Atmos. Meas. Tech. Discuss., 7, 11045–11085, 2014
www.atmos-meas-tech-discuss.net/7/11045/2014/
doi:10.5194/amtd-7-11045-2014

Atmospheric Measurement Techniques Discussions
Open Access

Bayesian cloud detection for MERIS, AATSR, and their combination

A. Hollstein, J. Fischer, C. Carbajal Henken, and R. Preusker

Institute for Space Sciences, Department of Earth Sciences, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6–10, 12165 Berlin, Germany

Received: 30 September 2014 – Accepted: 20 October 2014 – Published: 6 November 2014

Correspondence to: A. Hollstein (andre.hollstein@fu-berlin.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

A broad range of different of Bayesian cloud detection schemes is applied to measurements from the Medium Resolution Imaging Spectrometer (MERIS), the Advanced Along-Track Scanning Radiometer (AATSR), and their combination. The cloud masks were designed to be numerically efficient and suited for the processing of large amounts of data. Results from the classical and naive approach to Bayesian cloud masking are discussed for MERIS and AATSR as well as for their combination. A sensitivity study on the resolution of multidimensional histograms, which were post-processed by Gaussian smoothing, shows how theoretically insufficient amounts of truth data can be used to set up accurate classical Bayesian cloud masks. Sets of exploited features from single and derived channels are numerically optimized and results for naive and classical Bayesian cloud masks are presented. The application of the Bayesian approach is discussed in terms of reproducing existing algorithms, enhancing existing algorithms, increasing the robustness of existing algorithms, and on setting up new classification schemes based on manually classified scenes.

1 Introduction

Cloud masking of Earth observation measurements is an important and often crucial part of various remote sensing retrievals. This includes, but is not limited to, the retrieval of cloud and aerosol micro-physical parameters, the estimation of cloud cover, ocean colour retrievals, and in general, algorithms which include atmospheric correction schemes. Cloud masking algorithms differ widely in their complexity, computational requirements, and in their assumption on what a cloud is and which physical process is exploited for their detection. Implementation of specific algorithms are often application specific which makes the cloud masks as well application specific and generally complicates the inter-comparison of results from different cloud masks.

AMTD
7, 11045–11085, 2014

This paper emphasizes the application of Bayesian methods for the cloud masking of the complete 9.5 year time series of the Medium Resolution Imaging Spectrometer (MERIS) (Rast et al., 1999) and the Advanced Along-Track Scanning Radiometer (AATSR) (Llewellyn-Jones et al., 2001) on-board the Environmental Satellite (ENVISAT) satellite and is part of the European Space Agency (ESA) Cloud CCI (Climate Change Initiative) activity (Hollmann et al., 2013). Thus, the requirements for the cloud masking scheme, which is described in Sects. 2 to 5, are robustness, accuracy, and computational efficiency. Several applications of the Bayesian method are discussed in Sect. 7. Results for MERIS and AATSR separately are discussed, but with having the focus on their combination within the Synergy product, where AATSR measurements are mapped on the MERIS swath and their mutual overlap is used. This Synergy dataset in combination with the presented cloud mask will be used for the retrieval of cloud micro-physical parameters using the FAME-C algorithm which was described by Carbajal Henken et al. (2014).

Major challenges of cloud detection are validation, the correct classification of scenes with clouds in over mountains and over snow and ice covered areas, and the distinction between clouds and optically thick aerosol plumes such as dust storms. These points are discussed in detail in Sect. 7.

Common approaches to cloud masking are hierarchies of thresholds (e.g. Rossow and Garder, 1993; Schlundt et al., 2011), complex statistical models (e.g. Murtagh et al., 2003; Gómez-Chova et al., 2008), or other Bayesian approaches (e.g. English et al., 1999; Uddstrom et al., 1999; Merchant et al., 2005; Mackie et al., 2010a; Heidinger et al., 2012). A classification scheme for Bayesian cloud masks which helps to clearly distinguish the various approaches to Bayesian cloud masking is introduced in Sect. 3 and in addition, a short overview of the relevant literature using that scheme is given.

The here presented results are computational highly efficient and are very well suited for the processing of large amounts of data, which makes these results very well suited for future application to the Ocean Land Colour Instrument (OLCI) (Nieke, 2008) and

AMTD
7, 11045–11085, 2014

the Sea and Land Surface Temperature Radiometer (SLSTR) (Coppo et al., 2010) onboard the Sentinel-3 satellite (Miguel et al., 2007) and its operational follow-ups.

2 Bayesian inference for cloud masking

Bayes' theorem can be used to reverse joint probabilities. It is appealing to apply it to cloud masking since its theory is widely adopted, its implementation on a computer system is straightforward, and its results are probabilities which can be directly interpreted. The theorem allows the computation of the probability $P(C,\boldsymbol{F})$ that a particular measurement with feature $\boldsymbol{F}$ is affected by a cloud, if the occurrence probabilities $P(\boldsymbol{F},C)$ and $P(\boldsymbol{F},\overline{C})$ of the feature under cloudy and non cloudy conditions are known. Here, $P(a,b)$ denotes the occurrence probability of a under the condition of the occurrence of b.

With C being the case that a measurement is affected by clouds and $\boldsymbol{F}$ being a set of features associated with that measurement, $P(C,\boldsymbol{F})$ can be expressed as:

$$P(C,\boldsymbol{F}) = \frac{P(C)P(\boldsymbol{F},C)}{P(\boldsymbol{F})} = \frac{P(C)P(\boldsymbol{F},C)}{P(C)P(\boldsymbol{F},C)+P(\overline{C})P(\boldsymbol{F},\overline{C})}, \quad (1)$$

where $P(C)$ is the background probability of cloudiness and $\overline{C}$ is the negation of C, which states that a measurement is not affected by clouds. The occurrence probability of the feature $P(\boldsymbol{F})$ can be expressed in terms of the joint probabilities $P(\boldsymbol{F},C)$ and $P(\boldsymbol{F},\overline{C})$ since cloudyness and non-cloudiness are the only two considered classes for each measurement.

Evaluating Bayes' theorem involves only few arithmetic operations, such that a specific implementation can be very fast and efficient, which is of importance if large amounts of data are to be processed. Additional computations involve the feature $\boldsymbol{F}$ and the a priori joint probabilities $P(\boldsymbol{F},C)$ and $P(\boldsymbol{F},\overline{C})$, which is discussed in the following sections.

AMTD
7, 11045–11085, 2014

With an appropriate set of thresholds, one can convert the probability $P(C,\boldsymbol{F})$ into a cloud mask. For instance, any probability strictly higher than 50 % could be interpreted as cloud, but other thresholds or more classification bins could be used. This is discussed in more detail in Sect. 7.1, but this choice clearly depends on the target application and is independent from the Bayesian approach. Other applications, such as the construction of a cost function as described by English et al. (1999), are also viable alternatives.

Estimating the value of the background probability $P(C)$ is not discussed in detail in this paper and for all following applications a value of 0.5 is used. This choice basically states, that for each measurement an equal probability of it being cloudy or not cloudy is assumed. This assumption is of course neither valid on global or local scales and a rich body of knowledge about the spatial and temporal distribution of cloud occurrence probabilities exists. Such knowledge, typical in the form of external climatologies, could be used to estimate $P(C)$, but would eventually shift derived climatologies towards the external one, which would then effectively lead to circular arguments. This point might be of lower importance for some applications, i.e. operational processing by weather services, but within Cloud CCI it is planed to derive climatological datasets from the full MERIS and AATSR time series and circular arguments are best to be avoided. For the special case of $P(C) = 0.5$, Eq. (1) simplifies to:

$$P(C,\boldsymbol{F}) = \frac{P(\boldsymbol{F},C)}{P(\boldsymbol{F},C) + P(\boldsymbol{F},\overline{C})}. \tag{2}$$

Setting up a particular Bayesian cloud mask algorithm involves several decisions, such as specifying the measurement feature $\boldsymbol{F}$ and choosing a technique to estimate $P(\boldsymbol{F},C)$ and $P(\boldsymbol{F},\overline{C})$, which allows to group the various possible approaches to Bayesian cloud masking into distinct subgroups. This natural grouping allows to clearly separate the presented approach from other algorithms and is discussed in Sect. 3. In addition, a short overview about the relevant literature is given.

AMTD
7, 11045–11085, 2014

3 Classification of Bayesian cloud masks

Several papers on Bayesian approaches to cloud masking have been published in the past and fundamental differences between the various algorithms are often buried in the technical details of the particular paper. A nomenclature which aims to clearly separate different approaches to Bayesian cloud masking is discussed in the following.

Let the feature $\boldsymbol{F}$ from Eq. (1) be a set of n_F real numbers $\boldsymbol{F} = (F_1, \ldots, F_{n_F}) \in \mathbb{R}^{n_F}$, where the F_i are typically determined from measurements $\boldsymbol{M} \in \mathbb{R}^{n_M}$, auxiliary data $\boldsymbol{A} \in \mathbb{R}^{n_A}$, and external data $\boldsymbol{E} \in \mathbb{R}^{n_E}$. The components F_i are computed from prescribed feature functions f_i, which generally depend on all of the above introduced classes of data: $F_i = f_i(\boldsymbol{M}, \boldsymbol{A}, \boldsymbol{E})$. In the case of the MERIS and AATSR Synergy, the set of measurements $\boldsymbol{M}$ includes radiances and brightness temperatures for a single collocated pixel. Auxiliary $\boldsymbol{A}$ data is available with negligible computational cost, such as time stamps, geolocation, and data flags. External data might be a function of the available measurements $\boldsymbol{M}$ and auxiliary data $\boldsymbol{A}$ and its retrieval is by definition associated with non negligible computational cost. This category essentially introduces significant external knowledge about the measurement and common examples are online radiative transfer (RT) simulations, nontrivial interpolation in numerical weather prediction (NWP) data, or the use of climatologies.

Let us call the feature set $\boldsymbol{F}$ independent, if it is only a function of the measurements $\boldsymbol{M}$ and auxiliary data $\boldsymbol{A}$ and dependent if in addition external data $\boldsymbol{E}$ is exploited. Both classes can be further subdivided with respect to weak and strong dependence to describe $\boldsymbol{F}$ even more precise. A weakly dependent feature set could for example depend on interpolation in NWP data which is of negligible computational cost, while a strongly dependent feature set could depend on online RT with non negligible numerical cost. Strongly independent feature sets would then depend only on measurements $\boldsymbol{M}$, while weakly independent feature sets could in addition depend on auxiliary data $\boldsymbol{A}$.

This paper is focused on Bayesian cloud masks based on strongly independent features. Only MERIS and AATSR measurements and trivial functions operating on them

AMTD
7, 11045–11085, 2014

are used to construct the feature set. This class of features allows to implement a numerically highly efficient algorithm with simple opportunities to parallelization and vectorization. With no dependency on external data, the algorithm can be used in non operational environments, where the acquisition of NWP data can pose a significant effort. However, there is no obvious reason why the techniques which are discussed in the following section should be limited to the independent case.

The second major branch in Bayesian cloud masking schemes involves the computation of the joint probabilities $P(\boldsymbol{F}, C)$ and $P(\boldsymbol{F}, \overline{C})$. The classical approach aims at the direct computation of these two joint probabilities, while the naive approach treats the components F_i of the feature set $\boldsymbol{F}$ as statistically independent and decouples the joint probabilities on $\boldsymbol{F}$ into a product of joint probabilities in the F_i:

$$P(\boldsymbol{F}, C) = \prod_i P(F_i, C). \tag{3}$$

One can either construct the feature set very carefully, such that this strong assumption holds (e.g. follow Merchant et al. (2005) and their discussion on cloud texture and cloud top temperature), or simply accept its violation and the possible effects on the cloud masking scheme. Formally proving the statement of Eq. (3) seems to be only possible for a rather limited class of features.

Computing the joint probabilities in the classical approach can be greatly simplified by assuming an analytic form and estimating its parameters. Depending on the assumed form, for instance multivariate Gaussian (e.g. see Uddstrom et al., 1999; Merchant et al., 2005), the resulting cloud mask could be called classical Gaussian. As for the naive approach, it will be difficult to formally prove the validity of such assumptions.

The classical and naive approaches can be mixed if one or more subsets of the F_i are treated as statistically independent, such that the decoupling of $P(\boldsymbol{F}, C)$ and $P(\boldsymbol{F}, \overline{C})$ becomes partial. For this class of Bayesian cloud masks we propose to use the terms mostly naive if the majority of features are decoupled and mostly classical if the majority of features are not decoupled.

AMTD
7, 11045–11085, 2014

This paper is mainly concerned with the discussion of the classical and naive approach with an emphasis on the classical one. In conclusion, this paper is mostly concerned with the application of classical Bayesian cloud masks based on strongly independent features. As it will be shown later in the paper, the classical approach gives better results for the cloud masking in our scenario and the strongly independent feature set was chosen to allow the implementation of a very fast algorithm.

Cloud detection methods based on Bayesian probabilities have been used for cloud masking in the past and a short overview, but without the attempt to fully outline them, is given now. English et al. (1999) used Bayesian probability with strongly dependent features to derive a cost function for a 1-D var retrieval of cloudiness. The classification is based on the exploitation of microwave and infrared channels and in addition external data from NWP simulations. Uddstrom et al. (1999) used AVHRR channels, derived channels such as reflectance ratios and brightness temperature differences, and textural measures to construct strongly independent features. It was found, that textural measures are of most importance for night time measurements. The joint probabilities were separated by assuming a multivariate Gaussian form and were expressed in terms of mean values and associated covariances. Merchant et al. (2005) used night time thermal infrared measurements at 3.7, 11, and 12 µm to construct a mostly classical Bayesian cloud mask. Textural features were assumed to be independent from measurements in thermal channels and were separated when computing the joint probabilities. $P(C)$ was estimated from NWP data and the algorithm was discussed with an emphasis on operational NWP centers, such that these feature are likely only weakly dependent. Mackie et al. (2010a, b) discussed a mostly classical Bayesian algorithm with strongly dependent features for the 3.9, 11, and 12 µm channel of the SEVIRI instrument. External knowledge is introduced by NWP data and a fast radiative transfer model and textural features were separated from spectral features assuming independence. Heidinger et al. (2012) discussed a naive Bayesian cloud mask with strongly dependent features for the AVHRR instrument. A surface type classification using external MODIS data was used to maximize the detection rate. CALIPSO lidar

AMTD
7, 11045–11085, 2014

measurements were used as truth data to compute histograms from which the occurrence probabilities for each feature were estimated.

4 Construction of feature sets

Channels of the MERIS and AATSR instruments cover the spectral range from 412 nm to 12 µm and are referenced in this paper by their central wavelength, while for MERIS the unit of nm and for AATSR the unit µm is used. Figures 1 and 2 show examples of possible features for two particular interesting scenes at Greenland and the vicinity of the Korean peninsula. Each Figure shows an RGB image, various single channels, and a selection of trivial functions which combine two channels. Both figures include a panel with results of the non Bayesian Synergy cloud mask, which is briefly discussed in Sect. 6. Figure 1 shows a scene over Greenland with its center located at 59°31′12″ W and 79°0′0″ N with high and low clouds over a large ice or snow covered region. Figure 2 shows a scene in the vicinity of the Korean peninsula with its center located at 125°52′12″ E and 37°45′36″ N. This scene shows a pronounced dust storm mixed with a deck of clouds.

Strongly independent features are constructed using a single channel or any combination of channels in a trivial function. Such combinations have been called derived channels in the literature (e.g. Uddstrom et al., 1999). Considered here are all basic arithmetic operations $+, -, \times, /$ and in addition the index function $\mathrm{d}x(a,b) = (a-b)/(a+b)$, which can be used to create indices such as the normalized difference vegetation index (NDVI) (see Kriegler et al., 1969), the normalized difference snow index (NDSI) (see Hall et al., 2002), or other general channel indices. Even if well known and generally accepted combinations of channels and indices are used, it is unclear whether a specific combination is the best possible candidate for the particular data set and one has to rely on the experience of the involved experts.

In contrast to such approaches, an objective measure for any given set of feature functions is exploited to numerically search for the best possible set of feature functions.

AMTD
7, 11045–11085, 2014

Maximizing the Hanssen Kuipers skill score (see Hanssen and Kuipers, 1965; Woodcock, 1976) with respect to a given validation dataset is an appropriate metric for this problem. It is also sometimes referred to as Hansen Kuipers discriminant and is essentially the difference of the hit rate and the false alarm rate of the cloud mask with respect to a validation data source. It covers the range of −1 to +1, with +1 being a perfect representation of the validation source. From now on, only the term skill score is used.

Validation of cloud masks for MERIS and AATSR onboard ENVISAT is a difficult task since no generally accepted and available set of truth data exists. A generally used approach is to generate truth data by means of manual classification of images by human experts or the use of data from ground based stations. Converting a ground truth to a pixel by pixel truth can be complicated and possibly insufficient spatial coverage can limit the applicability of that approach. Consequently, most approaches for generating truth data for MERIS and AATSR are one way or the other based on the manual classification of sample data by human experts (e.g. Gómez-Chova et al., 2006, 2008; Schlundt et al., 2011). Such datasets can be called artificial truth since they are used as if they were truth's, but it is as least arguable whether such data sets are in fact truth's.

To demonstrate the feasibility of the Bayesian approach, results from the Synergy cloud mask (see Sect. 6 for a brief description) were chosen as source of artificial truth data and it is therefore assessed whether Bayesian cloud masks can reproduce this Synergy cloud mask. The major advantage of this approach is that large amounts of artificial truth data can be created without significant effort. Clearly, all shortcomings of this seeding algorithm will be present in this dataset and will limit the success of the application of the Bayesian technique.

Optimizing the choice for a particular set of feature functions is not straightforward, since this problem is noncontinuous with a varying number of free parameters. First, the number of feature functions has to be set. Then, for each feature, a feature function from the pool of considered functions has to be selected. The identity function, all four

AMTD
7, 11045–11085, 2014

basic arithmetic operations, and the index function are considered as feature functions. As a last step, the input channels for each feature function must be set. Depending on the chosen functions and channels, a maximum of $2 \times n_F$ channels can be included in the computation of a feature set with n_F elements.

Then, for a particular feature set, the prerequisites for computing the joint probabilities must be carried out, which is described in detail in Sect. 5. Once this step is completed, the Hanssen–Kuipers skill score for the selected set of validation data can be computed.

The only numeric optimization procedure that we are aware of, which is generally applicable to this situation, is a random search in the huge search space spanned by this outlined procedure. This is quite a different approach to that of a human expert, who would likely start an educated search, but might not attempt to cover the whole search space. The number of possible combinations depends on the number of chosen features and the number of available channels (22 in the case of the MERIS and AATSR Synergy) and can be estimated using the binomial coefficient. In the simplest case, where merely the identity function is used, no channel is used more than once, and four features are to be selected, the search space spans $\binom{22}{4} = 7315$ elements. If only functions of two channels are to be selected and re-selected and channels can be used multiple times, then the search space consists of $\binom{5\times(22^2-22)}{4} = \binom{2310}{4} \approx 1.2\times10^{12}$ entries. The enormous size of the search space makes it difficult to completely cover it by a search, but the random search can be allowed to run appropriately long such that a result of sufficient quality is obtained. One can expect that a large number of different sets of feature functions will essentially exhibit very similar classification skills. The considered feature function are not symmetric under a change of the parameter order, but the overall classification result might be approximately symmetric. This alone would decrease the search space by a factor of approximately 16. In addition, the classification results might be only weakly dependent with respect to the feature function itself, i.e. the index function $\mathrm{d}x(a,b)$ might be as effective as a ratio a/b, which would decrease the effective size of the search space.

AMTD
7, 11045–11085, 2014

The proposed random search might not be able to cover the complete search space, but with sufficient long runtime one will be able to find solutions with sufficiently high skill score. In addition, unusual combinations of channels might be found which wouldn't be considered in an educated search by a human expert. The features shown in Figs. 1 and 2 are frequently found in searches when results from the non Bayesian Synergy cloud mask are used as artificial truth.

The physical meaning of a certain feature set and why it might be better or worse than a different one is not discussed here and is also not within the scope of this paper. This knowledge is very useful for educated searches, but is not necessarily needed in this setup. However, for the experienced expert it is of course no surprise which channels are found to be successful by the optimization scheme.

Implementing such a search strategy is straightforward. A generator of random feature functions must be implemented and each of these instances can be tested for its skill score with respect to the artificial truth. This procedure is easy parallelizeable and one could store only results with higher skill score than some predefined value. At any given time of an ongoing search, one can sort these results and evaluate the top results.

5 Estimation of background joint probabilities

The background joint probabilities $P(\boldsymbol{F},C)$ and $P(\boldsymbol{F},\overline{C})$ could be computed in various ways, but here only the frequentistic approach based on sample data is considered. A sufficiently large number of already classified measurements is converted into its corresponding set of features and probability density histograms are produced from which the probabilities are estimated. In the naive Bayesian approach, as many one dimensional histograms as there are features are needed, while in the classical Bayesian approach a single n_F dimensional histogram is used. If these histograms are stored in a computer system, the handling of any reasonable number of one dimensional histograms poses no specific problem, while an array of dimension n_F grows rapidly in

AMTD
7, 11045–11085, 2014

memory with increasing number of bins n_B. For the sake of simplicity, the same number of bins is assumed for each particular dimension. With four bits per float and twenty bins per feature, one would need 0.6 GB to store a single histogram for four features, but already about 4883 GB for seven features. This limits the practical number of features for the classical Bayesian approach to about four to six at the time of writing this paper.

However, the main argument of Uddstrom et al. (1999) and Heidinger et al. (2012) against the use of the classical approach is, that one has generally not enough truth data available to robustly derive the histograms in a completely frequentistic way. This can be a valid point for real truth data which, is limited in principle, but not so much for artificial truth data. Here, the number of available data is merely a function of the available human labor for manual classification or computational resources if an existing cloud masking scheme is used to produce artificial truth data.

Both left panels of Figs. 3 and 4 show results of two dimensional histograms for MERIS and AATSR Synergy data. For both cases, almost one million spectra were used to compute both histograms. Shown is the difference of the histograms for C and $\overline{C}$. Both choices of features recreate the Synergy cloud mask reasonably well with a skill score of about 0.76. The cloud masking setup is discussed in detail in Sect. 7. The main point here is that with enough data points these histograms can be computed. The two dimensional case was chosen since this is simple to visualize.

Both right panels of Figs. 3 and 4 show remarkably similar histograms with just barely smaller skill score values of about 0.75, but only 1000 measurements were used to produce these histograms. A simple Gaussian smoothing filter was applied to both histograms and the Gaussian smoothing factor was chosen such that the skill score as a function of the Gaussian smoothing was maximized. This is the first main result of this paper. This numerical experiment shows, that at least for some sets of feature functions, the n_F dimensional histograms can be approximated by using very few data points and an appropriate Gaussian smoothing factor. The best smoothing factor for both cases was slightly different and is obtained from optimization. More detailed

AMTD
7, 11045–11085, 2014

results are shown in Sect. 7.1. In addition to the previously discussed parameters, e.g. the construction of features, classical Bayesian cloud masks are defined by the number of bins used in the histograms and the chosen Gaussian smoothing parameter, which is discussed in Sect. 7.1.

The Gaussian smoothing approach works reasonably well and is so far only justified by its actual success for a particular problem, where in fact sufficient amounts of artificial truth data is available. Its general application to situations with limited amounts of such data is therefore not very well justified. However, numerical experiments with the available data have shown that this approach yields remarkably good results. Other functional kernels have not been tested, but the Gaussian approach seems sufficient since the convoluted histograms yield nearly the same skill score as the original histograms. Success of this approach is likely based on the fact that the smoothing procedure distributes data to neighbour bins, but does not strongly change the defining spectral features of the measurements. In that, it implicitly creates data which could represent different viewing geometries or situations with slightly varying optical parameters. Hence, this approach is not justified by first principles, but rather with working examples which strengthen our expectations that this approach will work reasonably well for any other set of features.

6 Synergy cloud mask

The Synergy cloud mask is discussed in detail by Gómez-Chova et al. (2008) and is implemented as an external processor for the BEAM toolbox (Fomferra and Brockmann, 2005). It is based on radiative transfer simulations covering all spectral bands of MERIS and AATSR and statistical analysis of classified data by human experts. It has been applied to process the years 2007 to 2009 of the MERIS and AATSR time series, such that a large number of artificial truth data is available.

AMTD
7, 11045–11085, 2014

7 Application to MERIS, AATSR, and their synergistic product

If the computation of $P(\boldsymbol{F},C)$ and $P(\boldsymbol{F},\overline{C})$ is based on the frequentistic approach and artificial truth data, then three major applications of the technique become feasible. Results from existing algorithms can be reproduced using the Bayesian technique, which could potentially speed up and simplify the cloud masking of large amounts of data. With the Synergy cloud mask from Sect. 6 as example, this procedure is discussed in the following Sect. 7.1. If the existing algorithm is reproduced reasonably well, one can use this technique to further enhance the algorithm, which is discussed in Sect. 7.2. A simple example where data classified by a human expert is used to set up a Bayesian cloud mask is discussed in Sect. 7.3.

7.1 Reproduction of existing algorithms

A Bayesian cloud mask can be used to approximate independent algorithms, but with the advantage of possibly drastically decreased computation times. However, it is not obvious that a particular algorithm is reproducible to a sufficient extent with this technique. Artificial truth data from the Synergy cloud mask, which was shortly discussed in Sect. 4, is used as a test case and a large number of Bayesian cloud masks with different feature sets were created and ranked according to their skill score. The joint probability's were estimated using globally equally distributed data from the year 2007 and similarly distributed data from the year 2008 was used to compute the skill score, which is used to asses the ability of the cloud mask to reproduce the Synergy cloud mask. The presented results don't have to represent a global optimum since only a small fraction of the search space was covered in the finite search time. Depending on the number of features and the classical or naive Bayesian approach, a certain upper bound of skill scores for any test case was not exceeded, but many feature sets with similar skill score to that soft limit were found.

Figures 5 and 6 show the global distribution of skill scores for two classical Bayesian cloud masks based on sets of two and four features. The increase to four features

AMTD
7, 11045–11085, 2014

improves the results, although not dramatically for the mean global skill score. The two feature sets were the best candidates within the allowed search time for the full Synergy set of channels. The results are best for ocean areas and worst for areas with mountains (Nepal, west coast of northern America), deserts (Sahara, Arabian peninsula), and ice and snow covered areas (poles, Siberia). These are actually the areas where one naturally would expect major difficulties in detecting clouds. The local skill score in these areas were significantly improved by increasing the number of used features to four.

The data that was used to produce Figs. 5 and 6 was sorted and used to generate the overview shown in Fig. 7. Shown is the computed cloud probability from the two Bayesian cloud masks, but separated for the cloudy and non-cloudy group as classified by the Synergy cloud mask. The threshold of 0.5 cloud probability is also shown and was used as separation between the cloudy and non-cloudy class. This representation shows the cause of non unity skill score. Here, the misses (number of red points before crossing the blue line vs. those beyond) and the false alarms (number of green points after crossing the blue line vs. number of points before crossing) are quite similar. Figure 7 shows that the Bayesian cloud mask with four features exhibits a much smother distribution of probabilities and an decreased rate of misses, while the improvement of the false alarm rate is only minor. Also, the impact of changing the threshold value can be nicely seen. The overall skill score seems to be almost unaffected when changing the threshold. The false alarm rate decreases when the threshold in increased, but as the same time the rate of misses increases which would decrease the skill score.

Similar results can be achieved by using different combinations of feature functions and channels. An overview about results for the Synergy data, MERIS, and AATSR alone is given in Tables 1–3. Tables 1 and 2 show results for a classical Bayesian cloud masks with strongly independent feature sets for two and respectively four features. Table 3 shows results for naive Bayesian cloud masks with five strongly independent features. Classical Bayesian cloud masks based on two strongly independent features show best results if the complete Synergy channel set or MERIS alone is used. The

AMTD
7, 11045–11085, 2014

results for AATSR alone are significantly inferior. For the Synergy data set, the 11 µm channel in combination with a MERIS channel in the blue (412 and 442 nm) is found in all three top results. For MERIS alone, a combination of a channel in the blue and an index of red and short wave infrared channels is found in the top results. It is quite counter intuitive that the best results for MERIS are achieved with only three channels, while the algorithm had the freedom to select up to four channels. The best result for the set of Synergy channels included four channels, which relates more to the naive intuition that more channels carry more information and would therefore be better suited for the application. However, since the search space was not fully covered, a better solution for MERIS with four channels could still be found.

Table 2 shows similar results, but for classical Bayesian cloud masks based on a set of four features. Again, Synergy and MERIS results are significantly better than those for AATSR, while the Synergy results are only slightly better then those for MERIS alone. All possible feature functions are used within the results, but of course not all the time for any result.

Similar studies were also performed for higher numbers of features, but no results with significantly higher skill scores were found. The skill score results for using three features are positioned right in the middle of the two discussed results, such that four features seems to be the best choice to reproduce the Synergy cloud mask with a classical Bayesian cloud mask based on strongly independent features.

Similar searches for naive Bayesian cloud masks with strongly independent features were performed for five and 15 features and results for five features are shown in Table 3. The search with 15 features did not show significantly better result than the ones shown. In general, these results are not as successful in reproducing the Synergy cloud mask as the approaches with the classical Bayesian cloud mask. Skill scores for AATSR alone are smaller than for MERIS and Synergy and also generally smaller than for the classical approach with four features.

Concluding this aspect, it is possible to find feature sets that reproduce the Synergy cloud mask reasonably well even without covering the complete search space. For

AMTD
7, 11045–11085, 2014

a soft upper limit of the skill score, different feature sets with similar skill score can be found. This is actually not surprising and represents the fact that the same classification results in terms of skill score can be achieved with many different feature sets. From a technical point, it is then sufficient to choose one of those results with best skill scores, even if this might not be the absolute global maximum.

Next, the impact of the number of bins n_B, Gaussian smoothing value, and sample size of the artificial truth data set is discussed. The sensitivity of the Bayesian cloud mask in terms of skill score with respect to a certain feature set is shown in Figs. 8 and 9. Both Figures show skill scores for Synergy cloud mask artificial truth data with respect to number of bins, Gaussian smoothing factor, and sample size of the artificial truth data. Figure 8 shows an extreme case where only 100 randomly selected globally distributed cases were used as artificial truth. Again, the year 2007 was used as pool for the artificial truth and the year 2008 to compute the skill score. The skill scores of the cloud mask which is based on such a small sample size clearly depends on the sample itself. The procedure was repeated ten times and the achieved mean skill score is shown. The SD in the last digit is shown in parenthesis.

With no Gaussian smoothing applied, the skill score clearly decreases with increasing number of bins since the sample size is much to small for this resolution. Also, the impact of the sample is largest since the SD is highest. The skill score increases with increasing number of bins and Gaussian smoothing until a maximum is reached. With then increasing bin number and smoothing, the skill score decreases, but only slightly. In this case, an optimal set of bin size and smoothing can be found. If smaller vales are used, the skill scores are drastically reduced, but if larger values are used, the skill score decreases only slightly.

A similar sensitivity study is shown in Fig. 9, but here a much larger sample size of artificial truth data was used. Again, without Gaussian smoothing, the smallest number of bins shows the best results while with increasing number of bins the skill score decreased since the total number of bins grows with the fourth potential of the number

AMTD
7, 11045–11085, 2014

of bins. A large plateau of consistently stable and high skill score values is found for numbers of bins above 25 and Gaussian smoothing of above 0.9.

In both cases, for small and very large sample sizes of artificial truth data, the skill score decreases with increasing Gaussian smoothing for small numbers of bins. This clearly shows that too strong Gaussian smoothing can destroy information in an accurately estimated histogram, but distributes information in incomplete histograms such that it better represents the true probability density.

In general, one can not perform such studies to assess the optimal number of bins and value of Gaussian smoothing parameter, since only an insufficient number of artificial truth data might be available. The presented results from numerical experiments indicate that for four features and a sufficiently large sample of artificial truth data, a bin size of 40 with a Gaussian smoothing of 1.5 is a good choice. This result holds not only for the presented feature set, but also for many other sets which have been assessed during this research.

7.2 Enhancements of existing algorithms

It was shown so far that Bayesian cloud masks can be used to reproduce at least one existing cloud mask up to a certain extent. It is unclear however, what the limiting factors are in global skill score with respect to this particular cloud mask. A major contributor to this upper limit can be inconsistencies in the artificial truth data set. Examples are shown in panel a and b of Figs. 10 and 11, which actually show the surroundings of the scenes shown in Figs. 1 and 2. Both figures show some classification errors of the Synergy cloud mask. The top part of Fig. 10 shows a partly cloudy scene over a large ice or snow covered area, which is completely masked as cloudy (white areas in panel b). In addition, the arrow shaped land area in the lower part of the Figure (Brodeur peninsula on Baffin island) is clearly not cloudy, but is classified as cloudy. Similarly in Fig. 11, the complete dust storm east of the Korean peninsula is marked as cloudy. Such classification errors introduce inconsistencies which affect the produced histograms and are in general difficult to reproduce with an independent system.

AMTD
7, 11045–11085, 2014

The appearance of such errors does not mean that the algorithm should be abandoned and with it all the work that has been invested into developing it. Panels c and d in Figs. 10 and 11 show how the Bayesian cloud mask technique can be used to enhance this existing algorithm, when errors in the artificial truth data are manually corrected by an human expert. Synergy cloud mask results from these two orbits were corrected by hand and used as artificial truth to produce a classical Bayesian cloud mask based on four strongly independent features. The two orbits were then reprocessed and the resulting cloud masks and cloud probabilities are shown. Some artifacts at land and ice boundaries are still present, but the major classification errors were strongly reduced.

This result is merely shown as proof of concept for the enhancement of existing algorithms. The shown case was limited to only two scenes which were manually corrected and used as artificial truth for the Bayesian cloud mask, which is therefore only strictly applicable to these two scenes. In a realistic approach, one would need some knowledge on where the existing algorithm performs below the requirements. This poses no real limitation and will always be the case, otherwise one would have no incentive to improve the existing algorithm. These cases, e.g. limited to certain areas, known weather conditions, or certain periods of time could be excluded from the artificial truth data set, while other correctly classified results are still included. These introduced data gaps, or better representativity gaps, can then be filled with artificial truth data from manual classification. Such an approach can be used to focus the attention of the human experts to areas where their expertise is most strongly needed and to use their available labor in the most efficient way.

As it was discussed in Sect. 7.1, possibly many different feature sets can be used to recreate the algorithms which was used to produce the artificial truth data. This property can be used to produce much more robust cloud masking algorithms. If the seeding algorithm can not cope with missing data, if e.g. a certain needed channel is flagged as unusable or saturated, one can simply switch to a different Bayesian cloud mask which does not depend on that channel. The operational version of the cloud mask for the Cloud CCI project contains several ranked Bayesian cloud masks and if

AMTD
7, 11045–11085, 2014

the top mask fails to produce a result, a mask of lower rank is used until the last mask was used or the algorithm produced a result. This approach can greatly reduce the number of unprocessed measurements for a cloud masking scheme.

7.3 Cloud masks from manually classified data

Human experts can produce artificial truth data of high quality by careful manual classification of MERIS, AATSR, or Synergy images. It is of great advantage that the spatial resolution of MERIS and AATSR images is high enough such that spatial and spectral patterns together can be used to classify data points. Cloud shadows for instance, can be used to clearly distinguish clouds from snow and ice surfaces. In that respect, the algorithm itself is not based on spatial information, but it was surely used to create the artificial truth data. It is beyond the scope of this paper to produce a cloud mask with global applicability, but it should be demonstrated how straightforward such a procedure would be. The here presented results are then clearly applicable to OLCI and SLSTR onboard the upcoming Sentinel-3 satellite.

The same two orbits which were discussed in Sects. 4 and 7.2 are used for the procedure. Both orbits contain scenes which are in general difficult to classify accurately, such as clouds over a snow and ice covered region, cloud free snow and ice covered surfaces, or a pronounced dust storm. The manual classification setup was designed, such that no special computational knowledge is needed to perform the cloud classification. For each test orbit, image files containing several layers were created. The various image layers include an RGB image, contrast stretched gray-scale images from Synergy channels, and several feature functions which were found to be of good performance in the Bayesian framework. To classify actual pixels, the human expert has to color areas (e.g. blue color for cloud free and red color for cloudy) in a blank image layer. By adjusting the transparency of the single layers, each scene can be carefully inspected before a decision is made. The actual shape of the colored areas is of lower importance as well as the actual amount of classified areas. However, the total variability of possible cases and scenes should be included in the classification.

AMTD
7, 11045–11085, 2014

Results of this procedure are shown in Fig. 12. The leftmost two panels show an RGB view of the scene and with blue and red color the areas which were classified by the human expert. The actual amount of the classified area is small compared to the total size of the scene. Then, this dataset was used as artificial truth and a classical Bayesian cloud mask with four strongly independent features was set up to to process the two orbits. The resulting cloud masks are shown in the middle two panels, while the actual cloud probability is shown in the two rightmost panels.

The Bayesian cloud mask is clearly able to separate the clouds from the snow and ice underground, does not misclassify the land area (see Sect. 7.2), and is able to mostly separate clouds from the dust storm. Most importantly, the human expert does not need to be an expert on how to implement this mask, or how to design hierarchies of thresholds, but simply translates classification decisions into cloud mask results. These images can be stored for future enhancements of the artificial truth data set and as self describing documentation of the algorithm.

This approach is most straightforward when the spatial resolution of the instrument in question is high enough, such that the human expert can use the spatial pattern information to correctly classify cloudy from non-cloudy areas. For global applicability, a higher number of orbits with representative spatial and seasonal sampling should be included in the set of considered artificial truth data. Especially, complex cases such as scenes with ice, snow, sun glint, mountains, or dust storms should be included in the classification effort.

8 Conclusions

The application of the classical and naive Bayesian cloud masking technique to MERIS, AATSR, and their Synergy was discussed in detail. Bayesian cloud masks based on independent features are numerically highly efficient and are very well suited for the fast processing of large amounts of data. Within ESA's Cloud CCI project, it is intended

AMTD
7, 11045–11085, 2014

to apply this technique to a reprocessing of the 9.5 year time series of MERIS and AATSR measurements.

Sufficient amounts of artificial truth data and the frequentistic approach can be used to estimate multidimensional histograms for the estimation of background joint probabilities. Gaussian smoothing of appropriate width can be used to drastically reduce the actual amounts of truth data needed to compute histograms for the classical Bayesian approach. This post processing step greatly simplifies our ability to further explore the classical Bayesian approach.

Due to restrictions of modern computer hardware, the practical limit for the classical Bayesian approach is reached with six to seven features. This does not actually restrict its applicability, since trivial feature functions can be used which combine any number of measurements into a single feature.

It was found that classical Bayesian cloud masks with four strongly independent features are the best choice for the cloud masking of MERIS, AATSR, and their Synergy measurements when the Synergy cloud mask is used as benchmark. The classical approach gave significantly better results then the naive approach. MERIS and the MERIS-AATSR Synergy give very similar results in terms of cloud classification, while AATSR alone shows significantly smaller skill scores. The MERIS Oxygen-A absorption channel was found to be present in the best results when the set of selected feature functions and channels was numerically optimized.

The broad spectral range and the number of available channels within the Synergy data set can be used to set up Bayesian cloud masks with very similar classification skill, but based on different combinations of channels. This can be used to design cloud masking schemes which are robust against partially missing data.

It was shown how Bayesian cloud masks can be used to reproduce the results of existing algorithms, improve existing algorithms, and how to set up new classification schemes based on manual classification by human experts. Reproducing existing algorithms offers the perspective of increased numerical efficiency and processing robustness. The approach based on manual image classification is straightforward for

AMTD
7, 11045–11085, 2014

the human expert. Classified scenes can be stored and revisited if the produced cloud masks show misclassifications in certain areas or weather conditions. If errors are not traceable to errors in the manual classification, then additional scenes can be added to the set of artificial truth data to increase the chance of correct classification.

The presented results for MERIS and AATSR can be used to implement an accurate and highly efficient cloud masking scheme for OLCI and SLSTR onboard the upcoming Sentinel 3 satellite. Especially, the additional oxygen absorption channels from the OLCI instrument might be used within an improved and numerically efficient cloud classification algorithm.

Although this paper is focused on strongly independent Bayesian cloud masks, there is no apparent reason which prevents the application of the introduced techniques to the case of dependent Bayesian cloud masks. It is straightforward to include external information such as clear sky radiance estimators or NWP fields in the proposed optimization strategy for the construction of features. The application of Gaussian smoothing to derived histogram fields is then independent from external information and can be used to reduce the amounts of needed truth data.

Acknowledgements. This work has been funded by the European Space Agency (ESA) in the framework of the Climate Change Initiative (CCI) project and the German Federal Ministry of Education and Research (BMBF) in the framework of the HD(CP)^2 project.

References

Carbajal Henken, C. K., Lindstrot, R., Preusker, R., and Fischer, J.: FAME-C: cloud property retrieval using synergistic AATSR and MERIS observations, Atmos. Meas. Tech. Discuss., 7, 4909–4947, doi:10.5194/amtd-7-4909-2014, 2014. 11047

Coppo, P., Ricciarelli, B., Brandani, F., Delderfield, J., Ferlet, M., Mutlow, C., Munro, G., Nightingale, T., Smith, D., Bianchi, S., Nicol, P., Kirschstein, S., Hennig, T., Engel, W., Frerick, J., and Nieke, J.: SLSTR: a high accuracy dual scan temperature radiometer for sea and land surface monitoring from space, J. Mod. Optic., 57, 1815–1830, doi:10.1080/09500340.2010.503010, 2010. 11048

AMTD
7, 11045–11085, 2014

English, S., Eyre, J., and Smith, J.: A cloud-detection scheme for use with satellite sounding radiances in the context of data assimilation for numerical weather prediction, Q. J. Roy. Meteor. Soc., 125, 2359–2378, 1999. 11047, 11049, 11052

Fomferra, N. and Brockmann, C.: Beam-the ENVISAT MERIS and AATSR toolbox, in: MERIS (A)ATSR Workshop 2005, vol. 597, 13 pp., 2005. 11058

Gómez-Chova, L., Camps-Valls, G., Amorós-López, J., Guanter, L., Alonso, L., Calpe, J., and Moreno, J.: New cloud detection algorithm for multispectral and hyperspectral images: application to ENVISAT/MERIS and PROBA/CHRIS sensors, in: IEEE International Geoscience and Remote Sensing Symposium, IGARSS, 2757–2760, 2006. 11054

Gómez-Chova, L., Camps-Valls, G., Munoz-Marı, J., Calpe, J., and Moreno, J.: Cloud screening methodology for MERIS/AATSR Synergy products, in: Proc. 2nd MERIS/AATSR User Workshop, ESRIN, Frascati, 22–26, 2008. 11047, 11054, 11058

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J.: MODIS snow-cover products, Remote Sens. Environ., 83, 181–194, 2002. 11053

Hanssen, A. and Kuipers, W.: On the Relationship Between the Frequency of Rain and Various Meteorological Parameters: with Reference to the Problem of Objective Forecasting, Staatsdrukkerij-en Uitgeverijbedrijf, 1965. 11054

Heidinger, A. K., Evan, A. T., Foster, M. J., and Walther, A.: A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x, J. Appl. Meteorol. Clim., 51, 1129–1144, 2012. 11047, 11052, 11057

Hollmann, R., Merchant, C., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., De Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van Roozendael, M., and Wagner, W.: The ESA climate change initiative: satellite data records for essential climate variables, B. Am. Meteorol. Soc., 94, 1541–1552, 2013. 11047

Kriegler, F., Malila, W., Nalepka, R., and Richardson, W.: Preprocessing transformations and their effects on multispectral recognition. Sixth International Symposium on Remotes Sensing of Environment, Ann Arbor, MI, 6, 97–131, 1969. 11053

Llewellyn-Jones, D., Edwards, M., Mutlow, C., Birks, A., Barton, I., and Tait, H.: AATSR: Global-change and surface-temperature measurements from Envisat, ESA Bull.-Eur. Space, 105, 11–21, 2001. 11047

AMTD
7, 11045–11085, 2014

Mackie, S., Embury, O., Old, C., Merchant, C., and Francis, P.: Generalized Bayesian cloud detection for satellite imagery, Part 1: Technique and validation for night-time imagery over land and sea, Int. J. Remote Sens., 31, 2573–2594, 2010a. 11047, 11052

Mackie, S., Merchant, C., Embury, O., and Francis, P.: Generalized Bayesian cloud detection for satellite imagery, Part 2: Technique and validation for daytime imagery, Int. J. Remote Sens., 31, 2595–2621, 2010b. 11052

Merchant, C., Harris, A., Maturi, E., and MacCallum, S.: Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval, Q. J. Roy. Meteor. Soc., 131, 2735–2755, 2005. 11047, 11051, 11052

Miguel, A., Bruno, B., Jean-Loup, B., Mark, D., Florence, H., Ulf, K., Constantinos, M., Pierluigi, S., Bruno, G., and Jerome, B.: Sentinel-3 - the ocean and medium-resolution land mission for GMES operational services, ESA Bull.-Eur. Space, 131, 24–29, 2007. 11048

Murtagh, F., Barreto, D., and Marcello, J.: Decision boundaries using Bayes factors: the case of cloud masks, IEEE T. Geosci. Remote, 41, 2952–2958, 2003. 11047

Nieke, J.: Status of the optical payload and processor development of ESA's Sentinel 3 mission, Geoscience and Remote Sensing Symposium, 2008, IGARSS 2008, IEEE International, 4, 427–430, 2008. 11047

Rast, M., Bezy, J. L., and Bruzzi, S.: The ESA Medium Resolution Imaging Spectrometer MERIS a review of the instrument and its mission, Int. J. Remote Sens., 20, 1681–1702, doi:10.1080/014311699212416, 1999. 11047

Rossow, W. B. and Garder, L. C.: Cloud detection using satellite measurements of infrared and visible radiances for ISCCP, J. Climate, 6, 2341–2369, 1993. 11047

Schlundt, C., Kokhanovsky, A. A., von Hoyningen-Huene, W., Dinter, T., Istomina, L., and Burrows, J. P.: Synergetic cloud fraction determination for SCIAMACHY using MERIS, Atmos. Meas. Tech., 4, 319–337, doi:10.5194/amt-4-319-2011, 2011. 11047, 11054

Uddstrom, M. J., Gray, W. R., Murphy, R., Oien, N. A., and Murray, T.: A Bayesian cloud mask for sea surface temperature retrieval, J. Atmos. Ocean. Tech., 16, 117–132, 1999. 11047, 11051, 11052, 11053, 11057

Woodcock, F.: The evaluation of yes/no forecasts for scientific and administrative purposes, Mon. Weather Rev., 104, 1209–1214, 1976. 11054

AMTD
7, 11045–11085, 2014

Table 1. Best found results for feature sets of classical Bayesian cloud masks with two strongly independent features which recreate Synergy cloud mask results best. The results are separated for the Synergy of MERIS and AATSR and MERIS and AATSR alone. Channels are referenced by their central wavelength. MERIS channels carry the unit of nm while for AATSR channels µm is used.

n_F	instrument	skill score	feature set
2	Synergy	0.781	620–900 nm, 412–11 µm
2	Synergy	0.780	442–11 µm, 778–708 nm
2	Synergy	0.776	885–620 nm, dx(11 µm, 442 nm)
2	MERIS	0.781	412 nm, dx(885 nm, 865 nm)
2	MERIS	0.774	412 nm, dx(900 nm, 681 nm)
2	MERIS	0.773	442 nm, dx(900 nm, 708 nm)
2	AATSR	0.707	12 µm/0.55 µm, 3.7 µm/11 µm
2	AATSR	0.706	0.55 µm/3.7 µm, dx(3.7 µm, 12 µm)
2	AATSR	0.706	0.55 µm/12 µm, dx(12 µm, 3.7 µm)

AMTD
7, 11045–11085, 2014

Table 2. Similar as Table 1, but results for classical Bayesian cloud masks based on four strongly independent features are shown.

n_F	instrument	skill score	feature set
4	Synergy	0.826	1.6 µm, 681 nm/778 nm, dx(0.55 µm, 760 nm), dx(11 µm, 412 nm)
4	Synergy	0.821	412 nm, 12 µm, 753 nm × 1.6 µm, dx(11 µm, 12 µm)
4	Synergy	0.820	442 nm, 3.7–11 µm, 3.7 µm × 12 µm, dx(665 nm, 753 nm)
4	MERIS	0.822	412 nm, 900 nm × 510 nm, dx(760 nm, 620 nm), dx(885 nm, 865 nm)
4	MERIS	0.821	753 nm/510 nm, 442 nm × 412 nm, dx(865 nm, 753 nm), dx(885 nm, 760 nm)
4	MERIS	0.818	442 nm, 665 nm/900 nm, dx(560 nm, 510 nm), dx(865 nm, 885 nm)
4	AATSR	0.765	12 µm, 0.55 µm, 12 µm/3.7 µm, 0.55 µm/0.87 µm
4	AATSR	0.757	0.67 µm, 12 µm/3.7 µm, 0.87 µm/0.55 µm, 11 µm/0.67 µm
4	AATSR	0.757	0.55 µm/0.87 µm, 12 µm × 3.7 µm, dx(0.55 µm, 3.7 µm), dx(12 µm, 11 µm)

AMTD
7, 11045–11085, 2014

Table 3. Similar as Table 3, but results for naive Bayesian cloud masks based on five strongly independent features are shown.

n_F	instrument	skill score	feature set
5	Synergy	0.756	12 µm, 760 nm, 412 nm, 560 nm × 490 nm, dx(0.87 µm, 865 nm)
5	Synergy	0.751	681–900 nm, 11–412 nm, 0.87 µm/865 nm, 560 nm × 3.7 µm, dx(708 nm, 490 nm)
5	Synergy	0.750	778 nm, 560 nm, 11–412 nm, 900–620 nm, dx(1.6 µm, 442 nm)
5	MERIS	0.753	412 nm, 442 nm, 865 nm, 560 nm/490 nm, dx(681 nm, 900 nm)
5	MERIS	0.750	412 nm, 510–708 nm, dx(885 nm, 760 nm), dx(665 nm, 900 nm), dx(620 nm, 412 nm)
5	MERIS	0.749	760 nm, 412 nm, 865–490 nm, dx(900 nm, 708 nm), dx(681 nm, 778 nm)
5	AATSR	0.695	11 µm, 12 µm, 11–0.87 µm, 3.7 µm/11 µm, 12 µm × 0.55 µm
5	AATSR	0.692	0.55 µm, 11 µm, 3.7–12 µm, 11–0.87 µm, dx(11 µm, 12 µm)
5	AATSR	0.691	11 µm, 12–3.7 µm, 11 µm × 3.7 µm, dx(0.87 µm, 11 µm), dx(0.55 µm, 12 µm)

AMTD
7, 11045–11085, 2014

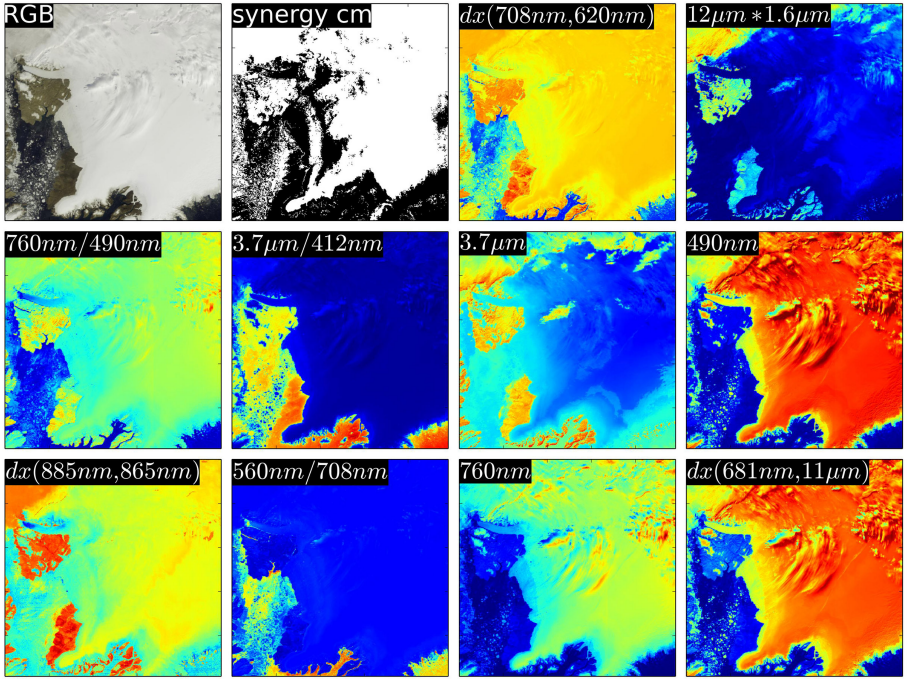

Figure 1. Several views of a scene over Greenland from the 17 July 2007 with the image centered at 59°31′12″ W and 79°0′0″ N. Single panels include a pseudo RGB view, results of the non Bayesian Synergy cloud mask (with white indicating clouds, see Sect. 6), as well as single channels and simple functions operating on two channels. The function dx denotes the index function and is defined as: d$x(a,b) = (a - b)/(a + b)$. Units are not shown and the color-scales are stretched to maximize the visible contrast.

AMTD
7, 11045–11085, 2014

Bayesian cloud detection for MERIS, AATSR, and their combination

A. Hollstein et al.

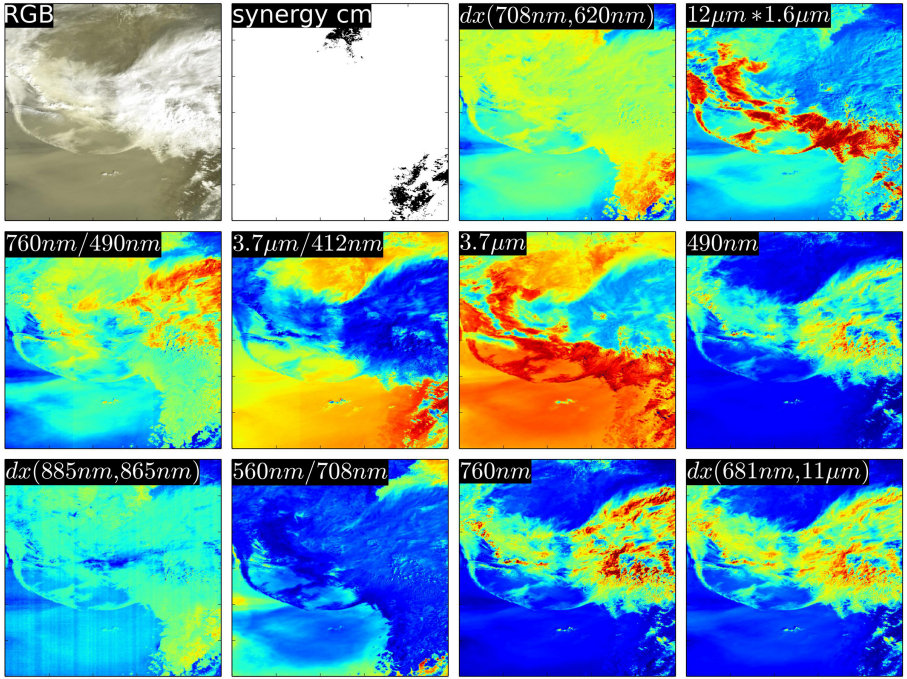

Figure 2. Similar as Fig. 1, but for a scene near the Korean peninsula. The center of the images is located at 125°52′12″ E and 37°45′36″ N.

AMTD
7, 11045–11085, 2014

Bayesian cloud detection for MERIS, AATSR, and their combination

A. Hollstein et al.

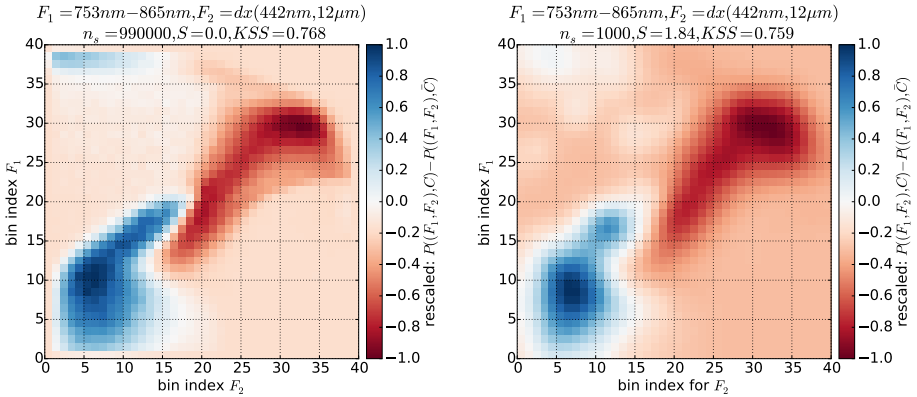

Figure 3. Difference of the two dimensional histograms for $P(F,C)$ and $P(F,\overline{C})$. The left panel show a direct results using 990 000 globally distributed measurements, while for the right panel only 1000 measurements were used. The histograms on the right side were post-processed using Gaussian smoothing with a width parameter of 1.84.

AMTD
7, 11045–11085, 2014

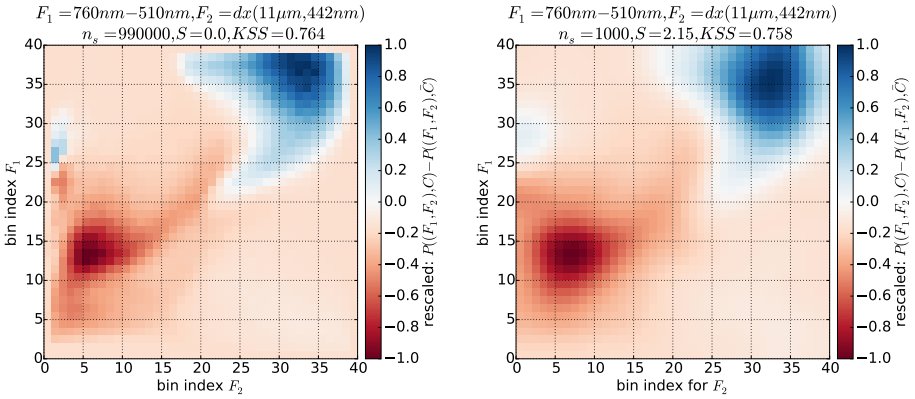

Figure 4. Similar as Fig. 3, but for a different set of Features and a different Gaussian smoothing factor of 2.15. This set of Features includes the MERIS Oxygen A band absorption channel.

AMTD
7, 11045–11085, 2014

Bayesian cloud detection for MERIS, AATSR, and their combination

A. Hollstein et al.

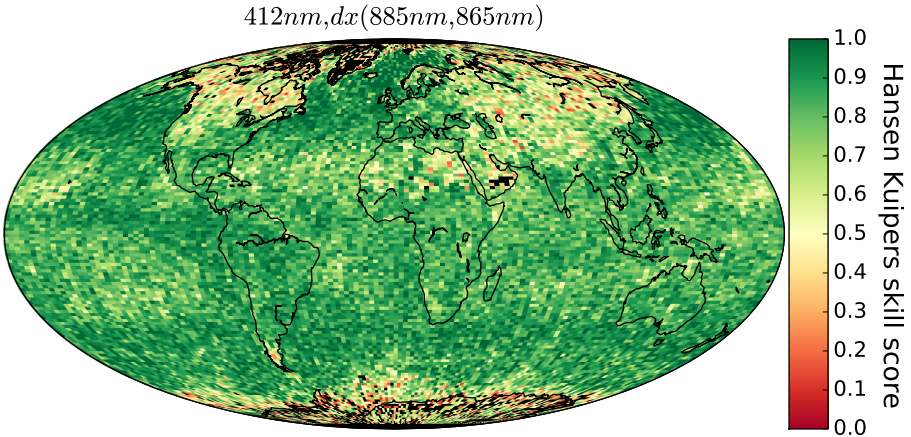

Figure 5. Global distribution of skill scores for a classical Bayesian cloud mask using only two strongly independent features. Data is shown for the year 2008 and the joint probabilities of the mask were estimated with data from the year 2007. The global skill score is 0.78 and the used features are shown in the title of the figure.

AMTD
7, 11045–11085, 2014

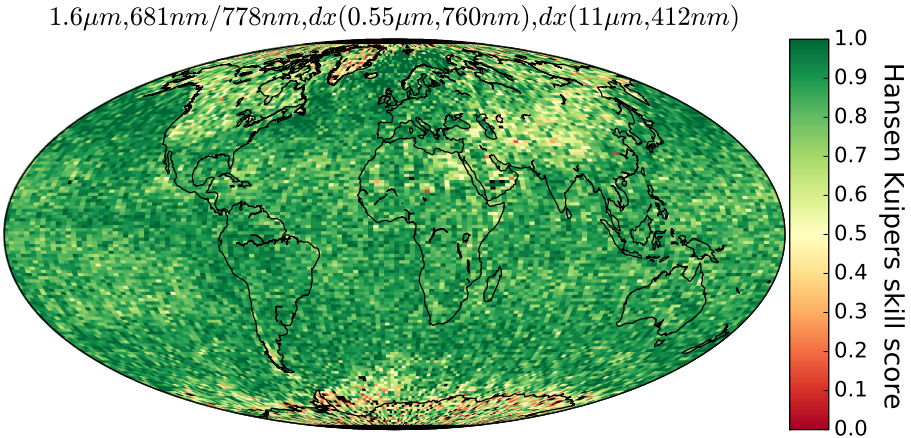

Figure 6. Similar as Fig. 5, but for a different classical Bayesian cloud mask based on four strongly independent features. The global skill score is 0.83.

AMTD
7, 11045–11085, 2014

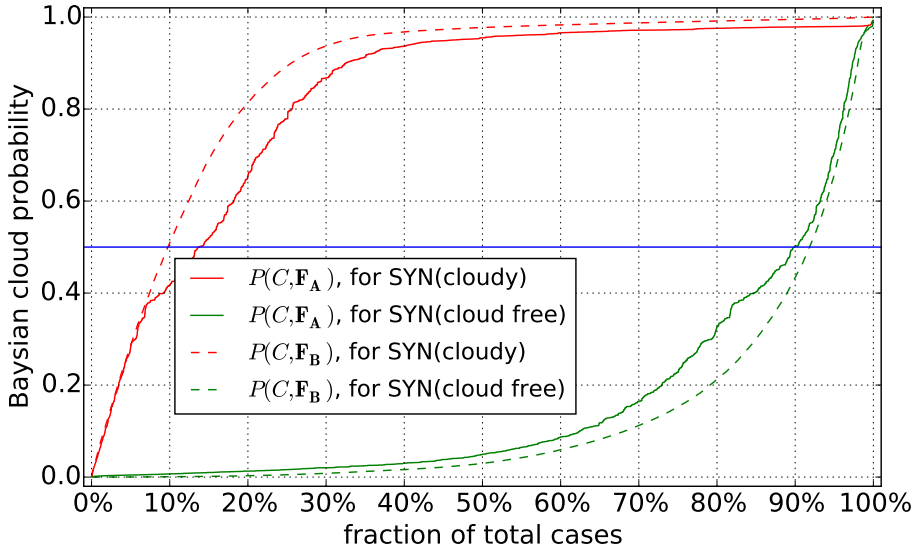

Figure 7. Cloud probability from the two classical Bayesian cloud masks from Fig. 5 (dashed line, $P(C,F_A)$) and Fig. 6 (solid line, $P(C,F_B)$) separated by cases which were labeled as cloudy (red) and non-cloudy (green) by the Synergy cloud mask. Same data as in Figs. 5 and 6 was used and the results were sorted for a better overview. The threshold of 0.5 cloud probability is marked with a blue line.

AMTD
7, 11045–11085, 2014

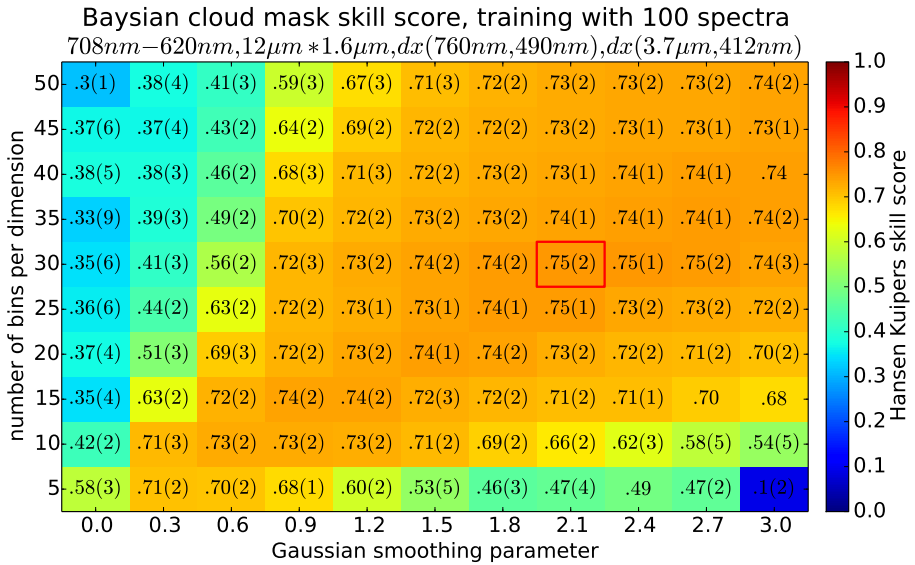

Figure 8. Skill score of a classical Bayesian cloud mask with four strongly independent features with respect to number of bins for each dimension of the underlying histograms and the applied Gaussian smoothing. Artificial truth data is taken from 2007 and skill scores were computed for the year 2008. Only 100 randomly selected and globally distributed spectra were used to compute the histograms. This selection was repeated ten times and mean values for the skill score are shown. The SD on the last significant digit is shown in parenthesis.

AMTD
7, 11045–11085, 2014

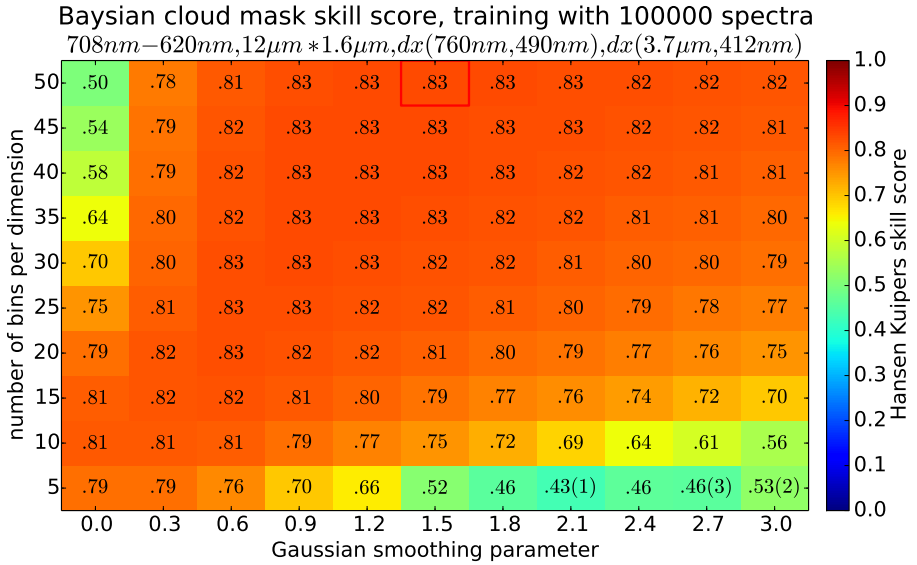

Figure 9. Similar as Fig. 8, but the sample size of the artificial truth was 1000 times larger with all in all 100 k cases.

AMTD
7, 11045–11085, 2014

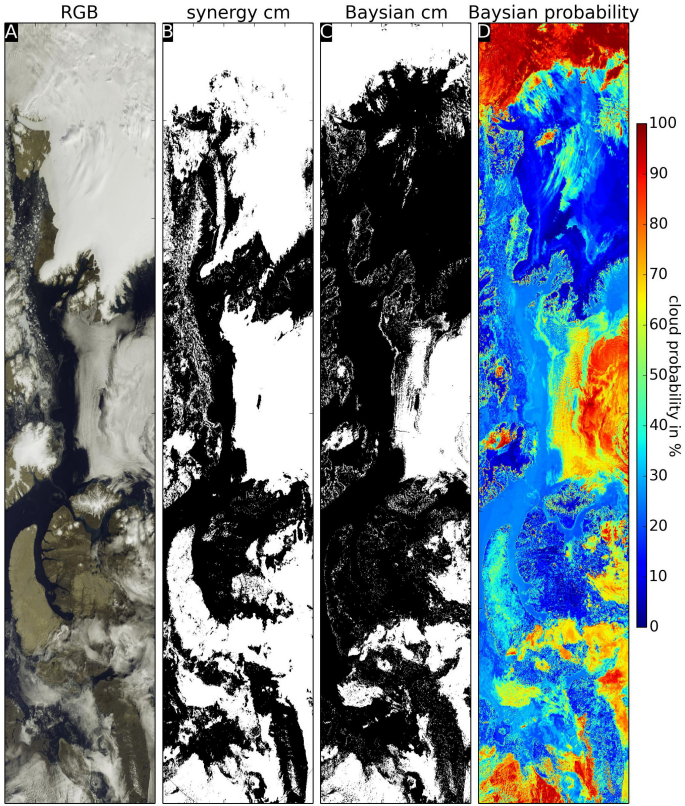

Figure 10. Panel **(a)** shows an RGB view of a larger area of the scene which is shown in Fig. 1. Panel **(b)** shows results of the non Bayesian Synergy cloud mask with some classification errors over the top snow and ice region and the arrow shaped land area in the bottom of the figure. Panel **(c)** shows results of a Bayesian cloud mask which is based on corrected artificial truth from this scene and the one shown in Fig. 11. Panel **(c)** shows the cloud probability results of this Bayesian cloud mask.

AMTD
7, 11045–11085, 2014

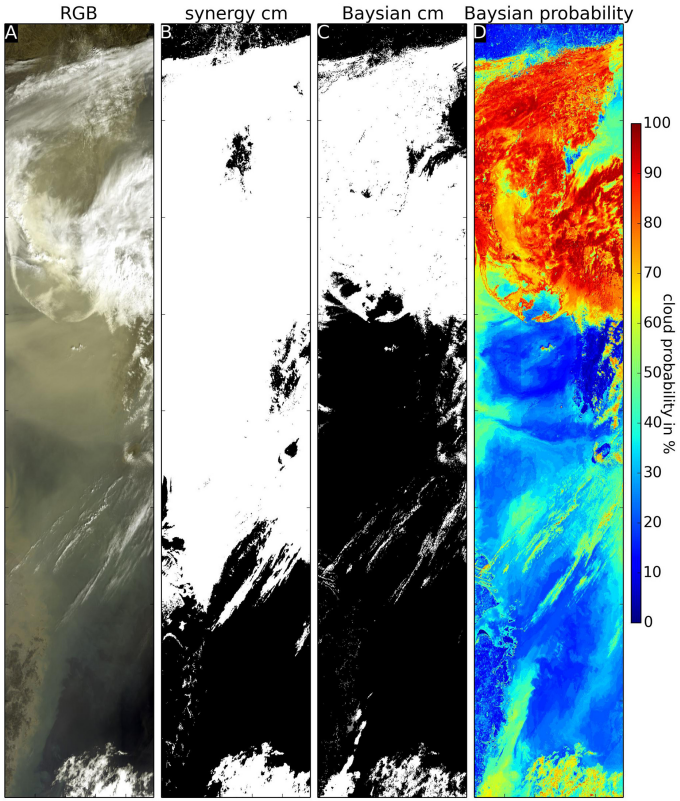

Figure 11. Similar as Fig. 10, but a larger area corresponding to Fig. 2 is shown. Panel **(b)** shows results of the non Bayesian Synergy cloud mask where the strong dust storm is completely classified as cloud. Panel **(c)** shows results of a Bayesian cloud mask which is based on corrected artificial truth from this scene and the one shown in Fig. 10. Panel **(c)** shows the cloud probability results of this Bayesian cloud mask.

AMTD
7, 11045–11085, 2014

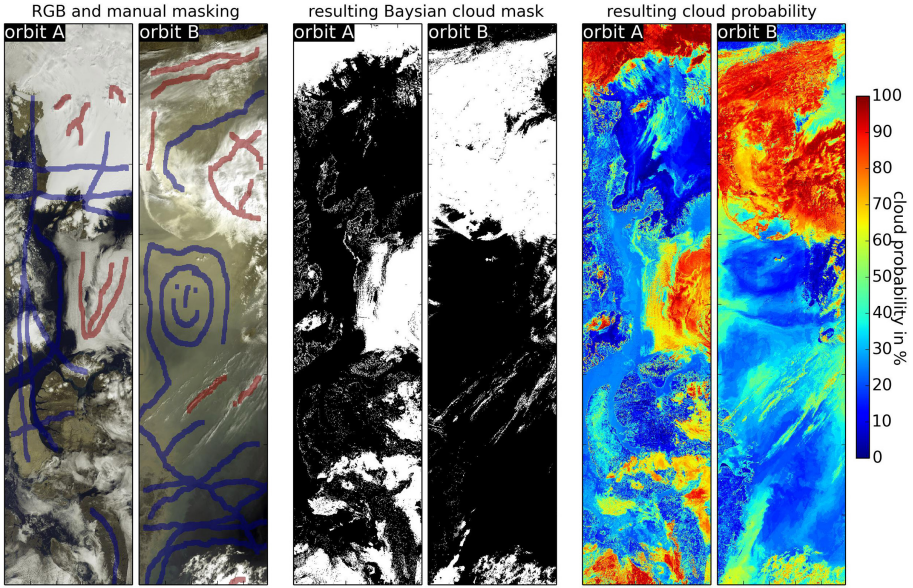

Figure 12. Manual classification of the scenes shown in Figs. 1 and 2. Shown are the cloudy and non-cloudy classification together with an RGB view for two scenes (two leftmost panels, blue = non-cloudy, red = cloudy), the resulting cloud mask (two middle panels), and the cloud probability (rightmost two panels).

AMTD
7, 11045–11085, 2014