EUMETSAT Satellite Application Facility on Climate Monitoring

Algorithm Theoretical Basis Document for the Cloud Probability of the NWC/PPS

NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudProbability, Issue 1, Rev. 0

Applicable to SAFNWC/PPS version 2018

Applicable to the following PGEs:

Acronym	Product ID	Product name	Version number
CMa-prob	NWC-152	Cloud Probability	1.0

Prepared by Swedish Meteorological and Hydrological Institute (SMHI)

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	SMHI		13 December 2018
Reviewed by	SAFNWC Project Team EUMETSAT		 15 November 2018
Authorised by	Anke Thoss, SMHI *SAFNWC PPS Manager*		13 December2018

Document Signature Table

	Name	Function	Signature	Date
Authors	Karl-Göran Karlsson Joseph Sedlar	CM SAF scientists		20/12/2016
Editor	Rainer Hollmann	Science Coordinator		
Approval	Rainer Hollmann	Science Coordinator		
Release	Martin Werscheck	Project Manager		

Distribution List

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

External Distribution

Company	Name	No. Copies
PUBLIC		1

Document Change Record

Issue/ Revision	Date	DCN No.	Changed Pages/Paragraphs
1.0	24/04/2015	SAF/CM/SMHI/ATBD/GAC/CMA-prob	Initial version, submitted for PCR 2.2
1.1	27/05/2016	SAF/CM/SMHI/ATBD/GAC/PBCM	Version submitted for DRR 2.2
1.2	02/08/2016	SAF/CM/SMHI/ATBD/GAC/PBCM	Updated version for DRR 2.2 Close-Out
1.3	20/12/2016	SAF/CM/SMHI/ATBD/GAC/PBCM	Version submitted for NWC SAF PPS 2018 PCR
1.4	17/10/2018	SAF/CM/SMHI/ATBD/GAC/PBCM	Version prepared for NWC SAF PPS 2018 ORR (equivalent to document NWC/CDOP3/PPS/SMHI/SCI/ATBD/1c)
1.0d	17/10/2018	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudProbability	Same as previous issue but transferred from CM SAF versions to NWC SAF
1.0	13/12/2018	NWC/CDOP3/PPS/SMHI/SCI/ATBD/CloudProbability	Updates after NWCSAF/PPS v2018 DRR. RID-001, 002, 003, 004, 005, 007, 008, 010 Other changes: Added a comment on no-data in twilight.

Table of Contents

List of Figures

List of Tables

1 The EUMETSAT SAF on Climate Monitoring

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

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

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

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

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

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

- Taking a major role in data set assessments performed by research organisations such as WCRP. This role provides the CM SAF with deep contacts to research organizations that form a substantial user group for the CM SAF CDRs,
- Maintaining and providing an operational and sustained infrastructure that can serve the community within the transition of mature CDR products from the research community into operational environments.

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

2 Introduction

This CM SAF Algorithm Theoretical Basis Document (ATBD) describes a new probabilistic cloud masking product - denoted CMA-prob – which has been developed by CM SAF during the CDOP-2 and CDOP-3 phases. It is based on Bayesian theory and it is complementary to the SAFNWC PPS cloud mask which was used when defining the CMSAF CLARA-A1 and CLARA-A2 datasets (i.e., the Fractional Cloud Cover product CM-11011). The idea is that on a longer term (CLARA-A3, and beyond) this new probabilistic cloud mask will replace the current one in order to improve the error characterisation of cloud masking and its influence on downstream cloud, surface radiation and surface albedo products. For CLARA-A2 only a demonstration product was provided for users to become acquainted with the product capability and for preliminary evaluation.

The ATBD generally follows the description of the method in Karlsson et al. (2015) but describes also several extensions which were introduced to improve the quantitative usefulness of the CMA-prob product. One of the most central new features is that the CMA-prob value of 50 % has been tuned to give optimal cloud detection everywhere regardless of underlying surfaces and observation conditions.

2.1 Applicable documents

Reference	Title	Code	Version	Date
AD 1	NWC SAF Product Requirements Document	NWC/CDOP3/SAF/AEMET/MGT/PRD	1.1	17/12/2018

2.2 Reference documents

Reference	Title	Code	Version	Date
RD 1	Algorithm Theoretical Basis Document for the Cloud Mask of the NWC/PPS	NWC/CDOP3/PPS/SCI/ATBD/CloudMask	2.1	13/12/2018
RD 2	Scientific and Validation Report for the Cloud Product Processors of the NWC/PPS	NWC/CDOP3/PPS/SMHI/SCI/VR/Cloud	2.0	13/12/2018

3 Theoretical description of the CMA-prob method

3.1 Background – problems with traditional cloud masking and suggested new approaches

For many years, the definition of fixed cloud masks or cloud masks with a small set of quality flags has been the most common way of solving the cloud screening problem in applications based on passive multispectral satellite imagery. Many examples of this exist in the literature, e.g., Dybbroe et al., (2005a, 2005b), Kriebel et al. (2003), Derrien and LeGleau (2005), Frey et al. (2008) and Pavolonis et al. (2005). The use of a fixed cloud mask is straightforward for downstream applications (e.g. for Sea Surface Temperature (SST), surface albedo, clear sky radiance and NDVI vegetation index retrievals) meaning that all cloudy pixels should be discarded in the retrieval of the actual parameter. However, the drawback is that no, or very limited, information about the uncertainty in the cloud screening is generally available with these methods. Consequently, the error characteristics are generally unknown even if internal parameter-specific algorithm uncertainties may be known. Furthermore, various cloud masks have generally been defined aiming for different purposes and applications. Consequently, the performance may vary considerably from method to method regarding whether the cloud screening is executed in a clear conservative way (i.e., defining clear pixels with high confidence) or in a cloud conservative way (i.e., defining cloudy pixels with high confidence). Instead, the desire to define a more flexible cloud mask, suitable for any (or at least most) downstream applications, has become increasingly important. Such a cloud mask can either be expressed as a cloud index (as suggested by Khlopenkov and Thrishchenko, 2007) or a cloud probability (Merchant et al., 2005) meaning that any user should be able to define the most suitable mode of operation. In other words, it could be used anywhere in the range from the clear conservative mode to the cloud conservative mode by just changing the tolerance level of the required cloud probabilities.

Although statistical and probabilistic (Bayesian) theory has been well established for decades (or even centuries), a problem has been to find appropriate observational references to represent the true global cloud occurrence from which a firm statistical cloud distribution database can be built. However, with the 2006 launch of the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) onboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite, the situation has improved considerably. CALIOP offers global cloud observations with higher detection sensitivity than any other passive instrument (Winker et al., 2009). Furthermore, observations can be matched simultaneously in time (however, restricted to certain conditions) to observations by current operational AVHRR sensors. This has triggered numerous studies examining AVHRR-based cloud detection methods in detail (e.g., Karlsson and Dybbroe, 2010, Karlsson and Johansson, 2013 and Stengel et al., 2014). It has also paved the way for more systematic attempts to provide cloud probabilities rather than fixed cloud masks (Heidinger et al., 2012, and Musial et al., 2014), and the CMA-prob development described here is another example of this. Recently also improved versions of the CALIOP cloud datasets have been utilised for in-depth studies of the cloud detection limits for methods based on AVHRR data (Karlsson et al., 2018). Their findings and their tools have been important for defining the final concept of the CMA-prob methodology (further outlined in Section 3.6 and 3.7).

3.2 Bayesian theory

Let us first recapitulate some fundamentals of the probabilistic statistical theory. The theory is based on the pioneering work by Thomas Bayes who, already in 1763, formulated his famous

theorem (nowadays referred to as Bayes' Theorem) for estimation of the posteriori probability of an event as a function of likelihoods (conditional probabilities) and a priori probabilities of other events. In the context of analysis of radiance feature vectors measured by satellite sensors, we may express Bayes' Theorem as follows: If $\boldsymbol{F}$ is a vector of satellite radiances or image features (e.g., brightness temperature differences or reflectances), we may denote the posteriori conditional probability that it is cloudy when $\boldsymbol{F}$ is given as $P(cloudy|\boldsymbol{F})$. In the same sense, we may denote the conditional probability that vector $\boldsymbol{F}$ occurs given it is cloudy as $P(\boldsymbol{F}|cloudy)$. If also introducing the overall probability (climatological mean) that it is cloudy as $\overline{P(cloudy)}$ and the overall probability that any given value of $\boldsymbol{F}$ occurs as $P(\boldsymbol{F})$, we may write Bayes' Theorem as follows:

$$P(cloudy|\boldsymbol{F}) = \frac{\overline{P(cloudy)}P(\boldsymbol{F}|cloudy)}{P(\boldsymbol{F})} \tag{1}$$

Despite its simple form, the solution to Eq. 1 is not easy to find in a situation with multispectral measurements (i.e., when the dimension of $\boldsymbol{F}$ is large). The estimation of parameters on the right hand side of Eq. 1 (especially $P(\boldsymbol{F}|cloudy)$) becomes increasingly difficult the more image features that are chosen. It then requires extraction from very large statistical training datasets to fully describe the dependence on individual image features and, in addition, also the effect of their mutual correlation. What complicates things even further is that even with one specific realisation of feature vector $\boldsymbol{F}$, probabilities may differ depending on different environmental situations (e.g. if the pixel measurement is made in winter or in summer, over land or over ocean, in mountainous terrain or over desert, etc.). Thus, the training process needs to take into account additional ancillary information for a correct description of the environmental conditions.

To reduce complexity of the problem some approximations may be utilised. One approach could be the entirely empirical approach of estimating $P(cloudy|\boldsymbol{F})$ directly from predefined Lookup Tables composed during training with some stratification based on ancillary data. Such a method has been demonstrated by Musial et al. (2014). Alternatively, some simplifications and approximations could be made to Eq. 1. One such simplification is denoted *The Naïve Bayesian approach*, and this is used for the CMA-prob method.

3.3 The CMA-prob Naïve Bayesian approach

If assuming that individual image feature components f_i in $\boldsymbol{F}$ are all independent (i.e., image features are uncorrelated), a simplification is possible so that individual probabilities may now be multiplied to get the total probability, following the fundamental statistical rule for "Compound Probability of Independent Events". Thus, Eq. 1 reduces to

$$P(cloudy|\boldsymbol{F}) = \frac{\overline{P(cloudy)}\prod_i P(f_i|cloudy)}{P(\boldsymbol{F})} \tag{2}$$

This approximation of Bayes' Theorem is denoted **the Naïve Bayesian approximation**.

The problem has now been reduced to estimating individual probabilities $P(f_i|cloudy)$ and then simply multiplying them. However, it must be emphasized that even if we have achieved

a very simple equation for calculation of the probabilities, the big scientific challenge lies in defining and estimating the conditional probabilities on the right-hand side of the equation. This includes the very fundamental choice of appropriate image features f_i. These must be chosen in an optimal way, with each individual feature having documented capability to provide essential information about cloud occurrence. The following sub-sections will describe the methodology used and the choice of optimal image feature components f_i .

We notice also that there must be a mutual inter-dependence between $P(f_i|cloudy)$ and $P(cloudy|f_i)$. More clearly, if knowing the conditional probability that it is cloudy given a certain image feature value, we can also calculate it the other way around from the same statistical training dataset (provided that both absolute and relative frequencies of cloud occurrences are stored). This fact will be utilised here when defining the method. It is often more practical and easier to understand results for individual image features if compiling statistical distributions of $P(cloudy|f_i)$. It also provides a natural link to development efforts trying to find optimal thresholds in multispectral thresholding schemes like the standard CMA method in the PPS cloud processing package. Remaining factors on the right hand side of Eq. 2 may also be calculated from training data. An estimation of the mean cloud occurrence $P(\overline{cloudy})$ is possible and the factor $P(\boldsymbol{F})$ may be estimated by summing contributions from both cloudy and clear cases and then compute the overall frequency for which any particular realisation of vector $\boldsymbol{F}$ occurs.

The Naïve Bayesian approximation has been successfully applied to many scientific applications (e.g., Kossin and Sitkowski, 2009) and it has also recently been applied to the AVHRR cloud screening problem (Heidinger et al., 2012). The main difference between CMA-prob and the latter method lies in the choice of image features and the used ancillary information.

3.4 Estimating conditional cloud probabilities from CALIPSO measurements

Having access to a system that can match and co-locate CALIPSO/CALIOP and NOAA/METOP AVHRR measurements makes it relatively straight-forward and simple to estimate conditional cloud probabilities, i.e. the frequency that it is cloudy at a certain image feature value $P(cloudy|f_i)$. However, as previously mentioned, this should only be done by using some restrictions on e.g. illumination conditions and the geographical coverage in order to avoid too broad distributions and distributions with a limited dynamical range of probabilities. If these restrictions on not considered, the final ability to separate cloudy from cloud-free radiances would be reduced (i.e., too often give cloud probabilities close to 50 %).

Figure 3-1 shows the estimated cloud probabilities as a function of the AVHRR visible reflectances from the 0.6 μm channel over tropical ocean surfaces (left panel) and over high-latitude homogeneous surfaces with permanent snow-cover (right panel). The distinction between low- and high latitudes is made at +/- 23.5º latitude and homogeneous surfaces are determined by applying a threshold of the local (11x11 pixels) variance in the digital elevation map (see RD 1). Information on snow cover is taken from NWP or Reanalysis datasets depending on whether results are derived for Nowcasting or for Climate applications. In Figure 3-1 snow information is taken from ERA-Interim data. The reference CALIPSO results are original CALIOP cloud masks (CLAY product version 4.1). It should also be mentioned that daytime conditions in Figure 3-1 are valid for solar zenith angles below 80 degrees.

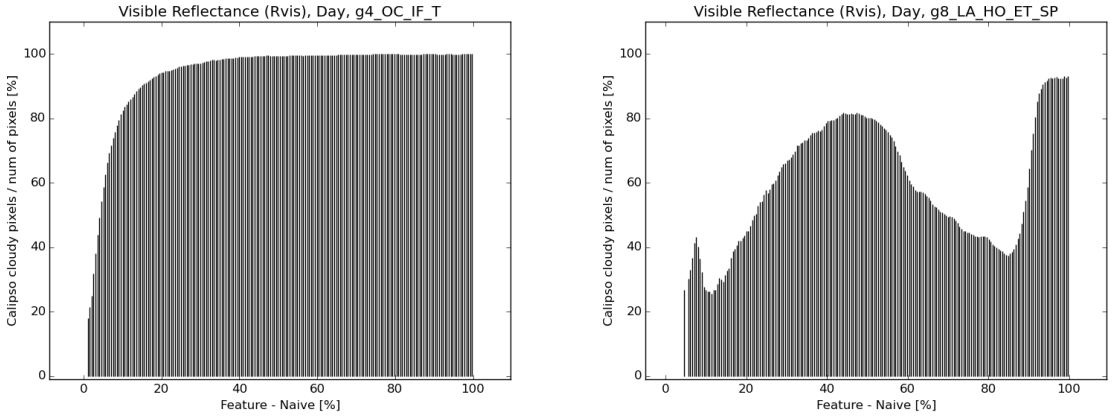

Figure 3-1 *Cloud probabilities estimated from global CALIPSO-CALIOP cloud data in the period 2006-2015 as a function of AVHRR 0.6 µm visible reflectances (Feature-Naïve in %, denoted Rvis).* ***Left:*** *Results over* ***Tropical ocean*** *surfaces.* ***Right:*** *Results over* ***High Latitude surfaces with permanent snow-cover****.*

From Figure 3-1 we conclude that cloud probabilities increase rapidly with reflectance over a very dark surface such as the ice-free ocean. Probabilities exceed 50 % already at a very low reflectance value (at approximately 5 % reflectance) and reach above the 80 % level at approximately 12 % reflectance. Thus, conditions for cloud-screening appear almost ideal. This is not the case for the snow-covered ground pixels in Figure 3-1 (right). Here, it is difficult to arrive at a robust, unique reflectance value where cloud probability exceeds 50 % (which would be needed for this image feature to be useful for cloud screening purposes). This occurs only for moderate to large reflectance values (30-70 % and above 90 %. For the inter-mediate region of high reflectances probabilities are actually well below 50 % and even down to 30 % for reflectivities between 85 % and 90 %. The reduced probability, especially in the interval 70-90 %, means that cloud-free snow surfaces have dominantly high reflectances in this interval. Consequently, cloud-free conditions are more likely to occur than cloudy conditions in this reflectance interval. We also notice that zero reflectances are never reached over these bright surfaces. Also, the small increase in cloud probabilities at low reflectances are likely to be caused by shadowing effects (i.e., cloud shadows on lower altitude clouds) or by very thin clouds observed at high solar zenith angles.

A similar situation is seen over the same surfaces for the infrared brightness temperature difference of the 11 µm channel with regard to the surface skin temperature (Figure 3-2, difference denoted Tirdiff). In this figure, positive values mean that measured brightness temperatures are ***colder*** than the surface temperature. Very good separability conditions are seen over tropical ocean surfaces, while they are problematic over snow-covered surfaces. Notice in particular the effect of near-surface temperature inversions (e.g., for negative temperature differences in Fig. 3-2 meaning that measurements are ***warmer*** than the surface temperature). We notice that clouds which are warmer than the surface may occur over tropical ocean (e.g., a small fraction of marine stratocumulus) as well as over snow-covered surfaces. However, such clouds occur much more frequently over snow-covered surfaces and while the temperature difference is generally small over ocean it is generally much larger over

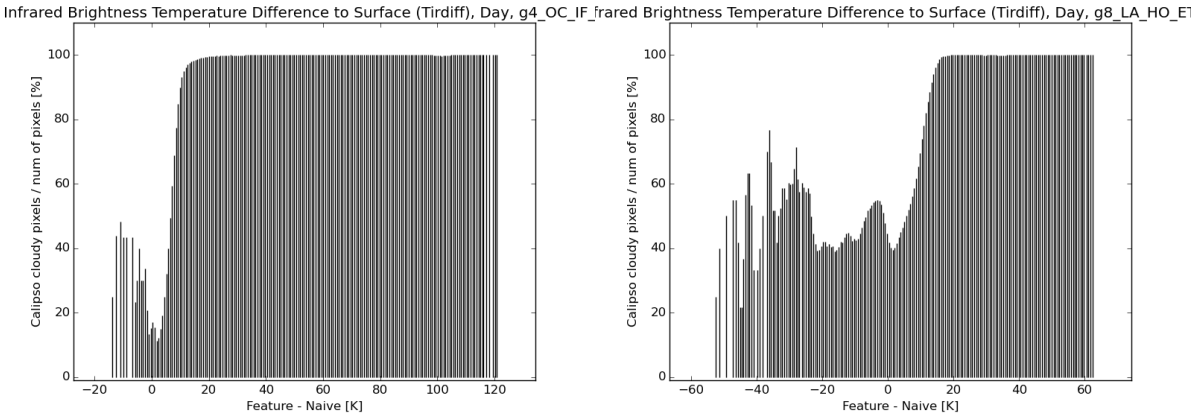

Figure 3-2 *Cloud probabilities estimated from CALIPSO-CALIOP cloud data in the period 2006-2015 as a function of the temperature difference between the ERA-Interim (Dee et al., 2011) surface skin reference temperature and the AVHRR 11 μm brightness temperature (Feature_Naive in K, denoted Tirdiff).* ***Left:*** *Results over* ***tropical ocean*** *surfaces during day.* ***Right:*** *Results over* ***High Latitude snow-covered*** *surfaces during day.*

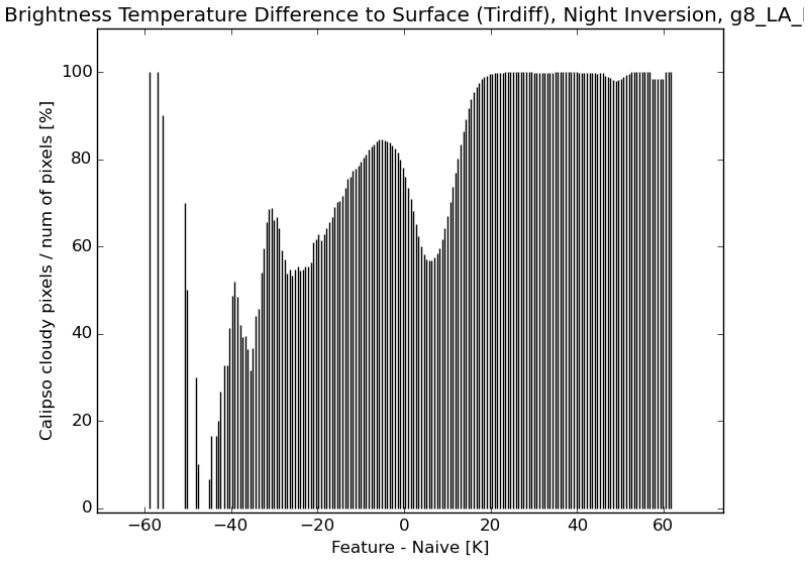

Figure 3-3 *Cloud probabilities estimated from CALIPSO-CALIOP cloud data in the period 2006-2015 as a function temperature difference between the ERA-Interim (Dee et al., 2011) surface skin reference temperature and the AVHRR 11 μm brightness temperature (Feature_Naive in K, denoted Tirdiff). Results are valid over* ***High Latitude snow-covered*** *surfaces during night and with strong near-surface inversions.*

snow-covered surfaces. The latter surfaces are dominated by cold and highly elevated surfaces in Antarctica and over Greenland. It is clear that the use of a single threshold at e.g. 5 K is capable of detecting the vast majority of clouds over tropical ocean while the same threshold applied over snow-covered surfaces will miss a large fraction of all existing clouds.

Statistics from situations with very strong near-surface inversions (i.e., when surface temperatures are colder than temperatures in the 950 hPa level) over snow-covered surfaces in

Fig. 3.3 are also interesting. In such situations, using this temperature difference feature for cloud screening is even worse. Clouds may here occur for any measured temperature difference but with probabilities which are not significantly different from 50 %. Only for clouds which are 20 K colder than the surface (generally medium- and high-level clouds) or possibly also clouds which are 5-15 K warmer than the surface (low-level inversion clouds) we get high enough cloud probabilities to detect clouds with certainty. We also notice a minimum in cloud probabilities at about 10 K instead of near 0 K which would be the anticipated value, i.e. a pixel that is neither warmer nor colder than the surface should be generally cloud-free. This indicates that the ERA-Interim surface temperature reference probably has a warm bias, i.e., that surface temperature inversions are probably stronger than ERA-Interim is capable of showing. This is supported by, e.g., long-time experiences of a warm bias for ECMWF-forecasted minimum 2-meter temperatures in boreal forest regions at high latitudes during winter (Hogan et al., 2017). Also the few cases of extremely large positive differences in Fig. 3.3 (up to 60 K) may be linked to errors in the surface temperature analysis but it can also come from navigation errors in coastal areas with steep orography.

We conclude from Figs. 3-1 to 3-3 that conditions for efficient cloud screening may be drastically different depending on the geographic location and the prevailing illumination conditions (i.e., day, night or twilight). This is one of the explanations for the very successful performance of simple bi-spectral VIS-IR cloud screening methods at low- to moderate latitudes (best exemplified by the results derived mainly from geostationary satellite data of the International Satellite Cloud Climatological Project – ISCCP – see Rossow et al., 1999 and Young et al., 2018). On the other hand, it also clearly illustrates potential serious limitations for the same methods over high latitudes and over the polar regions (as highlighted by Karlsson and Devasthale, 2018).

3.5 Definition of a basic sub-set of constrained AVHRR image features

The Naïve Bayesian CMA-prob method utilises estimated conditional cloud probabilities (introduced in the previous section) for a sub-set of image features. However, rather than to define them in their purest form (as illustrated in Figs 3-1 – 3-3) we have chosen to define them linked to pre-calculated dynamic image feature thresholds used by the Polar Platform System cloud software package (PPS, see Dybbroe et al, 2005a, 2005b) and in this particular case version PPS version 2018 [RD 1]. The reason for linking image features to pre-calculated thresholds is that the latter are defined across a wide range of environmental conditions (see Dybbroe et al., 2005a for more details). This concerns image feature variability due to the following factors: Solar and satellite geometry (direct angular dependence and dependence on scattering angles), prevailing atmospheric profiles of temperature and humidity, climatological ozone and aerosol amounts, topography and land cover, and spectral surface emissivities. Without considerating these factors when training the probabilistic classifier, the results risk being imprecise and most likely misleading. We claim that it is better to piggy-back on existing prepared threshold information, based on knowledge built on many years of cloud thresholding experience, than to try to train a classifier from scratch as a function of all the mentioned factors. The latter would require the creation of very large dimension Look-up Tables of statistical relations of cloudiness and image features and their respective dependencies on a wide range of environmental factors.

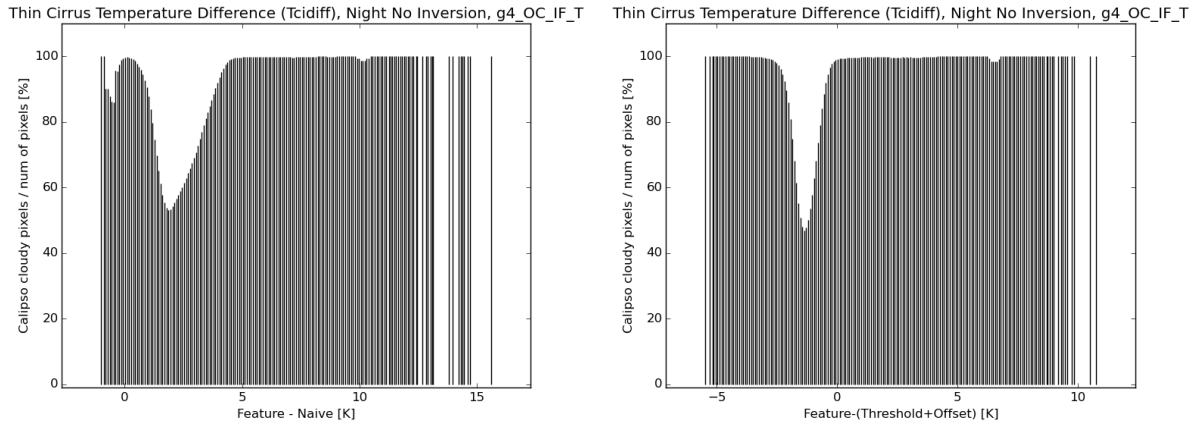

Figure 3-4 *Cloud probabilities estimated from CALIPSO-CALIOP cloud data in the period 2006-2015 as a function of AVHRR temperature differences between AVHRR channel 4 and 5 (denoted Feature in the plots) over Tropical ocean surfaces during night. Left panel shows results in original form and right panel if plotting results as a function of temperature differences related to PPS thresholds (consisting of dynamic threshold plus a tuning offset value).*

To illustrate the usefulness of this concept we consider one of the most commonly used AVHRR image features for detecting thin cirrus clouds (originally suggested by Inoue, 1987): the brightness temperature difference between AVHRR channels 4 and 5 at 11 μm and 12 μm, respectively. The main principle used for Cirrus detection is normally that the cloud transmissivity for thin ice clouds is higher in AVHRR channel 4 than in AVHRR channel 5, thus creating a positive brightness temperature difference between AVHRR channels 4 and 5. Figure 3-4 shows the cloud probabilities as a function of this temperature difference (Fig. 3-4, left) but also as a function of the temperature difference relative to the corresponding PPS threshold (Fig. 3-4, right).

In its original form (Fig. 3-4, left), we have two peaks in cloud occurrence where one is for differences close to zero, and the other for values exceeding approximately 4 K. Thus, the area with lower cloud frequencies between the peaks spans an interval of almost 4 K. In the alternative formulation (Fig. 3-4, right) results are much more distinctly organised, the interval with lower frequencies is reduced to only 2 K and the range of probability values have been enlarged (which is favourable for the probabilistic classification process). The latter circumstance is especially true for the leftmost part of the distribution. We interpret this as primarily an effect of being able to take into account the natural cloud-free contribution from atmospheric water vapour emission in the split-window channels. This emission is also able to create a discernible temperature difference in the absence of cirrus clouds explaining the broader and less decisive probability distribution in its original form for temperature differences below approximately 4 K. Resulting distributions after the coordinate change now clearly separates thin cirrus clouds to the right in the plot from the opaque clouds in the left part of the plot with cloud-free cases now concentrated around the changed x-coordinate value of around –1 K.

Table 3-1 *Spectral channels of the Advanced Very High Resolution Radiometer (AVHRR). Three different versions of the instrument are described as well as corresponding satellites.*

Channel Number	*Wavelength (µm) AVHRR/1 Tiros-N, NOAA-6,8,10*	*Wavelength (µm) AVHRR/2 NOAA-7,9,11,12,14*	*Wavelength (µm) AVHRR/3 NOAA-15,16,17,18 NOAA-19, Metop-A Metop-B*
1	0.58-0.68	0.58-0.68	0.58-0.68
2	0.725-1.10	0.725-1.10	0.725-1.10
3A	-	-	1.58-1.64
3B	3.55-3.93	3.55-3.93	3.55-3.93
4	10.50-11.50	10.50-11.50	10.50-11.50
5	Channel 4 repeated	11.5-12.5	11.5-12.5

Table 3-2 *Used AVHRR image features for day illumination probabilistic cloud masking.*

Feature name	Definition	Main cloud detection ability
Rvis	Over land: AVHRR channel 1 TOA reflectances minus PPS thresholds Over ocean: AVHRR channel 2 TOA reflectances minus PPS thresholds	Identification of bright clouds over dark Earth surfaces
Tirdiff	AVHRR channel 4 brightness temperatures minus ERA-Interim surface skin temperatures minus PPS thresholds	Identification of clouds which are significantly colder than the Earth surface
Rswir_3a *(morning orbit AVHRR/3)*	AVHRR channel 3A reflectances divided by AVHRR channel 1 reflectances	Identification of clouds with significant reflection in the visible near-infrared infrared region (in particular water clouds and thick multi-layered ice clouds over snow-covered surfaces)
Rvis37 *(afternoon orbit all AVHRRs and morning orbit AVHRR/2)*	AVHRR channel 3B reflectances minus PPS thresholds	Identification of clouds with significant reflection in the short-wave infrared region (water clouds and thick multi-layered ice clouds

With this background we now list in Tab. 3-2 and Tab. 3-3 a set of 8 constrained image features (related to original AVHRR channels described in Tab. 3-1) that will be used for the definition of the CMA-prob probabilistic cloud mask estimates. Four image features are selected for two scene categories: day (solar zenith angle [SZA] below 80º); Tab. 3-2 and night (SZA ≥ 89º; Tab. 3-3). For the twilight case (80º ≤ SZA < 89º) either the day or night approach is used depending on if non-zero reflectances are detected in AVHRR channel 1. Only one common feature (Tirdiff) is used both day and night but the underlying statistics achieved during the training of the method is separated into day and night categories. In order to account for geographical and topographical differences, we defined 14 geographical regions over which we trained the probabilistic classifiers. These regions are listed in Tab. 3-4. Ice cover information is taken from OSI SAF ice concentration data and snow information (snow depth) is taken from ERA-Interim re-analyses. The land-sea mask is taken from USGS landuse dataset and the labelling of whether or not the surface is dry is based on land emissivity climatologies (see RD 1 for more details on the use of ancillary datasets). The classifier was trained using the CALIPSO-CALIOP cloud product, denoted Cloud and Aerosol Layer Information product version 4.01. In summary, we trained the probabilistic classifier for 14 different surface regions, 2 illumination conditions and 4 AVHRR feature tests for each illumination class yielding a total of 112 unique, individual probabilistic estimates (probability distribution functions, PDF).

Table 3-3 *Used transformed AVHRR image features for night illumination probabilistic cloud masking.*

Feature name	Definition	Main cloud detection ability
Tirdiff	AVHRR channel 4 brightness temperatures minus ERA-Interim surface skin temperatures minus PPS thresholds	Identification of clouds which are significantly colder than the Earth surface
Tcidiff	AVHRR channel 4 brightness temperatures minus AVHRR channel 5 brightness temperatures minus PPS thresholds	Identification of thin cirrus clouds
Twdiff	(AVHRR channel 3b brightness temperatures minus AVHRR channel 4 brightness temperatures) minus PPS thresholds	Identification of water clouds
Texture_night	Over land: Not used (surface variability generally too large)! Over ocean: (Sum of local 3x3 pixel variances for AVHRR channel 4 brightness temperatures and AVHRR channel 3b and 5 brightness temperature differences) minus PPS thresholds	Identification of fractional or broken clouds over ocean

Table 3-4 *Geographical regions used when training the probabilistic classifier. Notice that areas with ice-cover are further sub-divided into areas with broken ice (marginal ice zone) and areas with 100 % ice cover. Similarly, snow covered surfaces are further subdivided into areas with seasonal and permanent snow cover. Consequently, the total number of defined regions or surfaces is 14.*

Geographical region	Definition
Polar ice-covered ocean	Ice-covered ocean based on OSI SAF ice concentration data
High-latitude ocean	Ice-free ocean at latitudes higher than 23.5°
Low-latitude ocean	Ocean at latitudes lower than 23.5°
High-latitude snow –covered mountains	Regions with high topography variability and with snow-cover (from ERA-Interim) at latitudes higher than 23.5°
High-latitude snow-free mountains	Regions with high topography variability without snow-cover at latitudes higher than 23.5°
High-latitude snow-covered land	Snow-covered land with low topography variability at latitudes higher than 23.5°
High-latitude snow-free land	Snow-free land with low topography variability and with vegetation cover (none-dry) at latitudes higher than 23.5°
Dry homogeneous land regions	Land areas with low topography variability and without homogeneous vegetation at latitudes lower than 23.5°
Dry mountainous regions	Land areas with high topography variability and without homogeneous vegetation at latitudes lower than 23.5
Low-latitude homogeneous vegetated regions	Vegetated land areas with low topography variability and without homogeneous vegetation at latitudes lower than 23.5at latitudes lower than 23.5°
Low-latitude mountain regions with vegetation	Vegetated mountain regions with high topography variability and with vegetation at latitudes lower than 23.5°

3.6 Training the classifier using CALIPSO-CALIOP cloud data with dependencies on CALIOP-estimated cloud optical thicknesses

To develop the CMA-prob classifier, we have taken advantage of the previously reported studies based on collocated NOAA/METOP AVHRR and CALIPSO orbits as described by Karlsson and Johansson (2013) and Karlsson and Håkansson (2018). In particular, the extension of the matchup database resulting from the study by Karlsson and Håkansson (2018) has been very important here. Some example results from this extended collocated dataset over 10 years (covering the period 2006-2015) have already been shown in the previous section. However, some important and necessary restrictions to the utilised information have been applied during the training process for being able to construct a probabilistic classifier capable of providing reasonable results.

A great advantage of the CALIPSO-CALIOP cloud products is their superior sensitivity for cloud detection compared to corresponding conditions for passive data, such as from the AVHRR sensor. However, this is also a problem when using this information as the basis for a statistical training of a probabilistic cloud masking method. More clearly, there is a risk for "over-training", i.e., that we force the method to try to detect clouds that are theoretically impossible to detect from AVHRR sensor data. As a result, the probabilistic cloud-screening method would then risk systematically creating artificial clouds in truly cloud-free areas since the cloud-free signal cannot be confidently separated from the cloudy signal for e.g., sub-visible cirrus clouds. Consequently, we need to find a way to restrict the used CALIOP-based cloud mask in the training process to include only those clouds we believe are potentially discernible in AVHRR images. In other words, we need to define the AVHRR cloud detection limit as accurately as possible.

This task can also be formulated as that we need to find the proper restricted CALIPSO cloud mask that is most accurately reproduced by the AVHRR-based cloud masking method. This means that we have to filter out the thinnest CALIOP-detected clouds from the CALIOP cloud mask up to a certain limit or threshold in cloud optical thickness where we then can find the best resemblance with the AVHRR-derived cloud mask. This limit in optical cloud thickness can be denoted "Cloud Detection Sensitivity" for the method and has previously been introduced as a useful concept by Karlsson and Håkansson (2018). We will use this concept here in the training and definition of the CMAprob method.

A practical solution to the desire of finding the optimal training of the method (according to the principles introduced above) can be found if also adding the following criterion:

> A binary cloud mask created by thresholding the cloud probability at the 50 % probability level should give maximum detection skill scores compared to a cloud mask derived by any other probability threshold.
>
> [1]

This criterion is not only necessary for solving the training problem but it also means that potential users of the product are given a clear recommendation on how to use the cloud probability results. Also, the above criterion should ideally be valid at every geographic location. A user can definitely adjust the threshold making classification results more clear-conservative (lowering the threshold) or more cloud-conservative (raising the threshold). However, the default threshold value at 50 % would be a reasonable starting point since the

user would then know that by using this threshold the detection skill score should theoretically be maximized. For any other threshold the rate of misclassifications will increase.

The most natural detection skill score to use in this context is the *Hit Rate* which simply is described as the percentage of correct cloudy and clear predictions with respect to the total number of predictions.

To illustrate how we can use this skill score in the training process we can study Fig. 3-5 taken from Karlsson and Johansson (2013). They introduced a plot on how the *Hit Rate* and other skill scores (in this example the *Kuipers Skill Score)* could vary as a function of the filtered cloud optical thickness. Basically, this shows how well the results of a specific method (in Fig. 3-5 cloud masks produced for the CLARA-A1 climate data record) agree with restricted (or original) CALIOP cloud masks. Notice that filtering does not mean that data is removed from the validation dataset. Instead, clouds with optical thicknesses below the filtered cloud optical thickness value are interpreted as being non-existent (i.e., changed into cloud-free).

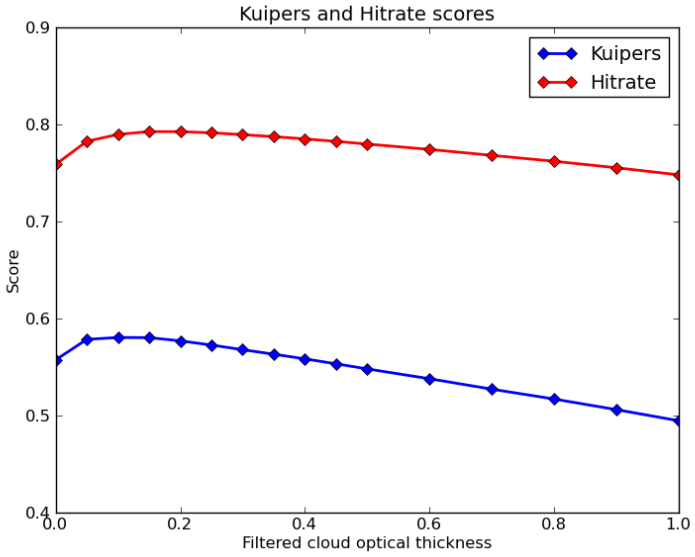

Figure 3-5 *Hitrate and Kuipers skill scores as a function of filtered CALIOP cloud masks (cloud optical thickness limits) for CLARA-A1/PPS 2014 cloud masks. Results derived from 99 collocated NOAA-18 and CALIPSO-CALIOP cloud masks in the period 2006-2009. (From Karlsson and Johansson, 2013).*

Figure 3-5 shows how skill scores first improves if filtering the thinnest (mostly non-detectable) clouds from the validation dataset but then decreases again after reaching a certain value of the filtered optical thickness. The decrease is explained by the fact that more correctly detected CALIOP-observed clouds are here converted to cloud-free (lowering the Hit rate) compared to how many missed clouds are converted into cloud-free (raising the Hit rate). At the maximum value for the Hit Rate we get the best fit with a restricted filtered CALIOP cloud mask. In this example this maximum occurs for a filtered cloud optical thickness value of approximately 0.2 (or slightly lower for the Kuipers score). This optimal filtered value is identical to the Cloud Detecting Sensitivity parameter introduced by Karlsson

and Håkansson (2018). They calculated it for all locations on the globe when validating results of the CLARA-A2 climate data record (Karlsson et al., 2016).

An important feature of the Hit Rate curve in Fig 3-5 is that it possesses important information about the probability of detecting a cloud with a certain optical thickness. If calculating the Hit Rate gradient from the values at finite intervals of filtered cloud optical thicknesses, we can deduce where the probability of detection reaches 50 %. More clearly, it means that, when the gradient is zero at the point of maximum Hit Rate, we have reached the situation when the probability of detection of clouds is exactly 50 %. If the Hit Rate stays constant after filtering (i.e., converting CALIOP clouds to cloud-free) a certain amount of clouds, it means that 50 % of those clouds were initially correctly classified. For intervals with lower cloud optical thicknesses the probability of detection is always lower and on the other side of the Hit Rate maximum it is always higher. If we imagine that our finite intervals (shown in Figure 3-5 as the distance between individual values on the curve) decrease towards infinitesimal widths we can then also say that we have a probability of detection of 50 % for cloud layers with an optical thickness of exactly this filtered cloud optical thickness value. We will utilise this property in the training process for CMA-prob since it has a direct link to criterion [1] above.

To find a solution which optimizes our cloud probability results, i.e., which gives us as small Cloud Detection Sensitivity values as possible everywhere on the globe, we need to add another level of complexity. The trick is to first do repeated training of the classifier using the full range of restricted cloud masks, i.e. full range of filtered cloud optical thicknesses. In the next step we then select the training statistics valid for that particular restricted cloud mask with a certain filtered optical thickness which coincides with the position of the Cloud Detection Sensitivity (i.e., position for maximum Hit Rate). Then we can be sure that the 50 % probability of detection is valid for this particular Cloud Detection Sensitivity which means that we are fulfilling criterion [1] above.

To explain this better we can consider two cases:

1. Hit rate peaks at smaller optical thicknesses than the filtered value used for the restricted cloud mask during training.
2. Hit rate peaks at larger optical thicknesses than the filtered value used during training.

Both cases mean that we are actually not succeeding very well in reproducing the restricted cloud mask that was used during training. The first case means that the used training statistics appears to be valid also to some extent for clouds with smaller optical thicknesses. Therefore, we should try to use a lower filtered optical thickness value during training. The second case means that a significant portion of the clouds are misclassified (missed) and we should therefore use a larger value for the optical thickness threshold of the restricted cloud mask.

In practise, this means that we have to carry out this evaluation separately for every defined surface or region according to Tab. 3-4. In this way we will get different optimal Cloud Detection Sensitivities for each surface or region. This will then reflect the different separabilities of cloudy and cloud-free conditions existing globally as a function of the underlying surface characteristics. Consequently, even if the 50 % cloud probability in the end result will mean the same everywhere (as a basis for a binary cloud mask), the probability is with respect to different kind of clouds with different layer optical thicknesses which will vary

with surface and region. The ambition is that by this method we will be able to identify clouds which are radiatively significant, i.e., being identified for having a sufficient contrast against the surface for each region. If combining all these results globally we would then get improved results compared to the case when we just unconditionally train our method against the original and unrestricted CALIOP cloud mask. To test if this is finally achieved, we have in the end to prove that overall validation results when validating against the unrestricted CALIOP cloud mask is better than results based on training exclusively using the original cloud mask. The latter case should give a higher rate of misclassifications if the concept is working.

3.7 Resulting sets of optimal training statistics for different Earth surfaces

The collected training dataset spans the period 2006-2015 and provides a reasonable global coverage over all seasons during that period. To notice, however, is that one year (2010) has been excluded from the training dataset to constitute an independent validation dataset. The training dataset is collected from almost 3000 NOAA-18 and NOAA-19 AVHRR Global Area Coverage (GAC) orbits and CALIOP pixels/samples at approximately 5 km horizontal resolution. The CALIPSO-CALIOP Cloud Layer (CLAY) product version 4.01 has been used. A more detailed description of this product is given by Karlsson and Håkansson (2018). The constrained training (i.e., image features now being related to PPS threshold information) is based on results from the PPS software version 2018 [RD 1]. This PPS version is highly advanced compared to the original method described by Dybbroe et al. (2005). The main new features of the method concern adaptations to global processing (e.g., over desert and Polar Regions) and a systematic use of prescribed MODIS-derived surface emissivity information.

Figures 3-6 – 3-9 below illustrate the selection of optimal training datasets over different surfaces and regions. Hitrate scores are here plotted by different curves representing training constellations using different restricted CALIOP cloud masks (explained in the legend). Notice that those training constellations giving highest cloud detection sensitivities (i.e., the capability to detect the thinnest cloud optical thicknesses) and which fulfil criterion [1] above are highlighted by thicker lines (in both figure and legend). In Figure 3-6 we study conditions for the surface having the best overall results of all surfaces globally, namely ice-free ocean surfaces over the extra-tropics. We deduce that over these surfaces we can use training data from a slightly restricted CALIOP cloud mask (τ-filtering at 0.05) during day and from the unrestricted original CALIOP cloud mask during night. Thus, the cloud detection sensitivity will then be 0.05 or even smaller over this surface.

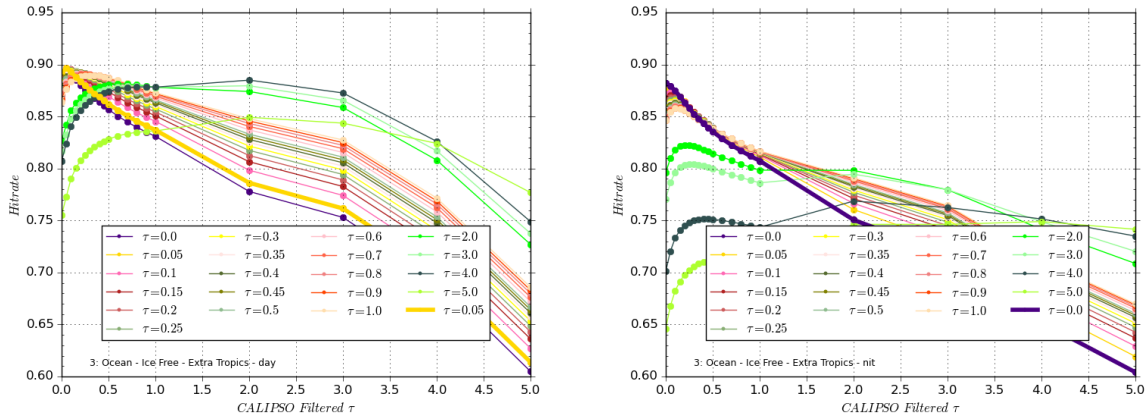

Figure 3-6 *Hit Rate as a function of filtered cloud optical thicknesses of the CALIOP cloud mask for different training scenarios (coloured curves) over ice free extratropical ocean during day (left) and night (right). The coloured curves describe results based on different restricted CALIOP cloud masks used when training. Thick lines denote possible solutions fulfilling the criterion that Hit Rate should be maximized for the same filtered optical thickness as was used during training.*

Conditions are more problematic over snow-free extratropical land surfaces as shown in Fig. 3.7. Best results are here found during daytime for a filtered cloud optical thickness of 0.6. Also solutions for larger thicknesses are proposed but highest Hit Rates are given for the lowest of the suggested solutions which should be chosen. Night time results are less decisive since we cannot find an obvious case where the Hit rate peak coincides with a particular trained filtered optical thickness. But the highest scores are still given for almost the same chosen filtered optical thickness as for daytime conditions and we decided to use this value also at night. It should be noted that the perfect solution at night maybe is not covered by the different training constellations we have tested. A finer resolution of restricted CALIOP masks could perhaps have suggested a cloud optical thickness value of 0.65.

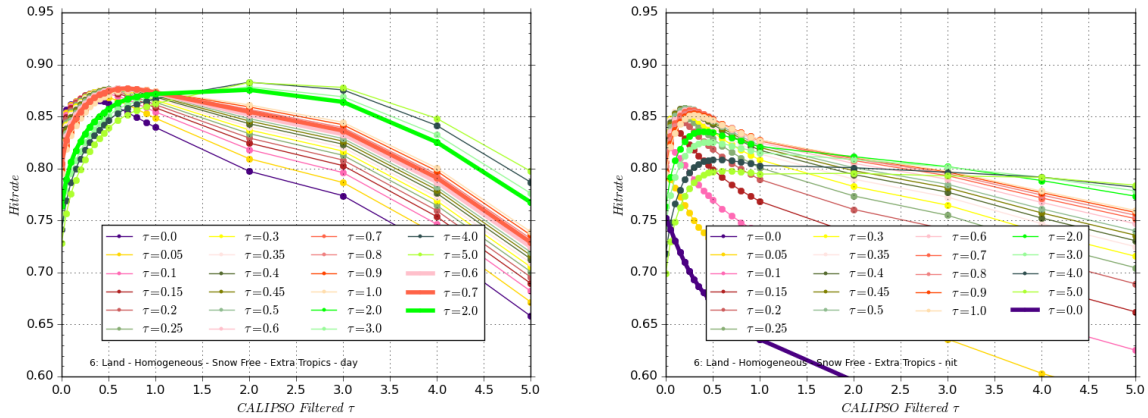

Figure 3-7 *Same as Figure 3-6 but for snow-free extratropical land surfaces during day (left) and night (right).*

Looking at results over well-known problematic areas, such as areas with permanent snow-cover (Greenland and Antarctica) in Fig. 3-8, the situation becomes less clear. Here it is difficult to find any clear guidance on what solution to choose when training. No clear peaks

in Hit rate seem to coincide with the chosen filtering level used when training. This indicates that, despite having seemingly homogeneous surface conditions, the true variability in results is very large over those surfaces. This can also be explained by existing contradicting results from different image features. If the necessary contrast needed for good cloud discrimination is found in one feature the situation can be completely the opposite for another image feature. This is a drawback of the Naïve Bayesian method, i.e. due to the multiplication of individual probabilities a near-zero value for one image feature will be able to compensate or neutralise results from other features.

The problem with the results in Fig. 3-8 is that no distinct peak can be found coinciding with the trained filtered cloud optical thickness. For daytime conditions a preference for using a cloud detection sensitivity of about 1.0 can be seen but it is a bit alarming to notice that Hit Rates for this value are not clearly decreasing for larger filtered optical thicknesses. This means that also for higher optical thicknesses almost 50% of all clouds will be missed. A better solution would be to choose a lower trained optical thickness value which is having a distinct peak but at a somewhat lower Hit Rate value. A reasonable compromise could be to use the value 0.3. The peak in Hit rate is still reasonably high and a clear capability to detect clouds with larger optical thicknesses is retained.

For night conditions in Fig. 3-8 (right), the results are even more confusing. At first sight a solution using the maximum filtered value of 5.0 is suggested. But in practice it means that we will never reach 50 % probability of detection for any value of the layer optical thickness in the studied interval since the curve for the filtered value of 5.0 is monotonically increasing over the full range of filtered optical thicknesses. There are, however, distinct peaks in Hit rate for other filtered optical thickness (e.g. in the range 0.2- 0.8) but then having much lower Hit Rates than what is indicted for the value 5.0 at the maximum filtering level. A compromise solution here could be to use the value 2.0 to acknowledge that situations should be more difficult than during daytime but not as extreme as suggested by the solution of using the value 5.0 for the restricted cloud mask. We conclude that over the most problematic surfaces we have to be careful using the general and idealized concept presented earlier since conditions are much more complex than over other Earth surfaces. To find the most optimally reproduced CALIOP cloud mask seems very difficult here which forces us to some compromise solutions.

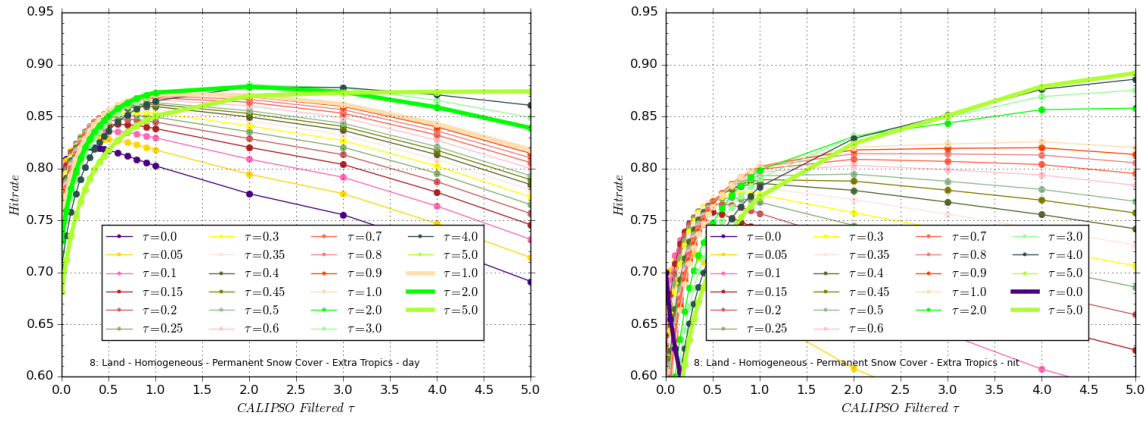

Figure 3-8 *Same as Figure 3-6 but for homogeneous land surfaces with permanent snow-cover during day (left) and night (right).*

Finally, we will also take a closer look at dry surfaces with sparse vegetation in Figure 3-9. We notice generally very high Hit Rates during daytime but also that optimal detection capabilities are reached for slightly higher optical thickness values compared to results for extra-tropical land surfaces in Fig. 3-7. Filtered optical thickness values increase here to 0.9 during day and similar (but less decisive) values during night. The observation that no clear solution can be found for the optimal cloud detection sensitivity during night probably indicates that conditions for desert surfaces are not as homogeneous and representative as initially assumed. The behaviour of the curves at night is very different compared to how it looks during day and over other surfaces. One reason for this could be the varying surface emissivities (influenced by both soil moisture conditions and surface material characteristics) in the infrared channels causing "cloud-like" temperature differences also in clear situations. Another reason could be that our reference surface skin temperatures from ERA-Interim do not capture minimum temperatures at night very well (which is also problematic for the polar winter surfaces described earlier in Fig. 3-8 right). However, the message is clear that the use of higher filtered optical thicknesses will yield improved results and less misclassifications. Thus, we can safely use the same value of the filtered optical thickness for training CMa-Prob as was selected for daytime conditions.

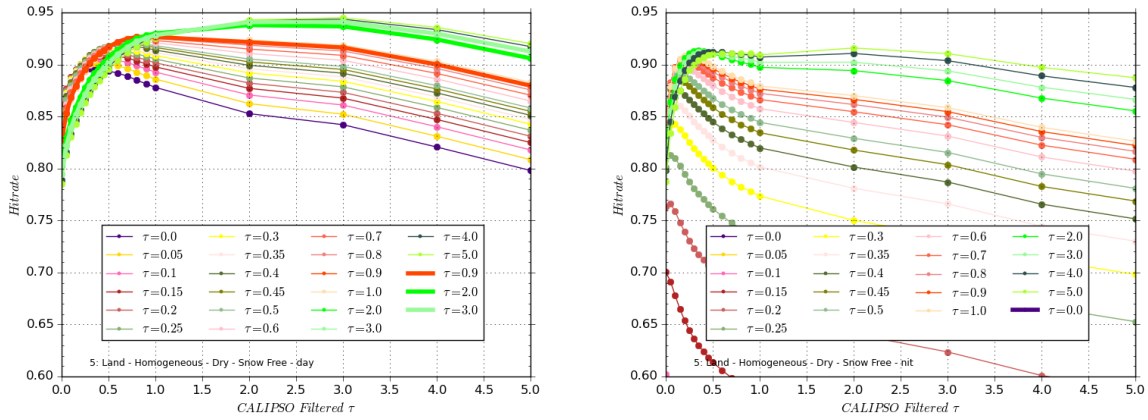

Figure 3-9 *Same as Fig. 3-6 but for dry land surfaces (dominated by deserts) with sparse vegetation during day (left) and night (right).*

We summarize the estimated optimal cloud detection sensitivities for all investigated surfaces in Tab. 3-5. The final statistics for the CMa-Prob classifier is compiled by choosing trained statistics for each optimal cloud detection sensitivity for each individual surface. For the

Table 3-5 *Estimated optimal Cloud Detection Sensitivities (i.e., lowest cloud layer optical thicknesses with probability of detection exceeding 50 %) for different Earth surfaces (defined in Table 3-4) and for different illumination categories. For these filtered optical thicknesses the best resemblance is achieved with the CALIOP cloud mask over different surfaces. Consequently, corresponding filtered CALIOP cloud masks are used to train the CMA-Prob classifier.*

SURFACE DESCRIPTION	Surface id	DAY	NIGHT	TWILIGHT
Marginal sea ice at high latitudes	G1	0.00	0.00	0.00
Sea ice at high latitudes	G2	0.10	1.00	0.50

Extra-tropical ice-free ocean	G3	0.00	0.00	0.00
Tropical ocean	G4	0.10	0.10	0.10
Dry homogeneous snow-free land	G5	0.90	0.90	0.90
Homogeneous, extra-tropical and snow-free land	G6	0.60	0.60	0.60
Homogeneous, extra-tropical land with seasonal snow	G7	0.15	1.00	0.15
Homogeneous extra-tropical land with permanent snow	G8	0.30	2.00	1.00
Rough, dry and snow-free land	G9	0.30	2.00	1.00
Rough, extra-tropical and snow-free land	G10	0.70	0.70	0.70
Rough, extra-tropical land with seasonal snow	G11	0.20	1.00	0.30
Rough, extra-tropical land with permanent snow	G12	0.30	2.00	1.00
Homogeneous, tropical land with vegetation	G13	0.20	0.10	0.20
Rough, tropical land with vegetation	G14	0.25	0.20	0.25

twilight category in Table 3-5 the CMa-Prob classifier selects either night-time or daytime statistics depending on if a detectable reflection signal can be found in AVHRR channel 1 at 0.6 micron. Thus, no specific twilight statistics is compiled but the used cloud detection sensitivity may still be different from the pure night and day case.

4 Final implementation of CMa-Prob and some demonstrated results

4.1 Demonstration of impact of using cloud detection sensitivity statistics instead of statistics based on original CALIOP cloud mask

Some results from selecting training statistics differently over various Earth surfaces are illustrated in this sub-section for some image features.

Figure 4-1 shows the achieved training statistics for the visible reflectance feature (Rvis in Tab. 3-2) over dry and homogeneous land surfaces (category G5 in Tab. 3-5). The left part of the figure shows the cloud probability distribution as a function of reflectance after training with the cloud detection sensitivity value 0.9 (according to Tab. 3-5). The right part shows statistics compiled using the original unfiltered CALIOP cloud mask. The latter shows that cloud occurrences are not negligible for low-to-moderate reflectivities even if the majority of clouds occur for higher reflectivities. However, after training against a restricted cloud mask, filtering clouds with optical thicknesses below 0.9, the cloud probabilities at low-to-moderate reflectivities are considerably reduced (Fig. 4-1 left). It means that some of the very thin clouds which previously were mixed up with the cloud-free reflectance are now treated as cloud-free cases. This improves the overall separability of cloudy and cloud-free cases. The peak of cloud occurrences seen at very low reflectances is most likely explained by shadows cast on other lower-level clouds.

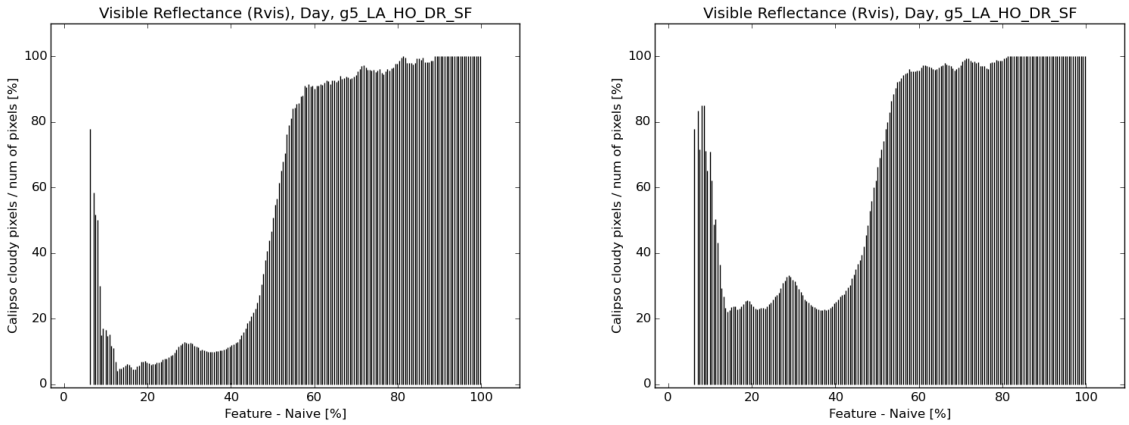

Figure 4-1 *Distribution of cloud occurrences (or cloud frequencies) as a function of the visible AVHRR reflectance at 0.6 micron over **dry homogeneous surfaces** (predominantly desert surfaces). Left: Statistics based on training with a CALIOP cloud mask filtered at optical thickness 0.9. Right: Statistics based on training with the original unfiltered CALIOP cloud mask.*

Figure 4-2 shows the same type of results as in Fig. 4-1 but now for the reflectance in the 3.7 micron channel (short-wave infrared). Again we can see how the removal of the thin clouds (i.e., meaning that we now interpret them as cloud-free) below optical thickness 0.9 increases the separability between cloudy and clear cases. Clouds are in this spectral region either weakly reflecting (ice clouds) or strongly reflecting (water clouds). The inter-mediate region in the figure (i.e., valid for reflectivites in the interval 15-35 %) is dominated by moderately reflecting desert surfaces. If not filtering out the thinnest clouds they are easily mixed up with the surface as seen by the non-zero cloud frequencies here. But after filtering, the risk of misclassifying (remaining) clouds is reduced. We repeat that by filtering out the thin clouds we reduce the capability to detect any of these clouds but on the other hand we will now minimise the risk of creating false clouds with more or less the same spectral signature. The use of the cloud detection sensitivity parameter for guiding us with the filtering procedure will guarantee that we gain more than we lose by this filtering procedure.

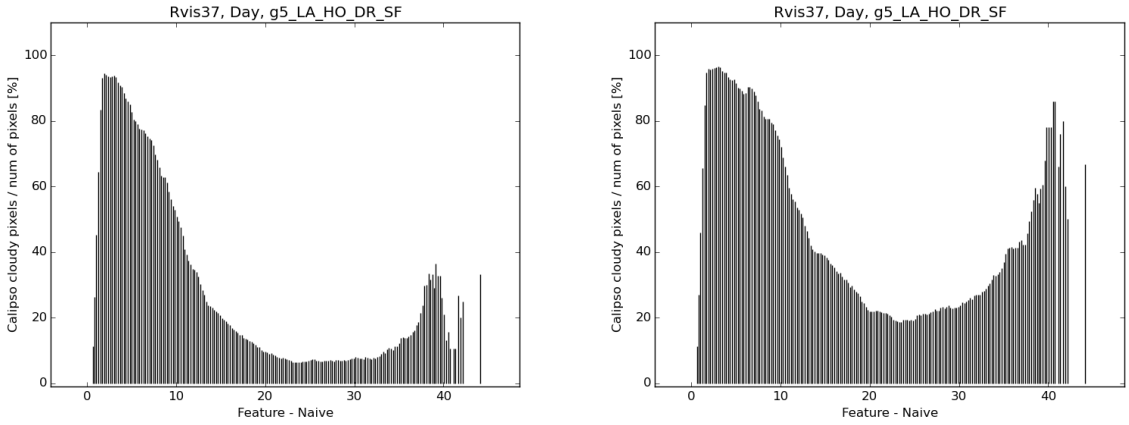

Figure 4-2 *Distribution of cloud occurrences (or cloud frequencies) as a function of the AVHRR reflectance at 3.7 micron (%) over **dry homogeneous surfaces** (predominantly*

*desert surfaces). **Left**: Statistics based on training with a CALIOP cloud mask filtered at optical thickness 0.9. **Right**: Statistics based on training with the original unfiltered CALIOP cloud mask.*

A final example of the effect of filtering is shown in Fig. 4-3 for the absolutely most problematic cloud detection condition encountered for AVHRR data: night-time cloud detection over cold and snow-covered surfaces (e.g., polar night over Greenland and Antarctica). Fig. 4-3 shows conditions when strong surface temperature inversions are present which is defined as when the surface temperature is colder than the temperature in the 950 hPa level. The figure shows the cloud probability as a function of the difference between the AVHRR brightness temperature at 11 micron and the surface temperature from ERA-Interim. The right part shows results when training against the unfiltered CALIOP cloud mask while the left part shows results when training against a CALIOP cloud mask filtered at optical thickness 2.0.

It is clear that for unfiltered training the probability of cloudy conditions is high for almost all conditions, except for cases when the surface temperature is much colder than the measured 11 micron brightness temperature (i.e., for large negative differences). However, probabilities are generally uncertain (i.e., close to 50 %) and only for restricted parts of the distribution do we find high and low probabilities which could contribute favourably to the cloud detection process. After filtering with a cloud optical thickness of 2.0, a large fraction of all clouds disappear (i.e., are now interpreted as cloud-free) and only clouds which are clearly colder than the ERA-Interim surface temperature remain with cloud probabilities clearly above 50 % (Fig. 4-3 left). It means that only optically thick and cold clouds are possible to detect with confidence over this surface type for this image feature.

The unfiltered results (Fig. 4-3, right) indicate some skill in identifying also clouds which are warmer than the surface (e.g., “black stratus”) for temperature differences near -20 K. The probabilities for remaining clouds of this type after the filtering is still non-zero but generally lower than 50 %. Successful identification of these clouds now depend on if also cloud probabilities are high enough for the other two infrared image features ((Twdiff and Tcidiff in Tab. 3-3). For the black stratus clouds Twdiff is important since it is normally showing a negative temperature difference between AVHRR channels at 3.7 microns and 11 microns as opposed to clear areas and ice clouds (the latter showing a positive temperature difference). But for very cold situations this typical cloud feature becomes less reliable which is caused by the transition from pure water clouds at higher temperatures to mixed phase clouds at cold temperatures. In addition, increasing radiometric noise at the 3.7 micron channel (leading to random positive and negative Twdiff values) will further decrease the usefulness of the Twdiff feature. In conclusion, cloud detection capabilities in AVHRR data over cold surfaces during the polar night remains as the most challenging of tasks for any cloud detection scheme.

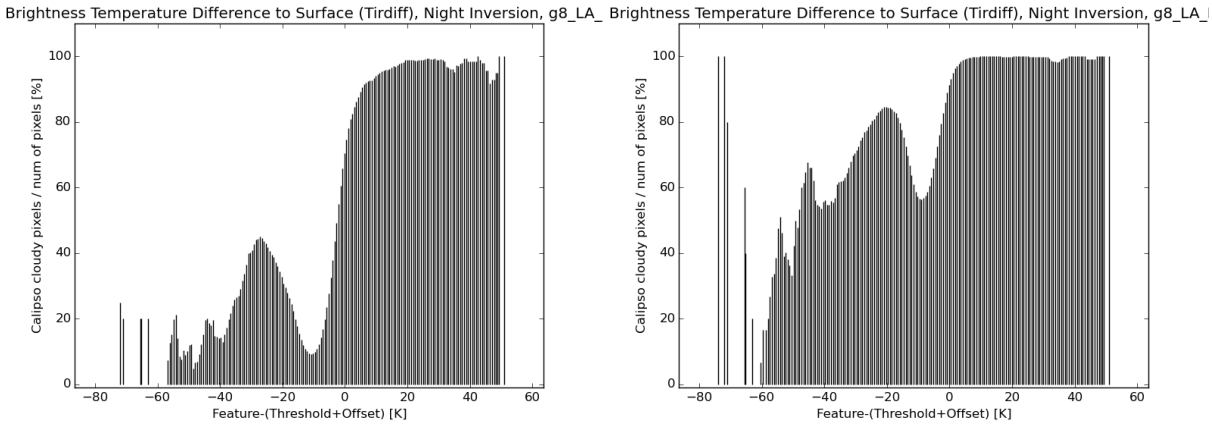

Figure 4-3 *Night-time distribution of cloud occurrences (or cloud frequencies) as a function of the difference between the AVHRR brightness temperature at 11 micron and the surface temperature from ERA-Interim over* ***permanently snow-covered surfaces****. These results were collected for situations with* ***strong surface temperature inversions****.* ***Left****: Statistics based on training with a CALIOP cloud mask filtered at optical thickness 2.0.* ***Right****: Statistics based on training with the original unfiltered CALIOP cloud mask.*

4.2 Treatment of data from satellites with the 1.6 micron channel replacing the 3.7 micron channel

Table 3-2 lists also the image feature Rswir_3a which is based on the reflectivity in the AVHRR channel at 1.6 microns divided by corresponding reflectivities in the AVHRR channel at 0.6 microns. Channel 3a is only available operationally (with only a few exceptions) from satellites operating in morning orbits (e.g., NOAA-17, METOP-A and METOP-B). Data from satellites in morning orbit can also be collocated with CALIOPSO-CALIOP data but only for positions close to latitude 70 degrees. This means that Rswir_3a statistics cannot be collected with global coverage. This is particularly problematic for the dry land surface category (G5 in Table 3-5) when considering that the surface reflectivity at 1.6 microns is high over desert surfaces risking bing mixed up with corresponding cloud reflectivities.

We have overcome this problem by training CMa-Prob against corresponding 1.6 micron radiances measured by the MODIS instrument carried by the Aqua satellite. This can be done since the spectral responses at this particular channel (and also for the channel at 0.6 micron used to calculate the reflectance quota) are very similar for the AVHRR and MODIS sensors. The Aqua satellite is part of the A-train and thus offers almost simultaneous and continuous observations with CALIPSO-CALIOP. We used one year of MODIS data (2010) for training the CMa-Prob method.

Figure 4-4 shows resulting daytime Rswir_3a probability distributions for desert surfaces (left) and for surfaces with permanent snow-cover (right). Results are based on training with the unfiltered CALIOP cloud mask. We notice a very distinctive cloud signature in both cases showing that this image feature can be used with great confidence in the cloud screening process. This feature appears actually more reliable for detection of clouds over desert surfaces than the previously described Rvis37 feature (see Fig. 4-2 right). Distributions for the latter has a much more serious mix between cloudy and clear radiances (e.g., in the interval

15-35 % reflectivity in Fig. 4-2). The high cloud detection capability of this feature over snow-covered surfaces is also clearly seen in Fig. 4-4 (right). This was the major reason for introducing this channel historically.

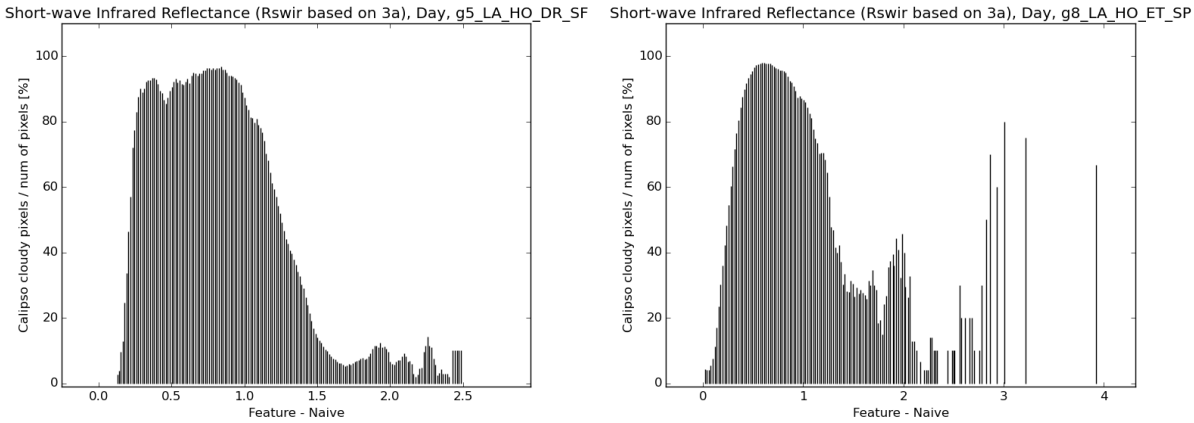

Figure 4-4 *Daytime distribution of cloud occurrences (or cloud frequencies) as a function of the reflectance quota (Rswir_3a in Tab. 3-2) between AVHRR-heritage channels at 1.6 micron and 0.6 microns based on Aqua Modis data.* ***Left****: Distribution over* ***dry surfaces*** *(surface category G5 in Tab. 3-5).* ***Right****: Distribution over* ***surfaces with permanent snow-cover*** *(surface category G8 in Tab. 3-5). All statistics are calculated from collocations with the original unfiltered CALIOP cloud mask.*

4.3 Product demonstration

The CMa-Prob product is available in PPS version 2018 for both high resolution AVHRR data (HRPT, 1 km) and reduced resolution Global Area Coverage (GAC, 5 km) data. It is also possible to produce CMa-Prob results for MODIS and VIIRS data, although currently restricted to only being based on information from AVHRR-heritage channels. For MODIS data, statistics is available trained on actual MODIS data for one full year (2010). In the case of VIIRS, the product is still based on statistics derived from AVHRR data. Representative training data for VIIRS from collocations with CALIPSO-CALIOP remains to be compiled in the near future.

Figure 4-5 illustrates an AVHRR GAC case with the full cloud probability result displayed as a greyscale image together with a colour composite image of the original radiances. We notice from visual inspection that the areas with high CMa-Prob cloud probabilities (white colours) correspond very well to cloud fields identified by visual inspection in the RGB composite for this particular case. However, noteworthy is that thin and broken cloud fields over the ocean surfaces are much more highlighted in the CMa-Prob image than in the colour composite.

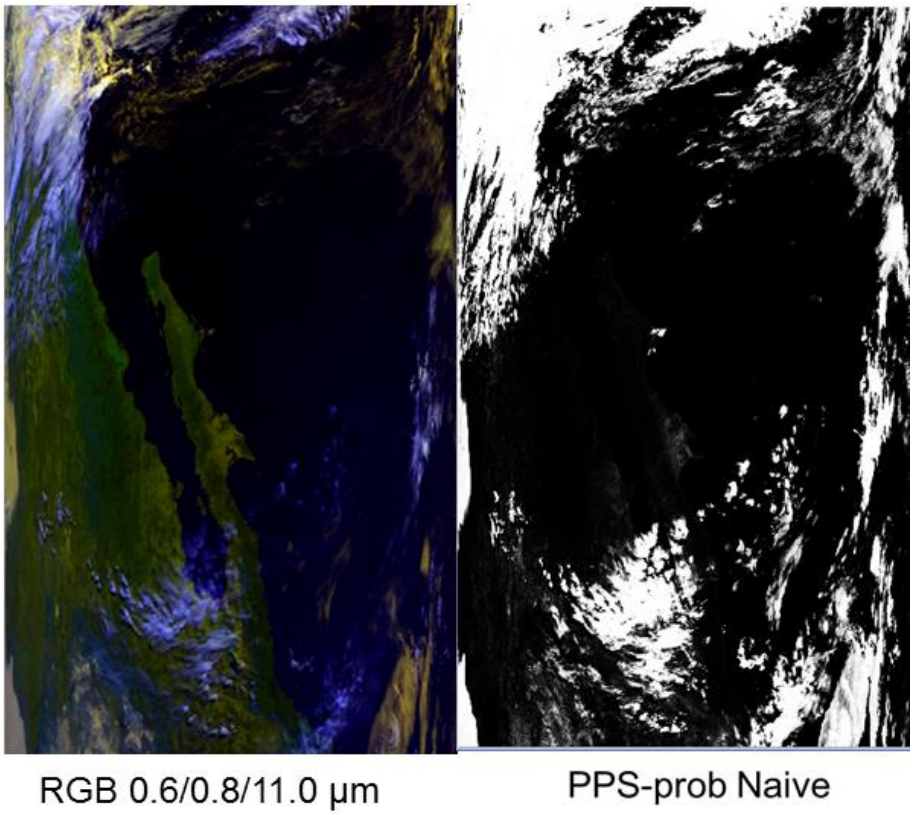

Figure 4-5 *Part of an original NOAA-18 AVHRR GAC scene in satellite projection over the North American west coast (with Gulf of California and Baja California in the center) registered in ascending mode (i.e., North is down, South is up) from 26 January 2010.* ***Left****: Colour composite with AVHRR channel 1 (red), channel 2 (green) and channel 4 (blue).* ***Right****: Corresponding CMa-Prob cloud probabilities (as greyscale image with range 0-100 %).*

This is mainly explained by the added cloud information coming from features **Rvis37** and **Texture_day** described earlier in Tab. 3-2. These features contain information from the 3.7 channel which is information that is not displayed by the colour composite in the leftmost panel of 5. Thus, the CMa-Prob results are clearly based on more information than what is displayed in the RGB representation in Fig. 4-5.

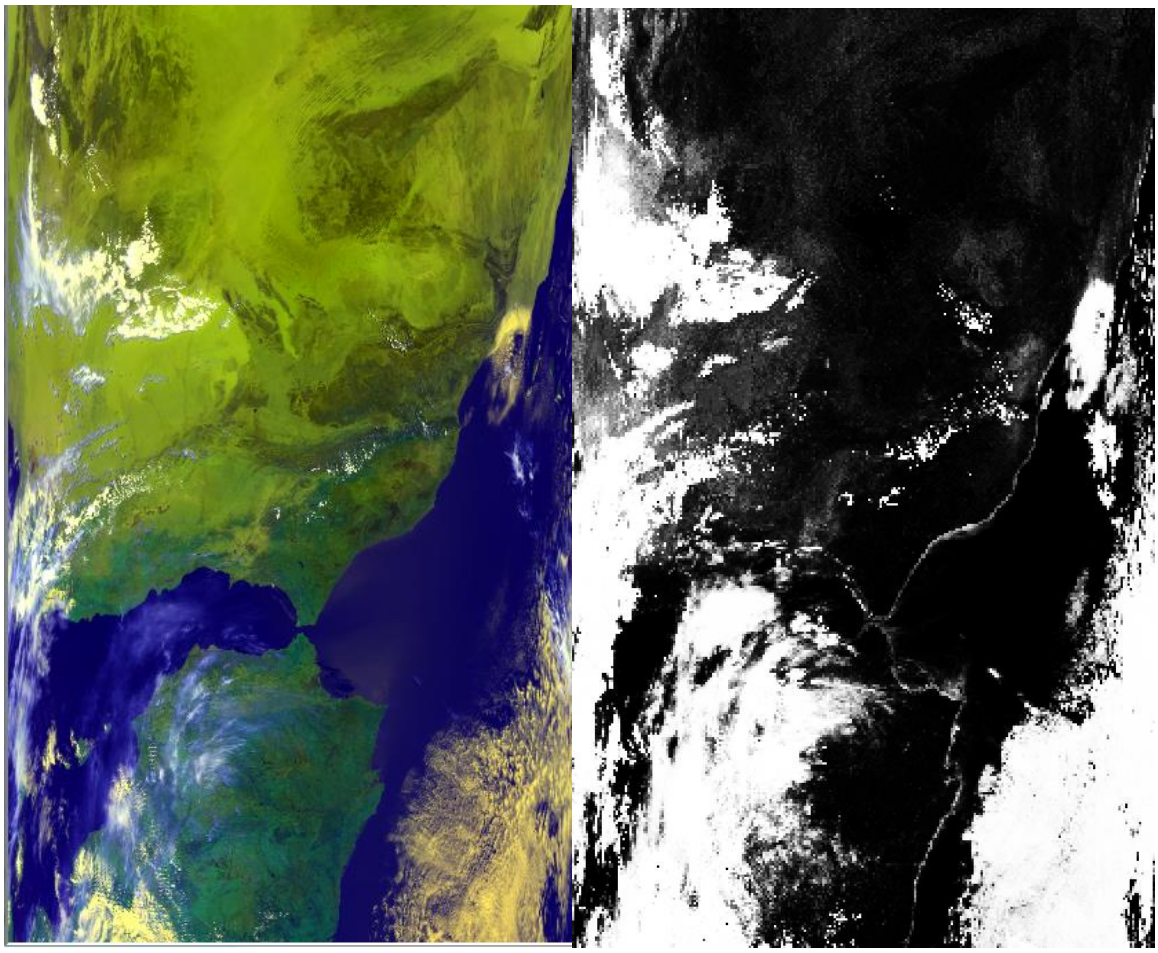

Figure 4-6 *Part of an original NOAA-18 AVHRR GAC scene in satellite projection over Spain and northern Africa registered in ascending mode (i.e., North is down, South is up) from 16 May 2007.* ***Left****: Colour composite with AVHRR channel 1 (red), channel 2 (green) and channel 4 (blue).* ***Right****: Corresponding CMa-Prob cloud probabilities (as greyscale image with range 0-100 %).*

Figure 4-6 shows a NOAA-18 case including portions of the desert regions of Northern Africa. This example illustrates how clouds and bright desert surfaces can be efficiently separated taking advantage of the full information content in all AVHRR channels. Despite observing over relatively bright desert surfaces (i.e., surface reflectances in visible channels are rather close to cloud reflectances here), the resulting cloud probabilities are distinctly at the zero level (black areas) for cloud free areas and close to 100 % (white areas) for cloudy areas. Again we find high cloud probabilities for thin cirrus cloud fields over Spain and the Mediterranean Sea. The most important information in this scene comes mainly from AVHRR channels 3 and 5, which are not displayed in the colour composite image. The problem of using contextual (texture) information in applications like this results in high cloud probabilities on the sea side of the coastlines (since texture features are only used over ocean surfaces). Consequently, probabilistic approaches also need to consider a special coastal treatment; this is included in the planned upcoming versions of CMa-Prob.

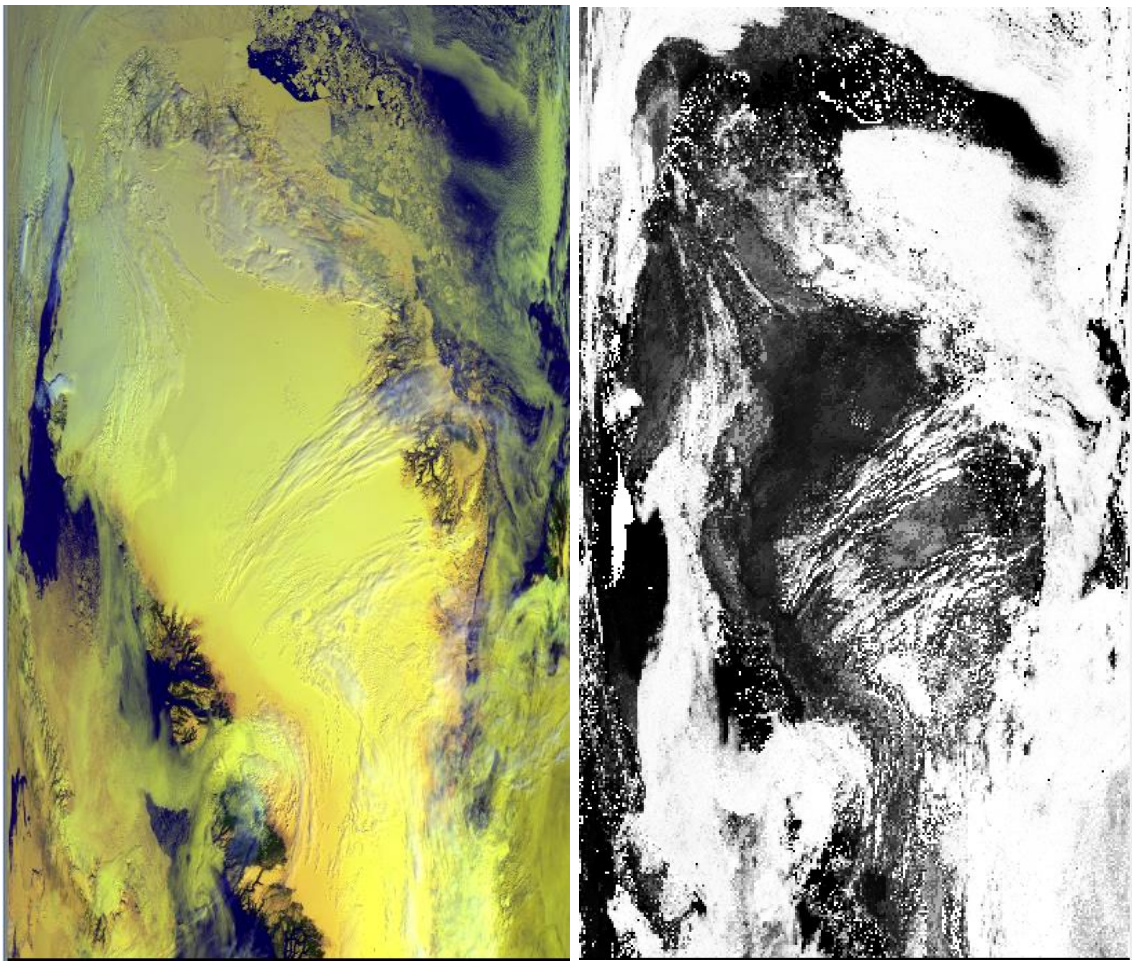

Figure 4-7 *Part of an original NOAA-17 AVHRR GAC scene in satellite projection over Greenland registered in descending mode (i.e., North is up, South is down) from 4 June 2009.* ***Left****: Colour composite with AVHRR channel 1 (red), channel 2 (green) and channel 4 (blue).* ***Right****: Corresponding CMa-Prob cloud probabilities (as greyscale image with range 0-100 %).*

Fig. 4-7 shows a NOAA-17 case over Greenland from 4 June 2009. This case clearly illustrates the strength of the ***Rswir_3a*** feature in Tab. 3-2. The colour composite shows how snow-covered surfaces and cloud features are hard to separate (if not using shadow effects) but taking into account also the information in AVHRR channel 3a makes this distinction very efficient in the CMA-prob image.

The previous figures have demonstrated results for rather well-illuminated and favourable atmospheric conditions. A final example is given in Figs. 4-8 and 4-9 showing a case where cloud separability conditions over different surfaces start to vary a lot and where the usefulness of different statistics over different surfaces is clearly visible in the results. Notice that Fig. 4-9 shows an overestimation of the differences between Greenland and the surrounding surfaces. This was done to highlight how probabilities may vary between different surfaces. According to Tab. 3-5 the recommended cloud detection sensitivity level for Greenland is 0.3 during daytime and 2.0 at night. These are the values used for the current CMa-Prob version in PPS version 2018. In Figs 4-8 and 4-9 a cloud detection sensitivity level

of 5.0 has been used to highlight the different probabilities over ice-free ocean and the Greenland snow cap.

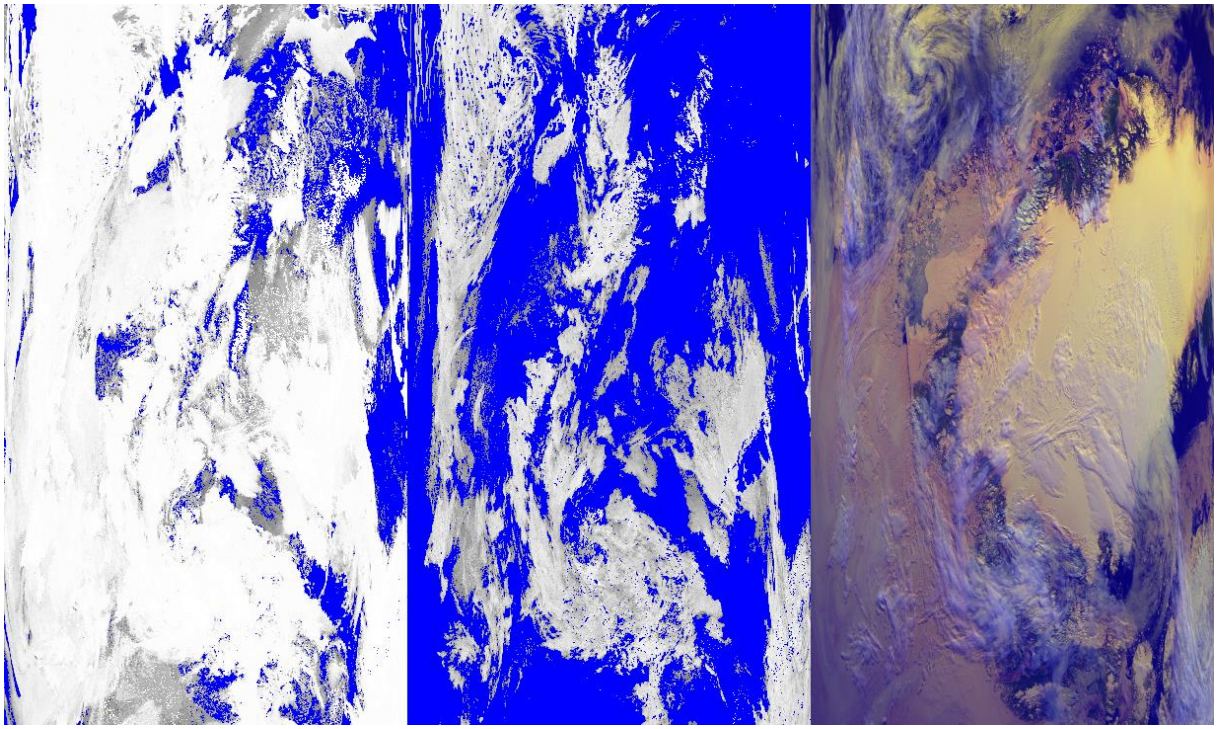

Figure 4-8 Left: *Resulting CMa-Prob probabilistic cloud mask in grayscales (blue = probability below 50 %) when training with unfiltered CALIOP mask.* ***Middle****: Same as left image but training with a CALIOP mask filtered at optical thickness 5.0.* ***Right:*** *RGB-image with AVHRR channels at 0.6 µm, 0.9 µm and 11 µm. AVHRR scene over Greenland (north is down!) from 22 June 2007 12:19 UTC.*

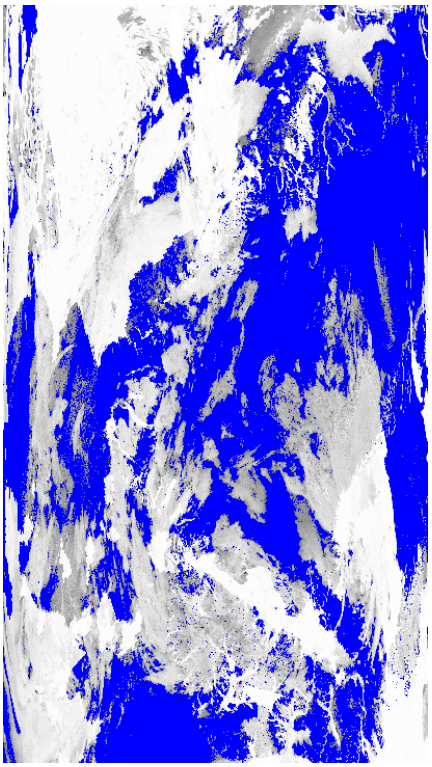

Figure 4-9 *Final cloud probability for the same case as in Figure 4-8 above. Blue areas have cloud probabilities below 50 %. Resulting probabilistic cloud mask is continuous but probability levels change when surface conditions for the classifier change from good to poor (e.g. along Greenland coast).*

Validation results for the CMa-Prob method are presented in RD 2.

5 Limitations and areas for future improvements

The CMA-prob method described in this document is the third prototype version and this time based on dynamic threshold information from PPS version 2018. The first version was described in the paper by Karlsson et al. (2015) and was based on dynamic threshold information from PPS version 2010. The second version (briefly described by Karlsson et al., 2016) was based on dynamic threshold information from PPS version 2014. Thus, it is clear that the method should be seen as an extension of the official PPS software and it cannot be run independently from PPS. Consequently, if continuing with this approach the method needs to be updated (new training) for every new release of the PPS method.

Statistical methods are always limited by the amount of training data being used. It is clear from probability distributions shown in Sections 3.6 and 3.7 that the current training dataset is not fully capable of providing very well-defined probability distributions over all Earth surfaces and for all conditions. Additionally, for some sensors like VIIRS the method is currently executable based exclusively on AVHRR-derived statistics. Thus, thorough training with real VIIRS data has still to be done. In that respect, it is very encouraging that results are still as good as being documented (RD 2). However, it is very clear that an improved amount of training data would be beneficial for future versions of the method. Consequently, it is planned to extend the training material with considerably more data from CALIPSO up to

present date. Also ways of transferring the results between different satellite sensors will be considered (e.g., by use of Spectral Band Adjustment Factors (SBAFs, Bhatt et al., 2016).

If the visual channels (0.6 μm and 0.9 μm) are missing, when expected, the Cloud Probability will be set to no-data. This could eg. sometimes happen in twilight.

6 References

Bhatt, R., Doelling, D. R., Scarino, B. R., Gopalan, A., Haney, C. O., Minnis, P., & Bedka, K. M. (2016). A consistent AVHRR visible calibration record based on multiple methods applicable for the NOAA degrading orbits. Parts I&II; Methodology. J. Atmos. Ocean. Tech., 33(11), 2499-2515.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delso, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quart. J. Roy. Meteor. Soc.*, **137**, 656, (doi = 10.1002/qj.828), 553—597.

Derrien, M.and H. LeGléau, 2005: MSG/SEVIRI cloud mask and type from SAFNWC. *Int. J. Remote Sens.,* **26**, 4707-4732.

Dybbroe, A., A. Thoss and K.-G. Karlsson, 2005a: NWC SAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling - Part I: Algorithm description, *J. Appl. Meteor*, **44**, pp. 39-54.

Dybbroe, A., A. Thoss and K.-G. Karlsson, 2005b: NWC SAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling - Part II: Tuning and validation, *J. Appl. Meteor*, **44**, 55-71.

Frey, R. A., S. A. Ackerman, Y. Liu, K. I. Strabala, H. Zhang, J. R. Key, and X. Wang, 2008: Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for collection 5, *J. Atm. Ocean. Tech*., **25**, 1057-1072.

Heidinger, A.K., A. T. Evan, M. Foster and A.Walther, 2012: A Naïve Bayesian Cloud Detection Scheme Derived from CALIPSO and Applied within PATMOS-x. *J. Appl. Meteor. Climatol.,* **51**, 1129-1144.

Hogan, R.J., M. Ahlgrimm, G. Balsamo, A. Beljaars, P. Berrisford, A. Bozzo, F. Di Giuseppe, R.M. Forbes, T. Haiden, S. Lang, M. Mayer, I. Polichtchouk, I. Sandu, F. Vitart and N. Wedi, 2017: Radiation in numerical weather prediction, ECMWF Technical Memorandum, **816**, ECMWF Research, Forecast and Copernicus Departments.

Inoue, T. 1987: A cloud type classification with NOAA 7 split-window measurements. *J. Geoph. Res.,* **92**, 3991–4000, DOI: 10.1029/JD092iD04p03991.

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Meirink, J. F., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R., 2017: CLARA-A2: The second edition of the CM SAF cloud and radiation data record from

34 years of global AVHRR data, *Atmos. Chem. Phys.*, **17**, 5809-5828, www.atmos-chem-phys.net/17/5809/2017, doi: 10.519/acp-17-5809-2017.

Karlsson, K.-G. and A. Devasthale, 2018: Inter-comparison and evaluation of the four longest satellite-derived cloud climate data records: CLARA-A2, ESA Cloud CCI V3, ISCCP-HGM and PATMOS-x, *Rem. Sens.*, **10**, 1567, doi:10.3390/rs10101567.

Karlsson K.-G. and A. Dybbroe, 2010: Evaluation of Arctic cloud products from the EUMETSAT Climate Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations. *Atmos. Chem. Phys.,* **10**, 1789–1807, 2010.

Karlsson, K.-G., and N. Håkansson, 2018: Characterization of AVHRR global cloud detection sensitivity based on CALIPSO-CALIOP cloud optical thickness information : demonstration of results based on the CM SAF CLARA-A2 climate data record. *Atmos. Meas. Tech*., **11**, 633–649. https://doi.org/10.5194/amt-11-633-2018

Karlsson, K.-G. and E. Johansson, 2013: On the optimal method for evaluating cloud products from passive satellite imagery using CALIPSO-CALIOP data: example investigating the CM SAF CLARA-A1 dataset. *Atmos. Meas. Tech.,* **6**, 1271–1286, www.atmos-meas-tech.net/6/1271/2013/ , doi:10.5194/amt-6-1271-2013.

Karlsson, K.-G., E. Johansson and A. Devasthale, 2015: Advancing the uncertainty characterisation of cloud masking in passive satellite imagery: Probabilistic formulations for NOAA AVHRR data, *Rem. Sens. Env.* , **158**, 126-139; doi:10.1016/j.rse.2014.10.028.

Kossin, J. P. and M. Sitkowski, 2009: An objective model for identifying secondary eyewall formation in hurricanes. *Mon. Wea. Rev.,* **137**, 876-892.

Kriebel, K. T., G. Gesell, M. Kästner and H. Mannstein, 2003: The Cloud Analysis Tool APOLLO: Improvements and Validations. *Int. J. Remote Sens*., 01/2003; **24**(2003):2389-2408. DOI:10.1080/01431160210163065.

Merchant, C. J., A. R., Harris, E. Maturi and S. MacCallum, 2005: Probabilistic physically-based cloud screening of satellite infra-red imagery for operational sea surface temperature retrieval, *Quart. J. Royal Met. Soc.,* **131**, 2735-2755.

Musial, J. P., F. Hüsler, M. Sütterlin, C. Neuhaus and S. Wunderle, 2014: Probabilistic approach to cloud and snow detection on Advanced Very High Resolution Radiometer (AVHRR) imagery. *Atmos. Meas. Tech.,* **7**, 799-822, www.atmos-meas-tech.net/7/799/2014/, doi:10.5194/amt-7-799-2014.

Pavolonis, M. J.; A. K. Heidinger and T. Uttal, 2005: Daytime global cloud typing from AVHRR and VIIRS: Algorithm description, validation, and comparisons. *Journal of Applied Meteorology*, **44**, Issue 6, 804-826.

Rossow, W. B., and R. A Schiffer, 1999: Advances in Understanding Clouds from ISCCP. Bull. Amer. Meteor. Soc., 80, 2261–2288.

Stengel, M., S. Mieruch, M. Jerg, K.-G. Karlsson, R. Scheirer, B. Maddux, J.F. Meirink, C. Poulsen, R. Siddans, A. Walther and R. Hollmann, 2013: The Clouds Climate Change Initiative: Assessment of state-of-the-art cloud property retrieval schemes applied to AVHRR heritage measurements, *Rem. Sens. Env.*, http://dx.doi.org/10.1016/j.rse.2013.10.035.

Winker, D.M., M.A.Vaughan, A. Omar, Y. Hu, K.A. Powell, Z. Liu, W.H. Hunt and S.A. Young, S.A, 2009: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms. *J. Atm. Ocean. Tech*. **26**:11, 2310-2323.

Young, A. H., K.R. Knapp, A. Inamdar, W. Hankins and W.B. Rossow, 2018. The International Satellite Cloud Climatology Project H-Series climate data record product, Earth Syst. Sci. Data, **10**, 1-11, https://doi.org/10.5194/essd-10-583-2018

7 Glossary

ATBD	Algorithm Theoretical Baseline Document
AVHRR	Advanced Very High Resolution Radiometer
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization (CALIPSO)
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CDOP	Continuous Development and Operations Phase
CLARA	CMSAF cLoud, Albedo and surface RAdiation dataset
CMA-prob	Cloud Mask (probabilistic)
CM SAF	Satellite Application Facility on Climate Monitoring
CPP	Cloud Physical Properties
DRI	Delivery Readiness Inspection
DWD	Deutscher Wetterdienst (German MetService)
ECMWF	European Centre for Medium Range Forecast
ECV	Essential Climate Variable
EPS	European Polar System
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FOV	Field of view
GAC	Global Area Coverage (AVHRR)
GCOS	Global Climate Observing System
IOP	Initial Operations Phase
ITCZ	Inter-Tropical Convergence Zone
KNMI	Koninklijk Nederlands Meteorologisch Institut
NASA	National Aeronautics and Space Administration
NDBC	National Data Buoy Center
NESDIS	National Environmental Satellite, Data, and Information System
NOAA	National Oceanic & Atmospheric Administration
NODC	National Oceanographic Data Center
NSIDC	National Snow and Ice Data Center
NWCSAF	Satellite Application Facility for Nowcasting
NWP	Numerical Weather Prediction
PPS	Polar Platform System
PRD	Product Requirement Document
PUM	Product User Manual
RMIB	Royal Meteorological Institute of Belgium

RMS	Root Mean Square
RSMAS	Rosenstiel School of Marine and Atmospheric Science
RSS	Remote Sensing Systems
SAF	Satellite Application Facility
SMHI	Swedish Meteorological and Hydrological Institute
SST	Sea Surface Temperature