An automatic cloud mask algorithm based on time series of MODIS measurements

A. Lyapustin,[1] Y. Wang,[1] and R. Frey[2]

Received 27 November 2007; revised 1 May 2008; accepted 21 May 2008; published 26 August 2008.

[1] Quality of aerosol retrievals and atmospheric correction over land 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 do not explicitly address the classical problem of cloud search, wherein the baseline clear-skies scene is defined for comparison. Here we report on a new land 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 multiangle implementation of atmospheric correction (*MAIAC*) algorithm for MODIS, relies on the 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.

Citation: Lyapustin, A., Y. Wang, and R. Frey (2008), An automatic cloud mask algorithm based on time series of MODIS measurements, *J. Geophys. Res.*, *113*, D16207, doi:10.1029/2007JD009641.

1. Introduction

[2] Cloud mask (CM) is a primary science algorithm that precedes detailed analysis of cloud, aerosol and land surface/ocean parameters from 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 the strong interdependence between human induced land changes and 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.

[3] 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. These algorithms utilize a generic land type classification but lack a priori knowledge about specific surface reflectance 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 spectral signatures of the surface, such as dense vegetation or deep water. Due to tremendous variability of clouds, their detection has always been problematic over brighter surfaces, especially snow. Because of similarities in spectral reflectances between snow and snow/ice clouds, and temperature inversions frequent in the low troposphere in wintertime, no particular set of spectral reflectance and brightness temperature tests may guarantee success over snow and ice.

[1]University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[2]Cooperative Institute for Meteorological Satellite Studies (SSEC), University of Wisconsin, Madison, Winconsin, USA.

Copyright 2008 by the American Geophysical Union.
0148-0227/08/2007JD009641

[4] On the other hand, contemporary polar-orbiting global observation imagers, such as AVHRR, MODIS, or future VIIRS, provide or will provide a daily global view of the Earth at the equator with multiple observations per day in midlatitude and polar regions. This opens up possibilities to add temporal and spatial dimensions into the cloud mask processing. For example, with a high frequency of observation, 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 a 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.

[5] We present a next step in the evolution of this idea - a new CM algorithm working with the time series of MODIS measurements and developed as a part of the multiangle implementation of atmospheric correction (*MAIAC*) algorithm [*Lyapustin and Wang*, 2007]. It uses covariance analysis to build reference clear skies images (*refcm*) and to accumulate a certain level of knowledge about the surface and its variability, thus constructing rather comprehensive comparison targets for cloud masking. The new algorithm has an internal surface classifier, producing a 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 bidirectional 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 increases the overall quality of cloud mask.

[6] The CM algorithm described in this paper was developed specifically for the purpose of atmospheric correction over land. Its goal is to identify clear conditions when measurements remain useful for surface retrievals, and cloudy conditions when no aerosol retrieval and atmospheric correction can be applied. In this regard, it differs from traditional cloud masks. For example, spatially homogeneous thin cirrus clouds are not masked out if the surface spatial variability is detectable from space and agrees with the known spatial pattern of *refcm*. The *MAIAC* CM produces only an integral cloud mask with values of CM_CLEAR for clear conditions with the mentioned caveat, and cloudy conditions (CM_POSSIBLY_CLOUD or CM_CLOUD), and it does not keep values of separate tests which could be useful for cloud classification. The new algorithm is not applicable for the global ocean and large spatially homogeneous inland water bodies with size exceeding ~25 km.

[7] 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. A brief technical background on implementation of the time series processing is given in section 2.

[8] Section 3 describes the algorithm constructing the reference clear-skies image (*refcm*). Section 4 and Appendix B present overall decision logic in cloud masking and provides the necessary technical detail. A dynamic surface classification algorithm is discussed in Appendix 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 5. 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. Implementation of a Time Series Processing

[9] In order to arrange time series processing, *MAIAC* first grids MODIS level 1B (L1B) calibrated and geolocated data to a regular 1 km grid. We use the MODIS land gridding algorithm [*Wolfe et al.*, 1998] with minor modifications that allow it to better preserve the anisotropy of signals in the gridded data when measured reflectance is high, for example over snow, thick clouds or water with glint. Next, *MAIAC* splits gridded data into 600 km *Tiles* representing fixed geographical regions of the world, and places the new *Tile* in the processing Queue. The Queue, designed for a "sliding window" algorithm, holds up to 16 days of successive gridded measurements. The current size of a Tile (600 km) is selected to fit the operational memory of our workstation. As a reminder, MODIS operational processing uses 1000 km Tiles.

[10] The *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, implement the sliding window algorithm, and dynamically update the Queue. The class functions are designed to easily handle processing in the various time-space scales, for example at the pixel-versus-block level, and for a single (last) day of measurements versus all available days in the Queue, or for a subset of days passing certain filters. The data storage in the Queue is efficiently organized using pointers, which avoids physically moving the previous data in memory when the new data arrive.

[11] The structure of the Queue is shown schematically in Figure 1. For every day of observations, MODIS measure-

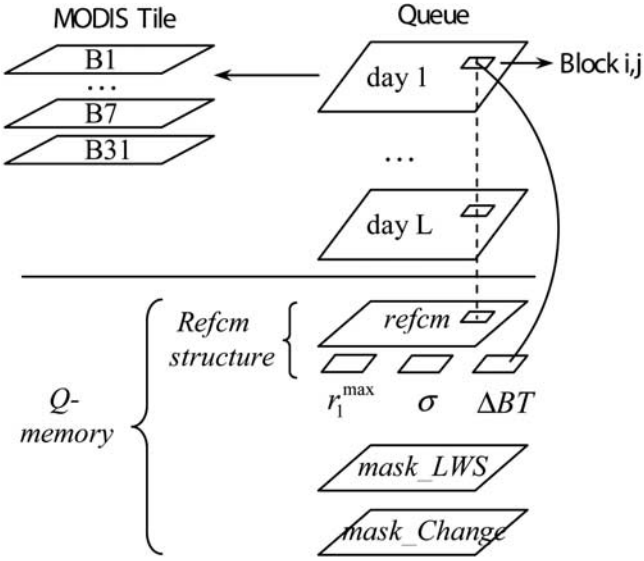

Figure 1. Structure of Queue for *MAIAC* CM algorithm. The Queue, designed for the sliding window algorithm, stores up to 16 days of gridded MODIS observations (600-km tile, 1-km grid cell). The CM algorithm uses MODIS bands 1–7 and band 31, which are stored as layers (double-indexed arrays) shown in the top left corner. A dedicated Q-memory is allocated to store the ancillary information for CM algorithm in *Refcm* structure, including a reference clear-skies image (*refcm*) and a set of statistical parameters for each surface block, such as maximal reflectance and standard deviation in band 1, and a brightness temperature contrast. This information is updated with latest measurements (day L) once given block is found cloud-free. The Q-memory also stores results of the dynamic land–water–snow classification and a surface change mask.

ments are stored as Layers for reflective bands 1–7 and thermal band 31. Besides storing gridded MODIS data (*Tiles*), the Queue has a dedicated memory (Q-memory) which accumulates ancillary information about every block of the surface in the *Refcm* data structure. This structure holds the reference clear-skies image *refcm* and statistical parameters of the band 1 reflectance and brightness temperature contrast (for detail, see section 3.1, item 4). The *Refcm* structure is updated each time the latest measurements for a given block are classified as cloud-free. The Q-memory stores the latest results of the dynamic land–water–snow classification (*mask_LWS*), and a surface change mask (*mask_Change*). It also keeps certain information for the other *MAIAC* algorithms related to the history of previous retrievals, for example parameters of the surface BRF model and albedo along with the date of the last retrieval (update day). If the Q-memory was not updated for a certain period of time (currently 45 days) because of persistent cloudiness, it is cleared and reinitialized.

3. Building Reference Clear Skies Image

[12] The clear-skies images of a particular 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 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.

[13] 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 a pair of images for which covariance is high, and caution is exercised to exclude correlated cloudy fields (see section 3.1, item 7). The algorithm calculates a block-level covariance between the *new Tile* and the previous *Tiles*, moving backward in the Queue until either the "head" of Queue is reached, in which case initialization fails and the algorithm would wait for the new data to continue, or clear conditions are found. The latter corresponds 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 the image *refcm* and the data structure *Refcm* are initialized, the algorithm begins to use the block-specific value of the brightness temperature contrast *Refcm*.ΔBT.

[14] 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 midlatitudes, where snow accumulation is usually low and day-to-day variations of reflectance during wintertime can be enormous.

[15] After initialization, the algorithm uses the *refcm* to compute covariance with the latest measurements. Once clear conditions are found, *refcm* and structure *Refcm* are 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 reinitialization which is performed each time when covariance of the latest *Tile* with *refcm* is found to be low (see section 4).

[16] The *MAIAC* CM algorithm uses seven 500 m resolution MODIS bands B1–B7 and a 1-km thermal band B31

Table 1. MODIS Data Used in *MAIAC* CM Algorithm[a]

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.

[a]LSC, land surface classification (land, water, and snow).

(Table 1). As mentioned earlier, data are initially reprojected and gridded to a 1-km grid.

[17] (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 using 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 about 10% more 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 semitransparent 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 better overall performance.

[18] (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 $\approx 2 \times 4$ km^2 for 1 km^2 nadir pixels. On the other hand, the success rate of the 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 masks using block sizes from 15–25 km and select the optimum for operational application.

[19] (3) In order to account for the effects related to 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, *refcm*1 for VZA $\approx$ 0–45° and *refcm*2 for VZA = 45–60°.

[20] (4) The clear-skies image is stored in the Q-memory in the *Refcm* structure, which also stores the maximal value (*Refcm*.r1$_{\max}$) and the variance (*Refcm*.σ_1) of reflectance in band 1 as well as the brightness temperature contrast (*Refcm*.ΔBT = BT$_{\max}$ − BT$_{\min}$) for each block. Analysis of MODIS data shows that thermal contrast (ΔBT) 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 thermal contrast is usually low (1–6 K) for flat terrain at MODIS 1 km resolution. It may increase significantly (10–20 K) when the block is a mixture of land, water or snow pixels. In partially cloudy conditions the contrast increases because BT$_{\min}$ is usually lower over clouds.

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

[22] (6) The allowed values of the cloud mask are clear (CM_CLEAR, CM_CLEAR_WATER, CM_CLEAR_SNOW), indicating surface type as well, possibly cloudy (CM_PCLOUD), and confidently cloudy (CM_CLOUD). Two more values of cloud mask are CM_SHADOW for pixels defined as cloud shadowed, and CM_GREY representing one 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.

[23] (7) A high covariance between two days alone does not guarantee clear conditions. Clouds "leak" into *refcm* in a number of different ways, especially over snow. From an extensive analysis of MODIS data over the world, from the 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 ($k = L - 1, \ldots 1$) stored in the Queue, the following conditions are used to reject any of the Tiles k for a given block:

[24] • High BT contrast, which usually indicates clouds: $\Delta\mathrm{BT}_k > \min(q.\Delta\mathrm{BT} + 15, 25)$ (K).

[25] • 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 very low wind speeds.

[26] • The number of pixels which passed the snow test is significantly different between observations L and k.

[27] 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 rejection conditions:

[28] • The difference of the block-average reflectances should not be too high:

$$|\langle r1_L\rangle - \langle r1_k\rangle| > 0.18.$$

This test filters infrequent cases when bright continuous clouds correlate spatially with much darker clear scenes.

[29] • The maximal reflectance of the block should not significantly exceed the *Refcm* value:

$$\mathrm{r1}_{\max,L} > \mathit{Refcm}.\mathrm{r1}_{\max} + 0.1 \text{ OR } \mathrm{r1}_{\max,k} > \mathit{Refcm}.\mathrm{r1}_{\max} + 0.1.$$

If *Refcm*.r1$_{\max}$ is not defined (*Refcm* not initialized), the rejection condition states that both Tiles should not be simultaneously bright ($\mathrm{r1}_{\max,L,k} > 0.2$) and cold ($\mathrm{BT}_{\min,L,k} <$ 277 K), which is indicative of clouds.

[30] • 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 tests on both days in order to mask a pixel as snow for the first time. To avoid snow detection over the tropics during the rainy season according to this logic, we use a common sense test that relatively low temperatures should be established for some period of time before snowfall. On the contrary, if the surface was warm (BT_{max} > 288 K) 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.

[31] • 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.

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

4. CM Algorithm Flowcharts

[33] The central idea of the *MAIAC* CM algorithm is to use the reference clear-skies images of the surface (*refcm*) and covariance analysis to identify clear 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. Even when covariance is high, a few pixels of the block may still be cloudy. Following covariance calculation, the algorithm looks for clouds at the pixel level. For regular surfaces, not covered by snow, cloud detection 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 ground brightness temperature 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 high covariance. For example, the bright–cold algorithm for pixel (*i,j*) of the low-covariance block is formulated as follows (see section B1 of Appendix):

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

The logic is shown in the general flowchart of the CM algorithm (Figure 2). 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. The letters in parentheses show spatial and temporal domains of operations, for example at pixel (P) and/or block (B) level, and using the data of the last Tile (LT) only or using the full time series of the Queue (Q).

[34] (1) As a first step, *Spectral tests* are performed for every pixel of the last *Tile* to detect clear vegetated, water, and possibly snow pixels as described in Appendix A. These tests are used to initialize and/or update pixel-level surface classifications and change masks (*mask_LWS*, *mask_Change*). The detected clear vegetated and water pixels are used to calculate brightness temperature of the ground (BT_G) and of the water (BT_W) for every block, which are required in further analysis.

[35] (2) Next, a *Cold/Thick cloud test* is performed for each pixel of the last *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 brightness temperature BT_{max} 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 = 15 K if BT_G is defined, or $BT_0 = BT_{max}$, Δ = 20 K 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.

[36] The *Thick* cloud test is designed to detect spectrally neutral bright and thick clouds. Usually, reflectance of such clouds decreases slowly with wavelength within the 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.

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

[38] The CM_CLOUD value of the cloud mask generated by these tests can be overwritten later if high covariance is found.

[39] (3) The subsequent processing path depends on whether *refcm* (*Refcm* structure) has been initialized. If not, then the algorithm tries to initialize *refcm* as described earlier in section 3 (module *initRefcm*). During initialization, the algorithm consecutively computes covariance of the last measurements with each of the earlier images of the Queue for a given block, until a 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*.

[40] If *refcm* was initialized earlier, then the algorithm calculates covariance between the new measurements and

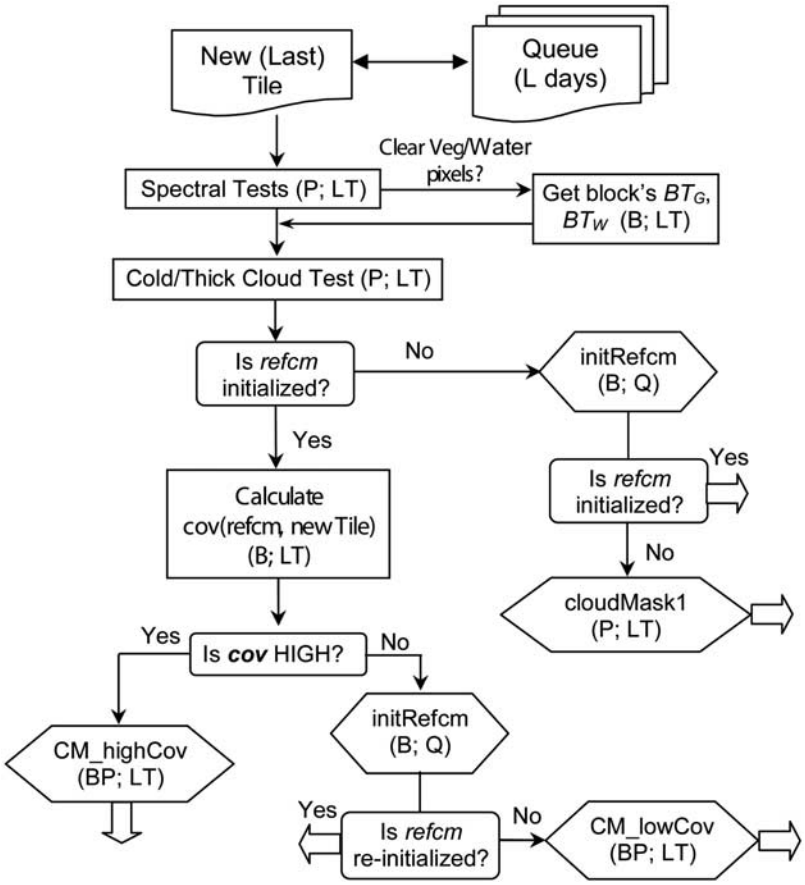

Figure 2. The general flowchart of CM algorithm. After spectral tests and cold/thick cloud test, the algorithm first tries to initialize the reference clear-skies image (*refcm*), and if fails, then it uses a backup cloud mask algorithm (module *cloudMask*1). If *refcm* is available, the algorithm calculates 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 reinitialize *refcm*, which takes care of rapid surface changes, as well as of possible errors in *refcm* caused by previously undetected clouds. In this diagram, rectangles represent separate functions, diamond shapes stand for separate subroutines (algorithms), and round-corner rectangles indicate decision (branching) points. The thick arrows show the points of exit from CM algorithm. The letters in parentheses indicate spatial and temporal domains of operations, for example, at pixel (P) or/and block (B) level, and using the data of the last tile only (LT) or using the full-time series of the Queue (Q). The pixel-level cloud mask is produced by modules *CM_highCov*, *CM_lowCov*, *cloudMask1*. Module *initRefcm* produces cloud mask only if initialization is successful. The *Refcm* structure is updated in modules *initRefcm* and *CM_highCov.*

the *refcm* image, and performs further processing depending on the value of covariance (modules *CM_highCov* and *CM_lowCov*). If covariance is low, the algorithm takes an intermediate step and attempts to reinitialize *refcm* with latest measurements. This conservative strategy serves to mitigate errors if partially or completely cloudy images bypassed filters of *initRefcm* and were used to update *refcm*. If this happened, the latest cloud-free image will not correlate well with *refcm*. Reinitialization also helps to update *refcm* when the surface reflectance changes rapidly.

[41] Following covariance calculations, the pixel-level cloud mask is produced by modules *CM_highCov* and *CM_lowCov* described in Appendix B. Although the logic implemented in these modules is somewhat different, the clouds in the absence of snow are detected similarly using the bright–cold algorithm.

[42] Each of modules *CM_highCov*, *CM_lowCov*, and *cloudMask1* produces pixel-level cloud mask for the new tile. Module *initRefcm* produces the 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.

5. Performance of *MAIAC* CM Algorithm

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

Before processing, we reprojected and gridded these data to a 1-km grid using the nearest neighbor method.

[44] 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 3 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. The covered periods are shown on the top. They 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.

[45] The first two columns of images show the initialization stage of the covariance algorithm. In this demonstration, module *CloudMask*1 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, in northern latitudes in summer months or during winter with stable snow cover. The last two columns of Figure 3 show a reliable detection of different types of clouds in summer and autumn.

[46] 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 peninsula, and Greenland. The testing was done for at least half a year of continuous data in each case.

[47] Figure 4 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 differing degrees 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 top left 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.

[48] In certain cases, when the land signature is indistinguishable from clouds both at the block and pixel levels, *MAIAC* algorithm masks the whole block as cloudy. This creates "blockiness," or mosaic effect in the CM image seen in Figure 4. The blockiness propagates into the downstream products using cloud mask as input, such as aerosol optical thickness and surface reflectance. Generally, these artifacts disappear in the higher level weekly or monthly averaged products. However, consistent blockiness over a particular region may bias the long-term statistics of environmental parameters. So far, the subsets of MODIS data used in the testing showed only a random blockiness without consistent pattern. This issue will be studied further in global tests of the algorithm.

[49] Figure 5 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] but more investigation is necessary. 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.

[50] A large-scale comparison of cloud mask products is shown in Figure 6 for a 1200 × 1200 km^2 region of the African Savannah (Zambia). This is a region of intense biomass burning in 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 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.

[51] A final example of the cloud mask comparison for the large (1800 × 1800 km^2) bright desert area of the Arabian Peninsula is shown in Figure 7 for days 145 and 207 of 2005. Here the *MAIAC* cloud mask is shown in the middle 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.

[52] These examples show that overall the new algorithm is demonstrating a high accuracy of cloud discrimination over land.

6. Conclusions

[53] The cloud mask algorithm described here implements relatively simple and straightforward logic. Using

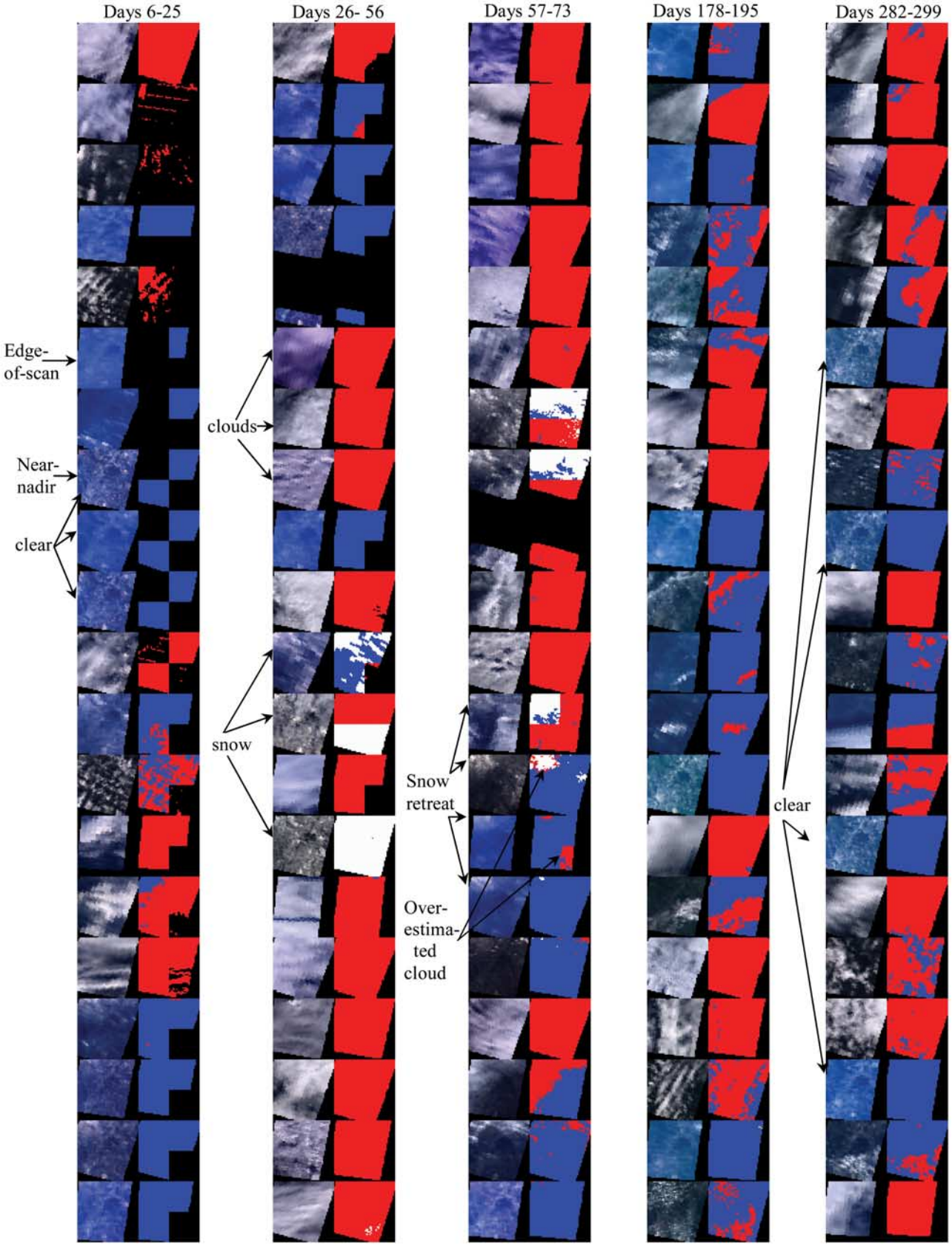

Figure 3. Examples of performance of *MAIAC* CM algorithm for 50 × 50 km^2 MODIS TERRA subsets for 2003, centered at GSFC, USA. Normalized gridded RGB images are shown on the left, cloud mask is shown on the right. CM legend: blue and white, clear over land and snow, respectively; red, cloudy; black, undefined (*Refcm* not initialized).

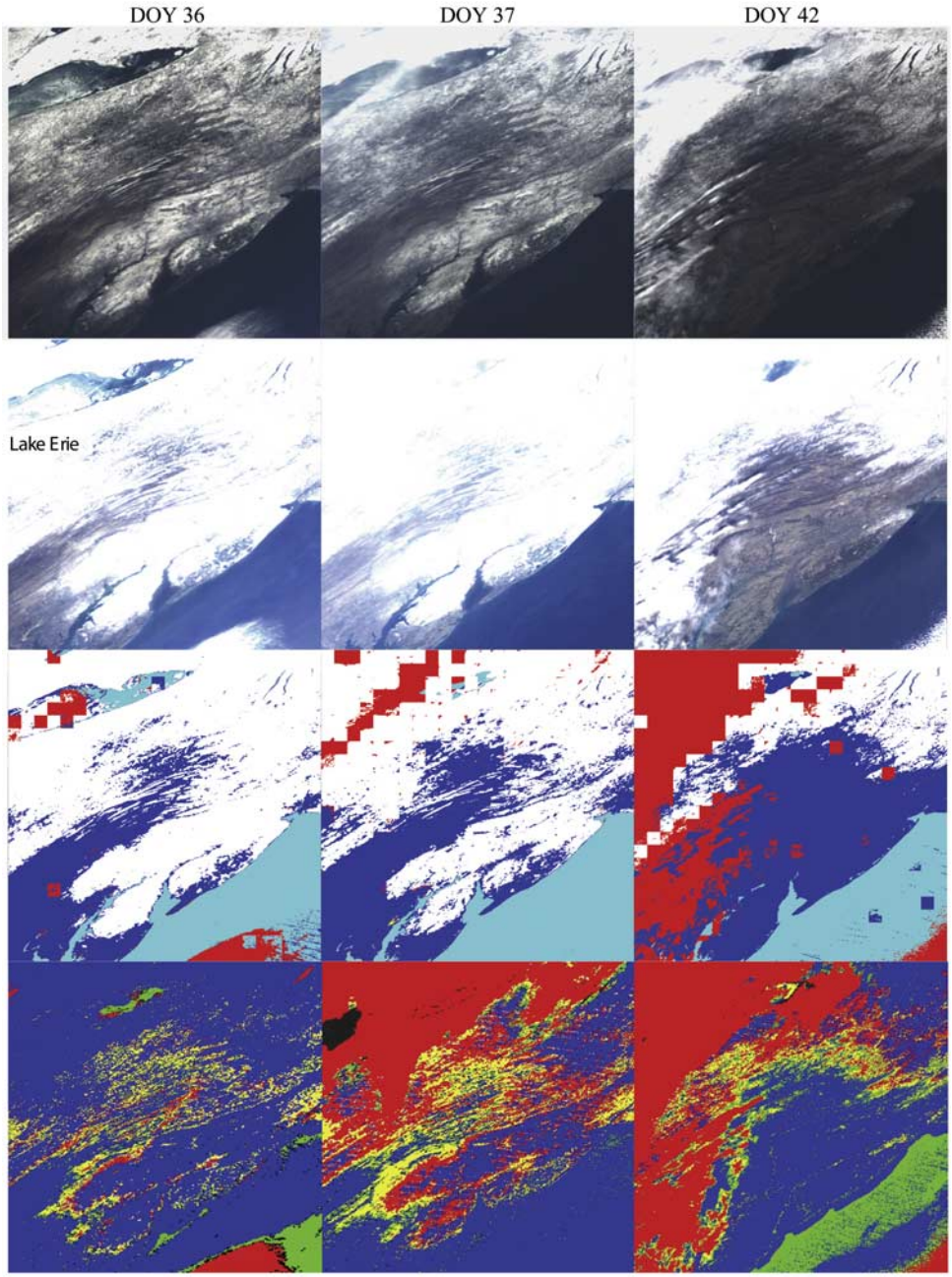

Figure 4. Example of *MAIAC* (third row) and MOD35 (bottom) cloud mask over snow from MODIS TERRA for days 36, 37, and 42 of 2005. The image shows one tile (600 × 600 km^2) for the north–east USA. The two RGB images have a different normalization helping visual distinction between snow and clouds. Legend for MOD35 CM: blue, clear; green, possibly clear; yellow, possibly cloudy; red, cloudy; black, undefined. Legend for MAIAC CM: blue, light blue, and white, clear (land, water, and snow, respectively); yellow, possibly cloudy; red, cloudy.

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 for every block of the land surface. The image- (block-) based processing helps to establish the ground brightness temperature. BT_G and *refcm* reflectance form a comparison target which is used for the pixel-level cloud detection using a simple bright-cold algorithm. A new approach has been developed for snow detection and cloud discrimination over snow.

[54] Because of the covariance component, the new algorithm works successfully over both dark and bright surfaces, including deserts and snow, as well as in both clear and hazy conditions. With dynamic updates of the reference clear skies image, the CM algorithm smoothly adjusts to the seasonal surface variations. Rapid surface change events (fire burns, snow fall, etc.) are accommodated through repetitive reinitialization of *refcm*. Keeping memory of the clear-skies image and of the essential statistical properties of

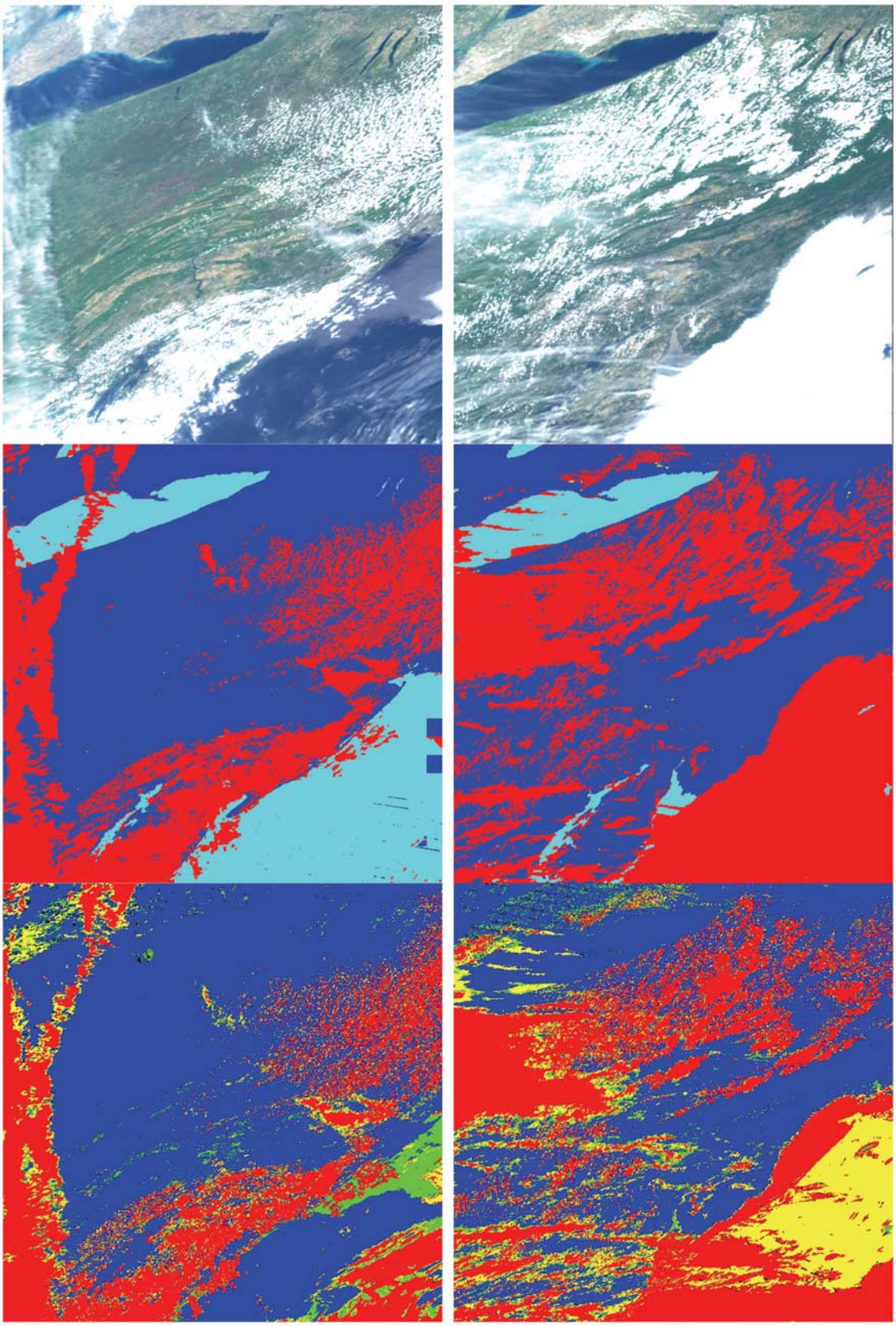

Figure 5. Example of *MAIAC* (center) and MOD35 (bottom) cloud mask from MODIS TERRA data for days 138 (left) and 152 (right) of 2005. The image shows the same tile (north–east USA) as in Figure 4.

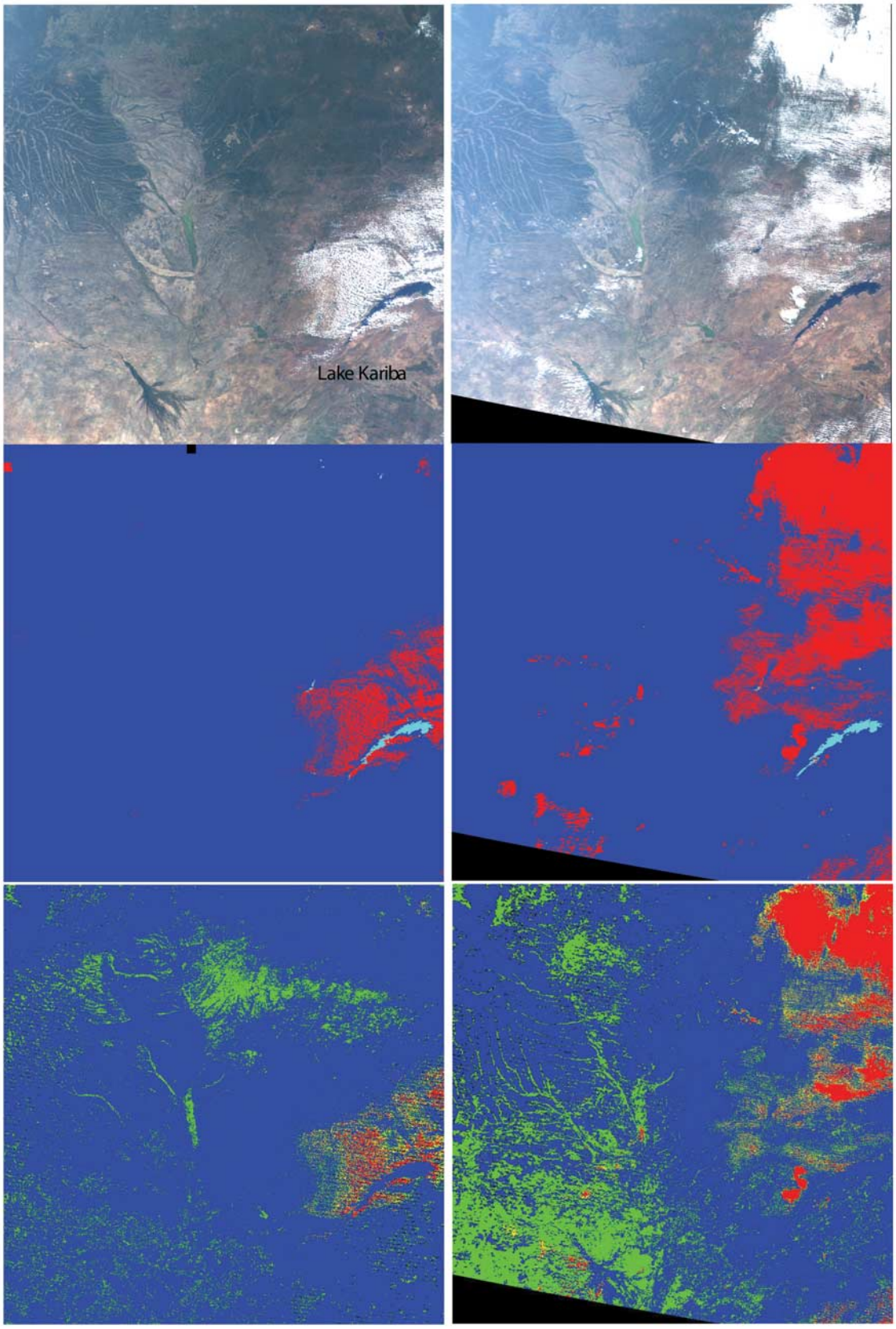

Figure 6. Example of *MAIAC* (center) and MOD35 (bottom) cloud mask at the beginning of dry season for Zambia, Africa, from MODIS TERRA data for days 130 (left) and 141 (right) of 2005. The image shows four tiles (1200 × 1200 km^2).

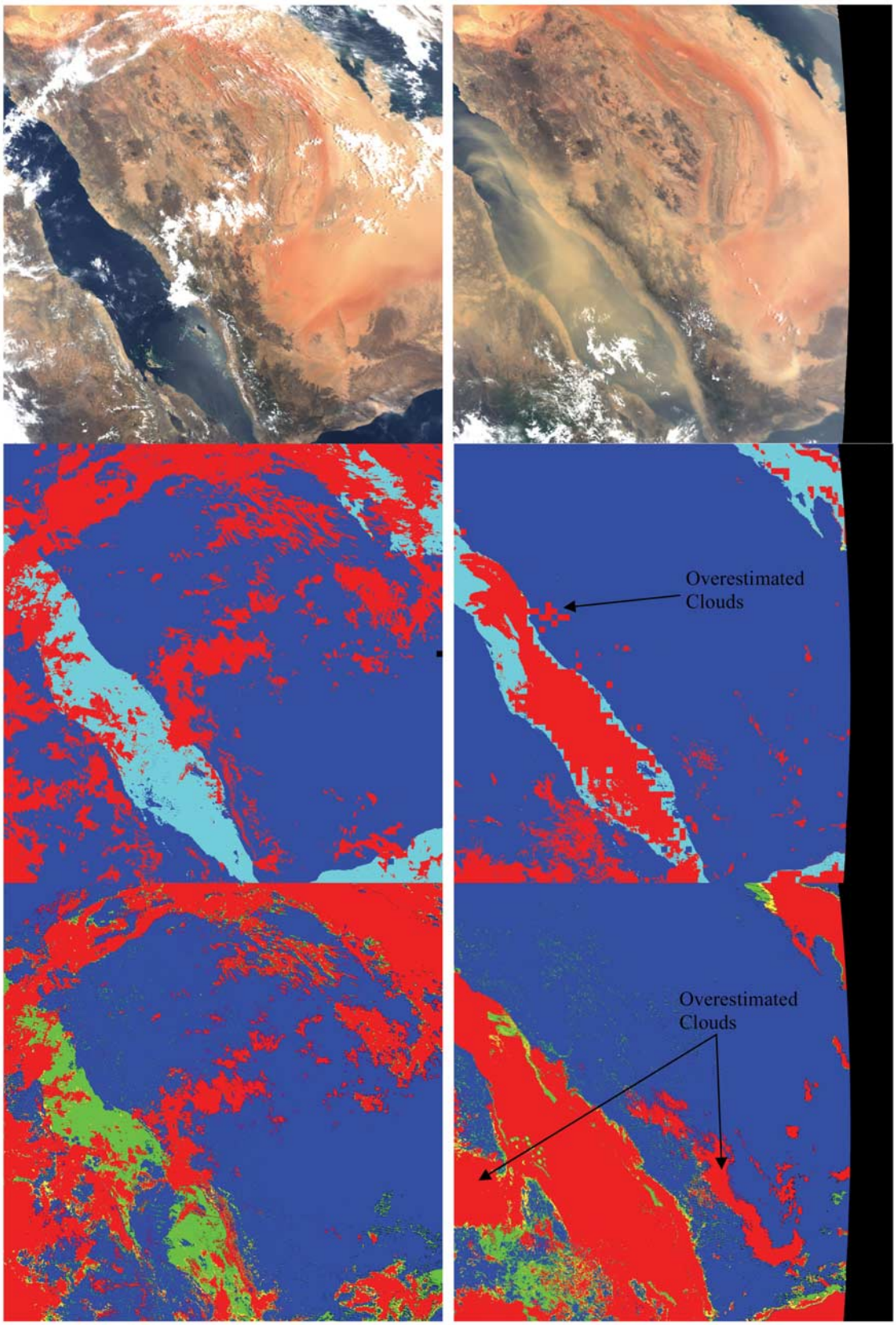

Figure 7. Example of *MAIAC* (top) and MOD35 (bottom) cloud mask for Arabian Peninsula from MODIS TERRA data for days 145 (left) and 207 (right) of 2005. The image shows nine tiles (1800 × 1800 km^2).

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 proves very helpful as it permits adapting the logic of cloud detection as a function of surface type.

[55] 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. *MAIAC* CM does not use the "possibly clear" category because clear surface regions are identified with covariance analysis.

[56] Changing variability of scenes as a function of view zenith angle and size of the footprint 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, data from a geostationary orbit would provide better performance conditions.

[57] The *MAIAC* CM algorithm was developed and optimized for practical needs of aerosol retrievals and atmospheric correction over land. It requires the surface to have some spatial variability, reproducible in the time series of measurements, and it is not applicable over large water bodies. The new algorithm does not mask ubiquitous thin cirrus or other semitransparent clouds that are homogeneous at a scale of 25 km or more, but corresponding MODIS measurements remain useful for the land discipline processing. On the other hand, because of covariance analysis and enhancements from the atmospheric correction algorithm, *MAIAC* CM is sensitive to the subpixel clouds and medium-size semitransparent clouds (<25 km). Unlike conventional CM algorithms, *MAIAC* requires an initialization time to construct *refcm* and acquire necessary statistics of surface reflectance and brightness temperature for each surface block (25×25 km^2) from cloud-free observations. Thus it may take from 2–3 days to several weeks depending on cloudiness for the algorithm to reach its optimal performance. Once *refcm* is initialized, the measurements for the initialization period can be reprocessed to produce a high-quality cloud mask.

[58] The *MAIAC* CM algorithm is still a work in progress. For example, capability to discriminate thick (opaque) aerosol plumes from clouds has not yet been developed. The algorithm has not been tested over shallow waters, which may display a relatively stable spatial pattern of reflectance due to depth variation. A global analysis is needed in order to fully optimize the spectral and brightness temperature tests. We will continue analysis of the information content of other MODIS channels to enhance *Refcm* structure for better cloud discrimination in partly cloudy conditions, especially over snow.

[59] The *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. Work has been initiated on a global intercomparison of *MAIAC* and MOD35 cloud mask in collaboration with the University of Wisconsin.

Appendix A: Spectral Tests and Dynamic Surface Classification

[60] The *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 the snow test does not necessarily distinguish between the snow/ice clouds and the snow, the snow detection uses a different logic described in section A2.

A1. Detection of Vegetation and Water

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

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

[63] 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.

[64] 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.

[65] 2. Water test: $r2 < 0.07$ and $r5 < 0.02$ and $r7 < 0.015$ and $NDVI \leq 0.2$, $\Rightarrow$ CM_CLEAR_WATER.

[66] 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 precalculated using the *Nakajima and Tanaka* [1983] model with mutual shadowing of waves and stored in a LUT. The algorithm can use a real time wind speed, if it becomes known operationally from independent sources.

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

[68] 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 (*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:

$$\mathrm{BT} > \mathrm{BT}_W - 1\mathrm{K},\ \text{AND r5} < \mathrm{r}_{glint} + 0.03,$$
$$\Rightarrow \text{CM_CLEAR_WATER},$$

or detect clouds if

$$\mathrm{BT} < \mathrm{BT}_W - 1\mathrm{K},\ \text{AND r5} > \mathrm{r}_{glint}, \Rightarrow \text{CM_CLOUD}.$$

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

$$\mathrm{BT} > 272.5\mathrm{K},\ \text{AND r5} < \mathrm{r}_{glint} + 0.03, \Rightarrow \text{CM_CLEAR_WATER}.$$

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

$$\mathrm{r1} < 0.12 \text{ AND } |r1 - r2| < 0.015, \Rightarrow \text{CM_CLEAR_WATER}.$$

A2. Detection of Snow

[69] Snow detection begins with the snow test:

[70] (3) Snow test: $NDSI = (\mathrm{r4} - \mathrm{r6})/(\mathrm{r4} + \mathrm{r6}) \geq 0.35$ 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 293 K (to avoid false snow detection in summer), and for the pixels inside the block with BT < 283 K over land or BT < 275 K 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 imagery, is usually darker than the specified threshold (r7 $\approx$ 0.03–0.09), although fresh snow and some types of snow/ice, for example on the high elevation slopes of Greenland, can be as bright as r7 $\approx$ 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 (Appendix B).

[71] 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 midlatitudes 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, partially snow-covered pixels are particularly difficult and are often misclassified as cloud.

[72] To filter out clouds that pass the snow test, we adopted a conservative approach whereby a pixel can be masked as snow for the first time only during initialization or reinitialization of *refcm*. The requirement of high covariance and a carefully designed set of filters, described in section 3.1 (item 7), is usually effective in separating clouds from snow. For snow, we reduce the high covariance threshold to 0.65 in order to accommodate the transient nature of snow in midlatitudes. 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 a Tile may change anytime while the Tile remains in the Queue. This postprocessing modification always increases the number of CM_CLEAR or CM_CLEAR_SNOW pixels. Once the snow is recorded at the Q-memory (in *mask_LWS*), the algorithm returns to the binary pixel-level logic of snow confirmation which is described in Appendix B.

[73] This 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 midlatitudes during the summer and in tropical regions generally, although it cannot completely eliminate this error. Despite being a conservative algorithm, our limited testing indicates that it still finds significantly more clear snow pixels than the MODIS cloud mask algorithm (MOD35) which is known to overestimate cloudiness over snow (D. Hall, NASA GSFC, personal communication of the first author, 2008).

Appendix B

B1. Module *CM_highCov*

[74] 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:

$$\text{IF } \left(\mathrm{BT}_{ij} < \mathrm{BT}_G - 4\right) \text{ AND } \left(\mathrm{r1}_{ij} > refcm.\mathrm{r1}_{ij} + 0.05 + dif\right)$$
$$\Rightarrow \text{CM_PCLOUD}.$$

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 remaining 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 warmest pixels of the block from a histogram analysis.

[75] In the presence of snow, amounts of 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 the warmest pixels, which are assumed to be cloud-free, $BT_{min} = \min\{BT(0.9)\} - 2$, and masks pixels which are colder than BT_{min} as CM_PCLOUD. The bright pixels of intermediate temperature between BT_{min} and 280 K are masked as CM_CLEAR_SNOW, and the value of the *mask_LWS* is set to MASK_SNOW. Those 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,
$mask_LWS_{ij} =$ 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 contain 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 the *refcm* image is high, these newly introduced snow pixels have a low error of false detection.

B2. Module *CM_lowCov*

[76] 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. Further processing depends on whether snow was previously detected (in *mask_LWS*). If not, then the algorithm uses the bright-cold algorithm to mask clouds:

IF $\left(BT_{ij} < BT_G - 4\right)$ AND $\left(r1_{ij} > refcm.r1_{ij} + 0.05\right)$
$\Rightarrow$ CM_PCLOUD.

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

[77] 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 the *snow* test, (2) snow was detected previously (*mask_LWS* = MASK_SNOW), and (3) its reflectance is close to the *refcm* value, $|r1_{ij} - refcm.r1_{ij}| \leq 0.05$. The last conservative requirement, which assumes rather stable snow conditions, is essential because otherwise a significant number of snow/ice clouds end up detected as snow.

[78] If any of the three conditions is not satisfied, and the pixels' band 1 reflectance exceeds 0.12, it is masked as CM_CLOUD. For this reason, darker surfaces partially covered by snow are often masked as cloud. This is a common misclassification error typical of all snow detection algorithms.

[79] **Acknowledgments.** 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 helpful discussion and suggestions on this paper.

References

Ackerman, S. A., K. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Guclusteringey (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, and H. Zhang (2006), 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., et al. (1998), AERONET-A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, *66*, 1–16.

Lyapustin, A., and Y. Wang (2007), *MAIAC* - MultiAngle Implementation of Atmospheric Correction for MODIS. Algorithm Theoretical Basis Document, 69 pp. (Available at: http://neptune.gsfc.n asa.gov/bsb/subpages/index.php?section = Projects&content = SHARM, section MAIAC ATBD)

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. Clim.*, *6*, 2341–2369.

Schaaf, C. B., et al. (2002), First operational BRDF, albedo nadir reflectance products from MODIS, *Remote Sens. Environ.*, *83*, 135–148.

Wolfe, R. E., D. P. Roy, and E. Vermote (1998), MODIS land data storage, gridding, and compositing methodology: Level 2 grid, *IEEE Trans. Geosci. Remote Sens.*, *36*, 1324–1338.

R. Frey, Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin, 1225 West Dayton Street, Madison, WI 53706, USA.

A. Lyapustin and Y. Wang, NASA Goddard Space Flight Center, Mail code 614.4, Bld. 33, Greenbelt, MD 20771, USA. (alexei.i.lyapustin@nasa.gov)