5 Validation

In this chapter, the performance of the developed cloud detection algorithm will be quantified by means of independent cloud coverage measurements. Beforehand, the exact objectives and principle limitations of cloud detection will be specified.
A universal cloud detection should satisfy each of the following five requirements:

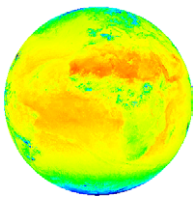
1) Each actually cloud free case shall be detected as cloud free.
2) Each pixel classified as cloud free shall be in fact cloud free.
3) Each actually cloud covered case shall be detected as cloudy.
4) Each pixel classified as cloudy shall be in fact cloud covered.
5) In average, the cloud detection shall produce bias-free results to be applicable for deriving cloud coverage statistics.

The first four points lead to the results that a pixel shall be categorized only as non-classifiable (undecided), if its actual cloud coverage is ambiguous, meaning that the pixel encloses fractional sub pixel cloud coverage. Due to pixel diameters of several kilometers, this would lead to a tremendous amount of non-classifiable pixels. Consequently, high classification rates are only obtainable at the expense of requirement 2) and 4), by classifying also a huge amount of pixels with fractional sub pixel cloud coverage.

Naturally, the results of a cloud detection algorithm distinctly depend on the underlying definition for clouds. Definitions based on e.g. the total water path, the impact on the radiation budget, or the visual recognizability are possible. The cloud definition used for the developed cloud detection scheme could be specified by the cloud-visibility within one of the channel combinations utilized for the manual classification (section 4. 2). Exemplarily the neural networks *nig_sea_nn* and *nig_sea_acs_nn* introduced in section 4. 2, have been confronted with simulated channel brightness temperature data from clouds with varying total water path and cloud top pressure. The post processed network outputs are shown in figure 5-1. High and therefore cold clouds are already detected at total water path values of only $3\,g/m^2$ to $6\,g/m^2$ with high confidence values. The *nig_sea_acs_nn* network, using BT_{ACSBTE}, is inferior in detection of low clouds even with large total water path. In all likelihood, low $BT_{ACSBTE} - BT_{108}$ differences are overestimated by this network. However, the improved capability of low level cloud detection by the *nig_sea_nn* network is associated with the fact that cloud free cases are classified with confidence values of only 0.90 to 0.95, at best. In addition the *nig_sea_acs_nn* network produces less frequent ambiguous results with low confidence values. On the basis of figure 5-1, it can be estimated that the developed cloud detection algorithm is sensitive to the total water path in the range of approximately $0.5\,g/m^2$ to $50\,g/m^2$. Assuming a homogenous cloud consisting of droplets with an effective radius of $r_{eff} = 6\,\mu m$, the corresponding optical thicknesses at $550\,nm$ τ_{550} can be estimated to 0.125 and 12.5, respectively by the following equation [*Stephens*, 1994].

$$\tau_{550} \approx \frac{3\,TWP}{2\,\rho_w\,r_{eff}} \qquad (5\text{-}1)$$

ρ_w stands for the density of water ($1\,g/cm^3$). In the following, the performance of three different cloud detection algorithms based on the developed neural networks will be analyzed: the $FUB_{no\ ACSBTE}$ algorithm, based on those neural networks that do not use BT_{ACSBTE}, the FUB_{ACSBTE} algorithm, based only on neural networks using BT_{ACSBTE},



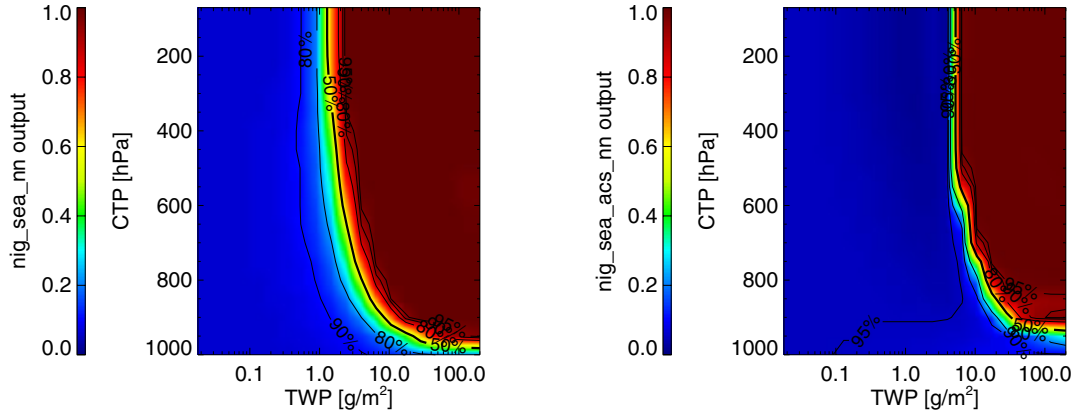


Figure 5-1: Post processed neural network output of the networks *nig_sea_nn* (**left**) and *nig_sea_acs_nn* (**right**) depending on the total water path and the cloud top pressure. XTRA simulated TOA brightness temperatures of a hypothetical water cloud with droplet effective radius of $6\mu m$ have been used as input data. Additionally the confidence levels 50%, 80%, 90%, and 95% are illustrated.

and the $FUB_{ACSBTE\,(night,\,twilight)}$ algorithm using BT_{ACSBTE} only at nighttime and under twilight conditions. The latter has been identified as the superior algorithm and is recommended for routine cloud detection.

5. 1 Data basis and validation methods

5. 1. 1 Synoptical observations

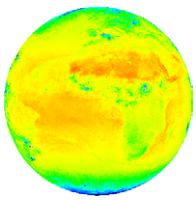
Synoptical (synop) observations are made according to a standardized procedure at fixed times of the day at the stations of the national synop observation networks. Parameters describing the current meteorological situation, including the total cloud coverage and the cloud base height (*CBH*) among others are reported. In general, both parameters are not automatically measured, but are manually estimated by a meteorologist. The total cloud coverage is given in octas. The observing meteorologist is supposed to report 0 or 8 octas only if the entire field of view is completely cloud free or cloud covered, respectively. If there is any fractional cloud coverage in the field of view, the total cloud coverage is

estimated between 1 and 7 octas as linear as possible, thus e.g. 4 octas corresponds to a total cloud coverage of 0.5. 9 octas shall be reported if there is fog in the field of view, or if the cloud coverage cannot be determined for any other reason. These cases have been excluded from the validation dataset.

The synop cloud coverage estimation profits from the powerful human pattern recognition capabilities, but the disadvantage is, that the results are only of subjective nature. Investigations e.g. of *Mohr* [1971] and of *Hahn et al.* [1995] prove that the synop cloud coverage estimation differs from linearity and varies from daytime to nighttime. In *Rossow et al.* [1993] it is also noted that actually "sky cover is not the same quantity as the earth cover seen from a satellite viewing the nadir point".

Many ground-based instruments like ceilometers, LIDARs, or cloud RADARs can determine with high reliability the cloudiness and also the cloud base height. But generally, results of these instruments represent only a narrow region above the instrument, while the synop observation is only limited by the range of vision, usually $30\,km$ to $50\,km$ [*Henderson-Sellers et al.*, 1987]. This becomes important when comparing cloud coverage measurements to satellite pixels with diameters of at least $3\,km$, where homogenous cloud coverage within a pixel cannot be presumed. Systems as described in *Dürr and Philipona* [2004], deriving the total cloud amount by measurements of the surface longwave downward radiation are only rare compared to synop stations.

Europe wide, there is a dense network of synop stations, generating three-hourly reports on the meteorological situation. As the station density in SEVIRI's field of view outside of Europe is much lower, the validation region has been limited to Europe. Additionally, it should be noted that the part of sea-based synop reports e.g. from ships or oil rigs is negligible compared to the part of land-based synop reports.



Altogether, 1378165 synop reports within the long-term validation phase from July, 1st 2004 to December, 31st 2004 and additionally 35022 within the short-term validation phase from June, 3rd 2004 to June, 8th 2004 have been used for validation purposes.
All analyzed synop data has been provided by Dr. E. Reimer, Freie Universität Berlin.

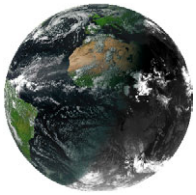
5. 1. 2 Statistical methods

To quantify the quality of a categorizing cloud mask (derived from the cloud detection output) by means of independent measurements like synop reports, several performance quantities can be calculated from the contingency table A given by table 5-1.

		cloud mask		
		cloud free	undecided	cloud covered
synop	cloud free	$A_{1,1}$	$A_{1,2}$	$A_{1,3}$
	undecided	$A_{2,1}$	$A_{2,2}$	$A_{2,3}$
	cloud covered	$A_{3,1}$	$A_{3,2}$	$A_{3,3}$

Table: 5-1: Contingency table A. The matrix elements $A_{n,m}$ represent frequencies of occurrence.

All synop reports with total cloud coverage of 0 or 1 octa are classified as cloud free. 7 and 8 octas are rated as cloud covered. The remaining reports with total cloud coverage between 2 and 6 octas are classified as undecided.
As the SEVIRI pixels are generally much smaller than the field of view of a synoptics and as there is no information on the spatial distribution and position of fractional cloud coverage within the synoptics' field of view, non of the undecided synop reports can be used for quantifying the performance of cloud detection and masking without restrictions. A synop reported cloud coverage of only 2 octas could e. g. properly be classified as cloud covered, if the SEVIRI pixel covers only the cloudy part of the synoptics' field of view.



The risk of accidentally observed complete cloud covered or cloud free pixels when synop reported 1 or 7 octas, respectively, is neglected in support of the total amount of validation data. The number of unambiguous synop reports with 0, 1, 7, or 8 octas utilized for quantifying the cloud mask performance amounts to 896865 within the long-term validation phase and to 21481 within the short-term validation phase.
Consequently for the contingency table applies:

$$A_{2,1} = A_{2,2} = A_{2,3} = 0 \tag{5-2}$$

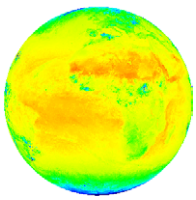
The undecided synop reports between 2 and 6 octas have been analyzed for qualitative estimation of the sensitivity of the cloud detection to fractional sub pixel cloud coverage. This will be described in section 5. 2. 5. Additionally, they are used for calculation of the bias between the mean cloud coverage derived from the cloud detection algorithm and from the synop reports.
In order to obtain a categorizing cloud mask, a confidence threshold c_t between 0.5 and 1.0 has been defined dividing the cloud detection output (representing cloud covered probabilities) into the categories clear sky, undecided, and cloud covered. The scheme to categorize the cloud detection output is shown in figure 5-2 and will be discussed more detailed in section 5. 1. 3.
The cloud mask performance has been quantified by the following statistical parameters:

- The probabiltiy of the cloud mask hitting a cloud free synop report in percentage:

$$\text{CM hits cloud free synop} = \frac{A_{1,1}}{\sum_i A_{1,i}} \cdot 100\,, \tag{5-3}$$

- the probabilty of the cloud mask hitting a cloud covered synop report in percentage:

$$\text{CM hits cloudy synop} = \frac{A_{3,3}}{\sum_i A_{3,i}} \cdot 100\,, \tag{5-4}$$



- the probability of the synop report confirming a cloud free cloud mask classification in percentage:

$$\begin{matrix}\text{synop confirms}\\ \text{cloud free CM}\end{matrix} = \frac{A_{1,1}}{\sum_i A_{i,1}} \cdot 100\,, \tag{5-5}$$

- the probability of the synop report confirming a cloud covered cloud mask classification in percentage:

$$\begin{matrix}\text{synop confirms}\\ \text{cloudy CM}\end{matrix} = \frac{A_{3,3}}{\sum_i A_{i,3}} \cdot 100\,, \tag{5-6}$$

- the probability of accordance of cloud mask and synop report often referred to as "proportion correct" (e.g. in *Stephenson* [2000]) in percentage:

$$\begin{matrix}\text{proportion}\\ \text{correct}\end{matrix} = \frac{1}{N}\sum_i A_{i,i} \cdot 100\,, \tag{5-7}$$

- the percentage of undecided cloud mask cases of overall N cases:

$$\text{undecided cases} = \frac{1}{N}\sum_i A_{i,2} \cdot 100\,, \text{ and} \tag{5-8}$$

- the generalized Kuipers skill score *KSS* as defined in *Wilks* [1995]:

$$KSS = \frac{N\sum_i A_{i,i} - \sum_i\left[\left(\sum_j A_{i,j}\right)\left(\sum_j A_{j,i}\right)\right]}{N^2 - \sum_i\left[\left(\sum_j A_{i,j}\right)^2\right]}\,. \tag{5-9}$$

In addition to these statistical parameters, the overall bias of all cloud detection outputs compared to the corresponding synop reports (including undecided cases) defined as difference between their mean values have been calculated.

The Kuipers skill score (also referred to as Hanssen and Kuipers discriminant, Peirce skill score, or true skill statistic) is often utilized to describe the performance of a forecast or measurement in relevance to an assumed truth. In contrast to the proportion correct which is heavily influenced by the most common category, the Kuipers skill score is also suitable for data with asymmetrical distributions (e.g. more cloudy than cloud free cases). Its value

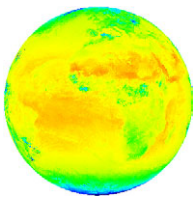
can be interpreted as accuracy of the forecast/measurement in predicting the correct category, relative to that of random chance. It is not interpretable as intuitively as e.g. the probability for correct cloud free measurements, but in exchange it is suitable to describe the cloud mask performance by only one number. The Kuipers skill score would be 1 only if the forecast was always true, 0 if the forecast was randomly chosen, and -1 if the forecast was always contrary to the truth. In *Woodcock* [1976], the Kuipers skill score is recommended to be the most suitable skill score for scientific and administrative purposes.

Referring to *BMRC* [2004], the non-parametric bootstrap method, introduced by *Efron* [1979], is ideally suited to estimate confidence intervals of verification scores like the Kuipers skill score. As briefly described in *BMRC* [2004], the non-parametric bootstrap procedure is quite simple: 1) Generate a bootstrap sample by drawing (with replacement) a random sample of N cloud detection output/synop pairs from the full set of N samples. 2) Compute the verification statistic (e.g. Kuipers skill score). 3) Repeat steps 1 and 2 numerous times. 4) Calculate e.g. standard deviation or confidence intervals of the verification statistic from the results of step 3. In the following, any margins of error concerning the quality of cloud detection or masking have been derived by calculating the standard deviation of 100 bootstrap samples.

5. 1. 3 Categorizing the output of the cloud detection algorithm

As described in the latter section, a confidence threshold c_t in the range of [1.0, 0.5] is utilized to categorize the network output in clear sky, undecided, and cloud covered cases. Therefore, the confidence value c_{onf} of the network output out is defined analog to equation (4-4) and the categorizing is done as shown in figure 5-2.

Based on the synop long-term validation dataset, the influence of c_t on the accordance of the cloud mask with the corresponding synop reports has been analyzed. The results are given in figure 5-3. The statistical parameters given by equation (5-3) to (5-9) are depicted



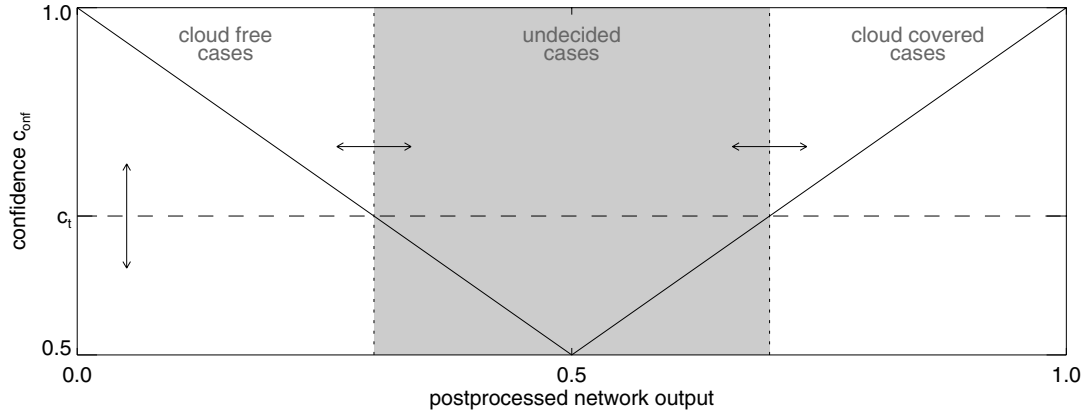
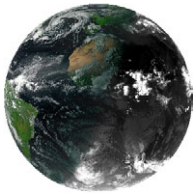


Figure 5-2: Scheme to categorize the cloud detection output by its confidence value c_{onf} (defined analog to equation (4-4)) and the confidence threshold c_t.

in respect to different values of c_t for the three algorithms $FUB_{noACSBTE}$, FUB_{ACSBTE}, and $FUB_{ACSBTE(night,twilight)}$. As expected, while decreasing c_t from 0.98 to 0.5, the ratio of undecided cases falls smoothly from more then 25 % to 0 %. With the decreasing ratio of undecided cases, the probability of the cloud mask hitting a cloud free synop report increases by about 20 %. The same applies to the probability of the cloud mask hitting a cloud covered synop report. Compared to this, the probability of the synop report confirming the cloud mask decreases only marginally at the same time. This means that the major part of cloud detection outputs with low confidence values can be categorized reliably, too. This is confirmed by a monotonic increasing Kuipers skill score and a monotonic increasing proportion correct which both reach their maxima at the confidence threshold 0.5. Thus, all further validation studies are based on this confidence threshold value.

It stands out that generally the $FUB_{ACSBTE(night,twilight)}$ algorithm produces best results, followed by the FUB_{ACSBTE} algorithm. The reasons for this behavior will be discussed in section 5. 2. 1.



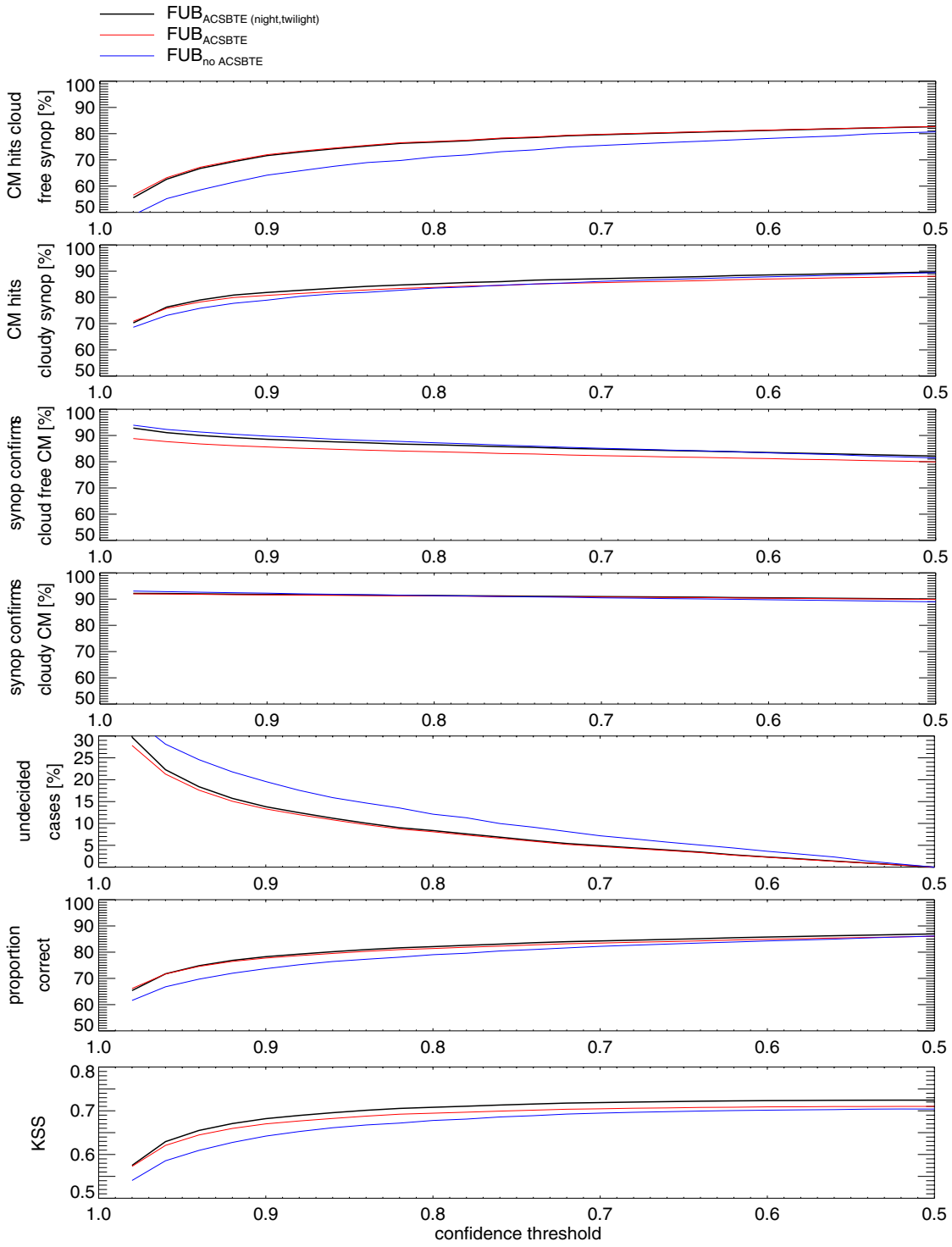
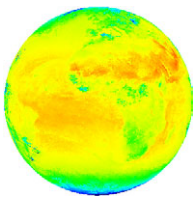


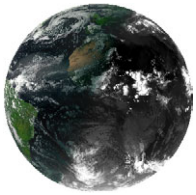
Figure 5-3: Cloud mask performance for the algorithms $FUB_{noACSBTE}$, FUB_{ACSBTE}, and $FUB_{ACSBTE(night,twilight)}$ dependent on the confidence threshold c_t varying from 0.98 to 0.5.



5. 2 Long-term validation

5. 2. 1 General performance

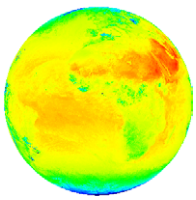
The overall performance of cloud detection and masking within the half year long-term validation phase from July, 1st 2004 to December, 31st 2004 was derived for the different FUB algorithms. Table 5-2 shows that at nighttime and even more under twilight conditions, the usage of BT_{ACSBTE} leads to better results in all investigated categories, except for the bias that does not change much. The Kuipers skill score increases from 0.634±0.001 to 0.658±0.001 at nighttime and from 0.655±0.002 to 0.699±0.002 at twilight. For the neural networks for twilight conditions, it seems that the usage of BT_{ACSBTE} can compensate the missing or ambiguous BT_{039} information. This behavior is contrary to that at daytime, where the usage of BT_{ACSBTE} leads to a larger negative bias and less performance in the categories "CM hits cloudy synop" and "synop confirms cloud free CM". As a result, the Kuipers skill score decreases from 0.858±0.001 to 0.811±0.001. Especially for low, warm clouds over land, where the BT_{ACSBTE} data is less accurate than over sea, the channels in the visible spectral region provide much more information on the cloud coverage than BT_{ACSBTE} does. An over-interpretation of BT_{ACSBTE} when creating the training dataset might be the reason for the fact that the additional usage of BT_{ACSBTE} reduces the cloud detection quality at daytime. As the daytime networks are only utilized for 30.8% of all cases, the disadvantages of the FUB_{ACSBTE} algorithm at daytime compared to the $FUB_{no\ ACSBTE}$ algorithm are compensated by the advantages at nighttime and twilight in the overall statistic. To combine the strengths of both, the $FUB_{ACSBTE(night,twilight)}$ algorithm was defined utilizing BT_{ACSBTE} information only at nighttime and twilight. This algorithm leads to best overall results. Its Kuipers skill score amounts to 0.724±0.001, its proportion correct is 86.96±0.04% and its bias of -0.0100±0.0003 is rather small.



situation	algorithm identifier	frequency of usage [%]	CM hits cloud free synop [%]	CM hits cloudy synop [%]	synop confirms cloud free CM [%]	synop confirms cloudy CM [%]	proportion correct [%]	*KSS*	bias (CM - synop)
night	$FUB_{no\ ACSBTE}$	51.6	75.60 ±0.09	87.77 ±0.06	80.04 ±0.09	84.73 ±0.07	82.98 ±0.05	0.634 ±0.001	0.0411 ±0.0005
	FUB_{ACSBTE}		77.98 ±0.09	87.84 ±0.06	80.62 ±0.10	86.02 ±0.06	83.96 ±0.05	0.658 ±0.001	0.0362 ±0.0005
twilight / sun glint	$FUB_{no\ ACSBTE}$	17.6	79.90 ±0.20	85.62 ±0.10	68.00 ±0.19	91.76 ±0.09	84.04 ±0.10	0.655 ±0.002	-0.0372 ±0.0008
	FUB_{ACSBTE}		82.95 ±0.18	86.97 ±0.10	70.89 ±0.20	93.02 ±0.08	85.86 ±0.09	0.699 ±0.002	-0.0399 ±0.0009
day	$FUB_{no\ ACSBTE}$	30.8	93.78 ±0.06	91.97 ±0.07	86.94 ±0.11	96.29 ±0.04	92.63 ±0.05	0.858 ±0.001	-0.0568 ±0.0004
	FUB_{ACSBTE}		93.66 ±0.08	87.40 ±0.08	80.89 ±0.13	96.03 ±0.05	89.67 ±0.06	0.811 ±0.001	-0.0844 ±0.0006
overall	$FUB_{no\ ACSBTE}$		81.76 ±0.07	88.64 ±0.04	80.43 ±0.07	89.49 ±0.04	86.14 ±0.04	0.704 ±0.001	-0.0072 ±0.0003
	FUB_{ACSBTE}	100	83.46 ±0.07	87.53 ±0.04	79.27 ±0.07	90.26 ±0.04	86.05 ±0.03	0.710 ±0.001	-0.0197 ±0.0004
	$FUB_{ACSBTE(n.,t.)}$		83.50 ±0.06	88.94 ±0.04	81.18 ±0.06	90.42 ±0.04	86.96 ±0.04	0.724 ±0.001	-0.0100 ±0.0003

Table: 5-2: Summarizing statistics describing the cloud detection/masking performance of the FUB algorithms during the whole long-term validation phase. The overall results for the FUB algorithm that is recommend for routine operation are gray-shaded.

In order to estimate the quality of cloud detection and masking over possibly snowy surfaces, 10679 synop reports from stations higher than $2000\,m$ have been analyzed. The $FUB_{ACSBTE(night,twilight)}$ algorithm produced the best results, although a Kuipers skill score of 0.67±0.01, a proportion correct of 81.0±0.6% and a bias of 0.087±0.005 indicate that discriminating clouds from snowy surfaces is not trivial and leads to an



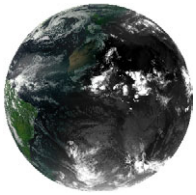
overestimation of cloudiness in these cases. In *Kästner and Kriebel* [2001] an alpine cloud climatology is described. For this publication the APOLLO cloud detection scheme was utilized. Its overestimation of cloudiness for non-vegetated (possibly snowy) surfaces is specified with 15%. According to *Rossow et al.* [1993], the ISCCP cloud detection algorithm seems to behave contrary to this, as there is an underestimation of cloudiness over polar land surfaces of about 22%.

5. 2. 2 Land/sea differences

situation	algorithm identifier	number of synop reports	CM hits cloud free synop [%]	CM hits cloudy synop [%]	synop confirms cloud free CM [%]	synop confirms cloudy CM [%]	proportion correct [%]	*KSS*	bias (CM - synop)
land	$FUB_{ACSBTE(n.,t.)}$	1237838	83.85 ±0.07	89.23 ±0.05	80.90 ±0.08	91.03 ±0.04	87.33 ±0.04	0.7308 ±0.0008	-0.0093 ±0.0004
sea		140327	81.12 ±0.20	85.86 ±0.16	83.15 ±0.18	84.08 ±0.17	83.66 ±0.12	0.6697 ±0.0025	-0.0157 ±0.0012

Table: 5-3: Summarizing statistics describing the overall cloud detection/masking performance of the $FUB_{ACSBTE(night,twilight)}$ algorithm depending on land/sea.

As different neural networks and different thresholds *IDT* and *EDT* (section 3. 2. 2) are used for land and sea surfaces, the land/sea dependency of the cloud detection and masking performance has been analyzed for the $FUB_{ACSBTE(night,twilight)}$ algorithm which is recommend for routine operation. The results are given in table 5-3. The column "number of synop reports" shows that only approximately 10% of the analyzed synop reports are assigned to sea surfaces according to SEVIRI's land/sea mask. The $FUB_{ACSBTE(night,twilight)}$ algorithm produces in nearly all analyzed categories better results over land surfaces. Solely, the probability of the synop report confirming a cloud



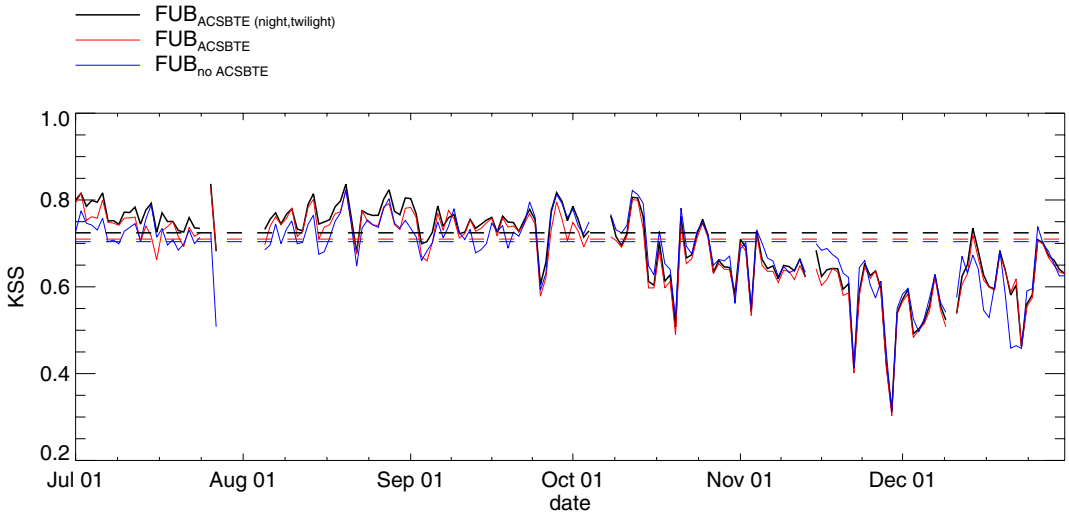
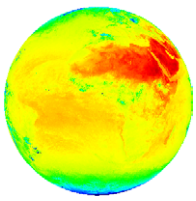


Figure 5-4: Seasonal evolution of cloud mask performance on a daily basis expressed by the Kuipers skill score. The overall values are given by the dashed lines.

free cloud mask classification is reduced slightly over land. This complies with the fact that most of the land-networks are more effective in classifying the samples of their training and test datasets with higher confidence levels.

5. 2. 3 Seasonal features

Using the Kuipers skill score as performance index, it was analyzed in which manner seasonal features affect the quality of cloud masking. Figure 5-4 shows, that the cloud masking performance is reduced in fall and winter with simultaneously increased variability. The lowered performance in these seasons can be a consequence of less contrasts between clouds and surfaces in the thermal channels due to colder surfaces or in the visible channels due to snowy surfaces. Furthermore, the daytime networks are less often utilized in fall and winter. The quality difference between the three tested algorithms shows no seasonal effect. For almost every day, the $FUB_{ACSBTE(night,twilight)}$ algorithm produces the best results.



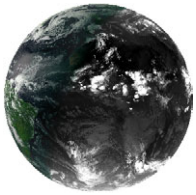
5. 2. 4 Influence of the cloud height

In section 5. 2. 1 it was already suggested that an overvaluation of BT_{ACSBTE} compared to the information from the visible channels while creating the training dataset might be the reason for the lowered performance of the daytime networks utilizing BT_{ACSBTE}. In the introduction of chapter 5, the neural networks *nig_sea_acs_nn* and *nig_sea_nn* have been confronted with simulated SEVIRI data. Consistently, it was shown that the usage of BT_{ACSBTE} is advantageous for the detection of clear sky cases with high confidence values but disadvantageous for the detection of low clouds.

All cloud covered synop reports from the long-term validation phase have been analyzed in respect to the estimated cloud base height (*CBH*) which is part of each synop report. When confronting the neural networks with measured data, the statistic in figure 5-5 shows that the usage of BT_{ACSBTE} only at daytime leads to a smaller probability for detecting low clouds. Overall, the best results for detecting clouds with *CBH* less than $1000m$ have been obtained using the $FUB_{ACSBTE(night,twilight)}$ algorithm. In addition, figure 5-5 shows that for clouds with *CBH* greater than $2500m$, the probability of detection is reduced again. This can be explained by the fact that these clouds are often cirrus clouds with low total water path and therefore hard to detect. The effect, that the probability for detection of high clouds is greater at nighttime than at daytime might be a result of differing synoptical cirrus detection quality.

5. 2. 5 Sensitivity to fractional cloud coverage

In section 5. 1. 2 it was suggested, that ambiguous synop reports with cloud coverage between 2 and 6 octas can be utilized for a qualitative analysis of the cloud detection algorithm's sensitivity to fractional sub pixel cloud coverage.



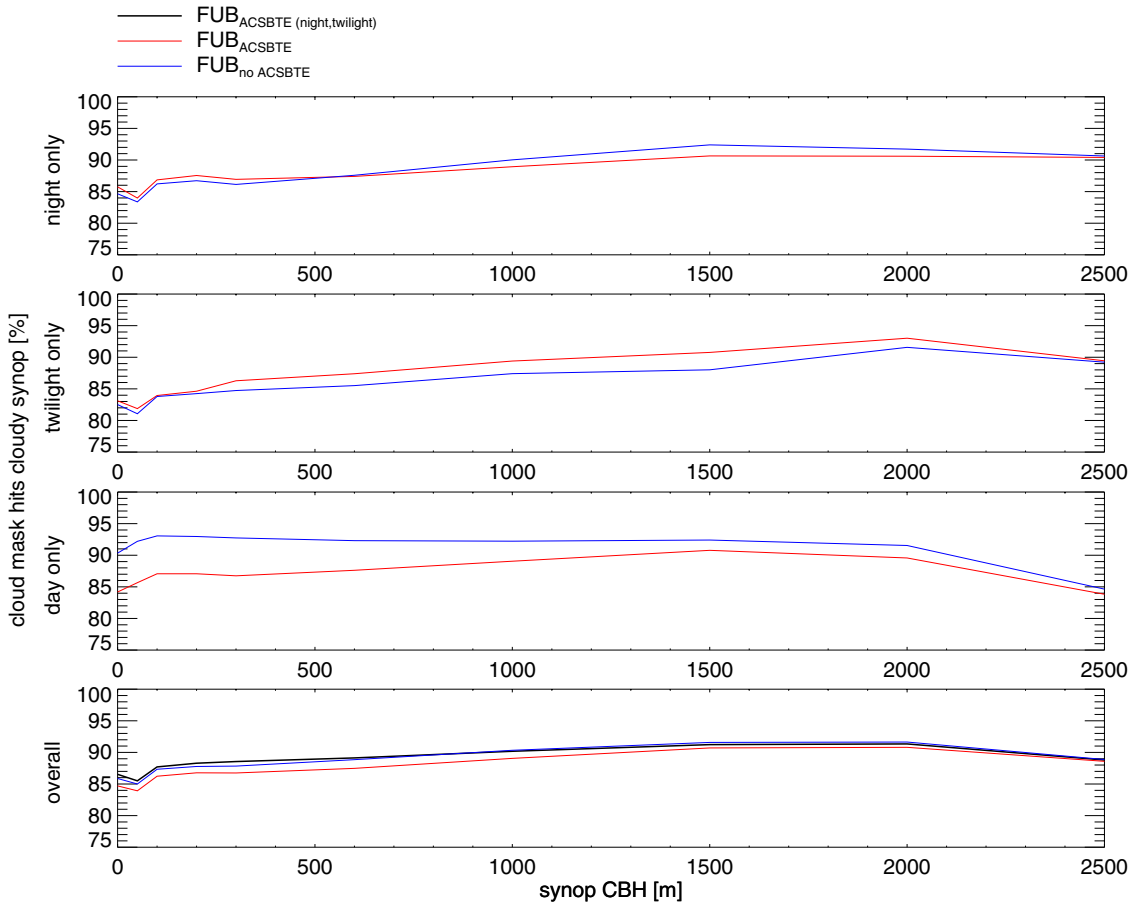
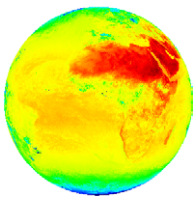


Figure 5-5: Influence of cloud base height (*CBH*) taken from synop observations on the probability of cloud detection. The x-axis gives the estimated minimum level of cloud base height, so that $2500\,m$ represents all clouds with a cloud base height greater equal $2500\,m$.

For unambiguous synop reports with cloud coverage of 0, 1, 7, or 8 octas, the probability of fractional cloudiness within the corresponding SEVIRI pixels is negligible. Assuming hypothetically, all SEVIRI pixels would have no fractional sub pixel cloud coverage, even for all ambiguous synop reports. In this case there should be a linear relationship between the synop reported cloud coverage and the mean corresponding cloud detection output in the range of 1 to 7 octas. Without doubt this hypothesis is erroneous, because the probability to stumble on fractional cloudiness within a SEVIRI pixel will have its maximum somewhere between 2 and 6 octas synop reported cloud coverage. For this reason, the deviation from linearity of the statistical relationship between synop and cloud



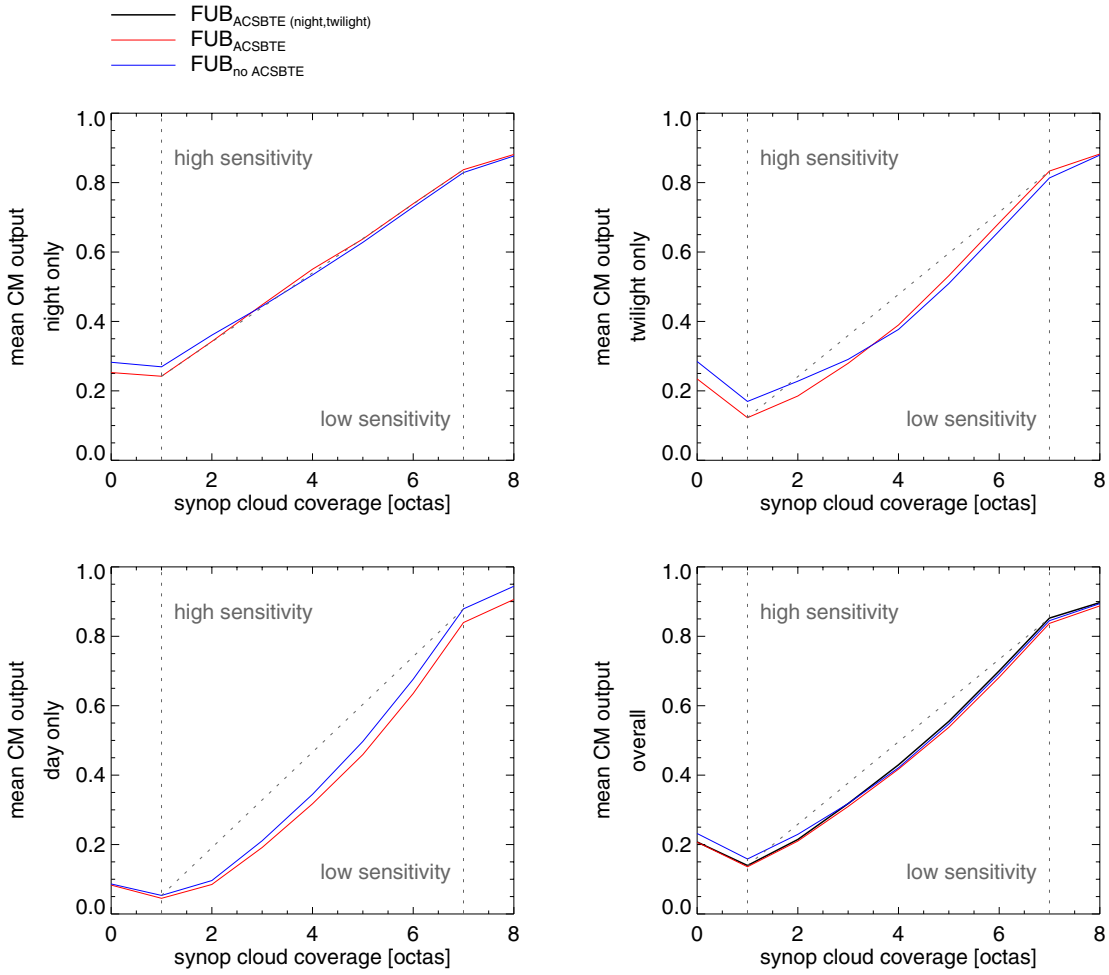
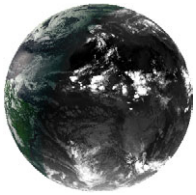


Figure 5-6: Sensitivity of cloud detection to sub pixel cloud fraction under daytime, twilight, and nighttime conditions and in the overall average. Curve progressions below the depicted linearity indicate low sensitivities while a curve progressions above linearity indicate the contrary. For reasons of simplicity in each plot the linearity is depicted only for the $FUB_{ACSBTE(night,twilight)}$ algorithm.

detection output allows qualitative conclusions on the cloud detection algorithm's sensitivity to sub pixel cloudiness. A curve progression below linearity indicates that sub pixel cloud fraction is more often interpreted as cloud free while a curve progression above linearity indicates the contrary. In the case of a linear relation, neutral sensitivity to sub pixel cloud fraction can be assumed.



Of course, the whole argumentation applies only if the synop observations between 1 and 7 octas have a linear relationship to the actual cloud coverage. As already mentioned in section 5. 1. 1, this assumption is only a rough approximation [*Mohr*, 1971]. Nevertheless, this investigation is suitable for a qualitative comparison of different algorithms among themselves.

Figure 5-6 shows that all compared algorithms have a lower sensitivity to fractional sub pixel cloudiness at daytime and twilight than at nighttime and in the overall average. This is one reason for the underestimation of the mean cloud coverage at daytime and twilight, expressed by negative bias values (table 5-2). A possible reason for non-linear sensitivity to fractional sub pixel cloud coverage is the non-linearity of BT_{039} to the sub pixel temperature composition (figure 2-6 and figure 2-22).

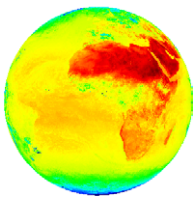
5. 3 Short-term validation

5. 3. 1 General performance

The *Institut für Weltraumwissenschaften* contributes to the cloud top pressure validation project of EUMETSAT (ITT 03/527). As incorrect cloud masking can cause large errors in cloud top pressure processing, the performance of the EUMETSAT cloud detection was analyzed by means of synoptical observations. The results have been compared to those from the FUB cloud detection.

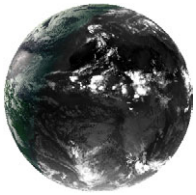
The investigated EUMETSAT cloud mask has been briefly introduced in chapter 1. A detailed algorithm description can be found in *EUMETSAT* [2004a]. As the EUMETSAT cloud mask is part of the EUMETSAT cloud analysis (CLA) product, it will be referred to as $EUMETSAT_{CLA}$ in the following.

The analyzed data extends over the total cloud top pressure validation project's second measurement phase from June, 3rd 2004 to June, 8th 2004. Table 5-4 summarizes the general performance of cloud detection and masking within this period. The results for the



situation	algorithm identifier	frequency of usage [%]	CM hits cloud free synop [%]	CM hits cloudy synop [%]	synop confirms cloud free CM [%]	synop confirms cloudy CM [%]	proportion correct [%]	*KSS*	bias (CM - synop)
night	$FUB_{no\ ACSBTE}$		72.2 ±0.8	92.6 ±0.5	91.0 ±0.5	76.1 ±0.7	82.2 ±0.4	0.648 ±0.009	0.110 ±0.004
	FUB_{ACSBTE}	31.4	82.8 ±0.6	90.2 ±0.5	89.8 ±0.6	83.4 ±0.6	86.4 ±0.4	0.730 ±0.007	0.063 ±0.004
	$EUMETSAT_{CLA}$		87.3 ±0.6	82.0 ±0.6	83.5 ±0.6	86.1 ±0.6	84.7 ±0.4	0.693 ±0.009	-0.040 ±0.004
twilight / sun glint	$FUB_{no\ ACSBTE}$		74.4 ±1.3	85.1 ±0.8	73.0 ±1.3	86.0 ±0.7	81.3 ±0.7	0.595 ±0.014	-0.031 ±0.006
	FUB_{ACSBTE}	14.7	92.6 ±0.8	86.8 ±0.8	79.2 ±1.2	95.6 ±0.5	88.8 ±0.6	0.794 ±0.011	-0.062 ±0.005
	$EUMETSAT_{CLA}$		90.1 ±1.0	86.4 ±0.8	78.2 ±1.2	94.2 ±0.6	87.7 ±0.6	0.765 ±0.011	-0.073 ±0.007
day	$FUB_{no\ ACSBTE}$		93.8 ±0.4	93.3 ±0.3	87.8 ±0.5	96.7 ±0.2	93.5 ±0.2	0.871 ±0.005	-0.062 ±0.002
	FUB_{ACSBTE}	53.9	93.4 ±0.3	87.8 ±0.4	79.7 ±0.6	96.3 ±0.2	89.7 ±0.3	0.812 ±0.005	-0.102 ±0.003
	$EUMETSAT_{CLA}$		91.5 ±0.4	86.1 ±0.4	77.1 ±0.6	95.2 ±0.2	87.9 ±0.3	0.776 ±0.006	-0.093 ±0.003
overall	$FUB_{no\ ACSBTE}$		82.5 ±0.4	91.8 ±0.2	86.8 ±0.4	88.9 ±0.3	88.1 ±0.2	0.743 ±0.005	-0.008 ±0.002
	FUB_{ACSBTE}	100	89.0 ±0.3	88.2 ±0.3	83.1 ±0.3	92.5 ±0.2	88.5 ±0.2	0.772 ±0.004	-0.049 ±0.002
	$FUB_{ACSBTE\,(n.,t.)}$		89.2 ±0.4	91.5 ±0.2	87.2 ±0.4	92.8 ±0.2	90.6 ±0.2	0.807 ±0.004	-0.026 ±0.002
	$EUMETSAT_{CLA}$		89.6 ±0.3	85.1 ±0.3	79.7 ±0.4	92.6 ±0.2	86.9 ±0.2	0.747 ±0.005	-0.075 ±0.002

Table: 5-4: Summarizing statistics describing the cloud detection/masking performance of the FUB algorithms and the $EUMETSAT_{CLA}$ algorithm within the whole short-term validation phase. The overall results for the $FUB_{ACSBTE\,(night,\,twilight)}$ and for the $EUMETSAT_{CLA}$ algorithm are gray-shaded.

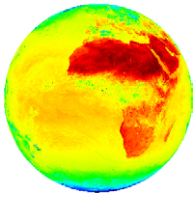


algorithms $FUB_{no\ ACSBTE}$, FUB_{ACSBTE}, and $FUB_{ACSBTE(night,twilight)}$ are similar to the results from the long-term validation phase given in table 5-2. Compared to this phase, the overall detection quality profits from a higher percentage of daytime cases and from larger Kuipers skill score values for the nighttime and twilight algorithms utilizing BT_{ACSBTE}. In the overall statistic, the FUB_{ACSBTE} algorithm is superior to the algorithm $FUB_{no\ ACSBTE}$. Nevertheless the $FUB_{ACSBTE(night,twilight)}$ algorithm produces the best results.

The overall performance of the $EUMETSAT_{CLA}$ algorithm measured by the Kuipers skill score amounts to 0.747±0.005 compared to 0.807±0.004 for the $FUB_{ACSBTE(night,twilight)}$, 0.772±0.004 for the FUB_{ACSBTE}, and 0.743±0.005 for the $FUB_{no\ ACSBTE}$ algorithm. The main weakness of the $EUMETSAT_{CLA}$ algorithm seems to be the under estimation of cloudiness manifested in an overall bias of -0.075±0.002 and lower probabilities of the cloud mask hitting a cloudy synop observation and of a synop report confirming a cloud free cloud mask pixel. This cloud free bias is partly caused by a low sensitivity to sub pixel cloud fraction. The $FUB_{ACSBTE(night,twilight)}$ algorithm has a similar sensitivity to sub pixel cloud fraction, but its overall bias amounts only to -0.026±0.002. The overall probabilities that the cloud mask hits a cloud free synop report, and that a synop observation confirms a cloudy cloud mask pixel are slightly higher for the $EUMETSAT_{CLA}$ than for the $FUB_{ACSBTE(night,twilight)}$ algorithm.

5. 3. 2 Land/sea differences

Analog to section 5. 2. 2, the land/sea dependency of the performance of cloud detection and masking has been analyzed within the short-term validation phase for the $EUMETSAT_{CLA}$ and for the $FUB_{ACSBTE(night,twilight)}$ algorithm. Table 5-5 shows that the land/sea dependency of the $FUB_{ACSBTE(night,twilight)}$ algorithm within the short-term validation phase is comparable to that within the long-term validation (compare table 5-3): The algorithm produces better results in nearly all categories over land surfaces. In principle the $EUMETSAT_{CLA}$ algorithm shows a similar behavior but the



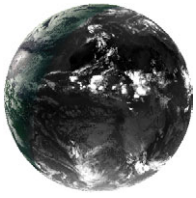
situation	algorithm identifier	number of synop reports	CM hits cloud free synop [%]	CM hits cloudy synop [%]	synop confirms cloud free CM [%]	synop confirms cloudy CM [%]	proportion correct [%]	*KSS*	bias (CM - synop)
land	$FUB_{ACSBTE(n.,t.)}$	31699	89.8 ±0.4	91.6 ±0.3	87.0 ±0.4	93.5 ±0.2	90.9 ±0.2	0.815 ±0.004	-0.027 ±0.002
	$EUMETSAT_{CLA}$		91.5 ±0.3	85.4 ±0.3	79.6 ±0.5	94.2 ±0.2	87.7 ±0.2	0.769 ±0.005	-0.084 ±0.002
sea	$FUB_{ACSBTE(n.,t.)}$	3323	84.4 ±1.1	89.8 ±1.0	89.0 ±1.1	85.4 ±1.1	87.1 ±0.8	0.742 ±0.015	-0.018 ±0.007
	$EUMETSAT_{CLA}$		75.4 ±1.4	81.9 ±1.3	80.3 ±1.4	77.2 ±1.3	78.7 ±1.0	0.573 ±0.020	0.005 ±0.010

Table: 5-5: Summarizing statistics describing the overall cloud detection/masking performance of the $FUB_{ACSBTE(night,twilight)}$ algorithm and the $EUMETSAT_{CLA}$ algorithm depending on land/sea.

performance differences between land and sea are much more pronounced. As extreme example, the probability of a synop report confirming a cloud covered cloud mask classification decreases from 94.2±0.2% over land surfaces to 77.2±1.3% over sea surfaces. The Kuipers skill score decreases from 0.769±0.005 over land surfaces to 0.573±0.020 over sea surfaces. The corresponding values for the $FUB_{ACSBTE(night,twilight)}$ algorithm are 0.815±0.004 for land surfaces and 0.742±0.015 for sea surfaces.

5. 3. 3 Influence of the cloud height

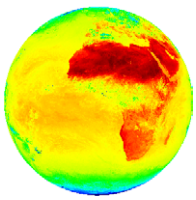
Analog to section 5. 2. 4, it was analyzed for the short-term validation phase, in which manner the probability of cloud detection depends on the cloud base height. Figure 5-7 shows, that the $EUMETSAT_{CLA}$ algorithm has the lowest overall detection rate for synoptical observed cloud coverage. This behavior is pronounced especially for low



clouds with cloud base height between $0\,m$ and $50\,m$, where the probability of detection decreases to less than $79\,\%$. The underestimation of low clouds could be a consequence of an overvaluation of one of the dynamic threshold tests utilizing estimated clear sky brightness temperatures. This would be conform with the underestimation of low clouds by the FUB_{ACSBTE} algorithm, as the usage of BT_{ACSBTE} can be interpreted as a dynamic threshold method. It was announced on the EUMETSAT website that the scene classification scheme was updated on August, 23rd 2004 with the effect that "more pixels are now classified as cloud during daytime" [*EUMETSAT*, 2004c]. A comparison to results of the updated cloud detection scheme, in particular with regard to the overall bias and the detection probability of low clouds, would be an interesting future task.
For all tested algorithms, high clouds with *CBH* greater than $2500\,m$ have been detected with similar probabilities. It is highly visible that the probability of detection is distinctly lower for these clouds than for clouds at medium heights. This might hint at weaknesses of both algorithms in the detection of thin cirrus clouds. Not without reason, the problem of thin cirrus detection is discussed in numerous publications [e.g. *Ackerman et al.*, 1990; *Hutchison and Choe*, 1996].

5. 3. 4 Visual cloud mask evaluation

Undoubtedly, it is possible to find excellent but also just good cases for any of the analyzed cloud masks for an exemplary visual comparison. The 6th of June 2004, 12:00 UTC has been chosen as a representative example. For this date, figure 5-8, figure 5-10, figure 5-12, and figure 5-14 show a quasi true color composite (TCC) image, an inverted and histogram equalized BT_{108} image, the corresponding $FUB_{ACSBTE(night,twilight)}$, and the $EUMETSAT_{CLA}$ cloud masking results, respectively. These images display huge parts of SEVIRI's full field of view. In the following, these images will be called "full disk"



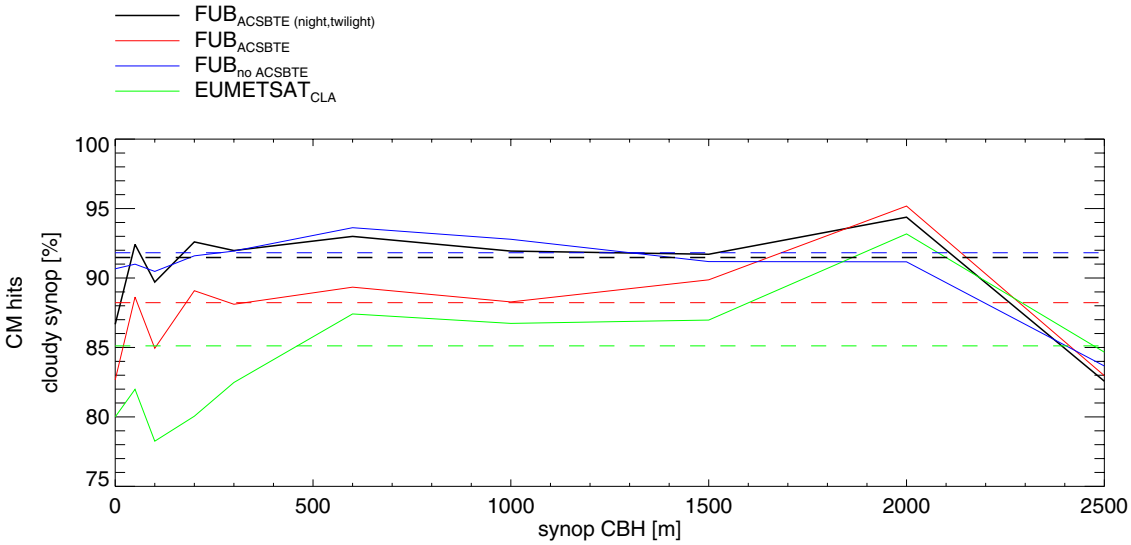


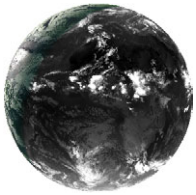
Figure 5-7: Influence of cloud base height from synop observations on the probability of cloud detection analog to figure 5-5 but for the short-term validation phase.

images. Corresponding views zoomed on the European region are given in figure 5-9, figure 5-11, figure 5-13, and figure 5-15. Most of all used synop stations are located in this region.

The full disk and the zoomed images are alternately arranged to enable easy comparison between the cloud masks, the BT_{108}, and the TCC images simply by turning the pages.

The zoomed images additionally show all analyzed synop reports of this point in time with unambiguous cloud coverage status. Triangles represent synop observation with 7 or 8 octas cloud coverage, while the squares indicate observations with 1 or 0 octas cloud coverage. The symbols are centered on the synop stations. In the cloud mask images, green symbols represent successful cloud masking while red symbols represent the opposite.

Some prominent features are highlighted in both TCC images (figure 5-8 and figure 5-9) and will be discussed in the following:



Full disk

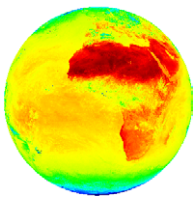
In region b) the actual status of cloud coverage is not apparent from the TCC image due to the high reflectance of xeric surfaces. In this region, the BT_{108} image shows cloud features that seem to be overestimated by the EUMETSAT algorithm.

The clouds above sea in region c) are optically very thin as they cannot be observed in the TCC image but only in the BT_{108} image. These clouds are overestimated by the EUMETSAT but underestimated by the FUB algorithm. The very high and thin cirrus clouds in region e) are also less pronounced by the FUB algorithm. However, as shown in some of the following examples, the underestimation of thin clouds is not a general characteristic of the FUB algorithm.

In accordance with section 5. 3. 3, low level clouds above relatively warm surfaces are often underestimated by the EUMETSAT algorithm, exemplarily highlighted by a) and f). Clouds of this type are also problematic for the FUB_{ACSBTE} algorithm.

Illustrated in region g), some low level clouds above cold sea are underestimated by the FUB algorithm. The sharp edge of the area erroneously classified as cloud free (near -50° N) is caused by switching to the twilight algorithm at low sun angles. Information from the channels in the visible spectral region is not available for this algorithm, additionally the contrast between BT_{ACSBTE} and BT_{108} of the low level clouds is very low. Unsolvable ambiguities are the result.

In region d) the EUMETSAT and the FUB algorithm produce different results. This might be caused by the high aerosol concentration in this region. The separation of clouds from aerosols in this case is ambiguous, so that the rating of the algorithms is not possible. Nevertheless, the patterns visible in the TCC image are better represented in the FUB cloud mask.



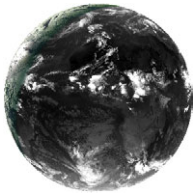
Europe

In some cases, low level clouds above sea surfaces are misinterpreted by both algorithms. Region a) and d) show a distinct underestimation by the EUMETSAT algorithm. An example for underestimation of those clouds by the FUB algorithm is given by region c). In all of these cases the particular algorithm seems to underestimate the information from the channels in the visible spectral region.

Region e) shows that the EUMETSAT algorithm is possibly more sensitive to a misaligned land/sea mask.

In respect to the illustrated synop observations, region b) is obviously the region with most differences between both algorithms. In this region the EUMETSAT algorithm often classifies optically thin clouds erroneously as cloud free. Detailed investigation by means of the channel combinations introduced in section 4. 2 suggests, that the missed clouds are optically thin, low or mid level water clouds. This hypothesis is verified by the corresponding synop reports with cloud base heights between $300\,m$ and $1500\,m$. It should be emphasized that these are definitely no cases with sub pixel cloud fraction. As mentioned before, only unambiguous synop reports have been analyzed for this comparison.

Overall, in this example it is highly visible that cloudy areas are represented less expanded in the EUMETSAT cloud mask image than in the corresponding FUB cloud mask image. It is suggested that the EUMETSAT algorithm is in average less sensitive to low and mid level clouds with low total water path. Therefore, fractional sub pixel cloud coverage of this type is also underestimated.



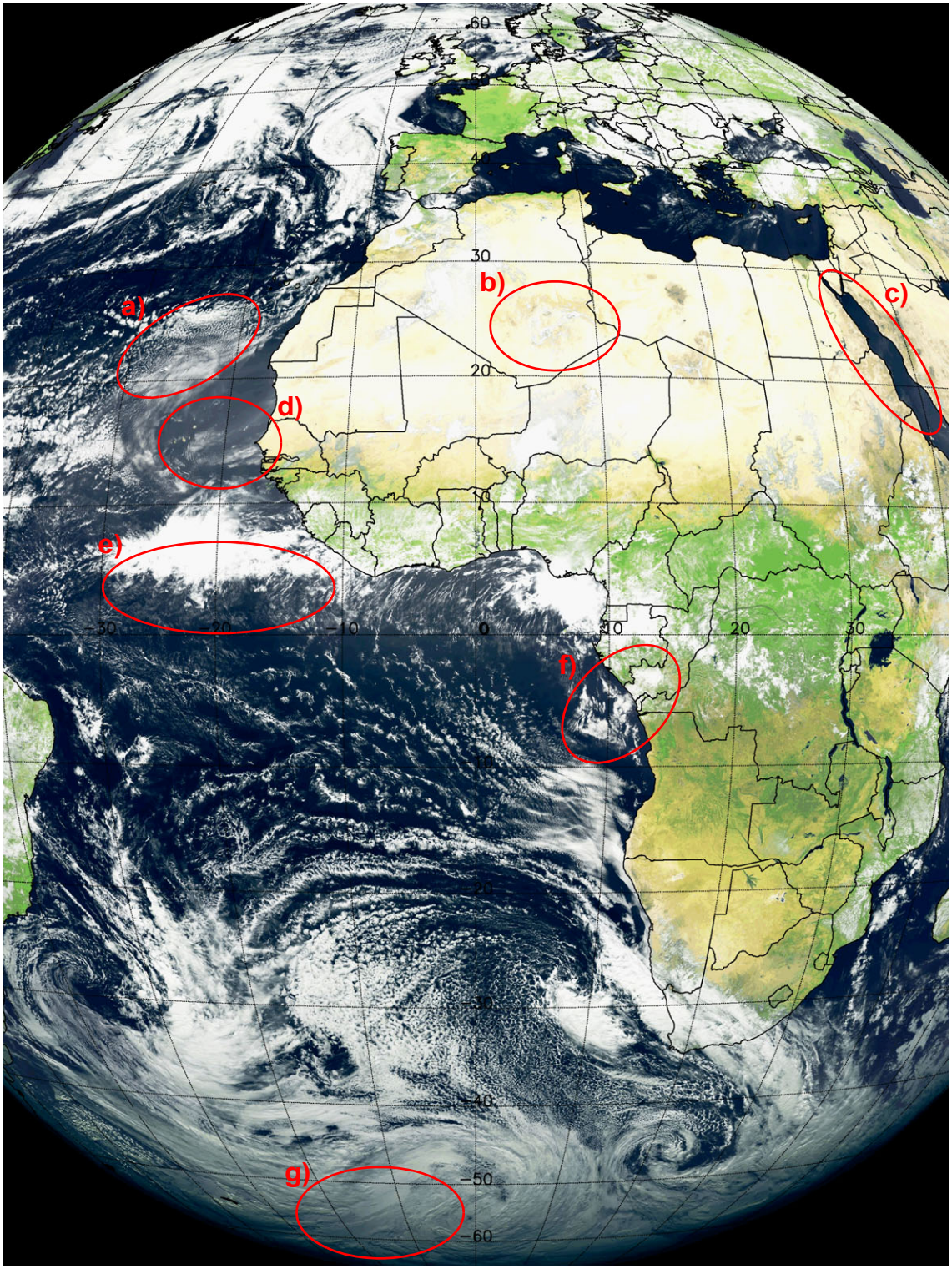
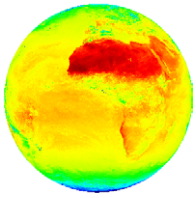


Figure 5-8: Full disk, true color composite, June, 6th 2004, 12:00 UTC.



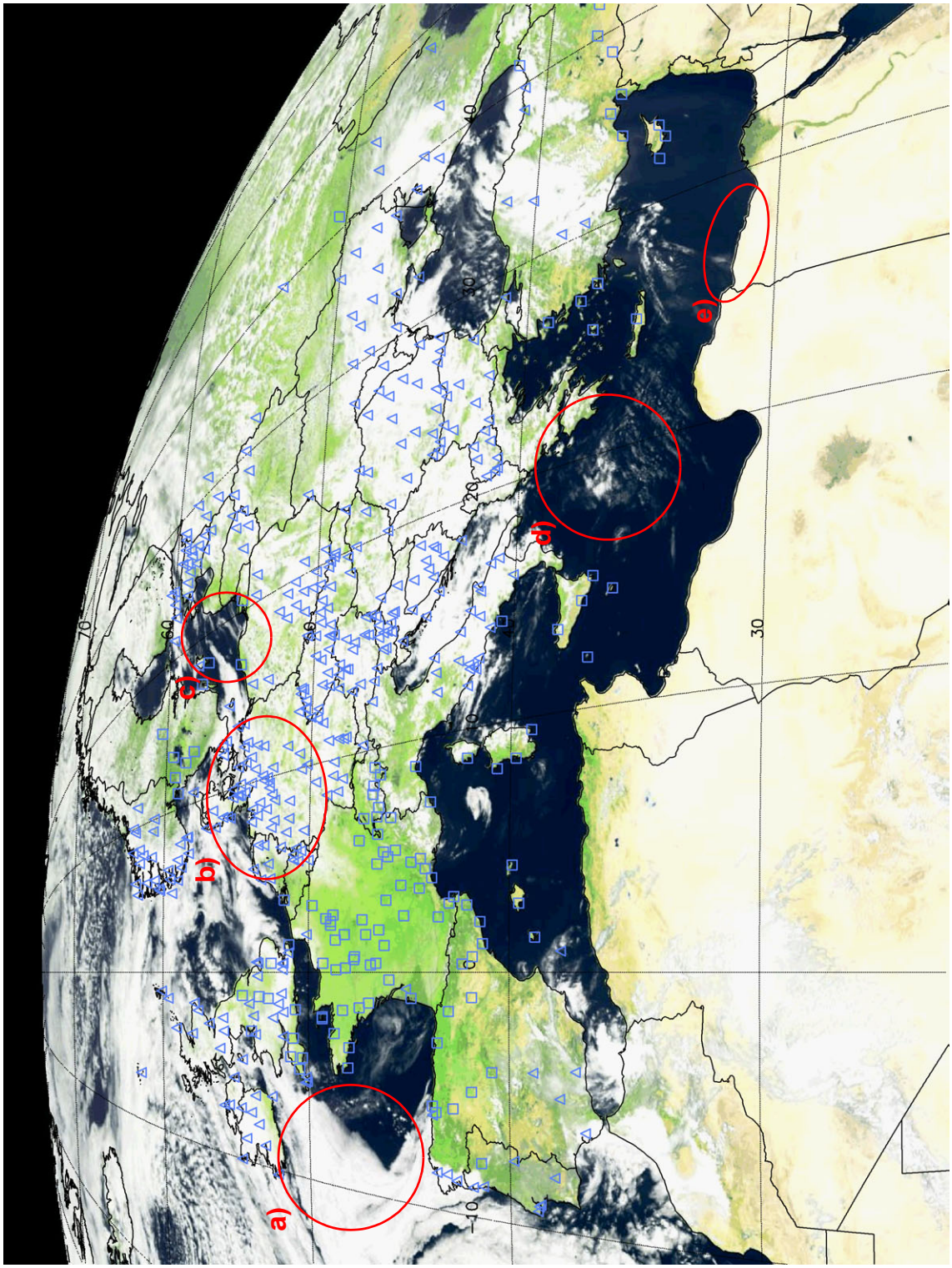
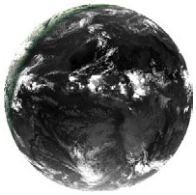


Figure 5-9: Europe, true color composite, June, 6th 2004, 12:00 UTC.



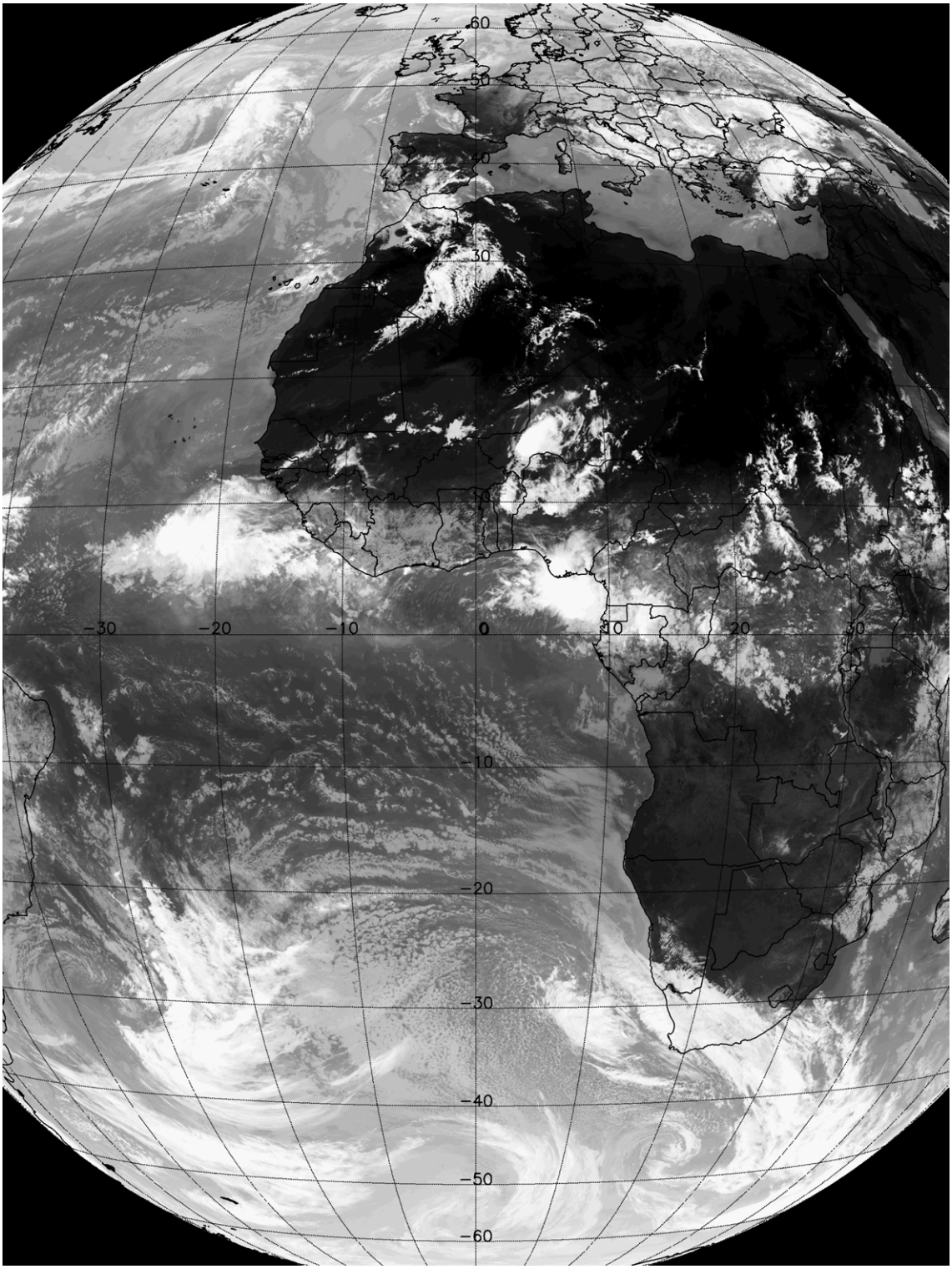
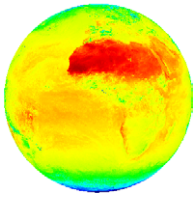


Figure 5-10: Full disk, BT_{108} (histogram enhanced), June, 6th 2004, 12:00 UTC.



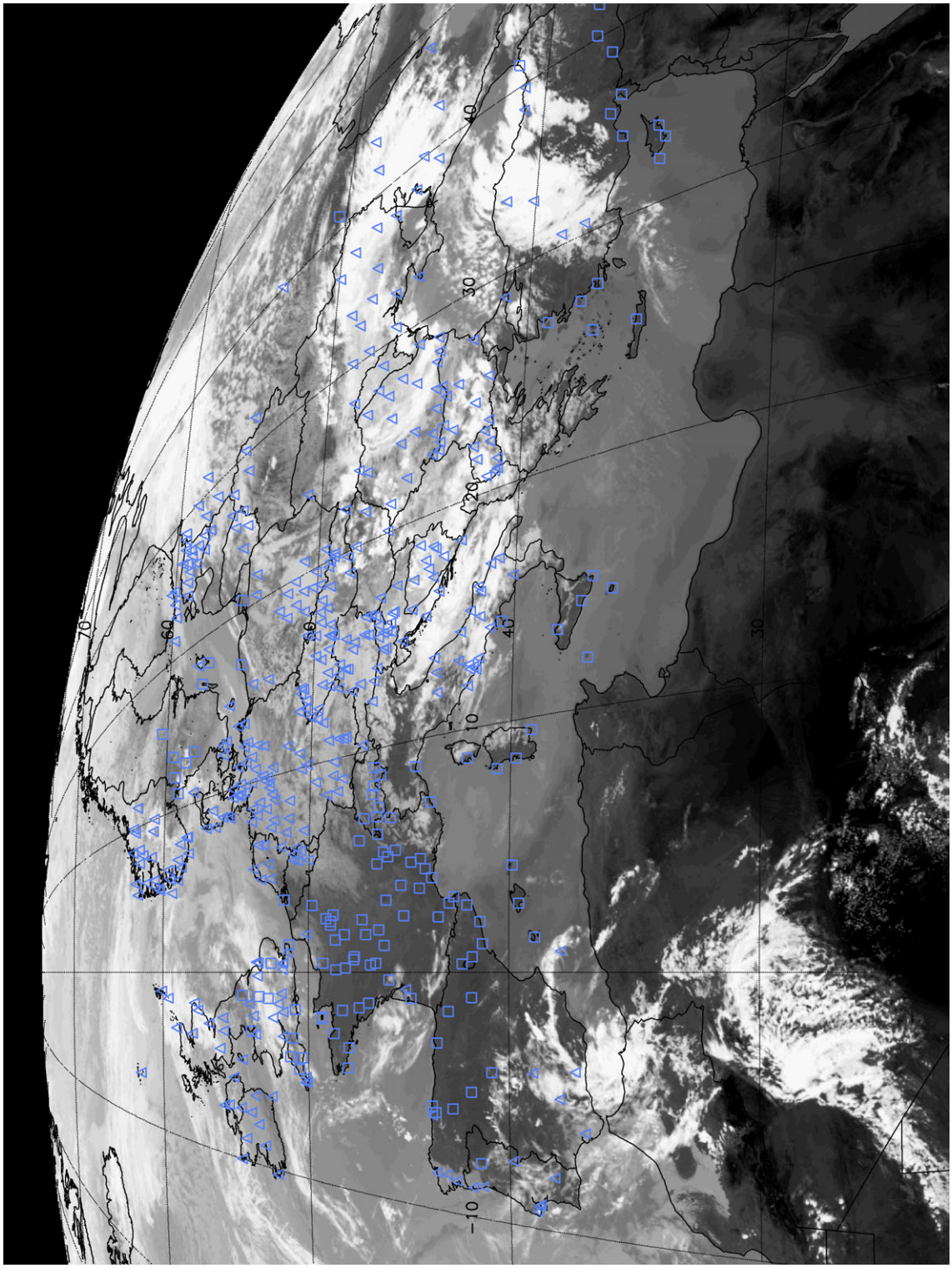
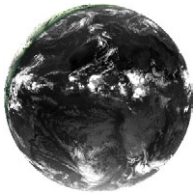


Figure 5-11: Europe, BT_{108} (histogram enhanced), June, 6th 2004, 12:00 UTC.



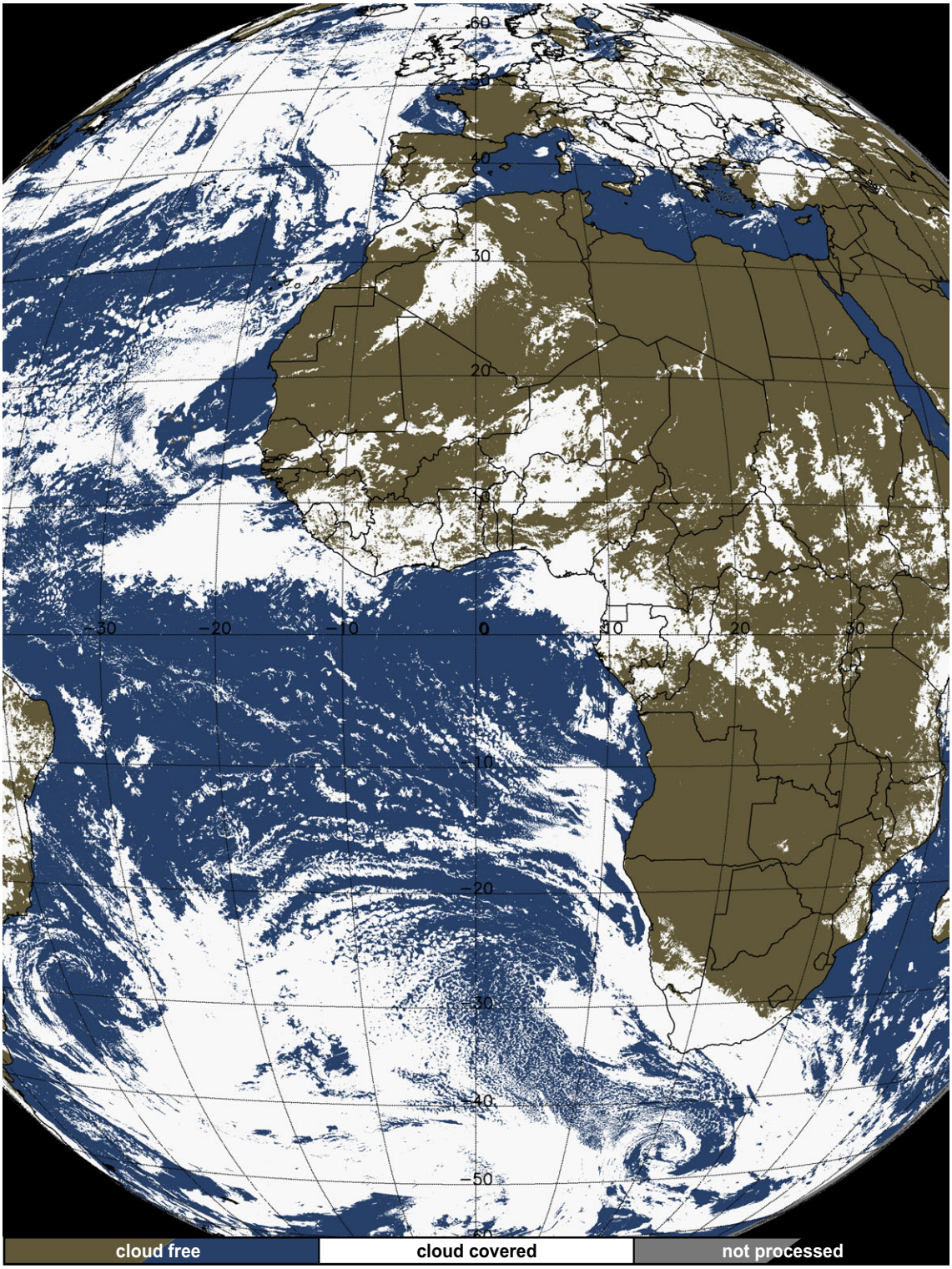
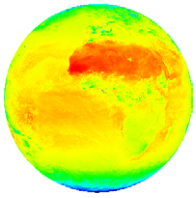


Figure 5-12: Full disk, $FUB_{ACSBTE\,(night,\,twilight)}$ cloud mask, June, 6th 2004, 12:00 UTC.



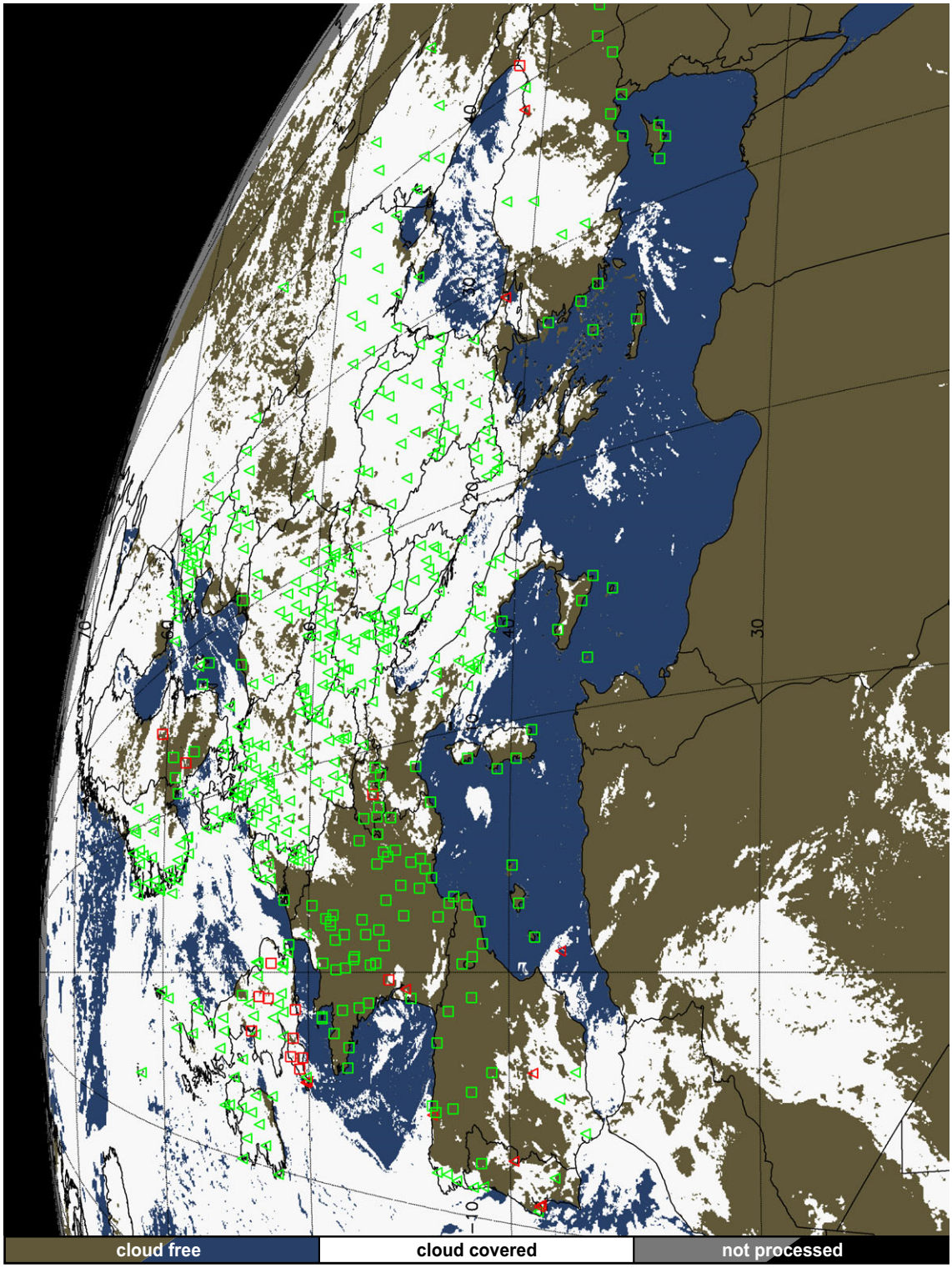
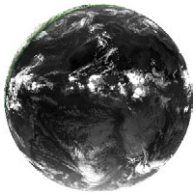


Figure 5-13: Europe, $FUB_{ACSBTE(night,twilight)}$ cloud mask, June, 6th 2004, 12:00 UTC.



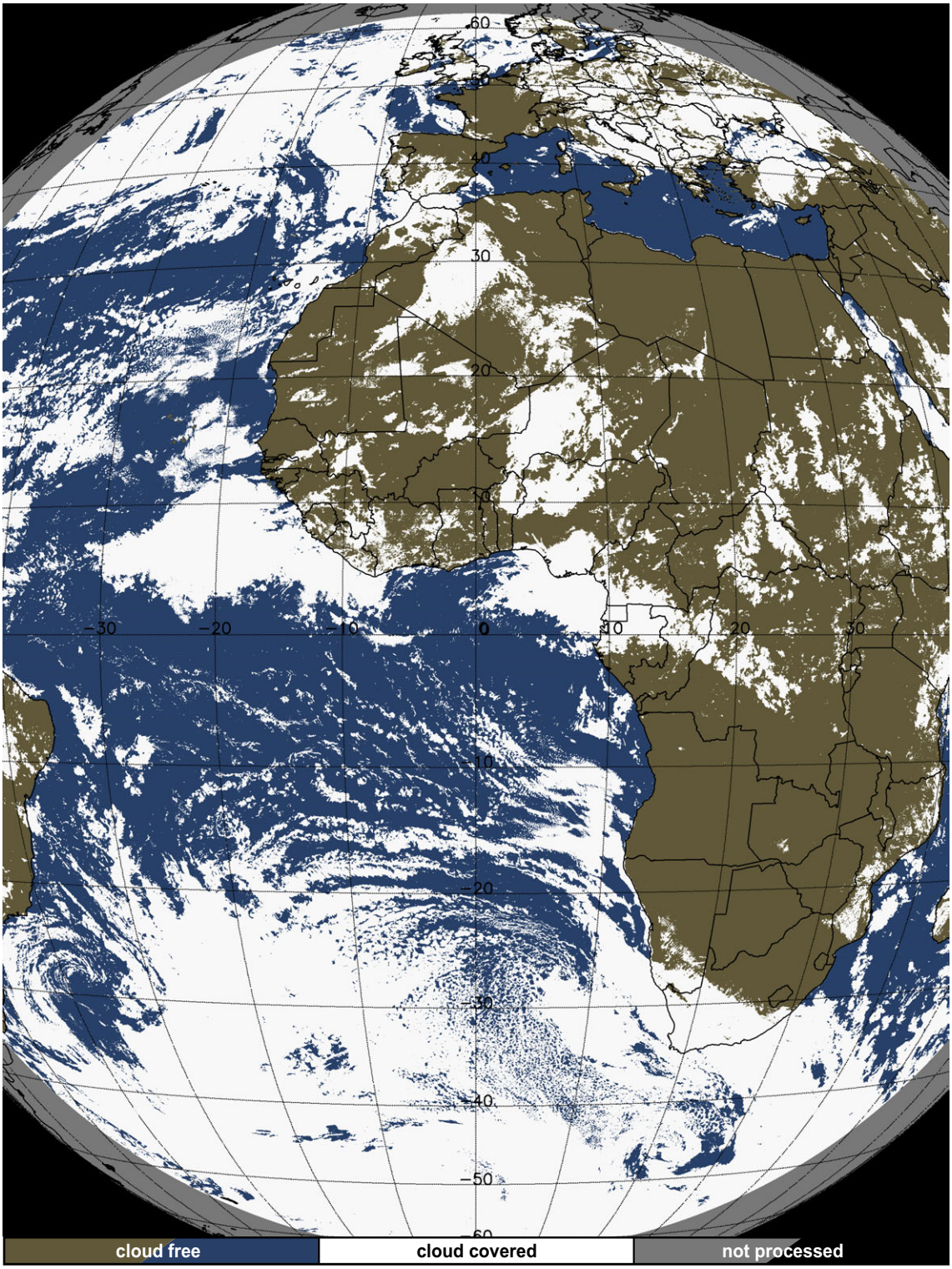
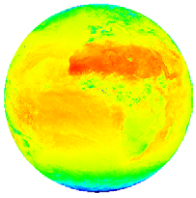


Figure 5-14: Full disk, $EUMETSAT_{CLA}$ cloud mask, June, 6th 2004, 12:00 UTC.



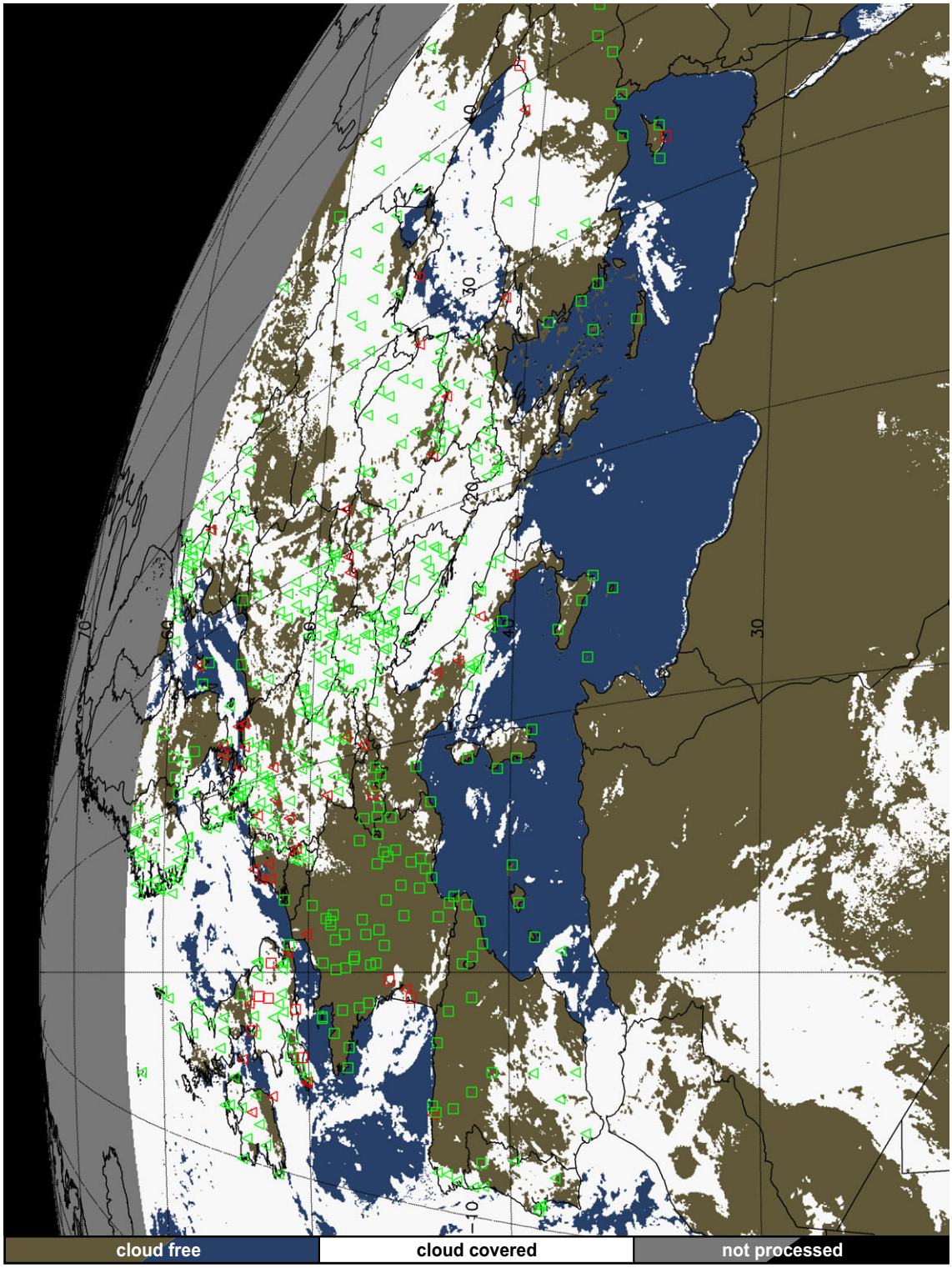


Figure 5-15: Europe, *EUMETSAT*$_{CLA}$ cloud mask, June, 6th 2004, 12:00 UTC.