MAIAC: MULTI-ANGLE IMPLEMENTATION OF

ATMOSPHERIC CORRECTION

FOR MODIS

ALGORITHM THEORETICAL BASIS DOCUMENT

(ver. 1.0)

Alexei Lyapustin and Yujie Wang

Goddard Earth Science and Technology Center UMBC, and

NASA GSFC, mail code 614.4 (Alexei.I.Lyapustin@nasa.gov)

November 12, 2008

TABLE OF CONTENT

Abstract. This document describes a new *MAIAC* algorithm for the aerosol retrievals and atmospheric correction of MODIS data over land. This algorithm explores advantages of the time series processing at synergistic level for cloud masking and aerosol-surface retrievals. The algorithm is generic and works globally over all surface types although aerosols are not currently retrieved over snow. *MAIAC* products include cloud mask, water vapor, aerosol optical thickness (AOT) at 0.47 μm and Angstrom parameter, surface spectral bidirectional reflectance factor (BRF), instantaneous BRF (*iBRF*), which is a specific reflectance for a given observation geometry, and albedo for MODIS land bands 1-7, and ocean bands 8-14L, which are not saturated over land. The BRF and albedo are derived from the time series of measurements, whereas the *iBRF* is derived from the last measurement only. All products are generated uniformly at 1 km resolution in gridded format. The suit of products is compliant with energy conservation principle in a sense that the radiative transfer computations with retrieved parameters reproduce measurements with high accuracy.

The cloud mask, aerosol retrievals, and atmospheric correction are completely new algorithms free from conventional assumptions that generally limit the accuracy of products. *MAIAC* algorithm uses up to 16 days of gridded MODIS measurements to make simultaneous retrievals of AOT and surface BRF/albedo. A requirement of consistency of the time series of retrieved BRF provides an additional constraint enhancing quality of aerosol and surface retrievals.

This ATBD describes the radiative transfer basis, and theoretical basis for the water vapor, cloud mask, aerosol and atmospheric correction algorithms. We provide examples of validation of water vapor and AOT with AERONET measurements, and initial comparisons of cloud mask, AOT and surface reflectance with MODIS operational products MOD35, MOD04 and MOD09/MOD43, respectively, for different regions of the world.

1. Introduction

The Earth Observing System [*NASA*, 1999] initiated high quality global Earth observations and operational aerosol retrievals and atmospheric correction over land. With the wide swath (2300 km) of MODIS instrument, the MODIS *Dark Target* algorithm [*Kaufman et al.*, 1997; *Remer et al.*, 2005; *Levy et al.*, 2007] currently complemented with the *Deep Blue* method [*Hsu et al.*, 2004] provides daily global view of planetary atmospheric aerosol. The MISR algorithm [*Martonchik et al.*, 1998; *Diner et al.*, 2005] makes high quality aerosol retrievals in 300 km swaths covering the globe in 8 days. The MODIS *Dark Target* method is the basis of MODIS Atmospheric Correction (AC) algorithm [*Vermote et al.*, 2002].

With MODIS aerosol and land programs being very successful, some algorithm issues remain unresolved. The current processing is pixel-based and relies on a single-orbit data. Such an approach produces a single measurement for every pixel characterized by two main unknowns, aerosol optical thickness (AOT) and surface reflectance (SR). This lack of information constitutes a fundamental problem of the remote sensing which cannot be resolved without a priori information. For example, MODIS *Dark Target* algorithm makes spectral assumptions about surface reflectance, whereas the *Deep Blue* method uses ancillary global database of surface reflectance. Both algorithms as well as MODIS AC algorithm assume a Lambertian surface model.

The surface-related assumptions in the aerosol retrievals may affect subsequent atmospheric correction in unintended way. For example, the *Dark Target* algorithm uses an empirical relationship to predict SR in the Blue (B3) and Red (B1) bands from the 2.1 μm channel (B7) for the purpose of aerosol retrieval. Obviously, the subsequent atmospheric correction will produce the same SR in the red and blue bands as predicted, i.e. an empirical function of $\rho_{2.1}$. In other words, the spectral, spatial and temporal variability of surface reflectance in the Blue and Red bands appears "borrowed" from band B7. This may have certain implications for the vegetation and global carbon analysis because the chlorophyll-sensing bands B1, B3 are effectively substituted in terms of variability by band B7, which is sensitive to the plant liquid water.

This ATBD describes a new recently developed generic aerosol-surface retrieval algorithm for MODIS. The Multi-Angle Implementation of Atmospheric Correction (*MAIAC*) algorithm simultaneously retrieves AOT and surface bi-directional reflection factor (BRF) using the time series of MODIS measurements.

MAIAC starts with accumulating 3 to 16 days of calibrated and geolocated level 1B (L1B) MODIS data. The multi-day data provide different view angles, which are required for the surface BRF retrieval. The MODIS data are first gridded to 1 km resolution in order to represent the same surface footprint at different view angles. Then, the algorithm takes advantage of the following properties of the atmosphere-surface system: 1) the surface reflectance changes little during accumulation period, and 2) AOT changes little at short distances (~25 km), because aerosols have a mesoscale range of global variability of ~50-60 km [*Anderson et al.*, 2003]. Under these generic assumptions, the system of equations becomes over-defined and formally can be resolved. Indeed, we define the elementary processing area as a block with the size of $N\sim25$ pixels (25 km). With K days in the processing queue, the number of measurements exceeds the number of unknowns

$$KN^2 > K+3N^2 \text{ if } K>3, \tag{1.1}$$

where K is the number of AOT values for different days, and 3 is the number of free parameters of the Li-Sparse Ross-Thick (LSRT) [*Lucht et al.*, 2000] BRF model for a pixel.

To simplify the inversion problem, the algorithm uses BRF, initially retrieved in B7, along with an assumption that the shape of BRF is similar between the 2.1 μm and the Red and Blue spectral bands:

$$\rho_{ij}^{\lambda}(\mu_0,\mu;\varphi) = b_{ij}^{\lambda}\rho_{ij}^{B7}(\mu_0,\mu;\varphi). \tag{1.2}$$

The scaling factor b is pixel-, wavelength-, and time-dependent. This physically well-based approach reduces the total number of unknown parameters to $K+N^2$. Below, factor b is called spectral regression coefficient (SRC).

The assumption (1.2) of similarity of the BRF shape is robust for most landcover types because the surface absorption coefficient, or inversely, surface brightness, is similar in the visible and shortwave infra-red (SWIR) spectral regions, and because the scale of macroscopic surface roughness, which defines shadowing, is much larger than the wavelength [*Flowerdew and Haigh*, 1995]. The accuracy of assumption (1.2) decreases when the visible and SWIR surface reflectances are significantly different. This issue will be discussed in detail in sec. … One obvious exception is snow, which is very bright in the visible wavelengths and dark in the SWIR. The principle of spectral similarity of the BRF shape was extensively tested and implemented in ATSR-2 [*Veefkind et al.*, 1998] and MISR [*Diner et al.*, 2005] operational aerosol retrievals.

The *MAIAC* algorithm is based on minimization of an objective function, so it can directly control the assumptions used. For example, the objective function is high if surface changed rapidly or if aerosol variability was high on one of the days. Such days are filtered and excluded from the processing.

From historical prospective, the new algorithm inherits from multiple concepts developed by the MISR science team, from using the rigorous radiative transfer model with non-Lambertian surface in aerosol/surface retrievals [*Diner et al.*, 1999; 2001] to the concept of image-based rather than pixels-based aerosol retrievals [*Martonchik et al.*, 1998]. The latter idea, in a different implementation, was proposed in the *Contrast Reduction* method by *Tanre et al.* [1988], who showed that consecutive images of the same surface area, acquired on different days, can be used to evaluate the AOT difference between these days.

MAIAC is a complex algorithm which includes water vapor retrievals, cloud masking, aerosol retrievals and atmospheric correction. The separate processing blocks are interdependent: they share the data through the common algorithm memory and may update each other's output. For example, the cloud mask is updated during both aerosol retrievals and atmospheric correction. Section 2 of this ATBD provides an overview of *MAIAC* processing. Water vapor algorithm is described in section 3. Section 4 presents the radiative transfer basis for the aerosol retrievals and atmospheric correction algorithm, which are described in sections 5-6, respectively. Section 7 describes *MAIAC* cloud mask algorithm. Validation of *MAIAC* AOT with AERONET measurements and surface BRF/albedo with AERONET-based Surface Reflectance Validation Network (*ASRVN*) products is described in Section 7. Finally, section 8 presents examples of *MAIAC* performance, discussion of operational implementation with parallel processing, and memory requirements. The ATBD is concluded with a summary.

2. *MAIAC* Overview

The block-diagram of *MAIAC* algorithm is shown in Figure 1.

1) The received L1B data are gridded, split in 600 km Tiles, and placed in a Queue with the previous data. The size of the Tile is selected to fit the operational memory of our workstation. As a reminder, MODIS uses 1000 km Tiles in operational processing. In order to limit variation of the footprint with changing view zenith angle (VZA), the resolution is coarsened by a factor of 2. For example, the grid cell size is 1 km for MODIS 500m channels B1-B7. We use the MODIS land gridding algorithm [*Wolfe et al.*, 1998] with minor modifications that allow us to better preserve the anisotropy of signal in the gridded data when measured reflectance is high, for example over snow, thick clouds or water with glint.

2) The column water vapor is retrieved for the last Tile using MODIS near-IR channels B17-B19 located in the water vapor absorption band 0.94 μm. This algorithm is a modified version of [*Gao and Kaufman*, 2003]. It is fast and has the average accuracy of ±5-10% over the land surface (see sec. 5). The water vapor retrievals are implemented internally to exclude dependence on other MODIS processing streams and unnecessary data transfers.

3) The time series of measurements helps to develop a high quality cloud mask (CM). It is based on the notion that the surface spatial pattern is stable and reproducible in the short time frame in cloud-free conditions, whereas clouds randomly disturb this pattern. The algorithm uses covariance analysis to identify cloud-free blocks. On this basis, it builds a reference clear-skies

image of the surface, which is further used in the pixel-level cloud masking. The *MAIAC* CM algorithm has an internal land-water-snow dynamic classification, which guides the algorithm flow.

Queue of K days

```
New          1. Grid L1B Data         2. TP: Retrieve      3. QB, TP: Cloud      Is B7 BRF
Granule  →   and Split in Tiles   →   Water Vapor      →   Mask (Covariance- →   known?
                                      (NIR Algorithm)      Based Algorithm)
```

No ↓   Yes ↓

Use MODIS *Dark Target* Algorithm

```
7. QP: Retrieve BRF and    6. TP: Retrieve      5. TP: Retrieve B3 AOT    4. QB: Retrieve Spectral
albedo in reflective bands ← Angstrom exponent ← using known surface   ← Regression Coefficients
(atmospheric correction)                         BRF, ρB=ρ7bB              in Blue band B3 (bB).
```

Figure 1. Block-diagram of *MAIAC* algorithm. The initial capital letters 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 (T) or using the full time series of the Queue (Q).

4) The *main algorithm* simultaneously retrieves the block-level AOT for *K*-days and N^2 values of the spectral regression coefficient b_{ij} for the Blue (B3) band. This algorithm turns on when the B7 BRF is known. Otherwise, *MAIAC* implements a simplified version of the MODIS *Dark Target* algorithm.

5) The AOT computed in the previous step has a low resolution of 25 km. On the other hand, knowledge of SRC provides the Blue band BRF from Eq. 2 at a grid resolution. With the surface boundary condition known, the Blue band AOT in this step is retrieved at high 1 km resolution from the last *Tile*.

6) The ratio of volumetric concentrations of coarse-to-fine aerosol fractions (schematically called *Angstrom exponent*) is calculated for the last *Tile* at the grid resolution. This parameter selects the relevant aerosol model and provides spectral dependence of AOT for the atmospheric correction. The AOT and Angstrom parameter retrievals are done simultaneously, which is indicated by two arrows between processing blocks 5 and 6.

7) Finally, surface BRF and albedo are retrieved at grid resolution from the *K*-day Queue for the reflective MODIS bands.

2.1 Implementation of Time Series Processing

The *MAIAC* processing uses both individual grid cells, also called pixels below, and fixed-size (25×25 km^2) areas, or blocks, required by the cloud mask algorithm and SRC retrievals. In order to organize such processing, we developed a framework of C++ classes and structures (algorithm-specific Containers). The class functions are designed to handle processing in the various time-space scales, for example at the pixel- *vs* block-level, and for a single (last) day of

measurements *vs* all available days in the Queue, or for a subset of days which satisfy certain requirements (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.

The structure of the Queue is shown schematically in Figure 2. For every day of observations, MODIS measurements are stored as Layers for reflective bands 1-13 for AC algorithm. The Queue stores the retrieval results required for atmospheric correction, such as water vapor and aerosol information. *MAIAC* uses bands 17-19 for water vapor retrievals, and bands 26, 31-32 for the cloud mask algorithm. These data, however, are not stored in the Queue, and the respective memory is released once the last *Tile* is processed. Besides storing gridded MODIS data (*Tiles*), the Queue has a dedicated memory (Q-memory) which accumulates ancillary information about every block and pixel of the surface for the cloud mask algorithm (*Refcm* data structure). It also keeps information related to the history of previous retrievals, for example spectral surface BRF parameters and albedo. Given the daily rate of MODIS observations, the land surface is a relatively static background. Therefore, knowledge of the previous surface state significantly enhances both the accuracy of the cloud detection, and the quality of atmospheric correction, for
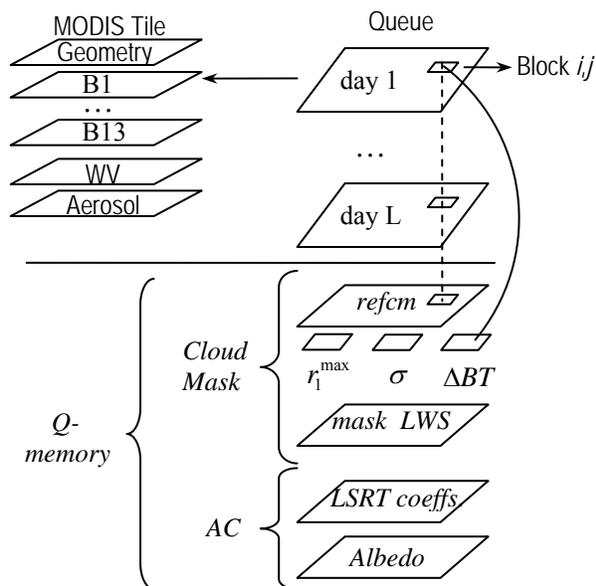
Figure 2. Queue Structure.

The Queue, designed for the sliding window algorithm, stores up to 16 days of gridded MODIS observations at 1 km resolution. The data are stored as Layers (double-indexed arrays) shown in the upper-left corner. A dedicated Q-memory is allocated to store the ancillary information for CM algorithm, such as a reference clear-skies image (*refcm*), block-level statistical parameters {$r1_{max}$; σ_1; ΔBT}, and results of dynamic Land-Water-Snow classification (*mask_LWS*). This information is updated with latest measurements (day L) once given block is found cloud-free, thus adapting to changing surface conditions. The Q-memory also stores results of previous reliable BRF retrievals for MODIS bands 1-13.

example, by imposing a requirement of consistency of the time series of BRF and albedo.

As frequency of MODIS observations increases polewards, the number of days (*Ndays*) stored in the Queue is a function of latitude. *Ndays*=16 at the equator decreasing to *Ndays*=6 (up to 30 daytime observations) at the poles. Because the memory requirement is high, the size of the *Tile* can be scaled to fit an operational memory of a particular workstation (e.g., 300-1000 km). The only requirement for the *Tile* size is that it should be a multiple of the block size (25 km).

3. Water Vapor Algorithm

Selected water vapor algorithm is a two-channel ratio algorithm based on the look-up tables. The column water vapor (*W*) retrievals are made over the land surfaces and the in-land or coastal waters with glint. The algorithm uses three MODIS water vapor channels with the following band center and width (nm): 17 (905, 30), 18 (936, 10), 19 (940, 50). The absorption is highest in band 18 and decreases in bands 19 and 17.

3.1 Theoretical Basis

The water vapor retrievals are based on the approximate Lambertian formula for the TOA radiance:

$$R_\lambda(\mu_0,\mu,\varphi) \cong R_\lambda^D(\mu_0,\mu,\varphi) + \frac{T_\lambda^{tot}(\mu_0,\mu)}{1-q_\lambda(\mu_0)c_{0,\lambda}}\rho_\lambda(\mu_0,\mu,\varphi), \qquad (3.1)$$

where R_λ^D is a path reflectance, $T_\lambda^{tot}(\mu_0,\mu) = \left\langle T^\downarrow(\mu_0)T^\uparrow(\mu)\right\rangle_\lambda$ is a total two-way atmospheric transmittance, $c_{0,\lambda}$ is spherical albedo of atmosphere, and $\rho_\lambda(\mu_0,\mu,\varphi)$, $q_\lambda(\mu_0)$ are surface BRF and albedo. The terms R_λ^D and $T_\lambda^{tot}(\mu_0,\mu)$ are spectrally integrated with the specific RSR of a given channel as described in section 4. The use of Lambertian approximation in this case is justified by the fact that the diffuse atmospheric transmission is much smaller in the 0.94 μm region than the direct transmission because the aerosol optical thickness is usually small and water vapor absorbs the diffuse light stronger than it absorbs the direct (unscattered) light. These factors reduce the weight of the diffuse radiance in the total signal, as well as the error due to its approximate modeling. For the same reason, the multiple scattering of light between the surface and the atmosphere can also be omitted, which gives:

$$R_\lambda(\mu_0,\mu,\varphi) \cong R_\lambda^D(\mu_0,\mu,\varphi) + T_\lambda^{tot}(\mu_0,\mu)\rho_\lambda(\mu_0,\mu,\varphi). \qquad (3.2)$$

If the surface reflectance is spectrally flat or changes little in a narrow spectral absorption interval of 0.9-0.94 μm, then a two-channel ratio algorithm can be used to derive *W*:

$$\frac{T_{18}^{tot}}{T_{19}^{tot}} = \frac{R_{18}-R_{18}^D}{R_{19}-R_{19}^D}, \text{ and } \frac{T_{19}^{tot}}{T_{17}^{tot}} = \frac{R_{19}-R_{19}^D}{R_{17}-R_{17}^D}. \qquad (3.3)$$

For a given view geometry, the solution is found by searching the LUT.

The MODIS water vapor channels have different effective absorption and different sensitivity under the same atmospheric conditions. The strong absorption channel 936 nm is most sensitive under dry conditions, while the weak absorption channel at 905 nm is most sensitive under humid conditions [*Gao and Kaufman,* 2003]. Given the atmospheric conditions, the derived *W* from the two ratios (Eq. 3.3) can be different. A mean water vapor is obtained as in [*Gao and Kaufman,* 2003]:

$$W = f_1 W_1 + f_2 W_2, \qquad (3.4)$$

where W_i are water vapor values derived from different channel ratios, and f_i are weighting functions. The weighting functions are related to the sensitivities of the two band pairs,

$$\eta_1 = \left| \Delta \frac{T_{18}^{tot}}{T_{19}^{tot}} \middle/ \Delta W \right|, \ \eta_2 = \left| \Delta \frac{T_{19}^{tot}}{T_{17}^{tot}} \middle/ \Delta W \right|, \text{ and are defined as normalized values:}$$

$$f_i = \eta_i/(\eta_1 + \eta_2). \tag{3.5}$$

The weighting functions are computed numerically from the ratios $T_{18}^{tot}(\mu_0, \mu, W)/T_{19}^{tot}(\mu_0, \mu, W)$, $T_{19}^{tot}(\mu_0, \mu, W)/T_{17}^{tot}(\mu_0, \mu, W)$ stored in the LUT.

The subsequent atmospheric correction in bands affected by water vapor absorption is performed using the weighted value of water vapor.

The WV algorithm uses following input data:
-   view geometry (SZA, VZA, relAZ);
-   measured reflectance B17, B18, B19.

The algorithm:
1) Initially, the aerosol parameters are unknown, and the atmospheric path reflectance, subtracted from measurements (in Eq. 3.3), corresponds to the background aerosol level assumed in the LUT calculations as τ^a=0.01 at λ=0.94 μm. The water vapor retrievals are performed over pixels that satisfy the condition: $R_{19} \geq 0.05$, which only excludes dark water in the off-glint region.
2) If the subsequent aerosol retrievals show high AOT at λ=0.94 μm (τ^a>1), then the water vapor retrievals should be repeated with the derived value of AOT. This part of algorithm is not currently implemented.

3.2 Look-Up Tables

The current LUT$_W$ is calculated using a vertical profile of the US 1976 Standard Model of Atmosphere. Given W, the total transmittance ratio is fairly insensitive to variations in the profiles of water vapor, temperature and pressure. The path reflectance in bands B17-B19, on the other hand, depends on the vertical profiles of water vapor and aerosol. Since both profiles are unknown, we assume that the water vapor has a profile of the US1976 Model, and that aerosols are uniformly distributed in the 0-2 km boundary layer, with small constant background level in the stratosphere. This model has a sufficiently high accuracy for atmospheric correction of MODIS spectral bands, carefully selected in the atmospheric windows.

The LUT$_W$ stores path reflectance $R_\lambda^D(\mu_0, \mu, \varphi, W, \tau^a)$ for three bands (B17-B19), and two transmittance ratios $T_{18}^{tot}/T_{19}^{tot}$, $T_{19}^{tot}/T_{17}^{tot}$, which depend on 4 parameters (μ_0, μ, W, τ^a). The LUT$_W$ was calculated with steps $\Delta\mu_0 = \Delta\mu$=0.02 for the range μ=1 - 0.4 (0° - 66.4^o), μ_0=1 - 0.34 (0° - 70.1^o), W=0 - 7.5 with step 0.3, and $\tau_{0.94}^a$={0.01, 0.2, 0.5, 1.0}. The step in azimuthal angle for the path reflectance is selected 10°. The angular resolution of LUT$_W$ is high enough to use nearest neighbor method, avoiding bi- or three-linear interpolation in angles.

3.3 AERONET Validation

To test developed algorithm, we processed 1 year (2003) of MODIS TERRA data subsetted for 156 AERONET stations globally. The AERONET v.2 water vapor algorithm [*Smirnov et al.*, 2004] is based on an accurate high spectral resolution model of atmospheric gaseous absorption developed by A. Lyapustin. This algorithm was validated against GPS retrievals. The comparison study showed an excellent agreement with zero intercept and a slope of 0.984 in the range of W $0-5$ cm [*Smirnov et al.*, 2004].

For the current study, we used AERONET data within ± 30 min of the MODIS TERRA overpass for validation. The retrieved data (W_R) were averaged over 9×9 km^2 region. Because water vapor is retrieved before CM algorithm, the clouds were filtered first using a very simple variance criterion: the point was considered potentially cloudy if the difference between the maximum and the average W_R over the 9×9 km^2 region exceeded 0.5 cm. This criterion, which filtered 12% of all retrievals, relies on a low local spatial variability of the atmospheric water vapor over relatively flat terrain. An additional implicit filter was availability of AERONET measurement for comparison, which is compliant with AERONET cloud mask.

The scatterplots of the retrieved water vapor against AERONET results are shown in Figure 3 for several sites around the globe. The last plot shows the summary comparison for 156 different AERONET sites. Over most AERONET sites, the correlation is very good. In the cloud-free conditions, the retrievals are in general unbiased, and accurate to 5-10%. The retrievals have lower accuracy (20-30%) over areas with red iron-rich soils (Canberra, Australia), whose reflectance changes considerably in the 0.9-1 µm spectral region due to absorption of the iron compounds. Similarly, lower accuracy has been reported in these cases by the operational near-IR water vapor algorithm MOD03 [*Gao and Kaufman*, 2003].

The plots of Figure 3 show that a fraction of retrievals is considerably lower than the AERONET data. The point-by-point analysis shows that the low retrievals are caused by either residual partial cloudiness, which leaked through our simple filter, or high aerosol optical thickness. When present in the atmosphere, clouds raise the effective reflecting boundary, and the algorithm, which is sensitive to the water vapor above the clouds, produces lower W_R.

Unless the surface is very bright, the aerosol scattering usually increases the measured signal and the band ratio, also causing lower retrieved values. Figure 4 shows the time series of the retrieved W_R against AERONET value, and AERONET aerosol optical thickness at $\lambda=1.06$ µm. The correlation between the W_R and AERONET is excellent except when the AOT is high. In the two examples for GSFC (USA) and Beijing (China), the aerosol outbreaks explain over 90% of cases when W_R was significantly lower than the AERONET value. These data show the need for aerosol correction on hazy days. The operational MODIS algorithm MOD03 did not implement this correction.

3.4 Alternative Algorithms

We have also studied two additional algorithms - an operational MODIS near-IR water vapor algorithm (MOD03), and an empirical analytical method, which is used in the current MODIS atmospheric correction algorithm (MOD09). The MOD03 algorithm uses 5 bands, the three water vapor channels and two window channels B2 (0.865 µm) and B5 (1.24 µm). The window channels are used to correct for spectral change of surface reflectance across spectral interval of $0.85 - 1.26$ µm assuming it changes linearly with wavelength. *Gao and Kaufman*, [2003]
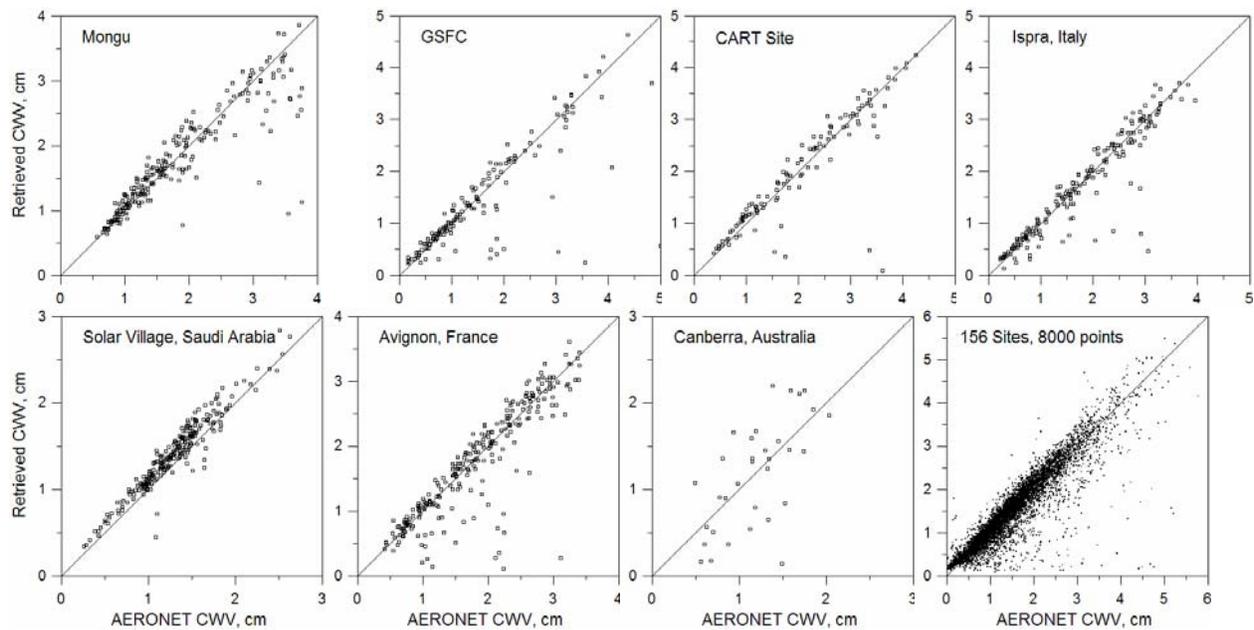
Figure 3. Comparison of retrieved column water vapor (CWV) with AERONET.
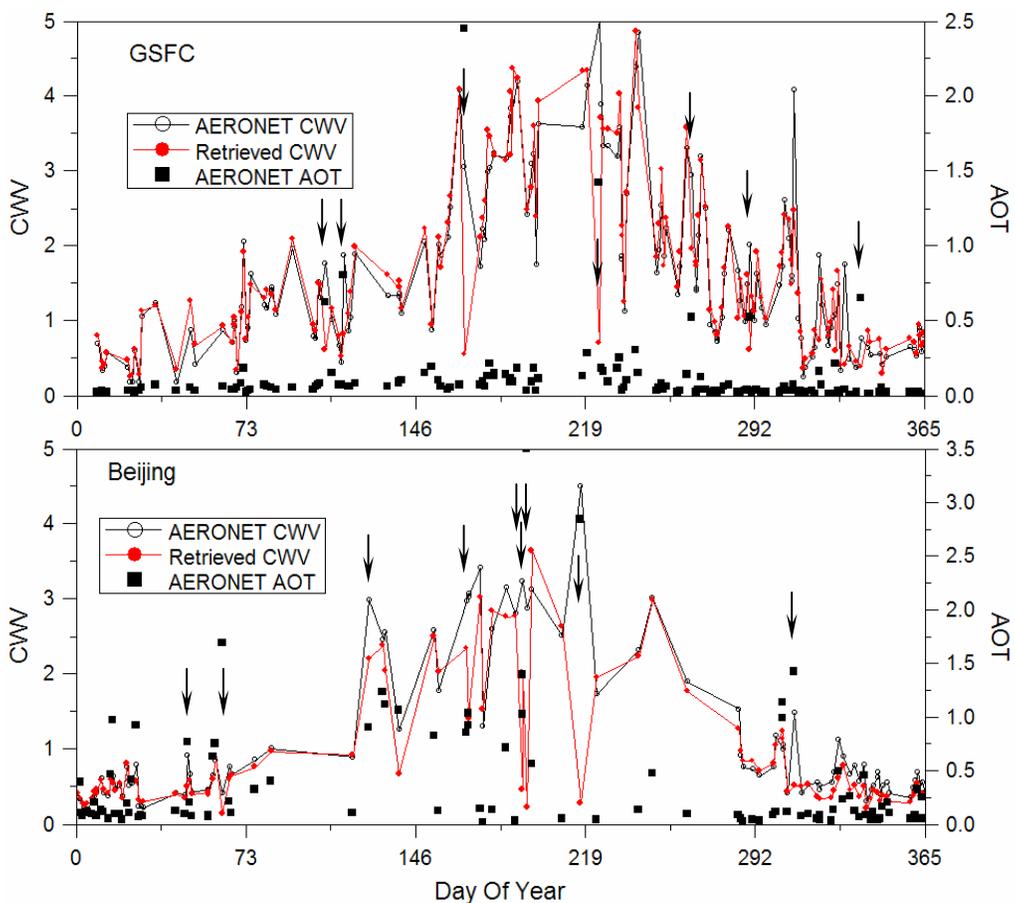
Figure 4. Time series of retrieved and AERONET column water vapor, and AERONET optical thickness at 1.06 μm. The arrows mark the days with high AOT when the retrieved W is significantly lower than the AERONET data.

assessed the accuracy of the linear model as 2.4-3.9% for most soils, rocks, vegetation, and snow, and 8.4% for the iron-rich soils.

Our comparison study with AERONET data showed that this method has a factor of 2-3 higher noise than the method we selected. This noise may be caused by the spectral variability of the surface reflectance. Indeed, if the linear spectral model is not a good predictor for the surface reflectance in $0.85 - 1.26$ µm region, then using the much wider spectral interval is expected to add noise into retrievals.

The second retrieval method used in MOD09 uses an empirical relation:

$$W = (a_1 \ln R + a_2 \ln^2 R)/m, \qquad (3.6)$$

where $\ln R = \ln(R_{18}/R_{19})$, m is airmass, and a_1, a_2 are band-dependent coefficients. This relation was used earlier for the column water vapor retrievals from space [*Bennartz and Fisher*, 2001]. Another empirical relation commonly used for water vapor retrievals from ground-based sunphotometry [*Michalsky et al.*, 1995] has a form:

$$T_{WV}/T_{Window} = \exp(-a[m_{WV}W]^b), \qquad (3.7)$$

where the left-hand side is the direct transmittance ratio in the water vapor absorption band and in the nearby atmospheric window, m_{WV} is the airmass of water vapor, and a, b are coefficients. We studied both equations (3.6) and (3.7). The coefficients were established by the least squares fit of the LUT transmittance ratio $T_{18}^{tot}(\mu_0,\mu,W)/T_{19}^{tot}(\mu_0,\mu,W)$, $T_{19}^{tot}(\mu_0,\mu,W)/T_{17}^{tot}(\mu_0,\mu,W)$. The relative accuracy of the parametric models is shown in Figure 5, where the ratio of LUT transmittance is called *ETratio*, and *Model 1* and *Model 2* represent equations (3.7) and (3.6), respectively. The W increases along the x-axes from 0.05 to 7.5 with step 0.3. At each W value, the zenith sun and view angles change from 0° to 60°. The transmittance ratio and the model
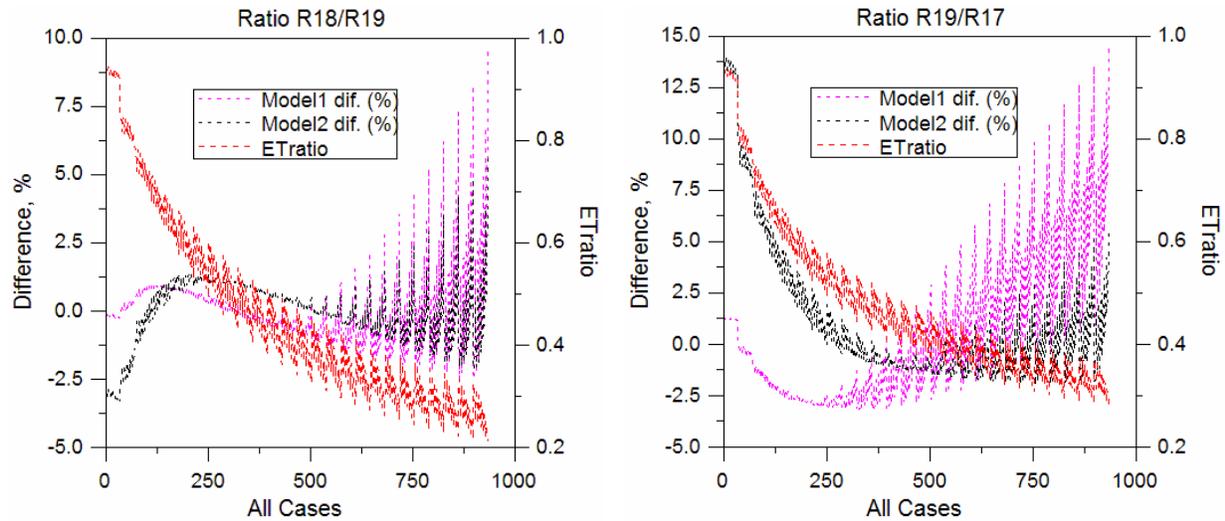
Figure 5. A relative accuracy of parametric models 1 (Eq. 5.7) and 2 (Eq. 5.6) fitting the two-way transmittance ratio for band combinations B18-B19, and B19-B17. The x-axis represents increasing column water vapor from 0.05 (at the left) to 7.5 (at the right), with all combinations of SZA and VZA from the range of 0-60°.

error have the comb-like structure. Each period of this structure corresponds to a fixed *W*-value and all combinations of SZA and VZA, with maximal error at SZA=VZA=60°.

Figure 5 shows that combining the two models, it is possible to sustain the accuracy of parameterization within 3-5%. However, the errors are rather high at large solar and view zenith angles, for which reason we did not pursue the parametric approach any further. Lack of possibility of aerosol correction is another limitation of parametric approach. On the other hand, our MODIS-based analysis using (Eq. 3.6) for water vapor retrievals showed that it works remarkably well, providing similar results with our selected approach and just a slightly lower correlation with AERONET data.

4. *MAIAC* CLOUD MASK ALGORITHM

Updates: Described below is the most recent version of the CM algorithm which was updated from the original version published by *Lyapustin et al.* [2008]. The following updates were made in year 2008:
- Added Cirrus test using MODIS band B26;
- Added cloud detection over water;
- Added B7 reference clear-sky image (the original version used only B1 reference image);
- Improved snow detection and cloud mask over snow;
- Significantly reduced cloud leak through a high covariance processing path;
- Added clear-sky restore algorithm in high covariance processing path.

4.1 Introduction

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.

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 (BT) 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 ground brightness temperature, 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.

MAIAC takes advantage of high observation rate of MODIS providing a daily global view of the Earth at the equator and multiple observations per day in mid-latitude and polar regions. At daily 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. 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. Uncertainty of the reference value, caused by natural variability and sensor noise, is directly calculated from the measurements.

MAIAC CM algorithm [*Lyapustin et al*, 2008] is a next step in the evolution of this idea. It uses the time series and spatial 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 rather comprehensive comparison target for cloud masking. The new algorithm has an internal

surface classifier, producing a dynamic land-water-snow (*LWS*) mask. It is 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 aerosol retrievals and atmospheric correction, which makes it a synergistic component of *MAIAC*.

MAIAC CM produces an integral cloud mask with values of CM_CLEAR for clear conditions and CM_PCLOUD (possibly cloud) or CM_CLOUD for cloudy conditions, and it does not keep values of separate tests. The new algorithm is applicable over land and inland water. It was not designed for the global ocean where it may be less accuracy than the MODIS operational CM algorithm (MOD35). The described algorithm has been updated since the initial version [*Lyapustin et al*, 2008] was reported.

4.2 Reference Clear Skies Image and Overall Logic of Cloud Detection

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,

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

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 of the land surface ($25{\times}25$ km^2 area used for *cov*-analysis). 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 sec.4.4). 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, 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*≥0.68) and several other conditions (sec. 4.4). After initialization, the algorithm uses the *refcm* to compute covariance with the latest measurements. Once 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 4.4).

In order to achieve high accuracy and global performance, *MAIAC* combines spectral pixel-level tests with spatio-temporal analysis involving *refcm*. The logic of CM algorithm is shown in the flowchart of Figure 6. Here, rectangles represent separate functions, diamond shapes stand for the separate subroutines (algorithms), and round-corner rectangles indicate decision (branching) points. The letters in parentheses show spatial and temporal domains of operations, for example
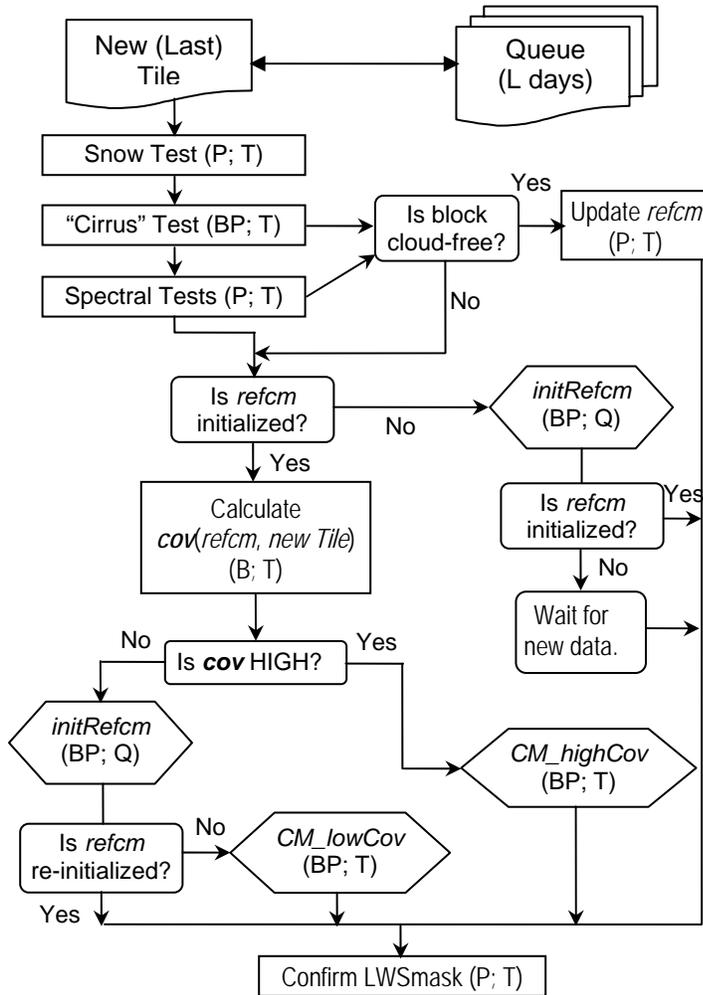
Figure 6. The general flowchart of CM algorithm. After "snow", "Cirrus" cloud and spectral tests, the algorithm first tries to initialize the reference clear-skies image (*refcm*), and if fails then it waits for the new data to arrive. If *refcm* is available, the algorithm calculates covariance between 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 re-initialize *refcm*, which takes care of rapid surface changes as well as of possible errors in *refcm* caused by previously undetected clouds. The pixel-level cloud mask is produced by modules *CM_highCov*, *CM_lowCov*. Module *initRefcm* produces cloud mask only if initialization is successful. The *Refcm* structure is updated after successful spectral tests and in modules *initRefcm* and *CM_highCov*.

Flowchart elements:

- New (Last) Tile
- Queue (L days)
- Snow Test (P; T)
- "Cirrus" Test (BP; T)
- Is block cloud-free? — Yes → Update *refcm* (P; T) — No
- Spectral Tests (P; T)
- Is *refcm* initialized? — No → *initRefcm* (BP; Q) — Yes
- Is *refcm* initialized? — Yes, No → Wait for new data.
- Calculate *cov*(*refcm, new Tile*) (B; T)
- Is *cov* HIGH? — No / Yes
- *initRefcm* (BP; Q)
- *CM_highCov* (BP; T)
- Is *refcm* re-initialized? — No → *CM_lowCov* (BP; T) — Yes
- Confirm LWSmask (P; T)

at pixel- (P) and/or block- (B) level, and using the data of the last Tile (T) only or using the full time series of the Queue (Q).

1. The processing starts with snow test, as its input is required by the *Cirrus* cloud test. The snow test does not uniquely detect cloud-free conditions because snow/ice clouds may have similar spectral features.

2. The *Cirrus* cloud spectral test uses MODIS band 26 to detect high clouds located above absorbing water vapor. This test is often insensitive to the low clouds, and its performance deteriorates when the atmosphere is dry. We developed an "anti-cirrus" test to detect clear areas of snow/ice in winter with low water vapor using band 26. If the block is declared cloud-free, then the *refcm* is updated and the algorithm proceeds to complete the land type classification (*confirm_LWSmask*).

3. The clear-skies spectral tests include a high NDVI test, a hot-bright desert test, and a dark water test. These tests rely on unique spectral signatures of surface and are used to initialize and/or update pixel-level surface classification mask (*LWSmask*). The detected clear land and water pixels are used to calculate brightness temperature of the ground (BT_G) and of the water

17

(BT_W) for every block, which are required in further analysis. If the block is declared cloud-free, then the *refcm* is updated and the algorithm proceeds to complete the land type classification.

Although providing reliable performance, these tests are applicable neither globally nor for all seasons of the year. If the generated cloud mask does not cover the full area of the block, the algorithm uses *refcm* for further processing.

4. The central idea of the *MAIAC* CM algorithm is to use *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 brighter and colder 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 (BC) algorithm for pixel (*i,j*) is formulated as follows:

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

The processing path depends on whether *refcm* (*Refcm* structure) has been initialized. If not, then the algorithm tries to initialize *refcm* (module *initRefcm*). If initialization is unsuccessful because of clouds, the algorithm waits for a new data to arrive. If *refcm* has been initialized, then the covariance is calculated between the new measurements and the *refcm* image. The further processing depends on the value of covariance (modules *CM_highCov* and *CM_lowCov*). If covariance is low, the algorithm takes an intermediate step and attempts to re-initialize *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*. Re-initialization also helps to update *refcm* when the surface reflectance has changed rapidly.

Following covariance calculations, the pixel-level cloud mask is produced by modules *CM_highCov* and *CM_lowCov*. Although the logic implemented in these modules is different, the clouds in the absence of snow are detected similarly using the bright-cold algorithm. Module *initRefcm* generates cloud mask only if initialization/re-initialization is successful. The *refcm* image is updated in the modules *initRefcm* and *CM_highCov*. It is also updated after spectral tests if all pixels of the block are cloud-free. The detail of spectral tests, surface classification scheme and algorithms implemented in modules *initRefcm*, *CM_highCov* and *CM_lowCov* are described next.

4.3 General Information

The *MAIAC* CM algorithm uses five 500 m resolution MODIS bands B1, B2, B4, B5, B7, a 1 km band B26, and 1 km thermal bands B31, B32 (Table 1). As mentioned earlier, data are initially re-projected and gridded to a 1 km grid.

1. The covariance analysis is currently performed for MODIS bands B1 (0.645 μm) and B7 (2.11 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 B1, 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 variable semi-transparent clouds. The darker red band B1 was found to provide better overall performance, especially in BC-test.

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*. BC test. LSC – Vegetation, Water. Land restore test.
B2	0.856	0.5	LSC – Vegetation, Water.
B4	0.554	0.5	LSC – Snow.
B5	1.242	0.5	Shadow detection. Land restore test. High cloud test (with B26).
B7	2.113	0.5	Covariance analysis, *refcm*. LSC – Vegetation, Water, Snow.
B26	1.38	1.	High cloud test. Detection of cloud-free blocks at low WV in atmosphere.
B31	11.030	1.	Covariance analysis, *refcm*. BC test. BT analysis. LSC. Dust detection.
B32	12.030	1.	Dust detection

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

2. The size of a block is 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.

3. The first version of *MAIAC* CM algorithm [*Lyapustin et al.*, 2008] maintained two reference clear-skies images, *refcm1* for VZA$\approx$0-45° and *refcm2* for VZA=45°-60° 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). Our later work showed that such an approach is redundant because a similar performance can be achieved with single *refcm*. The current version maintains one *refcm* which is updated with measurements from the range VZA≈0-50°. On the other hand, if *refcm* (in bands B1 and B7) has been initialized, it is used for covariance analysis with the new measurements for the full range of observation angles.

4. The clear-skies image is stored in the Q-memory in the *Refcm* structure, which also stores the maximal (r_{max}) the mean (r_{av}) and the contrast (r_{max}- r_{min}) values of reflectance in bands B1 and B7 for each block. It also stores the variance (σ) of B1 reflectance and the brightness temperature contrast ($\Delta BT = BT_{max} - BT_{min}$). 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.

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* re-initialization, and during aerosol retrievals and atmospheric correction.

6. The allowed values of the cloud mask are clear (CM_CLEAR, CM_CLEAR_WATER, CM_CLEAR_SNOW, CM_CLEAR_ICE), indicating surface type as well, possibly cloudy (CM_PCLOUD), and confidently cloudy (CM_CLOUD). One more value CM_SHADOW is used for pixels defined as cloud shadowed. 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. An example of *refcm* initialization is shown in Figure 7 for the 150 km area encompassing Lake Erie. Shown are two consecutive 8-day periods from the beginning of processing, May 4-12, and May 20-28 of 2005. MODIS measurements are displayed in columns RGB, R1, R7, R26 and BT(B31). Cloud mask (CM), *refcm* images in bands B1 and B7 (*refcm*1,7) and column water vapor (WV) are results of *MAIAC* processing. Initialization of the reference image starts on day 3. Most of *refcm* over land is initialized by day 8. The green-up of the land during this period of time is accompanied by reduction of reflectance in bands B1, B7. *Refcm* tracks these changes as can be seen at the end of the second 8-day period.

4.4 Clear-skies Spectral Tests and Dynamic Surface Classification

The *MAIAC* algorithm maintains a dynamic land-water-snow mask (*mask_LWS*) which guides the cloud mask algorithm and controls the processing 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 four stable values (MASK_LAND, MASK_WATER, MASK_SNOW, MASK_ICE) 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. The ice test (value MASK_ICE) is applied only over water. In this work, the term "land" implies any land surface other than water or snow.
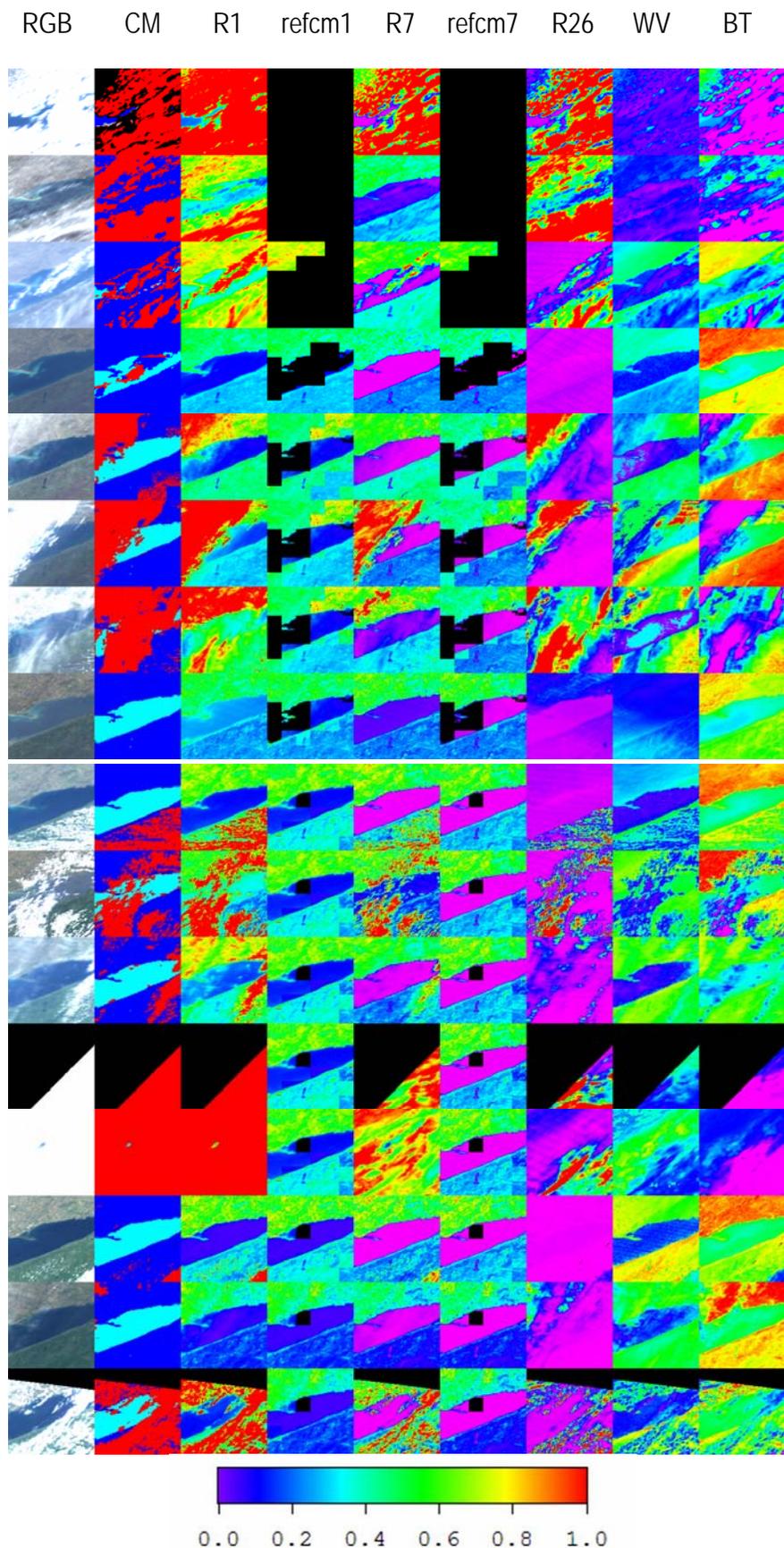
Figure 7. Illustration of *refcm* initialization for 150km area around Lake Erie (USA). Shown are two consecutive 8-day periods from the beginning of processing, May 4-12, and May 20-28 of 2005. Column 1 shows measured MODIS TERRA RGB data followed by *MAIAC* cloud mask. The red, blue and light blue colors of CM show clouds, clear land, and clear water, respectively. Columns 3-6 show measured MODIS reflectance in bands B1 and B7 and respective reference images. Also shown are MODIS measurements in B26 (Cirrus channel) and B31 (BT), and retrieved column water vapor (WV). The water vapor retrievals over water are only valid except in the glint region, and are not used in further processing. Columns 3-9 are drawn using rainbow color palette shown at the bottom for scale 0-1. The respective scales for the columns are: R1, refcm1 (0-0.2), R7, refcm7 (0-0.3), R26 (0-0.05), WV (0-2 cm), BT (260-300). Measurements, exceeding the upper scale limit, are shown in red.
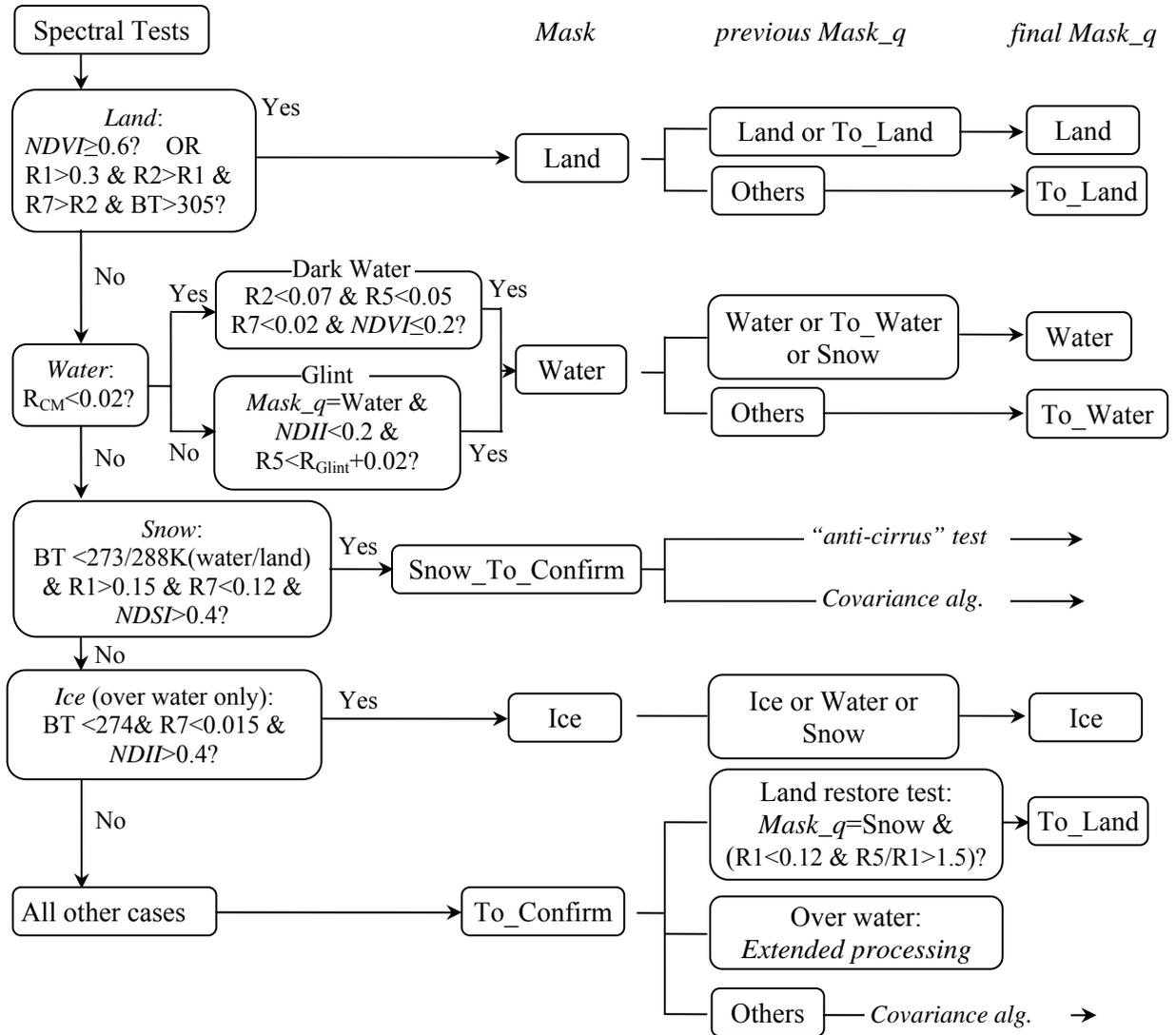
Figure 8. Diagram of spectral clear-skies tests and land type classification. Following spectral tests, the generated *mask* is compared with the previous q.*LWSmask* (*mask_q*) stored at the Queue. When tests are successful, the final mask is produced using the binary logic for the vegetated and water (including the ice-covered) pixels. An *Extended processing* is performed for known water pixels which didn't pass the current *Water* test. For the other pixels (*mask*=MASK_TO_CONFIRM), including those which passed the *Snow* test (*mask*=MASK_SNOW_TO_CONFIRM), the decision is delayed and will depend on the results of covariance analysis.

The classification scheme is shown in Figure 8. Initially, the algorithm performs land (high NDVI, bright hot desert), water, and ice tests. Although the *Snow* test is shown as integrated into Spectral tests of Figure 8, it is performed separately before *Cirrus* cloud test. The generated values of *mask* are compared with the previous value of *mask_q*=q.*LWSmask* stored in the Queue. The classification of vegetation and water is performed at the pixel-level using binary logic which either confirms the previous state of the surface, or detects change. Because the snow tests may not distinguish between the snow/ice clouds and the snow, the snow detection uses a different logic which relies on identification of clear conditions by covariance algorithm.

The snow and ice tests are similar to those developed for the ADEOS Global Land Imager (GLI) [*Stamnes et al.*, 2007], and are empirically adjusted based on MODIS data.

1. *High NDVI test*: $NDVI=(r2 - r1)/(r2 + r1)>0.6$.

If the pixel passes this test, the value of cloud mask is set to CM_CLEAR unless it has CM_CLOUD value from *Cirrus* test. If the previous value of land-water-snow mask (q.*LWSmask*) was MASK_LAND or MASK_TO_LAND, then the value of mask is validated: q.*LWSmask* =MASK_LAND. Otherwise, change is detected: q.*LWSmask* =MASK_TO_LAND. This test finds heavily vegetated pixels of the land. Sparsely vegetated regions, bare soil, rocks, sand etc., are classified as land after covariance analysis when the pixel is found cloud-free and it was not classified as either water or snow.

2. *Hot bright desert test*:   r1>0.3 AND r2>r1 AND r7>r2 AND BT>305

This test uses the fact that spectral reflectance of sand is increasing in the visible through shortwave infrared part of spectrum, whereas reflectance of clouds in the 2 μm region is always lower than that in the visible because of water\ice absorption. The high brightness temperature threshold increases confidence in selecting clear conditions. This test mainly aims at very homogeneous desert regions where covariance analysis is not successful.

3. *Water test*:   r2<0.07 & r5<0.02 & r7<0.015 & *NDVI*≤0.2 & BT>272.8 (dark 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 the *Nakajima and Tanaka* (1983) model with mutual shadowing of waves and is stored in a LUT. The algorithm can use a real time wind speed, if it becomes known operationally from independent sources.

For the known water (q.*LWSmask*=MASK_WATER), a separate test is conducted in the glint region:

$$NDII< 0.2 \text{ \& } r5<r_{glint}+0.02 \text{ \& } BT>272.8,$$

where Normalized Difference Ice Index *NDII*=(r5-r2)/(r5+r2).

In a similar manner as above, the value of cloud mask is set to CM_CLEAR unless *Cirrus* test has detected a cloud, and the value MASK_WATER is either validated with a new measurement or change is detected (MASK_TO_WATER).

Different factors may lead to failure of the *Water* test over an actual water, including elevated aerosol loading, undetected clouds, transition to land (drying of shallow water, drainage) etc. For this reason, an *Extended* processing is performed for the pixels which were previously classified as water (q.*LWSmask*=MASK_WATER) but did not pass the *Water* test this time (*mask*=MASK_TO_CONFIRM or MASK_SNOW_TO_CONFIRM). It starts with evaluation of the water brightness temperature BT_W for the block from the detected clear water pixels. Then, the land-restore test (r1<0.12 AND r5/r1>1.5) checks the possible transition to land for the off-glint viewing geometries. These pixels are declared CM_CLEAR and the value of mask changes to q.*LWSmask*=MASK_TO_LAND. For the rest of pixels, the water brightness temperature of the block is used either to confirm the clear water pixels, according to criteria:

$$BT \geq BT_W\text{-}0.5, \text{ AND } r5 < r_{glint} +0.05, \Rightarrow \texttt{CM\_CLEAR},$$

or detect clouds if

$$BT < BT_W\text{-}0.5K, \Rightarrow \texttt{CM\_CLOUD}.$$

If BT_W is unavailable, the brightness temperature is compared to a fixed threshold:
$$BT > 272.8K, \text{ AND } r5 < r_{glint} +0.03, \Rightarrow \texttt{CM\_CLEAR}.$$

Figure 7 shows an example of land-water classification. The initially un-initialized water mask is mostly correct by the day 8 of processing, and does not change after day 16 (May 20).

4. *Snow test*:

$$NDSI=(r4-r6)/(r4+r6) \geq 0.35 \text{ AND } r1 > 0.15 \text{ AND } r7 < 0.12 \text{ AND } BT < 288 \text{ (land)}/275 \text{ (water)}.$$

Because this test cannot discriminate between snow and ice clouds, the value of *mask* is set to MASK_SNOW_TO_CONFIRM, and value of cloud mask remains undefined. To avoid false snow detection in summer, this test is conducted if the maximal brightness temperature of the block and its nearest neighbors does not exceed +20° C (293K). The temperature threshold is high over land 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 B7 threshold ($r7 < 0.12$) 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 ($r7 \approx 0.02$-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 below. To account for snow brightening in the glint region, the B7 threshold is increased to 0.15 when relative azimuth is less than 35°.

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. Third, partially snow-covered pixels are particularly difficult and are often misclassified as cloud. 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 of *refcm* or re-initialization which responds to surface change following snowfall. The requirement of high covariance and a carefully designed set of filters, described in sec. 4.6, is usually effective in separating clouds from snow. 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 mid-latitudes during the summer and in tropical regions generally, although it cannot completely eliminate this error.

Once the snow is recorded at the Q-memory (q.*LWSmask*=MASK_SNOW) for a given block, the algorithm relies on identification of clear conditions by the *refcm*-analysis or by an "anti-cirrus" test (see sec. 4.5) in snow detection/confirmation. When covariance of the *new Tile* with reference image containing snow is high, the new snow pixels, which passed *Snow* test, become classified as snow. This is a relatively safe way to introduce new snow pixels under clear skies. On the contrary, the low covariance processing, which usually indicates presence of clouds, does

not add new snow. Moreover, a known snow pixel is detected as clear only if its reflectance is close to the *refcm* values. Despite being a conservative algorithm, our limited testing indicates that it 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).

5. *Ice test*:

$$NDII=(r5-r2)/(r5+r2)\geq0.4 \text{ AND } r7<0.015 \text{ AND } BT<273 \text{ (water only)}.$$

Because of the very low threshold on B7 reflectance, this test rarely misclassifies clouds as ice. If test is passed, the pixels is declared CM_CLEAR and the value of mask changes to q.*LWSmask*=MASK_ICE. Overall, this test captures only a small subset of fresh water or sea ice, which has a very wide range of spectral variability, from the spectrum typical of open water to the one of the old snow.

6. *The rest of pixels*, which didn't pass any of the above tests, and were not detected as clouds by the *Cirrus* test, are assigned the land mask value of MASK_TO_CONFIRM. Along with possibly snow pixels (MASK_SNOW_TO_CONFIRM), these pixels will be processed after *refcm*-based analysis.

4.5 *Cirrus* Cloud Test and *Anti-cirrus* Clear-Sky Test

Cirrus cloud test [*Gao*, 1993; *Ackerman et al*., 2006] uses MODIS band B26 (1.38 μm) located in the region of strong water vapor absorption. Even a small amount of atmospheric water vapor (less than 1 cm in the total path) is sufficient to absorb reflected solar radiation in this region. Thus, most of the Earth surface appears black in this channel. If high clouds are present, this channel becomes very bright as there is little water vapor above the clouds and most of reflected radiance escapes to space. This can be seen in Figure 2. Band B26 is very dark at medium and high WV, and becomes bright over high clouds which can be judged by the low brightness temperature.

We implemented a standard cirrus test with a global threshold 0.035 recommended by MODIS cloud team [*Ackerman et al*., 2006]:

$$\text{IF } r26>0.035 \Rightarrow CM\_CLOUD.$$

In agreement with MOD35, this test is not used if the surface height exceeds 2 km above the sea level.

When the atmosphere is very dry, channel 1.38 μm "sees" the surface. Ben-Dor [1994] used AVIRIS measurements and RT simulations to show that in this case the high signal in "cirrus" channel over bright surfaces would cause problems for cloud detection, strongly overestimating cloudiness. On the positive side, we notice that this signal has the same spatial variability and a high correlation with the neighbor "window" channel 1.2 μm (B5) (see Fig. 9), which can be used to detect clear skies. The correlation between these channels is also high over high clouds. These two cases are distinguished based on covariance between B1 and *refcm*, which is high in cloud-free conditions, and an additional constraint, $r5^{av}/r26^{av}>8$. Empirical analysis shows that for most of high clouds, this ratio is below 3-6, whereas typical clear-skies values over scenes with partial or full snow cover is in the range of 15-30.

This *Anti-cirrus* test is implemented before the *Cirrus* threshold test. Currently, it is used in winter when snow has been detected in a given block both earlier (in q.*LWSmask*) and in the latest *Tile*, and the maximal retrieved column water vapor in the block is below 0.4 cm. Our analysis shows that *Anti-cirrus* test is very successful over snow in cold clear winter days with temperatures below $-(10 \div 20^{\circ}$ C). It is fast and simple in comparison to the standard algorithm implemented in *CM_HighCov*, which is the main algorithm detecting clear-skies and snow.

Figure 9 shows that cold dry conditions are realized on days 3-4 and 7-8 of the first 8-day period when channel 26 brightens and acquires the same spatial image pattern as the visible and NIR channels. On the other days, this band is dominated by either water vapor absorption or clouds.
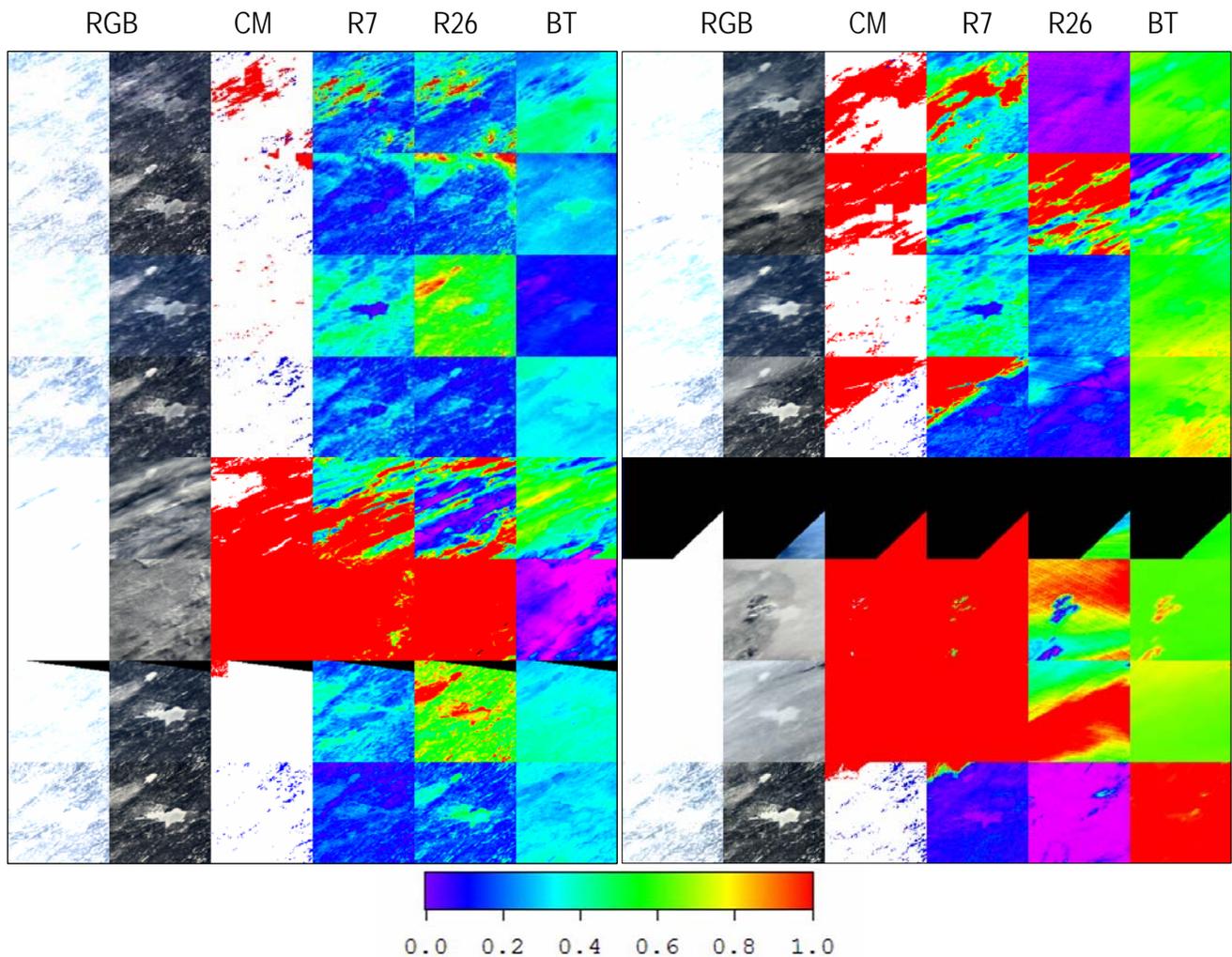
Figure 9. Illustration of "Anti-cirrus" test for 150km area north USA for two consecutive 8-day periods, DOY 27-35 and 36-44, 2005. At low temperature and low atmospheric moisture, band 26 "sees" the bright snow-covered surface and correlates spatially with VIS-NIR bands. These conditions realize during the cold period on days 3-4 and 7-8. The correlation disappears under warmer weather and higher moisture in the subsequent 8 days. The left two columns show differently normalized MODIS TERRA RGB images. The red, white and blue colors of *MAIAC* cloud mask indicate clouds, clear snow and clear land. Columns 4-6 are shown using rainbow color scheme with the following scales: (0-0.2) for B7, (0-0.05) for B26, (240-270) for BT.

Figure 9 also illustrates the general principles of cloud detection over snow based on B7 and B26. Because snow is dark in the shortwave IR region (B7), mixed or liquid water clouds usually appear brighter than the surface, whereas high clouds are usually well captured by *Cirrus* channel (B26).

4.6 Module *initRefcm( )*

This module runs when *refcm* is not initialized at the beginning of processing, or when low covariance was found with existing *refcm*. It requires at least two images to be stored in the Queue. The algorithm (Fig. 10) compares the *new Tile* and the previous *Tiles* stored in the Queue. It keeps moving in the backward direction in the Queue until either the "head" of Queue is reached, or the clear conditions are found. The processing is performed for each block separately.

1. It starts with computing B1 covariance between the *new Tile* and the previous *Tile*. If cov_{B1} is low (<0.68), then it continues with the other days from the Queue. Otherwise, if the surface is
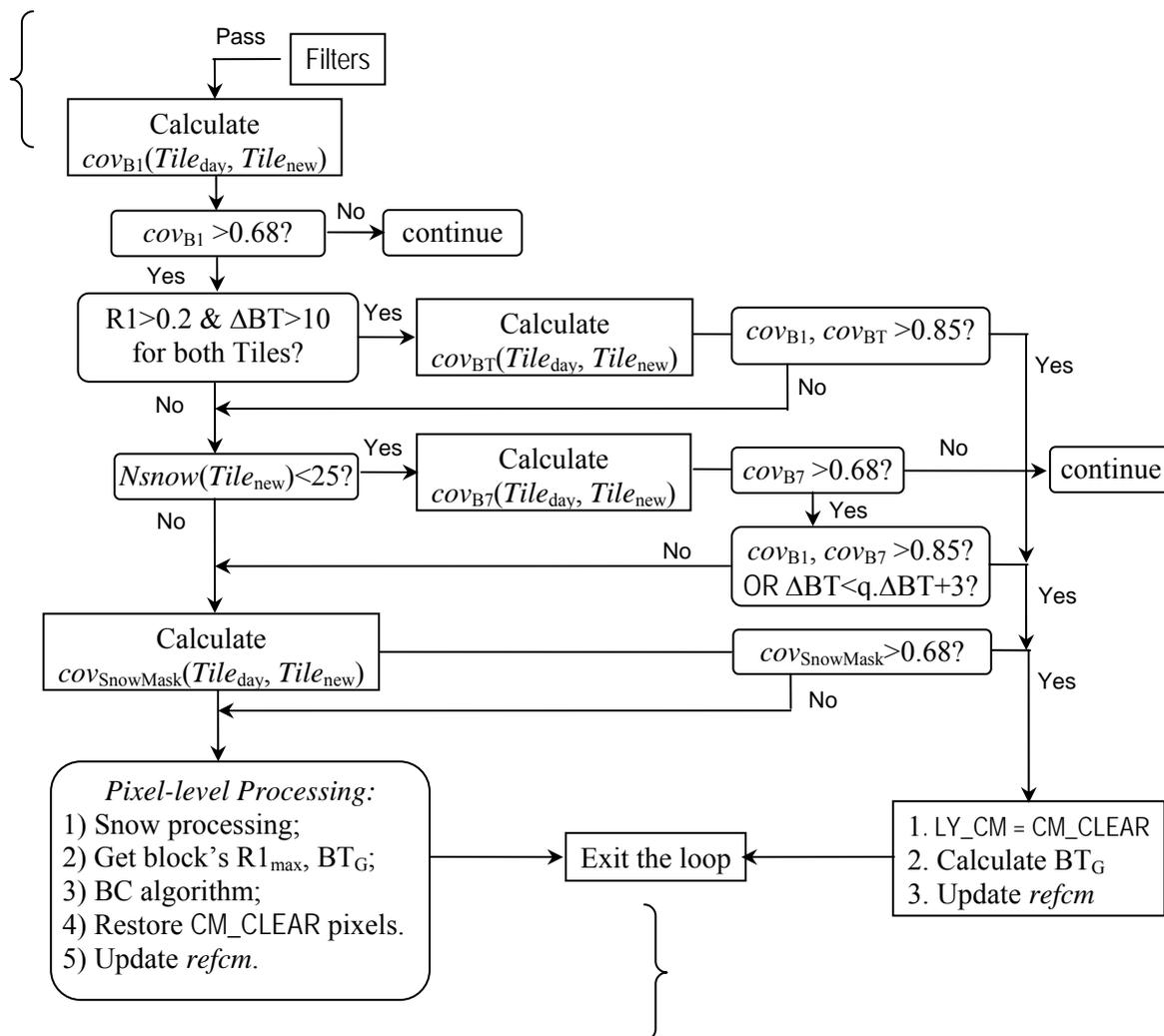
Figure 10. Module *initRefcm()* of *MAIAC* CM algorithm.

bright and has a high enough brightness temperature contrast, the BT covariance is calculated. If both B1 and BT are highly correlative (*cov*>0.85), then the initialization succeeds, all pixels of the block are declared CM_CLEAR and *Refcm* is updated. This pathway is successful over areas with mixed land type, coastal regions, swampy areas, or, more generally, areas with ground waters close to the surface, which creates a high and spatially stable BT contrast. When the BT contrast is low (several degrees), the results of BT covariance analysis become unpredictable.

2. If *Snow* test detects little or no snow, the B7 covariance is computed. As before, the processing continues with other days if cov_{B7} is low. If both B1 and B7 are highly correlative (*cov*>0.85), or cov_{B1}>0.85 and the BT contrast is low and agrees with the known value for a given block, then the initialization succeeds. When snow covers the ground, band B7 becomes dark and shows little spatial variability, so that the results of B7 covariance analysis become unpredictable.

3. If the number of detected snow pixels is high enough, the algorithm makes sure that the detected snow pixels on both days correlate spatially by computing covariance of the snow mask. This mask has a default value of 0 which switches to 1 if *Snow* test detected snow. If computed covariance is HIGH (≥0.68), the *refcm* initialization succeeds, as well as initialization of snow in q.*LWSmask* (MASK_SNOW).

4. Because covariance is HIGH in bands B1 (and B7 in the absence of snow), it is safe to assume at this stage that the conditions are mostly cloud-free although a few pixels may be cloudy. Next, the algorithm performs a pixel-level analysis. If *Snow* test detected snow pixels, a two-step *Snow* processing is performed first for every block.

Step 1: Mask reliable snow pixels as clear (CM_CLEAR) and compute minimal ($BT_{\min}^{snow}$) and brightness temperature and a minimal B1 reflectance ($r1_{\min}^{snow}$) of snow. The reliable pixels are defined as those which passed snow test in both *Tiles*, or in the *new Tile* and have been detected earlier in the q.*LSWmask*.

Step 2: Mask the rest of pixels, which passed *Snow* test, as follows:

For pixels of the *new Tile*:   IF(BT< $BT_{\min}^{snow}$-4) $\Rightarrow$ CM_PCLOUD (cold pixels);
                                ELSE $\Rightarrow$ CM_CLEAR.

For pixels, detected as snow earlier (q.*LSWmask*=MASK_SNOW):

IF(r1>min($r1_{\min}^{snow}$,0.12) & r7<0.12 & (BT≥$BT_{\min}^{snow}$-4 & BT<288)) $\Rightarrow$ CM_CLEAR.
ELSE $\Rightarrow$ CM_PCLOUD.

After snow processing, the unprocessed pixels are assumed snow-free. The algorithm evaluates ground brightness temperature (BT_G) unless it is known from clear-skies *Spectral* tests, and maximal clear-skies reflectance ($r1_{max}$) using histogram analysis for 90% pixels assuming that there are less than 10% clouds, which will be coldest and brightest in the block. For example, BT_G corresponds to the lowest value of 90% warmest pixels of the block. Then, the bright-cold algorithm detects clouds as follows:

IF(r1>$r1_{max}$+0.02 & BT<BT_G-2 & BT_G<293) $\Rightarrow$ CM_PCLOUD;
ELSE IF(r1>0.5 & BT<273) $\Rightarrow$ CM_PCLOUD;
ELSE $\Rightarrow$ CM_CLEAR.

5. In certain cases, the bright-cold algorithm makes a systematic error, masking clear pixels as cloudy. One frequent example is glaciers or melting snow in higher elevations, surrounded by a much warmer and darker land. Another example is salt playas south of Salt Lake City, Utah, which are very bright and which may be 10-15° colder than surrounding barren land, perhaps due to close ground waters and low absorption of sun light. Despite these areas perfectly fit the BC logic, they also display a remarkably stable spatial reflectance pattern which help identify these areas as clear. *MAIAC* performs a post-processing restoring clear value for the possibly cloudy pixels (CM_PCLOUD). It is based on the assumption that the B1 images between two *Tiles* have a linear relationship under clear-skies, r1(*new Tile*)=a+b×r1(*Tile*). We establish coefficients *a* and *b* using CM_CLEAR pixels, and then re-map CM_PCLOUD pixels if their B1 reflectance does not exceed predicted value with uncertainty of 0.04:

$$\text{IF}(\text{r1}(\textit{new Tile}) < \text{a+b×r1}(\textit{Tile}) + 0.04) \Rightarrow \text{CM\_CLEAR}.$$

6. 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, … 1) 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 very low wind speeds.
- The number of pixels which passed *Snow* test is significantly different between observations L and *k*.

- The expected reflectance test based on the number of detected snow pixels. This test rejects blocks if the average B1 reflectance exceeds expected value, $r1_{av}>r1_{Exp}+0.12$. The value $r1_{Exp}=\text{snowFrac}\times0.9+(1-\text{snowFrac})\times\min(0.15, r1_{\min}^{Block})$ is computed assuming snow reflectance of 0.9 and land reflectance of 0.15. The snow fraction is calculated as a ratio of the number of detected snow pixels to the total number of non-FILL_VALUE pixels in the block. This test mostly rejects bright liquid water and mixed-phase clouds.

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:

- The difference of the block-average reflectances should not be too high: $\left|\langle r1_L\rangle - \langle r1_k\rangle\right| > 0.25$.

  This test filters infrequent cases when bright continuous clouds correlate spatially with much darker clear scenes.
- Finally, a "sigma" test is used to filter Tiles with significantly higher or lower spatial variability than that of a clear-skies surface:

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

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 globally, the developed set of filters ensures a reliable initialization or re-initialization 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.

The `CM_CLEAR` pixels which passed *Snow* test are classified as snow in the q.*LWSmask*. This is the only way for the snow to be introduced in classification mask for the first time. Although this conservative approach may delay detection of fresh snow, it has a low error of cloud misclassification and a globally reliable performance.

4.7 Module *CM_highCov*
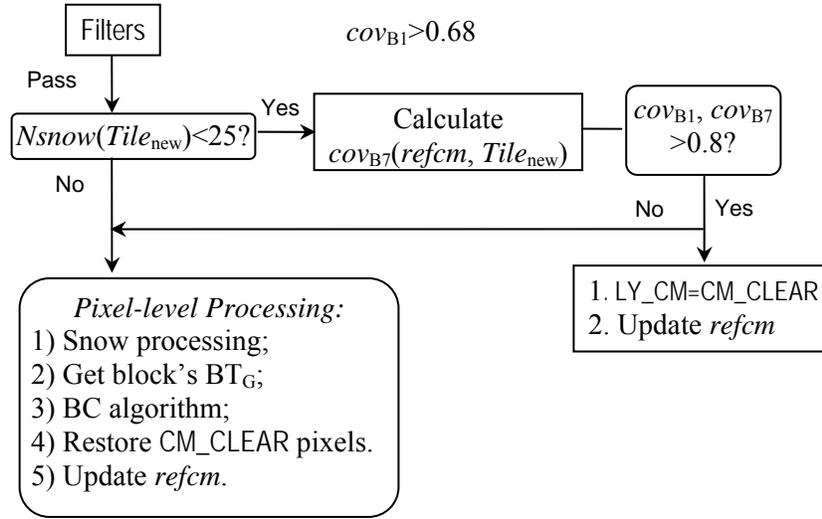
Figure 11. Module *CM_highCov*.

This module (Fig. 11) is called when *refcm* is initialized and covariance between *refcm* and *new Tile* for a given block is HIGH ($cov_{B1} > 0.68$). High covariance implies that the block is most probably clear although a few pixels may still be cloudy. This algorithm is a reduced version of *initRefcm*. First, the filter excludes clouds which correlate with a clear scene. The filter has two tests: the expected reflectance test based on the number of detected snow pixels, and the average scene brightness test with the exclusion condition of:

$$\left| \langle refcm_{B1} \rangle - \langle r1 \rangle \right| > 0.25 \text{ or } \left| \langle refcm_{B7} \rangle - \langle r7 \rangle \right| > 0.2 .$$

Next, if the number of detected snow pixels is low, it computes covariance in B7. If *refcm* and *new Tile* are highly correlative, the block is declared `CM_CLEAR` and *refcm* is updated. Otherwise, the algorithm performs a pixel-based analysis (steps 4-5 of *initRefcm*). It performs a two-step snow processing, computes BT_G, runs Bright-Cold algorithm followed by restoring of `CM_CLEAR` pixels, and finally, updates *Refcm* structure. The BC-algorithm is formulated as follows:

IF (BT<BT_G-4 & (r1>$refcm_{B1}$+0.05+Δ OR r7>$refcm_{B7}$+0.05))   ⇒ `CM_PCLOUD`;
 ELSE ⇒ `CM_CLEAR`.

Parameter $\Delta = \langle r1 \rangle - \langle refcm_{B1} \rangle$ (Δ=0 if the difference is negative) removes the average difference of reflectance helping detect clear conditions when the atmosphere is hazy.

4.8 Module *CM_lowCov*

This module is called when covariance is low ($cov_{B1} \leq 0.68$) and re-initialization did not succeed. Usually, this means that the block has clouds. Initially, the algorithm evaluates BT_G for a given block if it is undefined. BT_G is computed as an average value of BT_G of neighbor blocks for

which it was defined either by clear-sky Spectral Tests or by the module *CM_highCov*(). Next, the algorithm performs a pixel-based analysis comparing measured B1, B7 reflectance with respective *refcm* reflectance and measured BT with the ground temperature BT_G if it is available. The pixel-level algorithm first treats possible snow pixels, and then runs the rest of pixels using Bright-Cold algorithm:

IF(*mask_q*=Snow & *mask*=Snow_to_confirm & $\left|r1 - refcm_{B1}\right| < 0.1$ & $\left|r7 - refcm_{B7}\right| < 0.05$) $\Rightarrow$ CM_CLEAR;

(Known snow pixel which passed *Snow* test and whose reflectance is close to *refcm*);

ELSE IF(*mask_q*=Snow)
{   IF(r1<*refcm*_{B1}-0.05 & r7< *refcm*_{B7}+0.1) $\Rightarrow$ CM_CLEAR; (melting snow)
      ELSE $\Rightarrow$ CM_CLOUD}

(Known snow pixel which failed *Snow* test either because snow is melting or it is cloudy. Thresholds indicate that reflectance of melting snow drops in B1 and usually grows in B7);

ELSE IF(*mask*=Snow_to_confirm)
{   IF($\left|r1 - refcm_{B1}\right| < 0.05$ & $\left|r7 - refcm_{B7}\right| < 0.05$) $\Rightarrow$ CM_CLEAR;
      ELSE $\Rightarrow$ CM_CLOUD}

(Pixel passed *Snow* test, but snow has not been recorded in the Queue' land-water-snow mask. This may be a first snow which has not been confirmed by a high covariance, or an undetected cloud. Both cases will be masked as a cloud. Due to a threshold nature of *Snow* test and uncertainties of measurements related to surface change, footprint variability etc., we are also testing if this could be a clear pixel whose reflectance is close to the *refcm* values);

ELSE $\Rightarrow$ BC algorithm.

The Bright-Cold algorithm tries to utilize all available information, including ground brightness temperature and *refcm* reflectance in bands B1 and B7. The variability allowed by thresholds is reduced if less ancillary information is available. The algorithm is formulated as follows:

IF(BT_G is available)
{ IF (BT>BT_G-4 & r1<*refcm*_{B1}+0.05 & r7<*refcm*_{B7}+0.05) $\Rightarrow$ CM_CLEAR;
  ELSE IF (BT>BT_G & r1<*refcm*_{B1}+0.1 & r7<*refcm*_{B7}+0.1) $\Rightarrow$ CM_CLEAR;
  ELSE $\Rightarrow$ CM_CLOUD}
ELSE (BT_G is unavailable)
{ IF (r1<*refcm*_{B1}+0.05 & $\left|r7 - refcm_{B7}\right| < 0.05$) $\Rightarrow$ CM_CLEAR;

  ELSE IF (BT>298 & r1<*refcm*_{B1}+0.1 & $\left|r7 - refcm_{B7}\right| < 0.1$) $\Rightarrow$ CM_CLEAR;

  ELSE $\Rightarrow$ CM_CLOUD}

When the ground brightness temperature cannot be evaluated, BC algorithm constrains variation of B7 reflectance from both above and below. Due to water/ice absorption in the 2 μm region, clouds often reduce B7 reflectance over bright surfaces, and the low reflectance constraint on band B7 helps to filter clouds. This constraint is not necessary when BT_G is known and colder clouds are filtered based on lower brightness temperature.

This principle is illustrated in Figure 12 for the area of north-west Rocky Mountain with the ridge on the left and bright central plateau. The average altitude of plateau is 2 km at the average reflectance of 0.3 in B1 and ~0.4 in B7. High clouds on day 4 significantly reduce B7 reflectance. Overall, *MAIAC* accurately masks clouds except for the last day of 5-day sequence,

RGB	CM	R1	R7	R26	BT
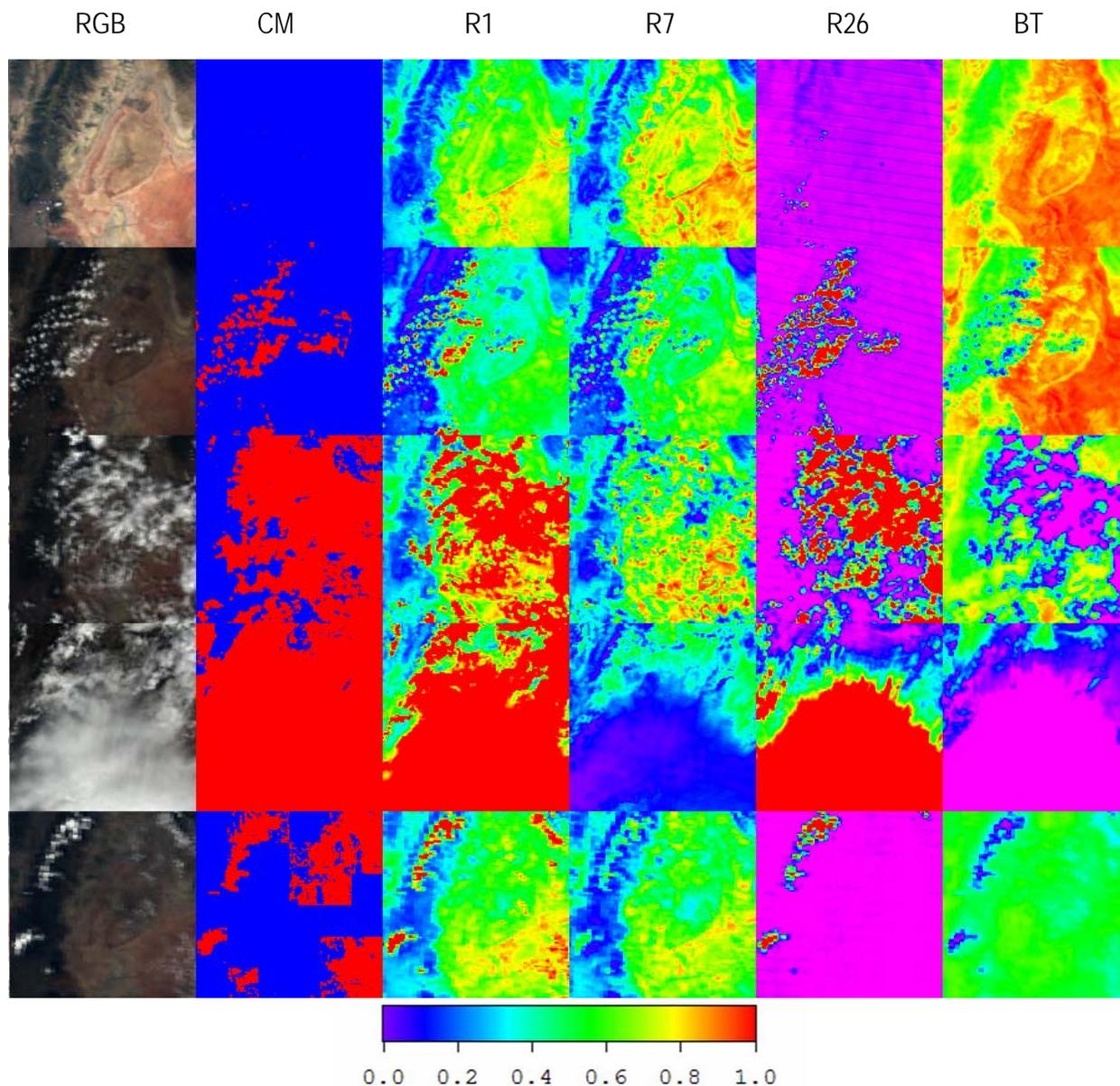
Figure 12. Illustration of *MAIAC* CM over bright area of north-west Rocky Mountains, end of August, 2007. Columns 3-6 have the following scales: (0-0.4) for B1, (0-0.5) for B7, (0-0.05) for B26, (260-310) for BT.

where the low-covariance algorithm somewhat overestimates cloudiness in blocks containing clouds.

4.9. Performance of CM Algorithm

Performance of the cloud mask algorithm has been extensively tested using ~9 years of MODIS TERRA data (2000 – august 2008) subsetted for areas of $50{\times}50$ km^2 for about 160 AERONET locations worldwide. The subsets are produced operationally by MODIS Adaptive Processing

System (MODAPS) and are automatically sent to our AERONET-based Surface Reflectance Validation Network (ASRVN) server [*Wang et al.*, 2008]. This testing showed globally robust performance. Some examples of cloud mask for these subsets, and its enhancement by an aerosol retrieval algorithm will be given later.

To gain insight on the large scale algorithm performance, we used MODIS TERRA data for the north-east USA, north-west Rocky Mountains, Middle East, Arabian peninsular, Africa, India etc., each covering the area of 600^2 to 1800^2 km^2. The testing was done for at least half a year of continuous data in each case. A separate analysis was done for the whole area of Greenland [*Lyapustin et al.*, 2008].

Examples of CM performance over Zambia, Africa, Arabian Peninsula, and north-east USA are shown in Figures 13-15, along with reprojected MODIS Collection 5 cloud mask MOD35. The following color legend was used: red – cloud, yellow – possibly cloud, green – possibly clear (MOD35 only), blue – clear land, white – clear snow, light blue – clear water (*MAIAC* only). One can see that over relatively dark surfaces cloud masks from *MAIAC* and MOD35 are generally similar. Due to covariance analysis and reference clear-sky image, *MAIAC* has a higher confidence in clear conditions (Fig. 13), which is important for land applications. It provides a better performance in difficult cases, such as dust storms (Fig. 14) or cases of high aerosol loading. Figure 15 shows that *MAIAC* performs significantly better over snow. This Figure shows a winter-spring season of 2005. The RGB image is scaled dynamically to show the full range of variability of measured reflectance. MOD35 correctly masks clouds, but it also often significantly overestimates cloudiness over snow in clear conditions. Currently, the work is underway on global comparison of *MAIAC* CM with MOD35 in cooperation with MODIS cloud group (P. Menzel, R. Frey, S. Dutcher, S. Ackerman et al).

Our analysis of the developed cloud mask on global subsets of AERONET data and large-scale target areas reveals no major problems. One of identified problems is frequent false cloud detection on the land-water boundaries. This problem may be inherent to our algorithm because gridding of data with variable footprint and center location at the land-water boundary produces large uncertainties and variability of gridded signal between observations. There are several ways to ameliorate this problem which we plan to address in the very near future.

Concluding this section, we would like to mention that the aerosol retrieval and atmospheric correction algorithms are the main arbiter to judge the performance of CM algorithm. Undetected clouds would bias the comparison statistics of retrieved aerosol optical thickness against AERONET measurements, and will explicitly show up in the images of corrected surface reflectance. On the other hand, over-detection of clouds would reduce statistics of AOT-AERONET matching and the rate of surface retrievals. From an extensive MODIS testing, we don't see these problems in the aerosol retrievals and atmospheric correction, which testifies to an adequate accuracy of developed CM algorithm, at least in the frame of the overall *MAIAC* processing.
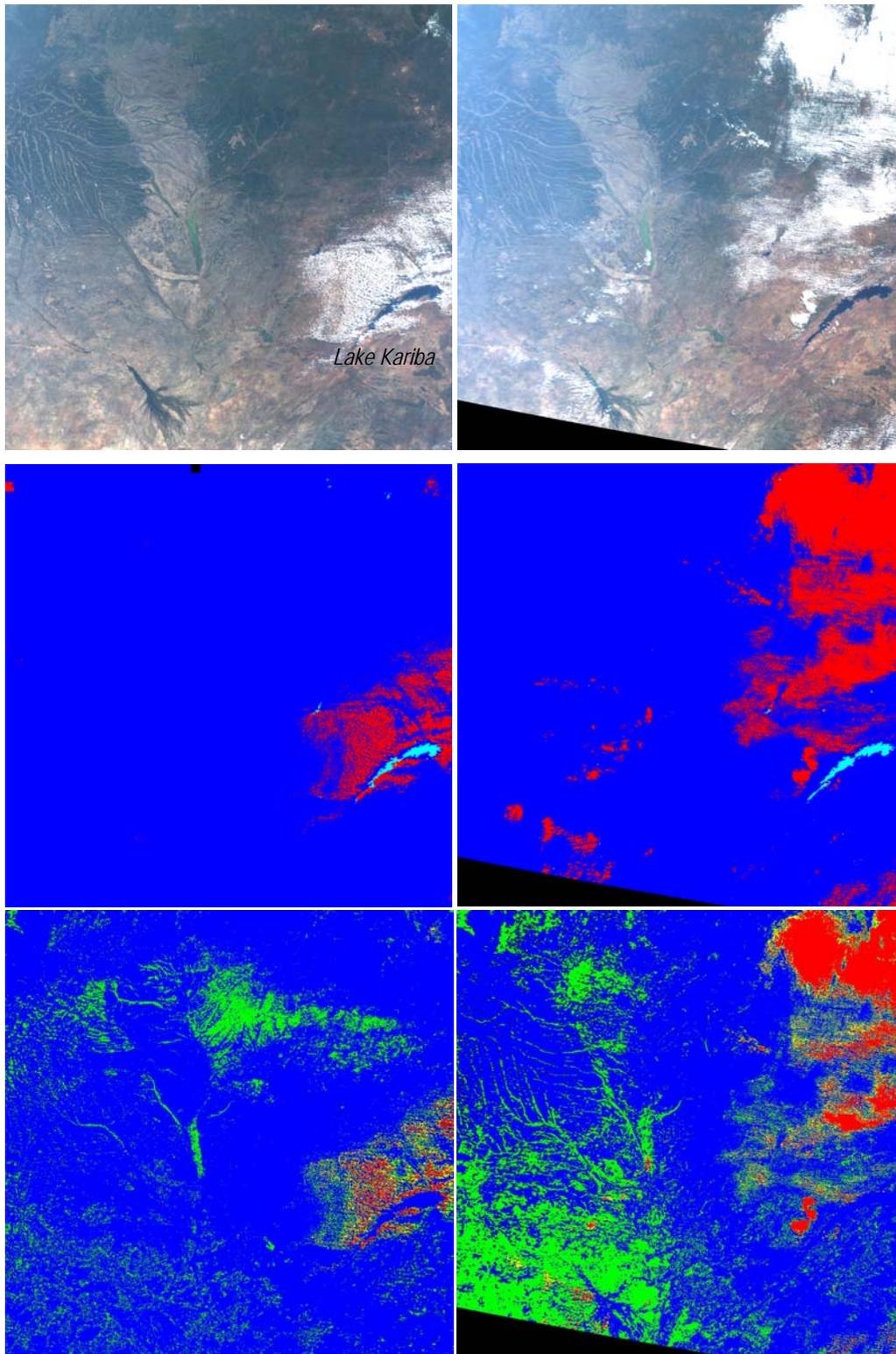
Figure 13. Example of *MAIAC* (moddle) and MOD35 (bottom) cloud mask at the beginning of dry season for Zambia, Africa, for days 130 (left) and 141 (right) of 2005. The image shows 4 Tiles (1200×1200 km^2). Color legend: red – cloud, yellow – possibly cloud, green – possibly clear (MOD35 only), blue – clear land, white – clear snow, light blue – clear water (*MAIAC* only).
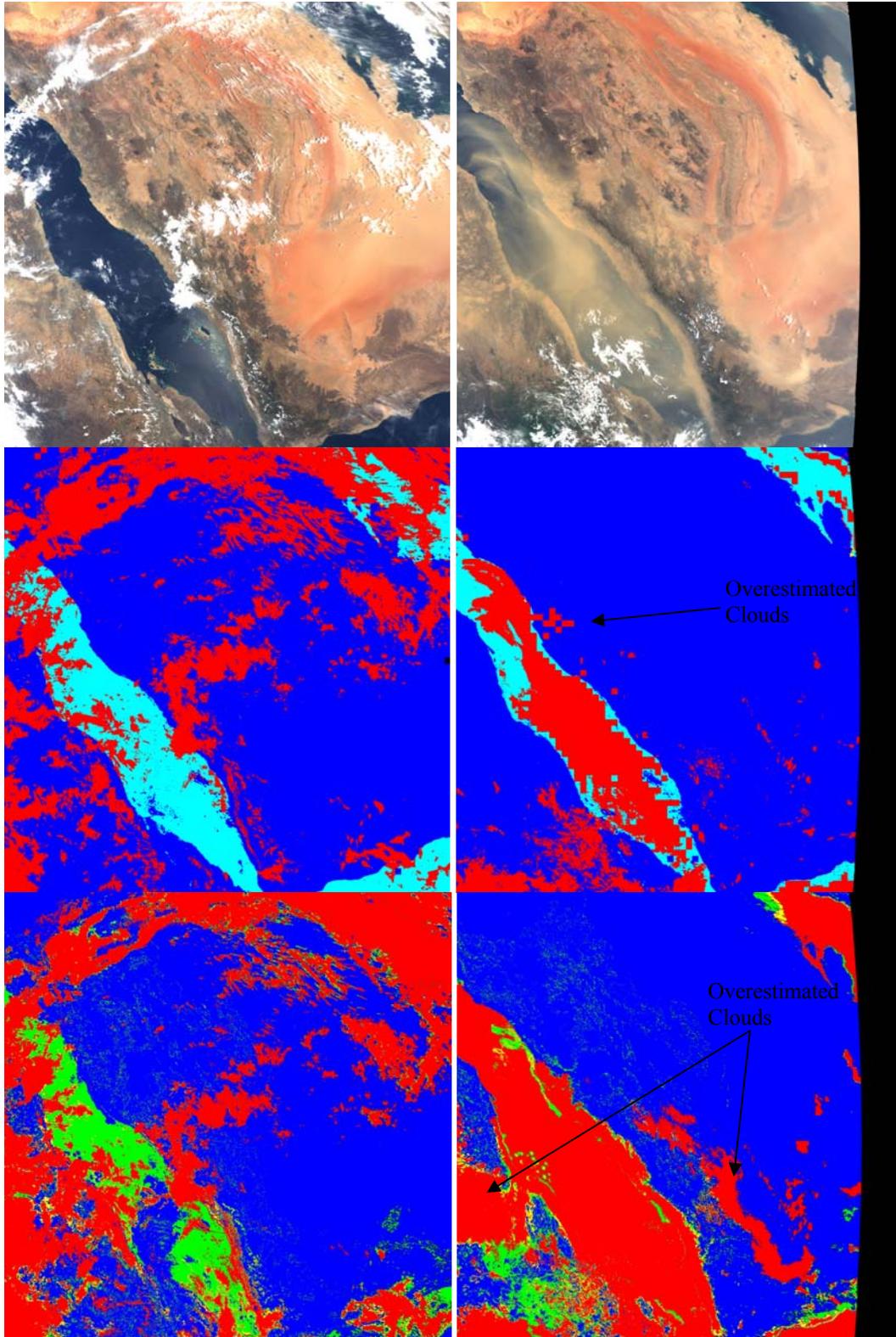
Figure 14. Example of *MAIAC* (middle) and MOD35 (bottom) cloud mask for Arabian Peninsula from MODIS TERRA data for days 145 (left) and 207 (right) of 2005 (1800×1800 km^2).

a) 2005-028
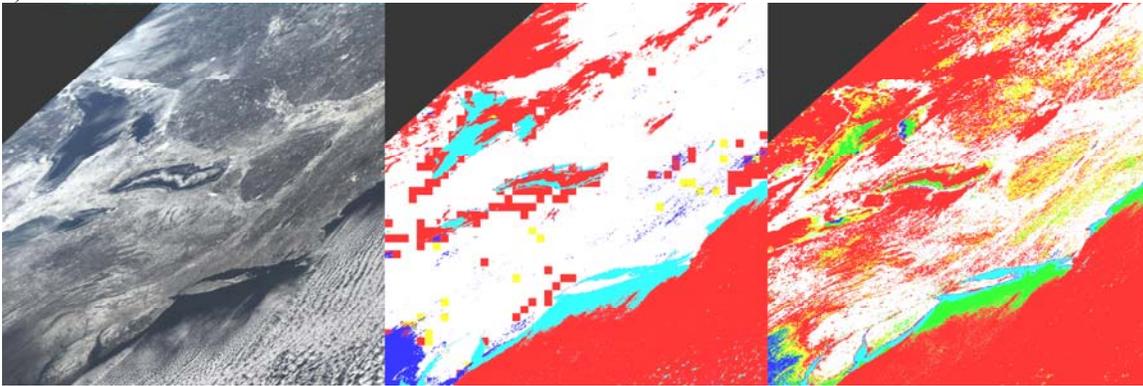
b) 2005-029
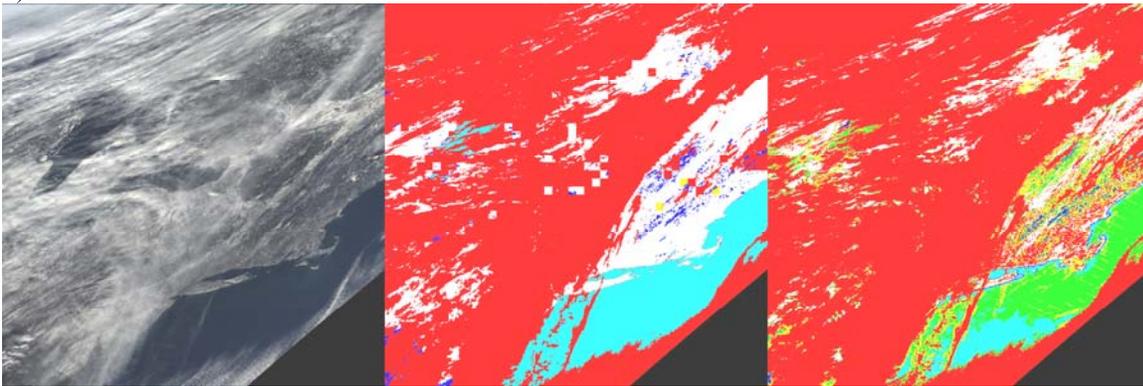
c) 2005-031
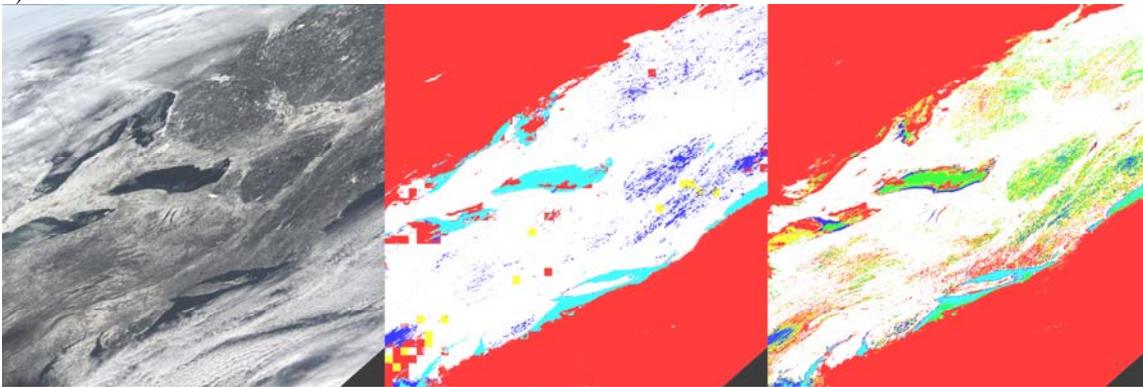
d) 2005-033
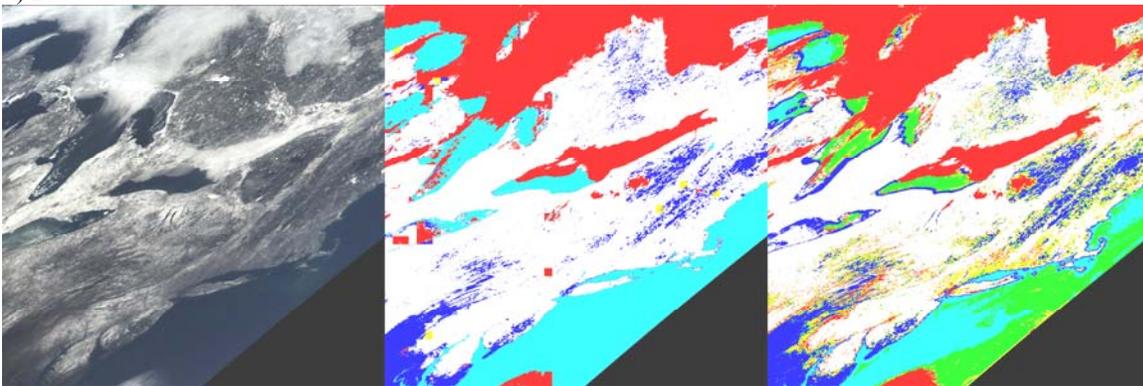
e) 2005-049
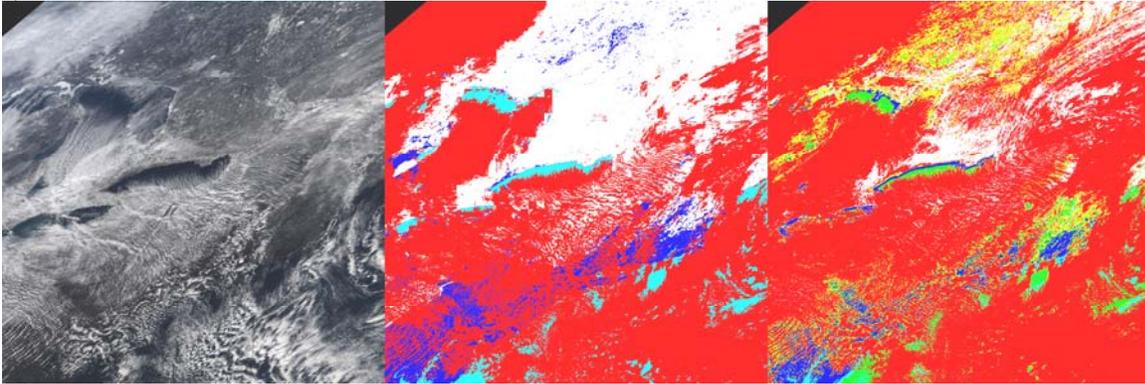
f) 2005-072
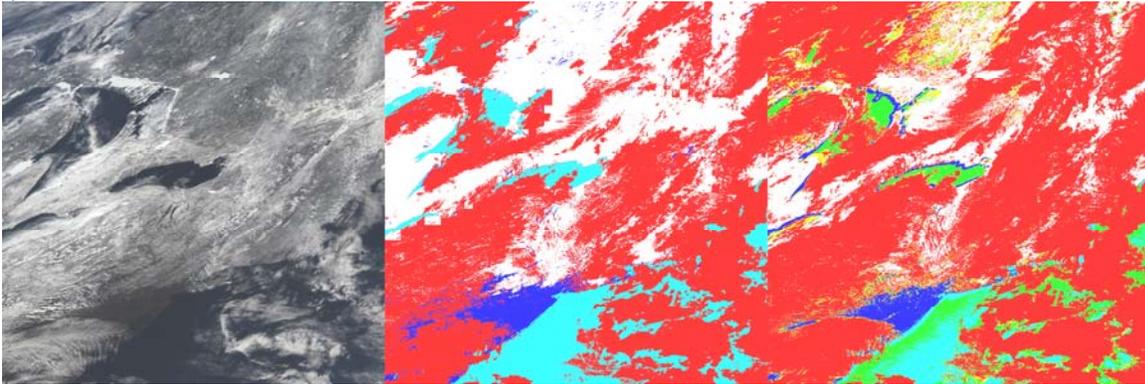
g) 2005-076
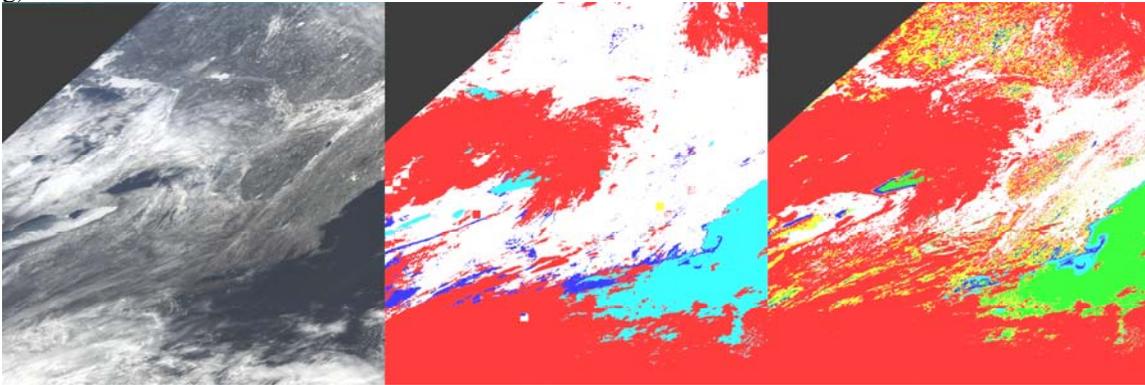
h) 2005-085
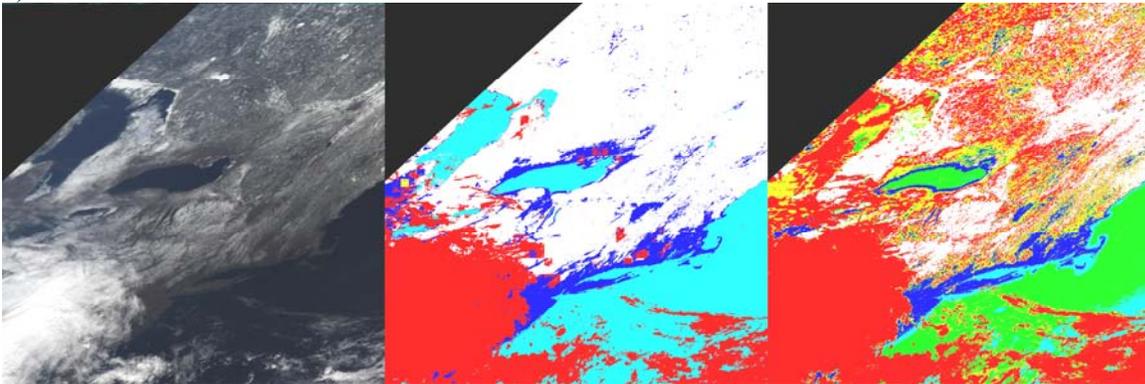
i) 2005-100
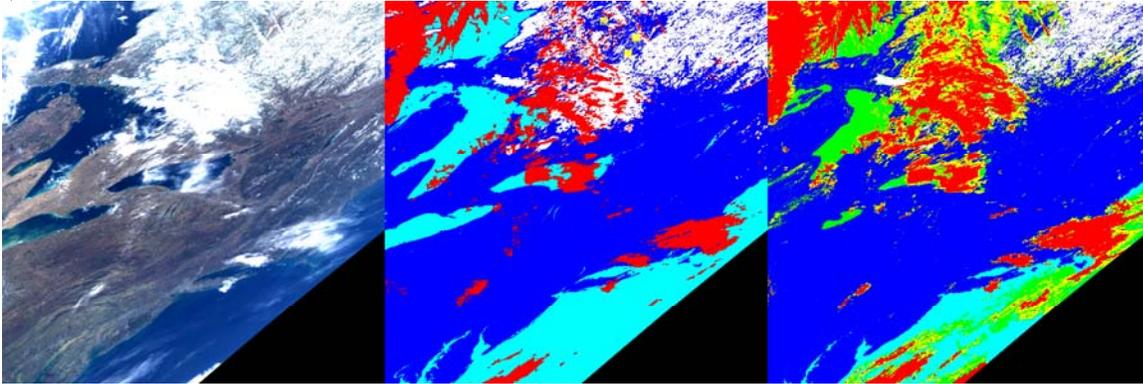
j) 2005-104
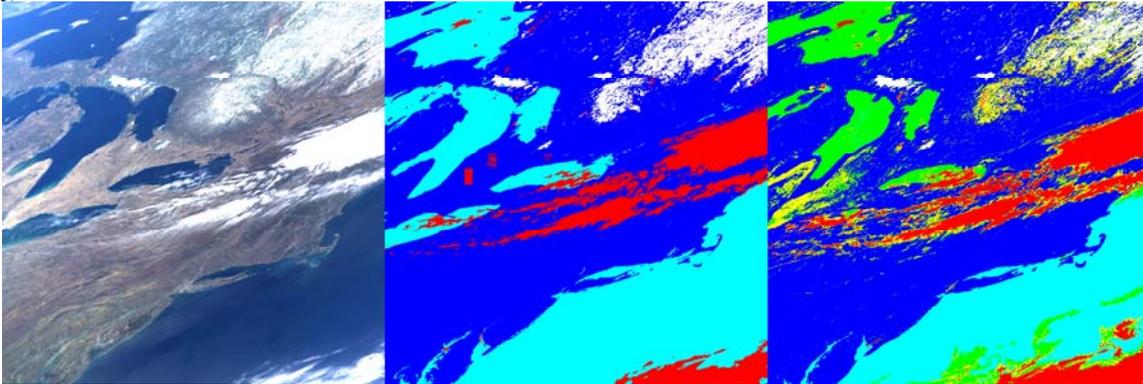
k) 2005-108
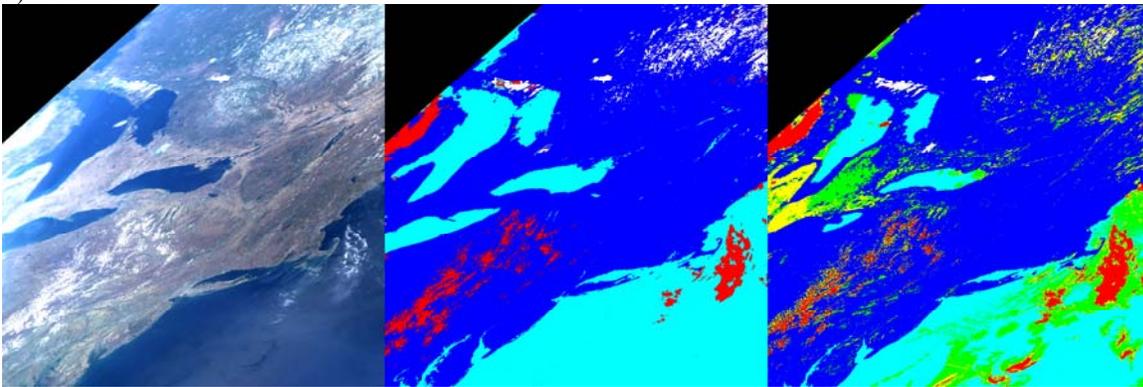
l) 2005-209
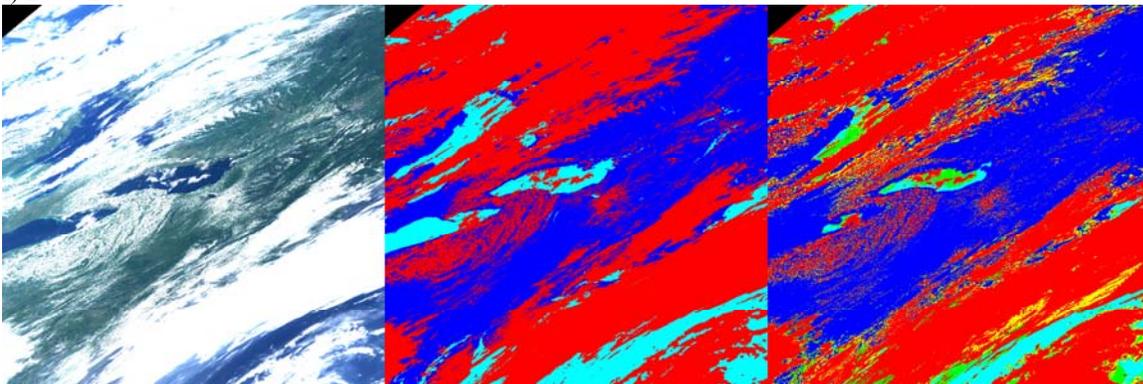
Figure 15. Example of *MAIAC* (middle) and MOD35 (right) cloud mask for north-east USA (1200^2 km^2).

5. *MAIAC* Radiative Transfer Basis

MAIAC resulted from an effort to develop an operational algorithm with explicit minimization where parameters of the surface BRF model can be calculated analytically from measurements. A similar approach developed by *Martonchik et al.* [1998] for MISR features a relatively small size of the look-up table (LUT) and a high efficacy, which is critically important for operational algorithm. We will be using a high accuracy semi-analytical formula for the top-of-atmosphere (TOA) radiance derived with the Green's function method [*Lyapustin and Knyazikhin*, 2001; *Lyapustin and Wang*, 2005].

The following notations are used below:

τ - atmospheric optical thickness (OT); τ^g, τ^w - absorption OT of 5 well-mixed gases (CO_2, CH_4, NO_2, CO, N_2O) and of water vapor;

s_0, s - incidence and view directions defined by pairs of zenith and azimuthal angles (θ, φ). For brevity, φ will also stand for the difference $\varphi - \varphi_0$;

μ_0, μ - cosines of the solar zenith angle (SZA) and view zenith angle (VZA) ($\mu = \cos\theta$). The z-axis is pointed downwards, so $\mu_0 > 0$ for the solar beam and $\mu < 0$ for the reflected beam.

πS_λ - extraterrestrial solar spectral irradiance;

ρ, q - surface bidirectional reflectance factor (BRF) and surface albedo;

c_0 - spherical albedo of the atmosphere.

The TOA radiance $L(s_0, s)$ is expressed as a sum of the atmospheric path radiance (*D*), and surface-reflected radiance (*L_s*), directly and diffusely transmitted through the atmosphere:

$$L(s_0, s) = D(s_0, s) + L_s(s_0, s)e^{-\tau/|\mu|} + L_s^d(s_0, s). \tag{5.1}$$

The surface-reflected radiance is written as:

$$L_s(s_0, s) \cong S_\lambda \mu_0 e^{-\tau/\mu_0} \{ \rho(s_0, s) + \alpha c_0 \rho_1(\mu)\rho_2(\mu_0) \} + \frac{\alpha}{\pi} \int_{\Omega^+} D_s(s_0, s')\rho(s', s)\mu' ds', \tag{5.2}$$

where D_s is path radiance incident on the surface, c_0 is spherical albedo of the atmosphere, and

$$\rho_1(\mu) = \frac{1}{2\pi} \int_{\Omega^+} \rho(s', s)ds', \quad \rho_2(\mu_0) = \frac{1}{2\pi} \int_{\Omega^-} \rho(s_0, s)ds. \tag{5.3}$$

α is a multiple reflection factor, $\alpha = (1 - q(\mu_0)c_0)^{-1}$, where q is surface albedo. The diffusely transmitted surface-reflected radiance at the TOA is calculated from L_s with the help of 1D diffuse Green's function of the atmosphere:

$$L_s^d(s_0, s) = \int_{\Omega^-} G^d(s_1, s)L_s(s_0, s_1)ds_1. \tag{5.4}$$

The function πG^d is often called bi-directional upward diffuse transmittance of the atmosphere. The method of its calculation was discussed in detail in [*Lyapustin and Knyazikhin*, 2001]. The surface albedo is defined as a ratio of reflected and incident radiative fluxes at the surface:

$$q(\mu_0) = F^{Up}(\mu_0) / F^{Down}(\mu_0), \tag{5.5a}$$

$$F^{Down}(\mu_0) = \pi S_\lambda \mu_0 e^{-\tau/\mu_0} + \int_{\Omega^+} D_s(s_0, s')\mu' ds' = F_s^{Dir}(\mu_0) + F_s^{Dif}(\mu_0), \qquad (5.5b)$$

$$F^{Up}(\mu_0) = \pi S_\lambda \mu_0 e^{-\tau/\mu_0} q_2(\mu_0) + \int_{\Omega^+} \mu' q_2(\mu') D_s(s_0, s') ds', \quad q_2(\mu_0) = \frac{1}{\pi}\int_{\Omega^-} \rho(s_0, s)\mu ds. \quad (5.5c)$$

These formulas give an explicit expression for the TOA radiance as a function of surface BRF. The accuracy of the above formulas is high, usually within a few tenths of a percent [*Lyapustin and Knyazikhin*, 2001]. Below we will use the TOA reflectance, which is defined as

$$R_\lambda = L_\lambda /(\mu_0 S_\lambda). \qquad (5.6)$$

5.1 Expression for the TOA reflectance using *LSRT* BRF model

Based on the described semi-analytical solution, TOA reflectance can be expressed as an explicit function of parameters of the BRF model. We are using a semi-empirical Li Sparse – Ross Thick (LSRT) BRF model [*Lucht et al.*, 2000]. This is a linear model, represented as a sum of Lambertian, geometric-optical, and volume scattering components:

$$\rho(\mu_0, \mu, \varphi) = k^L + k^G f_G(\mu_0, \mu, \varphi) + k^V f_V(\mu_0, \mu, \varphi). \qquad (5.7)$$

It uses predefined geometric functions (kernels) f_G, f_V to describe different angular shapes. The kernels are independent of the land conditions. The BRF of a pixel is characterized by a combination of three kernel weights, $\vec{K} = \{k^L, k^G, k^V\}^T$. The LSRT model is used in the operational MODIS BRF/albedo algorithm [*Schaaf et al.*, 2002].

The substitution of Eq. (5.7) into Eqs. (5.1-5.5) and normalization to the reflectance units gives the following expressions for the surface-reflected signal (the last two terms of Eq. (5.1)):

$$R_s(\mu_0, \mu, \varphi) = e^{-\tau/\mu_0} \{ k^L + k^G f_G(\mu_0, \mu, \varphi) + k^V f_V(\mu_0, \mu, \varphi) + \alpha c_0 \rho_1(\mu) \, \rho_2(\mu_0) \}$$
$$+ \alpha \mu_0^{-1} \{ k^L E_0^d(\mu_0) + k^G D_G^1(\mu_0, \mu, \varphi) + k^V D_V^1(\mu_0, \mu, \varphi) \}, \qquad (5.8)$$

$$R_s^d(\mu_0, \mu, \varphi) = e^{-\tau/\mu_0} \times \{ [k^L G^{av}(\mu) + k^G G_G^1(\mu_0, \mu, \varphi) + k^V G_V^1(\mu_0, \mu, \varphi)] +$$
$$\alpha c_0 [k^L G^{av}(\mu) + k^G G_G^{11}(\mu) + k^V G_V^{11}(\mu)] \, \rho_2(\mu_0) \}$$
$$+ \alpha \mu_0^{-1} \{ k^L E_0^d(\mu_0) G^{av}(\mu) + k^G H_G^1(\mu_0, \mu, \varphi) + k^V H_V^1(\mu_0, \mu, \varphi) \}. \qquad (5.9)$$

The surface albedo is written as:

$$q(\mu_0) = E_0^{-1}(\mu_0) \{ \mu_0 e^{-\tau/\mu_0} q_2(\mu_0) + k^L E_0^d(\mu_0) + k^G D_G^3(\mu_0) + k^V D_V^3(\mu_0) \}. \qquad (5.10)$$

Different functions of these equations represent different integrals of the incident path radiance (D_s) and atmospheric Green's function (G) with the BRF kernels. They were described in [*Lyapustin and Wang*, 2005] along with the method of numerical calculation. Below, we give only the integral expressions for these functions:

$$\rho_1(\mu) = k^L + k^G f_G^1(\mu) + k^V f_V^1(\mu), \qquad (5.11)$$

$$\rho_2(\mu_0) = k^L + k^G f_G^2(\mu_0) + k^V f_V^2(\mu_0), \qquad (5.12)$$

$$q_2(\mu_0) = k^L + k^G f_G^3(\mu_0) + k^V f_V^3(\mu_0), \qquad (5.13)$$

$$D_k^1(\mu_0,\mu,\varphi-\varphi_0)=\frac{1}{\pi}\int_0^1\mu'd\mu'\int_0^{2\pi}d\varphi'D_s(\mu_0,\mu',\varphi'-\varphi_0)f_k(\mu',\mu,\varphi-\varphi'),\qquad(5.14)$$

$$D_k^3(\mu_0)=\frac{1}{\pi}\int_0^{2\pi}d\varphi'\int_0^1\mu'f_k^3(\mu')D_s(\mu_0,\mu';\varphi')d\mu',\qquad(5.15)$$

$$G^{av}(\mu)=\int_{-1}^0d\mu_1\int_0^{2\pi}G^d(\mu_1,\mu,\varphi-\varphi_1)d\varphi_1,\qquad(5.16)$$

$$G_k^{11}(\mu)=\int_{-1}^0f_k^1(\mu_1)d\mu_1\int_0^{2\pi}G^d(\mu_1,\mu,\varphi-\varphi_1)d\varphi_1,\qquad(5.17)$$

$$G_k^1(\mu_0,\mu,\varphi-\varphi_0)=\int_{-1}^0d\mu_1\int_0^{2\pi}G^d(\mu_1,\mu,\varphi-\varphi_1)f_k(\mu_0,\mu_1,\varphi_1-\varphi_0)d\varphi_1,\qquad(5.18)$$

$$H_k^1(\mu_0,\mu,\varphi-\varphi_0)=\int_{-1}^0d\mu_1\int_0^{2\pi}G^d(\mu_1,\mu,\varphi-\varphi_1)D_k^1(\mu_0,\mu_1,\varphi_1-\varphi_0)d\varphi_1.\qquad(5.19)$$

The subscript k in the above expressions refers to either geometric-optical (G) or volumetric (V) kernels, and the supplementary functions of the BRF kernels are given by:

$$f_k^1(\mu)=\frac{1}{2\pi}\int_0^1d\mu'\int_0^{2\pi}f_k(\mu',\mu,\varphi'-\varphi)d\varphi',\qquad(5.20a)$$

$$f_k^2(\mu_0)=\frac{1}{2\pi}\int_{-1}^0d\mu_1\int_0^{2\pi}f_k(\mu_0,\mu_1,\varphi_1-\varphi_0)d\varphi_1,\qquad(5.20b)$$

$$f_k^3(\mu')=\frac{1}{\pi}\int_{-1}^0\mu d\mu\int_0^{2\pi}f_k(\mu',\mu,\varphi-\varphi')d\varphi.\qquad(5.20c)$$

The diffuse and total spectral surface irradiance are calculated from (5.5b) as:

$$E_0^d(\mu_0)=F^{Dif}(\mu_0)/(\pi S_\lambda),\quad E_0(\mu_0)=F^{Down}(\mu_0)/(\pi S_\lambda).\qquad(5.21)$$

Let us re-write equations (5.8-5.9) separating the kernel weights. First, separate the small terms proportional to the product $c_0\rho_2(\mu_0)$ into the non-linear term:

$$R^{nl}(\mu_0,\mu)=\alpha c_0\,\rho_2(\mu_0)\,e^{-\tau/\mu_0}\{e^{-\tau/|\mu|}\rho_1(\mu)+k^LG^{av}(\mu)+k^GG_G^{11}(\mu)+k^VG_V^{11}(\mu)\}.\qquad(5.22)$$

Second, collect the remaining multiplicative factors for the kernel weights:

$$F^L(\mu_0,\mu)=(e^{-\tau/\mu_0}+\alpha\mu_0^{-1}E_0^d(\mu_0))(e^{-\tau/|\mu|}+G^{av}(\mu)),\qquad(5.23)$$

$$F^k(\mu_0,\mu;\varphi)=\{e^{-\tau/\mu_0}f_k(\mu_0,\mu,\varphi)+\alpha\mu_0^{-1}D_k^1(\mu_0,\mu,\varphi)\}e^{-\tau/|\mu|}+$$
$$e^{-\tau/\mu_0}G_k^1(\mu_0,\mu,\varphi)+\alpha\mu_0^{-1}H_k^1(\mu_0,\mu,\varphi),\,k=V,\,G.\qquad(5.24)$$

With these notations, the TOA reflectance becomes:

$$R(\mu_0,\mu,\varphi)=R^D(\mu_0,\mu,\varphi)+k^LF^L(\mu_0,\mu)+k^GF^G(\mu_0,\mu,\varphi)+k^VF^V(\mu_0,\mu,\varphi)+R^{nl}(\mu_0,\mu).\qquad(5.25)$$

This equation, representing TOA reflectance as an explicit function of the BRF model parameters, provides the means for an efficient atmospheric correction.

Let us derive a modified form of this equation which is used in the aerosol retrievals. The last non-linear term of formula (5.25), which describes multiple reflections of the direct-beam sunlight between the surface and the atmosphere, is small ($R^{nl} \propto qc_0$), and can be neglected for simplicity of further consideration. The functions F^k are still weakly non-linear via parameter α, which describes multiple reflections of the diffuse incident sunlight. By setting $\alpha=1$, we omit this non-linearity and equation (5.25) becomes a linear function of the BRF parameters. With an additional assumption of spectral invariance of the BRF shape (Eq. (1.2)), formula (5.25) can be re-written for the pixel (i, j) and observation day k as:

$$R_{ij}^{k}(\lambda) \cong R^{D}(\lambda, \tau^{k}) + b_{ij}(\lambda)Y_{ij}(\lambda, \tau^{k}), \qquad (5.26)$$

where $b_{ij}(\lambda)$ is spectral regression coefficient for a given spectral band, and function

$$Y_{ij}(\lambda, \tau^{k}) = k_{ij}^{L,B7}F^{L}(\lambda, \tau^{k}) + k_{ij}^{G,B7}F^{G}(\lambda, \tau^{k}) + k_{ij}^{V,B7}F^{V}(\lambda, \tau^{k}) \qquad (5.27)$$

can be calculated from the look-up table (LUT) for a given geometry, AOT and wavelength, once the BRF parameters in band B7 for the pixel (i, j) are known. The pressure- and water vapor corrections of the LUT functions are performed with the algorithm described in sec. 5.3.

5.2 *MAIAC* Look-Up Tables

The LUT stores functions f_k^1, f_k^2, f_k^3, which depend on geometry of observations, and functions D_k^1, D_k^3, G^{av}, G_k^1, G_k^{11}, H_k^1, E_0^d, E_0, R^D, which depend on geometry, selected aerosol model and AOT. Index k refers to either volumetric (*V*) or geometric-optical (*G*) BRF kernel function. Following MISR algorithm [*Diner et al.*, 2001; *Diner et al.*, 1999], we store only a multiple-scattering path reflectance (R^{Dms}) in the LUT, and single-scattering part is calculated exactly for a given geometry, pressure and water vapor. The LUT is computed for a dense grid of VZA, SZA, and azimuthal angles ($\Delta\mu_0 = \Delta\mu$=0.02 for the range 0.4 – 1 (0° - 66.42°), and $\Delta\varphi$= 3°). Similarly to MISR processing, the algorithm uses the nearest neighbor angle for speed consideration because it avoids 3D interpolation in angles. In addition, the LUT stores spherical albedo of atmosphere c_0 which depends on aerosol model and AOT. The LUT is calculated for 11 AOT values, {0, 0.03, 0.1, 0.2, 0.3, 0.4, 0.6, 0.9, 1.4, 2, 3}. A linear interpolation is used to derive LUT functions for the required AOT. The grid density was selected empirically from considerations of accuracy and minimum required memory. For example, AOT retrievals from MODIS data using described AOT grid show insignificant difference from retrievals using four times or twice as dense AOT grid. The current size of the LUT for Rayleigh atmosphere, and for one fine and one coarse aerosol fraction for 7 MODIS land bands is 78 MB.

The radiative transfer calculations were done with LUT-generation software based on the code SHARM [*Lyapustin*, 2005] and the Interpolation and Profile Correction (IPC) method [*Lyapustin*, 2003]. The IPC method is designed for fast line-by-line calculations in the spectral interval of interest with flexible spectral resolution of 0.01 – 1 cm^{-1} and an accuracy of several tenths of a percent. The line-by-line calculations are then integrated directly with solar irradiance [*Kurucz*, 1997] and sensor' relative spectral response (RSR) function of MODIS (*X. Xiong*, personal communication). The radiative transfer model (RTM) included absorption of 6 major atmospheric gases (H_2O, CO_2, CH_4, NO_2, CO, N_2O) calculated for the HITRAN-2000 [*Rothman*

et al., 2003] database using a Voigt vertical profile, and the Atmospheric Environmental Research (AER) continuum absorption model [*Mlawer et al.*, 2006]. Because ozone absorption is corrected separately, it was not included in LUT calculations. The LUT is generated for a fixed column water vapor, τ^w=0.5 cm. The correction of LUT functions for the water vapor variations is done analytically, as described in section 3.3.

The LUT is calculated as follows: functions $R^{Dms}(\mu_0, \mu, \varphi)$, $E_0(\mu_0)$, $E_0^d(\mu_0)$, c_0, λ_C^{eff} (effective band center wavelength), and τ_λ^R, τ_λ^g (in-band effective Rayleigh and gaseous absorption optical thicknesses) are calculated first with RSR of sensor. For example, the atmospheric path reflectance is calculated using the following expression:

$$R^D(\mu_0, \mu, \varphi) = \int_{\Delta\lambda} S_\lambda R_\lambda^D(\mu_0, \mu, \varphi) h_\lambda d\lambda \bigg/ \int_{\Delta\lambda} S_\lambda h_\lambda d\lambda . \qquad (5.28)$$

The effective band center wavelength is defined as a wavelength for which monochromatic and narrow-band direct vertical transmittances of the aerosol-free atmosphere are equal:

$$\exp\{-\tau^R(\lambda_{Band}^{eff})\} = \int_{\Delta\lambda} S_\lambda \exp\{-\tau^R(\lambda)\} h_\lambda d\lambda \bigg/ \int_{\Delta\lambda} S_\lambda h_\lambda d\lambda . \qquad (5.29)$$

On the next step, the functions of kernels ($D_k^1(\mu_0, \mu, \varphi)$, $G^{av}(\mu)$, $G_k^1(\mu_0, \mu, \varphi)$, $G_k^{11}(\mu)$, $H_k^1(\mu_0, \mu, \varphi)$, k=v, go) are calculated using the monochromatic RT at the band center wavelengths λ_{Band}^{eff}, and with the optical thickness of the in-band gaseous absorption:

$$\tau_{\Delta\lambda}^g = -\ln\{\int_{\Delta\lambda} S_\lambda \exp\{-\tau^{Gas}(\lambda)\} h_\lambda d\lambda \bigg/ \int_{\Delta\lambda} S_\lambda h_\lambda d\lambda\} . \qquad (5.30)$$

Because functions of kernels are calculated for a large number of quadrature and view geometry angles, this approach is selected for its speed. The monochromatic solution provides a good accuracy because the gaseous absorption in MODIS bands B1-B16 is low. The in-band absorption optical thickness is calculated for the column water vapor W=0.5 cm, carbon dioxide concentration of 380 *ppm*, and concentration of four other major gases (CH_4, NO_2, CO, N_2O) corresponding to the US1976 Standard atmospheric model [*Kneizys et al.*, 1996]. Because ozone absorption in *MAIAC* algorithm is corrected separately, the LUT functions are calculated with zero ozone concentration. The values of λ_C^{eff}, τ_λ^R, τ_λ^g and column absorption optical thickness of water vapor τ_λ^w calculated for MODIS TERRA land bands (B1-B7) are shown in Table 1.

Following MODIS [*Remer et al.*, 2005] and MISR [*Diner et al.*, 2001] aerosol algorithms, the LUT is calculated for the fine and coarse aerosol fractions separately. Because of this, aerosols can be retrieved with various models constructed by mixing the fine and coarse aerosol modes in different proportions, while keeping the LUT size relatively small. Calculations for the aerosol mixtures are performed with linear mixing method (LMM) [*Wang and Gordon*, 1994] for all functions except path radiance. For high accuracy of calculations, we are using LMM for the single scattering path radiance, which is exact in this case, and a modified LMM [*Abdou et al.*, 1997] for the multiple scattering part. The modified method is remarkable for its high accuracy in conditions of high AOT or larger aerosol absorption, when the standard LMM breaks. The pressure- and water vapor corrections of the LUT functions are done with the algorithm described below [*Lyapustin et al.*, 2008a, submitted].

λ^{eff} , μm	0.6449	0.8556	0.4655	0.5535	1.2419	1.6290	2.1131
τ^R	0.05086	0.01622	0.19258	0.09474	0.00362	0.00122	0.00043
τ^g , W=0	1.32e-3	1.82e-5	2.06e-3	4.60e-4	2.90e-3	1.01e-2	1.80e-2
τ^w , 1 cm	3.62e-3	5.49e-3	5.20e-5	3.63e-4	3.62e-3	0.87e-3	1.63e-2
τ^w , 3 cm	1.03e-2	1.32e-2	1.60e-4	1.08e-3	1.00e-2	2.74e-3	4.23e-2
τ^w , 5 cm	1.63e-2	2.00e-2	2.60e-4	1.80e-3	1.58e-2	4.83e-3	6.45e-2

Table 1. The effective center wavelength of MODIS TERRA land bands, and the in-band optical thickness of Rayleigh scattering, of gaseous absorption, and of water vapor absorption for three different levels of column water vapor.

5.3 Surface Pressure (Height) and Water Vapor Correction of LUT Functions

A correction of the LUT functions for variations of surface pressure/height and atmospheric water vapor is a common task in the operational aerosol retrieval and atmospheric correction algorithms. Similarly to MISR algorithms, we calculate the single scattering path radiance at effective band center wavelengths analytically for a given surface pressure and water vapor. A new analytical method has been developed for the water vapor correction of the multiple scattering path radiance. It is described below. The surface-reflected radiance is corrected approximately using the two-way direct transmission function. The pressure correction is implemented via a wavelength shift from the band center, which achieves the required Rayleigh optical depth reduced by pressure. Our extensive numerical study shows that the accuracy of this approach is generally better than 1-4%.

The MODIS spectral channels were carefully selected to avoid or minimize absorption by water vapor and other atmospheric gases. As a result, bands B3 and B8 are not affected by the water vapor absorption. In bands B4, B9-B11, B14, and B16, the water vapor absorption is very weak. It causes less than 0.4% change in the LUT functions when the CWV changes from 0.05 cm to 5 cm. Thus, the water vapor correction is not performed on bands B3-B4, B8-B11, B14, and B16.

The surface height (pressure) defines the amount of molecular scattering in the atmospheric column, which is a function of Rayleigh optical thickness. τ^R rapidly decreases with wavelength approximately as $\lambda^{-4.09}$ (see Table 1). For bands B5-B7, the Rayleigh optical thickness is very low, and change of τ^R with surface pressure is negligible. For these bands, the surface pressure correction is not performed.

5.3.1 Correction of Multiple Scattering Path Radiance for Water Vapor Variations

We are using a perturbation technique, originally developed for the narrow-band and broad-band radiative transfer [*Lyapustin*, 2003]. Let I^m stand for the multiple scattering path radiance at a reference column water vapor W_0. Given the coefficients of absorption $k(z)$, scattering $\sigma(z)$, and extinction $\alpha(z) = \sigma(z) + k(z)$, and scattering function $\chi(\gamma)$, I^m is described by the following radiative transfer equation:

$$\mu \frac{\partial I^m}{\partial z} = -\alpha I^m + \frac{\sigma}{4\pi} \int \{I^m(s') + I^1(s')\} \chi(\gamma) ds' . \tag{5.31}$$

A change in the water vapor content $W=W_0+\delta W$ and related change in extinction $\alpha^\delta(z) = \alpha(z) + \delta\alpha(z)$ perturbs the multiple scattering radiance $I^m + \delta I^m$:

$$\mu \frac{\partial(I^m + \delta I^m)}{\partial z} = -(\alpha + \delta\alpha)(I^m + \delta I^m) + \frac{\sigma}{4\pi} \int (I^m + I^1 + \delta I^m + \delta I^1) \chi(\gamma) ds' . \tag{5.32}$$

Subtracting equation (5.31) from (5.32) and dividing the result by $\alpha^\delta(z)$, we obtain equation for the variation of multiple scattering radiance:

$$\mu \frac{\partial(\delta I^m)}{\partial \tau^\delta} = -(\delta I^m + \frac{\delta\alpha}{\alpha^\delta} I^m) + \frac{\omega^\delta}{4\pi} \int \{\delta I^m + \delta I^1\} \chi(\gamma) ds' , \tag{5.33}$$

where $\tau^\delta = \int \alpha^\delta(z) dz$ is an optical thickness of atmosphere with water vapor W. To evaluate the scattering integral, we assume that the angular dependence of variation $\delta I^m + \delta I^1$ is relatively small compared to that of phase function, and the variation term can be taken outside of the integral sign. Then, Eq. (5.33) turns into an ordinary differential equation

$$\mu \frac{\partial(\delta I^m(s))}{\partial \tau^\delta} + \delta I^m(s)(1 - \omega^\delta) = \omega^\delta \delta I^1(s) - \frac{\delta\alpha}{\alpha^\delta} I^m(s) . \tag{5.34}$$

This equation has constant coefficients ω^δ and $\frac{\delta\alpha}{\alpha^\delta}$ within homogeneous atmospheric layers [i, $i+1$], and a following solution on the interfaces of layers:

$$\{\delta I^m e^{\frac{\tau^\delta(1-\omega^\delta)}{\mu}}\}_{i+1} - \{\delta I^m e^{\frac{\tau^\delta(1-\omega^\delta)}{\mu}}\}_i = \frac{1}{\mu} \int_{\tau_i^\delta}^{\tau_{i+1}^\delta} (\omega^\delta \delta I^1 - \frac{\delta\alpha}{\alpha^\delta} I^m) e^{\frac{\tau'(1-\omega^\delta)}{\mu}} d\tau' . \tag{5.35}$$

To perform integration, we are using a linear approximation of known functions δI^1 and I^m within the homogeneous layer, $\delta I^1(\tau) = \delta I_i^1 + g_i(\tau - \tau_i^\delta)$, $I^m(\tau) = I_i^m + e_i(\tau - \tau_i^\delta)$. Along with the boundary conditions at the top (TOA) and bottom (BOA) of the atmosphere, this yields the following solution:

$$\delta I_{i+1}^m e^{\frac{\tau_{i+1}^\delta(1-\omega_i^\delta)}{\mu}} - \delta I_i^m e^{\frac{\tau_i^\delta(1-\omega_i^\delta)}{\mu}} = \beta_i J_0 + \gamma_i J_1 , \tag{5.36}$$

$$\delta I_0^m = 0 , \; \mu>0 \text{ (TOA, downward directions)}, \tag{5.36a}$$

$$\delta I_N^m = 0 , \; \mu<0 \text{ (BOA, upward directions)}, \tag{5.36b}$$

where

$$\beta_i = \omega_i^\delta [\delta I_i^1 - g_i \tau_i^\delta] - (\frac{\delta\alpha}{\alpha^\delta})_i [I_i^m - e_i \tau_i^\delta] ; \quad \gamma_i = \omega_i^\delta g_i - (\frac{\delta\alpha}{\alpha^\delta})_i e_i ;$$

$$J_0 = \frac{1}{\mu} \int_{\tau_i^\delta}^{\tau_{i+1}^\delta} e^{\frac{\tau'(1-\omega_i^\delta)}{\mu}} d\tau' ; \; J_1 = \frac{1}{\mu} \int_{\tau_i^\delta}^{\tau_{i+1}^\delta} \tau' e^{\frac{\tau'(1-\omega_i^\delta)}{\mu}} d\tau' .$$

Since only the TOA value of path radiance is stored in the LUT, regardless of the number of atmospheric layers used in the radiative transfer calculations, we can only use a solution for a

45

homogeneous (single layer) atmosphere. Let us denote $E = e^{\frac{\tau^\delta (1-\omega^\delta)}{\mu}}$, and $J_0 = \frac{(E-1)}{1-\omega^\delta}$, then the final expression for the correction term can be written as follows:

$$\delta I^m = \beta \frac{1 - \mu J_0 / \tau^\delta}{1 - \omega^\delta} \ (\mu < 0), \ \beta = \omega^\delta \delta I^1 - (\frac{\delta \alpha}{\alpha^\delta}) I^m. \tag{5.37}$$

To assess the accuracy of this very simple expression, we have performed extensive numerical simulations of path radiance in the MODIS TERRA bands for different aerosol types and atmospheric moisture. The spectral TOA reflectance was simulated using Eq. (5.28). The results of accuracy analysis for a typical continental aerosol model and a dust model from AERONET [*Holben et al.*, 1998] classification [*Dubovik et al.*, 2002] are shown in Figure 16. The continental aerosol is represented by the urban low absorption model for the Goddard Space Flight Center (GSFC), USA, and the dust is described by the model for the Solar Village, Saudi Arabia. The column water vapor in calculations was varied from 0.3 cm to 6 cm. The typical range of water vapor values for the USA mid-latitudes is 0.3-0.5 cm for the winter to 1.5-4.0 cm in the summer.
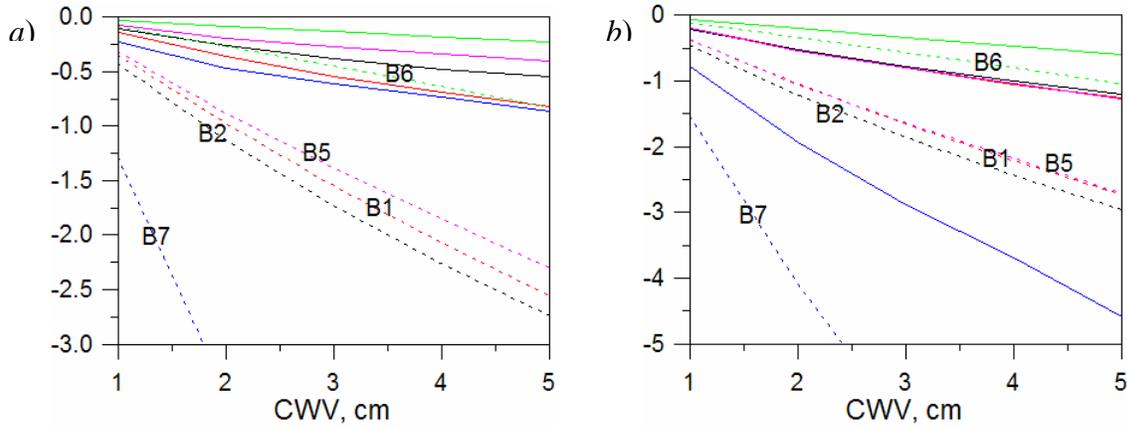
Figure 16. A relative error of path radiance (%) in the MODIS land bands with (solid lines) and without (dotted lines) water vapor correction. The uncorrected curves show the change in path radiance with water vapor relative to the baseline value of W=0.5 cm ($L(W) - L(0.5cm) / L(W) *100\%$). The aerosol model represents *a*) urban low absorption conditions (GSFC AERONET model) with moderate to low optical thickness (τ^a_{B1}=0.18, τ^a_{B7}=0.03), and *b*) dust (Solar Village AERONET model, Saudi Arabia), medium optical thickness (τ^a_{B1}=0.53, τ^a_{B7}=0.53).

The dotted lines show the change of uncorrected path radiance with water vapor relative to the baseline calculations at W=0.5 cm in %. In the MODIS TERRA red (B1) and near-IR (B2) bands, the error in path radiance due to WV stays within 1-3%. In the 2.1 μm region (B7), the error is much higher, 2-12%. The solid lines show the accuracy of path radiance with multiple scattering correction. The results are shown for a typical geometry of SZA=VZA=45°, and a relative azimuth 0°. Figure 3 shows that except B7, the accuracy of corrected path radiance is generally better than 0.5-1%. Summarizing, the MS correction reduces the error of the path radiance by a factor of 2-10, depending on viewing geometry.

5.3.2 Surface Pressure/Height Correction

The *MAIAC* LUT was calculated for the standard atmospheric pressure P_0=1013 *mB* and W_0=0.5 *cm*. The change in the surface elevation calls for pressure correction of the LUT functions. For purely Rayleigh and non-absorbing atmosphere, the pressure correction at wavelength λ_c can be achieved by sliding the wavelength from the band center in order to agree with the pressure-reduced Rayleigh optical depth:

$$\tau^R(\lambda) = \frac{P(z)}{P_0} \tau^R(\lambda_c). \tag{5.38}$$

In this case, the pressure correction is done via spectral interpolation between the given band and its longer wave neighbor. Because Rayleigh optical depth rapidly changes with wavelength, $\tau^R(\lambda) \approx a\lambda^{-4.09}$, such a correction requires only a small wavelength shift, e.g. $\approx$2.5% for $P(z)$=900 *mB*, or $\approx$5.3% for 800 *mB*. For this reason, this method of correction can be applied for the atmospheres with aerosol, because the associated changes of the aerosol phase function or absorption are small. This method, originally suggested by *R. Fraser*, is used by MODIS C5 aerosol algorithm [*Levy et al.*, 2007].

It needs to be emphasized that the necessary requirement for using this method is that the two neighbor bands should have a low and similar gaseous absorption. Table 1 shows that at a reference W_0=0.5 cm column water vapor the absorption is similar in MODIS bands B1-B4 (τ^g~10^{-3}). On the other hand, if a climatology column water vapor $W_0 \approx$1.5-2 cm were used to generate a LUT [*Levy et al.*, 2007], the green, red and NIR channels will have a significantly different absorption ($\Delta\tau^g$~10^{-2}). The differential spectral absorption will create an error during pressure correction which may be comparable in magnitude to the corrected effect of pressure (height) variations.

5.3.3 Linear Mixing Method

Let us describe LMM in application to the single-scattering (SS) path reflectance. First, the SS components, corresponding to the fine and coarse aerosol fractions, are calculated analytically for a given pressure and water vapor. For this purpose, we store the respective aerosol phase functions and spectral functions $h_f(\lambda)$, $h_c(\lambda)$, explained below, in the LUT.

Next, the function $R^{D^{ss}}$ is calculated for the mixture of the fine and coarse aerosol fractions using LMM:

$$R^{D^{ss}} = \sum f_i R^{D^{ss}}_i, \tag{5.39}$$

where $f_i = \tau_i^a / \tau^a$ is a fractional contribution to the total AOT $\tau^a = \sum \tau_i^a$. The AOT can be expressed using volumetric concentrations of fractions (C_{Vf}, C_{Vc}):

$$\tau^a(\lambda) = \tau_f^a(\lambda) + \tau_c^a(\lambda) = C_{Vf}h_f(\lambda) + C_{Vc}h_c(\lambda) = C_{Vf}(h_f(\lambda) + \eta h_c(\lambda)), \tag{5.40}$$

where $\eta = C_{Vc}/C_{Vf}$ is spectrally-independent ratio of volumetric concentrations of coarse and fine fractions, and $h_i(\lambda) = \tau_i^a(\lambda)/C_{Vi}$ is a fractional optical thickness per unitary concentration. For example, $\eta \approx$0.5 for the GSFC aerosol model at moderate AOT. Given the size distribution and refractive index (given $h_f(\lambda)$, $h_c(\lambda)$), the spectral slope of AOT (Angstrom parameter) is

defined by the ratio η. Given parameter η and total AOT, the volumetric concentration of the fine fraction is defined as $C_{Vf} = \tau^a(\lambda)/(h_f(\lambda) + \eta h_c(\lambda))$. Finally, the weighting factors of the LM method do not depend on the total AOT,

$$f_f(\lambda) = C_{Vf} h_f(\lambda)/\tau^a(\lambda) = h_f(\lambda)/(h_f(\lambda) + \eta h_c(\lambda)), \ f_c(\lambda) = 1 - f_f(\lambda). \tag{5.41}$$

5.3.4 Calculating TOA Reflectance from LUT given Surface Pressure and Water Vapor

The algorithm consists of several steps:

1) Calculate R^{Dss} analytically for a given geometry, AOT, surface pressure, and column water vapor for selected fine and coarse aerosol fractions, and add them using LMM (5.39) for a given ratio of volumetric concentrations η.

2.1) Calculate pressure-shifted effective center wavelength of the MODIS channel λ_1:
$\lambda = \lambda_1 (P/P_0)^{1/4.09}$, $\lambda_1 \leq \lambda \leq \lambda_2$;

2.2) Calculate LUT functions (R^{Dms}, D_k^1, D_k^3, G^{av}, G_k^1, G_k^{11}, H_k^1, E_0^d, E_0, c_0) for the required AOT for two wavelengths λ_1 and λ_2 for the fine and coarse aerosol fractions.

2.3) Perform pressure correction for all LUT functions by linear spectral interpolation of LUT-functions between λ_1 and λ_2 to λ.

3) Calculate multiple-scattering path reflectance using a modified LMM:

$$R^{D^{ms}} = R^{D^{ms}_R} + \sum \frac{\omega_{mix}}{\omega_i} e^{-\tau^a|\omega_i - \omega_{mix}|} f_i (R^{D^{ms}_i} - R^{D^{ms}_R}), \tag{5.42}$$

where $R^{D^{ms}_R}$ is the multiple scattering path reflectance due to Rayleigh scattering stored in the LUT, and $\omega_{mix} = \sum f_i \omega_i$ is the single scattering albedo of mixture.

4) Perform water vapor correction for R^{Dms}. Let τ^{W_0} and τ^W be the water vapor absorption optical thickness for $W_0=0.5$ cm, for which the LUT was calculated, and for W. Then, calculate

4.1) $\tau^\delta = \tau^a + \dfrac{P}{P_0}(\tau^R + \tau^g) + \tau^W$, $\dfrac{\delta\alpha}{\alpha^\delta} = \dfrac{\tau^W - \tau^{W_0}}{\tau^\delta}$, $\omega^\delta = \dfrac{\tau^a \omega^a + (P/P_0)\tau^R}{\tau^\delta}$, $\qquad$ (5.43)

and a variation of single scattering path radiance

$$\delta I^{ss} = I^{ss}(\lambda_1; P, W_0) - I^{ss}(\lambda_1; P, W); \tag{5.44}$$

4.2) correction term $\delta I(\lambda_1; P, W_0, W)$ using formula (5.37). The corrected path reflectance is:

$$R^D(\lambda_1; P, W) = R^{Dss}(\lambda_1; P, W) + R^{Dms}(\lambda_1; P, W_0) + \delta I(\lambda_1; P, W_0, W)/\mu_0. \tag{5.45}$$

4.3) Finally, the TOA reflectance is a sum of path reflectance (5.45), and a pressure-corrected surface-reflected signal R_s (sum of terms (5.2) and (5.4)), additionally corrected for the difference in the gaseous absorption via the two-way direct transmittance:

$$R(\lambda_1) = R^D(\lambda_1, P, W) + R_s(\lambda_1, P, W_0) \exp(-\{(\tau^W - \tau^{W_0})\tilde{m} - (1 - P/P_0)\tau^g m\}) \tag{5.46}$$

where $m = \mu_0^{-1} + |\mu|^{-1}$ is an atmospheric air mass and $\tilde{m} = |\mu|^{-1/2} + \mu_0^{-1/2}$ is an effective water vapor air mass.

5.4 Summary

The described algorithm renders an accurate representation of the TOA reflectance and its natural variability based on a rather compact look-up table. It takes advantage of the accumulated knowledge and developments of the MISR and MODIS science teams of the past 10-20 years. The radiative transfer model accounts for anisotropy of surface reflectance, variations of surface pressure/height, and for atmospheric absorption due to ozone, uniformly mixed gases and variable column water vapor. Aerosols are represented by several models as a combination of fine and coarse modes. Only a multiple scattering component of the atmospheric path reflectance is stored in the LUT, whereas the single scattering part is calculated using an exact formula "on the fly". The other radiative transfer functions are computed from LUT for a given aerosol model using the linear mixing and the modified linear mixing methods. The dense grid of solar and view angles makes possible to use nearest neighbor avoiding 3D interpolation in angles, which achieves a high speed required in the operational processing.

Our considerable accumulated experience of processing MODIS data shows that the LUT algorithm has a required accuracy for aerosol retrievals and atmospheric correction of MODIS data in the spectral range of 0.47 – 2.2 μm. Perhaps, a small improvement could be achieved by adding polarization to the shortwave channel 0.47 μm, which is a relatively straightforward task.

6. Aerosol Algorithm

The aerosol algorithm consists of two steps: deriving spectral regression coefficients (SRC), and retrieving AOT and Angstrom exponent. The SRC retrievals use parametric formula (5.26) and involve two assumptions which are verified by the algorithm: 1) the surface reflectance changes little during accumulation period; 2) AOT changes little at within the block area.

6.1 SRC Retrievals

Let us assume that the ancillary information for the aerosol retrievals, including water vapor, cloud mask, and surface BRF in band B7, is available. Let us also assume that gridded TOA MODIS reflectance data is available for $3 \leq K \leq 16$ cloud-free days, which form the processing Queue. Our goal is to derive the set of K AOT values for different days (orbits), and N^2 SRC values for the Blue band (B3) for a given 25 km block of the surface. The SRC algorithm is implemented in three steps:

1) Select the clearest day from the Queue;
2) Calculate the AOT difference for every day with respect to the clearest day, $\Delta \tau^k = \tau^k - \tau_0$;
3) Find AOT on the clearest day, τ_0. At this step, the algorithm simultaneously generates the full set of spectral regression coefficients.

The first task is solved as follows. Initially, the SRCs are calculated for every day and every pixel separately using formula (5.26) for AOT=0. For a given pixel, the coefficient b_{ij}^k is lowest on the clearest day because its value is increased by the path reflectance on hazier days. Therefore, the clearest day is selected as a day with the lowest on average set of coefficients b_{ij}^k in the block.

In the next step (2), the AOT difference between the day k and the clearest day is calculated independently for every day of the Queue by minimizing the difference

$$F_1^k = \frac{1}{N^2} \sum_{i,j} \{b_{ij}^{Clear} - b_{ij}^k (\Delta \tau^k)\}^2 = \min\{\Delta \tau^k\}. \tag{6.1}$$

The SRCs for the clearest day (b_{ij}^{Clear}) have been calculated for $\tau_0 = 0$ in step 1. When solving Eq. (5.57), SRCs for the day k are re-calculated for the increasing values of AOT from the LUT τ^k ($\Delta \tau^k = \tau^k - \tau_0 = \tau^k$) until the minimum is reached. This operation is equivalent to simultaneous removal of bias and "stretching" the contrast for a given block that minimizes the overall difference.

In step 3), AOT on the clearest day is found by minimization of *rmse* between the theoretical reflectance and the full set of measurements for K days and N^2 pixels:

$$F_2 = \sum_{K} \sum_{i,j} \{R_{ij}^{Meas,k} - R_{ij}^{Th,k}(\tau_0 + \Delta \tau^k)\}^2 = \min\{\tau_0\}. \tag{6.2}$$

To calculate theoretical reflectance with Eq. (5.26), one needs to know the coefficients b_{ij}. These are calculated using the first assumption described in Introduction, namely that the surface reflectance changes little during K days. Therefore, for a given pixel (*i,j*) and given value τ_0, the SRC can be found by minimizing the *rmse* over all days of the Queue:

$$F_{ij} = \sum_K \{R_{ij}^{Meas,k} - R_{ij}^{Th,k}(\tau^k)\}^2 = \min\{b_{ij}\}, \ \tau^k = \tau_0 + \Delta\tau^k, \tag{6.3}$$

which is solved by the least-squares method ($\partial F_{ij}/\partial b_{ij} = 0$) with the analytical solution:

$$b_{ij} = \sum_K [R_{ij}^{Meas,k} - R^D(\tau^k)]Y_{ij}(\tau^k)/\sum_K \{Y_{ij}(\tau^k)\}^2. \tag{6.5}$$

Thus, given the aerosol model, Eq. (6.5) becomes parameterized in terms of the only parameter τ_0. Equations (6.1) and (6.2) are positively defined quadratic forms which have unique solutions. To solve these equations numerically, *MAIAC* incrementally increases AOT (e.g. τ_0 in Eq. (6.2)) using the LUT entries, until the minimum is found. Because the discretization of LUT in AOT is relatively coarse, the algorithm finds the "bend" point, where function F_2 starts increasing, approximates the last three points, encompassing the minimum, with quadratic function, and finds the minimum analytically. The set of SRCs is calculated with the final value τ_0 from Eq. (6.5).

The algorithm controls positiveness of surface BRF by limiting the maximal retrieved AOT ($\tau^k = \tau_0 + \Delta\tau^k$). Specifically, we require that the path reflectance in the blue (B3) and "deep blue" (B8) bands does not exceed the minimal measured reflectance in the block ($R_{\lambda,\min}$),

$$R_{B3}^D(\tau_{B3,\max}^a) < R_{B3,\min}, \ R_{B8}^D(\tau_{B8,\max}^a) < R_{B8,\min}.$$

This algorithm was developed and optimized through a long series of trial and error. It requires at least three clear or partially clear days in the Queue for the inversion, with at least 50% of the pixels of the block being clear for three or more days. The algorithm has a self-consistency check, verifying whether the main assumptions hold. For the last day, this is done during step 2 processing. If the surface had undergone a rapid change during the accumulation period (e.g. a snowfall, or a large-scale fire, flooding or rapid landcover conversion, with the size of disturbance comparable to the block size), or if the AOT changes significantly inside a given block on day k, then the value of *rmse* $\sqrt{F_1^k}$ remains high. Currently, the algorithm excludes such days from the processing Queue based on a simple empirically established threshold $\sqrt{F_1^k} \geq 0.03$. In regular conditions, the value $\sqrt{F_1^k}$ is usually lower than 0.01-0.015.

The sensitivity of solution to surface reflectance and SRC reduces when the AOT grows. At high AOT on one or more days, the algorithm may retrieve very low value τ_0 resulting in positive errors in retrieved SRC. To avoid that, the *previous* days stored in the Queue are filtered for a given block if the average retrieved blue band AOT for a slant path is high, AOT$_{av}$/cosVZA>0.6. We also filter days in which AOT variation in the block is high (AOT$_{av}$- AOT$_{min}$ > 0.15). The last two filters are applied only over the dark surfaces (R$_{B7}$<0.2).

Retrieving SRCs is a well-optimized and a relatively fast process. For every block, the retrieved spectral regression coefficients are stored in the Q-memory, along with the band B7 LSRT coefficients. They are used as ancillary information for the aerosol retrievals at 1 km grid resolution, which are described next.

6.2 Aerosol Retrievals

This algorithm requires a set of aerosol models with increasing particle size and asymmetry parameter of scattering. The aerosols are modeled as a superposition of the fine and coarse fractions, each described by a log-normal size distribution. Some of the currently used models are listed in Table 2. For example, for the continental USA we are currently using the moderate absorption Fine model 2 and generic Coarse model 5. By varying the ratio of volumetric concentrations of coarse and fine fractions, $\eta = C_v^{Coarse}/C_v^{Fine}$, a wide range of asymmetry (size) parameter is simulated. The LUT is originally computed for the fine and coarse fractions separately. When *MAIAC* reads the LUT, it generates a series of mixed aerosol LUTs for several different values $\eta = \{0.5; 1.5; 2.5; 4; 8\}$, which are stored in the operational memory. In this sequence, value $\eta = 0.5$ gives a model that is close to the urban continental moderate absorption (GSFC) model from AERONET classification [*Dubovik et al.*, 2002], whereas the values $\eta = 2\text{-}5$ are more representative of the mineral dust. As described above in section 5.3, a modified linear mixing algorithm is used to compute required radiative transfer (RT) functions from the LUT.

Table 2. Aerosol Fractions Currently Used in *MAIAC* Algorithm: R_V, σ_V are the median volumetric radius and standard deviation, and (n_r, n_i) is refractive index.

N	Fraction Type	R_V (μm)	σ_V (μm)	n_r	n_i (0.4-1.2 μm)	n_i (2.2 μm)
1	Fine, very low abs.	0.12	0.35	1.41	0.004	0.001
2	Fine, low absorption	0.14	0.38	1.41	0.006	0.001
3	Fine, industrial/biomass burning, moderate abs.	0.12	0.35	1.47	0.012	0.003
4	Fine, biomass burning high absorption	0.14	0.35	1.47	0.02	0.005
5	Coarse, generic	2.9	0.75	1.41	0.004	0.001
6	Coarse, absorbing	2.9	0.75	1.41	0.012	0.003
7	Coarse, high absorb.	2.8	0.7	1.47	0.02	0.005
8	Coarse, mineral dust	2.2	0.6	1.56	0.0025	0.001
9	Cloud, liquid water	5.0	0.1	1.33	According to *Hale and Querry* [1973]	

With spectral regression coefficients retrieved, the surface BRF in every grid cell in the Blue band becomes known (Eq. 1.2). Further, the AOT and Angstrom parameter are retrieved at 1 km resolution from the last *Tile* of MODIS measurements.

For each pixel, the retrieval algorithm goes through a loop of increasing values of fractional ratio η, and using known surface BRF $\rho_{ij}^{B3}(\mu_0, \mu; \varphi) = b_{ij}^{B3} \rho_{ij}^{B7}(\mu_0, \mu; \varphi)$ it computes AOT (τ_{ij}) in the Blue band by fitting theoretical TOA reflectance to the measurement

$$R^{Theor,B3}(\eta; \tau_{ij}) = R_{ij}^{Meas,B3}. \tag{6.5}$$

In the next step, a spectral residual is evaluated using the Red (B1), and SWIR (B7) bands:

$$\chi_{ij} = \sum_{\lambda} \{R_{ij}^{Meas,\lambda} - R_{ij}^{Theor,k}(\tau^{\lambda}(\eta))\}^2 = \min\{\eta\} .$$ (6.6)

The procedure is repeated with the next value η until the minimum is found. Theoretical reflectance in (6.6) is computed with the LSRT BRF parameters from the previous cycle of atmospheric correction, which are stored in the Q-memory.

Because MODIS measurements provide only a spectral slice of information, *MAIAC* does not attempt MISR-like retrievals for multiple aerosol models with different absorption and sphericity of particles. Instead, it follows the MODIS *Dark Target* approach [*Levy* et al., 2007] where the aerosol fractions and their specific absorption properties are fixed regionally.

The spectral sensitivity of measurements to variations of the aerosol model in clear atmospheric conditions, especially at longer wavelengths, is limited. Currently, growth of the MODIS footprint with the scan angle is the main source of uncertainty in *MAIAC*'s knowledge of the surface spectral BRF. These errors, although small, can be costly if very asymmetric aerosol model with large AOT values is selected when the atmosphere is actually very clean. For these reasons, the full minimization procedure (6.5-6.6) is performed only when the retrieved optical thickness for the standard continental model (η =0.5) exceeds 0.5. Otherwise, a single value of η =0.5 is used and AOT is reported for these background conditions.

Because atmospheric correction (AC) is an integral part of *MAIAC* aerosol retrievals, we describe the AC algorithm next, followed by examples of retrievals and AERONET validation.

7. Atmospheric Correction Algorithm

Once the cloud mask is created and aerosol retrievals performed, the *MAIAC* algorithm filters the time series of MODIS measurements for every pixel and places the remaining clear-skies data in a "container". The filter excludes pixels with clouds and cloud shadows, as well as snow-covered and water pixels as detected by the CM algorithm during land-water-snow classification. Pixels with high AOT (>0.9), where sensitivity of measurements to the surface reflectance decreases, are also filtered. The container stores measurements along with the LUT-based RT functions for the cloud-free days of the Queue. If the number of available measurements exceeds 3 for a given pixel, then the coefficients of LSRT BRF model are computed.

7.1 Inversion for LSRT Coefficients

In the current operational MODIS land processing, the BRF is determined in two steps: first, the atmospheric correction algorithm derives surface reflectance for a given observation geometry using a Lambertian approximation [*Vermote et al.*, 2002]. Next, three LSRT coefficients are retrieved from the time series of surface reflectance accumulated for a 16-day period [*Schaaf et al.*, 2002]. The Lambertian assumption simplifies the atmospheric correction but creates biases in the surface reflectance which depend on the observation geometry and atmospheric opacity. It is known that Lambertian assumption creates a flatter BRF pattern while the true BRF is more anisotropic [e.g., *Lyapustin*, 1999].

The *MAIAC* algorithm derives LSRT coefficients directly by fitting the radiative transfer solution to the measured TOA reflectance accumulated over a 4-16 day period. The inversion is based on formula (5.25):

$$R(\mu_0,\mu,\varphi)=R^D(\mu_0,\mu,\varphi)+k^L F^L(\mu_0,\mu)+k^G F^G(\mu_0,\mu,\varphi)+k^V F^V(\mu_0,\mu,\varphi)+R^{nl}(\mu_0,\mu). \quad (7.1)$$

The quasi-linear form of equation (7.1) leads to a very efficient iterative minimization algorithm:

$$RMSE=\sum_j (r_j^{(n)}-F_j^L k^{L(n)}-F_j^V k^{V(n)}-F_j^G k^{G(n)})^2 =\min_{\{\vec{K}\}}, \; r^{(n)}=R-R^D-R^{nl(n-1)}, \quad (7.2)$$

where index *j* denotes measurements for different days, and *n* is the iteration number. Equation (7.2) provides an explicit least-squares solution for the kernel weights. In matrix form, the solution is written as:

$$\vec{K}^{(n)}=A^{-1}\vec{b}^{(n)}, \qquad\qquad (7.3)$$

where

$$A=\begin{bmatrix} \sum(F_j^L)^2 & \sum_j F_j^G F_j^L & \sum_j F_j^V F_j^L \\ \sum_j F_j^G F_j^L & \sum_j (F_j^G)^2 & \sum_j F_j^V F_j^G \\ \sum_j F_j^V F_j^L & \sum_j F_j^V F_j^G & \sum_j (F_j^V)^2 \end{bmatrix}, \; \vec{b}^{(n)}=\begin{bmatrix} \sum_j r_j^{(n)} F_j^L \\ \sum_j r_j^{(n)} F_j^G \\ \sum_j r_j^{(n)} F_j^V \end{bmatrix}.$$

In the first iteration, the small non-linear term is set to zero, $R_j^{nl(0)}=0$, and the multiple reflection factor α (see sec. 5) is set to one, $\alpha^{(0)}=1$. These parameters are updated once after the BRF coefficients are calculated in the first iteration. Except for snow-covered surfaces, the problem converges with high accuracy in two iterations because the non-linear terms are small. Current version of *MAIAC* algorithm does not make retrievals over snow.

Prior to inversion, the algorithm checks if the dataset has a sufficient angular sampling. The MODIS operational BRDF/albedo algorithm [*Schaaf et al.*, 2002] makes an inversion if at least 7 cloud-free observations are available during 16-day period. We studied this problem experimentally using MODIS data for a number of AERONET sites, varying the minimal required number of measurements (from 3 to 10) and testing different metrics of angular sampling. One metric used the magnitude of determinant of the inverse matrix *A* which shows how different the sampling angles are. Although such analysis is, perhaps, most straightforward theoretically, we found it often too restrictive. In the end, a simple criterion was chosen based on the range of cosine of the view zenith angle (μ_{max}-$\mu_{min} \geq 0.2$), which is usually sufficient to ensure robust and consistent retrieval. The described inversion algorithm is very fast.

7.2 Solution Selection and Update

Although the LSRT model leads to an efficient BRF retrieval algorithm, there are several caveats associated with this model. The LSRT kernels are not orthogonal, are not positive-only functions, and are normalized in a somewhat arbitrary fashion that is not linked to radiative transfer theory. These factors reduce the stability and uniqueness of the solutions, such that small perturbations in measurements may lead to significantly different solutions. The high goodness-of-fit at the measurement angles does not guarantee the correct shape of the retrieved BRF, and

may result in negative BRF values at other angles. The albedo, being an integral function of BRF, is especially sensitive to an incorrect BRF shape. For these reasons, we use several tests to remove unrealistic solutions.

The initial validation of the solution checks that the maximal difference over all days of the Queue between measured and computed TOA reflectance does not exceed a 0.05 threshold, $\left|R^{Meas} - R^{LSRT}\right| > 0.05$. If it does, the day (measurement) with the highest deviation is excluded from the Queue and the inversion is repeated. If the number of measurements goes below four after the exclusion, no retrieval will be made for this pixel in given band.

If a solution provides a good agreement with measurements within 0.05 for all days, the algorithm verifies that values of the direct-beam albedo (q) at SZA=15°, 45°, 60° are positive. Finally, the new solution must be consistent with the previous solution: $\left|q(45^o) - q^{Prev}(45^o)\right| < \Delta(\lambda)$, where Δ is the band-dependent threshold currently equal to 0.04 (blue), 0.05 (green and red), 0.1 (for spectral region of 0.8 – 1.6 μm) and 0.05 for the shortwave infrared band (2.1 μm). Consistency of the time series of BRF and albedo is characterized by a *status* index. Initially, the confidence in the solution is low (*status*=0). Each time the new retrieval agrees with the previous retrieval, *status* increases by 1. When *status*≥3, the retrieval is considered reliable.

The thresholds (0.05 and $\Delta(\lambda)$) in the LSRT inversion routine are selected, on one hand, tight enough to reject most of undetected clouds, which remain the dominant source of errors, and sufficiently loose, on the other hand, for the solution to adapt to the surface change. The most pervasive type of change is seasonal variations, related to the spring green-up and fall senescence at northern latitudes, or greenness variations caused by wet and dry seasons in tropics. The total seasonal variation of reflectance over vegetated surface is about several absolute percent in the visible bands (~0.03-0.1), and is significantly higher in the near infrared (~0.1-0.3). Threshold $\Delta(\lambda)$ for the daily variation was selected accordingly, and our analysis of a large volume of processed MODIS data confirms that *MAIAC* algorithm does not reject measurements when surface is changing, even in the agricultural regions characterized by a rapid reflectance change during harvesting.

When the new solution is validated, the coefficients of the BRF model and direct-beam albedo $q(45^o)$, stored in the Q-memory, are updated. The update is done with relaxation, designed to mitigate random noise of retrievals:

$$\vec{K}_\lambda^{New} = w\,\vec{K}_\lambda^{New} + (1-w)\,\vec{K}_\lambda^{Prev}. \tag{7.4}$$

The weight w depends on our confidence in previous solution, which increases with its *status*. The weight w=1 for the first retrieval (*status*=0), w=0.8 for the second retrieval, w=0.6 for the third consecutive retrieval, and w=0.5 thereafter. This method of update increases the quality of solution when the surface is relatively stable, but it delays the response of the solution to surface changes.

Often, the solution for some pixels or the full area cannot be produced because of lack of clear-skies measurements. In these cases, we assume that the surface does not change and the gaps are filled-in with the previous LSRT solution for up to a 32-day period. During short time intervals, this is the most natural way of gap-filling with specific BRF/albedo solution for a given pixel. In

most cases, the assumption of a stable surface is reasonable. The gap-filled pixels are marked as "Extended" in the quality assurance (QA) value, with parameter QA.*nDelay* giving the number of days since the last update.

7.3 Tracking Surface Change

During spring green-up or fall senescence, the surface may entirely change in a period of one or two weeks. This period is often accompanied by high cloudiness which may be caused by the onset of the rain season in tropics, or by the changing boundary layer meteorology due to thawed soil releasing the moisture into the warmer air at northern latitudes in spring. The surface BRF changes considerably during these periods not only by magnitude but also in shape. In these conditions, the response of the full Queue solution is delayed, often significantly, depending on cloudiness. To ameliorate this problem, we always make two LSRT inversions, one with the full Queue, and another using the last four clear days of the Queue. We compute the maximal difference of both solutions with measurements $dif = |R^{Meas} - R^{LSRT}|$ over the last four days. The four-day solution is less reliable and more prone to errors. Therefore, provided it has lower value of *dif*, we chose the 4-day solution only if surface change has independently been confirmed by measurements. The latter is accomplished as follows:

IF($R_{B7} < R_{B7}^{Th}$-0.03 & ($NDVI - NDVI^{Th}$)>THRESH_NDVI & $R_{B2} > R_{B2}^{Th}$+0.04) $\Rightarrow$ green-up.

IF($R_{B7} > R_{B7}^{Th}$+0.03 & ($NDVI - NDVI^{Th}$)< -THRESH_NDVI & $R_{B2} < R_{B2}^{Th}$-0.04) $\Rightarrow$ senescence.

Here, superscript *Th* indicates theoretical values which are computed using known LSRT parameters for a given day of observations. The *NDVI* threshold has a default value of 0.07, which is increased to 0.1 when the atmosphere is hazy (AOT>0.5).

Thus, when *i*) the 4-day LSRT solution gives a better fit to the last 4 days of measurements, and *ii*) surface change is confirmed by MODIS data, then the *status* of the pixel is dropped to zero and the new solution takes precedence.

7.4 *MAIAC* Surface Reflectance Products

MAIAC computes two main products at 1 km resolution for seven 500m MODIS bands, the set of BRF coefficients, and the surface albedo. The albedo is defined by Equation (5.5a) as a ratio of surface-reflected to incident radiative fluxes. Thus, it represents a true albedo at a given solar zenith angle in ambient atmospheric conditions, the value, which can be directly compared to ground-based measurements.

MAIAC also computes several derivative products useful for science data analysis and validation:
      1) *NBRF* - a BRF Normalized to the common geometry of nadir view and SZA = 45°. This product is analogous to MODIS NBAR (nadir BRF-adjusted reflectance) product (part of the MOD43 standard product suite). With the geometry variations removed, the time series of *NBRF* is useful for studying vegetation phenology, performing surface classification, etc.
      2) *IBRF* – an Instantaneous (or one-angle) BRF value for the specific viewing geometry of the last day of observations. In essence, *IBRF* is a bidirectional reflectance which would be measured if the atmosphere were absent. This product is calculated from the latest MODIS measurement using known BRF shape. To illustrate computation of *IBRF*, we re-write equation for the measured TOA reflectance as follows:

$$R(\mu_0, \mu, \varphi) = R^D(\mu_0, \mu, \varphi) + bR^{Surf}(\mu_0, \mu, \varphi), \qquad (7.5)$$

where R^{Surf} combines all surface related terms and is calculated using the current LSRT solution and retrieved aerosol data. b is spectrally-dependent scaling factor. Then,

$$IBRF_\lambda(\mu_0, \mu, \varphi) = b_\lambda BRF_\lambda(\mu_0, \mu, \varphi), \qquad (7.6)$$

where BRF_λ is calculated using LSRT model. This algorithm (Equation 7.6) will be referred to as scaling. This description was given for the purpose of illustration. In reality, R^{Surf} is a non-linear function so computing parameter b_λ and *IBRF* is done accurately using the formulas given in sec. 5.

The *IBRF* is computed for the land pixel if it is cloud-free, AOT is retrieved and it is not too high (AOT<1), and LSRT coefficients for the pixel are known. Based on its definition, *IBRF* is what is supposed to be produced by the operational MODIS atmospheric correction algorithm (product MOD09). Because *IBRF* is derived using the last day of measurements, it responds to surface changes immediately contrary to LSRT parameters (or NBRF) whose response may be delayed.

A list of *MAIAC* products and their operational MODIS counterparts are given in Table 3.

Table 3. *MAIAC* product suite

Product Name	Data Type	Descriptions	MODIS Product Counterpart
CloudMask	DFNT_UINT8	Cloud mask and Land-Water-Snow dynamic classification.	MOD35 cloud Mask
NBRF	DFNT_FLOAT32	Bi-directional reflectance factor normalized to SZA=45° and nadir view.	MOD43B4 Nadir BRDF-Adjusted Reflectance (NBAR)
Albedo	DFNT_FLOAT32	Surface albedo at a given solar zenith angle in ambient atmospheric conditions.	A combination of MOD43B3 black-sky and white-sky albedo weighted with respective relative direct and diffuse incident fluxes.
IBRF	DFNT_FLOAT32	Instantaneous (or one-angle) BRF for specific viewing geometry of the last day of observations.	MOD09 Surface reflectance
Kiso	DFNT_FLOAT32	The isotropic coefficient of LSRT model.	MOD43B1 BRDF/Albedo Model Parameters
Kvol	DFNT_FLOAT32	The volumetric coefficient of LSRT model.	MOD43B1 BRDF/Albedo Model Parameters
Kgeo	DFNT_FLOAT32	The geometric-optics coefficient of LSRT model.	MOD43B1 BRDF/Albedo Model Parameters
Snow Fraction	DFNT_FLOAT32	Sub-pixel snow fraction (0-1).	N/A
Optical snow grain size	DFNT_FLOAT32	Snow grain size which provides an equivalent spectral reflectance in spectral range 0.4-2.2 μm.	N/A

8. Retrieval Examples and AERONET Validation

8.1 Illustration of the Algorithm Performance

We will illustrate *MAIAC* performance using 50 km MODIS TERRA subsets for the GSFC site (Greenbelt, Maryland). Our usual method of analysis includes visualization of the input and output data in a form of time series, similarly to the data arrangement in the processing Queue. This way we can analyze both the overall quality of solution and complex interrelationships between different parameters. Figure 17 shows 4 different 15-day periods for the year of 2000. Eleven columns show MODIS TERRA top of atmosphere RGB and B7 reflectance, and *MAIAC* products, including cloud mask, RGB NBRF, RGB IBRF, spectral regression coefficient, retrieved AOT and aerosol model (fractional ratio), the Blue band NBRF, and the B7 IBRF. Because B7 LSRT retrievals play critical role in the total *MAIAC* processing, the B7 IBRF is shown next to the measured B7 reflectance for better assessment of the retrieval quality. The first period shows the beginning of processing starting from day 1. Because the surface is dark, the cloud mask initializes at the very beginning, whereas LSRT retrievals start with accumulation of 4 clear days. The SRC retrievals begin on the 4th day after initialization of B7 BRF, or on 11th day from start. Until that time, the aerosol retrievals are performed using constant value of SRC=0.3 in the Lambertian approximation. The SRC coefficients are fully initialized the week after the first period (days 72-78, not shown). The second and third sets show the middle and the end of the green-up period. During green-up, B7 reflectance reduces substantially followed by SRC increase. It remains high throughout the summer and begins to decrease in the second half of autumn (days 301-314).

The retrieved surface reflectance does not show artifacts from clouds or incorrectly derived aerosol.

8.2 Seasonal Surface Change and Spectral Regression Coefficient

The spectral relation between the SWIR and visible reflectance depends on the surface type. Laboratory measured spectra from ASTER (http://speclib.jpl.nasa.gov) and USGS (*Clark et al.*, 2003) Spectral Libraries show that the ratio of the reflectance in the visible spectrum to reflectance in the SWIR is a variable function for different types of soil and minerals. The range of ratio from these measurements is 0.05-3 for the blue, and 0.15-5 for the red band. Over the vegetated regions of the world, SRC changes with vegetation type, cover, and phenology, and usually has a strong seasonal cycle. Agricultural regions have a very large seasonal variability from crop growth to ripening to harvesting (soil exposure).

Two more examples are used here to illustrate dynamic nature of the surface spectral relationships (SRC), and adaptability of *MAIAC*. Figure 18 shows the area of Mongu, Zambia. The Mongu village can be seen as a bright spot in the middle of the subset. A lower area to the east is a floodplain, which is flooded by the waters of Zambezi river each year following the rain season. Days 116-131 show maximum of flooded area, identified by *MAIAC* cloud mask. Because of standing water on the left part of image, reflectance in band 7 is very low and SRC is high. Due to availability of water, grasses grow very fast and the area left of the flooded zone is very green. Vegetation reaches its maximal strength following water retreat (days 140-150). The surface begins to dry up by the end of the second and through the third periods, which correlates with the increase of reflectance in the Blue and SWIR bands and reduction of SRC. The last set of images (days 195-210) shows a stable dry period with high B3, B7 reflectance and low SRC.
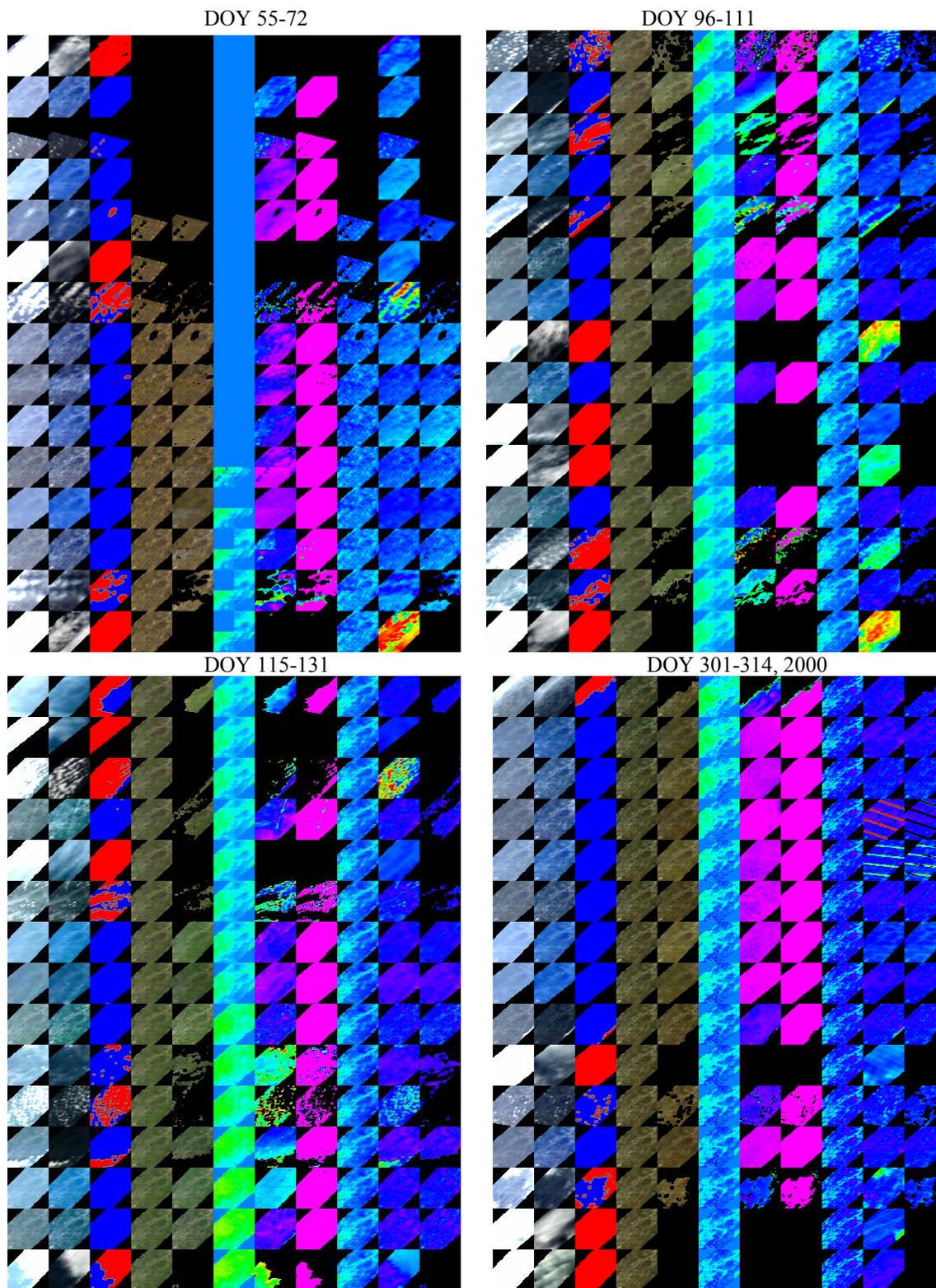
Figure 17. Illustration of *MAIAC* performance for 4 different 15-day periods of 2000 for GSFC. Columns have the following correspondence (scale): 1,2 - MODIS TERRA TOA RGB, 3 - CM, 4 – RGB NBRF, 5 – RGB IBRF, 6 – SRC (0-1), 7 – AOT (0-1), 8 – aerosol model (fractional ratio), 9– Blue band NBRF (0-0.1), 10 - TOA R7, 11 – B7 IBRF (0-0.3). The palette for monochromatic bands is the same as in Fig. 7.

DOY 116-131

DOY 148-163

DOY 164-178

DOY 195-210, 2003
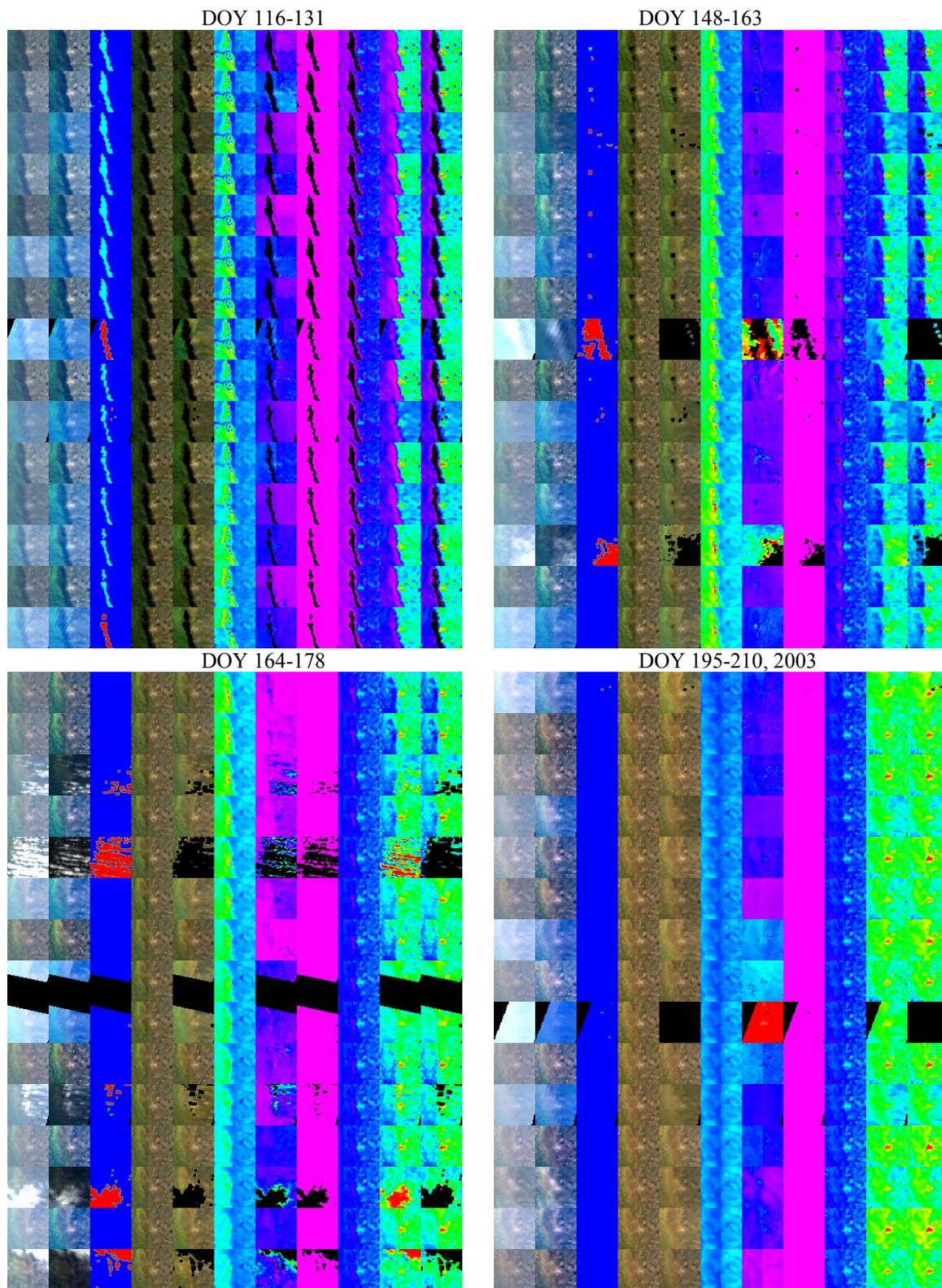
Figure 18. Seasonal trend of surface reflectance and SRC for 2003 over Mongu. Columns and scales are the same as in Figure 17.
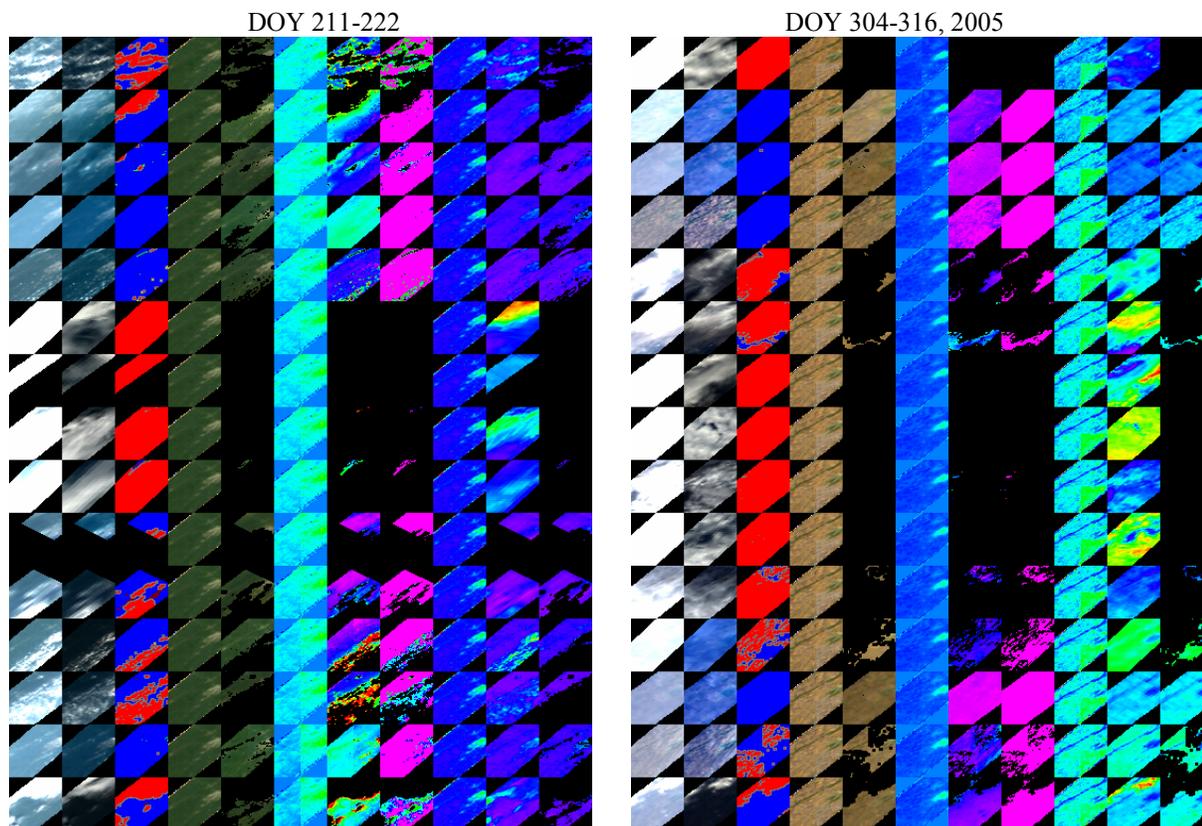
Figure 19. Seasonal trend of surface reflectance and SRC for 2005 over Bondville. Columns and scales are the same as in Figure 17.

Figure 19 shows the agricultural area of Bondville for the middle of summer with high greenness and end of autumn where bright soil is exposed. One can see a significant change in surface reflectance and spectral regression coefficient.

Finally, Figure 20 shows the difference in response of NBRF and IBRF to a rapid surface change caused by a fire. The two-day fire (days 243-244) is shown by an oval shape. The disturbance (burnt area) is clearly visible in the IBRF image the next day. It takes 4 days for the LSRT coefficients (or NBRF) to become adjusted to the surface change. This example shows that the IBRF product rather than NBRF should be used to detect rapid surface changes and disturbances.

8.3 Bright Surfaces

Figure 21 shows the time series of retrieved AOT (black line) for the bright desert site of Solar Village, Saudi Arabia. Overall, *MAIAC* retrievals agree well with AERONET (red line). Due to low cloudiness, the SRC and surface BRF retrievals for this site are remarkably stable with very low noise, which can be generalized to other bright surfaces. However, the scatterplot shows a high scatter caused by the high-frequency noise in our retrievals. A detailed analysis showed that this noise correlates with MODIS TERRA viewing geometry (forward *vs* back-scattering). *MAIAC* slightly underestimates AOT for the forward scattering directions and overestimates AOT for the backscattering observations. Adding dust non-sphericity to our look-up tables did not reduce this effect. Further investigation showed that the "culprit" is the difference in the BRF shape between B7 and the Blue band. Over brighter surfaces, reflectance in the SWIR band
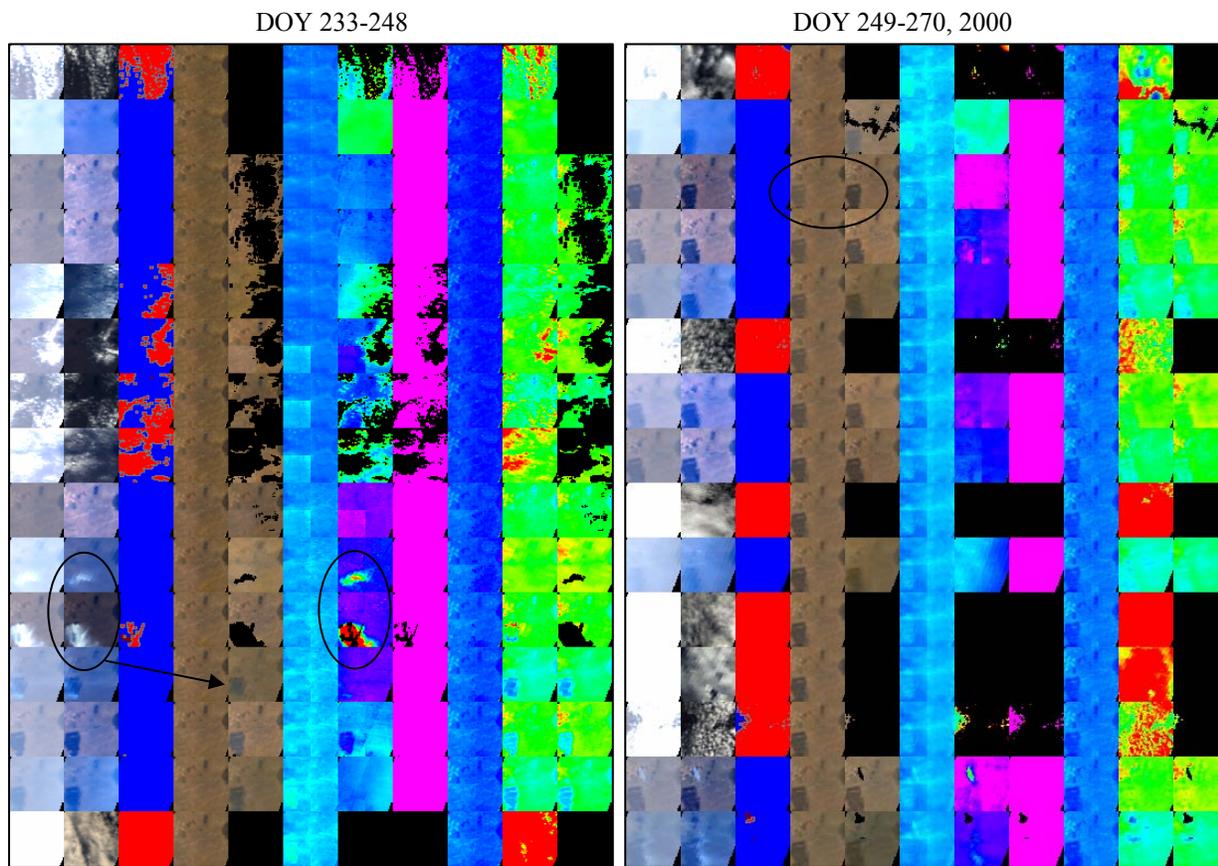
Figure 20. Illustration of difference in the response time of IBRF and NBRF to the rapid surface change caused by a fire in the area of Skukuza, Africa. Columns and scales are the same as in Figure 17.

may be 3-5 time higher than the blue band reflectance, for example 0.45-0.6 vs 0.1-0.15 for the Solar village. Because of low absorption (high reflectance), photons in the SWIR band may scatter many times on the microscopic surface roughness thus flattening the BRF shape. Therefore, although the single scattering BRF shape would be the same or very similar between SWIR and Blue bands, the difference in the multiple scattering would cause the angular dependence of the SRC. As an example, Figure 22 shows the angular dependence of SRC for the blue and red channels for the azimuthal plane 45-135°. This dependence was computed using LSRT coefficients from the AERONET-based Surface Reflectance Validation Network (ASRVN) dataset [*Wang et al.*, 2008]. The ASRVN is an operational processing system which performs atmospheric correction of MODIS data for the AERONET sites using AERONET aerosol and water vapor data. The surface retrieval algorithm is similar to the one used in *MAIAC*. Thus, ASRVN data render an accurate representation of the surface bidirectional reflectance in MODIS channels. Figure 22 shows that SRC depends on the viewing geometry for both bright and dark vegetated surfaces. However, this error has little effect over dark vegetation because of low magnitude of reflectance, whereas it becomes prominent over bright surfaces.

Because the Queue combines measurements from all directions, the retrieved SRC represents the average value typical of nadir view VZA=0°. Our algorithm underestimates SRC and surface BRF for the backscattering directions, which results in higher AOT values. This effect shows as

62
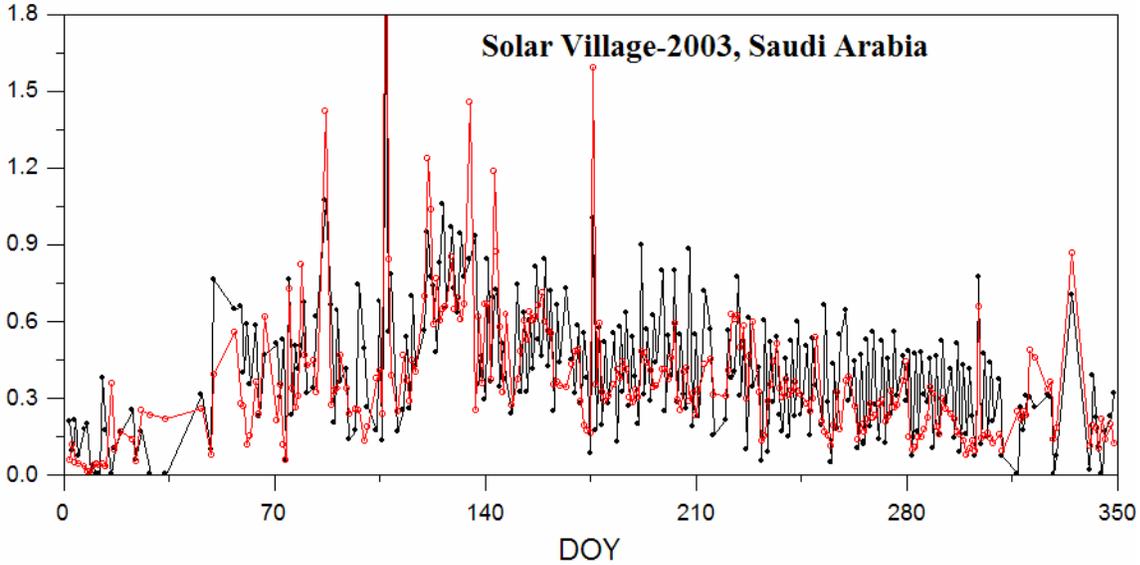
Figure 21. Time series of *MAIAC* (black) and AERONET (red) AOT for Solar Village, 2003.
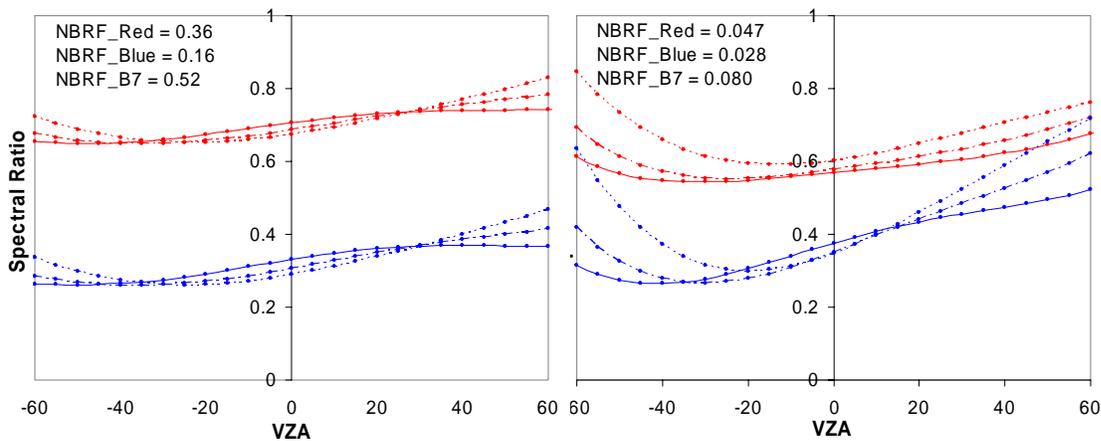
Figure 22. Spectral Ratio for the blue (BRF_{B3}/BRF_{B7}) and red (BRF_{B1}/BRF_{B7}) bands for the desert pixel of Solar Village (left) and vegetated pixel of Konza_EDC (right). The bidirectional reflectance was computed using LSRT parameters of ASRVN database [*Wang et al*, 2008] for the azimuthal plane 45-135°. Positive (negative) values of the view zenith angle represent back- (forward) scattering directions. The solid, dashed and dotted lines correspond to solar zenith angle of 30, 45 and 60°.

high scattering on all *MAIAC*-AERONET scatterplots over bright surfaces (see next section), for example over Mexico City and Sao Paulo. This effect appears in such locations as Bondville and Maricopa in spring and autumn when foliage sheds and reflectance of bright soil becomes dominant.

We believe that our current study will resolve the described problem. The idea is to find an angular correction function for SRC. We may assume that the shapes of BRF in the single scattering are the same in the Blue and SWIR bands, whereas the single scattering albedo is significantly lower in the blue band. The surface albedo in these bands, which should be unbiased due to equal angular coverage between the forward and the backscattering directions, is an integrated function of BRF and can be used to evaluate single scattering albedo. Once this is

done, the multiple scattering component of the SWIR BRF can be corrected, giving more accurate BRF shape for SRC retrievals.

8.4 NDVI Time Series

Figure 23 shows the NDVI time series using *MAIAC* NBRF and IBRF for locations of Bondville, GSFC (USA), Mongu (Zambia) and Cuiaba-Miranda (Brazil). The NDVI plots are built for individual pixels without averaging or pre-processing. For each location, we selected a bright (soil-dominated) and a green (vegetated) pixel. The NDVI time series has a very low noise and a highly reproducible interannual pattern. The IBRF NDVI is not normalized in the view geometry, and this causes scatter for the bright soil pixels. The periods where the NBRF NDVI does not change correspond to either high cloudiness or missing MODIS data when the 16-day Queue did not have enough clear-sky observations to make a new retrieval. These plots also show that the NBRF NDVI has a delayed response to the surface seasonal change, whereas IBRF NDVI tracks seasonal variations in a timely manner. The average delay of the NBRF NDVI, as can be seen from the stretched image c) is 4-7 days depending on cloudiness. This set of images shows different climate zones and vegetation types. This difference manifests itself in variable seasonal pattern and different magnitude of seasonal NDVI change.

8.5 AERONET Validation

Figure 24 shows scatterplots of *MAIAC* AOT (0.465 μm) *vs* AERONET AOT at 0.44 μm based on 8+ years of MODIS TERRA data (2000 – august 2008). A long dataset provides a good statistics for each plot accumulating from ~300 to ~900 points. Due to the wavelength difference, AERONET AOT is expected to be a little higher. For example, the difference for the GSFC site based on simulations should be about 5-7%. Following MODIS validation strategy [*Remer et al.*, 2005], AERONET v2.0 data were averaged over ±30 min interval of TERRA satellite overpass. *MAIAC* retrievals are averaged over 10 km area. *MAIAC* value is reported if the number of retrievals exceeds 20. Two pixels around a cloud or a clear-sky snow pixel were discarded. Cases of inhomogeneous aerosols were filtered according to the criterion: $AOT_{av} - AOT_{min} > 0.15$.

The top row scatterplots shows comparisons for relatively dark and vegetated surfaces, Goddard Space Flight Center (MD, USA), Bondville, (IL, USA), Ispra (Italy). A low absorption aerosol model (fine fraction #2, coarse fraction #5) was used for these sites. The middle row shows three megacities with population over 12 million people. To fit AERONET record for Moscow, we had to use a very low absorption model (fine fraction #1, coarse fraction #5). Beijing and Mexico City were processed with moderate-to-high absorption model (fine fraction #3, coarse fraction #6). Finally, the bottom row represents sites of biomass burning in Africa and Brazil. High absorption models (fine fraction #4, coarse fraction #7) typical of African Savannah biomass burning were used for Mongu, and a moderate-to-high absorption model (fine fraction #3, coarse fraction #6) was used for Cuiaba-Miranda and Skukuza. Our use of absorption models agrees with AERONET classification [*Dubovik et al.*, 2002] and with MOD04 aerosol model derived from AERONET data.

In addition to pollution, there are several dust storms per year over Beijing with dust blown from the nearest Gobi desert located to the north of the city. In this comparison, we did not use dust model. On average, *MAIAC* results compare well with the AERONET data for most of sites with high correlation. There is a considerable scattering for Mexico City, which has a bright surface.

a) Bondville, Bright pixel
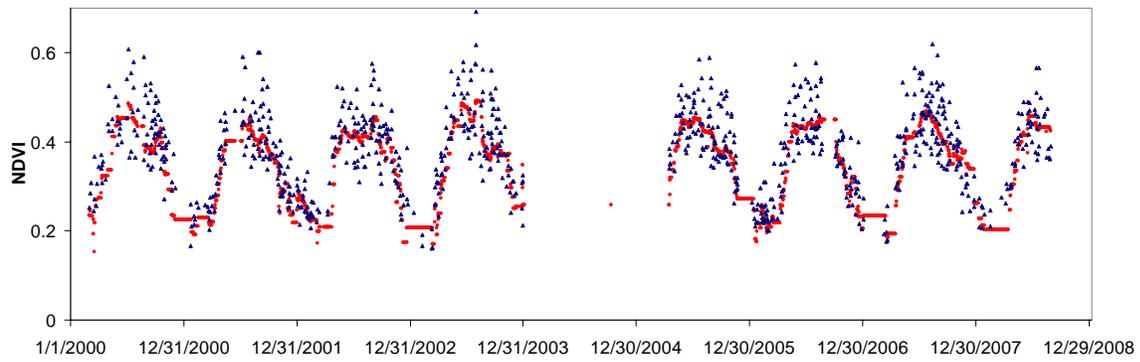
b) Bondville, Green pixel
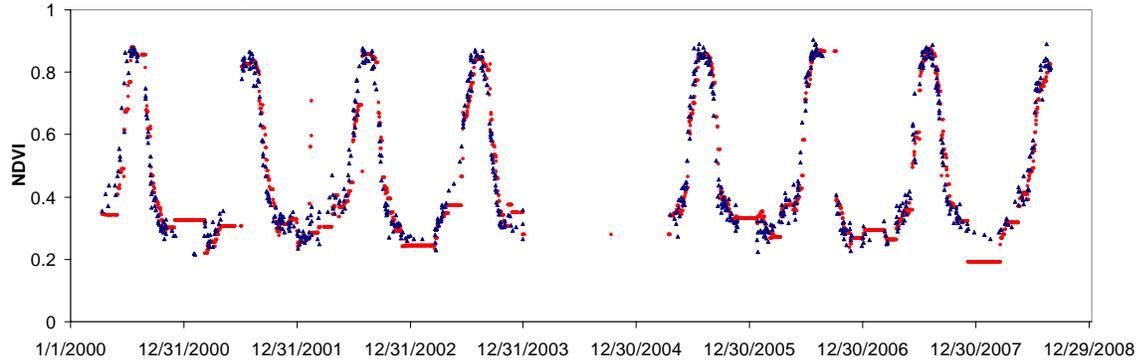
c) Bondville, Green pixel, 2002-2003

d) GSFC, Green pixel

65
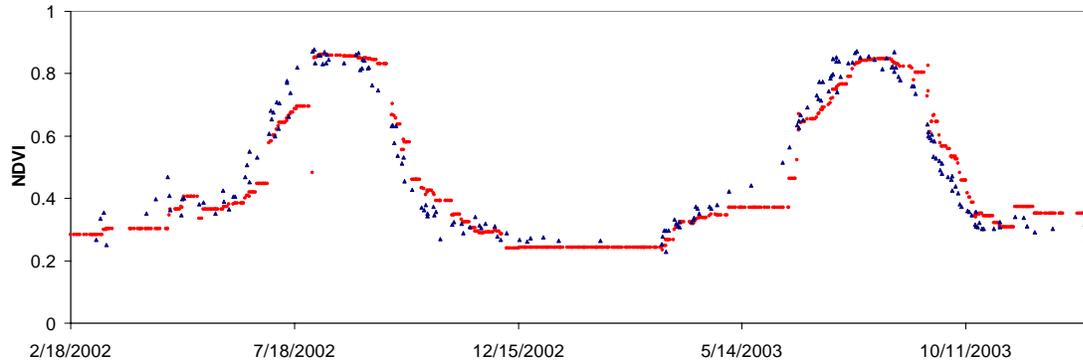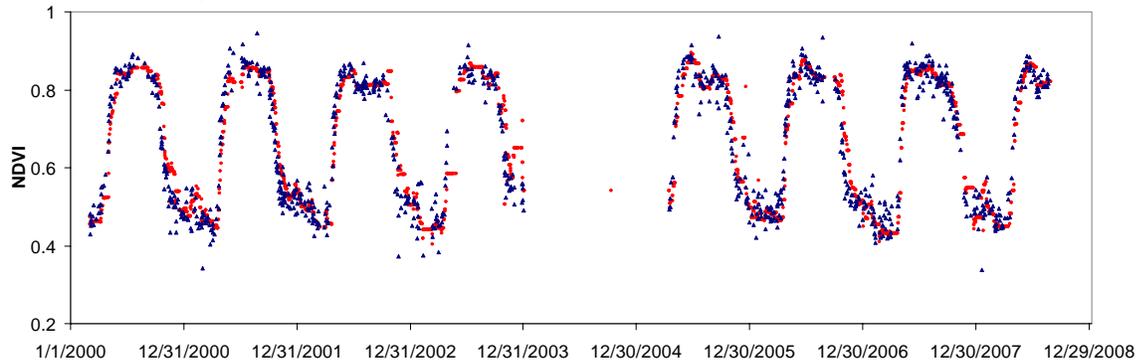
e) Mongu, Bright pixel
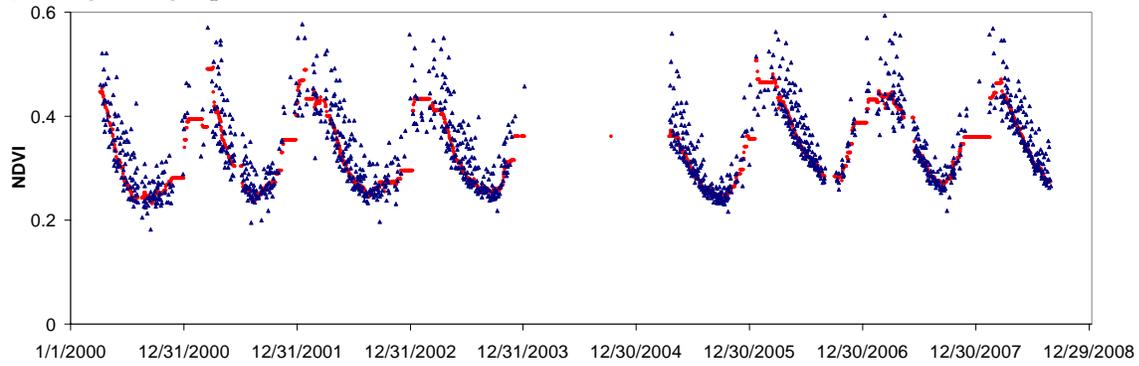
f) Mongu, Green pixel
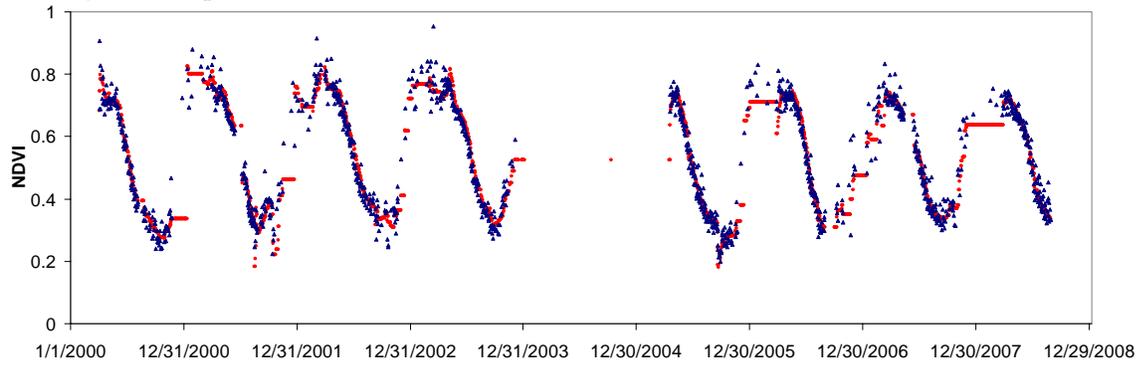
g) Cuiaba Miranda, Bright pixel
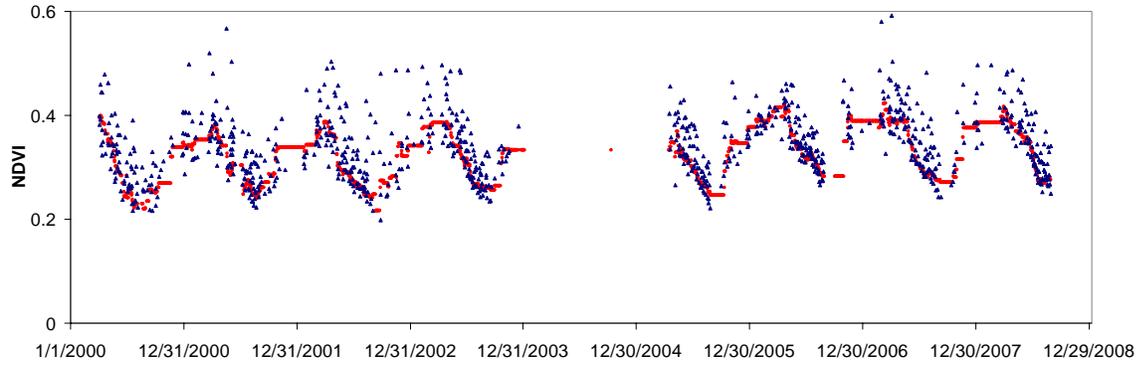
h) Cuiaba Miranda, Green pixel
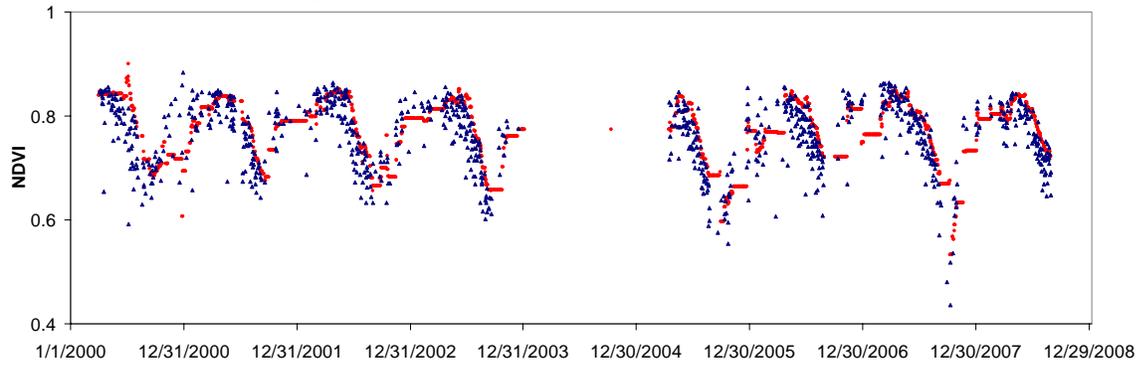
Figure 23. Time series of NBRF NDVI (red) and IBRF NDVI (blue).
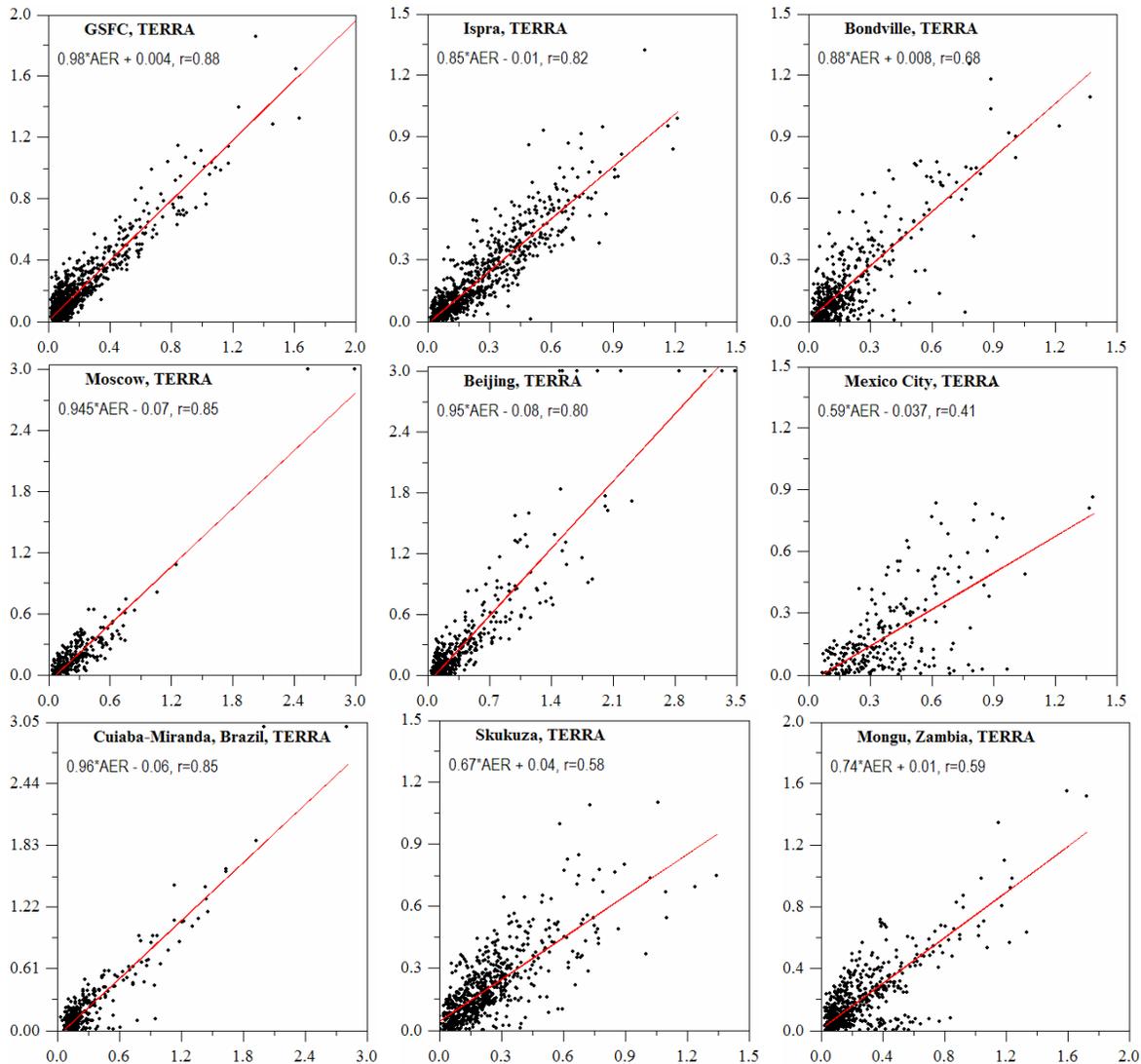
Figure 24. Scatterplots *MAIAC* AOT (0.47 μm) – AERONET AOT (0.44 μm) for MODIS TERRA (2000-august 2008).

This source of systematic uncertainty was explained previously in sec. 8.3. Strong scattering effect with large number of *MAIAC*'s low AOT values skews the slope of regression. A somewhat lower scattering, observed for Bondville, Skukuza and Mongu, has the same origin. The surface becomes rather bright in the dry season over Mongu and Skukuza, and in the late fall-early spring time period over Bondville.

8.6 Examples of Large-Scale Processing

We have evaluated *MAIAC* performance over the different world regions for an extended period of time. Typically, we order MODIS data for large areas of several thousand square kilometers for at least one year, and process the full set of data. Two examples of the large-scale AOT retrievals from MODIS TERRA are shown in Figures 25-26. Figure 25 shows smoke from biomass-burning during the dry season over an area of 1200×1200 km^2 in Zambia, Africa. The

TOA image for the day 205 shows dozens of small-to-large fires. The fine 1 km resolution allows *MAIAC* to resolve and trace plumes of the individual fires. The fire plumes disappear at the coarse 10 km resolution of operational MODIS aerosol product MOD04 shown on the inset. The comparison shows that the magnitude of MOD04 and *MAIAC* retrieved AOT and its spatial distribution is rather similar, although there are certain differences depending on the surface type and geometry of observations. This particular example shows that through significantly higher spatial resolution, *MAIAC* offers quantitatively new information about aerosols and their sources unavailable before. The gradient of AOT at 1 km resolution is high enough to implement an automatic delineation algorithm for the smoke plume detection, with the data that could be used in different applications, such as air quality.

Figure 26 shows *MAIAC* aerosol retrievals over a large portion of bright Arabian Peninsula (area 1800×1800 km^2) for day 207 of 2005 with rather complex conditions. On one hand, the dust is transported across the Red Sea from Sudan (Africa). The wind does not penetrate the mountains along peninsular's western shore. It is clear on the top of the mountain ridge, and the dust is concentrated along the shore, as can be seen both from the MODIS RGB image and from the AOT image. On the other hand, a separate internal dust storm has developed in the southern part
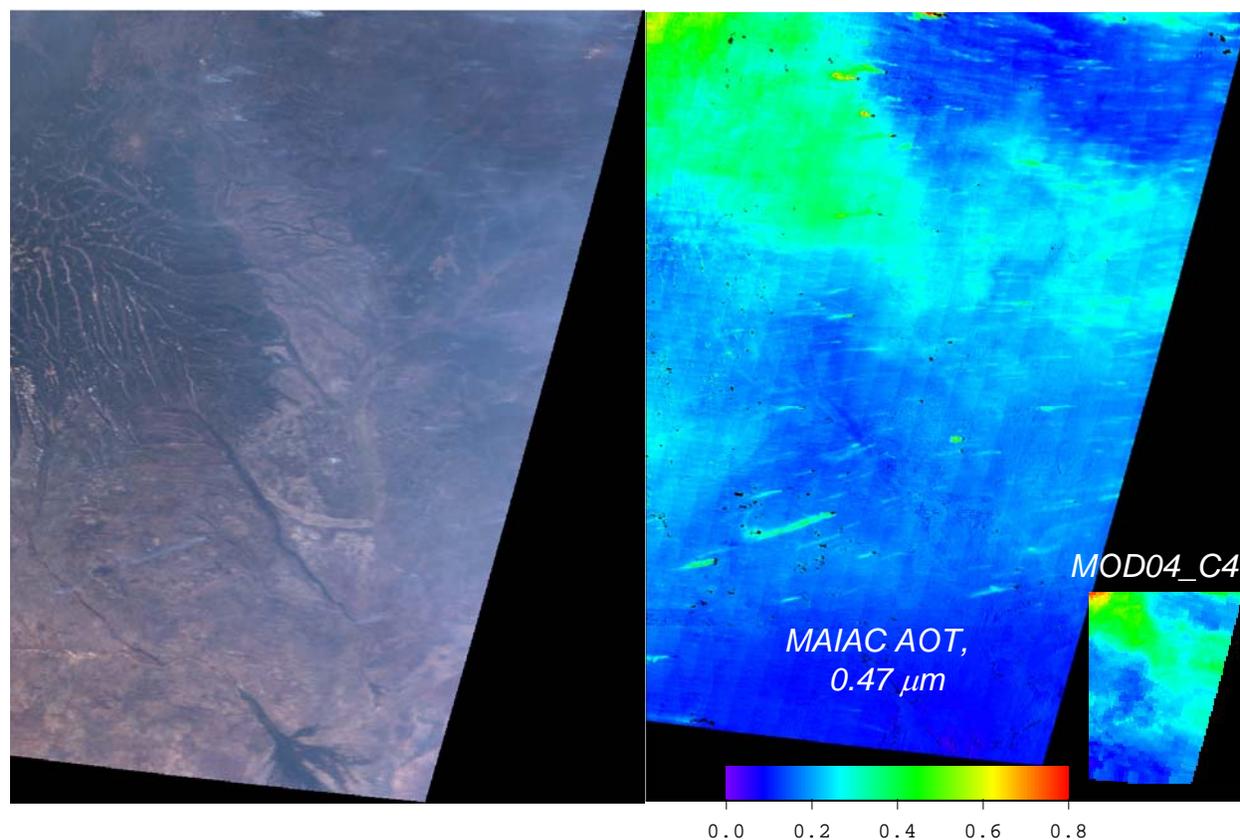
Figure 25. Fires during dry biomass-burning season in Zambia, Africa, for day 205 of 2005 (area 1200×1200 km^2). The 1km gridded MODIS TERRA TOA RGB image is shown on the left and *MAIAC*-retrieved AOT at 0.47 µm is on the right. The AOT scale is the same for MOD04 and *MAIAC*. The high resolution (1 km) of AOT product allows detecting and tracing individual fire plums. The inset shows result of the MODIS dark target algorithm MOD04_C4.
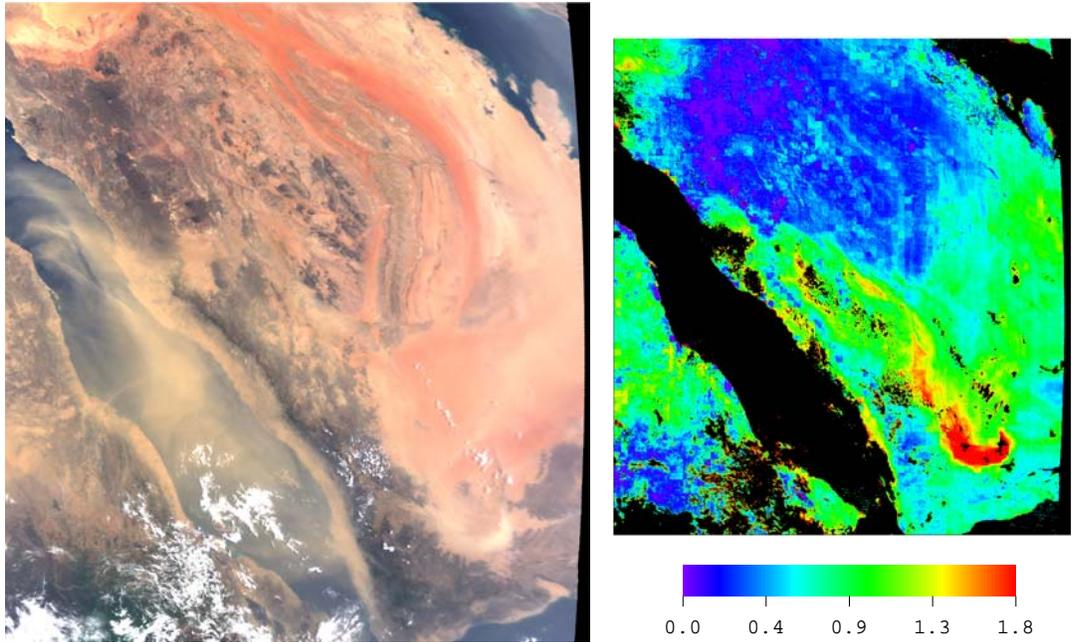
Figure 26. MODIS TERRA RGB TOA image and *MAIAC* AOT at 0.47 μm over Arabian Peninsula (area 1800×1800 km^2) for day 207 of 2005.
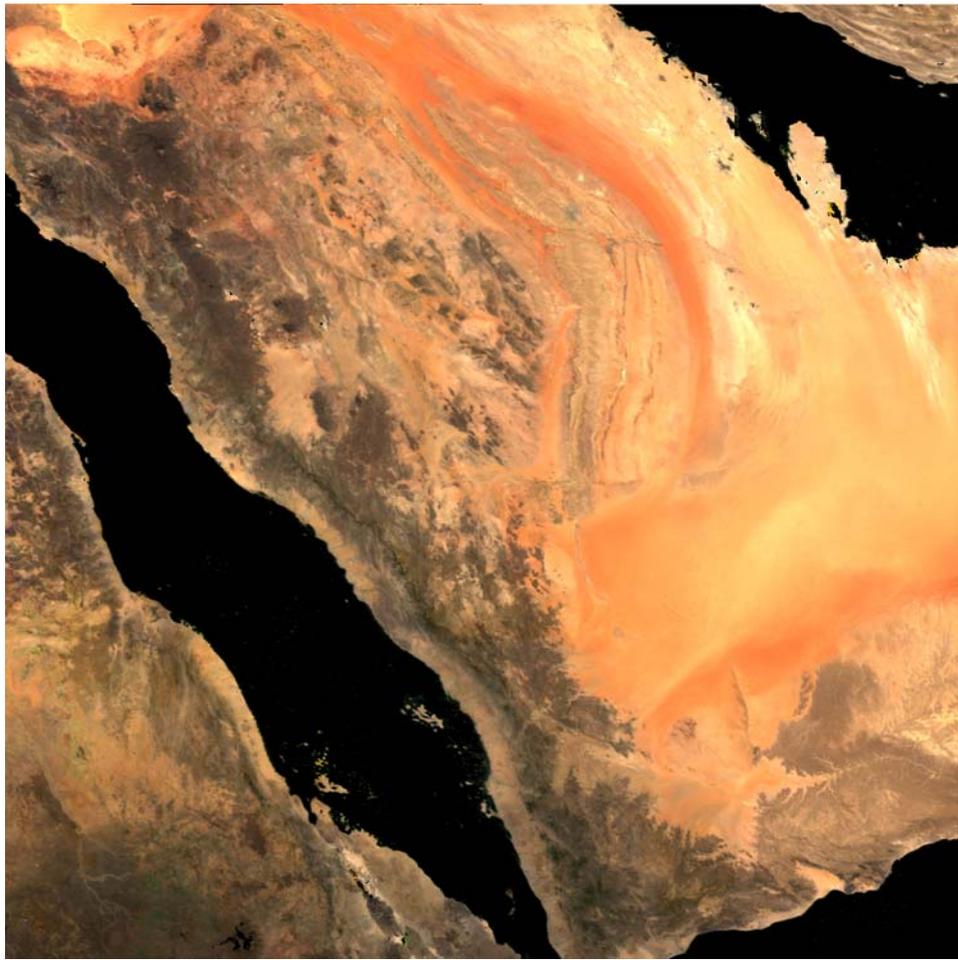
Figure 27. RGB image of surface NBRF (BRF for a fixed geometry, VZA=0°, SZA=45°) for Arabian Peninsula for day 184 of 2005. The image is built with equal weights for RGB bands.

69

DOY 150   201   224, 2007
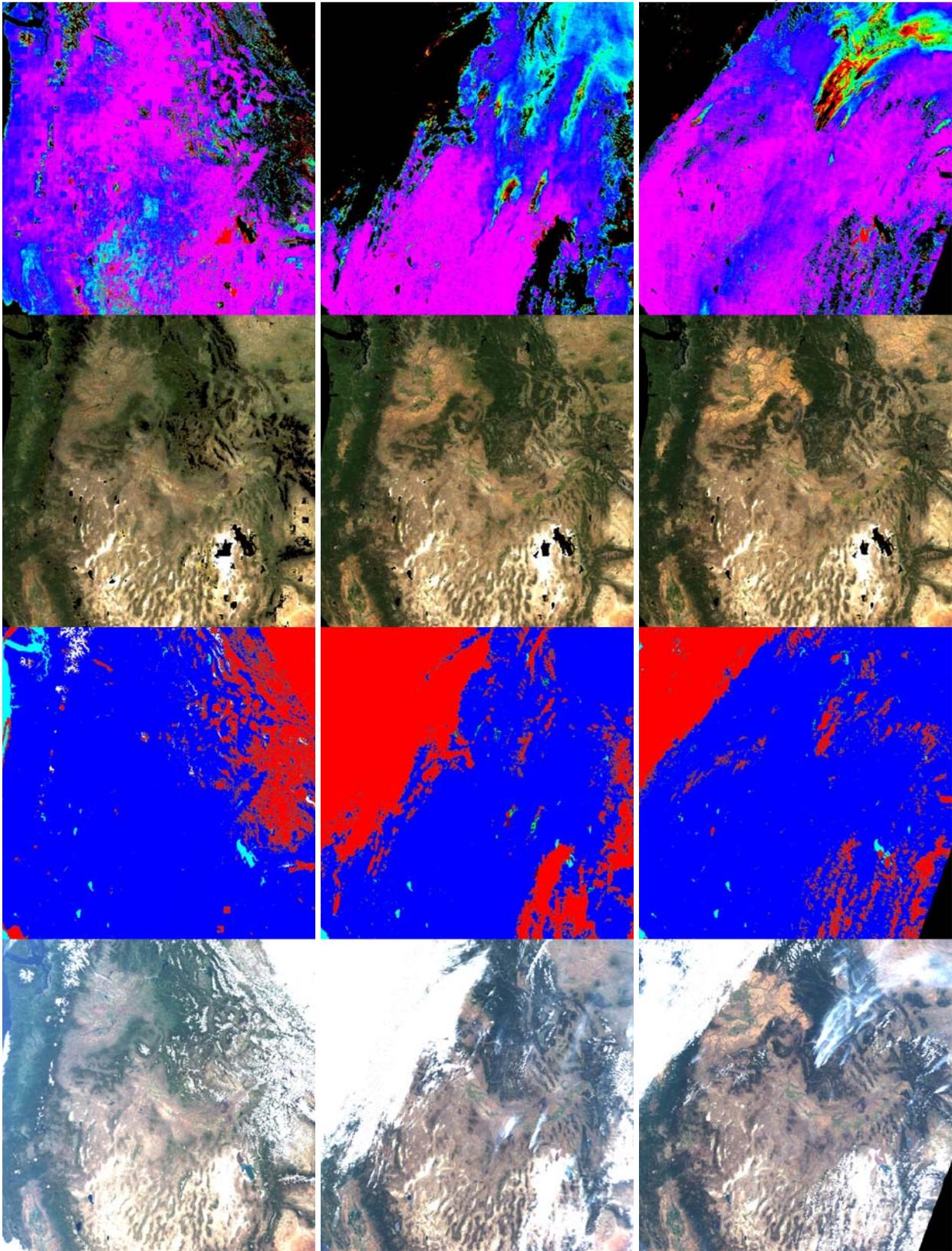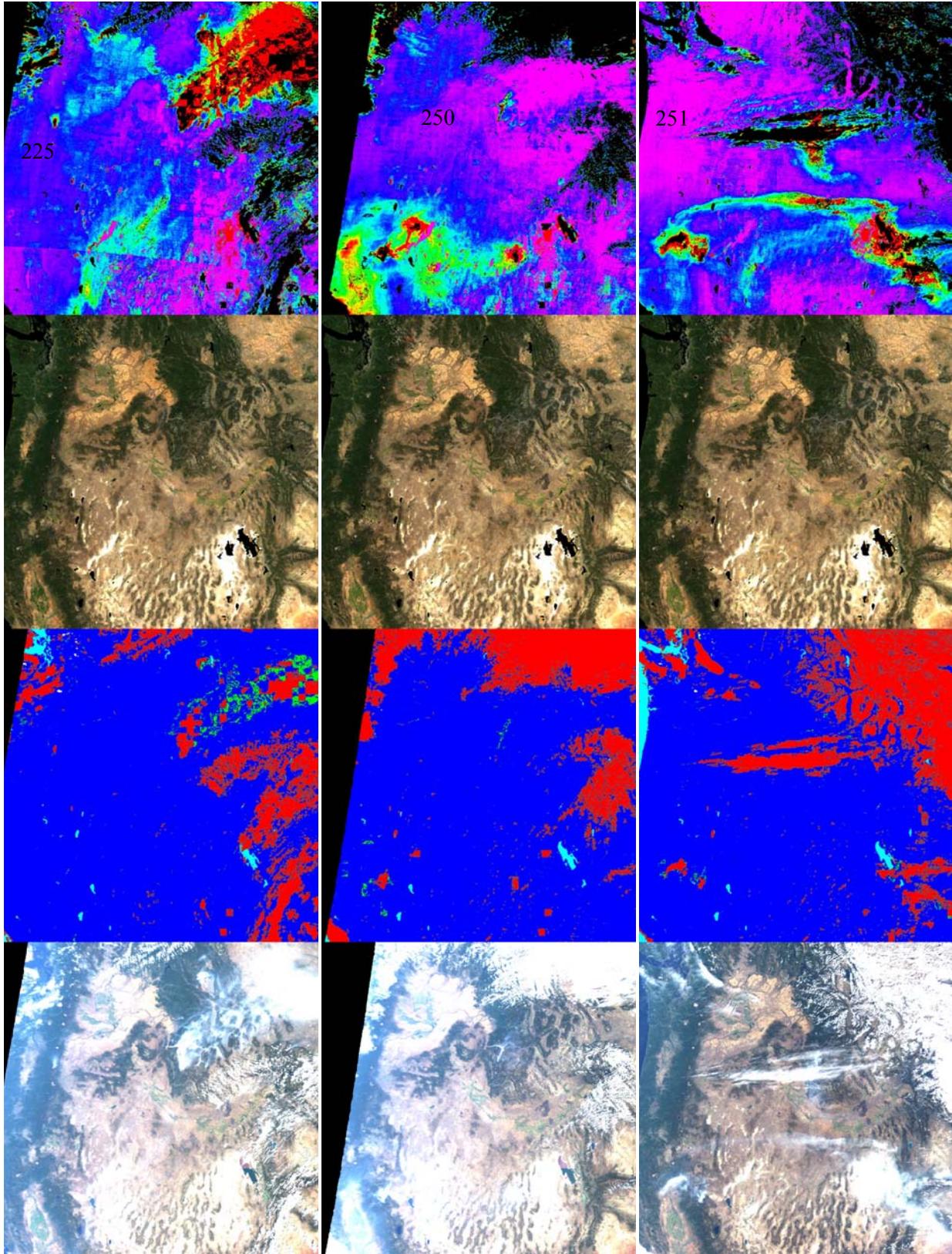
Figure 28. Fires in the North-Western Rocky Mountains in the summer of 2007. The images show AOT (scale 0-1: the red color shows AOT≥1), RGB NBRF, cloud mask and MODIS RGB images.

of peninsula, with winds carrying dust in the north-west direction. For comparison, Figure 27 shows the true color RGB image of the surface NBRF for this area. The bright surface feature, corresponding to the epicenter of the dust storms, is absent on the NBRF image, which confirms that this event is indeed a local dust storm.

A sequence of images with extensive fires is shown for the summer period of 2007 for the North-West Rocky Mountain region. This area was processed by the request of the US Forest Service. One can see glaciers of Mt. StHelen and Rocky Mountains on day 150. Most of surface is green except very bright areas to the south. There are several black spots in the NBRF image which remain un-initialized for the full time. The largest is the Salt Lake. The second largest, in the middle of very bright area on the south-west of Salt Lake City, is salt playa whose spectrum of reflectance is close to that of a snow: it is extremely bright in the visible (r~0.6-0.7) and very dark in the SWIR (r~0.04-0.1). Current *MAIAC* algorithm systematically gives wrong very high AOT values for these spots.
The surface dries out and becomes brighter by day 220, as clearly visible from the NBRF image. The fire have been active, with some intermissions, for about a two month period, and we are showing only several of them. With high MAIAC AOT resolution, one can distinguish different fire epicenters changing from day to day. We find the last image (day 251) particularly interesting with smoke plume in the middle of the image below two large cloud systems.

9. Summary

MAIAC is a new algorithm which uses a time series processing and combines an image- and pixel-level processing. It includes cloud mask and generic aerosol-surface retrieval algorithm. The suite of *MAIAC* products includes column water vapor, cloud mask, dynamic mask of standing water and snow, AOT at 0.47 μm and Angstrom exponent (or the ratio of volumetric concentrations of the coarse and fine fractions), and spectral surface reflectance metrics, which include LSRT coefficients, albedo, NBRF and IBRF. The suite of products is generated in a systematic and mutually consistent way to observe the energy conservation principle. In other words, the radiative transfer calculation with the given set of parameters closely corresponds to measurements. All products are produced in gridded format at the resolution of 1 km.

A high spatial resolution of *MAIAC* (1 km vs 10 km for operational MODIS aerosol product) allows a new type of analysis and applications. One demonstrated example is a possibility of detection and tracing fire plumes from biomass burning. A high resolution of 1 km makes this application possible, whereas most of the information disappears at coarse 10 km resolution. MAIAC provides high quality surface reflectance parameters. We demonstrated a high quality and reliability of NBRF, IBRF and NDVI at 1 km resolution.

We have started validation of different *MAIAC* products, and this work will be continued together with cross-comparison with current operational MODIS products.

This ATBD does not describe two additional *MAIAC* products, sub-pixel snow fraction and optical snow grain size. Use of the time series and knowledge of the surface spectral BRF prior to the onset of winter allows us to implement snow retrievals in the most simple and straightforward manner using linear unmixing approach. This algorithm is not yet validated.

The current performance of the algorithm is not yet fully optimized. Nevertheless, *MAIAC* is already sufficiently fast for operational processing: it takes ≈50 sec. of one single-core AMD

Opteron-64 processor to process one Tile (600×600 km^2) of MODIS data. The operational testing of *MAIAC* is planned to begin in the end of 2008 in collaboration with the University of Wisconsin and GSFC-based MODIS land processing team.

10. Remaining Tasks for 2009

Cloud Mask:

- Improve CM performance on the land-water boundaries.
- Perform global testing of algorithm, and extensive inter-comparison with MODIS CM product (MOD35) in collaboration with University of Wisconsin, Madison (P. Menzel, S. Ackerman, R. Frey, S. Dutcher et al.).

Aerosol Retrievals:

- Development of global geographically prescribed aerosol climatology model. The following work is underway: we are studying the AERONET-based classification [*Holben et al.*, 2001] and MODIS Collection 5 aerosol model [*Levy*, 2007] and plan to investigate MISR level 3 aerosol product, which provides an independent global aerosol climatology over land.

- The current version of *MAIAC* uses MODIS blue band (B3) as a reference channel for aerosol retrievals. The AERONET validation analysis shows that *MAIAC* may underestimate AOT over bright surfaces by 0.05-0.1. We plan to investigate the use of MODIS deep blue channel (B8) as a reference band as a possible solution to this problem.

- Currently, we use MODIS red (B1) and SWIR (B7) bands to evaluate coarse-to-fine mode fractional ratio (aerosol model). This band selection works well, based on comparison with AERONET, over dark surfaces. However, it often fails over bright surfaces, selecting the highest weight of coarse fraction and accordingly, very high AOT. We plan to analyze the use of MODIS deep blue channel (B8) instead of (or in addition to) the SWIR band. This channel is expected to improve aerosol model selection because surface at 0.412 μm is significantly darker, and uncertainties associated with surface reflectance will have little impact.

- Develop a correction method to compensate the angular bias in the spectral regression relationship over bright surfaces (see sec. 8.3).

Atmospheric Correction:

The climate change research, and global carbon analysis as its part, relies on accurate estimates of carbon balance of planet Earth. These estimates critically depend on accurate tracking of surface change, for example the date of onset and magnitude of spring green-up at northern latitudes. The early period of vegetation photosynthesis is particularly important because soil is not warmed up, and soil respiration, usually balancing carbon uptake by photosynthesis, is very

low. As we described above, even the 4-day LSRT solution may be delayed if surface changes rapidly and the cloudiness is high.

This problem can be solved if for every pixel we kept two sets of LSRT coefficients from the previous year (or season) in the memory. These solutions will represent the maximal and minimal greenness of the surface, which can be tracked in the time series using NDVI and magnitude of NBRF in several spectral bands. In this case, once the surface change has been detected from measurements, the LSRT will be modeled as a linear combination of the two solutions, exactly fitting any given day of the transition period with no delay. In fact, this is the only possible way to maintain the best accuracy of both aerosol retrievals and atmospheric correction. When the surface reflectance changes both in magnitude and shape of BRF during the accumulation period, then the Queue LSRT retrievals, strictly speaking, are not valid. This is a straightforward approach which only requires an additional memory for several spectral bands. We will implement this method in 2009.

REFERENCES

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

Abdou, W., J. V. Martonchik, R. A. Kahn, R. A. West, and D. J. Diner, 1997: A modified linear-mixing method for calculating atmospheric path radiances of aerosol mixtures. *J. Geophys. Res*., **102**, 16,883-16,888.

Anderson, T. L., R. J. Charlson, D. M. Winker, J. A. Ogren, K. Holmen, 2003: Mesoscale Variations of Tropospheric Aerosols, *J. Atm. Sci.*, **60**, 119-136.

Ben-Dor, E., 1995: A precaution regarding cirrus cloud detection from airborne imaging spectrometer data using the 1.38 μm water vapor band. *Rem. Sens. Environ*., **50**, 356-350.

Bennartz, R., J. Fischer, 2001: Retrieval of columnar water vapor over land from backscattered solar radiation using the Medium Resolution Imaging Spectrometer, *Rem. Sens. Environ*., **78**, 275-283.

Clark, R. N., G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley (2003). USGS Digital Spectral Library splib05a, USGS Open File Report 03-395.

Diner, D. J., W. Abdou, T. Ackerman et al., 2001: MISR level 2 aerosol retrieval algorithm theoretical basis, JPL D-11500, Rev. E, 105 pp.

Diner, D. J., J. Martonchik et al., 1999: MISR level 2 surface retrieval algorithm theoretical basis. *JPL D-11501*, Rev. D, 105 pp.

Diner, D. J., J. V. Martonchik, R. A. Kahn, B. Pinty, N. Gobron, D. L. Nelson, B. N. Holben, 2005: Using angular and spectral shape similarity constraints to improve MISR aerosol and surface retrievals over land, *Rem. Sens. Environ.*, **95**, 155-171.

Dubovik, O., B. Holben, T. F. Eck., A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanre, and I. Slutzker, 2002: Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci*., **59**, 590-608.

Gao, B. C., and Y. J. Kaufman, 2003: Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels. *J. Geophys. Res*., **108** (D13), 5389-5399, doi:10.1029/2002JD003023.

Hale, G. M., and M. R. Querry, 1973: Optical constants of water in the 200-nm to 200 μm wavelength region. *Appl. Optics*, **12**, 555-563.

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.

Holben, B., et al., 2001: An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *J. Geophys. Res*., **106**, 9807-9826.

Hsu, N.C, Tsay S. C., King, M. D., et al., 2005: Aerosol properties over bright-reflecting source regions. *IEEE Trans. Geosci. Remote Sensing*, **52**, 557-569.

Kahn, R. A., B. J. Gaitley, J. V. Martonchik, D. J. Diner, K. A. Crean, B. Holben, 2005: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on two years of coincident Aerosol Robotic Network (AERONET) observations, *J. Geophys. Res*., **110** (D10S05), doi: 10.1029/2005JD005706.

Kneizys, F. X., L. W. Abreu, G. P. Anderson, J. H. Chetwynd, E. P. Shettle, A. Berk, L. S. Bernstein, D. C. Robertson, P. Acharya, L. S. Rothman, J. E. A. Selby, W. O. Gallery, S. A. Clough, 1996: The MODTRAN 2/3 Report and LOWTRAN 7 Model. MODTRAN Report: Ontar Corporation, North Andover, MA, 261 pp.

Kurucz, R. L., 1997: The solar irradiance by computation. Available at http://cfaku5.cfa.harvard.edu/papers/irradiance/.

Lucht, W., Schaaf C. B., Strahler A. H., 2000: An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Trans. Geosci. Remote Sens.*, **38**, 977-998.

Lyapustin, A., and Yu. Knyazikhin, 2001: Green's function method in the radiative transfer problem. I: Homogeneous non-Lambertian surface, *Appl. Optics*, **50**, 3595-3501.

Lyapustin, A., 2003: Interpolation and Profile Correction (IPC) method for shortwave radiative transfer in spectral intervals of gaseous absorption. *J. Atmos. Sci.* **60**, 865-871.

Lyapustin A., D. Williams, B. Markham, J. Irons, B. Holben, and Y. Wang, 2005: A Method for Unbiased High Resolution Aerosol Retrieval from Landsat, *J. Atmos. Sci.*, **61**, 1233-1255.

Lyapustin, A., 2005: Radiative transfer code SHARM for atmospheric and terrestrial applications. *Appl. Optics*, **55**, 7765-7772.

Lyapustin, A., and Wang, Y., 2005: Parameterized Code Sharm-3D for Radiative Transfer Over Inhomogeneous Surfaces, *Appl. Optics*, **55**, 7602-7610.

Lyapustin, A., Y. Wang, J. Martonchik, J. Privette, B. Holben, I. Slutsker, A. Sinyuk, and A. Smirnov, 2006: Local Analysis of MISR Surface BRF and Albedo over GSFC and Mongu AERONET Sites. *IEEE Trans. Geosci. Remote Sens.*, **55**, 1707-1718.

Lyapustin, A., Y. Wang, R. Kahn, J. Xiong, A. Ignatov, R. Wolfe, A. Wu, B. Holben, C. Bruegge, 2007: Analysis of MODIS-MISR calibration differences using surface albedo around AERONET sites and cloud reflectance. *Rem. Sens. Environment*, MISR Special Issue (in press).

Lyapustin, A., Y. Wang, et al., Surface Pressure and Water Vapor Correction for Operational Aerosol Retrieval and Atmospheric Correction Algorithms, in preparation.

Martonchik, J. V., D. J. Diner, K. A. Crean, et al. 2002: Regional aerosol retrieval results from MISR. *IEEE Trans. Geosci. Remote Sensing*, **50** (7): 1520-1531.

Martonchik J. V., D. J. Diner, R. A. Kahn, et al. 1998: Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging. *IEEE Trans. Geosci. Remote Sens.*, **36**, 1212-1227.

McClain, E. P., 1993: Evaluation of CLAVR phase-I algorithm performance. *Final Report* 201-525.

Michalsky J., J., J. C. Liljegren, L. C. Harrison, 1995: A Comparison of Sun photometer derivations of total column water vapor and ozone to standard measures of same at the Southern Great Planes Atmospheric Radiation Measurement site. *J. Geophys. Res.*, **100** (D12), 25,995-26,003.

Middleton E. M., Drolet G., Huemmrich K. F., Lyapustin A., et al., 2005: Direct satellite inference of ecosystem light use efficiency for carbon exchange using MODIS on TERRA and AQUA. *IGARSS Proc.*, **1-7**, 2038-2051.

Mlawer, E. J., D. C. Tobin, and S. A. Clough, 2006: A Revised Perspective on the Water Vapor Continuum: The MT_CKD Model, in preparation.

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.

Remer, L., Y. Kaufman, D. Tanre et al., 2005: The MODIS Aerosol Algorithm, Products, and Validation, *J. Atm. Sci.*, **62**, 957-973.

Remer, L., D. Tanre, Y. Kaufman, R. Levy, S. Mattoo (2006). Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS: Collection 5. *MODIS ATBD*, available at http://modis-atmos.gsfc.nasa.gov/ MOD05_L2/atbd.html.

Rossow, W. B., and L. C. Garder, 1993: Cloud detection using satellite measurements of infrared and visible radiances for ISCCP. *J. Climate*, **6**, 2351-2369.

Rothman, L. S., Barbe, A., Benner, D. C., Brown, L. R., Camy-Peyret, C., Carleer, M. R., Chance, K. V., Clerbaux, C., Dana, V., Devi, V. M., Fayt, A., Flaud, J.-M., Gamache, R. R., Goldman, A., Jacquemart, D., Jucks, K. W., Lafferty, W. J., Mandin, J.-Y., Massie, S. T., Nemtchinov, V., Newnham, D. A., Perrin, A., Rinsland, C. P., Schroeder, J., Smith, K. M., Smith, M. A. H., Tang, K., Toth, R. A., Vander, Auwera J., Varanasi, P., Yoshino, K. (2003). The HITRAN Molecular Spectroscopic Database: Edition of 2000 Including Updates through 2001. *J. Quant. Spectrosc. Radiat. Transfer*, **82**, 5-55.

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**, 135-158.

Smirnov, A, B. N. Holben, A. Lyapustin, I. Slutsker, and T. F. Eck, AERONET processing algorithms refinement, AERONET Workshop, May 10 - 15, 2005, El Arenosillo, Spain.

Stamnes, K., W. Li, H. Eide, T. Aoki, M. Hori, R. Storvold, 2007: ADEOS-II/GLI snow/ice products – Part I: Scientific Basis. *Remote Sens. Environ.,* **111**, 258–273.

Tanre, D., P. Y. Deschamps, C. Devaux, M. Herman, 1988: Estimation of Saharan aerosol optical thickness from blurring effect in Thematic Mapper data. *J. Geophys. Res.*, **93**, 15,955-15,965.

Veefkind, J. P., G. de Leeuw, P. Durkee, 1998: Retrieval of aerosol optical depth over land using two-angle view satellite radiometry during TARFOX. *Geophys. Res. Lett.*, **25**, 3135-3138.

Vermote, E. F, El Saleous N.Z., C. O. Justice, 2002: Atmospheric Correction of MODIS Data in the Visible to Middle Infrared: First Results. *Rem. Sens. Environ.*, **83**, 97-111.

Wang, M., and H. R. Gordon, 1995: Radiance reflected from the atmosphere-ocean system: Synthesis from the individual components of the aerosol size distribution. *Appl. Optics*, **33**, 7088-7095.