An Automatic Cloud Mask Algorithm Based on Time Series of MODIS Measurements

A. Lyapustin, Y. Wang
University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, MD
R. Frey
Cooperative Institute for Meteorological Satellite Studies (SSEC), University of Wisconsin, Madison, WI

Popular Summary

Quality of aerosol retrievals and atmospheric correction depends strongly on accuracy of the cloud mask (CM) algorithm. The heritage CM algorithms developed for AVHRR and MODIS use the latest sensor measurements of spectral reflectance and brightness temperature and perform processing at the pixel level. The algorithms are threshold-based and empirically tuned. They don't explicitly address the classical problem of cloud search, wherein the baseline clear-skies scene is defined for comparison. Here, we report on a new CM algorithm which explicitly builds and maintains a reference clear-skies image of the surface (refcm) using a time series of MODIS measurements. The new algorithm, developed as part of the Multi-Angle Implementation of Atmospheric Correction (MAIAC) algorithm for MODIS, relies on fact that clear-skies images of the same surface area have a common textural pattern, defined by the surface topography, boundaries of rivers and lakes, distribution of soils and vegetation etc. This pattern changes slowly given the daily rate of global Earth observations, whereas clouds introduce high-frequency random disturbances. Under clear skies, consecutive gridded images of the same surface area have a high covariance, whereas in presence of clouds covariance is usually low. This idea is central to initialization of refcm which is used to derive cloud mask in combination with spectral and brightness temperature tests. The refcm is continuously updated with the latest clear-skies MODIS measurements, thus adapting to seasonal and rapid surface changes. The algorithm is enhanced by an internal dynamic land-water-snow classification coupled with a surface change mask. An initial comparison shows that the new algorithm offers the potential to perform better than the MODIS MOD35 cloud mask in situations where the land surface is changing rapidly, and over Earth regions covered by snow and ice.

An Automatic Cloud Mask Algorithm Based on Time Series of MODIS Measurements

A. Lyapustin, Y. Wang

University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, MD

R. Frey

Cooperative Institute for Meteorological Satellite Studies (SSEC), University of Wisconsin, Madison, WI

Abstract. Quality of aerosol retrievals and atmospheric correction depends strongly on accuracy of the cloud mask (CM) algorithm. The heritage CM algorithms developed for AVHRR and MODIS use the latest sensor measurements of spectral reflectance and brightness temperature and perform processing at the pixel level. The algorithms are threshold-based and empirically tuned. They don't explicitly address the classical problem of cloud search, wherein the baseline clear-skies scene is defined for comparison. Here, we report on a new CM algorithm which explicitly builds and maintains a reference clear-skies image of the surface (*refcm*) using a time series of MODIS measurements. The new algorithm, developed as part of the Multi-Angle Implementation of Atmospheric Correction (*MAIAC*) algorithm for MODIS, relies on fact that clear-skies *images* of the same surface area have a common textural pattern, defined by the surface topography, boundaries of rivers and lakes, distribution of soils and vegetation etc. This pattern changes slowly given the daily rate of global Earth observations, whereas clouds introduce high-frequency random disturbances. Under clear skies, consecutive gridded images of the same surface area have a high covariance, whereas in presence of clouds covariance is usually low. This idea is central to initialization of *refcm* which is used to derive cloud mask in combination with spectral and brightness temperature tests. The *refcm* is continuously updated with the latest clear-skies MODIS measurements, thus adapting to seasonal and rapid surface changes. The algorithm is enhanced by an internal dynamic land-water-snow classification coupled with a surface change mask. An initial comparison shows that the new algorithm offers the potential to perform better than the MODIS MOD35 cloud mask in situations where the land surface is changing rapidly, and over Earth regions covered by snow and ice.

1. Introduction

Cloud mask (CM) is a primary science algorithm that precedes detailed analysis of cloud, aerosol and land surface/ocean parameters from the global observing space borne sensors. Accuracy of cloud detection has a significant impact on aerosol retrievals and atmospheric correction. At the global scale, undetected clouds introduce a positive bias in aerosol concentration and increase land albedo, whereas regional and seasonal biases correlated with cloudiness affect spatial distribution and temporal changes of these parameters. Given strong interdependence between human induced land changes, aerosols and clouds, accurate cloud identification has a growing importance in our understanding of the Earth system, and the role of anthropogenic factors in the modern climate and its changes.

The heritage cloud mask algorithms for the low-orbiting sensors, including the AVHRR CLAVR [*McClain*, 1993] and MODIS algorithm [*Ackerman et al.*, 1998; 2006], use the *latest* sensor measurements of spectral reflectance and brightness temperature and perform processing at the *pixel* level. In these cases, the algorithms lack apriori knowledge about the surface and have to deal with large uncertainties caused by wide natural variability of both land surface and clouds.

Even identification of clear pixels in cloud-free conditions is challenging when measurements do not exhibit explicit signatures of the surface, such as spectral signatures of dense vegetation or of deep water. Due to tremendous variability of clouds, their detection has always been problematic over brighter surfaces, especially over snow. Because of similarity of spectral reflectance between the snow and snow/ice clouds, and temperature inversions frequent in the low troposphere in wintertime, no particular set of spectral reflectance - brightness temperature tests may guarantee a success over snow and ice.

On the other hand, contemporary polar-orbit global observation imagers, such as AVHRR, MODIS, or future VIIRS, provide or will provide a daily global view of the Earth at equator with multiple observations per day for the mid-latitude and polar regions. This opens a possibility to add temporal and spatial dimensions into the cloud mask processing. For example, with a high frequency of observations, the land surface can be considered as a static or slowly changing background contrary to ephemeral clouds, which offers a reliable way of developing the "comparison target" for the CM algorithm. The modern memory and processing power of computers allow deriving and storing a sufficient volume of ancillary information about every surface pixel in order to optimize performance of cloud masking globally. An early example of such an approach is the ISCCP CM algorithm [*Rossow and Garder*, 1993] developed for geostationary platforms. It builds the clear-skies composite map from the *previous* measurements and infers CM for every pixel by comparing current measurement with the clear-skies reference value. The uncertainty of the reference value, caused by the natural variability and sensor noise, is directly calculated from the measurements.

We present a next step in evolution of this idea – a new CM algorithm working with the time series of MODIS measurements, and developed as a part of the Multi-Angle Implementation of Atmospheric Correction (*MAIAC*) algorithm [*Lyapustin and Wang*, 2007a, b]. It uses covariance analysis to build a reference clear skies image (*refcm*) and to accumulate a certain level of knowledge about the surface and its variability thus constructing a rather comprehensive comparison target for the cloud masking. The new algorithm has an internal surface classifier, producing dynamic land-water-snow (*LWS*) mask, and a surface change mask. These are an integral part of *MAIAC* guiding both cloud masking and further aerosol-surface reflectance retrievals when the surface changes rapidly as a result of fires, floods or snow fall/ablation. The cloud mask generated by the CM algorithm is updated during atmospheric correction, which makes it a synergistic component of *MAIAC*. The atmospheric correction uses a time series of MODIS measurements, acquired at different angles, to retrieve a surface bi-directional reflectance factor (BRF). It requires consistency between the latest retrieved BRF and the previous solution unless land surface change has been detected. This complex approach is found capable of catching sub-pixel and semi-transparent clouds that are difficult to detect otherwise.

Utilizing the time series in operational processing of remote sensing data is a novel concept although prototypes have been in existence, such as the MODIS BRF/albedo algorithm based on accumulation of 16-day time series of gridded MODIS surface reflectance data [*Schaaf et al.*, 2002]. In the *MAIAC* algorithm, use of the time series of MODIS data is key to all major components of processing, including cloud mask, joint aerosol-surface retrievals, and atmospheric correction. In order to arrange a time series processing, *MAIAC* first grids MODIS level 1B (L1B) calibrated and geolocated data to a regular 1 km grid [*Wang and Lyapustin*, 2007], splits them into 600 km *Tiles* and places the new *Tile* in the Queue, which can hold up to 16 days of previous measurements. The size of Tile is scaled with the operational memory of our workstation, and it

can be set to the value of 1000 km used in the MODIS operational processing. Further *MAIAC* processing uses both individual grid cells, interchangeably called pixels here, and fixed-size (25×25 km^2) areas, called blocks. In order to organize such processing, we developed a framework of C++ classes and structures (algorithm-specific Containers) to store *Tiles* of MODIS data and dynamically update the Queue. The class functions are designed to easily handle processing in the time-space coordinates, for example at the pixel- *vs* block-level, and for a single (last) day of measurements *vs* for all available days in the Queue, or for a subset of days passing certain filters. The data storage in the Queue is organized in a very efficient way using pointers, which avoids physically moving the previous data in memory when the new data arrive.

A section 2 describes the algorithm constructing the reference clear-skies image (*refcm*). Section 3 and Appendix B present an overall decision logic in cloud masking and provides the necessary technical detail. A dynamic surface classification algorithm is discussed in section A. Examples of the *MAIAC* cloud mask for 50×50 km^2 subsets of MODIS TERRA data for different global locations are given in section 4. This section also shows several examples of a large-scale comparison of *MAIAC* CM with the cloud mask of the operational MODIS algorithm (MOD35). It should be mentioned that this paper describes the results of the *MAIAC* CM algorithm alone, not enhanced by additional filtering and cloud shadow detection of the atmospheric correction algorithm, which will be described separately. The paper concludes with a Summary.

2. Building Reference Clear Skies Image

The clear-skies images of the same surface area have a common textural pattern, defined by the surface topography, boundaries of rivers and lakes, distribution of soils and vegetation etc. This pattern changes slowly compared with the daily rate of global Earth observations. Clouds randomly change this pattern, which can be easily detected by covariance analysis. The covariance is a metric showing how well the two images X and Y correlate over an area of $N \times N$ pixels,

$$\mathrm{cov} = \frac{1}{N^2}\sum_{i,j=1}^{N}\frac{(x_{ij}-\bar{x})(y_{ij}-\bar{y})}{\sigma_x\sigma_y},\ \sigma_x^2 = \frac{1}{N^2}\sum_{i,j=1}^{N}(x_{ij}-\bar{x})^2 .$$

A high covariance of two images usually implies cloud-free conditions in both images, whereas low covariance usually indicates presence of clouds at least in one of the images. A rapid surface change or significant variation of aerosol density in the area may also reduce covariance. Because covariance removes the average component of the signals, this metric is equally successful over the dark and bright surfaces and in both clear and hazy conditions if the surface spatial variability is still detectable from space.

The core of the *MAIAC* CM algorithm is initialization and regular update of the reference clear-skies image for every block. The *refcm* is initially built from the pair of images for which covariance is high, and caution is exercised to exclude correlated cloudy fields. The algorithm calculates a block-level covariance between the *new Tile* and the previous *Tiles*, moving backwards in the Queue until either the "head" of Queue is reached, or the clear conditions are found. The latter correspond to high covariance ($cov \geq 0.68$) and low brightness temperature contrast in the block for both days, $\Delta BT = BT_{max} - BT_{min} < \Delta_1$. The initial value of threshold Δ_1 is currently defined as $\Delta_1 = 7$ K for pure water blocks, and $\Delta_1 = 25 + dT(h)$ K for pure land or mixed land-water, land-snow, or land-water-snow blocks. Factor dT(h) accounts for the surface height variations in the block, and is defined for an average lapse rate, $dT(h) = 0.0045(h_{max} - h_{min})$, where *h*

(km) is surface height over the sea level. Once *refcm* is initialized, the algorithm begins to use the block-specific value of BT contrast.

When snow is detected, the high covariance threshold is slightly reduced from 0.68 to 0.65 in order to accommodate for a rapid change of snow cover at mid-latitudes, where snow accumulation is usually low and day-to-day variations of reflectance during wintertime can be enormous.

After initialization, the algorithm uses *refcm* to compute covariance with the latest measurements. Once the clear conditions are found, *refcm* is updated. With this dynamic update, the *refcm* adapts to the gradual landcover changes related to the seasonal cycle of vegetation. The rapid surface change events (e.g. snowfall/ablation) are handled through repetitive re-initialization which is performed each time when covariance of the latest *Tile* with *refcm* is found to be low (see section 3).

2.1 Algorithm Technical Detail

The *MAIAC* CM algorithm uses seven 500 m resolution MODIS bands B1-B7 and a 1 km thermal band B31 (Table 1). As mentioned earlier, data are initially re-projected and gridded to 1 km grid.

1. The covariance analysis is currently performed for MODIS band B1 (0.645 μm). We have extensively studied the use of B6 (1.629 μm) and B5 (1.242 μm). B6 was initially our band of choice because it has low molecular absorption and aerosol extinction, as well as high land surface variability and spatial contrasts. We have conducted an independent covariance analysis for 1 year of MODIS TERRA data subsetted for 50×50 km^2 areas around 156 AERONET [*Holben et al.*, 1998] locations globally. From considered MODIS bands B2 (0.856 μm), B5, B6 and B7 (2.113 μm), band B6 provided by about 10% more of the high covariance cases. On the other hand, this band has not been working properly on MODIS AQUA, and it has developed occasional problems since 2006 on MODIS TERRA. We have also found that bands B5 and B6 often cannot detect significant spatial aerosol variability or variable semi-transparent clouds. Spatial uniformity of aerosols within the block area is one of the major requirements of *MAIAC* inversion method, and B1 was found to provide a better overall performance.

2. The size of a block is currently selected as 25×25 pixels (km^2) for two reasons. First, this size is large enough to capture a variety of spatial variability scales (geologic, topographic, ecologic etc.) required for covariance analysis. Second, it is sufficiently large to capture surface variability at the edge of scan where the MODIS pixel size grows to ≈2×4 km^2 for 1 km^2 nadir pixels. On the other hand, the success rate of covariance algorithm to select clear blocks in conditions of broken cloudiness is higher for smaller blocks. The MISR CM algorithm [*Diner et al.*, 1999], which extensively uses covariance analysis, works with the block size of 17.6 km. We plan to evaluate the global performance of cloud mask for smaller blocks 15-25 km, and select the optimum for operational application.

3. In order to account for the effects related to the scan angle variation, e.g. pixel size growth, surface BRF effect or reduction of contrast at higher view zenith angles (VZA), two reference clear-skies images are maintained by the algorithm, *refcm1* for VZA≈0-45° and *refcm2* for VZA=45°-60°.

4. In addition to the clear-skies image, the *refcm* structure stores the maximal value ($refcm.r1_{max}$) and the variance ($refcm.\sigma$) of reflectance for the block. The algorithm has a separate allocated storage at the Queue level for some parameters of every pixel and block from the previous data or retrievals. One such parameter is the brightness temperature contrast ($q.\Delta BT=BT_{max}-BT_{min}$) stored for every block. MODIS data show that the BT contrast is a rather stable metric of a given land area in clear conditions. For the pure land blocks, containing no water or snow pixels, the BT contrast is usually low (1-6 K) for flat terrains at MODIS 1 km resolution of thermal bands. It may increase, sometimes significantly (10-20 K) when the block is a mixture of land, water or snow pixels. In partially cloudy conditions the BT contrast increases because BT_{min} is usually lower over clouds.

5. The algorithm keeps a two-level cloud mask, the standard mask at the grid (1 km) resolution (CM), and another one at the block resolution (CM_COV). The CM_COV mask is used to efficiently control the algorithm flow for *refcm* re-initialization, and during aerosol retrievals and atmospheric correction.

6. The allowed values of cloud mask are clear (CM_CLEAR, CM_CLEAR_WATER, CM_CLEAR_SNOW), indicating surface type, possibly cloudy (CM_PCLOUD), and confidently cloudy (CM_CLOUD). Two more values of cloud mask are CM_SHADOW for pixels defined as a cloud shadow, and CM_GREY representing 1 dilated pixel on cloud edges, which is used for aerosol retrievals and atmospheric correction. The covariance component of our algorithm, which offers a direct way to identify clear conditions, renders another commonly used value of cloud mask – "possibly clear" – redundant.

7. A high covariance between two days alone does not guarantee clear conditions. Clouds "leak" into *refcm* in a number of different and often unimaginable ways, especially over snow. From an extensive analysis of MODIS data over the world, from Amazon region to Greenland, we designed a set of filters that achieve a rather satisfactory selection of clear conditions:

Prior to calculating covariance between the last Tile (L) and the previous Tiles (k=L-1, … 0) stored in the Queue, the following conditions are used to reject any of the Tiles k for a given block:

- High BT contrast, which usually indicates clouds: $\Delta BT_k > \min(q.\Delta BT+15, 25)$ (K).
- Small time difference between observations: $t_L - t_k < 200$ min. This test was introduced to exclude correlation of the same cloudy fields on stagnant days with zero or very low winds.
- The number of pixels which passed the snow test is significantly different between observations L and k.

When no snow is detected by the snow test in both Tiles L and k and the calculated covariance is high, the following two tests also serve as a rejection condition:

- The difference of the block-average TOA reflectances should not be too high: $\left|\langle r1_L\rangle - \langle r1_k\rangle\right| > 0.18$. This test filters infrequent cases when bright continuous clouds correlate spatially with much darker clear scene.
- The maximal reflectance of the block should not significantly exceed the *refcm* value:
  $r1_{max,\,L} > refcm.r1_{max}+0.1$ OR $r1_{max,\,k} > refcm.r1_{max}+0.1$.
  If $refcm.r1_{max}$ is not defined, as during *refcm* initialization, the condition states that both Tiles should not be simultaneously bright ($r1_{max,\,L,k} >0.2$) and cold ($BT_{min,\,L,k}<277$K), which is indicative of clouds.

- Overcast conditions with snow/ice clouds producing high covariance are ubiquitous over the tropical regions and at northern latitudes, especially in winter. These are filtered by our snow classification algorithm (see Appendix A) which requires high covariance between observations and a significant number of pixels passing snow test on both days in order to mask a pixel as snow for the first time. To avoid snow detection over tropics during the rain season according to this logic, we use common sense that a relatively low temperature should establish for some period of time before snowfall. On the contrary, if the surface was warm (BT_{max}>288K) during any of the previous clear or partially clear days, when the ground brightness temperature (BT_G) can be established, then most probably we have the case of spatially correlated snow/ice clouds. The 16-day time series of measurements usually provides an assessment of BT_G for any given block from previous days.

- Finally, we are using a "sigma" test to filter Tiles with significantly higher or lower spatial variability than that of a clear-skies surface:

$$|\sigma_L - refcm.\sigma| > \Delta\sigma \text{ OR } |\sigma_k - refcm.\sigma| > \Delta\sigma, \text{ where } \Delta\sigma = \max(refcm.\sigma, 0.015).$$

This test is particularly useful over very homogeneous snow-covered regions lacking terrain features, such as inner regions of the Greenland ice sheet, where clouds usually have significantly higher amplitude of spatial variation.

According to testing performed, the developed set of filters ensures a reliable initialization or re-initialization of the reference clear-skies image of the surface globally with the low percentage of errors. However, this set of filters will continue to be revised as new exceptions are found.

3. CM Algorithm Flowcharts

The central idea of *MAIAC* CM algorithm is to use the reference clear-skies image of the surface (*refcm*) and covariance analysis to identify clear skies and cloudy blocks, which are usually characterized by high and low covariance, respectively. The low covariance of the latest Tile with *refcm* may be caused by presence of clouds, dense inhomogeneous aerosols, or a rapid surface change on a scale comparable to the block size. For regular surfaces, not covered by snow, cloud detection in the low-covariance blocks is based on a simple postulate that clouds are usually colder and brighter than the surface. The reference surface reflectance for every pixel is provided by the *refcm* clear-skies image, whereas an estimate of the BT_G comes either from the clear land pixels detected by spectral tests for a given block, or from the cloud-free neighbor blocks, identified by the high covariance. Cases of surface change are processed with the help of Land-Water-Snow mask (*mask_LWS*) and land change mask (*mask_Change*) derived dynamically by an internal surface classifier and stored in the Queue memory.

The general flowchart of CM algorithm is shown in Figure 1. Here, rectangles represent separate functions, diamond shapes stand for the separate subroutines (algorithms), and round-corner rectangles indicate decision (branching) points. The thick arrows show the points of exit.

1. As a first step, spectral tests are performed for each block to detect clear vegetated, water and possibly snow pixels as described in Section 4. These tests are used to initialize and update pixel-level surface classification and change masks (*mask_LWS*, *mask_Change*). The detected clear vegetated and water pixels are used to calculate brightness temperature of the ground and of the water for the block (BT_G, BT_W), which are required in further analysis.

2. Next, a *Cold/Thick cloud test* is performed for each pixel of the Tile. The *Cold* test is designed to find clouds which are rather cold as compared to either the BT_G if the latter is available, or to the maximal BT found among a given block and its four closest neighbors. This test is formulated as follows:

$$r1 > 0.15 \text{ AND } BT < BT_0-\Delta, \Rightarrow \text{CM\_CLOUD.}$$

where $BT_0 = BT_G$, Δ=15K if BT_G is defined, or $BT_0 = BT_{max}$, Δ=20K if BT_G is undefined. This is a relative test, which does not use an absolute BT threshold, and thus can be used with the full range of temperatures from the poles to the equator. The threshold Δ is selected to be high enough to avoid a possible error due to temperature inversions in the boundary layer.

The *Thick* cloud test is designed to detect spectrally neutral bright and thick clouds. Usually, reflectance of such clouds decreases slowly with wavelength within spectral range of 0.4-1.25 μm. The test is formulated as:

$$r1 > 0.2 \text{ AND } r3 > r1 \text{ AND } r1 > r5, \Rightarrow \text{CM\_CLOUD}$$

where MODIS bands 3 and 5 represent wavelengths of 0.47 μm and 1.24 μm.

If the pixel passed the *snow* test, neither of the described cloud tests is used, and the algorithm delays decision till the covariance analysis, which may confirm cloud-free conditions and snow on the surface.

The cloud mask, generated by these tests, can be overwritten later if high covariance is found.

3. Further processing path depends on whether *refcm* has been initialized. If not, then the algorithm tries to initialize *refcm* as described earlier in sec. 2 (module *initRefcm*). During initialization, the algorithm consecutively computes covariance of the last measurements with each of the earlier images of the Queue, until the high covariance is found and clear conditions in both images are confirmed. If initialization is unsuccessful because of clouds, the algorithm runs a backup pixel-level algorithm *cloudMask1*.

If *refcm* was initialized earlier, then the algorithm calculates the covariance between the new measurements and the *refcm* image, and performs further processing depending on the value of covariance (modules *CM_highCov* and *CM_lowCov*). The algorithm pursues a conservative strategy admitting that sometimes partially or completely cloudy images can bypass filters of *initRefcm*() and update *refcm*. If this happens, the latest cloud-free image will not correlate well with *refcm*. To reduce such possible errors, our algorithm attempts to re-initialize *refcm* with latest measurements each time the calculated covariance with existing *refcm* is found low. This approach also helps to re-initialize *refcm* when the surface reflectance changes rapidly.

Each of modules *CM_highCov*, *CM_lowCov*, *cloudMask1* is producing cloud mask for the new tile. Module *initRefcm* produces cloud mask only if it finds high covariance. The *refcm* image and block-parameters of the *refcm* structure are updated in the modules *initRefcm* and *CM_highCov*. The surface classification scheme and algorithms implemented in modules *CM_highCov* and *CM_lowCov* are described in the Appendix.

4. Performance of *MAIAC* CM Algorithm

Performance of the cloud mask algorithm has been tested using 1 year (2003) of MODIS TERRA data subsetted for 156 AERONET locations worldwide. The data were received in the swath format with resolution aggregated to 1 km in all bands for the subset area of 50×50 km^2. Before

processing, we re-projected and gridded these data to 1 km grid using the nearest neighbor method.

The analysis was conducted by a visual comparison of the generated cloud mask against the true color MODIS RGB image and, if required, the brightness temperature image. Figure 2 gives an illustration of the algorithm performance for the area centered at the NASA Goddard Space Flight Center (Greenbelt, Maryland, USA) for different seasons of the year. Each column contains 20 consecutive images covering the Julian day period shown on the top. The periods covered are not even because data were missing for some days while on others the area could be observed twice from different orbits. The reproducible spatial pattern of the surface on cloud-free days, clearly visible in these image sequences, is a key to the success of the developed algorithm.

The first two columns of images show the initialization stage of the covariance algorithm. In this demonstration, module *CloudMask1* was turned off. Blocks 2 and 3 with the largest contrast are initialized first, and block 4 with the lowest contrast is initialized last. The cloud-free blocks are identified accurately by the covariance analysis, and performance of algorithm is robust regardless of the surface brightness. For example, the first three columns show rather accurate mapping of clear conditions over variable snow. Generally, the algorithm tends to overestimate cloudiness during snowmelt when surface changes are rapid and no reliable comparison target can be established based on previous measurements. It achieves best performance when the surface condition is stable, for example over deserts and arid regions, or, in northern latitudes, in summer months or during winter with stable snow cover. The last two columns of Figure 2 show a reliable detection of different types of clouds in summer and autumn.

To gain insight on the large-scale algorithm performance, we used the 2004-2005 MODIS TERRA data for northeastern USA, Southern Africa (Zambia), Amazon region (Brazil), Arabian peninsular, and Greenland. The testing was done for at least half a year of continuous data in each case.

Figure 3a shows a case of cloud detection over receding snow for three winter days (36-37, 42) of 2005 for the north-east USA. The area of the image is 600×600 km^2. The two RGB images have a different normalization, helping visual distinction between snow and clouds. The *MAIAC* cloud mask is shown on the right, and the MODIS Collection 5 (MOD35) reprojected and gridded cloud mask is shown on the bottom. The conditions represent a different degree of cloudiness over the land. It is entirely clear on day 35. *MAIAC* CM algorithm gives an accurate overall classification. Thin ice on Lake Erie is partly misclassified as clouds. It is not as bright as snow in the visible bands, and has a higher than snow reflectance in the shortwave infrared (2.1 μm). The same holds true for the block of land and some pixels in the transitional zone from snow to land which are masked as clouds. As explained earlier, the error is expected in these cases. On day 36, *MAIAC* accurately detects a cloud stretching across Lake Erie. There are two large cloud systems on day 42, in the left upper and left bottom parts of the image, captured well by the algorithm. These images also show the strong retreat of the snow line by day 42, and a high quality of snow mapping by *MAIAC*. The MOD35 product accurately detects clouds, but it also overestimates cloudiness over snow on all three days, with the highest error on day 37.

Figure 3b compares the cloud mask of the two algorithms for the late spring of 2005 for the same region. Over land, the accuracy is similar. Some difference exists with regards to thin cirrus, or otherwise semitransparent clouds. *MAIAC* CM does not explicitly try to mask these clouds. Created for the purpose of aerosol retrievals and atmospheric correction, the algorithm maximizes

the volume of data available for the atmospheric correction. Our study shows that achievable accuracy of surface reflectance retrievals through thin cirrus is sufficiently high [*Lyapustin and Wang*, 2007]. Another notable difference is cloud detection over the water. One should bear in mind that the current algorithm has been developed for the land applications. Cloud detection over water at this stage is rudimentary, and presented examples may display a number of artifacts.

A large-scale comparison of the cloud mask products for the African Savannah (Zambia) is shown in Figure 4 for the area of 1200 km. This is the region of intense biomass burning later on during the dry season. The *MAIAC* and MOD35 cloud masks are generally comparable. *MAIAC* is a little more sensitive, detecting more clouds. One stark difference is the large number of "possibly clear" pixels in MOD35 when the algorithm cannot declare clear conditions with confidence. This category is not used in *MAIAC*, which has a very straightforward and reliable covariance criterion and ancillary *refcm* data to identify clear conditions. This feature is particularly appealing to land applications, sometimes significantly increasing the volume of measurements, which may be confidently used in the atmospheric correction and in further applied analysis.

A final example of the cloud mask comparison for the large (1800 km) bright desert area of Arabian Peninsula is shown in Figure 5 for days 145 and 207 of 2005. Here, the *MAIAC* cloud mask is shown on the top of image, and the MOD35 product is shown on the bottom. Except for a few small differences, the products agree quite well for day 145. On day 207, MOD35 overestimates cloudiness masking the dust storm areas as clouds.

The new algorithm is demonstrating a high accuracy of cloud discrimination over land.

5. Conclusions

The cloud mask algorithm described here implements relatively simple and straightforward logic. Using covariance calculations, the algorithm has a reliable metric to identify cloud-free conditions at the block level, and to define the reference clear-skies image (*refcm*) of every block of the land surface. It works successfully over both dark and bright surfaces, including deserts and snow, as well as in both clear and hazy conditions. Dynamic updates of the reference clear skies image allows the CM algorithm to smoothly adjust to the seasonal surface variations. Rapid surface change events (fire burns, snow fall etc.) are accommodated through repetitive re-initialization of *refcm*. Keeping memory of the clear-skies image and of the essential statistical properties of reflectance and brightness temperature for every land surface block strongly enhances the probability for correct cloud detection in difficult cases. An internal dynamic land-water-snow classification is also very helpful as it permits adapting logic of cloud search as a function of surface type.

Comparison with the MODIS Collection 5 cloud mask shows that *MAIAC* CM improves performance over snow and deserts. The accuracy of *MAIAC* CM and MOD35 over regular land surfaces is found to be similar. On the other hand, the MOD35 product often leaves some doubt for the land disciplines in a form of "possibly clear" pixels. When the high quality of input data is required, as in the MODIS BRDF/albedo algorithm, such pixels are often discarded from processing thus noticeably reducing the volume of data. With a reliable criterion for identification of clear surface regions, the use of "possibly clear" category in *MAIAC* CM is no longer necessary.

Variability of the footprint with the view zenith angle is one of largest sources of uncertainty in the time series cloud screening. Although our gridding algorithm reduces the resolution of MODIS

500 m bands to 1 km, it still cannot eliminate the resultant noise. In this sense, geostationary orbit would provide optimal performance conditions.

The *MAIAC* CM algorithm is still a work in progress. We are not sure if the time series analysis may bring any advance in cloud detection over water, which lacks spatial variability. However, the use of *refcm* should help cloud detection over the in-land, especially shallow, waters, which may have strong seasonality. This part of the algorithm is yet to be developed. A global analysis is needed in order to fully optimize the spectral and brightness temperature tests. On the other hand, we will continue analysis of the information content of other MODIS channels to enhance *refcm* structure for better cloud discrimination in partly cloudy conditions. Interestingly, we found that the MODIS "cirrus" channel B26 (1.38 μm) does not add new information in our algorithm and thus it is not used in *MAIAC* cloud detection over land.

MAIAC CM algorithm offers the potential to perform better than the MODIS MOD35 cloud mask in situations where the land surface is changing rapidly, and over Earth regions covered by snow and ice. The work has been initiated on the global inter-comparison of *MAIAC* and MOD35 cloud mask in collaboration with the University of Wisconsin.

Acknowledgements. The work of Dr. Lyapustin and Dr. Wang was supported by the NASA EOS Science (Dr. D. Wickland) grant. The first author is grateful to Dr. P. Menzel for comments on this paper.

REFERENCES

Ackerman, S. A., Strabala K., Menzel, W. P., Frey, R. A., Moeller, C. C., Guclusteringey, L. E. (1998): Discriminating clear-sky from cloud with MODIS. *J. Geophys. Res.*, *103*, 141-157.

Ackerman, S., K. Strabala, P. Menzel, R. Frey, C. Moeller, L. Gumley, B. Baum, S. Seemann, H. Zhang, Discriminating clear-sky from clouds with MODIS. Algorithm Theoretical Basis Document (MOD35). Ver. 5.0, October 2006, 129 pp.

Diner, D. J., L. Di Girolamo, and E. Clothiaux, 1999: MISR Level 1 cloud detection algorithm theoretical basis, Jet Propulsion Laboratory, JPL D-13397, Rev. B, 38 pp.

Holben, B. N., T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, A. Smirnov, 1998: AERONET-A Federated Instrument Network and Data Archive for Aerosol Characterization, *Rem. Sens. Environ.*, **66,** 1-16.

Lyapustin, A., Y. Wang, *MAIAC* – Multi-Angle Implementation of Atmospheric Correction for MODIS. Part I: Aerosol Retrievals. *J. Geophys. Res.*, to be submitted (2007).

Lyapustin, A., Y. Wang, *MAIAC* – Multi-Angle Implementation of Atmospheric Correction for MODIS. Part II: Atmospheric Correction. *J. Geophys. Res.*, to be submitted (2007).

McClain, E. P., 1993: Evaluation of CLAVR phase-I algorithm performance. *Final Report* 201-424.

Nakajima, T., and M. Tanaka, 1983: Effect of wind-generated waves on the transfer of solar radiation in the atmosphere-ocean system. *J. Quant. Spectrosc. Radiat. Transfer* **29**, 521-537.

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

Roy, D. P., P. E. Lewis, C. O. Justice, 2002: Burned area mapping using multi-temporal moderate spatial resolution data – a bi-directional reflectance model-based expectation approach. *Remote Sens. Environ.*, **83**, 263-286.

Schaaf C. B., Gao F., Strahler A.H., et al., 2002: First operational BRDF, albedo nadir reflectance products from MODIS, *Rem. Sens. Environ.*, **83**, 135-148.

Wang, Y., and A. Lyapustin: The impact of Bowtie Effect and View Angle Discontinuity on MODIS Swath Data Gridding, *IEEE Trans. Geosci. Rem. Sens. Letters*, submitted (2007).

APPENDIX

A. Spectral Tests and Dynamic Surface Classification

MAIAC algorithm maintains a dynamic land-water-snow mask (*mask_LWS*) which guides the cloud mask algorithm and controls the path and selection of the surface BRF model during aerosol-surface reflectance retrievals. It also helps processing algorithms to adjust to surface changes, such as snow fall/ablation, flooding etc. It has three stable values (MASK_LAND, MASK_WATER, MASK_SNOW) and two transitional values used when surface change is detected (MASK_TO_LAND, MASK_TO_WATER). For example, value MASK_TO_LAND represents transition from snow or water to land. In this work, the term "land" implies any land surface other than water or snow. In order to handle surface change, the algorithm uses a supporting mask indicating stability of state (*mask_Change*) which has values of MASK_STABLE and MASK_CHANGE. The vegetation and water classification is performed at the pixel-level using binary logic, which is enforced by the BT-analysis. Because snow test does not necessarily distinguish between the snow/ice clouds and the snow, the snow detection uses a different logic described in Sec. A.2.

A.1 Detection of Vegetation and Water

When the new Tile is received, two tests are used to detect clear pixels and validate or change the status of mask *mask_LWS*:

1. High NDVI test: $NDVI=(r2 - r1)/(r2 + r1)>0.6, \Rightarrow$ CM_CLEAR.

If the pixel passes this test, the value of cloud mask is set to CM_CLEAR. If the previous value of land-water-snow mask was MASK_LAND or MASK_TO_LAND, then the value of mask is validated: *mask_LWS* =MASK_LAND, *mask_Change* =MASK_STABLE. Otherwise, change is detected: *mask_LWS* =MASK_TO_LAND, *mask_Change* =MASK_CHANGE.

This test finds heavily vegetated pixels of the land. Sparsely vegetated regions, bare soil, rocks, sand etc., are classified as land during covariance analysis, when the pixel is cloud-free, covariance at the block-level is high, and the pixel was not classified as either water or snow.

2. Water test: $r2<0.07$ AND $r5<0.02$ AND $r7<0.015$ AND $NDVI\leq0.2, \Rightarrow$ CM_CLEAR_WATER.

The water test is conducted only for the off-glint geometries, which are defined according to a condition $r_{glint}<0.02$, where r_{glint} is a Cox-Munk glint reflectance for the wind-ruffled water surface calculated at wind speed of 7 m/s. Theoretical reflectance is pre-calculated using *Nakajima and Tanaka* (1983) model with the mutual shadowing of waves, and stored in the LUT. The algorithm can use a real time wind speed, if it becomes known operationally from independent sources.

In the similar manner as above, the value MASK_WATER is either validated with the new measurement or change is detected if spectral tests find a signature of vegetation or soil.

Different factors may lead to failure of the *water* test for a given pixel: elevated aerosol loading, clouds, transition to land (drying of shallow water, drainage) etc. For this reason, additional processing is performed for the pixels which were earlier classified as water (q. *mask_LWS* =MASK_WATER) but did not pass the *water* test this time. First, the land-restore test ($r5/r1>1.2$) checks the possible transition to land. Next, the brightness temperature of water for the block

(BT_W), evaluated from the detected water pixels, is used either to confirm the clear water pixels, according to criteria:

$$BT > BT_W\text{-}1K, \text{ AND } r5 < r_{glint} + 0.03, \Rightarrow \text{CM\_CLEAR\_WATER},$$

or detect clouds if

$$BT < BT_W\text{-}1K, \text{ AND } r5 > r_{glint}, \Rightarrow \text{CM\_CLOUD}.$$

If BT_W is unavailable, the brightness temperature is compared to a fixed threshold:

$$BT > 272.5K, \text{ AND } r5 < r_{glint} + 0.03, \Rightarrow \text{CM\_CLEAR\_WATER}.$$

Last, the algorithm tests pixels of medium brightness, which may be the result of aerosols enhancing reflectance over the water:

$$r1 < 0.12 \text{ AND } |r1 - r2| < 0.015, \Rightarrow \text{CM\_CLEAR\_WATER}.$$

A.2 Detection of Snow

Snow detection begins with the snow test:

3. Snow test: $NDSI = (r4 - r6)/(r4 + r6) \geq 0.35$ AND $r1 > 0.15$ AND $r7 < 0.12$.

This test is performed for every pixel of the block if the maximal brightness temperature of the block and its nearest neighbors does not exceed 293K to avoid false snow detection on summer, and for the pixels inside the block with BT<283K over land or BT<275K over water. The temperature thresholds are high because snow-free patches of land surface can be very warm in spring, while the snow amount still being significant to warrant snow detection. The last condition (r7<0.12) serves to filter some of the mixed-phase clouds which are abundant and often have a higher reflectance at wavelength of 2.1 μm. The snow, as seen in MODIS channels, is usually darker than the specified threshold (r7≈0.03-0.09), although fresh snow and some types of snow/ice, which can be found for example on the high elevation slopes of Greenland, can be as bright as r7≈0.15-0.20. The band 7 threshold also filters some pixels partially covered by snow. To classify these pixels as snow, we have a different mechanism, described later (Sec. B).

Overall, snow detection is a difficult problem. First, snow/ice clouds often pass the snow test so it alone cannot guarantee reliable snow detection. Second, snow in the mid-latitudes during winter is often short-lived, which gives rise to high variability of surface brightness in time and thus requires changing the logic of *refcm* update and of the overall use of the reference comparison target. Third, snow, partially covering the area of a pixel, is particularly difficult case and is often prone to misclassification as cloud.

To filter clouds, which pass the snow test, we adopted a conservative approach by which the pixel can be masked as snow for the first time only during initialization or re-initialization of *refcm*. The requirement of high covariance and a carefully designed set of filters, described in sec. 2.1, for the most part help to separate clouds. For snow, we reduce the high covariance threshold to 0.65 in order to accommodate the transient nature of snow in mid-latitudes. When the high covariance conditions are satisfied, the cloud mask for the earlier Tile *k*, which had correlated with the last Tile, is reset to the value CM_CLEAR_SNOW. In this sense, the cloud mask of any given block of Tile may change anytime while the Tile remains in the Queue. This post-processing modification always increases the number of CM_CLEAR or CM_CLEAR_SNOW pixels. Once the snow is

recorded at the Q-level (in q.*Mask_LWS*), the algorithm returns to the binary pixel-level logic of snow confirmation which is described in section B.

The described conservative strategy, which requires two clear days in the Queue, may delay detection of fresh snow up until the high covariance with the later Tile is found, during which time it will be masked as clouds. On the other hand, it dramatically reduces misclassification of high clouds in mid-latitudes in the summer, and in tropical regions, although it cannot completely eliminate this error. In our limited testing, the new algorithm, despite being conservative, still finds significantly more clear snow pixels than the MODIS cloud mask algorithm (MOD35) which is known to overestimate cloudiness over snow.

B.1 Module *CM_highCov*

This module is called when *refcm* is initialized and covariance between *refcm* and new Tile for a given block is high (*cov* > 0.68, or 0.65 for snow) indicating clear-skies conditions although a few pixels may still be cloudy. If the BT contrast is low, then all pixels of the block are masked as CM_CLEAR, the ground BT is calculated, and *refcm* is updated. If it exceeds a threshold of q.ΔBT+3, the algorithm masks possibly cloudy pixels that exceed the *refcm* reflectance and at the same time are colder than the ground brightness temperature with an offset:

$$\text{IF } (BT_{ij} < BT_G - 4) \text{ AND } (r1_{ij} > refcm.r1_{ij} + 0.05 + dif) \Rightarrow \text{CM\_PCLOUD.}$$

Dif is the difference between the average reflectance of the *last Tile* and of *refcm* over the common area of a given block. It is designed to remove bias caused by differences in the view geometry or aerosol concentration. The rest of pixels are masked as CM_CLEAR. If BT_G remains undefined after spectral tests, it is first evaluated as a minimal value for 90% of the hottest pixels of the block using histogram analysis.

In the presence of snow, which may change rapidly, comparison of reflectance is not helpful, and the logic changes. Our algorithm calculates the minimal brightness temperature of the block for 90% of warmest pixels, which are assumed to be cloud-free, $BT_{min} = min\{ BT(0.9) \} - 2^{\circ}$, and masks pixels which are colder than BT_{min} as CM_PCLOUD. The bright pixels of intermediate temperature between BT_{min} and 280K are masked as CM_CLEAR_SNOW, and the value of the q.*mask_LWS* is set to MASK_SNOW. The pixels, darker than 0.2, are masked as clear:

IF $r1_{ij} > 0.2$
{
  IF $BT_{ij} < BT_{min}$ OR $BT_{ij} > 280 \Rightarrow$ CM_PCLOUD.
  ELSE $\Rightarrow$ CM_CLEAR_SNOW, q.*mask_LWS*$_{ij}$= MASK_SNOW.
}
ELSE $\Rightarrow$ CM_CLEAR.

Because the covariance is high, and a given block for both days is known from the *snow* test to have snow, the described procedure is a relatively safe way to introduce new snow pixels which did not pass *snow* test, and were not detected as snow before. As mentioned earlier, the *snow* test often misses partially covered snow pixels, which have relatively high reflectance in band 6 and are not so bright in band 4. It also misses some forms of snow/ice with high reflectance in band 7 (2.1 μm). Because covariance with thoroughly selected *refcm* image is high, thus introduced snow has a low commission error.

B.2 Module *CM_lowCov*

This module is called when covariance is low ($cov \leq 0.68$) indicating that the block may be cloudy or partly clear. It starts with evaluation of BT_G for a given block if it is undefined. The further processing depends whether or not the snow was detected before (at the Q-level). If not, then the algorithm uses the Bright-Cold algorithm to mask clouds:

$$\text{IF } (BT_{ij} < BT_G - 4) \text{ AND } (r1_{ij} > refcm.r1_{ij} + 0.05) \Rightarrow \text{CM\_PCLOUD.}$$

In this case, the average brightness of the scene is not subtracted because it can be biased by clouds.

When there is snow on the ground, the logic is reversed as compared to the case of high covariance. This algorithm does not add new snow: the pixel can be declared as CM_CLEAR_SNOW only if it satisfies three conditions: 1) it passes *snow* test, 2) snow was detected before (*q.mask_LWS* = MASK_SNOW), and 3) its reflectance is close to *refcm* value, $\left| r1_{ij} - refcm.r1_{ij} \right| \leq 0.05$. The conservative last requirement, which assumes rather stable snow conditions, is essential because otherwise a significant number of snow/ice clouds end up detected as snow.

If one of three conditions is not satisfied, and the pixels's band 1 reflectance exceeds 0.12, it is masked as CM_CLOUD. For this reason, the darker surface partially covered by snow is often masked as cloud. This is a common error of comission typical of all snow detection algorithms.

Table 1. MODIS Data Used in *MAIAC* CM Algorithm

MODIS band	λ_C (μm)	Nadir resolution (km)	Primary Use
B1	0.645	0.5	Covariance analysis, *refcm*. LSC – Vegetation, Water. Land restore test. Thick bright cloud test.
B2	0.856	0.5	LSC – Vegetation, Water.
B3	0.466	0.5	Thick bright cloud test.
B4	0.554	0.5	LSC – Snow.
B5	1.242	0.5	Shadow detection. Land restore test. Thick bright cloud test.
B6	1.629	0.5	LSC – Snow.
B7	2.113	0.5	LSC – Vegetation, Water, Snow.
B31	11.030	1.	BT analysis, *refcm*. LSC.

LSC – Land Surface Classification (land, water and snow).

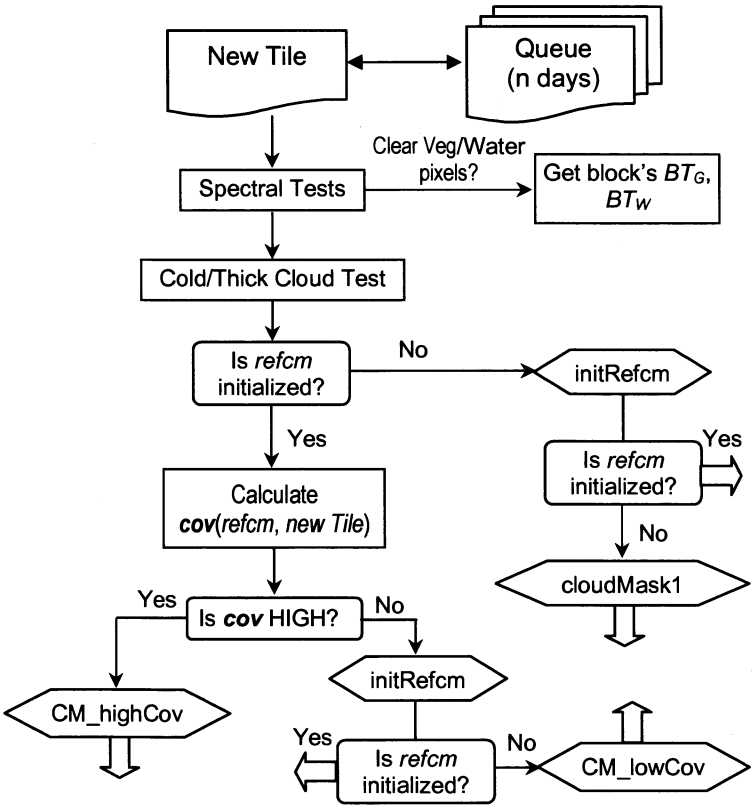

Figure 1. The general flowchart of CM algorithm. After spectral tests, the algorithm first tries to initialize the reference clear-skies image (*refcm*), and if fails then it uses a backup cloud mask algorithm (module *cloudMask1*). If *refcm* is available, the CM algorithm calculates covariance between the latest image and *refcm*, and carries on further analysis depending on whether covariance is high or low. In case of low covariance, the algorithm takes an intermediate step trying to re-initialize *refcm*, which takes care of rapid surface changes as well as of possible errors in *refcm* caused by previously undetected clouds.

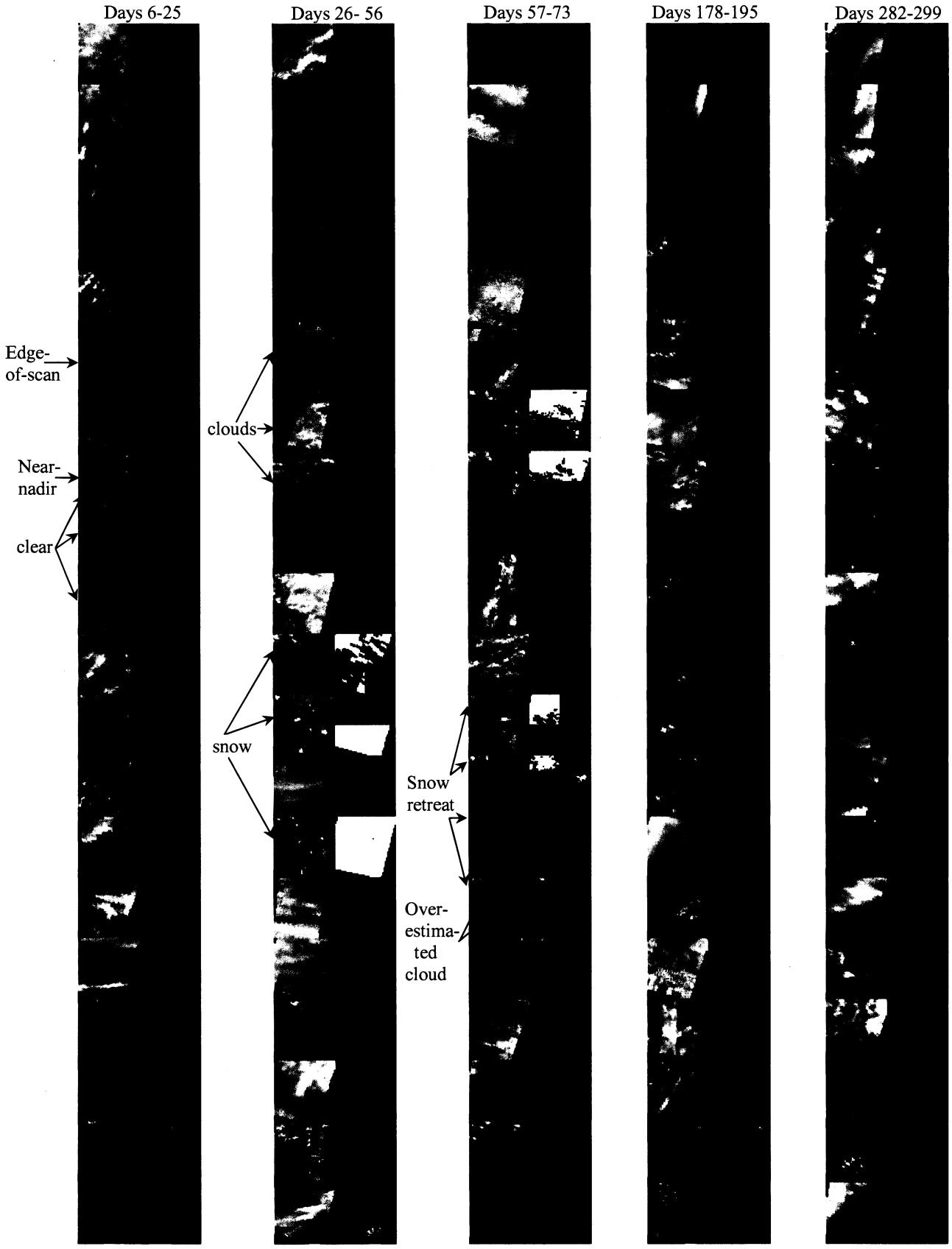
Days 6-25
Days 26- 56
Days 57-73
Days 178-195
Days 282-299
Edge-of-scan
Near-nadir
clear
clouds
snow
Snow retreat
Over-estima-ted cloud

Figure 2. Examples of performance of *MAIAC* CM algorithm for 50×50 km^2 MODIS TERRA subsets for 2003, centered at GSFC, USA. Normalized gridded RGB images are shown on the left, cloud mask is shown on the right. CM legend: Blue and White – clear over land and snow, respectively; Red – cloudy.

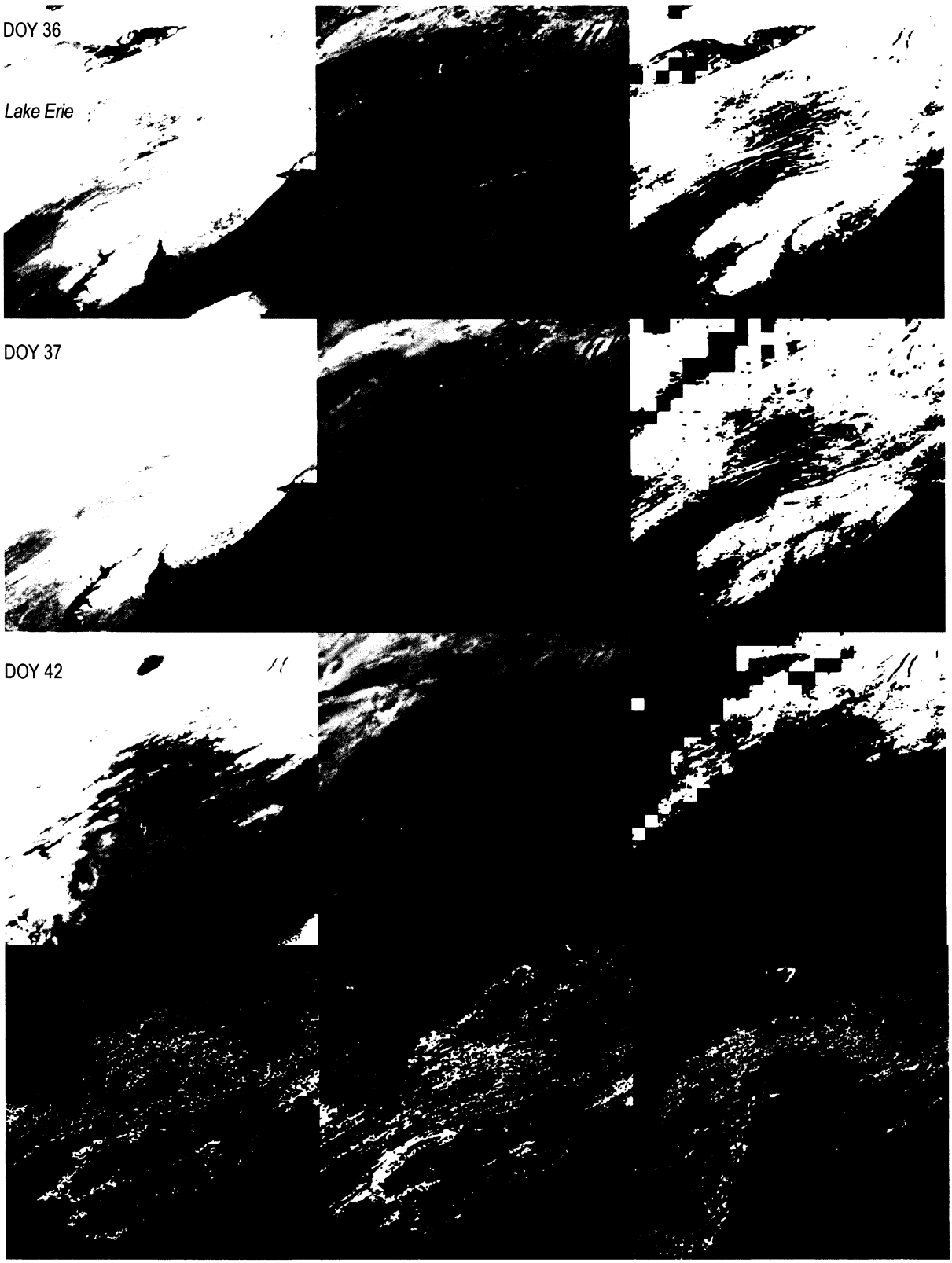
DOY 36
Lake Erie
DOY 37
DOY 42

Figure 3a. Example of *MAIAC* and MOD35 (bottom) cloud mask over snow from MODIS TERRA data for days 36, 37 and 42 of 2005. The image shows 1 Tile (600×600 km^2) for the north-east USA. The two RGB images have a different normalization helping visual distinction between snow and clouds. MOD35 cloud mask for days 36, 37 is shown at the bottom. Legend for MOD35 CM: Blue – clear, Green – possibly clear, Yellow – possibly cloudy, Red – cloudy. Legend for MAIAC CM: Blue, Light Blue, and White – clear (land, water and snow, respectively), Yellow – possibly cloudy, Red – cloudy.

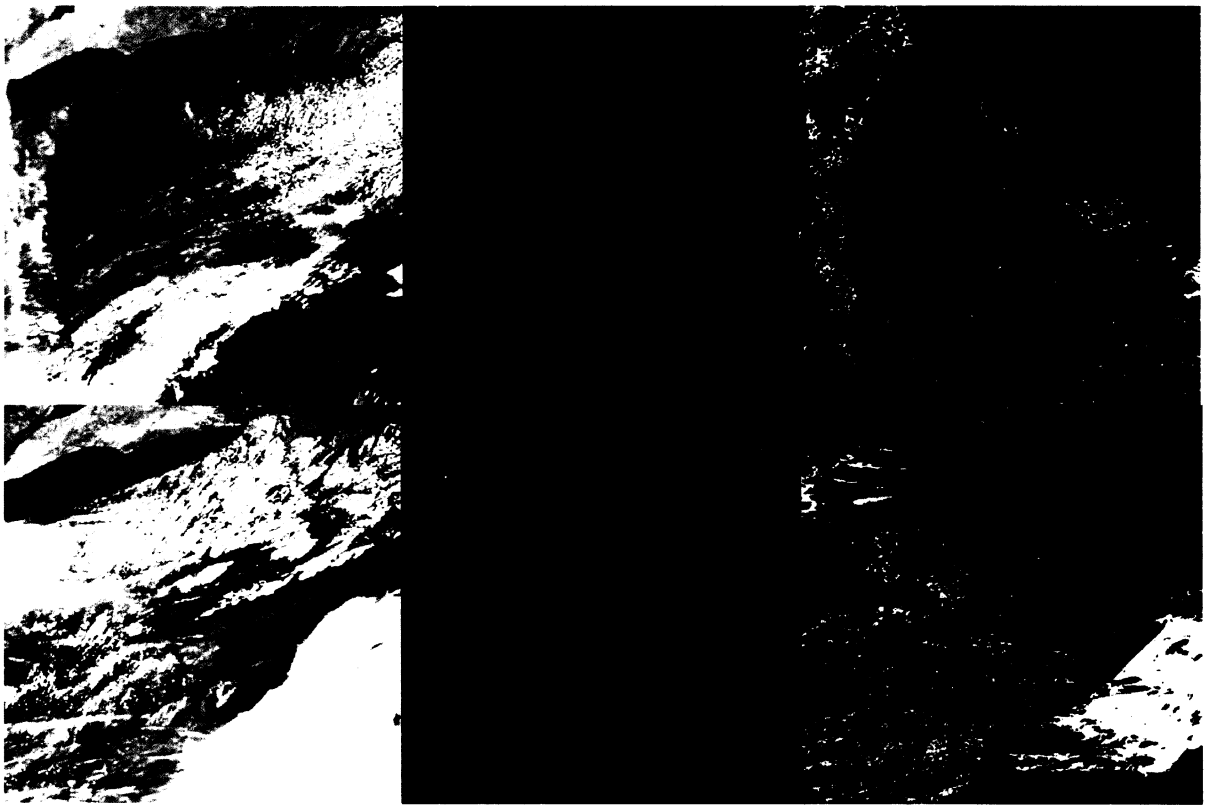

Figure 3b. Example of *MAIAC* (center) and MOD35 (right) cloud mask from MODIS TERRA data for days 138 (top) and 152 (bottom) of 2005. The image shows the same Tile (north-east USA) as in Figure 3a.

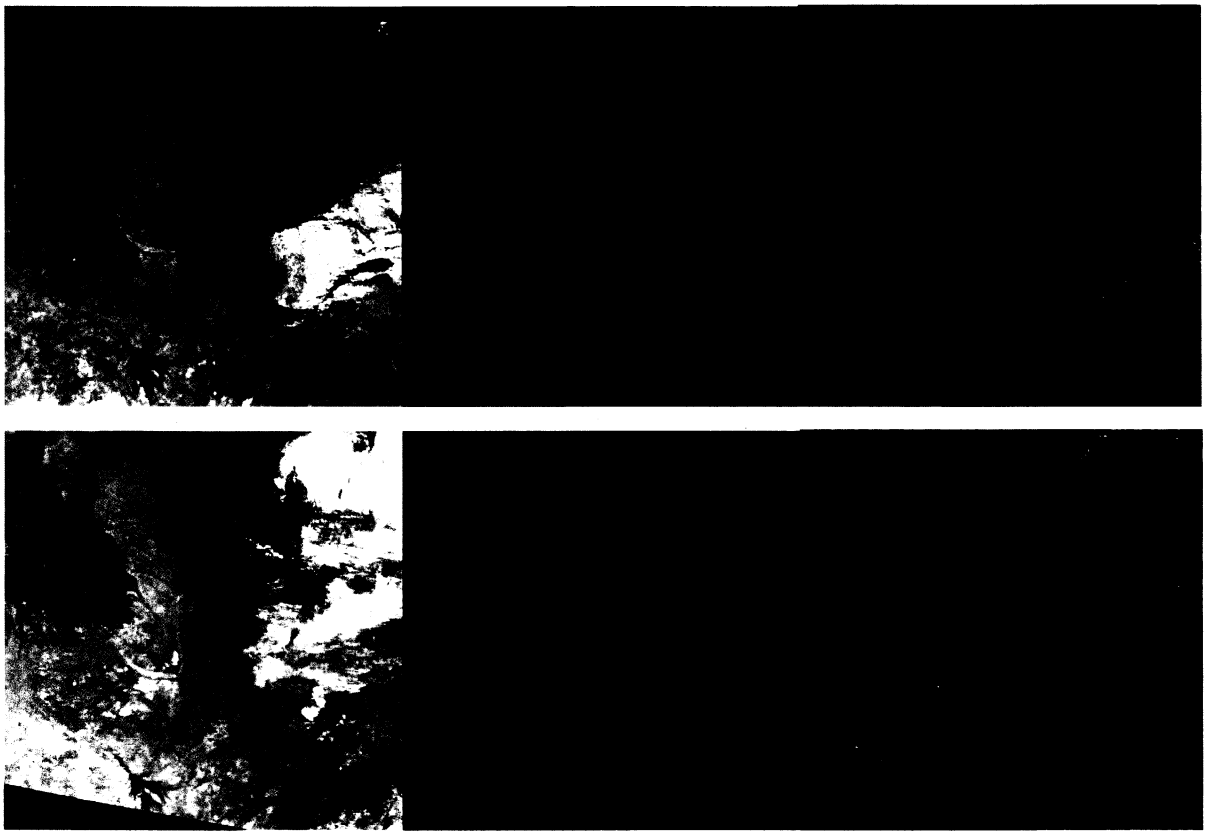

Figure 4. Example of *MAIAC* (center) and MOD35 (right) cloud mask at the beginning of dry season for Zambia, Africa, from MODIS TERRA data for days 130 (top) and 141 (bottom) of 2005. The image shows 4 Tiles (1200×1200 km^2).

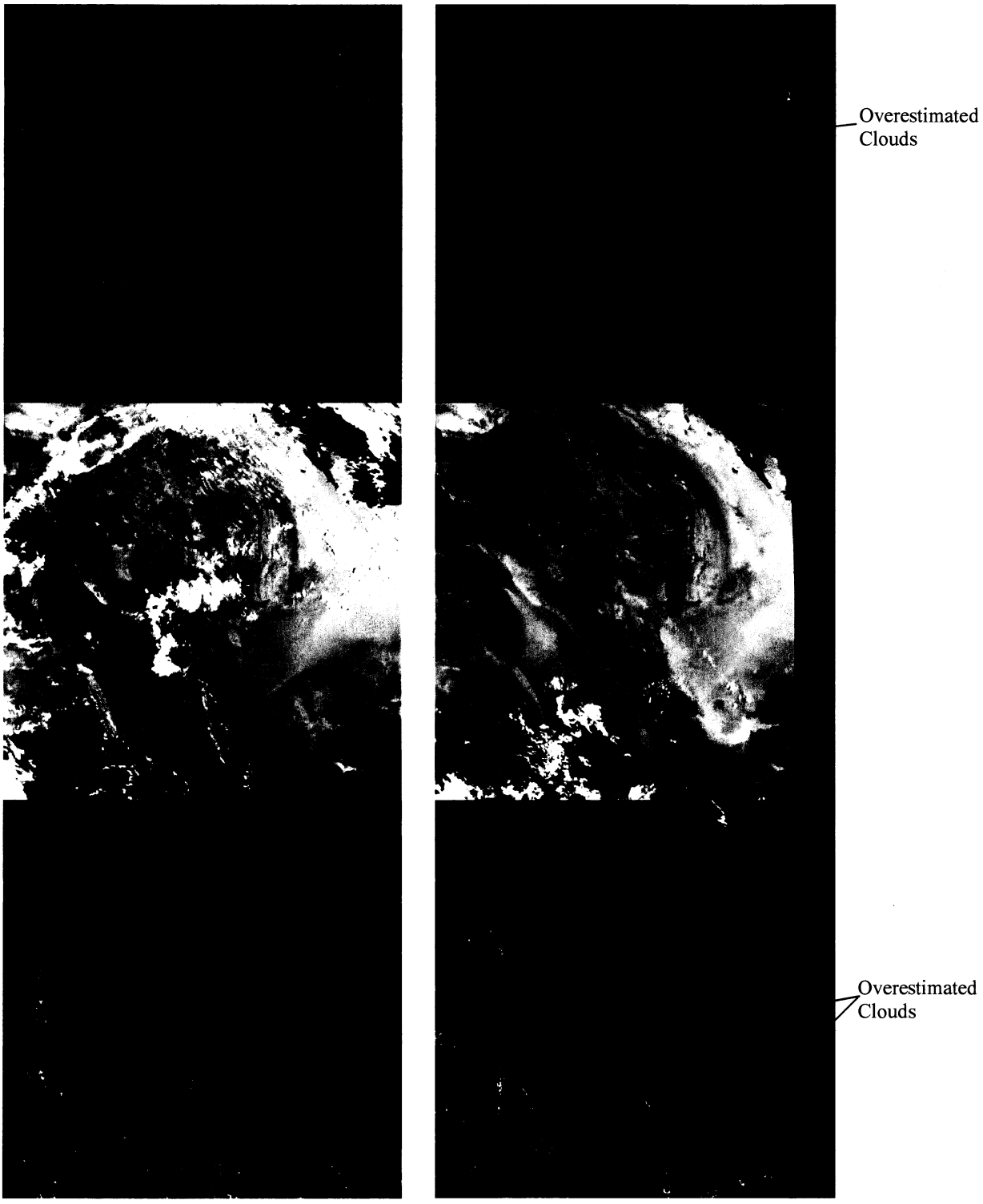

Figure 5. Example of *MAIAC* (top) and MOD35 (bottom) cloud mask for Arabian Peninsula from MODIS TERRA data for days 145 (left) and 207 (right) of 2005. The image shows 9 Tiles (1800×1800 km^2).