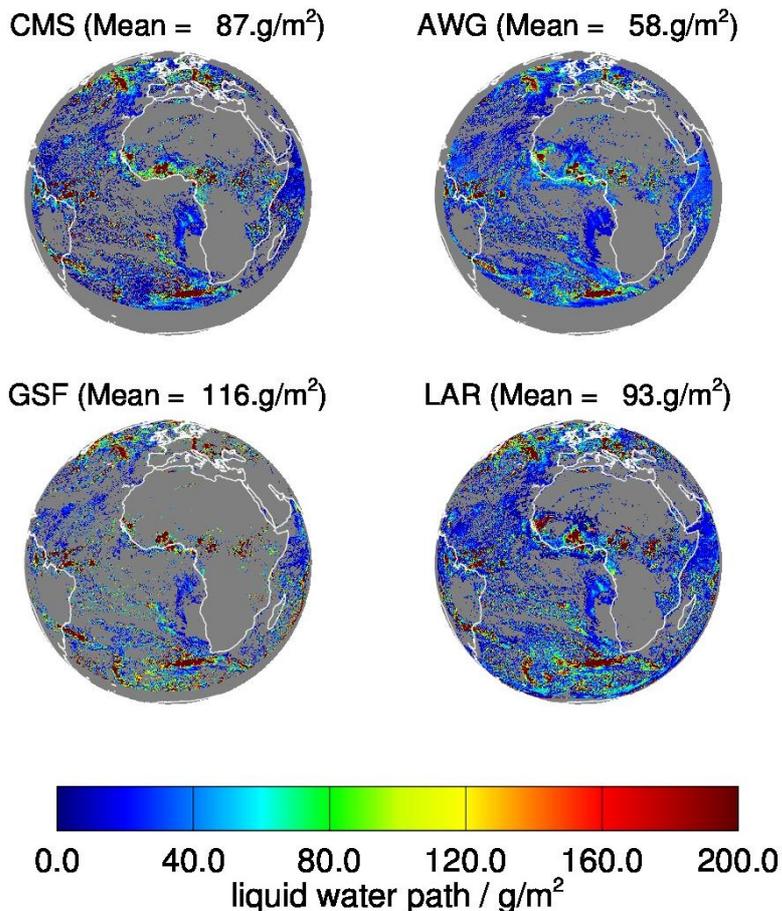

Figure 30: Same as Figure 29, but only for ice clouds (left) and water clouds (right) separately.

4.2   Validation of SEVIRI products

Courtesy: Ulrich Hamann (KNMI, MeteoSwiss), Andi Walther (SSEC), Ralf Bennartz (SSEC) and the CREW team

List of open questions for the retrieval validation working groups (extendable)
- What validation sources are best suited for validation the different cloud parameters?
- What ground based sensors or sensor networks are best suited?
- How can measurement campaigns optimally contribute to validation of cloud properties?
- How can we coordinate the validation between data providers?
- How can we improve the validation over different climate regions and surface types?
- What role can reanalysis data play in validation practices?
- Should and could CREW support the development of centralized validation database comprising active instrument, ground based and campaign observations (generalize AVACS)?
- Did the differences between the different data providers decrease?
- What is the best practice…
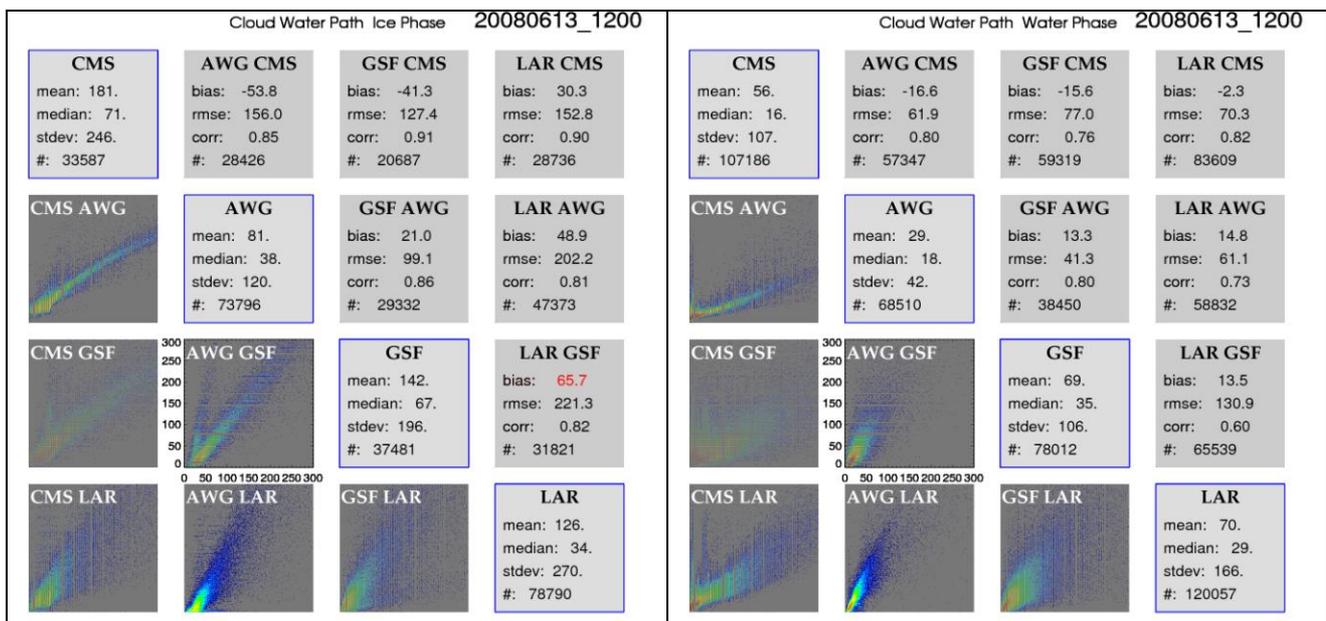 1. ... to take care of different sensitivities (active/passive sensors)?
  2. ... to take care of different spatial resolution?
  3. ... to take care of different viewing geometries?
  4. ... to include the uncertainty information in the inter-comparison/validation?
- Is there an issue with respect to differences in spectral response functions?
- Should we use of satellite observation simulators?
- Can we improve the inter-comparison between geostationary and polar satellite products?

While chapter 4.1 shows the results on an inter-comparison study this chapter compares the findings with other, independent observations from sensors of the ATRAIN satellite chain. Relevant data products are cloud top height from CPR and CALIPSO, cloud optical products from MODIS and liquid water path from AMSR-E. For the purpose of the comparison the new EUMETSAT validation tool AVAC-S was used and enhanced (Bennartz, 2010). The basic strategy is to store all products from all groups and sensors on the nominal MSG grid. Therefore, all ATRAIN sensor data are transformed in a suitable way to the SEVIRI grid. Within the AVAC-S software package, several data levels were defined. Level "S" denotes all source data, those usually comes in HDF-4 or other binary data format. Level "B" denotes the collected data set of all ATRAIN data. AVAC-S also provides suitable tools for visualization. For the purpose of this workshop a new data level "C" was created. These files contain the data provided by the participants of this workshop in a similar way to level-B data. All data from level "B" and "C" are stored in HDF-5 format.

In order to quantify the accuracy of the SEVIRI cloud top height (CTH) retrievals the datasets are validated against independent observations from sensors of the ATRAIN satellite constellation, namely the cloud and aerosol lidar CALIOP (Winker et al., 2009) and cloud profiling radar CPR (Stephens et al., 2002). Both instruments were launched in 2006.

The CALIOP and CPR data were reprojected on the SEVIRI grid by using the nearest neighboring values. The AVAC-S software can correct for the parallax effect of the SEVIRI viewing zenith angle. The SEVIRI sensor scans the observation disk every 15 min. The scan starts in the south and takes 12 min until it reaches the northernmost point. The CALIOP and CPR data is matched with the SEVIRI observations for which the time shift is smallest, leading to a maximal observation time difference of 7.5 min. In case a SEVIRI algorithm only provides Cloud Top Pressure (CTP) and not the CTH, CTP values were transformed to CTH values using temperature profiles of the ECMWF.

In Figure 31 the CTH retrievals of the SEVIRI algorithms are compared with the CALIOP and CPR retrievals. The upper right panel shows the CPR backscatter profile and the CTH retrieved from the CPR data. For cloud free regions the noise level of the radar signal can be observed. Additionally the lower right panel shows the CTH from CALIOP in green and the mean of all SEVIRI algorithms in black.
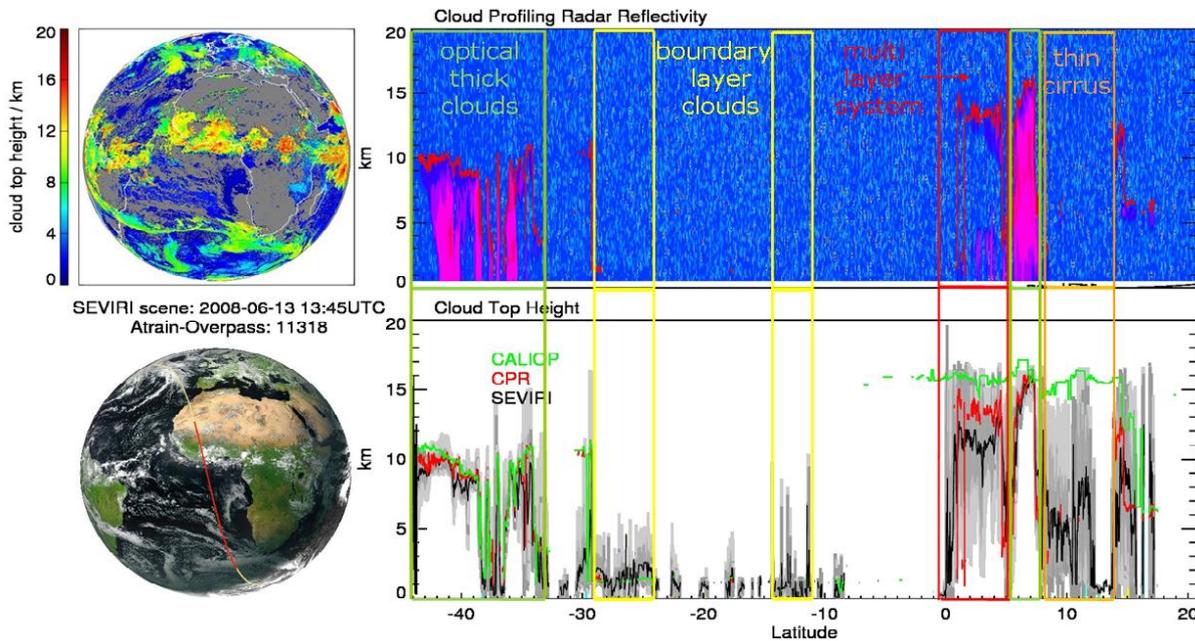

Figure 31: Validation of the Cloud Top Height (CTH) retrievals using SEVIRI with CALIPSO and CPR for 13-06-2008 at 13:45 UTC or ATRAIN overpass 11318. The upper left panel shows the mean CTH of all 11 SEVIRI algorithms. The lower left diagram shows the false color composite of the SEVIRI disk with the path of the ATRAIN satellite constellation in yellow and the part of the track that is shown on the right in red. At the upper right the CPR radar reflectivity and the CTH derived from CPR (red) is shown. The lower right diagram shows the CTH derived from CALIOP (green) and CPR (red) observations. The mean of the SEVIRI algorithms is shown in black. Light grey shows the range from minimum to maximum of the SEVIRI retrievals, dark grey indicates the multi algorithm standard deviation. Areas of optical thick clouds are marked in green, multi layer clouds in red and optically thin clouds in orage.

In order to understand the differences of the CTHs the sensitivities of the observing systems have to be considered. CALIOP is the most sensitive to cloud particles and is able to detect clouds with a very small optical depth. The radar system CPR is less sensitive than CALIOP. Therefore it is expected, that the CTH of CPR usually is below the CTH of CALIOP. Both systems are active and hence have a high vertical resolution compared to passive sensors. It is 30 m to 60 m for CALIOP and 500 m for CPR. In contrast the SEVIRI sensor is a passive instrument. The radiance measured by SEVIRI originates from different levels, hence the retrieved CTH is a radiatively effective one. Due to the high sensitivity of CALIOP, it is expected that the CTH of passive imager retrievals (SEVIRI) is lower than the CTH of CALIOP and might be similar to the CTH of CPR.

In Figure 32, we investigate the individual CTH retrievals the region with multi layer clouds and thin cirrus clouds marked in red and orange respectively. In the red area, there are two different cloud layers according to the CPR reflectivity, see the upper right panel of Figure 32. The CALIOP signal indicates that the cloud top of the upper cloud layer is located at 17 km. The CTH retrieved by the CPR is 1 to 3 km lower than the CTH of CALIOP. A second cloud layer is located between 0 and 4 km. The mean of the SEVIRI results is located in between the two cloud layers at about 12 km, and the standard deviation of the SEVIRI results is as large as 3 to 5 km. In the orange region, CALIOP is able to detect an optically thin cloud layer in around 16 km. The sensitivity of CPR is not sufficient to catch this layer. No other lower cloud layer is indicated by the CPR reflectivity. Therefore we call this a thin cirrus cloud layer situation. Due to the small cloud optical depth of the cirrus layer, the CTH detected by CALIOP and the radiatively effective CTH retrieved by the SEVIRI algorithms are different. The large spread of SEVIRI CTHs indicates that this is a challenging situation for passive imager retrievals.
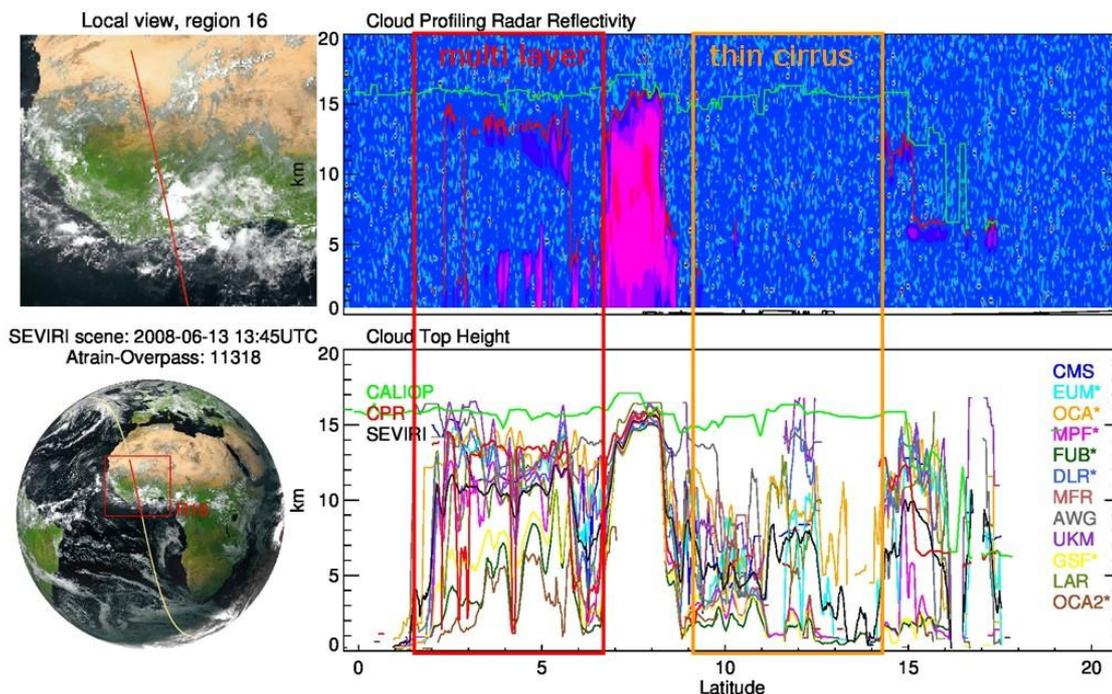
Figure 32: Similar to Figure 31, but for a smaller region focusing on a region with multi layer clouds and thin cirrus clouds. The upper left figure is replaced by a close up of the false color composite. In the lower right figure, not only the mean of the SEVIRI algorithms is shown, but also the results of all individual algorithms (smoothed horizontally by 7 pixels). The groups are CM SAF (CMS), EUMETSAT retrievals (EUM, OCA and MPF), Free University of Berlin (FUB), German Aerospace Center (DLR), Meteo France (MFR), University of Madison Wisconsin (AWG), UK MetOffice (UKM), NASA Goddard Space Flight Center (GSF), and NASA Langley Research Centre (LAR). OCA2 is a special product of the OCA algorithm, it is the cloud top height of the second cloud layer. Groups that do not submit a cloud top height, but a cloud top pressure (that we converted to cloud top height according to ECMWF data) are marked with a star *.

In Figure 33 we investigate the effect of the common mask filter on the histogram of CTH. In Figure 33a the histograms of the original datasets are shown. The CALIOP datasets shows cloud occurance maxima in 16km, 11km, 6.5km and 1.5km. CPR and the SEVIRI datasets have according cloud maxima, but somewhat lower. When looking at the datasets reduced to the common mask, in particular the occurrence of thin cirrus at 16km is strongly reduced. Figure 33b shows the histograms of the individual SEVIRI algorithms. There are some differences among the SEVIRI algorithms in reproducing the cloud occurrence maxima, e.g., the cloud occurrence maxima of the boundary layer clouds of LAR, MFR and DLR are at higher altitudes than those of the other algorithms.
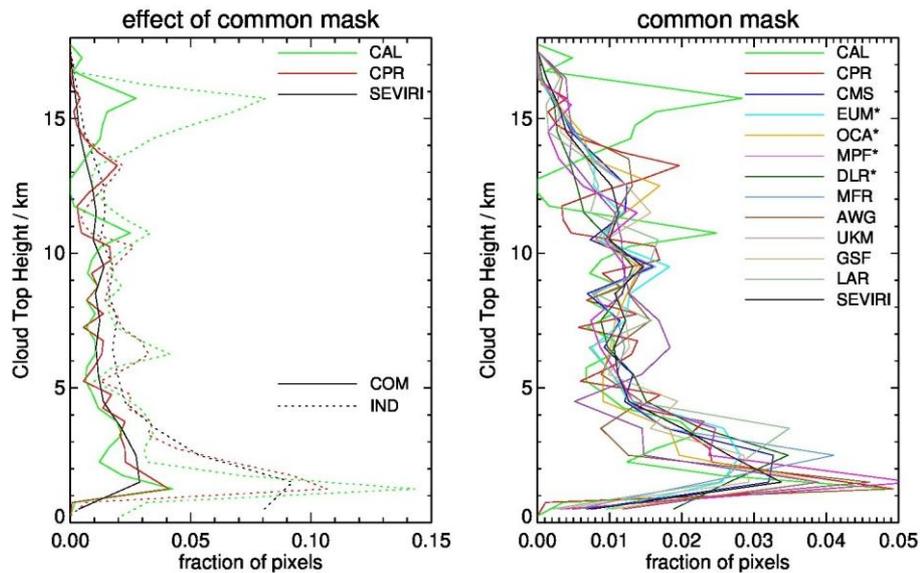
Figure 33: Histograms of the CTH for 13 June 2008, 12:00–15:30UTC (A-train overpasses 11 317– 11 319). Panel (a) shows the histograms of the complete CALIOP and CPR dataset and the average of the SEVIRI algorithm histograms as dotted lines. The histograms using the common mask filtering are shown as solid lines. In panel (b) the histograms of the individual algorithms are shown using the common mask filtering. For multi-layer cloud situations only the uppermost CTH is considered.

Summary Report 2014 on the Inter-comparison and Validation studies of CREW (Cloud Retrieval Evaluation Workshop)

Validation: Cloud Top Height, CALIOP - SEVIRI, Overpass 11317-11319, com. mask
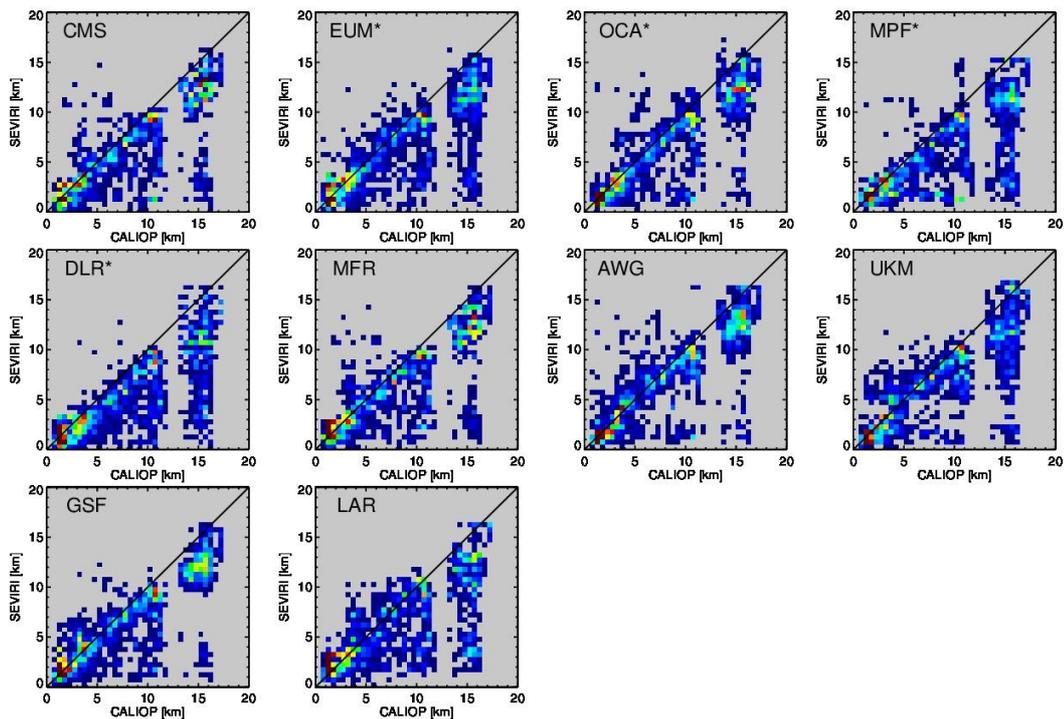
Figure 34: Scatter plots of the cloud top height SEVIRI datasets against the CALIOP dataset for 13 June 2008, 12:00–15:30UTC (A-train overpasses 11 317–11 319). Most of the points are on the lower right side showing that the SEVIRI algorithms derive lower CTH than CALIOP.

Validation: Cloud Top Height, CPR - SEVIRI, Overpass 11317-11319, com. mask
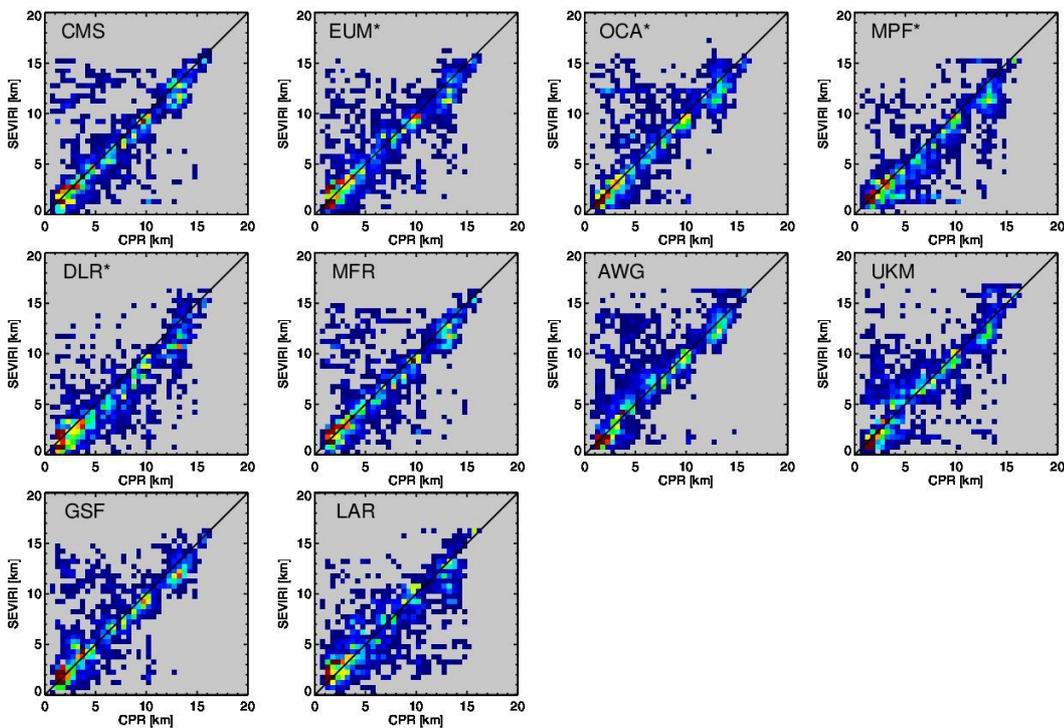
Figure 35: Scatter plots of the cloud top height SEVIRI datasets against the CPR dataset for 13 June 2008, 12:00–15:30UTC (A-train overpasses 11 317–11 319). Most of the points are on the lower right side showing that the SEVIRI algorithms derive lower CTH than CALIOP.
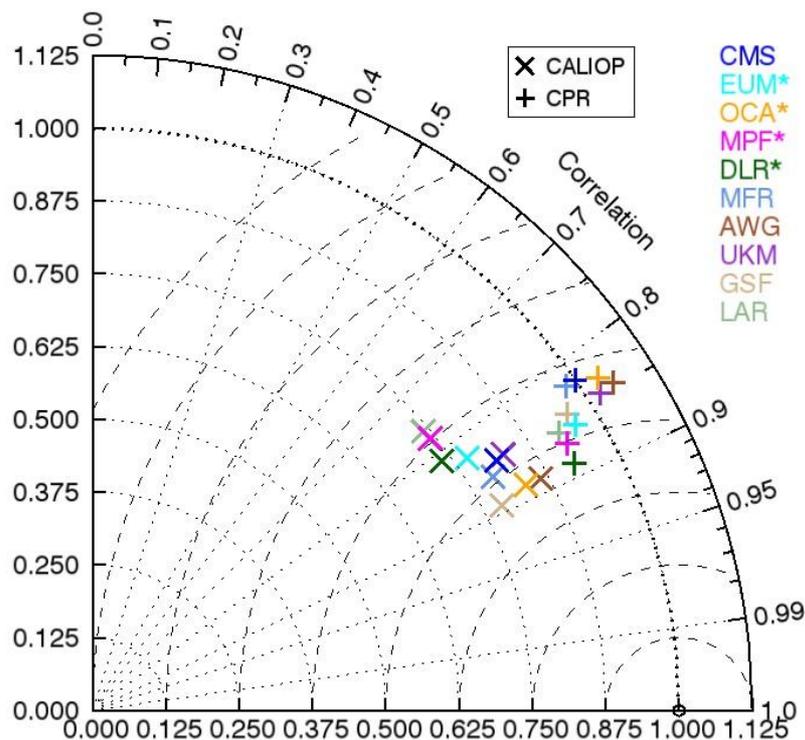
Figure 36: Taylor diagram for CALIOP and CPR. The Taylor diagram shows the standard deviation of the SEVIRI retrieval divided by those of the reference sensor as radial coordinate and the cosine of the correlation coefficients of these datasets as angle. The diagram shows the comparison to the CALIOP and CPR dataset. The standard deviations of the SEVIRI datasets are smaller than the one of CALIOP and comparable to the one of CPR. The correlation coefficients are for both active sensors between 0.77 and 0.90.
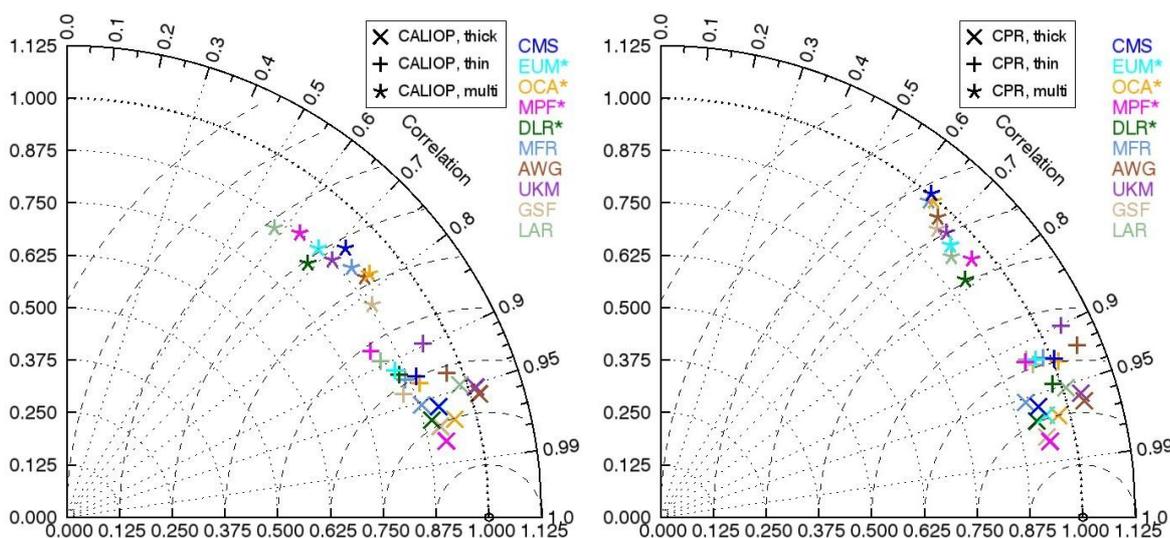
Figure 37: Taylor diagram for CALIOP (left) and CPR (right). The statistics are calculated separately for optically thick (cod > 3) and thin (cod < 3) single layer clouds as well as for multi-layer clouds.
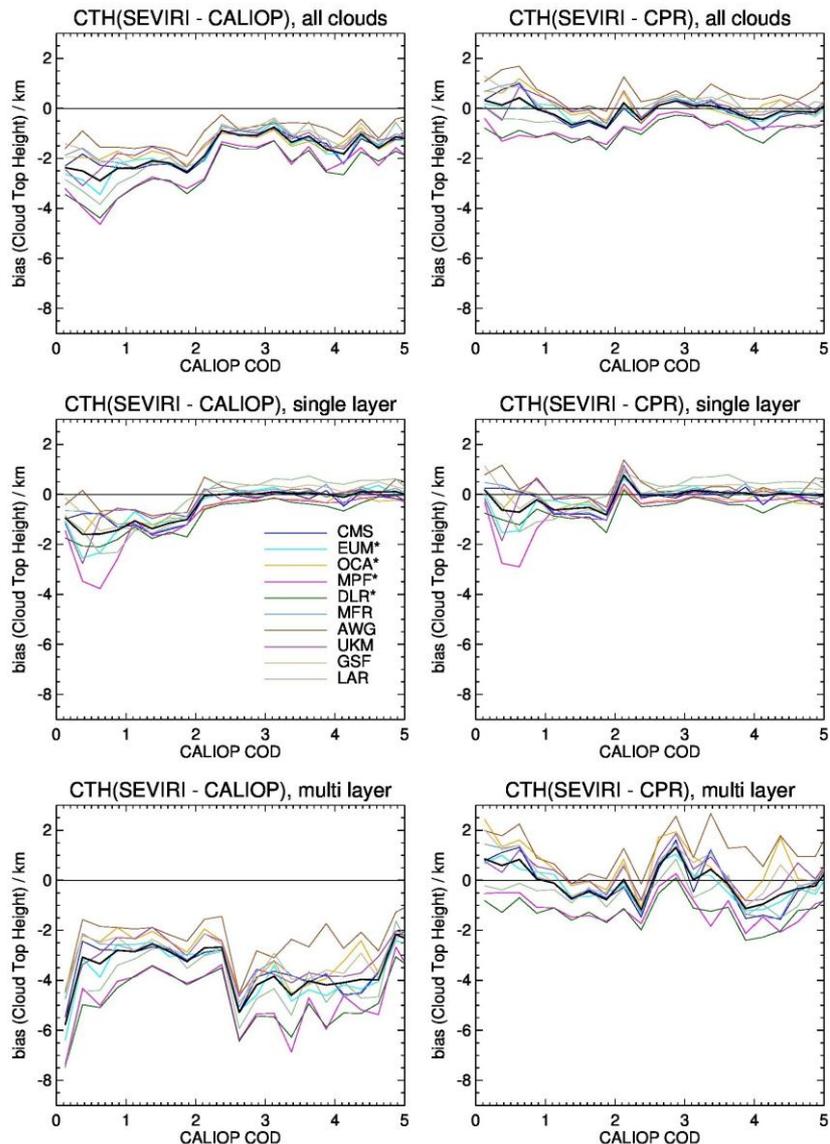
Figure 38: Differences of the CTH between the SEVIRI and CALIOP and CPR in dependence of the CALIOP cloud optical depth for 13 June 2008, 12:00–15:30UTC. Uppermost row shows the results for all clouds, second row for single layer and the third row for multi-layer clouds. The first and third column show the comparison to CALIOP CTH, the second and fourth one the comparison to CPR CTH. All statistics are calculated for the common mask.

Plans for the future

At the moment only the cloud top height and pressure have been evaluated. Further investigations are probably quite interesting. Following issues have to be solved before doing so:

Validation of the cloud mask

CALIOP is the most sensitive instrument for clouds compared to passive imagers and radar. As CALIOP also provides information about the cloud optical depth, its dataset is ideally suited to investigate the detection limits of passive imagers. The first step here is to evaluate the definitions of the cloud mask of the individual groups. In particular, the definition of the cloud mask for broken

clouds and optically thin clouds matters in this respect. The second step is to implement the 333m and 1km products into the AVACS software. Due to the construction of the CALIOP products, the clouds detected by CALIOP product with high horizontal resolution are removed from the signal before the product with a rougher horizontal resolution is processed. (That means that clouds in the 1km products are not present in the 5km product.) Therefore, the usage of the 5km product alone (as it is implemented in the AVACS software at the moment) is not sufficient to validate the cloud detection of passive imagers. An intelligent strategy has to be designed to map the 333m, 1km and 5km CALIOP product to the SEVIRI grid before this kind of validation can be done.

4.3   Validation of Polar Satellite products

4.3.1   Results of Cloud_cci's Round Robin exercise

Courtesy: Martin Stengel (DWD) and the Round Robin team

List of open questions for the retrieval validation working groups (extendable)
Same as in Section 0.

Scope
Within the ESA Cloud CCI Round Robin exercise, cloud retrieval data from different retrieval schemes were analysed, validated and intercompared. With this, an objective and critical view on the quality of the retrieved cloud properties and the performances of corresponding state-of-the-art retrieval schemes have been accomplished. To ensure comparability, satellite radiances (MODIS-AQUA and AVHRR-NOAA18) and auxiliary information (ERA-Interim) were prescribed for all participating schemes. For AVHRR the used measurements are based on Heidinger et al. (2010) inter-calibration. For MODIS, Collection 5 Level1b data was used. A complete report is given in the document Product Validation and Intercomparison Report (PVASR, 2011)

Participating Schemes
1) The Cloud from AVHRR Extended (CLAVR-x) processing system is hosted at NOAA at University of Wisconsin. CLAVR-x is the basis of the PATMOS-x climate data set (Pavolonis et al., 2005; Walther et al., 2011).
2) The Climate Monitoring Satellite Application Facility (CM SAF) scheme is based on two components: the Cloud Physical Properties (CPP) algorithm developed at KNMI (Roebeling et al., 2006) and the Polar Platform System (PPS) cloud processing package developed by SMHI (Dybbroe et al., 2005a+b).
3) The Oxford RAL retrieval of Aerosol and Cloud (ORAC) algorithm (Poulsen et al., 2010 and Watts et al., 1998) is an optimal estimation retrieval method that can be used to determine both aerosol and cloud properties from visible and infrared satellite radiometers. The cloud retrieval has thus far been applied to ATSR-2, AATSR, MODIS, SEVIRI and AVHRR.

Results
All Figures and figure captions of this subsection have been take from and are in detail discussed in Stengel et al., 2013.
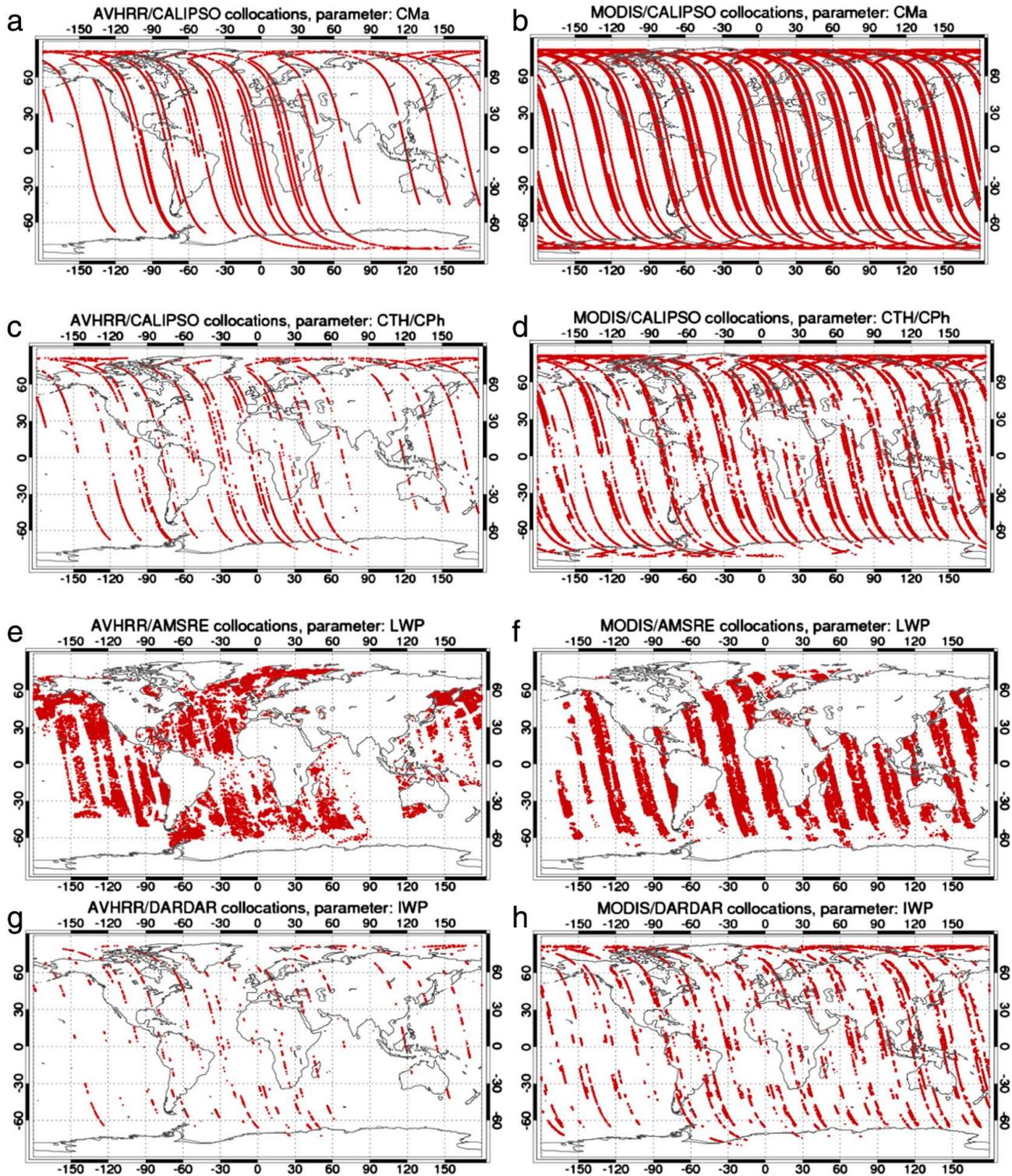
Figure 39: Global map of spatial occurrences of AVHRR- (left column) and MODIS-based (right column) retrievals, after collocation with CALIPSO (panels a, b for CMa and c, d for CTH and cloud phase (CPH)),with DARDAR (panels g, h for IWP), and with AMSR-E (panels e, f for LWP). The collocated pixels of available orbits of all five days are shown. For a better visualization they are plotted enlarged compared to the original pixel size on the ground. Figure and Caption are taken from Stengel et al. (2013).
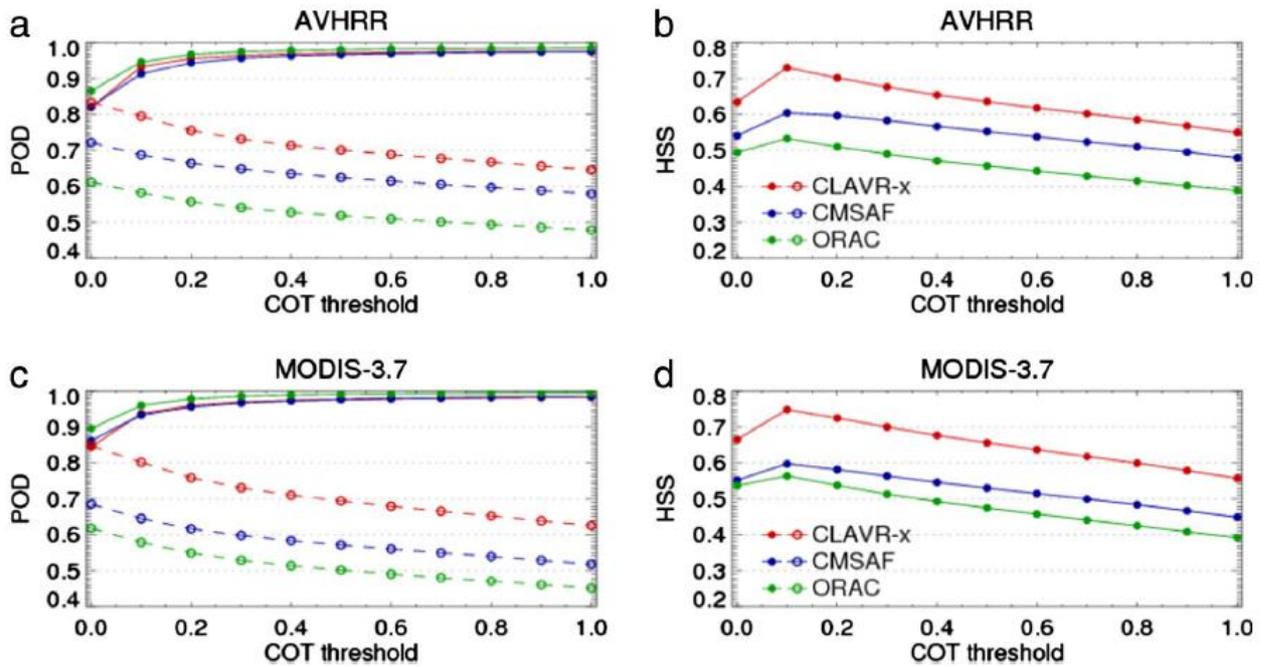
Figure 40: Left panels: Probability of detection (POD) for clear-sky (dashed lines) and cloudy scenes (solid lines) as function of COT threshold, applied to CALIPSO COT estimates, used for separating cloudy and clear-sky conditions as reference for AVHRR (a) and MODIS (c). Right panels: Heidke Skill Score (HSS) for cloud detection as function of COT threshold for AVHRR (b) and MODIS (d). Figure and Caption are taken from Stengel et al. (2013).
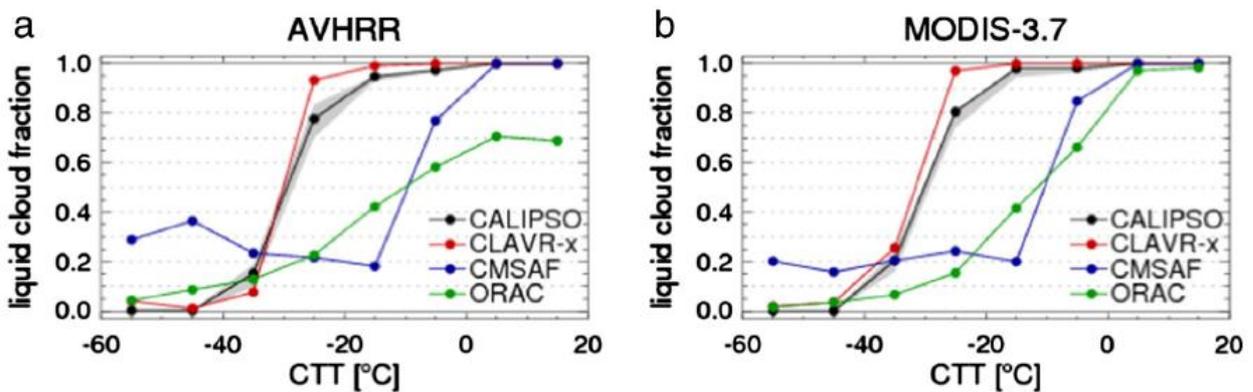
Figure 41: Liquid cloud top phase occurrence as function of cloud top temperature (CTT) for AVHRR (a) andMODIS-3.7 (b). The CALIPSO lines represent phase and temperature taken below the cloud top at 0.25 COT threshold, while the gray shaded areas show the variation of these CALIPSO statistics if calculated between cloud top and down to a COT threshold of 1.0. Figure and Caption are taken from Stengel et al. (2013).
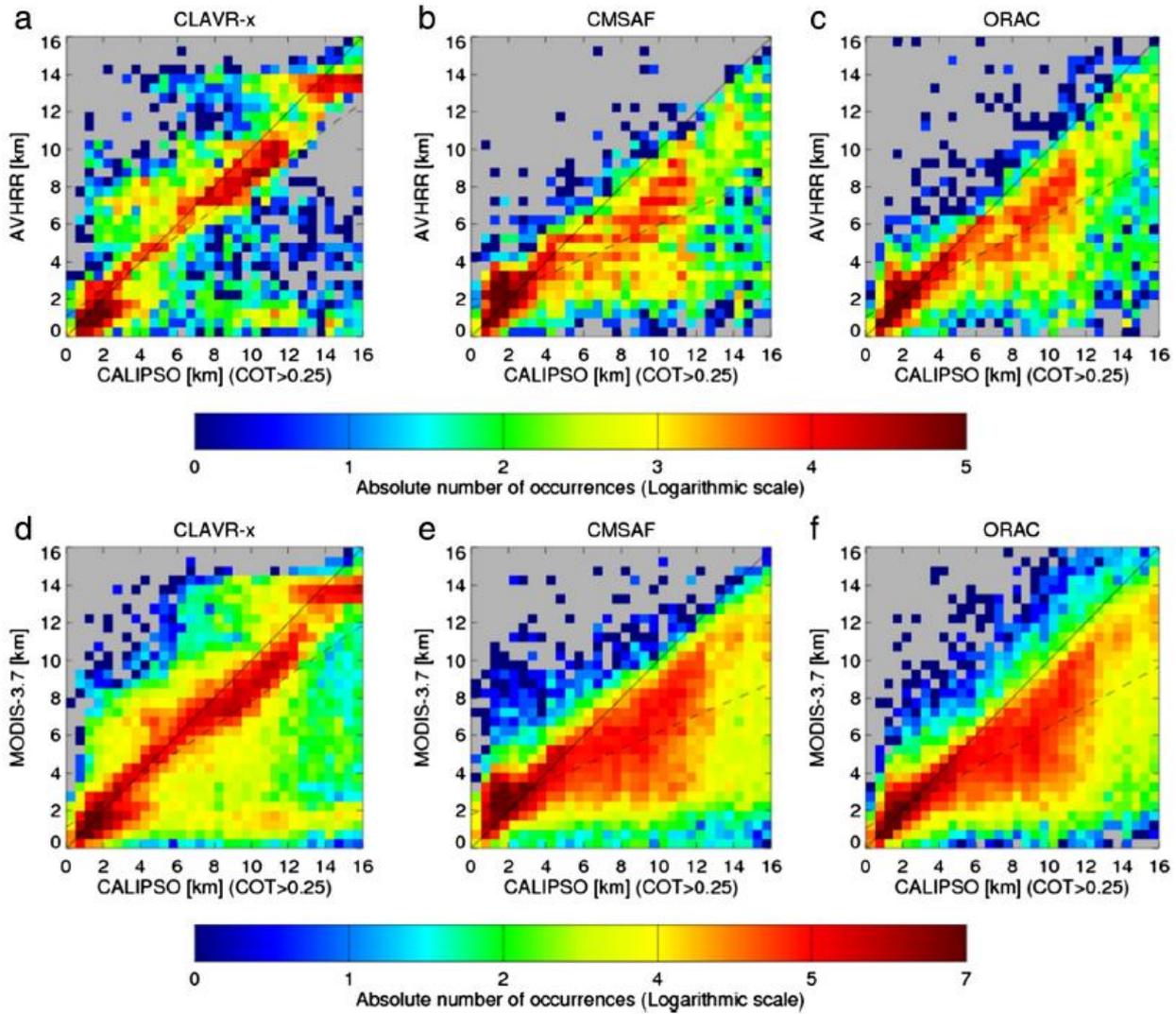
Figure 42: Two-dimensional frequency distribution of cloud top height derived from CALIPSO, AVHRR (top panels) and MODIS (bottom panels). Linear regression lines are also shown (dashed lines). For these comparisons CALIPSO cloud top height was determined for that cloud level for which the level-to-cloud-top COT reached 0.25. The dependence of the CTH deviations on the COT threshold is indicated in Fig. 5. Figure and Caption are taken from Stengel et al. (2013).
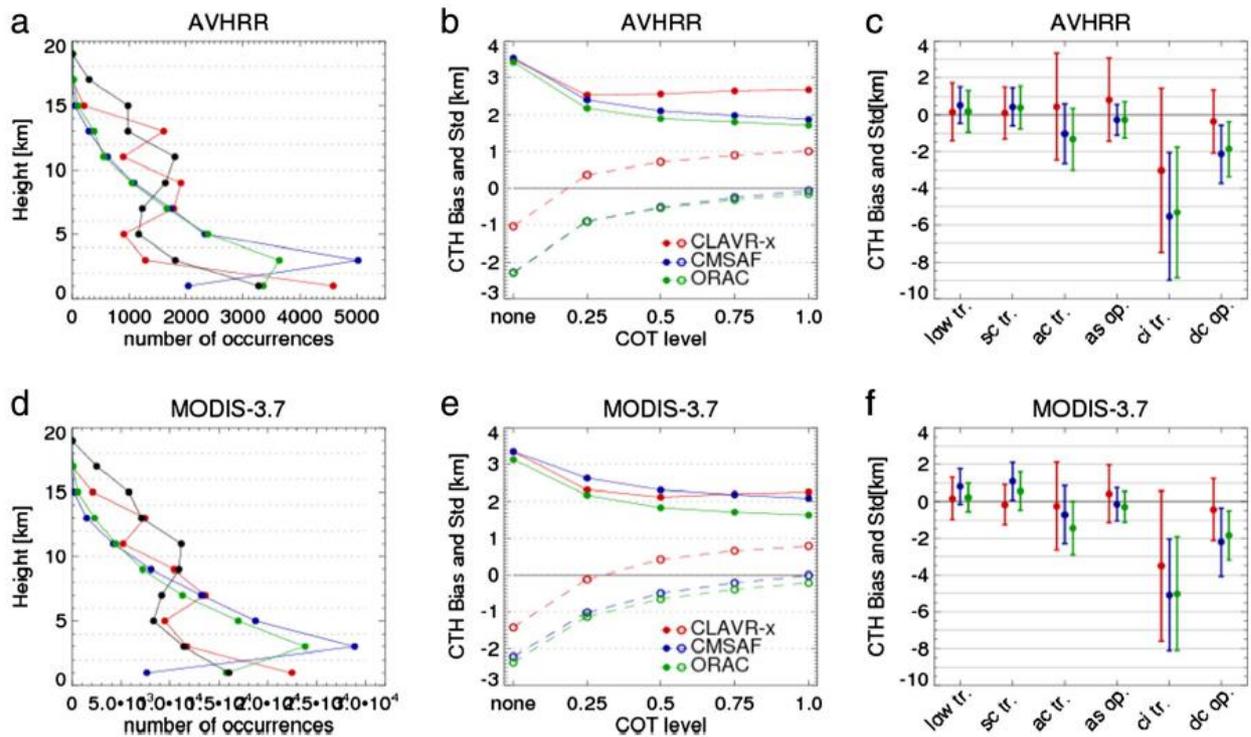
Figure 43: Left panels: One-dimensional histograms of cloud top height (CTH) derived from CALIPSO(black lines) and AVHRR (a) andMODIS-3.7 (d)with a bin size of 2 km. A COT threshold of 0.25 was applied to CALIPSO profiles. Middle panels: Dependence of standard deviation (Std, solid lines) and Bias (dashed lines) for AVHRR (b) and MODIS (e) CTH retrievals on CALIPSO CTH taken from different level-to-cloud-top COT with 'none' indicating the uppermost cloud level in CALIPSO. Right panels: Standard deviation (Std, shown as error bars) and bias for AVHRR (c) and MODIS (f) CTH retrievals for different cloud type (as given by CALIPSO) with low tr.: low overcast transparent, sc tr.: transition stratocumulus, ac. tr.: altocumulus transparent, as. op.: altostratus opaque, ci. tr.: cirrus transparent, dc. op.: deep convective opaque. Figure and Caption are taken from Stengel et al. (2013).
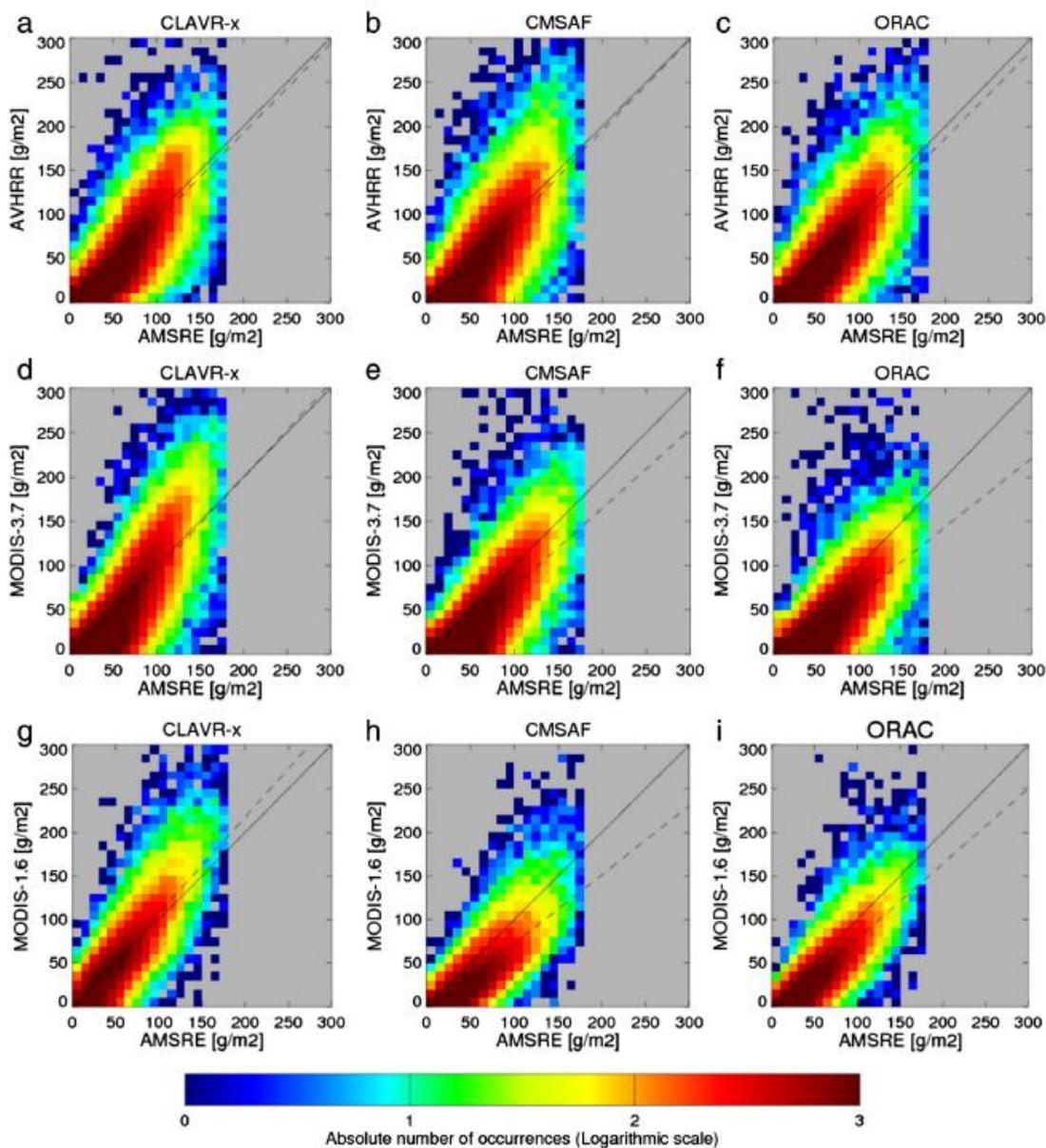
Figure 44: Two-dimensional histograms of AVHRR-based (top panels), MODIS-3.7 (middle panels), MODIS-1.6 (bottom panels) liquid water path (LWP) for CLAVR-x (left panels), CMSAF (middle panels) and ORAC (right panels) compared against the LWP derived from AMSR-E. Figure and Caption are taken from Stengel et al. (2013).
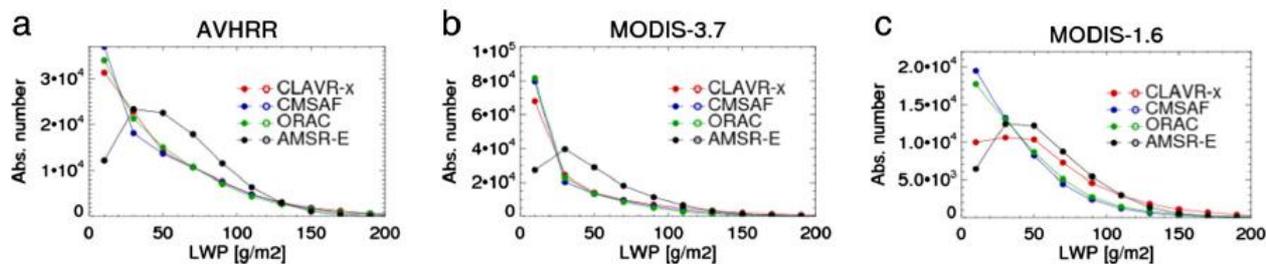
Figure 45: One-dimensional histograms of LWP retrievals for AVHRR (a), MODIS-3.7 (b), MODIS-1.6 (c) with AMSR-E data being also shown. Figure and Caption are taken from Stengel et al. (2013).
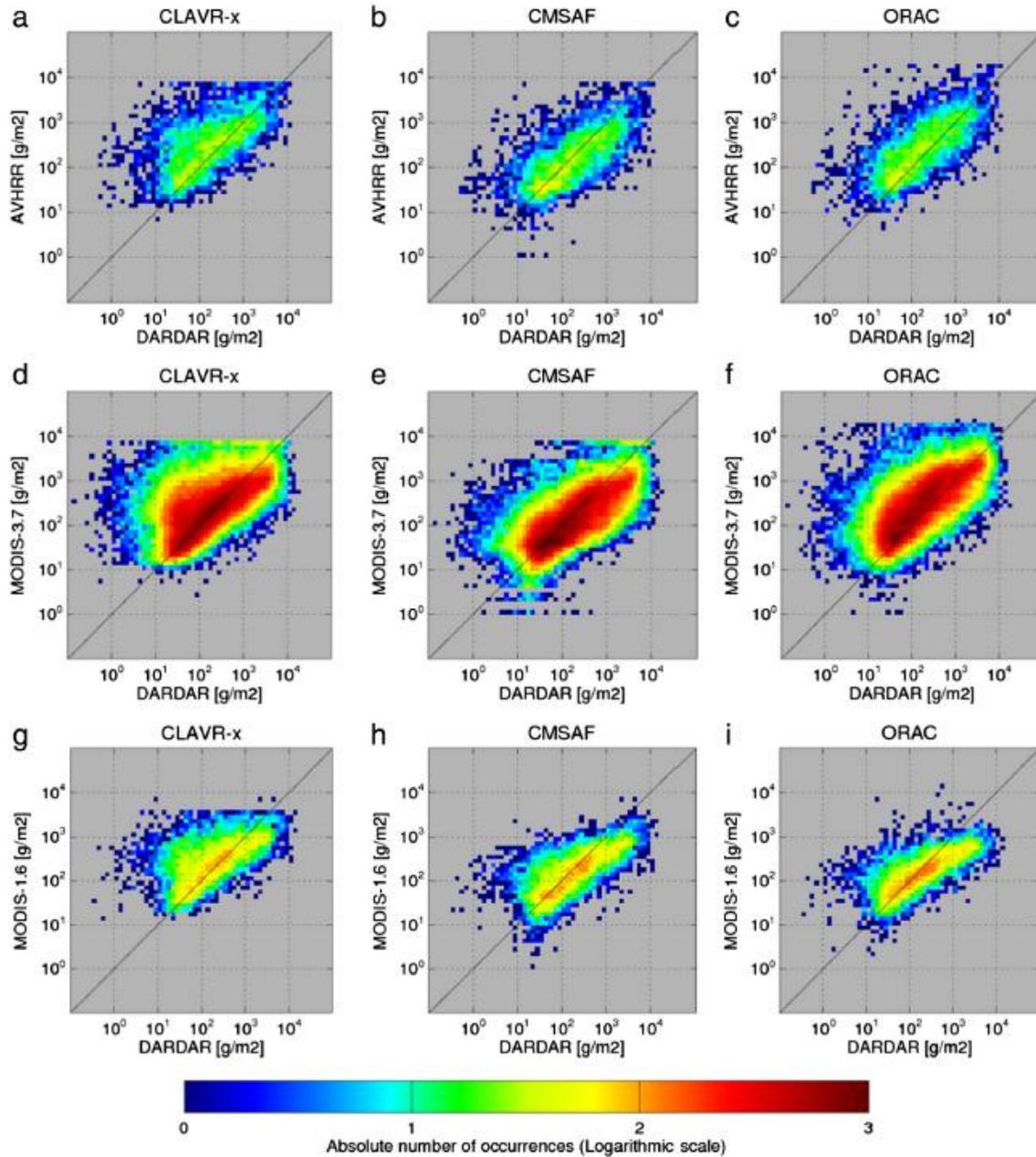
Figure 46: Two-dimensional histograms of AVHRR-based (top panels), MODIS-3.7 (middle panels), MODIS-1.6 (bottom panels) ice water path (IWP) for CLAVR-x (left panels), CMSAF (middle panels) and ORAC (right panels) compared against the IWP derived from DARDAR. Figure and Caption are taken from Stengel et al. (2013).
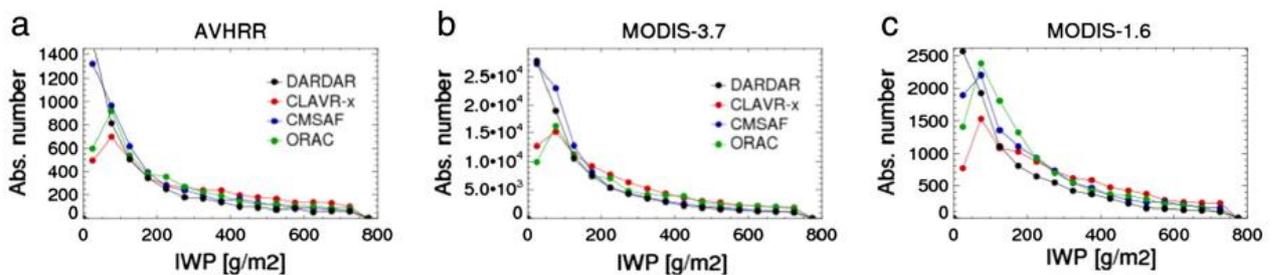
Figure 47: One-dimensional histograms of IWP retrievals for AVHRR (a), MODIS-3.7 (b), MODIS-1.6 (c) with DARDAR data being also shown. Figure and Caption are taken from Stengel et al. (2013).

5   Level-2: Assessment of cloud parameter uncertainty estimates

Courtesy: Ulrich Hamann (KNMI, MeteoSwiss), Jan Fokke Meirink (KNMI) and Phil Watts (EUMETSAT)

List of open questions for the retrieval uncertainty working groups (extendable)
- Can we provide uncertainty estimates?
- Which components of uncertainty are taken into account and how should they be reported?
- How can the cloud retrieval community go towards standard definitions for uncertainties?
- How can the uncertainty estimates be validated?
- What are first experiences with uncertainty assessments?
- Should CREW provide multi ensemble algorithm statistics and how can they be used?

There are plans to investigate the error estimate of the retrievals at the 4th CREW next year. The uncertainty estimate of the individual algorithms will be compared to the spread of the multi algorithm ensemble. Ideally the spread of the algorithm results should be small, when the error estimates of the individual algorithms is small, too. In this section we present a preliminary assessment of error estimates. We compare the error estimates as calculated by the OCA algorithm with the standard deviation of the multiple algorithm ensemble.

The retrieval uncertainty for many cloud physical parameters is dependent on the cloud optical depth. For example, a common method for the retrieval of the cloud optical depth and effective radius is the Nakajima-King approach (Nakajima, 1990) using one channel in the visible and one in the near infrared wavelength region. The relation of the cloud optical depth and the effective radius on the reflectivities is illustrated in Figure 48. The reflectivity of the absorbing channel increases less than the reflectivity of the non-absorbing channel with increasing optical depth. The near infrared channel has also a stronger sensitivity to the particle size as the non-absorbing channel. In Figure 48 the uncertainty of the reflectivities is schematically represented by a grey and a red shaded area. For clouds with a small optical depth, it is obvious that reflectivities of clouds with different effective radii are very similar. Furthermore, the bidirectional reflectance distribution function of the earth surface, that itself is not perfectly known, influences reflectivities for small optical depths. Therefore, we expect a large uncertainty of the effective radius, when the optical depth is small. For optically thick clouds, the change of the reflectivities with increasing optical depth is very small, but the reflectivity of the absorbing channels depends strongly on the effective radius. Hence we expect a high uncertainty for the retrieved optical depth and low uncertainty for the effective radius.

In the following we examine the standard deviation of the CREW algorithm ensemble and the error estimate of OCA algorithm. These datasets are different means to estimate the uncertainty of the retrieved effective radius. The error estimate of the OCA algorithm is based on the residual of the optimal estimation retrieval. Therefore, the algorithm takes into account the uncertainty of the satellite measurement, the estimated effect of cloud inhomogeneity and the surface reflectivity, but does not include the uncertainty of some retrieval assumptions like details of the parameterization of the optical properties of the cloud particles, in particular the choice of the ice crystal shape, remaining uncertainties from trace gas and temperature profiles as well as aerosols. The standard deviation of the multi algorithm ensemble partly includes these kinds of uncertainties, in case that different assumptions were made in the different retrievals. In case that all retrievals make the same assumption, the uncertainty of this assumption is not reproduced in the algorithm standard deviation neither. Considering this, we do not expect a one-to-one correspondence of the two uncertainty estimates, but nevertheless a positive correlation.
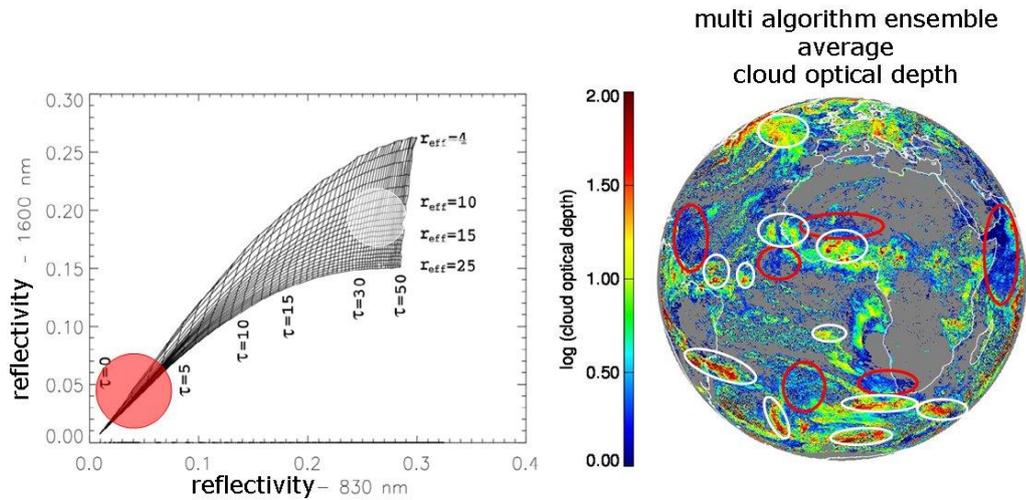
Figure 48: The Nakajima-King method describes the simultaneous retrieval of the effective radius and the optical depth with one non-absorbing channel (here 830nm) and one absorbing channel (here 1600nm), modified from Zinner (2005). Two uncertainty regimes are marked: In red the uncertainty for optically thin clouds and in white for optically thick clouds. On the right hand side the multi algorithm ensemble average of the optical depth is plotted for 13-06-2008 at 12:00UTC. The corresponding uncertainty regimes are marked in the same colors as in the Nakajima King plot.

Figure 49 shows the standard deviation of the effective radius of the CREW algorithm ensemble on the left hand side and the error estimate of the OCA algorithm on the right hand side. The same areas with optically thin clouds as in are marked. Both datasets indicate large uncertainties in the marked regions with optically thick clouds, as the discussion of the Nakajima-King suggests. At the southern edge of the SEVIRI disk there is a sharp increase of the uncertainty estimate of OCA. During the time of the observation, 2008-06-13 12UTC, the sun is close to or below the horizon in this region. Hence the OCA algorithm does not make use of the solar channels, and in consequence the uncertainty of the retrieval is larger in this region compared to the rest of the disk. The multi algorithm standard deviation increases here, too, but with a less sharp transition, as the different algorithms have different cutoff thresholds of the solar zenith angle for the usage of the solar channels. Both datasets show that the effective radii in marine strato-cumulus regions west of Angola and in the North Atlantic as well as over tropical Africa can be retrieved with the lowest uncertainty. In this region, the clouds are mainly water clouds. The retrieval of the effective radius of water droplets is easier than the one for ice crystals, as the shape of water droplets is well known, but there are various types of ice crystals and the assumptions in the retrieval methods of the ice crystal shape may differ. Another reason for the low retrieval uncertainty in these regions is the horizontal homogeneity of the marine strato-cumulus. Therefore no pronounced effects of the three dimensional radiative transfer or subpixel inhomogenities are expected here.
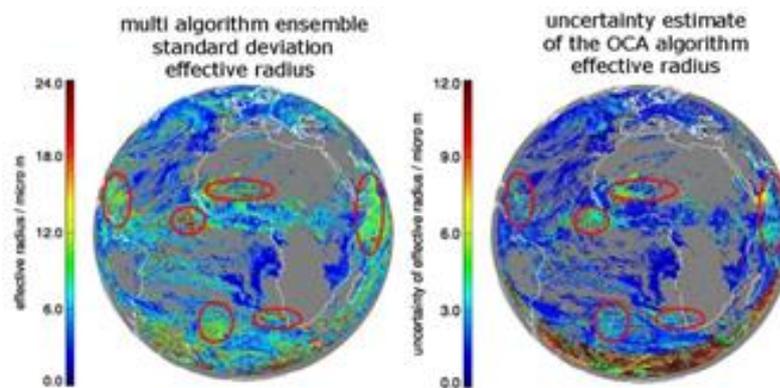
Figure 49: The left hand side shows the multi algorithm ensemble standard deviation of the effective radius for 2008-06-13 12UTC. On the right hand side the uncertainty estimate of the OCA algorithm is shown. The same areas of low optical depth as in Figure 48 are marked.

In Figure 50, we examine the error estimates of the cloud optical depth. The multi algorithm standard deviation is shown on the left hand side and the error estimate of the OCA algorithm on the right hand side. As expected from the discussion of the Nakajima-King method, the uncertainty for the retrieved cloud optical depth is largest for optically thick clouds, compare the marked areas in Figure 48 and Figure 50. This effect can be observed for both the OCA error estimate and the multi algorithm standard deviation. Furthermore the both uncertainties rise at the southernmost part of the SEVIRI disk for the same reason as for the effective radius, the lack of solar observations. Tendentially, lower retrieval uncertainties are noticed for clouds over the ocean. Two possible reasons are the good contrast between clouds and the dark ocean and the well known reflection properties of the ocean. In contrast, the reflection properties of various land surface types vary much more.

In summary, we conclude that the uncertainty for the effective radius retrieval is largest when the optical depth is small. Furthermore the uncertainty is larger for ice clouds than for water clouds. The uncertainty for the cloud optical depth is largest for optically thick clouds. The retrievals for both cloud properties have a higher uncertainty, if no solar observations are available. All findings can be observed in the multi algorithm standard deviation and in the uncertainty estimate of the OCA retrieval.
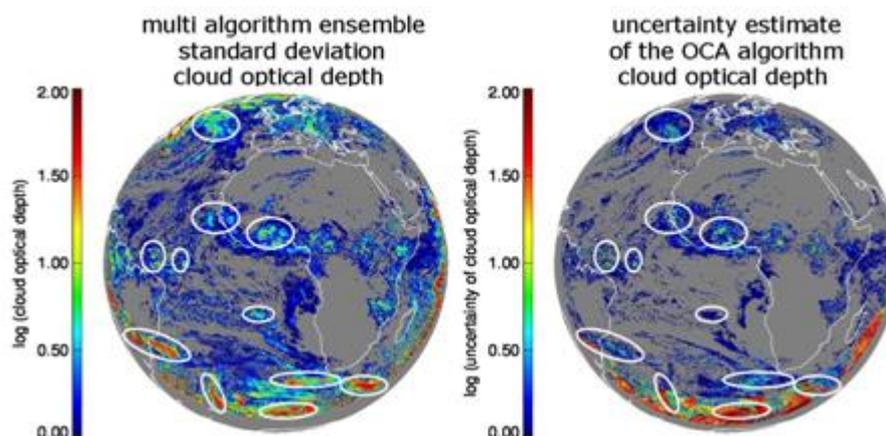
Figure 50: The left hand side shows the multi algorithm ensemble standard deviation of the cloud optical depth for 2008-06-13 12UTC. On the right hand side the uncertainty estimate of the OCA algorithm is shown. The same areas of high optical depth as in Figure 48 are marked.

In order to compare the uncertainty estimate of the OCA algorithm with a deviation between the OCA and the CPR data set a small sensitivity analysis was performed.
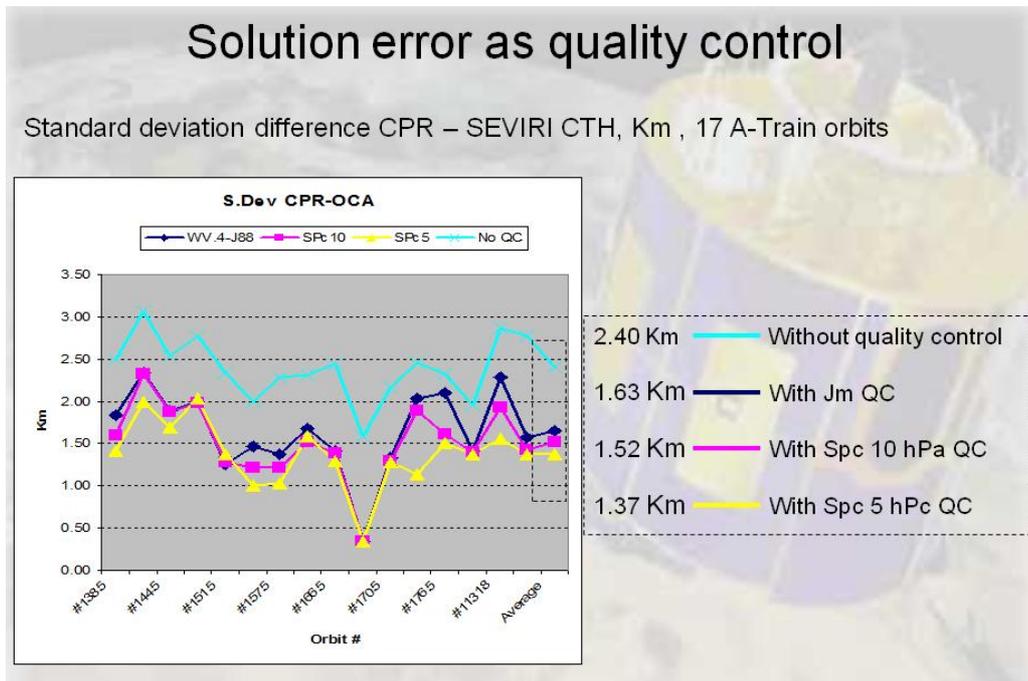
Figure 51: The figure shows standard deviations of retrieved CTH from SEVIRI compared to a small selection of CPR orbits. Results without quality control filtering applied (cyan) show the effect of handling (in OCA) multi-layer cloud with a single layer cloud model. By applying a filter on the solution cost (Jm) these are effectively removed. Subsequent application of a filter on the estimated CTH error further reduces the standard deviations in several orbits. Results after a filter to remove cases with an estimated error of >10 hPa are shown magenta, and with an estimated error >5 hPa are shown yellow.

The threshold values (5 and 10 hPa) have to be (much) smaller than the apparent errors (~1.5 km or therefore ~70 hPa) to have any effect. This is partly no doubt due to unavoidable collocation, sampling and parallax effects on the comparison of CPR and SEVIRI, but it is also true that the estimated error can only to date modelling a few of the known error sources in the retrieval system.

6   Level 3: Evaluation of Aggregation Methods

Courtesy: Ralf Bennartz (SSEC), Nadia Smith (SSEC), Theo Steenbergen (EUMETSAT), Brent Maddux, and Rob Roebeling (EUMETSAT)

List of open questions for the Level 2 to 3 aggregation methods working groups (extendable)
- Who are the users that we are addressing?
- How can we better serve the weather and climate model community?
- What are the most important cloud modeling problems in climate science and how can cloud observation help to understand them?
- What is our current ability to estimate climatologic trends of cloud properties?
- What is the best practice to generate a long term cloud data record (e.g. continuity between different sensors, or how to deal with sensor degradation)?
- How can we, or should we, conserve consistency between the Level 2 and Level 3 products for the different cloud parameters?
- What are the key elements to consider in the aggregation of geophysical data?
- Are there any useful uncertainty/error estimates to use in aggregation methods?
- How can we use a reference climate model as background (ERA-Interim) for testing the performance of an aggregation method?
- What systematic errors can a reference climate model help identify?
- How do filtering rules (nr. of observations per grid box, viewing and solar angles, cloud mask, day/night) affect the level 3 results?
- What are the optimal space and time aggregation settings?
- How should we use and represent the number of observations used in the aggregated product relative to the total number of available observations?

6.1   Comparison of Level-3 aggregations using identical Level-2 data

In this section we compare three Level-3 products for the month September 2012. The three level 3 products were derived from the same Level-2 products, however, using different aggregation methods. The monthly Level-3 products are compared for CTP, CTT, CER, and COT. We restricted our comparison to products that were aggregated on a 1 degree rectangular grid. The three Level-3 products evaluated are:

1. MODIS MAIN: This is the Level-3 product aggregated at EUMETSAT using a classical averaging approach
2. MODIS LAADS: The Level-3 products as provided by the MODIS LAADS team
3. MODIS GEWEX-CA: The Level-3 product generated with the aggregation methods prepared for and used within the GEWEX-CA.

The Level-3 products have been generated from MODIS Level 2 products. The MODIS L2 products (Collection 5.1) have been downloaded per ftp from the NASA LAADS portal. The availability of the downloaded files (nominally 288 files per day) was 99.64%; files were missing for 13 Sep (15 files) and 27 Sep (16 files).

Statistics are provided for three segments (sea/ocean, land and coastal area), as well as all three combined. To distinguish the coastal area, a distance of 150 km over sea/ocean to the nearest shore line has been used.

Notes:

- The number of observations for the MODIS CREW products has been derived by counting the number of individual scans per grid box. No threshold value to the number of observations has been used.
- The number of observations is not included in the MODIS LAADS products for CER and COT.
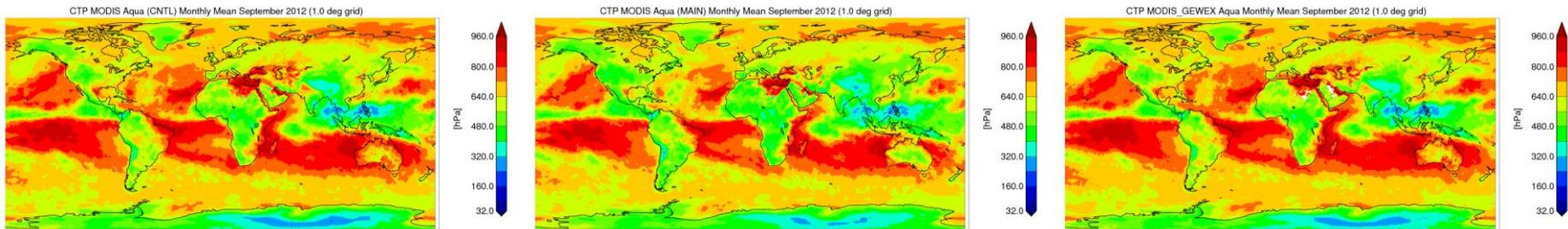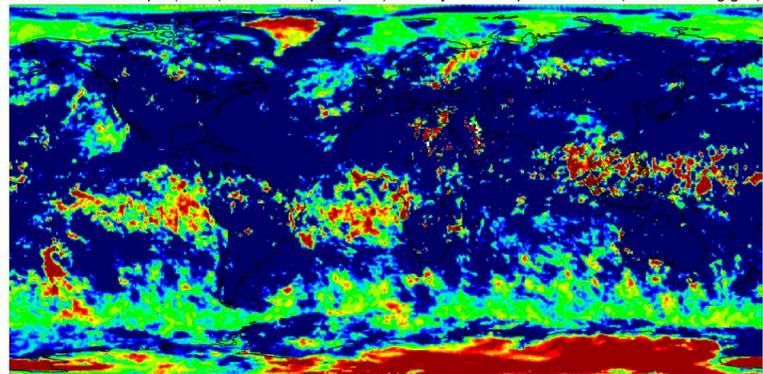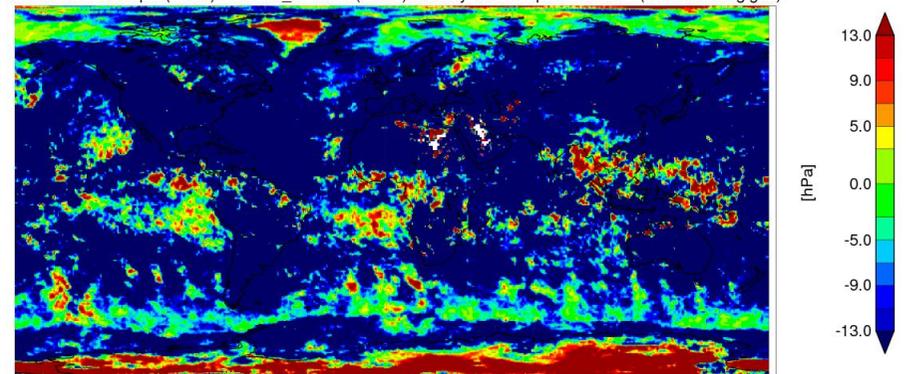
Figure 52: Upper panel monthly mean CTP for September 2012 from i) MODIS LAADS, ii) MODIS EUMETSAT, and iii) MODIS GEWEX. Lower panel shows the difference MODIS EUMETSAT- MODIS LAADS (Left) and MODIS EUMETSAT- MODIS GEWEX (Right).

Table 1: Inter-comparison results monthly mean CTPs for September 2012 from MODIS EUM (EUMETSAT) relative to MODIS LAADS or MODIS GEWEX.

	Min (hPa)	Max (hPa)	Mean (hPa)	Std Dev (hPa)	Mean N_obs	Bias (hPa)	RMSE (hPa)	Valid Grids (%)
Sea/Ocean								
EUM	265.1	971.6	680.1	100.9	12,314	-	-	-
LAADS	260.5	967.9	695.1	100.3	12,355	-15.0	24.7	58
GEWEX	265.9	972.3	699.7	100.2	12,355	-19.6	29.8	58
Coast								
EUM	261.6	966.7	616.7	107.6	12,031	-	-	-
LAADS	254.8	959.6	644.2	108.8	12,074	-27.4	42.4	17
GEWEX	257.2	957.0	650.3	109.3	12,075	-33.6	48.6	17
Land								
EUM	242.4	924.8	517.6	103.0	7,897	-	-	-
LAADS	237.8	930.4	537.9	119.6	7,940	-20.4	46.7	25
GEWEX	247.3	922.5	543.0	120.4	8,270	-23.7	50.0	24
All								
EUM	242.4	971.6	628.9	123.2	11,167	-	-	-
LAADS	237.8	967.9	647.4	125.4	11,209	-18.4	34.7	100
GEWEX	247.3	972.3	653.5	125.1	11,323	-23.0	39.1	99
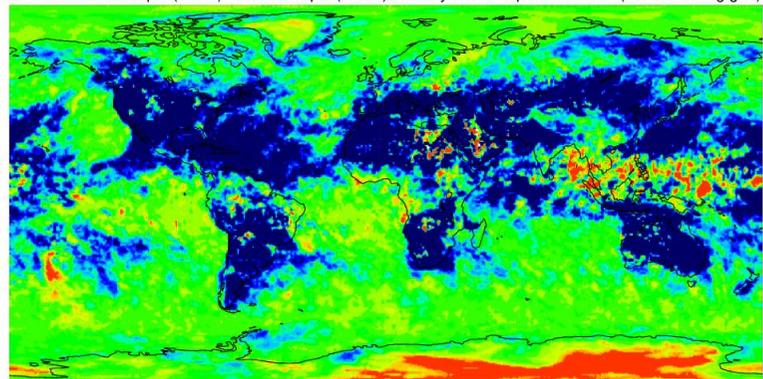
Figure 53: Upper panel monthly mean CTT for September 2012 from i) MODIS LAADS, ii) MODIS EUMETSAT, and iii) MODIS GEWEX. Lower panel shows the difference MODIS EUMETSAT- MODIS LAADS (Left) and MODIS EUMETSAT- MODIS GEWEX (Right)

Table 2: Inter-comparison results monthly mean CTTs for September 2012 from MODIS EUM (EUMETSAT) relative to MODIS LAADS or MODIS GEWEX.

	Min (K)	Max (K)	Mean (K)	Std Dev (K)	Mean N_obs	Bias (K)	RMSE (K)	Valid Grids (%)
Sea/Ocean								
EUM	223.5	296.4	263.3	13.75	12,314	-	-	-
LAADS	222.3	296.7	264.5	14.06	12,355	-1.22	2.25	58
GEWEX	223.7	296.6	264.9	14.03	/	-1.62	2.65	58
Coast								
EUM	223.4	301.8	258.3	12.26	12,031	-	-	-
LAADS	223.4	302.1	260.7	13.14	12,074	-2.41	3.95	17
GEWEX	169.9	301.9	261.2	13.22	/	-2.88	4.58	17
Land								
EUM	206.8	297.9	245.5	19.54	7,896	-	-	-
LAADS	202.6	301.7	247.6	21.89	7,939	-2.06	4.57	25
GEWEX	135.4	299.5	248.7	22.02	/	-2.10	7.88	24
All								
EUM	206.8	301.8	258.0	16.90	11,166	-	-	-
LAADS	202.6	302.1	259.7	17.72	11,208	-1.63	3.28	100
GEWEX	135.4	301.9	260.4	17.55	/	-1.95	4.77	99

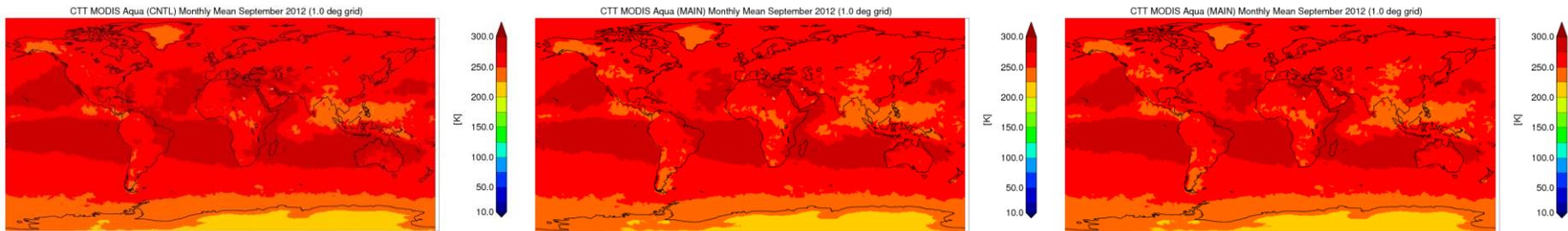**Figure 54: Upper panel monthly mean COT for September 2012 from i) MODIS LAADS, ii) MODIS EUMETSAT, and iii) MODIS GEWEX. Lower panel shows the difference MODIS EUMETSAT- MODIS LAADS (Left) and MODIS EUMETSAT- MODIS GEWEX (Right)**

Table 3: Inter-comparison results monthly mean COTs for September 2012 from MODIS EUM (EUMETSAT) relative to MODIS LAADS or MODIS GEWEX.

	Min (-)	Max (-)	Mean (-)	Std Dev (-)	Mean N_obs	Bias (-)	RMSE (-)	Valid Grids (%)
Sea/Ocean								
EUM	0.72	100.00	14.13	7.36	/	-	-	-
LAADS	0.77	100.00	15.19	8.72	/	-1.07	7.72	58
GEWEX	0.77	100.0	9.24	5.42	/	+4.89	8.35	58
Coast								
EUM	0.50	100.00	14.66	6.63	/	-	-	-
LAADS	1.38	65.75	18.49	7.94	/	-3.83	9.11	16
GEWEX	1.29	40.27	10.85	4.28	/	+3.81	7.67	16
Land								
EUM	0.05	100.00	15.71	9.73	/	-	-	-
LAADS	0.94	66.48	18.77	7.76	/	-3.06	11.76	17
GEWEX	0.94	45.39	10.84	4.34	/	+4.86	11.59	17
All								
EUM	0.05	100.00	14.52	7.76	/	-	-	-
LAADS	0.77	100.00	16.45	8.57	/	-1.93	8.86	91
GEWEX	0.77	100.00	9.83	5.10	/	+4.69	8.93	91

CER MODIS Aqua (MAIN) Monthly Mean September 2012 (1.0 deg grid)
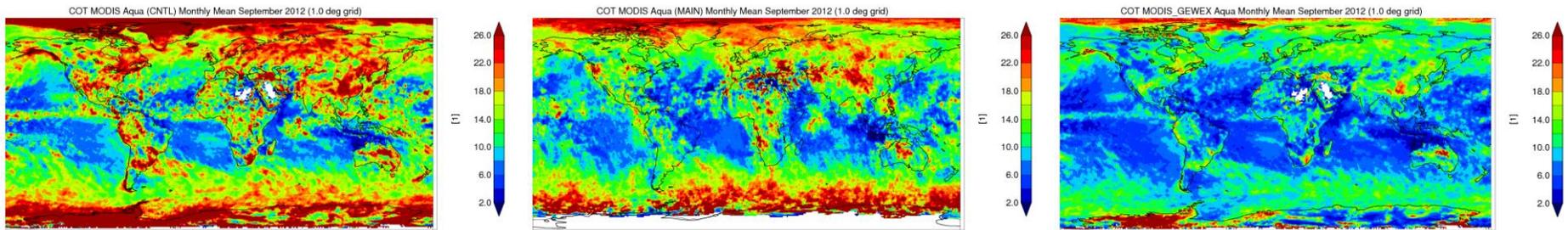
CER MODIS Aqua (CNTL) Monthly Mean September 2012 (1.0 deg grid)

Delta CER MODIS Aqua (MAIN) - MODIS Aqua (CNTL) Monthly Mean September 2012 (1.0 vs 1.0 deg grid)
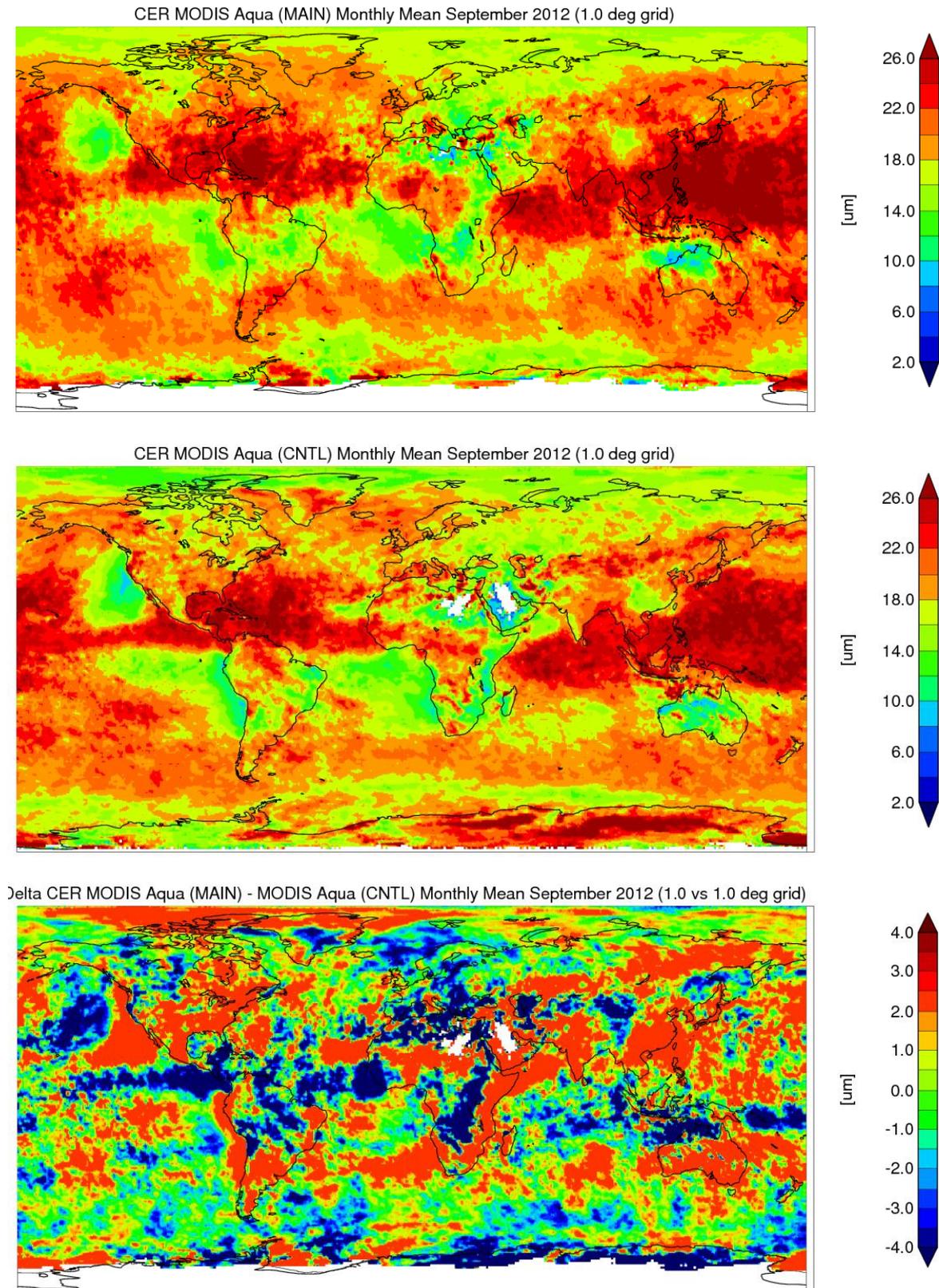
Figure 55: Monthly mean CER from MODIS MAIN (Top), MODIS (LAADS) (Middle), and the difference MODIS MAIN - MODIS LAADS (Lower).

Summary Report 2014 on the Inter-comparison and Validation studies of CREW (Cloud Retrieval Evaluation Workshop)

Table 4: Inter-comparison results of CER for MODIS CREW vs MODIS LAADS, September 2012 monthly mean.

	Min (K)	Max (K)	Mean (K)	Std Dev (K)	Mean N_obs	Bias (K)	RMSE (K)	Valid grids (%)
Sea/Ocean								
MAIN	7.51	43.63	19.67	3.62	/			
LAADS	4.60	34.53	19.53	3.53	/	+0.14	2.87	58
Coast								
MAIN	7.20	40.08	19.22	3.43	/			
LAADS	7.10	37.31	18.74	3.36	/	+0.48	3.63	16
Land								
MAIN	5.86	40.27	18.55	3.34	/			
LAADS	4.08	37.67	18.14	3.83	/	+0.40	4.81	17
All								
MAIN	5.86	19.38	3.56	2.77	/			
LAADS	4.08	19.13	3.60	2.79	/	+0.25	3.45	91

6.2   Analysis of Level-3 aggregations sensitivities

Courtesy: Nadia Smith, Ralf Bennartz

Aggregation is a mechanism for reducing data into descriptive information that depict the patterns and processes of geophysical phenomena over time. It is a means with which to project satellite retrievals (Level-2) from their instrument-specific domain to a uniform space-time domain (Level-3). An aggregation method can vary greatly depending on the parameter or application in question. The key elements (decision steps) that make up a robust uniform space-time aggregation method are described in Smith et al. (2013). The following results characterize Level-3 data by highlighting some of their sensitivities to differences in aggregation methods.

Figure 47 addresses the bias that could result from using a uniform grid in data evaluation and comparisons. CrIS brightness temperature (BT) is convolved to VIIRS band M15 (10.8 µm) and aggregated into a mean with standard deviation (SD) on a 1-degree grid (Figure 47a). Similarly, VIIRS BT from band M15 is aggregated into a mean with SD on a 1-degree grid (Figure 47b).
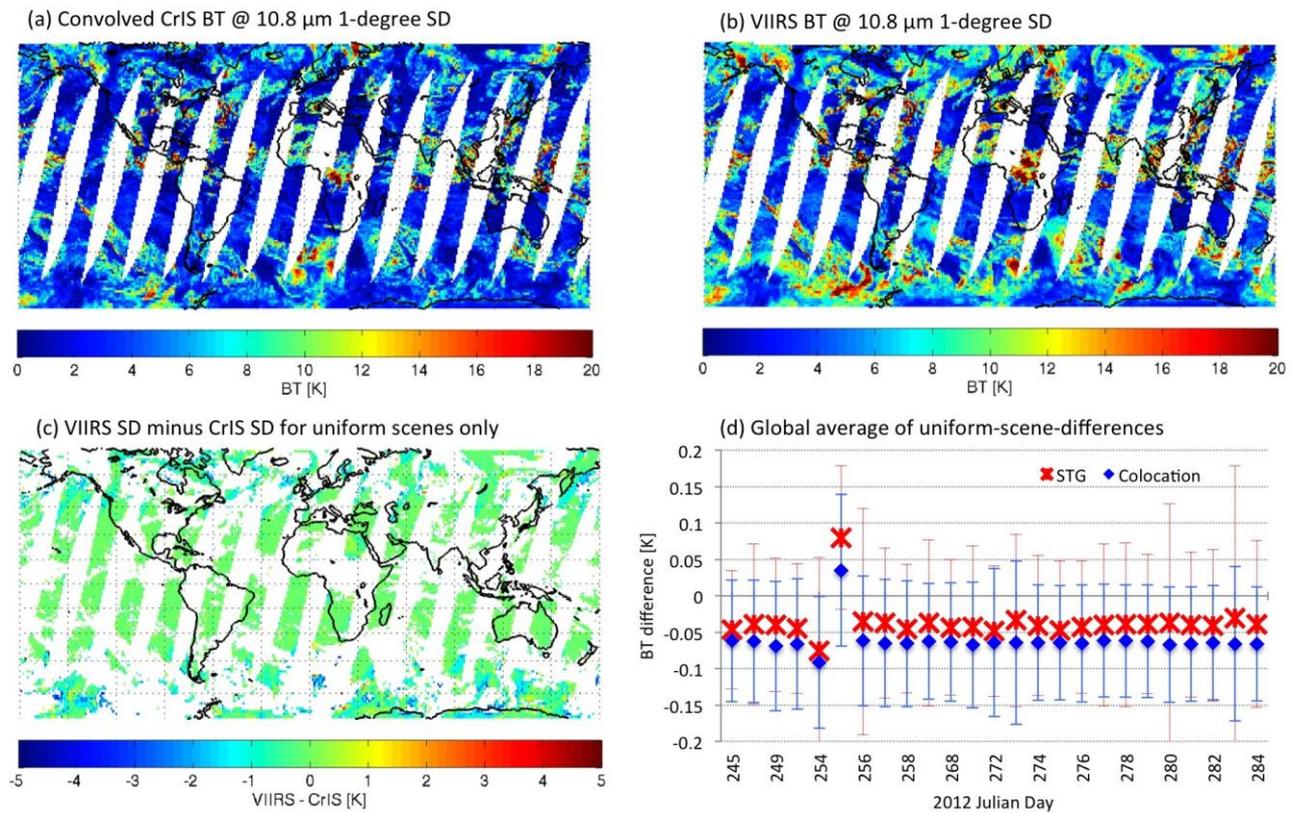
Figure 47: Colocation bias caused by 1-degree uniform equal-angle global grid. (a) Standard deviation (SD) of CrIS brightness temperature (BT) spectrum convolved to VIIRS band M15 (10.8 µm). (b) SD of VIIRS BT for band M15 (10.8 µm). (c) Difference of VIIRS SD (from b) and CrIS SD (from a) for uniform scenes (identified as those grid cells with a SD < 1K). (d) Daily global average of SD difference (from c, indicated as the red star), compared to the same procedure but from a one-on-one instrument-specific colocation scheme (indicated as the blue diamond).

For each instrument, the SD is evaluated and grid cells are filtered to include uniform (cloudy or clear) scenes only. The latter are cells with a SD < 1K. A global average is then calculated from the uniform-scene difference (VIIRS mean BT minus CrIS mean BT, Figure 47c). The latter is then

compared to the global average of uniform-scene-differences derived from a one-on-one instrument specific colocation scheme (VIIRS pixels are collocated exactly to each CrIS field-of-view). By evaluating BT, as apposed to a retrieval parameter, we are able to capture instrument differences more acutely. These results indicate that for zonal (or global) statistics, if aggregation is considered carefully and done correctly, the co-location bias caused by a uniform equal-angle grid (as apposed to a one-on-one instrument-specific colocation scheme) is negligible.

The propagation of uncertainty from Level-2 (L-2) to Level-3 (L-3) is examined in Figure 48 for retrievals of cloud effective radius (for water clouds only) from MODIS Collection 5 (C5) products. The term "uncertainty" is an expression of doubt regarding the correctness of the result. Uncertainty can result from both systematic and random effects. L-2 uncertainty is determined here by the retrieval product uncertainty flag (a by-product of the retrieval method), whereas L-3 uncertainty is determined by the standard deviation (SD) about the mean per uniform grid cell. By changing the threshold with which L-2 retrievals are filtered (prior to aggregation), the propagation of uncertainty can be measured. The results below demonstrate the sensitivity to quality filtering both in terms of the magnitude of propagated uncertainty and in the spatial coverage of the L-3 result. L-2 input with higher (lower) uncertainty results in L-3 with higher (lower) uncertainty and larger (smaller) coverage.
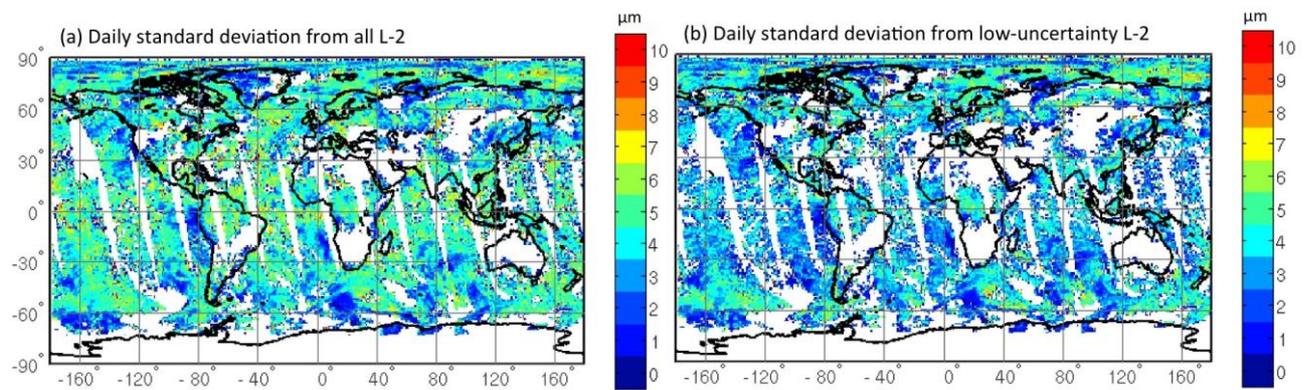

Figure 48: Propagation of uncertainty from Level-2 (L-2) to Level-3 (L-3) on a 1-degree grid. L-2 retrievals are cloud effective radius (for water clouds only) from MODIS Collection 5 (C5) with a retrieval uncertainty flag. Grid cell standard deviation (SD) is a measure of the L-3 uncertainty. (a) Daily Standard Deviation (SD) using all L-2 retrievals (no filtering). (b) Daily SD using low-uncertainty L-2 retrievals only (filtering retrievals with uncertainty < 10%).

The effect of grid resolution on the calculation of zonal statistics is examined in Figure 49. A day of daytime MODIS C5 CTP retrievals was aggregated into high cloud (< 440hPa) statistics on 1-, 2-, and 10-degree grids, respectively. For each of these sets, zonal statistics were calculated, namely a mean of all grid cell means, and a mean of all grid cell SD (error bars in Figure 49). The four zones considered here were Northern Hemisphere (NH) mid-latitude (MidLat, 60˚ to 30˚), NH Tropics (30˚ to 0˚), Southern Hemisphere (SH) Tropics (0˚to -30˚), and SH MidLat (-30˚ to -60˚). The error bars depict the average uncertainty associated with each zone. As expected the 10-degree resolution grid cells have the highest average uncertainty. However, the difference in zonal means from grid cells with different spatial resolutions is not significant for CTP.

This said, the reader should bear in mind that, grid cell estimates (or means) at 10-degree spatial resolution will require much larger (or widespread) change before a change in the grid mean is observed, i.e., the 10-degree zonal means will depict much smaller changes over time than the 1-degree zonal means. If the objective is to use L-3 products to map change at high temporal frequency, then a smaller grid resolution should be used instead.
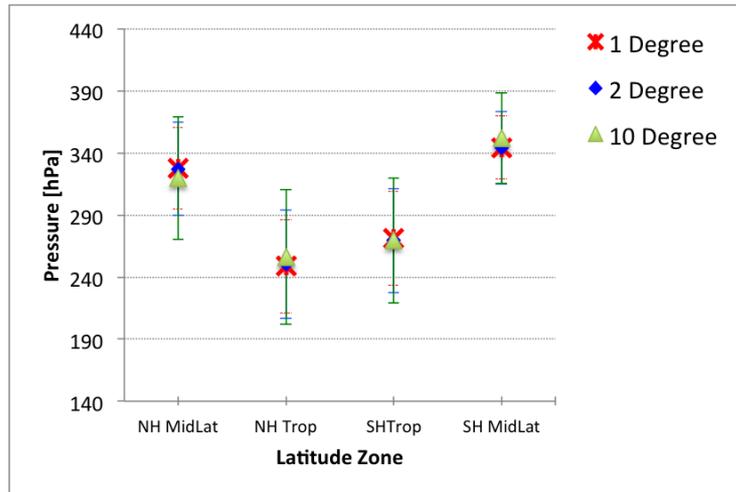
Figure 49: Sensitivity of zonal statistics to grid cell size. Results are for a day of cloud top pressure (CTP) retrievals from MODIS C5 products subset to high clouds (< 440 hPa) and aggregated to 1-, 2-, and 10-degree resolution grids, respectively. For each grid, a mean and SD of high clouds is calculated. Zonal statistics are calculated for Northern Hemisphere (NH) mid-latitude (MidLat, 60° to 30°), NH Tropics (30° to 0°), Southern Hemisphere (SH) Tropics (0° to -30°), and SH MidLat (-30° to -60°). A mean of grid cell means (zonal mean) and mean of grid cell SD (error bars) are calculated for each zone using grids of 1-, 2-, and 10-degrees respectively.

Figure 50 addresses the question of a monthly mean. Should a monthly mean be calculated when only a few daily means are available? How many days does it take to make up a monthly mean?
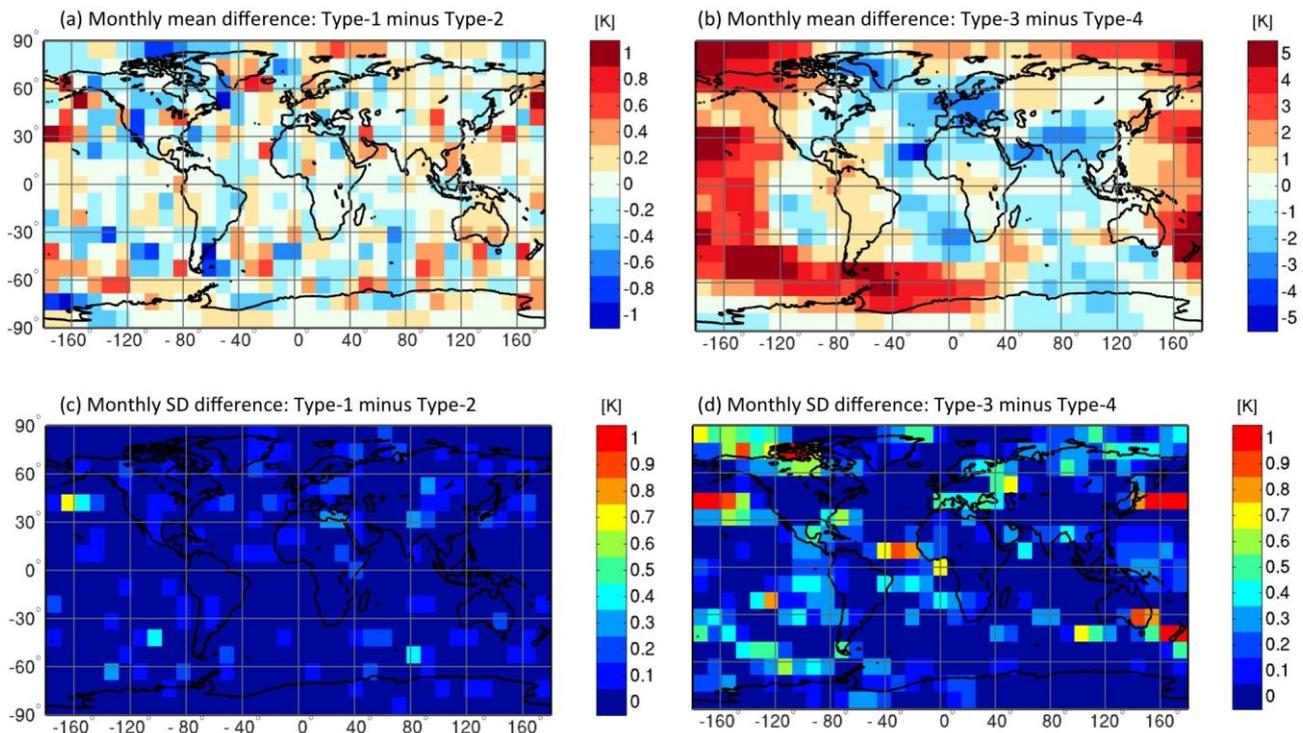
Figure 50: Sensitivity of monthly means to aggregate of daily means. Type-1 is a monthly mean from all even-numbered days in February 2012. Type-2 is from all odd-numbered days. Type-3 is from the first 15 days, and Type-4 is from the last 15 days. (a) Difference between Type-1 and Type-2 monthly means, (b) Difference between Type-3 and Type-4 monthly means, (c) Difference between Type-1 and Type-2 monthly mean SD, (d) Difference between Type-3 and Type-4 monthly mean SD.

Four types of monthly means for February 2012 were investigated; Type-1 is a monthly mean from all even-numbered days, Type-2 is a monthly mean from all odd-numbered days, Type-3 is a monthly mean from 1–14 February (first half), Type-4 is a monthly mean from 15–29 February (second half). Figure 50 (a) and (c) depict the difference of Type-1 minus Type-2 monthly mean and SD, respectively. Figure 50 (b) and (d) depict the difference of Type-3 minus Type-4 monthly mean and SD, respectively. These results demonstrate a strong dependence on the sample of days used to represent a monthly mean. If days are distributed evenly across the month, then random differences arise in the comparison of two different L-3 products. However, if a limited sequence of days is used, and the sequence differs between two L-3 products, systematic differences arise. A mean from a sequence of days can capture an extreme weather event that that misrepresents the monthly mean conditions. This leads to the conclusion that if a few days of data are available for a month and they are closely space together in time, then a monthly mean should not be calculated because systematic differences can be introduced. However, if the same number of days is distributed evenly across the time period, a monthly mean can be calculated. Moreover, monthly means from different sources can only be compared if the sequence of days making up the monthly mean from each source is distributed evenly over the time period (differences between the products will be due to random effects only). The monthly mean SD is a measure of the average uncertainty associated with the L-3 product. A difference of monthly mean SD between two products, highlight the differences in uncertainty between L-3 products. The lowest and most uniform L-3 uncertainty is for a monthly mean of evenly spaced daily means. The largest uncertainty is for a monthly mean from a short sequence of days.

The use of standard statistical methods helps characterize and quantify uncertainty in L-3 products. A "best practice" aggregation method should be instrument-independent, preserve information, reduce uncertainty (due to both random and systematic effects), as well as generate products that are at a space-time scale appropriate to the application. Scale greatly affects the type and magnitude of geophysical change that can be detected over time. As demonstrated here, the space-time-gridding (STG) framework described in Smith et al. (2013) allows for a transparent, iterative and dynamic approach to the development of "best practices" for generating cloud data records.

7   References

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

Bugliaro, L., Zinner, T., Keil, C., Mayer, B., Hollmann, R., Reuter, M., and Thomas, W.: Validation of cloud property retrievals with simulated satellite radiances: a case study for SEVIRI, Atmos. Chem. Phys, 11, 5603–5624, 2011.

Chang, F.-L., Minnis, P., Lin, B., Khaiyer, M. M., Palikonda, R., and Spangenberg, D. A.: A modified method for inferring upper troposphere cloud top height using the GOES 12 imager 10.7 and 13.3 µm data, Journal of Geophysical Research, 115, D06 208, 2010.

Cess, R. D., Potter, G., Blanchet, J., Boer, G., Ghan, S., Kiehl, J., Le Treut, H., Li, Z.-X., Liang, X.-Z., Mitchell, J., et al.: Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models, Science, 245, 513–516, 1989.

Derrien, M. and Le Gléau, H.: MSG/SEVIRI cloud mask and type from SAFNWC, International Journal of Remote Sensing, 26, 4707–4732, 2005.

Derrien, M. and Le Gléau, H.: Improvement of cloud detection near sunrise and sunset by temporal-differencing and region-growing techniques with real-time SEVIRI, International Journal of Remote Sensing, 31, 1765– 1780, 2010.

Derrien, M. and Le Gléau, H.: Algorithm Theoretical Basis Document for "Cloud Products" (CMa-PGE01 v3.2,27 CT-PGE02 v2.2 & CTTH-PGE03 v2.2), NWC SAF document: SAF/NWC/CDOP/MFL/SCI/ATBD/01, 2013.

Dybbroe, A., et al.: NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling - Part I: Algorithm description, J. Appl. Meteor, 44, pp. 39-54, 2005a.

Dybbroe, A., et al.: NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling - Part II: Tuning and validation, J. Appl. Meteor, 44, 55-71, 2005b.

Ewald, F., Bugliaro, L., Mannstein, H., and Mayer, B.: An improved cirrus detection algorithm MeCiDA2 for SEVIRI and its evaluation with MODIS, Atmospheric Measurement Techniques, 6, 309–322, 2013.

Eyre, J. R. and Menzel, W. P.: Retrieval of cloud parameters from satellite sounder data: A simulation study., Journal of Applied Meteorology, 28, 267–275, 1989.

Francis, P., Hocking, J. and Saunders, R. W.: Improved diagnosis of low-level cloud from MSG SEVIRI data for assimilation into Met Office limited area models, Proceedings of the 2008 EUMETSAT Meteorological Satellite Conference, Darmstadt, 2008.

Hamann, U., Walther, A. Baum, B., Bennartz, R., Bugliaro, L., Derrien, M, Francis, P, Heidinger, A., Joro, S., Kniffka, A., Le Gléau, H., Lockhoff, M., Lutz, H.-J., Meirink, J. F., Minnis, P., Palikonda, R., Roebeling, R., Thoss, A., Platnick, S., Watts, P. and Wind, G.: Remote sensing of the cloud top pressure/height from SEVIRI: Analysis of ten current retrieval algorithms. Atmos. Meas. Tech. Discuss., 7, 401–473, 2014, www.atmos-meas-tech-discuss.net/7/401/2014/, doi:10.5194/amtd-7-401-2014, 2014.

Heidinger, A. K. and Pavolonis, M. J.: Gazing at cirrus clouds for 25 years through a split window. Part I: Methodology, Journal of Applied Meteorology and Climatology, 48, 1100–1116, 2009.

Heidinger, A., Pavolonis, M., Holz, R., Baum, B. A., and Berthier, S.: Using CALIPSO to explore the sensitivity to cirrus height in the infrared observations from NPOESS/VIIRS and GOES-R/ABI, Journal of Geophysical Research: Atmospheres (1984–2012), 115, 2010.

Heidinger, A., et al.: Deriving an intersensor consistent calibration for the AVHRR solar reflectance data record, Int. J. Remote Sensing, 31, 6493–6517, doi:10.1080/01431161.2010.496472, 2010.

King, M. D., Platnick, S., Hubanks, P. A., Arnold, G. T., Moody, E. G., Wind, G., and Wind, B.: Collection 005 change summary for the MODIS cloud optical property (06 OD) algorithm, MODIS Atmosphere, p. 8, 2006.

Lutz, H.-J., Wilson, H. K., Heinemann, T., Samain, O., Gustafsson, J., Holmlund, K., and de Smet, A.: MSG Meteorological Products Extraction Facility Algorithm Specification Document, Eumetsat Document Nr. EUM/MSG/SPE/022, 2011.

Meerkötter, R. and Bugliaro, L.: Diurnal evolution of cloud base heights in convective cloud fields from MSG/SEVIRI data, Atmos. Chem. Phys, 9, 1767–1778, 2009.

Meirink, J.F., R.A. Roebeling and P. Stammes, Inter-calibration of polar imager solar channels using SEVIRI, Atmos. Meas. Tech., 6, 2495-2508, doi:10.5194/amt-6-2495-2013, 2013.

Menzel, W., Smith, W., and Stewart, T.: Improved cloud motion wind vector and altitude assignment using VAS, Journal of Climate and Applied meteorology, 22, 377–384, 1983.

Menzel, W. P., Frey, R. A., Zhang, H., Wylie, D. P., Moeller, C. C., Holz, R. E., Maddux, B., Baum, B. A., Strabala, K. I., and Gumley, L. E.: MODIS global cloud-top pressure and amount estimation: Algorithm description and results, Journal of Applied Meteorology and Climatology, 47, 1175–1198, 2008.

Minnis, P., Yost, C. R., Sun-Mack, S., and Chen, Y.: Estimating the top altitude of optically thick ice clouds from thermal infrared satellite observations using CALIPSO data, Geophysical Research Letters, 35, 2008.

Minnis, P., Sun-Mack, S., Trepte, Q. Z., Chang, F.-L., Heck, P. W., Chen, Y., Yi, Y., Arduini, R. F., Ayers, K., Bedka, K., Bedka, S., Brown, R., Gibson, S., Heckert, E., Hong, G., Jin, Z., Palikonda, R., Smith, R., Smith, W. L. J., Spangenberg, D. A., Yang, P., Yost, C. R., and Xie, Y.: CERES Edition 3 Cloud Retrievals, AMS 13th Conf. Atmos. Rad. 2010, Portland, 2010.

Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., Spangenberg, D. A., Arduini, R. F., Trepte, Q. Z., Smith, W. L., Ayers, J. K., Gibson, S. C., Miller, W. F., Chakrapani, V., Takano, Y., Liou, K.-N., Xie, Y., and Yang, P.: CERES edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data, Part I: Algorithms, IEEE Trans. Geosci. Remote Sens., 49, 4374–4400, 2011.

Moseley, S.: Changes to the Nimrod cloud top height diagnosis, Met Office Forecasting Research Technical Report, 424., 2003.

Pavolonis, M. J., et al.: Daytime global cloud typing from AVHRR and VIIRS: Algorithm description, validation, and comparisons, J. Appl. Meteor., 44(6), 804-826, 2005

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Rieˊdi, J. C., and Frey, R. A.: The MODIS cloud products: Algorithms and examples from Terra, IEEE Transactions on Geoscience and Remote Sensing, 41, 459–473, 2003.

Poulsen C. A., et al.: Cloud retrievals from satellite data using optimal estimation, evaluation and application to ATSR, in preparation, 2010.

PVASR – ESA Cloud CCI Product Validation and Intercomparison Report: available at www.esa-cloud-cci.org, 2011.

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

Rossow, W., Mosher, F., Kinsella, E., Arking, A., Desbois, M., Harrison, E., Minnis, P., Ruprecht, E., Seze, G., Simmer, C., et al.: ISCCP cloud algorithm inter-comparison, Journal of Climate and Applied Meteorology, 24, 877–903, 1985.

Saunders, R., Francis, R., Francis, P., Crawford, J., Smith, A., Brown, I., Taylor, R., Forsythe, M., Doutriaux- Boucher, M., and Millington, S.: The exploitation of Meteosat Second Generation Data in the Met Office, Proceedings of the 2006 EUMETSAT Meteorological Satellite Conference, Helsinki, 2006.

Schmetz, J., Holmlund, K., Hoffman, J., Strauss, B., Mason, B., Gaertner, V., Koch, A., and Van De Berg, L.: Operational cloud-motion winds from Meteosat infrared images, Journal of Applied Meteorology, 32, 1206–1225, 1993.

Seemann, S.W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L.: Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance measurements, Journal of Applied Meteorology and Climatology, 47, 108–123, 2008.

Smith, N. Menzel, W. P., Weisz, E., Heidinger, A. K., and Baum, B. A.: A uniform space-time gridding algorithm for comparison of satellite data products: characterization and sensitivity study, Journal of Applied Meteorology and Climatology, 52, 255–268, 2013.

Stengel, M., et al., The Clouds Climate Change Initiative: Assessment of state-of-the-art cloud property retrieval schemes applied to AVHRR heritage measurements, Remote Sensing of Environment, http://dx.doi.org/10.1016/j.rse.2013.10.035, 2013.

Stubenrauch, C., Rossow, W., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B. C., Menzel, P. Minnis, P., Pearl, C. Platnick, S., Poulsen, C. Riedi, J.m Sun-Mack, S. Walther, A., Winkler, D., Zeng, S. and Zhao, G.: Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX Radiation Panel, Bulletin of the American Meteorological Society, 94, 1031–1049, doi:10.1175/BAMS–D–12–00 117.1, 2013.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.): IPCC, 2007: Climate Change 2007: The Physical Science Basis, Cambridge University Press, Cambridge, 996pp, 2007.

Walther, Andi, and Andrew K. Heidinger. "Implementation of the Daytime Cloud Optical and Microphysical Properties Algorithm (DCOMP) in PATMOS-x." Journal of Applied Meteorology & Climatology 51.7, 2012

Watts P.D., et al.: Study on cloud properties derived from Meteosat Second Generation Observations. Eumetsat Report, http://www.eumetsat.de/en/area2/publications/rep_cloud.pdf, 1998.

Watts, P., Bennartz, R., and Fell, F.: Retrieval of two-layer cloud properties from multispectral observations using optimal estimation, Journal of Geophysical Research: Atmospheres (1984–2012), 116, 2011.

Wind, G., Platnick, S., King, M. D., Hubanks, P. A., Pavolonis, M. J., Heidinger, A. K., Yang, P., and Baum, B. A.: Multilayer cloud detection with the MODIS near-infrared water vapor absorption band, Journal of Applied Meteorology and Climatology, 49, 2315–2333, 2010.

Annex 1: CREW project website

For further information please have a look on our project website

<p align="center">http://www.icare.univ-lille1.fr/crew</p>

describing the intention and goals of the CREW project, the datasets and scientific methods, and the participating institutes. It also gives a short summary of the CREW meetings, and provides reports, programs, and the participant lists of these meetings.

Annex 2: Datasets and Products

A.1 The CREW retrieval database

In the framework of the Cloud Retrieval Evaluation Workshop (CREW), a common cloud retrieval database was built to investigate strengths and weaknesses of currently available cloud property retrieval algorithms using passive imager observations. The cloud properties stored in the CREW database are listed in Table 1.

Table 5: List of cloud parameters in the CREW database

Acronym	Property	Units
CMB	Cloud Mask	[%]
CTH	Cloud Top Height	[m]
CTP	Cloud Top Pressure	[hPa]
CTT	Cloud Top Temperature	[K]
CPH	Cloud Phase	[water, ice, mixed]
COD	Cloud Optical Depth	[-]
REF	Particle Effective Radius	[μm]
LWP	Liquid Water Path	[gm^{-2}]
IWP	Ice Water Path	[gm^{-2}]
CTY	Cloud Type	[-]

In this paper the main focus is on the SEVIRI datasets, but also polar orbiting sensors such as the MODerate resolution Imaging Spectroradiometer (MODIS) onboard the EOS-Terra and Aqua satellites and the Advanced Very High Resolution Radiometer (AVHRR) as well as Multi-angle Imaging SpectroRadiometer (MISR), POLarization and Directionality of the Earth's Reflectances (POLDER) and Atmospheric InfraRed Sounder (AIRS) retrievals are included in the database. The database is complemented with cloud measurements that serve as a reference, including Advanced Microwave Scanning Radiometer for EOS (AMSR-E) observations and the active instruments CPR on Cloudsat and CALIOP on CALIPSO.

The CREW database contains five days of data, see Table 2. During these days the NOAA-18 satellite was aligned with A-train orbit for several core hours. In this paper we focus on 13 June 2008, as the dataset is most complete for this day.

Table 6: List of days of the retrieval inter-comparison.

Day	Month	Year	Hours with alignment between A-Train and NOAA-18
13	June	2008	12:00-15:30
17	June	2008	22:15-24:00
18	June	2008	00:00-01:45
22	June	2008	10:30-12:15
03	July	2008	10:00-12:00

In total, twelve institutions from Europe and USA participated in the CREW inter-comparison and validation of their SEVIRI datasets. This paper investigates the ten datasets providing cloud top height or cloud top pressure retrievals. The acronyms and contact persons of the participating institutions are listed in Table 3.

Table 7: List of Research groups that provided cloud parameter retrievals for the Common Database. For each groups the satellite instruments and cloud parameters for which data are provided are given. Where the codes from the cloud parameters are: cmk= cloud mask; cph= cloud phase;ctt= cloud top temperature; ctp= cloud top pressure; cth= cloud top height; cod = cloud optical depth; ref= effective radius; lwp= liquid water path; iwp= ice water path; cty= cloud type, and cs= convective signature. The coverage F means Full disc and S means swath data along the satellite path. COCS makes retrievals for ice clouds only (cloud phase = I) and the retrieved optical thickness is limited those smaller than 3 (cloud phase = F*). The EIM algorithm makes retrievals for water clouds only (cloud phase = W).

Acro.	Institute	Contact	Sensor	cmk	cph	ctt	ctp	cth	cod	ref	lwp	iwp
AWG	Wisconsin	Heidinger, Walther	SEVIRI	F	F	F	F	F	F	F	F	F
CMS	DWD	Stengel, Lockhoff, Kniffka	SEVIRI	F	F	F	F	F	F	F	F	F
COX	DLR	Kox	SEVIRI	F	I			F	F*			
DLR	DLR	Bugliaro	SEVIRI	F	F	F	F		F		F	F
EIM	Uni Marburg	Kühnlein	SEVIRI	F	W				F	F		
MPF	Eumetsat	Joro	SEVIRI	F	F	F	F					
EUM	Eumetsat	Lutz	SEVIRI	F	F		F					
FUB	FU Berlin	Preusker	SEVIRI	F				F				
GSF	Nasa Goddard	Platnick	SEVIRI	F	F	F	F		F	F	F	F
GSF	Nasa Goddard	Platnick	MODIS	S	S	S	S	S	S	S	S	S
KNM	KNMI	Roebeling, Meirink	SEVIRI	F		F			F	F	F	F
KNM	KNMI	Roebeling, Meirink	MODIS	S		S			S	S	S	S
KNM	KNMI	Roebeling, Meirink	AVHRR	S		S			S	S	S	S
LAR	NASA Langley	Minnis	SEVIRI	F	F	F	F	F	F	F	F	F
LAR	NASA Langley	Minnis	MODIS	S	S	S	S	S	S	S	S	S
MFR	Meteo France	Le Gleau, Derrien	SEVIRI	F		F	F	F				
OCA	Eumetsat	Watts	SEVIRI	F	F		F		F	F		
RMB	KMI	Ipe	SEVIRI	F	F				F			
SMH	SMHI	Dybbroe	MODIS	S		S	S	S				
SMH	SMHI	Dybbroe	AVHRR	S		S	S	S				
UKM	UK Met Office	Francis	SEVIRI	F	F	F	F	F		F		
ULI	Univ. Lille	Riedi	POLDER	S					S			

Annex 3: Digital dataset of analysis results

The results summarized in these reports are supported by a large set of JPEG images. The full datasets of images and tables that have been generated for this study can be found at the FTP site of the Space Science and Engineering Center (SSEC) (ftp://ftpush.icare.univ-lille1.fr/crew/plots/). There are several levels of specifications for the images that are mirrored in the name convention and directory structure.

The first level in the directory is the product **(results/<product>/)**. For the MSG comparison the possible first levels are:

- ***cmb***   Cloud mask binary
- ***cod***   Cloud optical depth
- ***cph***   Cloud phase
- ***cth***   Cloud top height
- ***ctp***   Cloud top pressure
- ***ctt***   Cloud top temperature
- ***lwp***   Liquid water path
- ***ref***   Effective radus

Additionally, RGB composites and more detailed overview of the special regions can be found in the subdirectories rgb and ovw. All other subdirectories will be explained in relevant chapters.

The next directory level is the image type **(results/<product>/<type>/).** Results images are grouped as follows. There are in general three different types of images:

1. Mapped visualization images of the parameters. (Directory name ***img***).
2. Distribution of frequency as one-dimensional histograms (***his***). Not existing for cloud mask and cloud phase.
3. Combined 2D frequency distribution histogram matrixes for all two algorithms pairs. (Subdirectory name is ***cor***). These images show individual scatter plot-like sub-images at the lower left part and the statistical parameters for one-by-one comparison bias, correlation and root mean square error at the upper right part. Statistics for each data set are written in the main diagonal of the matrix.
4. Tables summarize the findings for each cut and product as JPEG files and as CSV tables importable in excel format. (Directory ***tables***).

Under this subdirectory level overview images are stored, which combine the results of all available algorithms. Those images show comparisons between two specific algorithms, and are stored in a further directory level ***single***

The file name convention includes further level of specifications. Seperator between the levels are a '-'. The files are organised using the following convention

(C<cut number>-<product>-<type>-<group>-<mask>-<yyyymmdd>-<hhmm>.jpg), where:

1. First point is the region of analysis. All analysis were done for the entire globe with a five times lower resolution (742 x 742 pixels instead of 3712 x 3712 pixels) and for special regions, called CUTs, with the full resolution. The region CUTs are described later in this

document. Global disk are named as *C00* (cut zero). That means that all files with *C00* include results of the entire globe.

2. Second part of the name specifies again the product.
3. Third part of the name specifies again the image type.
4. Forth part can be *ALL* for overview images or the acronyms of two groups (such as *AWG* and *FUB* for Algorithm working group of NOAA/NESDIS and *Freie Universität* algorithm).
5. Next part of the name indicates if a common cloud mask (*COM*) or an individual cloud mask (*IND*) was used.
6. Finally, the day and the time are specified at the end of the file name.

In this way the general file structure is:

results/<*product*>**/**<*type*>**/C**<*cut number*>**-**<*product*>**-**<*type*>**-**<*group*>**-**<*mask*>**-**<*yyyymmdd*>**-**<*hhmm*>**.jpg**

To give an example, the file "*results/cth/cor/C03-cth-cor-ALL-IND-20080613_1345.jpg*" refers to the 2D histogram matrix of cloud height for the region Cut 3, individual cloud mask for the 13:45 UTC scene of 13th June 2008.

We have added the name as well as the file subdirectory at the lower left edge of all images in this report.

The software was designed to automatically generate the images. So, that we could not go through all images to check all files for graphical issues, such as character size, or the proper location of the color bar, etc. In case you find any errors please contact Ulrich Hamann (ulrich.hamann@gmx.net) or Andi Walther (andi.walther@ssec.wisc.edu).

Annex 4: Acronyms

Through this document and within the result image package a system of acronyms and file name conventions is used that is important to know.

AIRS	Atmospheric InfraRed Sounder
AMSR-E	Special Sensor Microwave/Imager
AVHRR	Advanced Very High Resolution Radiometer
CA	Cloud Amount
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CLOUDSAT	Cloud satellite mission operated by NASA
LWP	cloud Liquid Water Path
COT	Cloud Optical Thickness
COD	Cloud Optical Depth
CPH	Cloud thermodynamic PHase
ECMWF	European Centre for Medium-Range Weather Forecasts
ESA	European Space Agency
EUMETSAT	Europe's Meteorological Satellite Organisation
GOES	Geostationary Operational Environment Satellite
IASI	Infrared Atmospheric Sounding Interferometer
IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
ISCCP	International Satellite Cloud Climatology Project
LUT	Look Up Table
MODIS	Moderate Resolution Imaging Spectroradiometer (NASA/Terra, Aqua)
METEOSAT	Meteorological satellite
MSG	Meteosat Second Generation
MWR	MicroWave Radiometer
NASA	National Aeronautics and Space Administration
NIR	Near-infrared
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction model
RTM	Radiative Transfer Model
SEVIRI	Spinning Enhanced Visible and Infrared Imager
TOA	Top of Atmosphere
VIS	Visible
WV	Water-Vapor