The Time Series Technique for Aerosol Retrievals Over Land from MODIS: Algorithm MAIAC

Alexei Lyapustin and Yujie Wang
University of Maryland Baltimore County, NASA Goddard Space Flight Center, Greenbelt, MD

1. Introduction

Atmospheric aerosols interact with sun light by scattering and absorbing radiation. By changing irradiance of the Earth surface, modifying cloud fractional cover and microphysical properties and a number of other mechanisms, they affect the energy balance, hydrological cycle, and planetary climate [*IPCC*, 2007]. In many world regions there is a growing impact of aerosols on air quality and human health.

The Earth Observing System [*NASA*, 1999] initiated high quality global Earth observations and operational aerosol retrievals 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.

With MODIS aerosol program being very successful, there are still several unresolved issues in the retrieval algorithms. 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 composed from minimal monthly measurements with Rayleigh correction. Both algorithms use 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 chapter 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 invariants 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~25 pixels (25 km). With K days in the processing queue, the number of measurements exceeds the number of unknowns

$$KN^2 > K+(3N)^2 \text{ if } K>3, \tag{1}$$

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:

$$\rho_{ij}^{\lambda}(\mu_0, \mu; \varphi) = b_{ij}^{\lambda} \rho_{ij}^{B7}(\mu_0, \mu; \varphi). \tag{2}$$

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 (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]. 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. The algorithm combines the block-level and the pixel-level processing, and produces the full set of parameters at 1 km resolution.

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 chapter provides an overview of *MAIAC* processing. Section 3 presents the radiative transfer basis for the aerosol retrievals and atmospheric correction algorithm, which are described in sections 4-5, respectively. Section 6 presents *MAIAC* cloud mask algorithm. Finally, section 7 presents examples of *MAIAC* performance and results of AERONET validation. The chapter 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 [*Lyapustin and Wang*, 2007]. 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.

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 boundary condition known, the Blue band AOT in this step is retrieved at 1 km resolution.

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 1-7 and thermal band 31, all of which are required by the CM algorithm. 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 example, by imposing a requirement of consistency of the time series of BRF and albedo.

3. 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]. Below, τ is atmospheric optical thickness, πS_λ is spectral extraterrestrial solar irradiance, and $s=(\mu=\cos\theta, \varphi)$ is a vector of direction defined by zenith (θ) and azimuthal (φ) angles. The z-axis is pointed downwards, so $\mu_0>0$ for the solar beam and $\mu<0$ for the reflected beam. 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:

$$L(s_0,s) = D(s_0,s) + L_s(s_0,s)e^{-\tau/|\mu|} + L_s^d(s_0,s)\,. \qquad (3)$$

The surface-reflected radiance is written as:

$$L_s(s_0,s) \cong S_\lambda \mu_0 e^{-\tau/\mu_0}\{\rho(s_0,s) + \alpha c_0 \rho_1(\mu)\rho_2(\mu_0)\} + \frac{\alpha}{\pi}\int_{\Omega^+} D_s(s_0,s')\rho(s',s)\mu' ds'\,, \qquad (4)$$

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

$$\rho_1(\mu) = \frac{1}{2\pi}\int_{\Omega^+}\rho(s',s)ds'\,,\ \rho_2(\mu_0) = \frac{1}{2\pi}\int_{\Omega^-}\rho(s_0,s)ds\,. \qquad (5)$$

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:

$$L_s^d(s_0,s) = \int_{\Omega^-} G^d(s_1,s)L_s(s_0,s_1)ds_1\,. \qquad (6)$$

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:

$$q(\mu_0) = F^{Up}(\mu_0)/F^{Down}(\mu_0)\,, \qquad (7a)$$

$$F^{Down}(\mu_0) = \pi S_\lambda \mu_0 e^{-\tau/\mu_0} + \int_{\Omega^+} D_s(s_0, s')\mu' ds' = F_s^{Dir}(\mu_0) + F_s^{Dif}(\mu_0)\,, \tag{7b}$$

$$F^{Up}(\mu_0) = \pi S_\lambda \mu_0 e^{-\tau/\mu_0} q_2(\mu_0) + \int_{\Omega^+} \mu' q_2(\mu') D_s(s_0, s') ds'\,,\ q_2(\mu_0) = \frac{1}{\pi}\int_{\Omega^-} \rho(s_0, s)\mu ds\,. \tag{7c}$$

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

$$R_\lambda = L_\lambda / (\mu_0 S_\lambda)\,. \tag{8}$$

3.1 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:

$$\rho(\mu_0, \mu, \varphi) = k^L + k^G f_G(\mu_0, \mu, \varphi) + k^V f_V(\mu_0, \mu, \varphi)\,. \tag{9}$$

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. (9) into Eqs. (3-7) and normalization to the reflectance units gives the following expressions for the surface-reflected signal (the last two terms of Eq. (3)):

$$\begin{aligned} R_s(\mu_0, \mu, \varphi) = {} & e^{-\tau/\mu_0}\{k^L + k^G f_G(\mu_0, \mu, \varphi) + k^V f_V(\mu_0, \mu, \varphi) + \alpha c_0 \rho_1(\mu)\ \rho_2(\mu_0)\} \\ & + \alpha\mu_0^{-1}\{k^L E_0^d(\mu_0) + k^G D_G^1(\mu_0, \mu, \varphi) + k^V D_V^1(\mu_0, \mu, \varphi)\}\,, \end{aligned} \tag{10}$$

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

The surface albedo is written as:

$$q(\mu_0) = E_0^{-1}(\mu_0)\{\mu_0 e^{-\tau/\mu_0} q_2(\mu_0) + k^L E_0^d(\mu_0) + k^G D_G^3(\mu_0) + k^V D_V^3(\mu_0)\}. \tag{12}$$

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:

$$\rho_1(\mu) = k^L + k^G f_G^1(\mu) + k^V f_V^1(\mu)\,, \tag{13}$$

$$\rho_2(\mu_0) = k^L + k^G f_G^2(\mu_0) + k^V f_V^2(\mu_0)\,, \tag{14}$$

$$q_2(\mu_0) = k^L + k^G f_G^3(\mu_0) + k^V f_V^3(\mu_0)\,, \tag{15}$$

$$D_k^1(\mu_0,\mu,\varphi-\varphi_0)=\frac{1}{\pi}\int_0^1 \mu' d\mu' \int_0^{2\pi} d\varphi' D_s(\mu_0,\mu',\varphi'-\varphi_0) f_k(\mu',\mu,\varphi-\varphi'), \tag{16}$$

$$D_k^3(\mu_0)=\frac{1}{\pi}\int_0^{2\pi} d\varphi' \int_0^1 \mu' f_k^3(\mu') D_s(\mu_0,\mu';\varphi') d\mu', \tag{17}$$

$$G^{av}(\mu)=\int_{-1}^0 d\mu_1 \int_0^{2\pi} G^d(\mu_1,\mu,\varphi-\varphi_1) d\varphi_1, \tag{18}$$

$$G_k^{11}(\mu)=\int_{-1}^0 f_k^1(\mu_1) d\mu_1 \int_0^{2\pi} G^d(\mu_1,\mu,\varphi-\varphi_1) d\varphi_1, \tag{19}$$

$$G_k^1(\mu_0,\mu,\varphi-\varphi_0)=\int_{-1}^0 d\mu_1 \int_0^{2\pi} G^d(\mu_1,\mu,\varphi-\varphi_1) f_k(\mu_0,\mu_1,\varphi_1-\varphi_0) d\varphi_1, \tag{20}$$

$$H_k^1(\mu_0,\mu,\varphi-\varphi_0)=\int_{-1}^0 d\mu_1 \int_0^{2\pi} G^d(\mu_1,\mu,\varphi-\varphi_1) D_k^1(\mu_0,\mu_1,\varphi_1-\varphi_0) d\varphi_1. \tag{21}$$

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:

$$f_k^1(\mu)=\frac{1}{2\pi}\int_0^1 d\mu' \int_0^{2\pi} f_k(\mu',\mu,\varphi'-\varphi) d\varphi', \tag{22a}$$

$$f_k^2(\mu_0)=\frac{1}{2\pi}\int_{-1}^0 d\mu_1 \int_0^{2\pi} f_k(\mu_0,\mu_1,\varphi_1-\varphi_0) d\varphi_1, \tag{22b}$$

$$f_k^3(\mu')=\frac{1}{\pi}\int_{-1}^0 \mu d\mu \int_0^{2\pi} f_k(\mu',\mu,\varphi-\varphi') d\varphi. \tag{22c}$$

The diffuse and total spectral surface irradiance are calculated from (7b) as:

$$E_0^d(\mu_0)=F^{Dif}(\mu_0)/(\pi S_\lambda),\ E_0(\mu_0)=F^{Down}(\mu_0)/(\pi S_\lambda). \tag{23}$$

Let us re-write equations 10-11 separating the kernel weights. First, separate the small terms proportional to the product $c_0\rho_2(\mu_0)$ into the non-linear term:

$$R^{nl}(\mu_0,\mu)=\alpha c_0\,\rho_2(\mu_0)\,e^{-\tau/\mu_0}\{e^{-\tau/|\mu|}\rho_1(\mu)+k^L G^{av}(\mu)+k^G G_G^{11}(\mu)+k^V G_V^{11}(\mu)\}. \tag{24}$$

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

$$F^L(\mu_0,\mu)=(e^{-\tau/\mu_0}+\alpha\mu_0^{-1}E_0^d(\mu_0))(e^{-\tau/|\mu|}+G^{av}(\mu)), \tag{25}$$

$$\begin{aligned}F^k(\mu_0,\mu;\varphi)&=\{e^{-\tau/\mu_0} f_k(\mu_0,\mu,\varphi)+\alpha\mu_0^{-1}D_k^1(\mu_0,\mu,\varphi)\}\,e^{-\tau/|\mu|}+\\ &\qquad e^{-\tau/\mu_0}G_k^1(\mu_0,\mu,\varphi)+\alpha\mu_0^{-1}H_k^1(\mu_0,\mu,\varphi),\ k=V,\ G.\end{aligned} \tag{26}$$

With these notations, the TOA reflectance becomes:

$$R(\mu_0,\mu,\varphi)=R^D(\mu_0,\mu,\varphi)+k^L F^L(\mu_0,\mu)+k^G F^G(\mu_0,\mu,\varphi)+k^V F^V(\mu_0,\mu,\varphi)+R^{nl}(\mu_0,\mu). \tag{27}$$

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 27, 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 α=1, we omit this non-linearity and equation (27) becomes a linear function of the BRF parameters. With an additional assumption of spectral invariance of the BRF shape (Eq. (2)), formula (27) can be re-written for the pixel (*i, j*) and observation day *k* as:

$$R_{ij}^{k}(\lambda) \cong R^{D}(\lambda, \tau^{k}) + b_{ij}(\lambda) Y_{ij}(\lambda, \tau^{k}), \tag{28}$$

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

$$Y_{ij}(\lambda, \tau^{k}) = k_{ij}^{L,B7} F^{L}(\lambda, \tau^{k}) + k_{ij}^{G,B7} F^{G}(\lambda, \tau^{k}) + k_{ij}^{V,B7} F^{V}(\lambda, \tau^{k}) \tag{29}$$

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 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. The pressure- and water vapor corrections of the LUT functions are performed with the algorithm described in [*Lyapustin and Wang*, 2007].

4. Aerosol Algorithm

The aerosol algorithm consists of two steps: deriving SRCs, and retrieving AOT and Angstrom exponent. The SRC retrievals use parametric formula (28).

4.1 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≤*K*≤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 (28) 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

$$F_1^k = \frac{1}{N^2}\sum_{i,j}\{b_{ij}^{Clear} - b_{ij}^k(\Delta\tau^k)\}^2 = \min\{\Delta\tau^k\}. \qquad (30)$$

The SRCs for the clearest day (b_{ij}^{Clear}) have been calculated for τ_0=0 in step 1. When solving Eq. (30), 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:

$$F_2 = \sum_K\sum_{i,j}\{R_{ij}^{Meas,k} - R_{ij}^{Th,k}(\tau_0 + \Delta\tau^k)\}^2 = \min\{\tau_0\}. \qquad (31)$$

To calculate theoretical reflectance with Eq. (28), 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 and given value τ_0, the SRC can be found by minimizing the *rmse* over all days of the Queue:

$$F_{ij} = \sum_K\{R_{ij}^{Meas,k} - R_{ij}^{Th,k}(\tau^k)\}^2 = \min\{b_{ij}\},\ \tau^k = \tau_0 + \Delta\tau^k, \qquad (32)$$

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

$$b_{ij} = \sum_K[R_{ij}^{Meas,k} - R^D(\tau^k)]Y_{ij}(\tau^k) / \sum_K\{Y_{ij}(\tau^k)\}^2. \qquad (33)$$

Thus, given the aerosol model, Eq. 31 becomes parameterized in terms of the only parameter τ_0. Equations (31) and (32) are positively defined quadratic forms which have unique solutions. To solve these equations numerically, the *MAIAC* algorithm incrementally increases the AOT (e.g. τ_0 in Eq. (31)) 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. (33).

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. 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 $\sqrt{F_1^k}$ ≥0.03. In regular conditions, the value $\sqrt{F_1^k}$ is usually lower than 0.01-0.015.

Retrieving SRCs is a well-optimized and a relatively fast process. Nevertheless, in order to reduce the total processing time, *MAIAC* makes these retrievals with the period of two-three days, which is sufficient to track a relatively slow seasonal variability of the land surface. 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.

4.2 Aerosol Retrievals

With spectral regression coefficients retrieved, the surface BRF in every grid cell in the Blue band becomes known (Eq. 2). Further, the AOT and Angstrom parameter are retrieved at 1 km resolution from the last Tile of MODIS measurements.

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. For example, for the continental USA we are currently using the weak absorption model with the following parameters for the fine (*F*) and coarse (*C*) fractions: median radius R_v^F =0.14 μm, R_v^C =2.9 μm; standard deviation σ^F =0.38 μm σ^C =0.75 μm; spectrally independent real part of refractive index n_r=1.41, and imaginary part of refractive index n_i={0.0044, 0.0044, 0.0044, 0.002, 0.001} at wavelengths {0.4, 0.55, 0.8, 1.02, 2.13} μm, respectively. Parameter n_i is linearly interpolated between five grid wavelengths. 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 different values η ={0.2; 0.5; 1; 2; 5, 10}, which are stored in the operational memory. In this sequence, value η =0.5 gives a model that is close to the urban continental weak absorption (GSFC) model from AERONET classification [*Dubovik et al.*, 2002], whereas the values η=2-5 are more representative of the mineral dust. Following the MISR aerosol algorithm [*Diner et al.*, 1999; 2001; *Martonchik et al.*, ch. 4], a modified linear mixing algorithm [*Abdou et al.*, 1998] is used to mix the LUT radiative transfer (RT) functions for the fine and coarse fractions. This algorithm retains high accuracy with increasing AOT and aerosol absorption.

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

$$R^{Theor,B3}(\eta; \tau_{ij}) = R_{ij}^{Meas,B3}. \qquad (34)$$

In the next step, a spectral residual is evaluated using on the Blue (B3), Red (B1), and SWIR (B7) bands:

$$\chi_{ij} = \sum_{\lambda} \{R_{ij}^{Meas,\lambda} - R_{ij}^{Theor,k}(\tau^{\lambda}(\eta))\}^2 = \min\{\eta\}. \qquad (35)$$

The procedure is repeated with the next value η until the minimum is found. Theoretical reflectance in (35) 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*, 2007; also ch. 5] 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 (34-35) is performed only when the retrieved optical thickness for the standard continental model (η =0.5) exceeds 0.3. Otherwise, a single value of η =0.5 is used and AOT is reported for these background conditions.

The retrieval examples given below were generated using the single low absorption aerosol model described above. In future, the aerosol model will be geographically prescribed based on the AERONET climatology [*Holben et al*., 2001]. The following work is underway: we are studying the AERONET-based classification used in the MODIS aerosol algorithm [*Levy*, 2007] and plan to investigate a MISR level 3 aerosol product, which provides an independent global aerosol climatology over land.

5. Atmospheric Correction

Determination of spectral surface BRF is an integral part of *MAIAC*'s aerosol retrieval process. Below, we describe the atmospheric correction algorithm implemented in *MAIAC*.

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 pixels as detected by the CM algorithm during land-water-snow classification. We also filter out pixels with high AOT (>0.9) where sensitivity of measurements to surface reflectance decreases. 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. The retrieval diagram is shown in Figure 3.

5.1 Inversion for LSRT Coefficients

In the operational MODIS land processing, BRF is determined in two steps: first, the atmospheric correction algorithm derives surface reflectance for a given observation geometry using Lambertian approximation [*Vermote et al.*, 2002], and 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 [*Lyapustin*, 1999].

MAIAC algorithm derives LSRT coefficients directly by fitting the measured TOA reflectance accumulated for a period of 4-16 days. The inversion is based on equation (27) derived earlier. This equation provides an explicit parameterization of TOA reflectance in terms of the BRF model parameters $\vec{K}=\{k^L, k^G, k^V\}^T$.

The quasi-linear form of equation (27) leads to a very efficient iterative minimization algorithm:

$$RMSE = \sum_j (r_j^{(n)} - F_j^L k^{L(n)} - F_j^V k^{V(n)} - F_j^G k^{G(n)})^2 = \min_{\{\vec{K}\}}, \; r^{(n)} = R - R^D - R^{nl(n-1)}, \qquad (36)$$

where index j lists measurements for different days, and n is the iteration number. Equation (36) provides an explicit least-squares solution for the kernel weights. In a matrix form, the solution is written as:

$$\vec{K}^{(n)} = A^{-1}\vec{b}^{(n)}, \qquad (37)$$

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. 3) 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 in all conditions because the non-linear terms are small. The described algorithm is very efficient computationally.

5.2 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, they are not positive-only functions, and they are normalized in somewhat arbitrary fashion not linked to the radiative transfer. These factors reduce the stability of solution upon small perturbation of measurements and may lead to non-uniqueness of solution. The high goodness-of-fit at the angles of measurements does not always 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 the peculiarities of a particular BRF shape. For these reasons, we developed several tests to remove unrealistic solutions.

The initial validation of solution (see Figure 3) checks that the maximal difference over all days of the Queue between measured and computed TOA reflectance does not exceed a specified threshold ($\left|R^{Meas} - R^{LSRT}\right|$>0.08). The day (measurement) with higher deviation is excluded from the Queue and the retrieval is repeated.

If the solution provides a good agreement with measurements for all days, the algorithm verifies that values of the direct-beam albedo (q: also function $q_2(\mu_0)$ in Eq. (7c)) at SZA=15°, 45°, 60° are positive. Finally, the new solution must be consistent with the previous solution: $\left|q(45^o) - q^{\mathrm{Prev}}(45^o)\right|$<Δ(λ). Δ is the band-dependent threshold currently equal to 0.04 (blue), 0.05 (green and red), 0.1 (NIR and shortwave infrared bands, B4-B6), and 0.06 for band B7. The thresholds are relatively loose to allow variations in the solution for surface reflectance. The consistency of the time series of BRF and albedo is characterized by a parameter *status*. Initially, the confidence in the solution is low (*status*=0). Each time, when the new retrieval agrees with the previous retrieval, the value of status increases by 1. When the *status* reaches the value of 4, the retrieval is considered reliable.

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

$$\vec{K}_\lambda^{New} = (\vec{K}_\lambda^{New} + \vec{K}_\lambda^{\mathrm{Pr}\,ev})/2. \tag{38}$$

This updating method increases the quality of BRF and albedo product when the surface is relatively stable, but it delays response of solution to the surface changes.

Often, the solution for some pixels or for the full area cannot be found because of lack of clear-skies measurements. In these cases, *MAIAC* assumes that the surface does not change and fills in the gaps with the previous solution for up to 32-day period. This is the most natural way of gap-filling with specific solution for a given pixel under the assumption of stable surface. The gap-filled pixel is marked as “Extended” in the quality assurance (QA) value with parameter QA.*nDelay* giving the number of days since the last reliable solution.

5.3 *MAIAC* Surface Reflectance Products

MAIAC computes two main products at 1 km resolution for seven 500m MODIS bands, i.e. a set of BRF coefficients and the surface albedo. The albedo is defined by Equation (7a) as a ratio of surface-reflected to incident radiative fluxes. Thus, it represents a true albedo at a given solar zenith angle (SZA) in ambient atmospheric conditions, the value which can be directly compared to the ground-based measurements.

MAIAC also computes several derivative products useful for science data analysis and validation:
1) *NBRF*: a Normalized BRF, which is computed from LSRT parameters for the common geometry of nadir view and SZA = 45^o. This product is similar to MODIS NBAR (nadir BRF-adjusted reflectance) product (part of MOD43 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 for specific viewing geometry of the last day of observations. This product is calculated from the latest MODIS measurements assuming that the shape of BRF, known from previous retrievals, has not changed. To illustrate computation of *IBRF*, let us re-write equation (27) for the measured TOA reflectance as follows:

$$R(\mu_0, \mu, \varphi) = R^D(\mu_0, \mu, \varphi) + cR^{Surf}(\mu_0, \mu, \varphi), \tag{39}$$

where R^{Surf} combines all surface related terms and can be calculated using previous solution for BRF (BRF_λ) and retrieved aerosol information. c is spectrally-dependent scaling factor. Then,

$$IBRF_\lambda(\mu_0, \mu, \varphi) = c_\lambda BRF_\lambda(\mu_0, \mu, \varphi). \tag{40}$$

Below, this algorithm will be referred to as scaling. This description was given as an illustration. In reality, R^{Surf} is a non-linear function so that parameter c_λ and *IBRF* are computed accurately using formulas of sec. 3.

The algorithm computing scaling coefficient (and *IBRF*) is shown in Figure 3 on the right. First, the algorithm filters out measurements which differ from theoretically predicted TOA reflectance based on previous solution (R_Q^{LSRT}) by more than factor of $\Delta(\lambda)$. Then, scaling coefficients are computed, and the consistency requirement is verified as follows: $0.8<c_\lambda<1.2$. If all conditions are satisfied and the *status* of pixel is high (*status*≥4), then the Q-memory is updated with the scaled solution:

$$\vec{K}_{\lambda}^{New} = \frac{c_{\lambda} + 1}{2} \vec{K}_{\lambda}^{\Pr ev} . \quad (41)$$

The time series processing is intrinsically controversial when surface changes rapidly. On the one hand, one needs all available cloud-free measurements and maximal time window in order to reduce the *RMSE.* This approach, which mitigates the noise of measurements, including that of gridding and residual clouds, and which ensures robust BRF shape, is best for stable periods, for example for natural ecosystems in summer time in mid-latitudes. On the other hand, detecting and tracking surface changes like spring green-up or fall senescence requires the least possible number of days in the Queue. Such retrievals tend to have more spatial and spectral noise. Moreover, it is difficult to assess reliability of such solutions when the surface reflectance changes daily with possible data gaps due to clouds. Based on our experience, the combination of one-day solution (*IBRF*) and 16-day solution (*NBRF*) with an update from the last day of measurements (41) combine both the accuracy and an ability to track surface changes.

Besides faster response to the surface change, our update strategy assures fast removal of the retrieval artifacts, mainly residual clouds, which turn out to be the most common problem.

While the response of the 16-day solution may still be delayed, the *IBRF* tracks spectral changes immediately. The update of Q-memory with latest measurements with equation (41) was found to significantly accelerate response of the LSRT coefficients ($\vec{K}_{\lambda}$), and hence of NBRF, to changing surface conditions. Overall, the *IBRF* is better suited for analysis of the fast surface processes and detection of the rapid surface changes.

6. *MAIAC* Cloud Mask

The *MAIAC* cloud mask is a new algorithm making use of the time series of MODIS measurements and combining an image- and pixel-based processing. With a high frequency of MODIS observations, the land surface can be considered as a static or slowly changing background contrary to ephemeral clouds. This offers a reliable way of developing the "comparison clear-skies 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. The uncertainty of the reference value, caused by the natural variability and sensor noise, is directly calculated from the measurements.

The *MAIAC* cloud mask is a next step in evolution of this idea. It uses covariance analysis to build reference clear skies images (*refcm*) and to accumulate a certain level of knowledge about every pixel of the surface and its variability, thus constructing rather comprehensive comparison targets for cloud masking. The reference image contains a clear-skies reflectance in MODIS band 1 (0.645 μm). In order to account for the effects related to scan angle variation, e.g. pixel size growth, surface BRF effect or reduction of contrast at higher view zenith angles (VZA), two reference clear-skies images are maintained by the algorithm, *refcm1* for VZA$\approx$0-45^o and *refcm2* for VZA=45^o-60^o. In addition to *refcm*, the Q-memory also stores the maximal value ($r1_{max}$) and the variance (σ_1) of reflectance in band 1 as well as the brightness temperature contrast ($\Delta BT=BT_{max}-BT_{min}$) for each 25×25 km^2 block. Analysis of MODIS data shows that thermal contrast (ΔBT) is a rather stable metric of a given land area in clear conditions. In partially cloudy conditions, the contrast increases because BT_{min} is usually lower over clouds.

The new CM algorithm has an internal surface classifier, producing a dynamic land-water-snow (*LWS*) mask, and a surface change mask. These are an integral part of *MAIAC* guiding both cloud masking and further aerosol-surface reflectance retrievals when the surface changes rapidly as a result of fires, floods or snow fall/ablation. The cloud mask generated by the CM algorithm is updated during aerosol retrievals and atmospheric correction, which makes it a synergistic component of *MAIAC*. This complex approach increases the overall quality of cloud mask.

Below, we briefly describe the algorithm constructing the reference clear-skies image (*refcm*), and an overall decision logic in cloud masking. Further details of the algorithm can be found in [*Lyapustin et al.*, 2008].

6.1 Building Reference Clear Skies Image

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

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

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

The core of the *MAIAC* CM algorithm is initialization and regular update of the reference clear-skies image for every block. The *refcm* is initially built from a pair of images for which covariance is high, and caution is exercised to exclude correlated cloudy fields. The algorithm calculates a block-level covariance between the *new Tile* and the previous *Tiles*, moving backwards in the Queue until either the “head” of Queue is reached, in which case initialization fails and the algorithm would wait for the new data to arrive, or until clear conditions are found. The latter corresponds to high covariance ($cov\geq 0.68$) and low brightness temperature contrast in the block for both days, $\Delta BT=BT_{max}- BT_{min}<\Delta_1$. The initial value of threshold Δ_1 is currently defined as $\Delta_1=25+dT(h)$ K. Factor dT(h) accounts for the surface height variations in the block and is defined for an average lapse rate, $dT(h)=0.0045(h_{max}-h_{min})$, where h (km) is surface height over the sea level. Once the image *refcm* is initialized, the algorithm begins to use the block-specific value of the brightness temperature contrast Q.ΔBT, which is stored in the Q-memory.

After initialization, the algorithm uses the *refcm* to compute covariance with the latest measurements. Once clear conditions are found, *refcm* and block-parameters $\{r1_{max};\ \sigma_1;\ \Delta BT\}$ are updated. With this dynamic update, the *refcm* adapts to the gradual landcover changes related to the seasonal cycle of vegetation. The rapid surface change events (e.g., snowfall/ablation) are handled through repetitive re-initialization which is performed each time when covariance of the latest *Tile* with *refcm* is found to be low.

Following the covariance calculation, the algorithm looks for clouds at the pixel level. For regular surfaces, not covered by snow, cloud detection is based on a simple postulate that clouds are usually colder and brighter than the surface:

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

where BT_{ij} is measured brightness temperature and $r1_{ij}$ is measured B1 reflectance. 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 a *Dense Vegetation* (or high NDVI) spectral test for a given block, or from the cloud-free neighbor blocks, identified by high covariance.

The final values of the *MAIAC* CM are clear (CM_CLEAR, CM_CLEAR_WATER, CM_CLEAR_SNOW), indicating surface type as well, possibly cloudy (CM_PCLOUD), and confidently cloudy (CM_CLOUD). The value of CM_SHADOW is used for pixels defined as cloud shadows. Shadows are detected with a simple threshold algorithm which compares the latest MODIS measurement (ρ^{meas}) with predicted reflectance (ρ^{pred}) based on the LSRT coefficients from the previous retrievals:

$$\text{IF } \rho^{meas} < \rho^{pred} - 0.12 \Rightarrow CLOUD\_SHADOW.$$

The shadow algorithm uses MODIS band 5 (1.24 μm), which has little atmospheric distortion and is bright over land so that the change of reflectance due to cloud shadow is easy to detect above the noise level.

The covariance component of *MAIAC* algorithm, which offers a direct way to identify clear conditions, renders another commonly used value of cloud mask – “possibly clear” – redundant.

6.2. Performance of *MAIAC* CM Algorithm

The algorithm performance has been tested at scales of 600-1800 km using the 2004-2005 MODIS TERRA data for northeastern USA, Southern Africa (Zambia), Amazon region (Brazil), Arabian peninsula, and Greenland. The testing was done for at least half a year of continuous data in each case, using visual analysis and comparison with MODIS Collection 5 operational cloud mask (MOD35).

Figure 4 shows a case of cloud detection over receding snow for three winter days (36-37, 42) of 2005 for the north-east USA. The area of the image is 600×600 km^2. The two RGB images have a different normalization, helping visual distinction between snow and clouds. The *MAIAC* cloud mask is shown on the right, and the MODIS Collection 5 (MOD35) reprojected and gridded cloud mask is shown at the bottom. The conditions represent different degrees of cloudiness over the land. Day 35 is entirely cloud-free over land. *MAIAC* CM algorithm gives an accurate overall classification. Thin ice on Lake Erie is partly misclassified as clouds. It is not as bright as snow in the visible bands, and has a higher than snow reflectance in the shortwave infrared (2.1 μm). The same holds true for the block of land and some pixels in the transitional zone from snow to land, which are masked as clouds. As explained earlier, the error is expected in these cases. On day 36, *MAIAC* accurately detects a cloud stretching across Lake Erie. There are two large cloud systems on day 42, in the left upper and left bottom parts of the image, captured well by the algorithm. These images also show the strong retreat of the snow line by day 42, and a high quality of snow mapping by *MAIAC*. The MOD35 product accurately detects clouds, but it also overestimates cloudiness over snow on all three days, with the highest error on day 37.

Figure 5 compares the cloud mask of the two algorithms for the late spring of 2005 for the same region. Over land, the accuracy is similar. Some difference exists with regards to thin cirrus, or otherwise semitransparent clouds. *MAIAC* CM does not explicitly try to mask these clouds. Created for the purpose of aerosol retrievals and atmospheric correction, the algorithm maximizes the volume of data available for the atmospheric correction. Our study shows that achievable accuracy of surface reflectance retrievals through thin cirrus is sufficiently high [*Lyapustin and Wang*, 2007] but more investigation is necessary. Another notable difference is cloud detection over the water. The current version of algorithm had been developed for the land applications and cloud detection over water at this stage is rudimentary.

A large-scale comparison of cloud mask products is shown in Figure 6 for a 1200×1200 km^2 region of the African Savannah (Zambia). This is a region of intense biomass burning in the dry season. The *MAIAC* and MOD35 cloud masks are generally comparable. *MAIAC* is a little more sensitive, detecting more clouds. One stark difference is the large number of “possibly clear” pixels in MOD35 when the algorithm cannot declare clear conditions with confidence. This category is not used in *MAIAC*, which has a covariance criterion and ancillary *refcm* data to identify clear conditions. This feature is particularly appealing to land applications, sometimes significantly increasing the volume of measurements, which may be confidently used in the atmospheric correction and in further applied analysis.

A final example of the cloud mask comparison for the large (1800×1800 km^2) bright desert area of the Arabian Peninsula is shown in Figure 7 for days 145 and 207 of 2005. Here, the *MAIAC* cloud mask is shown in the middle of image and the MOD35 product is shown on the bottom. Except for a few small differences, the products agree quite well for day 145. On day 207, MOD35 overestimates cloudiness masking the dust storm areas as clouds.

These examples show that *MAIAC* CM algorithm demonstrates a high accuracy of cloud discrimination over land. If offers potential improvements to the operational MODIS cloud mask when land surface changes rapidly, and over bright snow and ice.

7. *MAIAC* Examples and Validation

The *MAIAC* performance has been tested extensively for different world regions using 50×50 km^2 subsets of MODIS TERRA data centered on AERONET sites. Because the algorithm is synergistic and the quality of aerosol retrievals and atmospheric correction are mutually dependent, the testing includes analysis of all main components of the algorithm. We are using both visual analysis, which remains unsurpassed in complex quality assessments of imagery products, and direct validation of aerosol retrievals by AERONET described in sec. 7.1. One example of such testing is shown in Figure 8 for the Goddard Space Flight Center (GSFC), USA site. The image shows 15 successive MODIS TERRA observations for the end of June - early July (Fig. 8a), and for December (Fig. 8b) of 2000. The left two columns show MODIS top of atmosphere RGB reflectance. The images are normalized differently to help visual separations of clouds and aerosols from the surface signal. With the viewing geometry varying, the images collected at nadir have a better spatial resolution and contrast than those observed at the edge of scan (VZA~55^o) despite aggregation to 1 km. This source of noise notwithstanding, the images display a well-reproducible spatial pattern in cloud-free conditions, which is the basis of the cloud mask algorithm. Because of the re-projection, the top-left and bottom-right corners appear not covered by measurements.

Columns 3-7 and 9 show products of *MAIAC* processing: RGB NBRF (normalized BRF for the standard viewing geometry of VZA=0^o, SZA=45^o); cloud mask; RGB IBRF (instantaneous, or one-angle, BRF for the viewing geometry of latest observation); Blue band AOT (scale 0-1); spectral regression coefficients for the Blue and Red bands (scale 0.1-0.8); NBRF and MODIS L1B TOA reflectance for band B7 (2.1 μm) (scale 0-0.4).

The last column shows the ratio of volumetric concentrations of the coarse and fine aerosol fractions (C_v^C / C_v^F). In these retrievals, values of C_v ratio η={0.5; 1; 2; 4; 10} were used, which are represented by colors magenta, blue, light blue, green and yellow, respectively. One more model used in the retrievals was a liquid water cloud model with the median droplet diameter of 5 μm. The cloud model was used in aerosol retrievals together with aerosol models to detect residual clouds. Cases when this model was selected are shown in red. The aerosol model with η=10, which is a usually unrealistic combination of the fine and coarse fractions, was also used for cloud detection. Although this model may provide a similar to a cloud spectrally neutral extinction, it absorbs in the shortwave spectral region whereas the cloud does not. Overall, the aerosol retrievals provide a valuable enhancement to the cloud mask (yellow and red colors in the last column). Figure 8 shows detection of small popcorn clouds in summer (8a) and semitransparent clouds in winter (8b), as well as extensions on the cloud boundaries, which are difficult to detect by any specialized cloud mask algorithm.

The NBRF and IBRF images are shown as true color RGB composites. The RGB images are produced using the Red (B1), Green (B4) and Blue (B3) channels with equal weights. One can see that the quality of atmospheric correction is generally good although there are still some artifacts related to aerosol retrievals, such as color distortions in the IBRF image (Fig. 8a, third row). The work is ongoing to resolve this and some other remaining issues of *MAIAC* algorithm.

Columns 7-8 in Figure 8 show the derived spectral regression coefficients in the Blue and Red bands. These retrievals are temporally consistent during the short time interval. A comparison of the summer and winter seasons shows an obvious seasonal trend of SRCs. These retrievals can be validated indirectly by comparing aerosol results with the AERONET measurements.

7.1 AERONET Validation

Figures 9-10 show scatterplots of *MAIAC* AOT *vs* AERONET AOT in the Blue and Red bands. Following MODIS validation strategy [*Remer et al.*, 2005], AERONET measurements are averaged over ±30 min interval of TERRA satellite overpass. *MAIAC* retrievals are averaged over 20 km area.

Figure 9 gives a comparison for the GSFC site. The overall agreement is good with relatively high correlation coefficient (*r*~0.78) and slopes of regression, which are close between the Blue and Red bands (0.88 and 0.85). The offset is positive in both bands (0.053 and 0.033, respectively). It can be explained by a limited sensitivity of the method at low AOT values, by residual cloud contamination, and by snow contamination in wintertime. The identification of pixels partially covered by snow is a very complex problem, and a small fraction of undetected snow may notably increase the retrieved AOT. In fact, snow is a significant source of bias in the *MAIAC* retrievals, as shown in the right plots, where winter days with snow on the ground were manually filtered. Snow filtering reduces the bias by a factor of 2 in the Blue band; it also increases correlation coefficient and slope of regression.

Figure 10 compares *MAIAC* AOT with AERONET data for several large cities of the world with medium-to-high levels of pollution, including Moscow, Beijing, Mexico City and Sao Paulo. In addition to pollution, there are several dust storms per year over Beijing with dust blown from the nearest Gobi desert to the north of the city. *MAIAC* results compare well with the AERONET data for Moscow and Beijing with high correlation. There is considerably more scattering in the cases of Mexico City and Sao Paulo. One can see that the slopes of regression are high (~0.9-0.95) for

Moscow and Sao Paulo. In these cases, the *MAIAC* aerosol model with relatively low absorption works reasonably well. However, the slope of regression drops down to ~0.6-0.7 for Beijing and Mexico City, indicating that aerosol is significantly more absorbing.

There is some correlation of noise in the retrievals over Mexico City and Sao Paulo with the viewing geometry, specifically in the forward *vs* backward scattering directions. *MAIAC* tends to underestimate AOT in the forward scattering directions and overestimate it for the backscattering view geometry. One possible explanation of this behavior is an uncompensated surface BRF effect. Over bright surfaces, such as Mexico City and Sao Paulo, the shape of BRF seems to be more anisotropic in the visible bands than in the SWIR, which is also brighter. We plan to address this issue in further research.

7.2 Examples of *MAIAC* Aerosol Retrievals

At present, we have evaluated performance of *MAIAC* 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 11-12. Figure 11 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.

Another example of *MAIAC* aerosol retrievals over a large portion of bright Arabian Peninsula (area 1800×1800 km^2) for day 207 of 2005 is shown in Figure 12. The conditions are rather complex on this day: 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. The conditions are 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 of peninsula, with winds carrying dust in the north-west direction. For comparison, Figure 13 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.

8. Concluding Remarks

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. We plan to apply *MAIAC* to study aerosol and their sources over large urban centers to complement the air quality analysis.

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 second half of 2008 in collaboration with the University of Wisconsin and GSFC-based MODIS land processing team.

Acknowledgements. The work of Dr. Lyapustin and Dr. Wang was supported by the NASA EOS Science (Dr. D. Wickland) grant.

REFERENCES

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.

Diner, D. J., J. Martonchik et al., 1999: MISR level 2 surface retrieval algorithm theoretical basis. *JPL D-11401*, Rev. D, 104 pp.

Diner, D. J., W. Abdou, T. Ackerman et al., 2001: MISR level 2 aerosol retrieval algorithm theoretical basis, JPL D-11400, Rev. E, 104 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.*, **94**, 155-171.

Flowerdew, R. J., and J. D. Haigh, 1995: An approximation to improve accuracy in the derivation of surface reflectances from multi-look satellite radiometers. *Geophys. Res. Lett.* **22**, 1693-1696.

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), 4389-4399, doi:10.1029/2002JD003023.

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, AERONET-A Federated Instrument Network and Data Archive for Aerosol Characterization, *Rem. Sens. Environ.*, **66,** 1-16, 1998.

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., 2004: Aerosol properties over bright-reflecting source regions. *IEEE Trans. Geosci. Remote Sensing*, **42**, 557-569, 2004.

Intergovernmental Panel on Climate Change: Climate Change 2007: Synthesis Report, 52 pp. Available at: http://www.ipcc.ch/ipccreports/ar4-syr.html.

Kaufman, Y. J., D. Tanre et al., 1997: Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *J. Geophys. Res*., **102**, 17051-17067.

Levy, R. C., L. Remer, S. Mattoo, E. Vermote, and Y. J. Kaufman 2007: Second-generation algorithm for retrieving aerosol properties over land from MODIS spectral reflectance. *J. Geophys. Res*., **112**, D13211, doi: 10.1029/2006JD007811.

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. I., Atmospheric and geometrical effects on land surface albedo, *J. Geophys. Res.*, **104**, 4123-4143, 1999.

Lyapustin, A., and Yu. Knyazikhin, Green's function method in the radiative transfer problem. I: Homogeneous non-Lambertian surface, *Appl. Optics*, **40**, 3495-3501, 2001.

Lyapustin, A., and Wang, Y., 2005: Parameterized Code Sharm-3D for Radiative Transfer Over Inhomogeneous Surfaces, *Appl. Optics*, **44**, 7602-7610.

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

Lyapustin, A., Y. Wang, R. Frey: An Automatic Cloud Mask Algorithm Based on Time Series of MODIS Measurements*, J. Geophys. Res*., in press, (2008).

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.

NASA (1999). M. King, & R. Greenstone (Eds.), *EOS reference handbook: a guide to Earth Science Enterprise and the Earth Observation System* (p. 355). Greenbelt, MD: EOS Project Science Office, NASA/Goddard Space Flight Center.

Remer, L., Y. Kaufman, D. Tanre et al., The MODIS Aerosol Algorithm, Products, and Validation, *J. Atm. Sci.*, **62**, 947-973, 2005.

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

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

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**, 15,955-15,964.

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.

Wolfe, R. E., Roy, D. P., and E. Vermote 1998: MODIS Land Data Storage, Gridding, and Compositing Methodology: Level 2 Grid. *IEEE Trans. Geosci. Remote Sens*., **36**, 1324–1338.

FIGURES

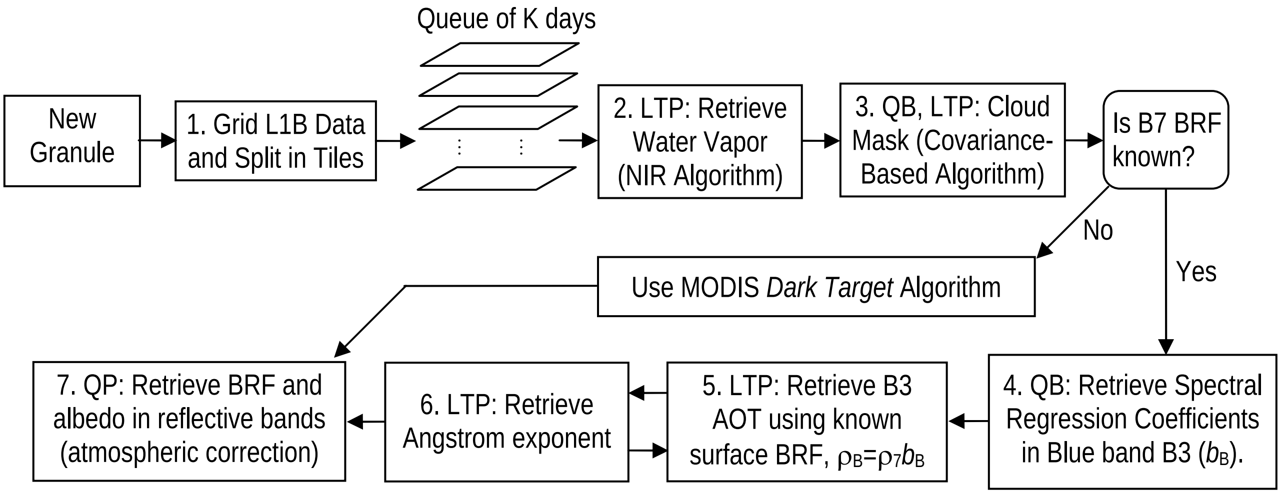

Figure 1. Block-diagram of *MAIAC* algorithm. The initial capital letters indicate spatial and temporal domains of operations, for example at pixel- (P) or/and block- (B) level, and using the data of the last Tile only (LT) or using the full time series of the Queue (Q).

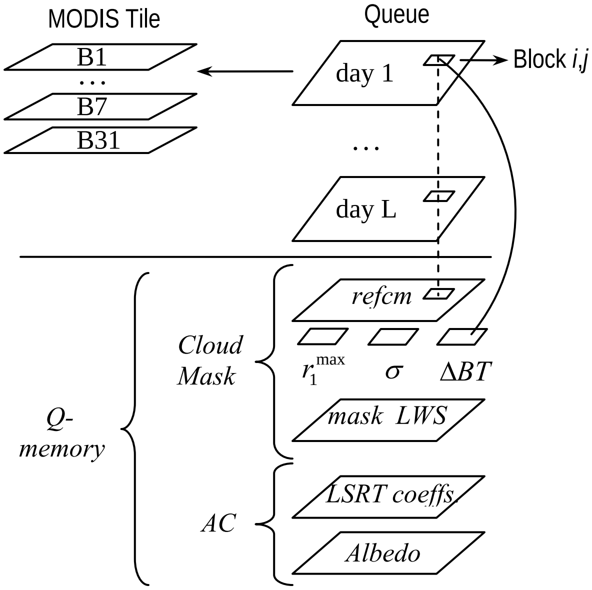

Figure 2. Structure of Queue for *ASRVN* processing. The Queue, designed for the sliding window algorithm, stores up to 16 days of gridded MODIS observations at 1 km resolution. The CM algorithm uses MODIS bands 1-7 and band 31, which are stored as Layers (double-indexed arrays) shown in the upper-left 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 $\{r1_{max}; \sigma_1; \Delta BT\}$, 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 1-7.

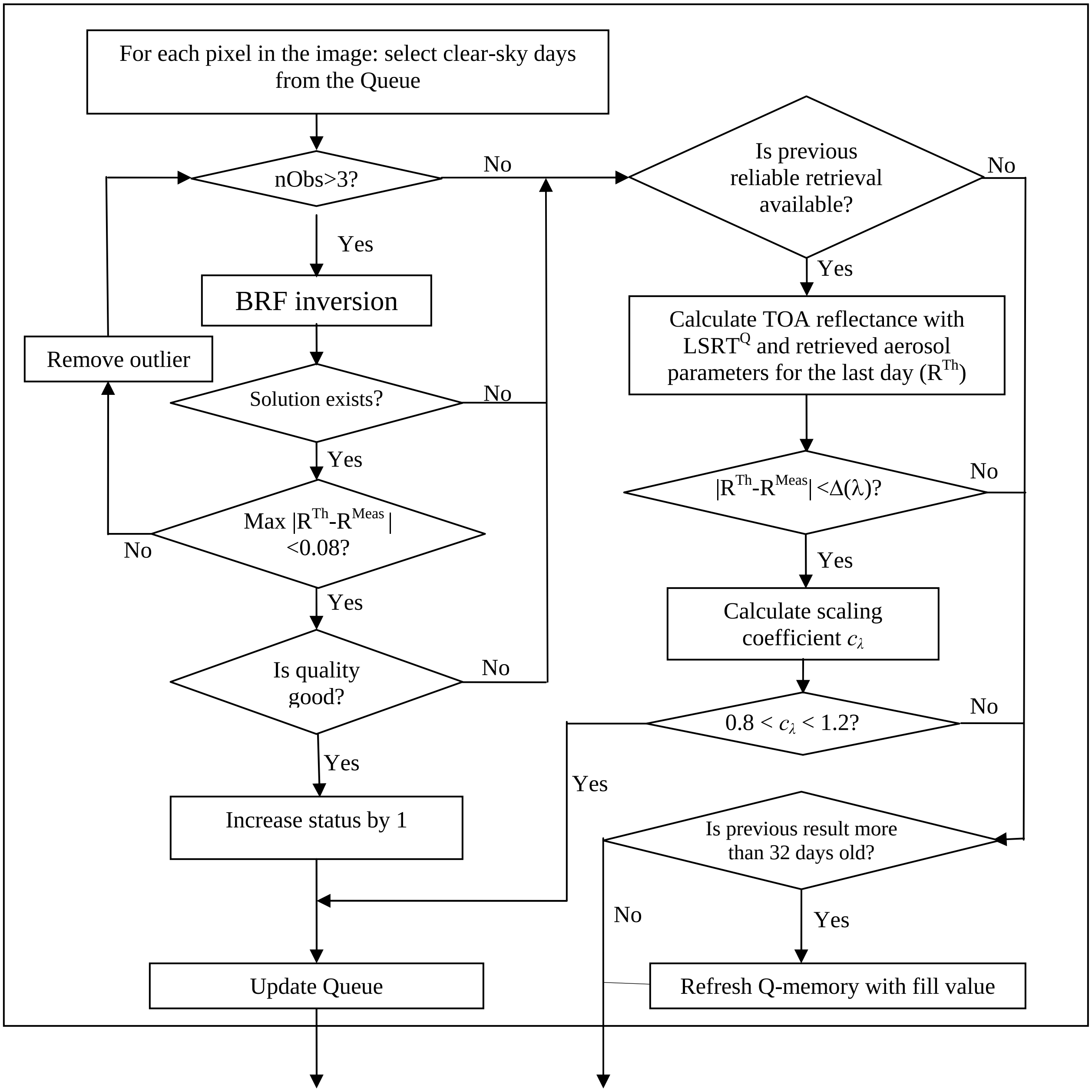

Figure 3. Block-diagram of *MAIAC* atmospheric correction algorithm.

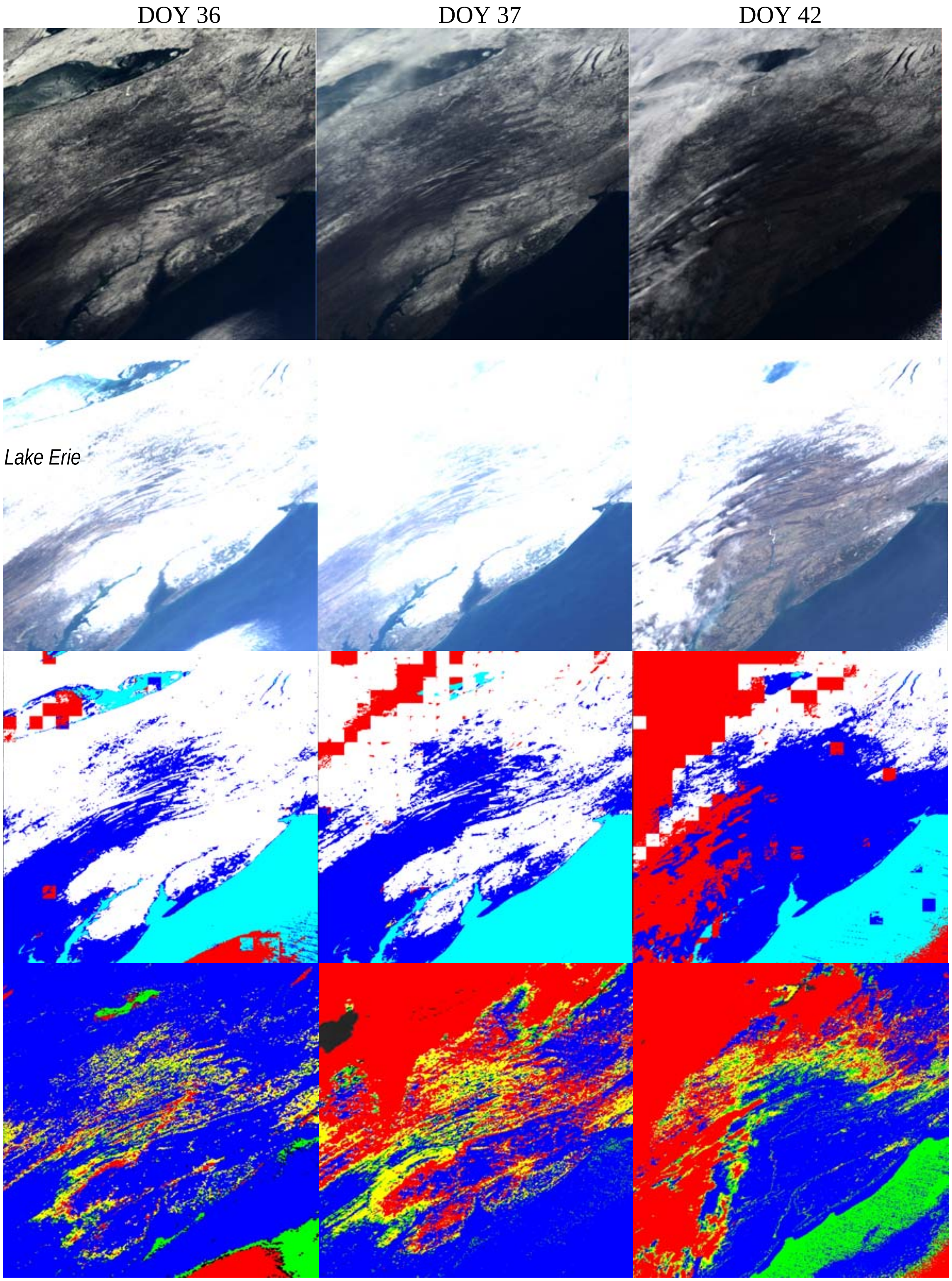

Figure 4. Example of *MAIAC* (3d row) and MOD35 (bottom) cloud mask over snow from MODIS TERRA for days 36, 37 and 42 of 2005. The image shows 1 Tile (600×600 km^2) for the north-east USA. The top two RGB images have a different normalization helping visual distinction between snow and clouds. Legend for MOD35 CM: Blue – clear, Green – possibly clear, Yellow – possibly cloudy, Red – cloudy, Black - undefined. Legend for *MAIAC* CM: Blue, Light Blue, and White – clear (land, water and snow, respectively), Yellow – possibly cloudy, Red – cloudy.

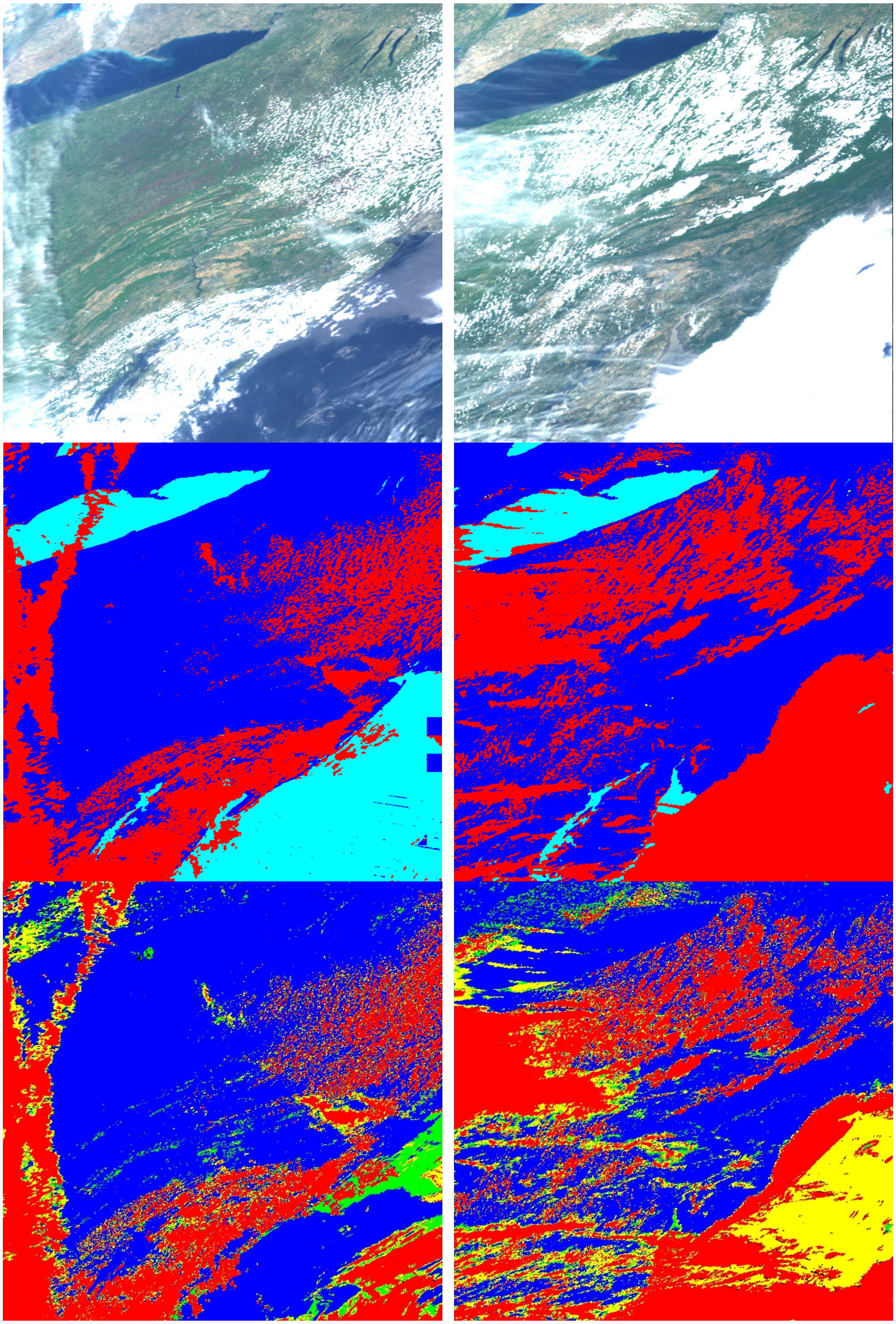

Figure 5. Example of *MAIAC* (middle) and MOD35 (bottom) cloud mask from MODIS TERRA data for days 138 (left) and 152 (right) of 2005. The image shows the same Tile (north-east USA) as in Figure 4.

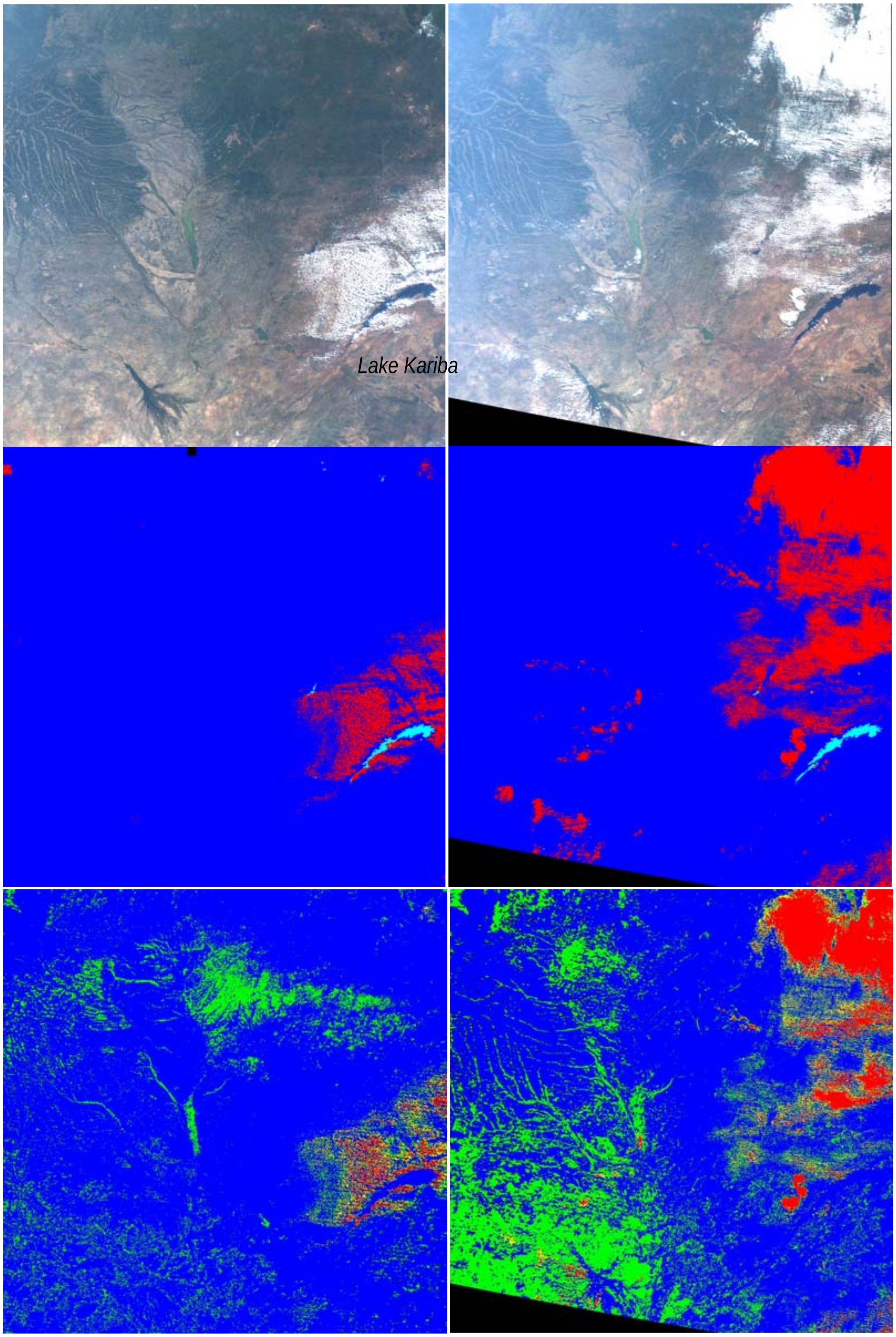

Figure 6. Example of *MAIAC* (middle) and MOD35 (bottom) cloud mask at the beginning of dry season for Zambia, Africa, from MODIS TERRA data for days 130 (left) and 141 (right) of 2005. The image shows 4 Tiles (1200×1200 km^2).

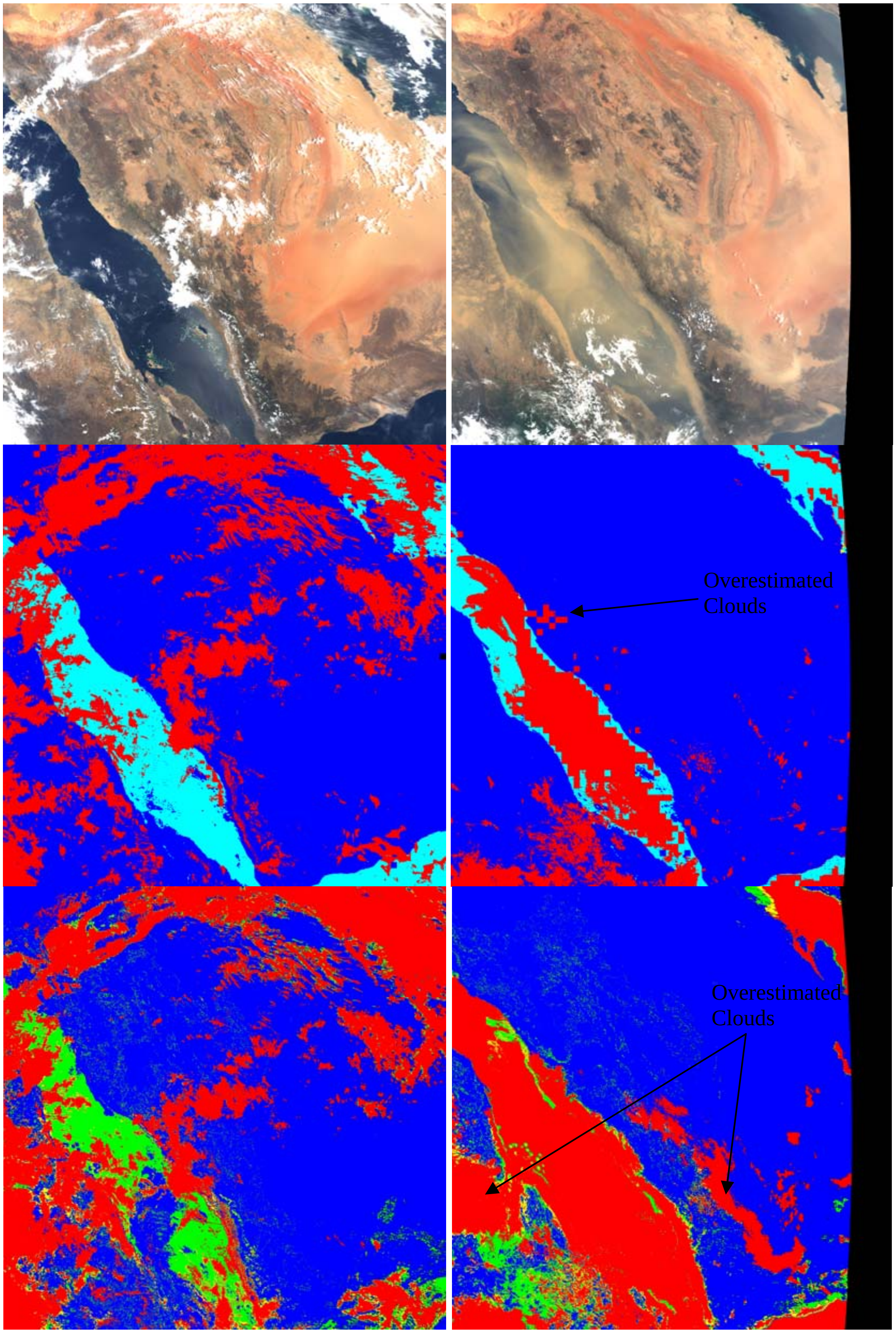

Figure 7. Example of *MAIAC* (middle) and MOD35 (bottom) cloud mask for Arabian Peninsula from MODIS TERRA data for days 145 (left) and 207 (right) of 2005. The image shows 9 Tiles (1800×1800 km^2).

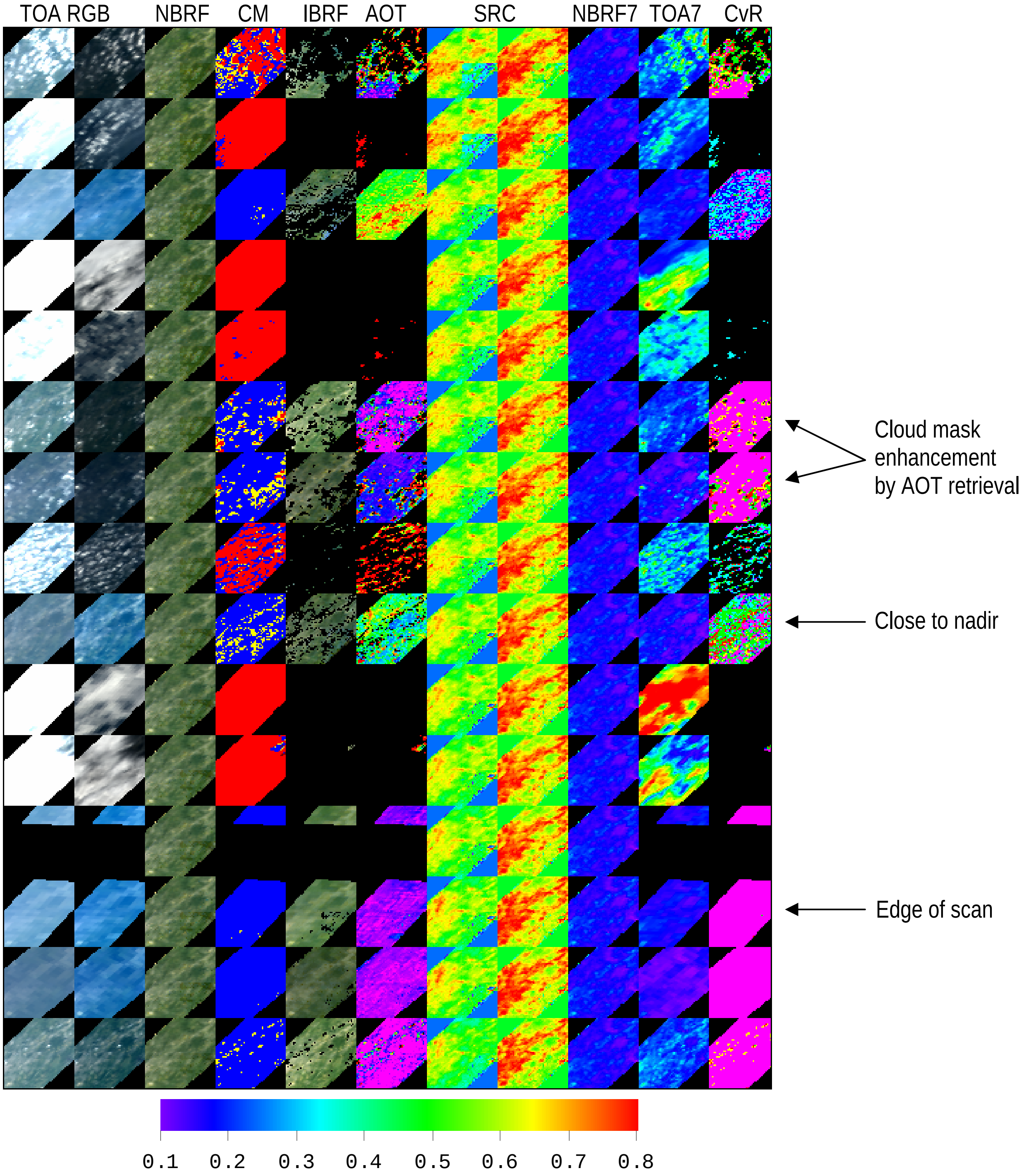

Figure 8a. Example of *MAIAC* processing for 50 km MODIS TERRA subsets for the GSFC site. Shown are 15 consecutive observations for days 175-189 of 2000. The left two columns show differently normalized TOA RGB MODIS gridded reflectance. Next, shown are the following *MAIAC* products: RGB NBRF; cloud mask; RGB IBRF; AOT at 0.47 µm (scale 0-1); spectral regression coefficients for the blue and red bands (scale 0.1-0.8); NBRF and MODIS L1B TOA reflectance for band B7 (2.1 µm) (scale 0-0.4); ratio of volumetric concentrations (C_v^C / C_v^F). The color bar is shown for SRCs.

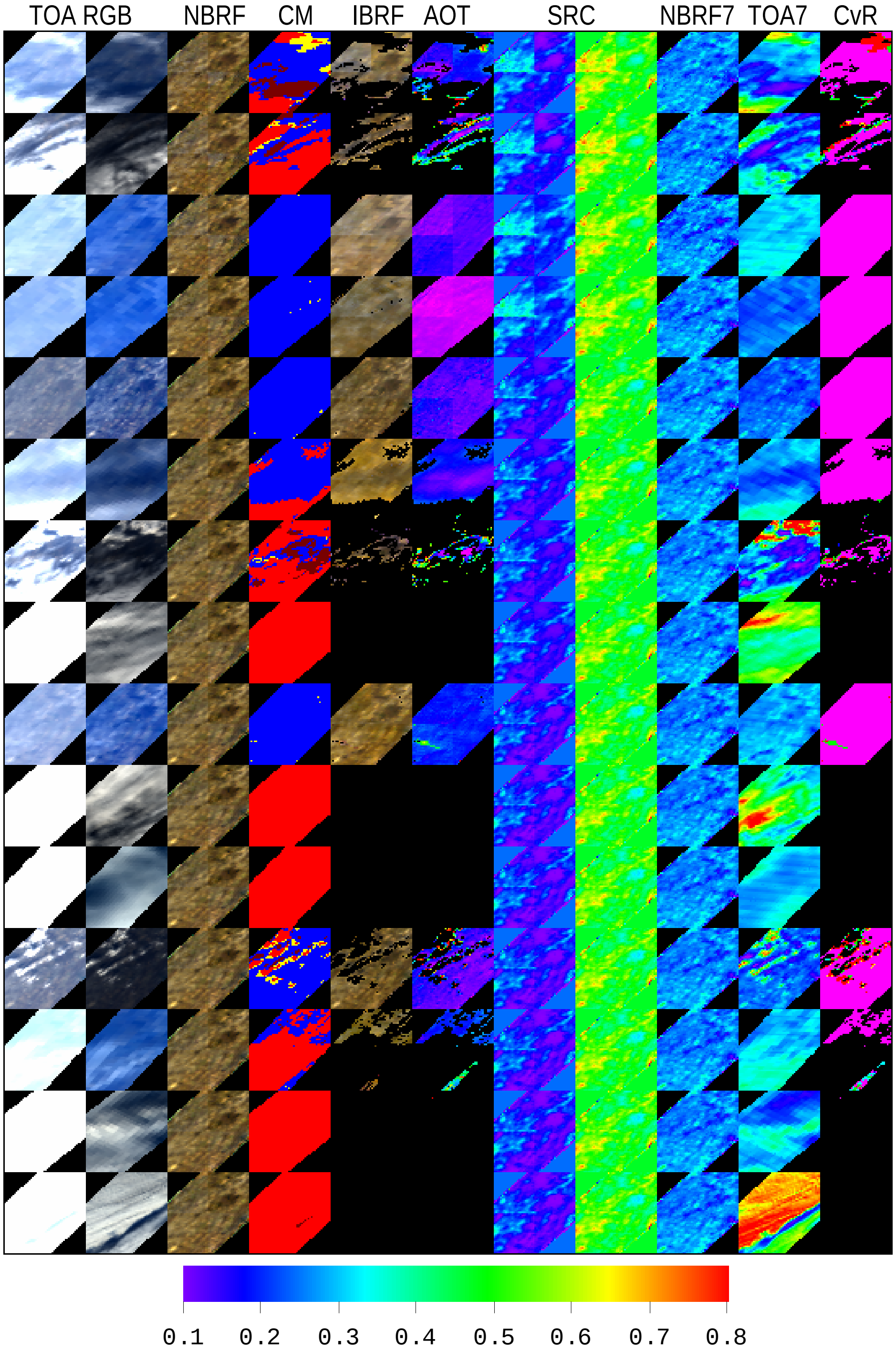

Figure 8b. Example of *MAIAC* processing for the GSFC site, days 337-349 of 2000. The dark red color in CM image shows detected cloud shadows.

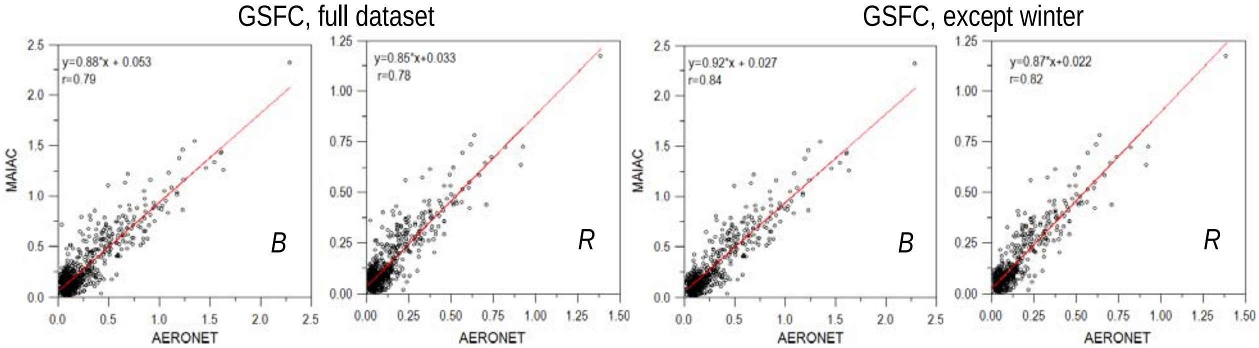

Figure 9. Comparison of *MAIAC* AOT with AERONET data for GSFC (USA) using MODIS TERRA data for 2000-2007. The left plots for both Blue (*B*, 0.47 μm) and Red (*R*, 0.64 μm) bands shows the full dataset (740 points). The right plot shows the reduced dataset where winter days with snow on ground were filtered (587 points).

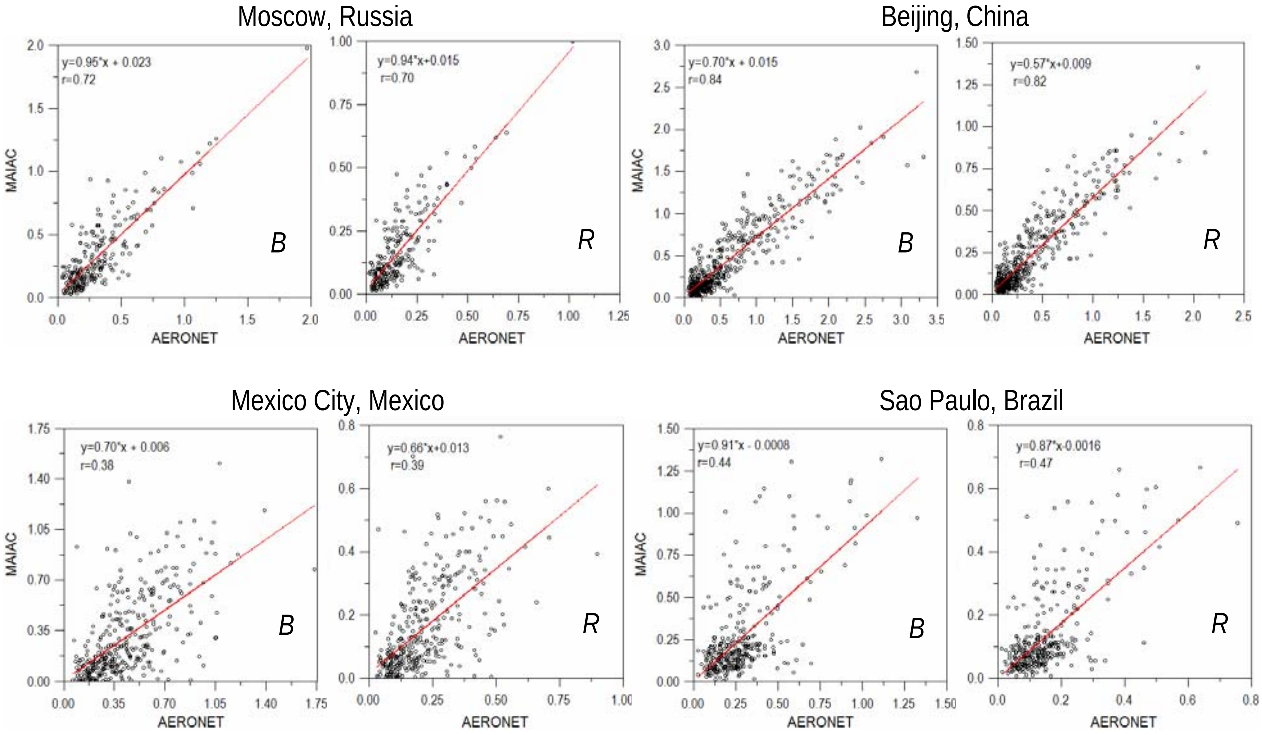

Figure 10. Comparison of *MAIAC* AOT with AERONET data for Moscow (2001-2007, 238 points), Beijing (2001-2007, 504 points), Mexico City (2000-2007, 342 points) and Sao Paulo (2000-2007, 327 points), using MODIS TERRA data.

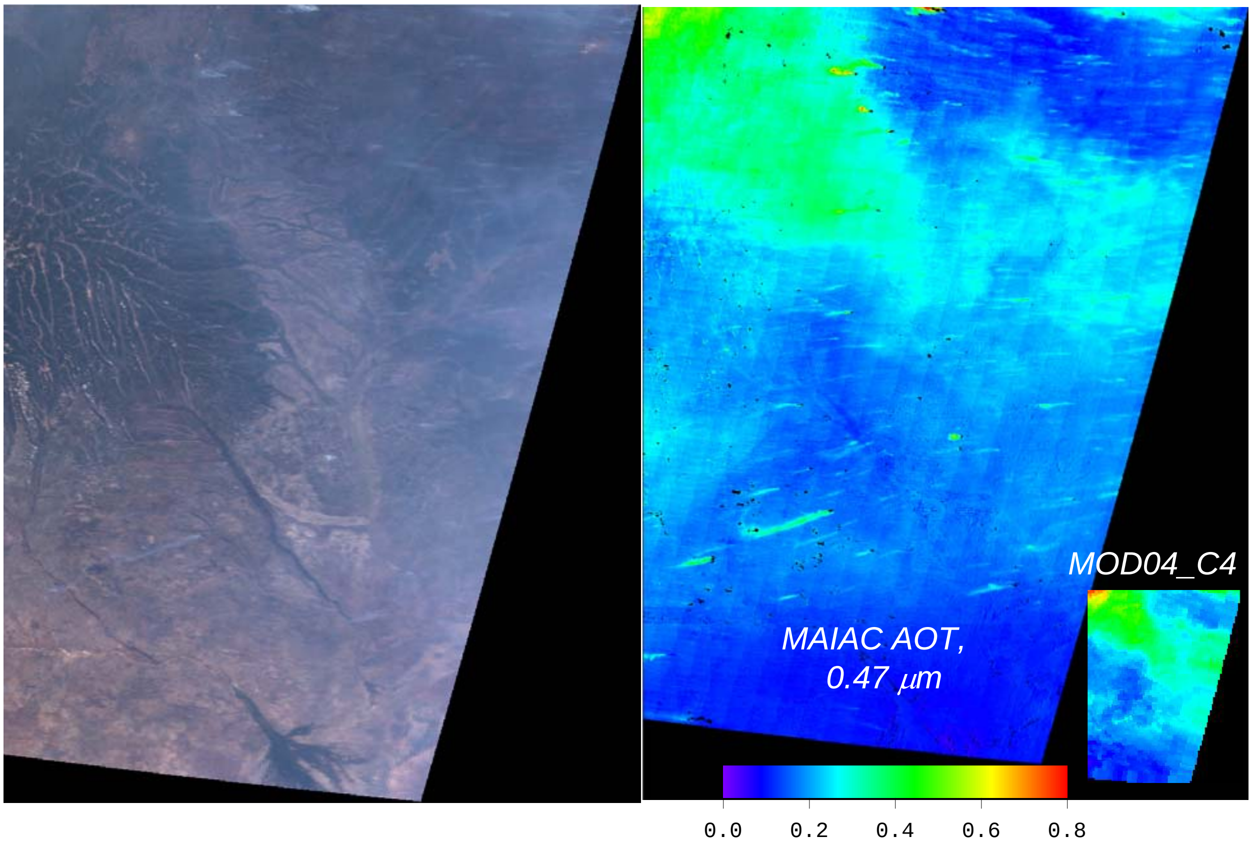

Figure 11. Fires during dry biomass-burning season in Zambia, Africa, for day 205 of 2005 (area 1200×1200 km^2). The 1km gridded MODIS TERRA TOA RGB image is shown on the left and *MAIAC*-retrieved AOT at 0.47 μm is on the right. The AOT scale is the same for MOD04 and *MAIAC*. The high resolution (1 km) of AOT product allows detecting and tracing individual fire plums. The inset shows result of the MODIS dark target algorithm MOD04_C4.

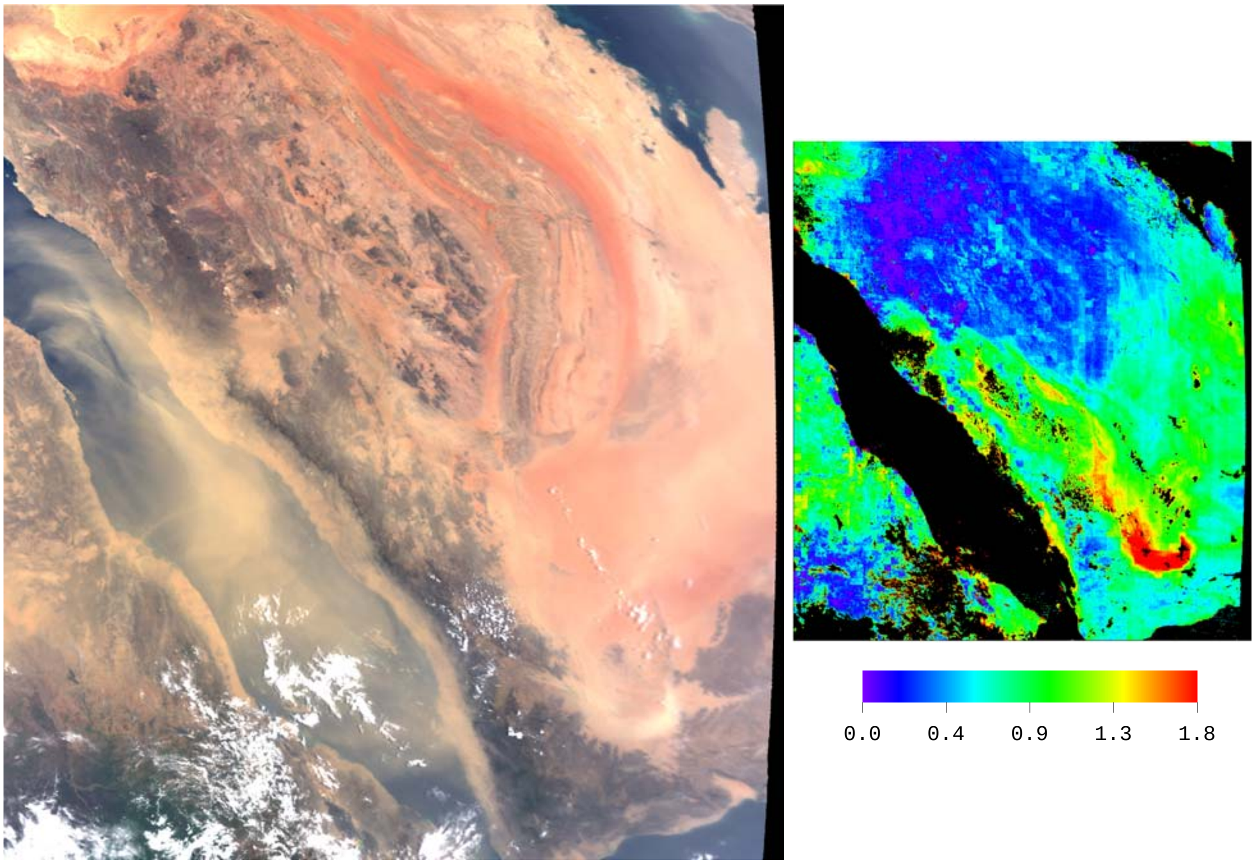

Figure 12. MODIS TERRA RGB TOA image and *MAIAC* AOT at 0.47 μm over Arabian Peninsula (area 1800×1800 km^2) for day 207 of 2005.

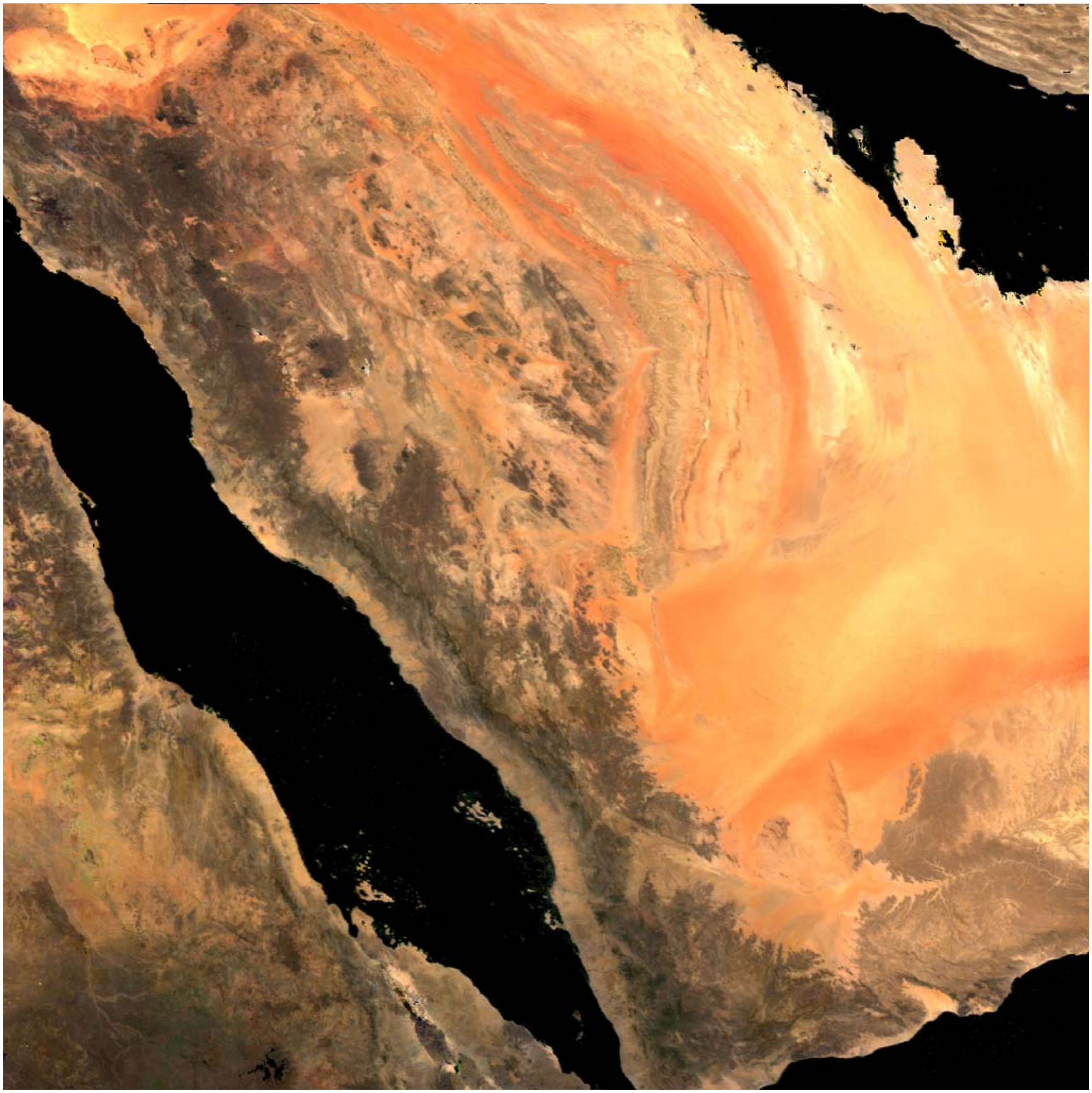

Figure 13. RGB image of surface NBRF (BRF for a fixed geometry, VZA=0^o, SZA=45^o) for Arabian Peninsula for day 184 of 2005. The image is built with equal weights for RGB bands.