MAIAC: MULTI-ANGLE IMPLEMENTATION OF
ATMOSPHERIC CORRECTION
FOR MODIS

ALGORITHM THEORETICAL BASIS DOCUMENT
(ver. 0)

Alexei Lyapustin and Yujie Wang
Goddard Earth Science and Technology Center UMBC, and
NASA GSFC, mail code 614.4 (alyapust@pop900.gsfc.nasa.gov)

This research and development was funded by
NASA Earth Observing System Science grant
(Dr. D. Wickland)

February 7, 2007
# TABLE OF CONTENT

1. **Introduction** ................................................................. 3
   1.1 Overview of MAIAC Algorithm ........................................ 4

2. **MAIAC Aerosol Retrieval Algorithm** .................................. 6
   2.1 Retrieve Spectral Regression Coefficients (SRC) ..................... 6
      2.1.1 Spatial and Seasonal Variability of SRC ....................... 8
      2.1.2 Sensitivity of Spectral Regression Coefficients to Aerosol Model .. 9
   2.2 Retrieve AOT in the Blue and Red Bands, and Angstrom Exponent .... 10
   2.3 AERONET Validation .................................................... 12
   2.4 Examples of MAIAC Aerosol Retrievals ................................ 12

3. **Atmospheric Correction Algorithm** .................................... 20
   3.1 BRF Retrieval in MODIS Band B7 .................................... 20
   3.2 Atmospheric Correction in MODIS Bands 1-6, and 8-14 ................ 23
   3.3 Examples of Large-Scale processing .................................. 25

4. **Radiative Transfer Basis of MAIAC Algorithm** ....................... 31
   4.1 Parameterized Expression for TOA Reflectance with BRF Model ....... 32
   4.2 MAIAC Look-Up Tables ............................................... 34
   4.3 Surface Pressure (Height) and Water Vapor Correction of LUT Functions ... 36
      4.3.1 Correction of Multiple Scattering Path Radiance for Water Vapor Variations ... 36
      4.3.2 Surface Pressure/Height Correction ............................ 39
      4.3.3 Linear Mixing Method ............................................ 38
      4.3.4 Calculating TOA Reflectance from LUT given Surface Pressure and Water Vapor .... 40

5. **Water Vapor Retrieval Algorithm** .................................... 42
   5.1 Theoretical Basis ..................................................... 42
   5.2 Look-Up Tables ....................................................... 43
   5.3 AERONET Validation ................................................ 43
   5.4 Alternative Algorithms .............................................. 44

6. **MAIAC Cloud Mask Algorithm** ....................................... 48
   6.1 Introduction .......................................................... 48
   6.2. Algorithm Technical Detail .......................................... 49
   6.3 Algorithm Flowcharts ................................................ 50
      6.3.1 Land-Water-Snow Mask ....................................... 51
      6.3.2 Module initRefcm() ........................................... 52
      6.3.3 Module CM_highCov() ...................................... 54
      6.3.4 Module CM_lowCov() ....................................... 54
      6.3.5 Module cloudMaskI() ...................................... 55
   6.4. Performance of CM Algorithm ...................................... 56
   6.5. Memory Requirements .............................................. 58
   6.6. Near-Future Tasks ................................................ 58

7. **Software Architecture** .................................................. 66

REFERENCES ................................................................. 67
This document describes a new MAIAC algorithm for the aerosol retrievals and atmospheric correction of MODIS data over land. The algorithm is generic and works globally over all surface types with temporary exception of snow. MAIAC products include cloud mask, water vapor, aerosol optical thickness (AOT) at 0.47 µm and 0.66 µm and Angstrom parameter, surface spectral bidirectional reflectance factor (BRF), direct BRF, 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 direct BRF is derived from the last measurement only. All products are generated uniformly at 1 km resolution in gridded format. The suite of products is fully 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 also give 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 contemporary paradigm of remote sensing from the whisk-broom low-orbit sensors (AVHRR, MODIS etc.) is pixel-based and relies on a single-orbit data. It produces a single measurement for every pixel characterized by two main unknowns, AOT and surface reflectance. This lack of information constitutes a fundamental problem of the remote sensing which cannot be solved without either a priori assumptions, e.g. about spectral regression coefficients for the Dark Target method (Remer et al., 2006; Vermote et al., 2002), or ancillary data, such as a global surface reflectance database for the Deep Blue method (Hsu et al., 2004). The approximate nature of these a priori constraints and use of common simplifications in the radiative transfer model, such as Lambertian model of surface reflectance, limit the accuracy and/or applicability of the current operational aerosol/atmospheric correction algorithms.

The surface-related assumptions in the aerosol retrievals have an unintended effect on the subsequent atmospheric correction. For example, MOD04 and MOD09 algorithms make aerosol retrievals in the Blue (B3) and Red (B1) channels based on estimating surface reflectance from the 2.1 µm channel (B7)

$$\rho^B_3 = 0.25\rho^B_7, \rho^B_1 = 0.5\rho^B_7.$$  (1.1)

It is easy to see that after atmospheric correction the derived surface reflectance in bands B1 and B3 will be described by the relationship (1.1). In other words, the spectral, spatial and temporal variability of surface reflectance in the Blue and Red bands is “slaved” to each other, and ultimately, to the reflectance in band 7. This may have negative implications for the global vegetation and carbon analysis because the chlorophyll-sensing bands are effectively substituted by the band B7, sensitive to the plant liquid water. Using relationship (1.1) also produces such
artifacts as correlation of retrieved AOT with surface brightness (e.g., Lyapustin et al., 2004), because regression coefficients usually grow for brighter, especially artificial (urban) surfaces.

It should be mentioned that MODIS Collection 5 algorithms MOD04 and MOD09 use more sophisticated relationships than (1.1) with empirical correction for the view geometry (BRF effect) and surface greenness. This may remedy described problems somewhat but it does not eliminate their roots.

A generic method, which we developed, retrieves AOT and surface BRF simultaneously using MODIS measurements accumulated for 3 to 16-day interval. The multi-day data provide different view angles which are required for the surface BRF retrieval. The TOA measurements are aggregated and gridded first in order to represent the same surface footprint at different view angles. Then, our method takes advantage of two invariants of the atmosphere-surface system: 1) the surface reflective properties (BRF) change little during accumulation period, and 2) AOT can be assumed constant at short distances (~20 km), because aerosols have a mesoscale range of global variability ~50-60 km (Anderson et al., 2003). Under these generic assumptions, the system of equations becomes over-defined and formally can be resolved. Indeed, our algorithm defines 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 unknown

\[ KN^2 > K+(3N)^2 \text{ if } K>3, \]  

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 of spectral invariance of BRF shape between the 2 \( \mu \)m and the red and blue spectral bands:

\[ \rho^B_{\nu}(\mu, \mu, \varphi) = b^B_{\nu} \rho^B_{\nu}(\mu, \mu, \varphi), \]  

The scaling factor is pixel- and wavelength-dependent. This physically well-based approach reduces the total number of unknown parameters to \( K+N^2 \).

The 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 aerosol variability was high on one of days. Such days are 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 MISR science team, from using of a rigorous mathematical 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 last idea, with 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.

1.1 Overview of MAIAC Algorithm

The block-diagram of MAIAC algorithm is shown in Figure 1.1. 1) The received L1B data are gridded, split in Tiles (~600-1000 km) and Blocks (25 km), and placed in a Queue with the
previous data. In order to limit variation of the footprint with changing 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. 2) The column water vapor is retrieved for the last tile using MODIS near-IR bands B17-B19. The algorithm has the average accuracy of ±5-10% over the land surface. 3) Using the time series of measurements helps to develop a high quality cloud mask based on the notion that the surface spatial structure at the landscape level is highly reproducible in the cloud-free conditions. 4) B7 BRF and albedo are retrieved from queue at a grid resolution. 5) 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) and Red (B1) bands. 6) The AOT is retrieved in the Blue and Red bands at grid resolution using known surface BRF, Eq. (1.3). 7) The ratio of volumetric concentrations of coarse-to-fine aerosol fractions (Angstrom exponent) is calculated for the last tile at the block resolution. This parameter selects the relevant aerosol model and assesses spectral dependence of AOT for the atmospheric correction. 8) Finally, surface BRF and albedo are retrieved at grid resolution from the $K$-day queue for all reflective MODIS bands except B7 and water vapor bands (B17-B19).

The water vapor retrievals are implemented internally to exclude dependence on other MODIS processing streams and unnecessary data transfers. Also, this algorithm is fast. The other

![Figure 1. Block-diagram of MAIAC algorithm.](image)

The LTP, QB and QP abbreviations are used to discriminate between the different time- and space-dimensions of processing (LT – last tile vs Q – $K$-day Queue of blocks (B) or pixels (P)). 1) The received data are gridded, split in Tiles (~600-1000 km) and Blocks (25 km), and placed in a Queue with previous data. 2) Water vapor is retrieved from the last tile at grid resolution. 3) Cloud Mask is generated at block and grid resolution. 4) B7 BRF and albedo are retrieved from Queue at grid resolution. 5) The main algorithm simultaneously retrieves AOT for $K$-days and $N^2$ values of the spectral regression coefficient $b_{ij}$ for the Blue (B3) and Red (B1) bands. 6) The AOT is retrieved in the Blue and Red bands at grid resolution using known surface BRF. 7) The ratio of volumetric concentrations of coarse-to-fine aerosol fractions (Angstrom exponent) is calculated for the last tile at the block resolution. 8) Finally, surface BRF and albedo are retrieved at grid resolution from $K$-day queue for the rest of reflective MODIS bands.
components of \textit{MAIAC} algorithm are interrelated, use a common data structure, and are optimized for the total performance. \textit{MAIAC} has an internal land-water-snow mask and a mask of surface stability/change, which guide the algorithm flow.

The current performance of the algorithm is not yet fully optimized. Nevertheless, \textit{MAIAC} is already sufficiently fast for operational processing: it takes $\approx 45$ s of one single-core AMD Opteron-64 processor to process one Tile ($600 \times 600$ km$^2$) with atmospheric correction of MODIS land bands.

\section{2. \textit{MAIAC} Aerosol Retrieval Algorithm}

This algorithm consists of two parts: 1) retrieving spectral regression coefficients (SRC) for the blue and red bands from multi-day measurements, and 2) retrieving AOT in these bands along with Angstrom parameter for the last day. The aerosol model (specific fine and coarse fractions, defining single scattering albedo) is prescribed regionally based on aerosol climatology provided by AERONET \citep{Holben1998, Dubovik2002}, which is similar to MOD04 algorithmic approach \citep{Remer2006}.

\subsection{2.1 Retrieve Spectral Regression Coefficients}

SRC retrievals use an assumption that the BRF shape is spectrally invariant between the visible (B1, B3) and shortwave IR (SWIR) MODIS band B7,

$$\rho(\lambda; \mu_0, \mu, \varphi) = b(\lambda)\rho_{\text{B7}}(\mu_0, \mu, \varphi).$$

\begin{equation}
(2.1)
\end{equation}

This assumption holds for the surfaces of similar brightness in the visible and SWIR regions. Indeed, the geometric-optics (direct-beam) component of BRF, depending on distribution of shadows and sunlit areas in the field-of-view, is the same because it is defined by the geometric size of 3D structures (surface roughness, size of trees, size of buildings etc.), which is much larger than the wavelengths of interest. The shape of volumetric (diffuse) component of BRF is the result of single and multiple scattering in the layered media (e.g. vegetation). It depends on the average number of internal scattering events, controlled by the single scattering albedo of leaves (for vegetation), and will also be similar for the surfaces of similar brightness. This assumption has been successfully used in the aerosol retrievals in the ATSR-2 \citep{Veefkind1998} and MISR \citep{Diner2005} algorithms.

The SRC retrievals use formula (4.26, sec. 4). According to this formula, the TOA reflectance over pixel $(i,j)$ on the day $k$ can be parameterized as a function of SRC:

$$R_{\text{B7}}(\lambda) \cong R^{\text{D}}(\lambda, \tau^k) + b_y(\lambda)Y_y(\lambda, \tau^k),$$

\begin{equation}
(2.2)
\end{equation}

where function $Y_y(\lambda, \tau^k)$ is calculated from the LUT for a given geometry, wavelength, AOT, and B7 BRF parameters for the pixel $(i,j)$ (Eq. (4.27), sec. 4).

Let us assume that the gridded L1B MODIS reflectance data are available for $3 \leq K \leq 16$ cloud-free days (Figure 2.1). Our goal is to derive the set of $K$ different AOT values over the block of $N \times N$ pixels, and $N^2$ SRC values ($b_y(\lambda)$) for the Blue and Red spectral bands (B3, B1). Implicit in equation (2.1) are two assumptions that 1) AOT is approximately constant in the block for any
The 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, \( \tau_0 \). At this step, the algorithm simultaneously generates the full set of spectral regression coefficients.

The first task is solved using measurements in the Blue band (B3). Initially, SRCs are calculated for every day and every pixel from measured reflectance for AOT=0 using formula (2.1). 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.

On the next step, the AOT difference with the clearest day is calculated independently in the Blue and Red bands. This problem is solved separately for every day \( k \) by minimizing the difference

\[
F_1^k = \frac{1}{N^2} \sum_{i,j} \left( b_{ij}^{\text{Clear}} - b_{ij}^k (\Delta \tau^k) \right)^2 = \min \{\Delta \tau^k\}. \tag{2.3}
\]

The SRCs for the clearest day are calculated once for \( \tau_0 = 0 \). Then, the coefficients for the day \( k \) are calculated using the LUT for the increasing AOT \( \Delta \tau^k \) until the minimum is reached.

On step 3), AOT on the clearest day is found by minimization of \( \text{rmse} \) between the theoretical reflectance and the full set of measurements, including Blue and Red bands, \( K \) days, and \( N^2 \) pixels:

\[
F_2 = \sum_{B3,B1} \sum_K \sum_{i,j} \left( R_{ij}^{\text{Mean},k}(\lambda) - R_{ij}^{\text{Th},k}(\tau_0^k) \right)^2 = \min \{\tau_0\}, \quad \tau_0^k = \alpha_{B3} \tau_0 + \Delta \tau_0^k. \tag{2.4}
\]

The minimization is performed for a selected aerosol model which defines spectral dependence of AOT (\( \alpha_\lambda \)). For example, if \( \tau_0 \) is AOT in the blue band on the clearest day (\( \alpha_{B3} = 1 \)), then \( \alpha_{B1} \tau_0 \) is a respective AOT in the Red band (\( \alpha_{B1} < 1 \) for the non-dust models).

To calculate theoretical reflectance with Eq. (2.2), one needs to know coefficients \( b_{ij}(\lambda) \). These are calculated using the assumption 2) that the surface reflectance does not change during \( K \).
days. Therefore, for a given offset $\tau_0$, the SRC is found by minimizing the $\text{rmse}$ over all days of the queue for a given pixel:

$$F_{ij}(\lambda) = \sum_k \{R_{ij}^{\text{Meas},k}(\lambda) - R_{ij}^{\text{Th},k}(\tau_0^k)\}^2 = \min\{b_{ij}(\lambda)\},$$

(2.5)

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

$$b_{ij}(\lambda) = \sum_k \{R_{ij}^{\text{Meas},k}(\lambda) - R_{ij}^{\text{Th},k}(\tau_0^k)\}H_{ij}(\tau_0^k) / \sum_k \{H_{ij}(\tau_0^k)\}^2.$$

(2.6)

For every $\tau_0$, the SRCs are calculated first in the Blue and Red bands from Eq. (2.6), followed by calculation of the objective function $F_2$. Thus, finding minimum of this function provides spectral regression coefficients for all pixels of the block, and AOT for the clearest day in the Blue and Red bands. The algorithm keeps SRC values for further AOT retrievals, which are performed at grid (1 km) resolution from the last Tile.

This algorithm was developed and optimized through a long series of trial and errors. For example, initially we tried to retrieve the offset $\tau_0^B$ independently in the Blue and Red bands. It turned out that in this case about 30% solutions were physically inconsistent, with $\tau_0^B > \tau_0^R$ for the non-dust aerosol models. On the other hand, joint minimization of an objective function in the Blue and Red bands, and use of constraint $\tau_0^R = \alpha_{B1}\tau_0^B$, $\alpha_{B1}<1$, gives physically robust solutions with $\tau_{k,B}^R < \tau_{k,B}^B$ for every day of the queue.

The algorithm 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 if the main assumptions (1-2) hold. This is done on step 2. If the surface had undergone a rapid change during 16-day interval (e.g. snowfall, fire, flooding), or the AOT is strongly inhomogeneous inside a given block on day $k$, then the value of $\text{rmse} \sqrt{F_{1,k}^k}$ remains high and can be filtered. Currently, the algorithm excludes such days from the processing queue based on criterion $\sqrt{F_{1,k}^k} \geq 0.03$.

Retrieving spectral regression coefficient is the longest part of the total processing (see Figure 1.1 sec.1). For this reason, these retrievals are currently made once every five days, which is sufficient for tracking a relatively slow seasonal variability of surface reflectance. For every block, the retrieved spectral regression coefficients are stored in the memory, along with band 7 BRF coefficients. They are used as ancillary information for aerosol retrievals for the last day of measurements.

### 2.1.1 Spatial and Seasonal Variability of SRC

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).

In order to make aerosol retrievals, MODIS MOD04 and MOD09 algorithms attempt to prescribe SRCs empirically (see Remer et al., 2005; Remer et al., 2006; Vermote et al., 2002). For example, MODIS Collection 4 algorithms used values $\frac{1}{4}$ and $\frac{1}{2}$ globally for the blue and red bands. The latest Collection 5 algorithms use more complicated relationships, empirically corrected for the surface “greenness” using NDVI, and for variations in view geometry, with average values shifted towards 0.3 and 0.55. We found similar higher values for the dense vegetation for the GSFC region from analysis of ETM+ data (Lyapustin et al., 2004). Still, the practice of “guessing” the true relationship between SWIR and visible reflectance causes several artifacts in current products, such as correlation of retrieved AOT with surface brightness and artificial correlation of blue and red surface reflectance with SWIR reflectance. Because reflectance in the blue/red and SWIR band was shown to be a good indicator of vegetation health or water stress condition, the latter may negatively impact land analysis applications.

Figure 2.2 shows an example of SRC spatial and seasonal variability for Bondville (USA), retrieved with MAIAC algorithm. Bondville is an agricultural area at mid-latitudes with strong seasonal cycle. One can see that SRCs for the city area of Urbana-Champaign (IL) (bright spot) change little throughout the year. However, SRCs for surrounding agricultural area change dramatically: they are a factor of 2-3 lower compared to the city area in the early summer, become similar in September-October, and then decrease again. This variability is related to the fractional cover and phenology of vegetation - change of plant leaf chlorophyll and moisture content through the summer-fall seasons.

Another example for the NASA GSFC area (Greenbelt, Maryland) is shown in Figure 2.3. One can see the urban north Washington, DC, and a 95-295 highway corridor with suburbs of southern Baltimore as high values of SRC (e.g. 0.8-1.2 for the Red). The rest of residential and vegetated area of subsets have lower SRC values (0.35-0.8).

2.1.2 Sensitivity of Spectral Regression Coefficients to Aerosol Model

Unlike MISR, MODIS does not really provide a multi-angle dimension to help selecting aerosol model. Therefore, the aerosol model (fine and coarse fractions) must be prescribed regionally based on climatology of AERONET measurements as in MOD04 aerosol algorithm. At the same time, retrieving AOT in the MODIS blue and red bands makes it possible to evaluate an Angstrom parameter, or fractional concentrations of fine and coarse aerosol modes. To make it possible, retrieved spectral regression coefficients should not depend on aerosol model used in the retrievals.

We have studied this dependence experimentally. The look-up tables were generated for three very different aerosol models from AERONET classification, the weak absorption continental (GSFC) model, the spherical dust model (Solar Village), and strongly absorbing biomass burning model (Mongu, African Savannah) (Dubovik et al., 2004). The retrieved SRCs for the GSFC area are shown in Figure 2.3. There are minor differences in the spatial distribution of coefficients, but their magnitude turns out to be quite close regardless of the aerosol model. On the contrary, the retrieved block-level values of AOT differ by the factor of up to 2 between the models.

This means that 1) MAIAC algorithm (Eqs. 2.3-2.6) provides a correct partitioning of solar energy between the surface and the atmosphere, and 2) SRC coefficients can be derived with
static aerosol model in which fractional ratio is fixed. Then, they can be used for unbiased retrievals of AOT and Angstrom parameter from MODIS blue- and red-band measurements, using variable fractional ratio and searching for its optimal value.

2.2 Retrieve AOT in the Blue and Red Bands and Angstrom Exponent
With spectral regression coefficients retrieved, the surface BRF in every grid cell in the Blue and Red bands becomes known (Eq. 2.1), and AOT can be retrieved at a grid resolution (1 km) from the latest measurements. These retrievals use an aerosol model which is geographically prescribed based on AERONET climatology. We are currently using regional aerosol classification similar to MOD04 algorithm (Remer et al., 2006). Because AERONET network does not cover some parts of the world, and has a sparse coverage in many regions with strong aerosol sources (as India, China, etc.), the MOD04 provides only a preliminary classification, which should be detailed further. One possible additional source of global aerosol information is an emerging MISR aerosol climatology (Martonchik et al., 2002; Kahn et al., 2005) which we plan to analyze and incorporate in our retrievals in the near future.

Our current algorithm first retrieves AOT in the Blue and Red bands independently. These values are used to derive the ratio of volumetric concentrations of coarse and fine fractions $\eta$ (see Eq. 4.40, sec. 4), which defines spectral dependence of AOT with selected fine and coarse aerosol fractions ($h_f(\lambda)$, $h_c(\lambda)$):

$$\eta = (\tau_{\text{Red}}^h h_f^{\text{Blue}} - \tau_{\text{Blue}}^h h_f^{\text{Red}}) / (\tau_{\text{Blue}}^c h_c^{\text{Red}} - \tau_{\text{Red}}^c h_c^{\text{Blue}}).$$  (2.7)

Once parameter $\eta$ is known, the AOT can be calculated at any wavelength for the purpose of atmospheric correction, using the Blue band AOT and Eq. (4.40).

![Figure 2.3](image.png)

Figure 2.3 Spectral Regression Coefficients retrieved with different aerosol models (Continental, high absorption Biomass Burning, and Dust) in the Blue and Red bands from MODIS TERRA 50 km subsets for the GSFC area, USA. The GSFC is located at the center of the subsets. The results are shown for different seasons of 2003.
Currently, we are retrieving a single value of parameter $\eta$ for the block, using average values of AOT for the cloud-free (CM_CLEAR) pixels in the Blue and Red bands. This approach is selected from consideration that while the aerosol concentration (AOT) may be highly variable, the aerosol model (type) is defined by regional sources and a long-range transport and changes slowly in space. For the (CM_GREY) pixels, bordering clouds, this parameter is evaluated at the pixels level.

2.3 AERONET Validation

Figure 2.4 shows examples of comparison of MAIAC (black) and MOD04 (blue) MODIS TERRA retrievals and AERONET (blue) measurements for the different world regions of 2003. At this time (Feb. 2007), only MOD04 Collection 4 data are available for comparison, and the MODIS Deep Blue Algorithm (Hsu et al., 2004) has not yet began operational production. The AERONET points represent the closest measurement within ±40 min of TERRA overpass. Satellite retrievals are averaged over 50 km (MOD04) and 20 km (MAIAC) area.

Generally, the AOT agreement between MAIAC, MOD04 and AERONET is very good. We did not do any filtering of MAIAC results over 20 km box before averaging. MAIAC values are sometimes higher than AERONET, usually because of residual clouds (e.g., Mongu, wet season days>280). The early-year differences (e.g. GSFC, days 60-70) are explained by snow conditions, for which MAIAC algorithm is not yet fully developed.

MAIAC retrievals for the USA and European locations were done with low-absorption GSFC AERONET model. This may explain somewhat lower MAIAC AOT for Ispra (Italy) as compared to AERONET.

MAIAC retrievals agree well with AERONET for the bright desert area of Solar Village, where MOD04 algorithm does not work. The high-frequency noise in our retrievals correlates with MODIS TERRA viewing geometry (forward vs back-scattering). Because spherical dust model was used in our retrievals, one possible reason for this noise is non-sphericity of aerosols, which flattens aerosol phase function in the back-scattering directions. Another possible reason is uncompensated surface BRF effect, have the LSRT model had a systematic error bias for the desert surfaces. We plan to resolve these issues in further research.

The MOD04 product also has a good agreement with AERONET measurements. Usually, it offers fewer points for the comparison because of filtering of measurements in the retrieval algorithm, and because it requires at least five retrievals in the 50 km area to be used as a validation point (C. Ichoku, personal communication, 2007). Presently, MOD04 validation points are better filtered for the residual cloudiness.

The MAIAC-AERONET agreement can be further improved by adding analysis of aerosol variability in 20 km block, and by averaging AERONET measurements for the time period of ±30 min of satellite overpass. Both steps are implemented in MOD04 aerosol validation.

2.4 Examples of MAIAC Aerosol Retrievals

Figures 2.5-2.6 show several examples of large-scale Blue-band AOT retrievals from MODIS TERRA. Figure 2.5 shows smoke from biomass-burning during the dry season over an area of 1200×1200 km² in Zambia, Africa. The TOA image for the day 205 shows dozens of small-to-
large fires. The fine 1 km resolution allows us to resolve and trace plumes of the individual fires including the smallest. The fire plumes disappear in the coarse 10 km resolution of MOD04. An image for day 233 shows sharp boundaries of smoke transport, clearly visible in the TOA RGB image. These boundaries are well distinguished in both MOD04 and MAIAC AOT images. A cloud in the middle of dense smoke, with clear smoke density gradient towards the cloud boundaries, is shown on image for day 238. The Angstrom exponent does not increase towards the cloud, i.e. the aerosol particle size remains constant, which may point to aerosols as a nucleation source for cloud formation.

These comparisons show that magnitude of MOD04 and MAIAC retrieved AOT is essentially the same for this region of the world. Analysis of images shows that the spatial distribution of AOT is generally similar, although there are certain differences depending on the day (geometry and season) of observations.

MAIAC aerosol retrievals over a large portion of bright Arabian Peninsula (area $1800 \times 1800$ km$^2$) for days 134, 158 and 207 are shown in Figure 2.6. There is no MOD04 product for this area, and results from Deep Blue algorithm are not yet available for comparison (Feb. 2007).

The TOA RGB image on day 134 shows dust storm with transport across the Red sea. Retrieved AOT values are consistent on both sides of the Red sea, with average optical thickness around 1. The aerosol retrievals were disallowed in storm epicenter in the top-middle part of the image by MAIAC cloud mask algorithm, which needs to be further developed for such applications. On the other hand, CM performance for day 207 with smaller dust storm was correct.

MODIS TERRA RGB and retrieved AOT images for day 158 show dust transport from Africa into the northern part of peninsula. The conditions on day 207 are particularly interesting: on one hand, the dust is carried across the Red sea from Sudan (Africa). The wind does not penetrate mountains along peninsular’ western shore. It is clear on the top of mountain ridge, with the dust concentrated along the shore, as can be seen both from MODIS RGB image and from AOT image. On the other hand, another dust storm has developed in the southern part of peninsula, with winds carrying dust in the north-west direction. For comparison, Figure 2.7 shows the normalized surface BRF for the nadir view and solar zenith angle of 45° (NBRF) image for this area for day 184. The bright surface features, corresponding to the epicenters of dust storms on days 134 and 207 are absent on the surface NBRF image, which proves that these events are indeed local dust storms.
Figure 2.4. Comparison of MAIAC (black), MOD04 Collection 4 (blue) and AERONET (red) AOT at 0.47 μm from MODIS TERRA for different world locations for 2003. The AERONET points represent the closest measurement within ±40 min of TERRA overpass. Satellite retrievals are averaged over 50 km (MOD04) and 20 km (MAIAC) area. The MOD04 product is unavailable for the bright desert region of Solar Village, where retrievals are not performed.
Figure 2.5a. Fires during dry biomass-burning season in Zambia, Africa, for day 205 of 2005 (area 1200×1200 km²). The 1km gridded MODIS TERRA TOA RGB image is shown on the left, MOD04_C4 at the middle, 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 plumes.
Figure 2.5b. Boundaries of smoke transport in Zambia, day 233, 2005. MODIS TERRA RGB image is on the left, \textit{MAIAC} and MOD04\_C4 AOT at 0.47 \( \mu \)m is on the right.

Figure 2.5c. Cloud in the middle of dense smoke cloud, Zambia, day 238, 2005, from MODIS TERRA. MOD04\_C4 AOT at 0.47 \( \mu \)m is overlaid on RGB image, and \textit{MAIAC} AOT is shown on the right.
Fig. 2.6 MODIS TERRA RGB TOA images and MAIAC AOT at 0.47 $\mu$m over part of Arabian Peninsula (area 1800×1800 km$^2$) for days 134 (a), 158 (b), and 207 (c) of 2005.
Figure 2.7. RGB image of surface NBRF (BRF for a fixed geometry, $VZA=0^\circ$, $SZA=45^\circ$) for Arabian Peninsula for day 184 of 2005. The image is built with equal RGB weights. The oval area, which should be more green (see Fig. 3.4) shows the delayed response of MAIAC surface retrievals.
3. Atmospheric Correction Algorithm

This section describes algorithm for BRF retrieval over land in the shortwave IR MODIS band B7 (sec. 3.1), and the general atmospheric correction algorithm for the other land (B1-B6) and unsaturated ocean reflective bands (B8-B14) (sec. 3.2).

3.1 BRF Retrieval in MODIS Band 7

This retrieval provides shape of surface BRF for accurate aerosol retrievals in the Blue and Red MODIS bands (see sec. 2).

For the most cases, aerosol effect in band 7 can be neglected. On the other hand, correction of absorption by well-mixed gases (CO2 and CH4) and by water vapor (see Table 4.1 of sec. 4) is required.

In order to reduce the size of look-up tables and speed-up processing, the atmospheric correction in band B7 is performed in the Lambertian approximation,

\[ R(\mu_0, \mu, \varphi) \cong R^D(\mu_0, \mu, \varphi) + T^\text{tot}(\mu_0, \mu)\rho(\mu_0, \mu, \varphi), \]  

(3.1)

where \( R^D \) is an atmospheric path reflectance and \( T^\text{tot}(\mu_0, \mu) \) is a two-way atmospheric transmittance, calculated for the background aerosol and different values of atmospheric water vapor. Because AOT is usually very low in the 2.1 \( \mu \text{m} \) channel, this formula has generally a high accuracy except when a considerable amount of dust is present in the atmosphere. The background aerosol AOT in band 7 is assumed 0.0055. A separate LUT (“LUT_B7.bin”) has been calculated for band 7, which only contains path reflectance and the two-way atmospheric transmittance.

The surface anisotropic reflectance is modeled with a semi-empirical Li Sparse – Ross Thick (LSRT) BRF model (Lucht et al., 2002) represented as a sum of Lambertian, geometric-optical, and volume scattering components:

\[ \rho(\mu_0, \mu, \varphi) = k^L f^L(\mu_0, \mu, \varphi) + k^V f^V(\mu_0, \mu, \varphi). \]  

(3.2)

This model uses predefined geometrical functions (kernels) \( f^L, f^V \) 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, \( \vec{K} = \{k^L, k^G, k^V \} \).

The BRF parameters are derived for every gridded pixel \((i, j)\) from the available cloud-free measurements covered by a 16-day sliding window. Given geometry and water vapor, the algorithm calculates “measured” BRF from Eq. (3.1) for every cloud-free day \(1 \leq k \leq 16\), \( \rho^k_{ij} = (R^\text{meas.}_{ij} - R^D_{ij}) / T^\text{tot}_{ij} \). Then, three kernel weights are derived by minimizing \( \text{rmse} \), similarly to MODIS BRDF/Albedo algorithm (Schaaf et al., 2002):

\[ F_{ij} = \sum_k (\rho^k_{ij} - k^L f^L_{ij} - k^V f^V_{ij} - k^G f^G_{ij})^2 = \min \{k^L_{ij}, k^V_{ij}, k^G_{ij} \}. \]  

(3.3)

This equation is expanded into three explicit equations derived from:

\[ \frac{\partial F_{ij}}{\partial k^L_{ij}} = 0, \quad \frac{\partial F_{ij}}{\partial k^V_{ij}} = 0, \quad \frac{\partial F_{ij}}{\partial k^G_{ij}} = 0. \]  

(3.4)
The solution is written analytically in the matrix form:

\[ \hat{K} = A^{-1} \hat{b}, \]  

where

\[ A = \begin{bmatrix} \sum_k (f_G^k)^2 & \sum_k f_G^k f_L^k & \sum_k f_V^k f_L^k \\ \sum_k f_G^k f_L^k & \sum_k (f_L^k)^2 & \sum_k f_V^k f_G^k \\ \sum_k f_V^k f_L^k & \sum_k f_V^k f_G^k & \sum_k (f_V^k)^2 \end{bmatrix}, \quad \hat{b} = \begin{bmatrix} \sum_k \rho_{ij}^k f_G^k \\ \sum_k \rho_{ij}^k f_L^k \\ \sum_k \rho_{ij}^k f_V^k \end{bmatrix}. \]  

The inverse of the 3×3 matrix A is calculated analytically. The algorithm requires at least 5 cloud-free measurements (days) for the inversion. The view angles should be sufficiently different to ensure reliable retrieval result. This is verified by checking the determinant of matrix A, which should not be too small (currently \( \det(A) \geq 10^{-3} \)).

The quality of retrieval is tested next using the criterion of positive direct beam albedo (see sec. 4, Eq. 4.5 and 4.13), calculated with derived BRF parameters:

\[ q(\mu_0) = q_2(\mu_0) = \frac{1}{\pi} \int \rho(s_0, s) \mu ds. \]  

Four values of albedo are tested, calculated for the solar zenith angles of 0°, 30°, 50°, 65°. This test is designed to eliminate retrievals which may provide low \( \text{rmse} \), i.e. a good fit to measurements for the specific view angles of MODIS observations, but would give a wrong BRF shape in general. Because albedo is an integrated function of BRF, albedo is sensitive to the overall BRF shape, and this test ensures selection of the reliable solution.

Conceptually, the B7 BRF algorithm is based on the principles of continuity of the BRF time series and allowance for the rapid surface change (Figure 3.1). For every gridded pixel, the algorithm keeps previous set of coefficients in memory. The new measurement is checked initially for consistency with predicted reflectance. The high difference values allow us to filter the undetected clouds and cloud shadows. If the new solution passes test criteria, it is used to update the previous solution. On cloudy days, the LSRT parameters from the memory are used to fill-in gaps for the NBRF (BRF, calculated for fixed geometry, VZA=0; SZA=45°) and albedo.

This retrieval strategy emphasizes the fact that the surface changes gradually and at a much slower rate than the frequency of observations from space. An obvious advantage of this approach is that once initialized, it provides a gapless coverage with BRF model specific for a given pixel (surface type), and a continuous time series of BRF and albedo.

An example of B7 albedo for the GSFC site for different seasons is shown in Figure 3.2, left, and retrieved LSRT parameters are shown on the right. These retrievals are based on the subsets of MODIS TERRA data collected around AERONET sites globally in 2003 by MODAPS. The data are gridded at 1 km resolution. Because of re-projection and edge-of-scan issues, MODIS data do not cover the area of 50×50 km² completely. For specific days, the albedo is shown only for the area where the measurements are available. On the other hand, the BRF model parameters are shown for the full area ever covered by MODIS measurements using fill-in values from the prior retrievals, where necessary. The first queue (days 11-27) is empty during the first several days, when the algorithm was accumulating cloud-free data in order to make an initial retrieval. One
Figure 3.1. Block-diagram of BRF retrieval algorithm for MODIS band B7.

Return F. means return Fail-value. Thresholds: $\Delta_1=0.05$; $\Delta_2=0.03$. (Byte) Parameter Status, stored at
the queue-level, counts how many successive retrievals were done for a given pixel. If Status$\geq 4$, the
retrieval is called reliable. NBRF is BRF calculated for fixed geometry, VZA=0°, SZA=45°.

The algorithm consists of 3 parts, 1) Initial Quality Check, 2) BRF retrieval, 3) Final Quality Check.

The Initial Quality Check makes sure that i) at least 5 days are available for BRF retrieval, and ii) the
new measurement is consistent with the prior reliable retrieval if the latter is available (Status$_{ij}$$\geq 3$). In this case, test A filters undetected clouds when the measured reflectance $\rho^M$ significantly exceeds
predicted reflectance $\rho^P$ (by at least $\Delta_1$). Test B filters possible cloud shadows, when the measured
reflectance is significantly lower than the predicted value. This test is applied only to pixels that are
not very dark, so reduction in the measured reflectance is significant. If Status$_{ij}$ is low (prior reliable
BRF retrieval is unavailable), the algorithm goes directly to the BRF retrieval part.

The BRF Retrieval part first calculates three best-fit coefficients of the LSRT model. Then the
algorithm checks that the quality of retrieval is good, e.g. the difference between the measurements
and the LSRT model for every day of the queue does not exceed $\Delta_2$. If this condition is violated, the
point (day) with maximal deviation is excluded from the queue, and the retrieval is repeated.
The Final Quality Check is designed to eliminate unrealistic solutions. The problem of unrealistic solutions originates in the poor conditioning of the LSRT BRF model. Specifically, 1) the kernels of this model are not orthogonal to each other. This reduces stability of solution upon perturbation of measurements, and may lead to the non-uniqueness of solution; 2) The Geometric-Optic (GO) and Volumetric (V) kernels are not positive-only functions, and have negative-value regions. As a result, certain combinations of view geometries may generate solutions with high goodness-of-fit at the angles of measurements, but with negative BRF (or albedo) at some other angles. The albedo, being an integral function of the BRF, is particularly sensitive to the “wrong” solution (or shape of BRF).

Tests C and D are designed to eliminate the unrealistic solutions. Test C verifies that the direct-beam albedos at SZA=15°, 45°, 65° and NBRF are non-negative. Test D verifies that the direct-beam albedo for the geometry of the last day of the queue is not significantly lower than the lowest measured reflectance of the queue. Both tests were tuned up empirically on the extensive tests with MODIS TERRA data for the AERONET sites globally.

If tests C and D are successful, the algorithm either checks consistency of the new albedo with the previous queue-level albedo for this pixel (if prior reliable retrieval is available), or checks RMSE between the measurements and the retrieved LSRT model, and then updates the queue-value of albedo and vector of LSRT coefficients $\vec{K}$.

can see that the algorithm tracks well the gradual seasonal changes, including spring green-up and fall senescence and defoliation.

3.2 Atmospheric Correction in MODIS Bands 1-6, and 8-14

In this step, BRF parameters $\vec{K} = \{k^L, k^G, k^V\}^T$ and surface albedo ($q(\mu_0)$) are calculated from TOA reflectance using retrieved atmospheric aerosol and water vapor. The retrieval is based on formula (4.25) (sec. 4) for the TOA reflectance:

$$R(\mu_0, \mu, \varphi) = R^D(\mu_0, \mu, \varphi) + k^L F^L(\mu_0, \mu) + k^G F^G(\mu_0, \mu, \varphi) + k^V F^V(\mu_0, \mu, \varphi) + R^{nl}(\mu_0, \mu). \quad (3.8)$$

BRF parameters are derived with the least-squares method:

$$F = \sum \{r_j^{(n)} - k^{L(n)} F_j^{L(n-1)} - k^{V(n)} F_j^{V(n-1)} - k^{G(n)} F_j^{G(n-1)}\}^2 = \min\{k^{L(n)}, k^{V(n)}, k^{G(n)}\}, \quad (3.9)$$

where $r_j^{(n)} = R_j^{\text{meas}} - R_j^D - R_j^{nl(n-1)}$, $n$ is iteration number, and index $j$ counts days in the queue.

Eq. (3.9) is expanded into three explicit equations derived from:

$$\frac{\partial F}{\partial k^L} = 0, \quad \frac{\partial F}{\partial k^G} = 0, \quad \frac{\partial F}{\partial k^V} = 0. \quad (3.10)$$

The solution is written analytically in the matrix form:

$$\vec{K} = A^{-1} \vec{b}, \quad (3.11)$$

with
Figure 3.2 An example of B7 albedo for the GSFC (USA) site (left), retrieved from MODIS TERRA subsets for 2003. The start and end days are given on the top of each queue. Three images on the right show the LSRT model parameters for the queue which starts on the day 247. For the imaging purpose, the queues were truncated. Note that the \( G \)- and \( V \)-kernel weights can be negative. Nevertheless, calculated BRF and albedo are positive and fit a reproducible and continuous time series.
Note, that equation (3.9) is iterative. In the first iteration, the small non-linear term is set to zero, \( R_{nl}^{(0)} = 0 \), and the factor of multiple reflections \( \alpha \) is set to one, \( \alpha^{(0)} = 1 \). These parameters are updated once after the kernel weights, and reflectance factors \( \rho_1(\mu), \rho_2(\mu_0), \) and albedo \( q(\mu_0) \), are calculated in the first iteration. The problem converges with high accuracy in two iterations in all conditions because the non-linear terms are small. Convergence and stability of this algorithm was extensively tested (Lyapustin et al., 2006).

The block-diagram of the algorithm is the same as for band 7, except for the “Initial Quality Check”. The check is not possible for the visible bands because of high variability caused by aerosols.

If BRF (NBRF) and albedo are retrieved from the time series of measurements for a given pixel, then the direct BRF is retrieved using the last measurement. In this sense it is equivalent to the MOD09 surface reflectance. Let us re-write the expression (4.25, sec. 4) for the TOA reflectance by explicitly separating the direct reflected and transmitted solar beam:

\[
R(\mu_0, \mu, \varphi) = R^B(\mu_0, \mu, \varphi) + \mu_0 \rho(\mu_0, \mu, \varphi) \exp(-\tau[\mu_0^{-1} + \mu]) + R_{Surf}(\mu_0, \mu, \varphi). \tag{3.13}
\]

The last term combines all surface reflected terms with radiation scattered at least once in the atmosphere. The direct BRF \( \rho(\mu_0, \mu, \varphi) \) is calculated from (3.13), whereas the surface term \( R_{Surf} \) is computed using the retrieved AOT and the previous reliable BRF retrieval for a given pixel \( K = \{k^L, k^G, k^V\}^T \). The direct BRF does not have the temporal inertia of NBRF/albedo because it depends mostly on the last day measurement. This parameter is sensitive to specific surface conditions on a given day of measurements, and it should be used, for example, for detecting rapid surface change (Roy et al., 2002) or vegetation stress in light use efficiency (LUE) studies (Middleton et al., 2004).

### 3.3 Examples of Large-Scale Processing

The MAIAC algorithm was tested using about 1 year of 2005 MODIS TERRA data for the northeast USA (600×600 km²), Zambia, Africa (1200×1200 km²) and Arabian Peninsula (1800×1800 km²). These regions have very different brightness and seasonal variability of surface, different aerosol conditions and dynamics of cloud cover.

An example of MAIAC surface NBRF for the USA (day 184) are shown in Figure 3.3 (bottom-right). The NBRF is BRF calculated for the nadir view angle and fixed solar zenith angle (VZA=0⁰, SZA=45⁰). The images are built from the red, green and blue bands with equal weight. For comparison, this image also shows the MOD09A1 8-day composite surface reflectance product (E. Vermote) for two closest days, 177 and 185. The MAIAC NBRF image for day 177 is
not shown because it is practically the same as for day 184, with the only difference that it has slightly more fill-values (black pixels). The white spots in the MOD09A1 product are unfiltered clouds. MOD09A1 should be analyzed along with the QA flag which is not shown here. One can see considerable changes in greenness of MOD09A1 on the left part and center of images. The comparison shows that MAIAC product has less noise as compared to MOD09A1. There are also differences in the color: the MAIAC’s reflectance has a closer resemblance of the MODIS RGB images than the MOD09A1 reflectance. This change of color in MOD09A1 may have resulted from the fixed relationship between the blue and the red surface reflectance imposed by the MOD09 aerosol retrieval algorithm (see Eq. 1.1, sec. 1).

Figures 2.7 and 3.4 compares MAIAC NBRF and MOD09A1 surface reflectance for the Arabian peninsula for close days 184 and 177. The arrows in Fig. 3.4 (MOD09A1) show the changed color and some of the artificial borders that should have resulted from compositing (for comparison, see the MODIS RGB image, Fig. 2.6b). On the other hand, MOD09A1 shows correctly the green vegetated spot after rains in the bottom-left part of the image. Because this area was persistently covered by clouds during 2-week period, MAIAC NBRF shows rapid green-up with a delay of about 11-12 days. However, the green-up is shown correctly in the MAIAC direct_BRF product (not shown here).

Figure 3.5a,b shows the surface reflectance comparisons for Zambia. In addition to MOD09A1, it also shows the MOD43 NBAR product. The NBAR is BRF for a fixed geometry of (VZA=SZA=0°). MOD43 is a conservative algorithm using only high quality (QA) input data provided by MOD09. This explains a large fraction of fill values (white spots) in the image. Because of practically cloud-free conditions during the long dry season, the quality of 8-day composite surface reflectance is high.

To further illustrate capabilities of MAIAC algorithm, we created a movie of the NBRF time series for the Zambia region, covering the dry season of 2005 from day 126 to day 280. The first several frames, showing initialization of the algorithm, have the time step of 2-3 days. After initialization, which takes 10-14 days, the time step is approximately 10 days. One can see a gradual phenology change in the region associated with growing soil moisture deficit. The movie can be found along with ATBD at http://neptune.gsfc.nasa.gov/bsb/subpages/index.php?section=Projects&content=SHARM.

Shown comparison is a simple qualitative demonstration of imaging capabilities of different methods of atmospheric correction. In the near future, we will perform comparisons of the time series of spectral surface reflectance, which will give a more reliable assessment of the accuracy (noise) and precision of the atmospheric correction results.

The main validation analysis will use the AERONET-based Surface Reflectance Validation Network (ASRVN). We developed ASRVN as an automated data collection and processing system implemented on a dedicated workstation (Lyapustin et al., 2007). It operationally receives the sensor’s L1B data (currently MODIS and MISR) from MODAPS and Langley DAAC, and aerosol and water vapor data from AERONET server. After successful test of data integrity and completeness, the ASRVN performs an atmospheric correction for each sensor, producing a sensor-specific suite of BRF/albedo and derivative products (vegetation index, radiative flux). The ASRVN products are stored in both swath and gridded formats for the product comparison and cross-sensor cal-val analysis.
Figure 3.3 The MOD09A1 8-day composite surface reflectance product for days 177 (top-left) and 185 (top-right), MODIS TERRA TOA RGB image (day 177, bottom-left), and MAIAC NBRF (day 184, bottom-right). The images are built from the red, green and blue bands with equal weight. The white spots in the composite product are unfiltered clouds.
Figure 3.4 The MOD09A1 8-day composite surface reflectance product for days 177 for Arabian Peninsula (cnf. MAIAC NBRF, Fig. 2.7).
Figure 3.5a The MODIS TERRA RGB image (left), MOD09A1 8-day composite surface reflectance (top right) and MAIAC NBRF (bottom right) for day 145, 2005, Zambia.
Figure 3.5b The MOD43B NBAR (top-left), MOD09A1 8-day composite surface reflectance (top-right) MODIS TERRA RGB image (bottom-left), and MAIAC NBRF (bottom-right) for day 225, 2005, Zambia. The white color in MOD43 product shows fill values.
4. Radiative Transfer Basis of MAIAC Algorithm

The MAIAC algorithm is based on a high accuracy semi-analytical formula derived with the Green’s function method (Lyapustin and Knyazikhin, 2001; Lyapustin and Wang, 2005). This formula expresses the TOA radiance as an explicit function of parameters of the surface BRF model, which is required for both aerosol retrievals over non-Lambertian surface and for the atmospheric correction. The following notations are used below:

- \( \tau \) - atmospheric optical thickness (OT); \( \tau^w \), \( \tau^u \) - absorption OT of 5 well-mixed gases (CO\(_2\), CH\(_4\), NO\(_2\), CO, N\(_2\)O) and of water vapor;
- \( s_0, s \) - incidence and view directions defined by pairs of zenith and azimuthal angles (\( \theta, \varphi \)). For brevity, \( \varphi \) will also stand for the difference \( \varphi = \varphi_0 - \varphi \);
- \( \mu_0, \mu \) - 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.
- \( \pi S_{\lambda} \) - extraterrestrial solar spectral irradiance;
- \( \rho, q \) - 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:

\[
L(s_0, s) = D(s_0, s) + L_s(s_0, s)e^{-\tau_0} + L_s^d(s_0, s). \tag{4.1}
\]

The surface reflected radiance is written as:

\[
L_s(s_0, s) \equiv S_\lambda \mu_0 e^{-\tau_0} \{ \rho(s_0, s) + \alpha \epsilon \rho_1(\mu) \rho_2(\mu) \} + \frac{\alpha}{\pi} \int \frac{\mu}{\Omega} D_s(s_0, s') \rho(s', s) \mu ds'. \tag{4.2}
\]

where \( D_s \) is path radiance incident on the surface, and

\[
\rho_1(\mu) = \frac{1}{2\pi} \int_{\Omega} \rho(s', s) ds', \quad \rho_2(\mu) = \frac{1}{2\pi} \int_{\Omega} \rho(s_0, s) ds. \tag{4.3}
\]

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

\[
L_s^d(s_0, s) = \int_{\Omega} \pi G^d(s_1, s)L_s(s_0, s_1)ds_1. \tag{4.4}
\]

The function \( \pi G^d \) is often called bi-directional upward diffuse transmittance of the atmosphere. The surface albedo is defined as a ratio of reflected and incident radiative fluxes at the surface:

\[
q(\mu_0) = F_{Up}(\mu_0) / F_{Down}(\mu_0), \tag{4.5a}
\]

\[
F_{Down}(\mu_0) = \pi S_\lambda \mu_0 e^{-\tau_0} + \int_{\Omega} D_s(s_0, s') \mu ds' = F_{Dir}(\mu_0) + F_{Dif}(\mu_0), \tag{4.5b}
\]

\[
F_{Up}(\mu_0) = \pi S_\lambda \mu_0 e^{-\tau_0} q_2(\mu_0) + \int_{\Omega} \mu q_2(\mu) D_s(s_0, s') ds', \quad q_2(\mu_0) = \frac{1}{\pi} \int \rho(s_0, s) \mu ds. \tag{4.5c}
\]
The described solution explicitly expresses the TOA radiance as a function of surface BRF, and uses parameterization of the multiple reflections of sunlight between the surface and the atmosphere. The accuracy of the above formulas is high, usually within a few tenths of a percent (Lyapustin and Knyazikhin, 2001). Below we will use a TOA reflectance (unitless), which is defined as

\[ R_\lambda = L_\lambda / (\mu_0 S_\lambda). \] (4.6)

4.1 Parameterized Expression for TOA Reflectance with BRF Model

Using the above solution, the TOA reflectance can be expressed as an explicit function of parameters of the BRF model. The land surface anisotropic reflectance is modeled with a semi-empirical Li Sparse – Ross Thick (LSRT) BRF model (Lucht et al., 2000) represented as a sum of Lambertian, geometric-optical, and volume scattering components:

\[ \rho(\mu_0, \mu, \varphi) = k^L + k^G f^G_G(\mu_0, \mu, \varphi) + k^V f^V_\nu(\mu_0, \mu, \varphi). \] (4.7)

This model uses predefined geometric functions (kernels) \( f^G_G, f^V_\nu \) 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, \( \tilde{K} = \{k^L, k^G, k^V\}^T \). This model is used in the MODLAND processing, and it has been shown to fit well the diverse BRF shapes of the real world (e.g., Schaaf et al., 2002; Lyapustin et al., 2006).

The substitution of Eq. (4.7) into Eqs. (4.1) – (4.5), and normalization to the reflectance units gives the following expressions for the surface-reflected signal (the last two terms of Eq. (4.1)):

\[
\begin{align*}
R_\lambda(\mu_0, \mu, \varphi) &= e^{-\mu_0} \{ k^L + k^G f^G_G(\mu_0, \mu, \varphi) + k^V f^V_\nu(\mu_0, \mu, \varphi) + \alpha c_0 \rho_1(\mu) \rho_2(\mu_0) \} \\
&\quad + \alpha \mu_0^{-1} \{ k^L E^d_0(\mu_0) + k^G D^G_0(\mu_0, \mu, \varphi) + k^V D^V_\nu(\mu_0, \mu, \varphi) \}, \tag{4.8}
\end{align*}
\]

\[
\begin{align*}
R^d_\lambda(\mu_0, \mu, \varphi) &= e^{-\mu_0} \times \{ [k^L G^{av}(\mu) + k^G G^G_G(\mu_0, \mu, \varphi) + k^V G^V_\nu(\mu_0, \mu, \varphi)] + \\
&\quad \alpha c_0 [k^L G^{av}(\mu) + k^G G^{av}_G(\mu) + k^V G^{av}_\nu(\mu) \rho_2(\mu_0)] \} \\
&\quad + \alpha \mu_0^{-1} \{ k^L E^d_0(\mu_0) G^{av}(\mu) + k^G H^G_G^{av}(\mu_0, \mu, \varphi) + k^V H^V_\nu(\mu_0, \mu, \varphi) \}. \tag{4.9}
\end{align*}
\]

The surface albedo is written as:

\[ q(\mu_0) = E_0^{-1}(\mu_0) \{ \mu_0 e^{-\mu_0} q_2(\mu_0) + k^L E^d_0(\mu_0) + k^G D^G_0(\mu_0) + k^V D^V_\nu(\mu_0) \}. \] (4.10)

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

\[
\begin{align*}
\rho_1(\mu) &= k^L + k^G f^G_G(\mu) + k^V f^V_\nu(\mu), \tag{4.11}
\rho_2(\mu_0) &= k^L + k^G f^G_G(\mu_0) + k^V f^V_\nu(\mu_0), \tag{4.12}
q_2(\mu_0) &= k^L + k^G f^G_G(\mu_0) + k^V f^V_\nu(\mu_0), \tag{4.13}
\end{align*}
\]
\[
D_k^1 (\mu_0, \mu, \varphi - \varphi_0) = \frac{1}{\pi} \int_0^1 \mu' d\mu' \int_0^{2\pi} d\varphi' D_s (\mu_0, \mu', \varphi' - \varphi_0) f_k (\mu', \mu, \varphi - \varphi'),
\]

\[
D_k^2 (\mu_0) = \frac{1}{\pi} \int_0^{2\pi} d\varphi' \int_0^1 \mu f_k^2 (\mu') D_s (\mu_0, \mu'; \varphi') d\mu',
\]

\[
G^m (\mu) = \int_0 d\mu_1 \int_0^{2\pi} G^d (\mu_1, \mu, \varphi - \varphi_1) d\varphi_1,
\]

\[
G_k^{11} (\mu) = \int f_k^1 (\mu_1) d\mu_1 \int G^d (\mu_1, \mu, \varphi - \varphi_1) d\varphi_1,
\]

\[
G_k^1 (\mu_0, \mu, \varphi - \varphi_0) = \int d\mu_1 \int G^d (\mu_1, \mu, \varphi - \varphi_1) f_k (\mu_0, \mu_1, \varphi_1 - \varphi_0) d\varphi_1,
\]

\[
H_k^1 (\mu_0, \mu, \varphi - \varphi_0) = \int d\mu_1 \int G^d (\mu_1, \mu, \varphi - \varphi_1) D_k^1 (\mu_0, \mu_1, \varphi_1 - \varphi_0) d\varphi_1.
\]

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:

\[
f_k^1 (\mu) = \frac{1}{2\pi} \int_0^{2\pi} d\varphi' \int f_k (\mu', \mu, \varphi' - \varphi) d\varphi',
\]

\[
f_k^2 (\mu_0) = \frac{1}{2\pi} \int_0^{2\pi} d\mu_1 \int f_k (\mu_0, \mu_1, \varphi_1 - \varphi_0) d\varphi_1,
\]

\[
f_k^3 (\mu') = \frac{1}{\pi} \int d\mu d\mu_1 f_k (\mu', \mu, \varphi - \varphi) d\varphi.
\]

The diffuse and total spectral surface irradiance are calculated from (4.5b) as:

\[
E_0^d (\mu_0) = F^{D,0} (\mu_0) / (\pi S_\lambda), \ E_0 (\mu_0) = F^{T,0} (\mu_0) / (\pi S_\lambda).
\]

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

\[
R^{al} (\mu_0, \mu) = c_0 \rho_2 (\mu_0) e^{-\mu_0} \{ e^{-\mu} | \rho_1 (\mu) + k^G G^G (\mu) + k^G G^G (\mu) + k^V G^V (\mu) \}.
\]

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

\[
F^L (\mu_0, \mu) = ( e^{-\mu_0} + \alpha_0 \rho_1 (\mu) ) ( e^{-\mu} | G^G (\mu) ) + F^X (\mu_0, \mu),
\]

\[
F^X (\mu_0, \mu, \varphi) = \{ e^{-\mu_0} f_X (\mu_0, \mu, \varphi) + \alpha_0 \rho_1 (\mu) \} e^{-\mu} | G^G (\mu, \mu, \varphi) + \alpha_0 \rho_1 (\mu) | H^X (\mu_0, \mu, \varphi),
\]

where index \( X \) refers to volumetric (\( V \)) or geometric-optical (\( G \)) kernels. With these notations, the TOA reflectance becomes an explicit function of the surface BRF model parameters:

\[
R (\mu_0, \mu, \varphi) = R^{al} (\mu_0, \mu, \varphi) + k^L F^L (\mu_0, \mu) + k^G F^G (\mu_0, \mu, \varphi) + k^V F^V (\mu_0, \mu, \varphi) + R^{al} (\mu_0, \mu).
\]
MAIAC algorithm uses this equation for an efficient atmospheric correction.

Let us give a modified form of equation (4.25) which is used in the aerosol retrievals. The last (non-linear) term of this formula, which describes multiple reflections of the direct-beam sunlight between the surface and the atmosphere, is small \((R_{nl} \propto \tau c)\), and can be neglected for simplicity of further consideration. The functions \(F_{\tau}\) are also weakly non-linear via parameter \(\tau\), which describes the multiple reflections of the diffuse incident sunlight. By setting \(\tau = 1\), we omit this non-linearity and equation (4.25) becomes a linear function of the BRF parameters. Let us re-write Eq. (4.25) for a specific pixel \((i, j)\) on \(k\)th day of observation. Index \(k\) will cumulatively describe an aerosol optical thickness \((\tau_k)\) and MODIS viewing geometry \((\mu_\theta, \mu_\varphi)^k\) on day \(k\) for the pixel \((i, j)\). With an additional assumption of spectral invariance of the BRF shape (Eq. (3)), formula (4.25) can be re-written as:

\[
R^k_i(\lambda) \approx R^0(\lambda, \tau_k^k) + b_y(\lambda)Y_y(\lambda, \tau_k^k),
\]

where \(b_y(\lambda)\) is spectral regression coefficient for the Blue or Red bands. A function

\[
Y_y(\lambda, \tau_k) = k_y^{L,0\lambda}F^L(\lambda, \tau_k^k) + k_y^{G,0\lambda}F^G(\lambda, \tau_k^k) + k_y^{V,0\lambda}F^V(\lambda, \tau_k^k)
\]

is easily calculated from the LUT for a given geometry, AOT and wavelength, once the BRF parameters in band \(B7\) for the pixel \((i, j)\) \{\(k_y^{L,B7}, k_y^{G,B7}, k_y^{V,B7}\)\} are known.

### 4.2 MAIAC Look-Up Tables

The LUT stores functions \(f_X, f_X^1, f_X^2, f_X^3\), which depend on geometry of observations only, and functions \(D_X^1, D_X^3, G^1, G^2, G^3, H^1, E^0, E_0, R_{Dms}\), which depend on geometry, selected aerosol model and AOT. Index \(X\) 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 pressure and water vapor. The LUT is computed for a dense grid of VZA, SZA, and azimuthal angles \((\Delta \mu_\theta = 0.02\) for the range \(0.4 – 1 \left(0^\circ - 66.42^\circ\right)\), and \(\Delta \varphi = 3^\circ\)). Similarly to MISR processing, the algorithm uses the nearest neighbor for speed consideration because it avoids 3D interpolation in angles. The angular dimension of LUT is 31\(\times\)31\(\times\)61=58,621 entries. In addition, the LUT stores function \(c_0\) depending on aerosol model and AOT. The LUT is calculated for 17 AOT values, \{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.5, 0.6, 0.8, 1.0, 1.2, 1.6, 2.0, 2.5\}. A linear interpolation is used to derive LUT functions for the required AOT. Current size of the LUT for one aerosol fraction for 7 MODIS land bands and 8 ocean bands is 81 MB. We plan to have 7-8 total aerosol fractions in the MAIAC algorithm, which will translate into \(\approx 600\) MB size of the LUT.

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 (X. Xiong, personal...
communication). The radiative transfer model (RTM) included absorption of 6 major atmospheric gases (H\textsubscript{2}O, CO\textsubscript{2}, CH\textsubscript{4}, NO\textsubscript{2}, CO, N\textsubscript{2}O) 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. Also, LUTs are generated for a fixed column water vapor, $\tau^w=0.5$ cm. The correction of LUT functions for the water vapor variations is done analytically, as described in sec. 4.3.1.

The LUT is calculated as follows: functions $R^{D\mu}(\mu_0, \mu, \varphi)$, $E_0(\mu_0)$, $E_0^d(\mu_0)$, $c_0$, $\lambda^\text{eff}_C$ (effective band center wavelength), and $\tau^R$, $\tau^g$ (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:

$$R^\varphi(\mu_0, \mu, \varphi) = \int S_\lambda R^\varphi(\mu_0, \mu, \varphi) h_\lambda d\lambda \int S_\lambda h_\lambda d\lambda .$$

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:

$$\exp\{-\tau^R(\lambda^\text{eff}_\text{Band})\} = \int S_\lambda \exp\{-\tau^g(\lambda)\} h_\lambda d\lambda \int S_\lambda h_\lambda d\lambda .$$

On the next step, the functions of kernels ($D^i_{go}(\mu_0, \mu, \varphi)$, $D^i_v(\mu_0, \mu, \varphi)$, $G^\text{av}(\mu)$, $G^i_{go}(\mu_0, \mu, \varphi)$, $G^i_v(\mu_0, \mu, \varphi)$, $G^i_{go}(\mu)$, $G^i_v(\mu)$, $H^i_{go}(\mu_0, \mu, \varphi)$, $H^i_v(\mu_0, \mu, \varphi)$) are calculated using the monochromatic RT at the band center wavelengths $\lambda^\text{eff}_\text{Band}$, and with the optical thickness of the in-band gaseous absorption:

$$\tau^g_{\lambda\Delta} = -\ln\{\int S_\lambda \exp\{-\tau^g(\lambda)\} h_\lambda d\lambda \int S_\lambda h_\lambda d\lambda \} .$$

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\textsubscript{4}, NO\textsubscript{2}, CO, N\textsubscript{2}O) 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 $\lambda^\text{eff}_C$, $\tau^R$, $\tau^g$ and column absorption optical thickness of water vapor $\tau^w$ calculated for MODIS TERRA land bands (B1-B7) are shown in Table 4.1.

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. This allows retrievals with various aerosol 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 a single scattering part of the 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 low values of the aerosol single scattering albedo, 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., in preparation).

<table>
<thead>
<tr>
<th>$\lambda^{\text{eff}}$, $\mu$m</th>
<th>0.6449</th>
<th>0.8556</th>
<th>0.4655</th>
<th>0.5535</th>
<th>1.2419</th>
<th>1.6290</th>
<th>2.1131</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau^R$</td>
<td>0.05086</td>
<td>0.01622</td>
<td>0.19258</td>
<td>0.09474</td>
<td>0.00362</td>
<td>0.00122</td>
<td>0.00043</td>
</tr>
<tr>
<td>$\tau^g$, $W=0$</td>
<td>1.32e-3</td>
<td>1.82e-5</td>
<td>2.06e-3</td>
<td>4.60e-4</td>
<td>2.90e-3</td>
<td>1.01e-2</td>
<td>1.80e-2</td>
</tr>
<tr>
<td>$\tau^w$, 1 cm</td>
<td>3.62e-3</td>
<td>5.49e-3</td>
<td>5.20e-5</td>
<td>3.63e-4</td>
<td>3.62e-3</td>
<td>0.87e-3</td>
<td>1.63e-2</td>
</tr>
<tr>
<td>$\tau^w$, 3 cm</td>
<td>1.03e-2</td>
<td>1.32e-2</td>
<td>1.60e-4</td>
<td>1.08e-3</td>
<td>1.00e-2</td>
<td>2.74e-3</td>
<td>4.23e-2</td>
</tr>
<tr>
<td>$\tau^w$, 5 cm</td>
<td>1.63e-2</td>
<td>2.00e-2</td>
<td>2.60e-4</td>
<td>1.80e-3</td>
<td>1.58e-2</td>
<td>4.83e-3</td>
<td>6.45e-2</td>
</tr>
</tbody>
</table>

Table 4.1. 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.

### 4.3 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% (Lyapustin et al., in preparation).

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. $\tau^R$ rapidly decreases with wavelength approximately as $\lambda^{-4.09}$ (see Table 4.1). For bands B5-B7, the Rayleigh optical thickness is very low, and change of $\tau^R$ with surface pressure is negligible. For these bands, the surface pressure correction is not performed.

### 4.3.1 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:

\[
\mu \frac{\partial I^m}{\partial z} = -\alpha I^m + \frac{\sigma}{4\pi} \int \{ I^m(s') + I^1(s') \} \chi(\gamma) ds'.
\] (4.31)

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 \( I^m + \delta I^m \):

\[
\mu \frac{\partial (I^m + \delta I^m)}{\partial z} = -(\alpha + \delta \alpha)(I^m + \delta I^m) + \frac{\sigma}{4\pi} \int (I^m + I^1 + \delta I^m + \delta I^1) \chi(\gamma) ds'.
\] (4.32)

Subtracting equation (4.31) from (4.32) and dividing the result by \( \alpha \), we obtain equation for the variation of multiple scattering radiance:

\[
\mu \frac{\partial (\delta I^m)}{\partial z^\delta} = -\delta \alpha I^m + \left( \frac{\omega}{\alpha} \right) \int \{ \delta I^m + \delta I^1 \} \chi(\gamma) ds',
\] (4.33)

where \( z^\delta = \int \alpha^\delta(z) dz \) 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. (4.33) turns into an ordinary differential equation

\[
\mu \frac{\partial (\delta I^m)}{\partial z^\delta} + \delta I^m(1 - \omega^\delta) = \omega^\delta \delta I^1(1 - \omega^\delta) - \frac{\delta \alpha}{\alpha^\delta} I^m(s). \] (4.34)

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

\[
\{ \delta I^m e^\mu \}_{i+1} - \{ \delta I^m e^\mu \}_{i} = \frac{1}{\mu} \int {\tau^i}^{(1 - \omega^\delta)} \left( \omega^\delta \delta I^1 - \frac{\delta \alpha}{\alpha^\delta} I^m \right) e^\mu d\tau'.
\] (4.35)

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

\[
\delta I^m_{i+1} e^\mu = \delta I^m_i e^\mu = \beta_i J_0 + \gamma_i J_1, \]
(4.36)

\[
\delta I^m_0 = 0, \mu > 0 \text{ (TOA, downward directions)},
\]
(4.36a)

\[
\delta I^m_N = 0, \mu < 0 \text{ (BOA, upward directions)},
\]
(4.36b)

where

\[
\beta_i = \omega_i^\delta [\delta I^1_i - g_i \tau^\delta_i] - \left( \frac{\delta \alpha}{\alpha^\delta} \right)_i [I^m_i - e_i \tau^\delta_i]; \quad \gamma_i = \omega_i^\delta g_i - \left( \frac{\delta \alpha}{\alpha^\delta} \right)_i e_i;
\]
\[ J_0 = \frac{1}{\mu} \int_{\tau_0^{(1-a^\delta)}}^{\tau} e^{-\mu} \, d\tau'; \quad J_1 = \frac{1}{\mu} \int_{\tau_0^{(1-a^\delta)}}^{\tau} \tau e^{-\mu} \, d\tau'. \]

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

\[ \delta I^m = \beta \frac{1}{\omega^\delta} \frac{\mu J_0 / \tau^\delta}{1-\omega^\delta} (\mu<0), \quad \beta = \omega^\delta \delta I^1 - \left( \frac{\delta \omega}{\omega^\delta} \right) I^m. \quad (4.37) \]

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. (4.28). The results of accuracy analysis for a typical continental aerosol model and a dust model from AERONET \cite{Holben et al., 1998} classification \cite{Dubovik et al., 2002} are shown in Figure 4.1. The continental aerosol is represented by the urban low absorption model for the Goddard Space Flight Center, USA (GSFC), and the dust is described by the model for the Solar Village, Saudi Arabia. The column water vapor in calculations 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.

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 \( \mu m \) 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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4_1.png}
\caption{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 \( (\tau_B^{a_B}=0.18, \tau_B^{a_B}=0.03) \), and b) dust (Solar Village AERONET model, Saudi Arabia), medium optical thickness \( (\tau_B^{a_B}=0.53, \tau_B^{a_B}=0.53) \).}
\end{figure}
relative azimuth 0°. Figure 4.1 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.

4.3.2 Surface Pressure/Height Correction

The MAIAC LUT was calculated for the standard atmospheric pressure $P_0=1013$ mB and $W=0.5$ 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 $\lambda_c$ could be achieved by sliding the wavelength from the band center in order to agree with the pressure-reduced Rayleigh optical depth:

$$\tau^R(\lambda) = \frac{P(z)}{P_0} \tau^R(\lambda_c).$$  \hspace{1cm} (4.38)

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. ≈2.5% for $P(z)=900$ mB, or ≈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. A necessary requirement for using this method is that the two neighbor bands should have a low and similar gaseous absorption. Table 4.1 shows that at zero water vapor the absorption is similar in MODIS bands B1-B4 ($\tau^g \sim 10^{-3}$). This method, originally suggested by R. Fraser, is used by MODIS aerosol algorithm (L. Remer, personal communication).

4.3.3 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 functions $h_f(\lambda)$, $h_c(\lambda)$, explained below, in the LUT.

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

$$R^{DS} = \sum f_i R^{DS}_i ,$$  \hspace{1cm} (4.39)

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

$$\tau^a(\lambda) = \tau^a_f(\lambda) + \tau^a_c(\lambda) = C_{vf} h_f(\lambda) + C_{vc} h_c(\lambda) = C_{vf} (h_f(\lambda) + \eta h_c(\lambda)) ,$$  \hspace{1cm} (4.40)

where $\eta = C_{vc} / C_{vf}$ is spectrally-independent ratio of volumetric concentrations of coarse and fine fractions, and $h_i(\lambda) = \tau^a_i(\lambda) / C_{vi}$ is a fractional optical thickness per unitary concentration. For example, $\eta=0.5$ for the GSFC aerosol model at moderate AOT. Given the size distribution and refractive index (functions $h_f(\lambda)$, $h_c(\lambda)$ are fixed), the spectral slope of AOT (Angstrom
parameter) is defined by the ratio $\eta$. Given parameter $\eta$ and total AOT, the volumetric concentration of the fine fraction is defined as $C_{ij} = \tau^a(\lambda)/(h_j(\lambda) + \eta h_c(\lambda))$. Finally, the weighting factors of the LM method do not depend on the total AOT,

$$f_j(\lambda) = C_{ij} h_j(\lambda) / \tau^a(\lambda) = h_j(\lambda)/(h_j(\lambda) + \eta h_c(\lambda)),$$

$$f_c(\lambda) = 1 - f_j(\lambda).$$

(4.41)

### 4.3.4 Calculating TOA Reflectance from LUT given Surface Pressure and Water Vapor

The algorithm consists of several steps:

1) Calculate $R^{D_{ss}}$ analytically for a given geometry, AOT, surface pressure, and column water vapor for selected fine and coarse aerosol fractions, and add them using LMM (4.39) for a given ratio of volumetric concentrations $\eta$.

2.1) Calculate pressure-shifted effective center wavelength of the MODIS channel:

$$\lambda = \lambda_i (P/P_0)^{1/4.09}, \lambda_1 \leq \lambda \leq \lambda_2;$$

2.2) Calculate LUT functions ($R^{D_{ms}}$, $D_X^1$, $D_X^3$, $G^av$, $G_X^1$, $G_X^1$, $H_X^1$, $E_0^d$, $E_0$, $c_0$) for the required AOT for two wavelengths $\lambda_1$ and $\lambda_2$ for the fine and coarse aerosol fractions.

2.3) Perform pressure correction for all LUT functions by linear spectral interpolation of LUT-functions between $\lambda_1$ and $\lambda_2$ to $\lambda$.

3) Following MISR algorithms, calculate multiple-scattering path reflectance using a modified LMM:

$$R^{D_{ms}} = R^{D_R} + \sum \frac{\omega_{mix}}{\omega_i} e^{-\tau^m[\omega_{mix}]} f_i (R^{D_{ms}^i} - R^{D_{ms}^R}),$$

(4.42)

where $R^{D_{ms}}$ 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. The MLM algorithm retains high accuracy for the high AOT and highly absorbing aerosol fractions, where the standard LM method breaks down (Abdou et al., 1997).

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

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

(4.43)

and a variation of single scattering path radiance

$$\delta I^{ss} = I^m(\lambda_i; P, W_0) - I^m(\lambda_i; P, W);$$

(4.44)

4.2) correction term $\delta I(\lambda_i; P, W_0, W)$ using formula (4.37). The corrected path reflectance is:

$$R^D(\lambda_i; P, W) = R^{D_{ms}}(\lambda_i; P, W) + R^{Dms}(\lambda_i; P, W_0) + \delta I(\lambda_i; P, W_0, W) / \mu_0.$$

(4.45)

4.3) Finally, the TOA reflectance is a sum of path reflectance (4.45), and a pressure-corrected surface-reflected signal $R_s$ (sum of terms (4.8) and (4.9)), additionally corrected for the difference in the gaseous absorption via the two-way direct transmittance:
\[ R(\lambda_1) = R^{D}(\lambda_1, P, W) + R_s(\lambda_1, P, W_0) \exp(-\{\tau^W - \tau^{W_0} - (1 - P / P_0)\tau^S\}m) \]  \hspace{1cm} (4.46)

where \( m = m_0^{-1} + |\mu|^{-1} \) is an atmospheric air mass.
5. Water Vapor Retrieval Algorithm

Selected water vapor algorithm is a two-channel ratio algorithm based on the LUTs. The column water vapor (CWV) 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.

5.1 Theoretical Basis

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

\[ R_\lambda (\mu_0, \mu, \varphi) \approx R_\lambda^D (\mu_0, \mu, \varphi) + \frac{T_\lambda^\text{tot} (\mu_0, \mu)}{1 - q_\lambda (\mu_0)} \rho_\lambda (\mu_0, \mu, \varphi), \tag{5.1} \]

where \( R_\lambda^D \) is a path reflectance, \( T_\lambda^\text{tot} (\mu_0, \mu) = (T^+(\mu_0)T^+(\mu))_\lambda \) 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 \( R_\lambda^D \) and \( T_\lambda^\text{tot} (\mu_0, \mu) \) are spectrally integrated with the specific RSR of a given channel as described in sec. 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 \( \mu \text{m} \) 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:

\[ R_\lambda (\mu_0, \mu, \varphi) \approx R_\lambda^D (\mu_0, \mu, \varphi) + T_\lambda^\text{tot} (\mu_0, \mu) \rho_\lambda (\mu_0, \mu, \varphi). \tag{5.2} \]

If the surface reflectance is spectrally flat or changes little in a narrow spectral absorption interval of 0.9-0.94 \( \mu \text{m} \), then a two-channel ratio algorithm can be used to derive CWV:

\[ \frac{T_{18}^\text{tot}}{T_{19}^\text{tot}} = \frac{R_{18} - R_{18}^D}{R_{19} - R_{19}^D}, \quad \text{and} \quad \frac{T_{19}^\text{tot}}{T_{17}^\text{tot}} = \frac{R_{19} - R_{19}^D}{R_{17} - R_{17}^D}. \tag{5.3} \]

For a given view geometry, the CWV 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 CWV from the two ratios (Eq. 3) can be different. A mean water vapor (W) is obtained as in (Gao and Kaufman, 2003):

\[ W = f_1 W_1 + f_2 W_2, \tag{5.4} \]
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,

$$\eta_1 = \left| \frac{\Delta T_{18}^{\text{tot}}}{T_{19}^{\text{tot}}} / \Delta W \right|, \quad \eta_2 = \left| \frac{\Delta T_{19}^{\text{tot}}}{T_{17}^{\text{tot}}} / \Delta W \right|,$$

and are defined as normalized values:

$$f_i = \eta_i / (\eta_1 + \eta_2). \quad (5.5)$$

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

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

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. 5.3), corresponds to the background aerosol level assumed in the LUT calculations as $\tau^a=0.01$ at $\lambda=0.94 \mu m$. 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 $\lambda=0.94 \mu m$ ($\tau^a>1$), then the water vapor retrievals should be repeated with the derived value of AOT (see sec. 5.5). This part of algorithm is not currently implemented.

5.2 Look-Up Tables

The current LUT$_W$ is calculated using a vertical profile of the US 1976 Standard Model of Atmosphere. Given the CWV, 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 gives an adequate CWV accuracy for the problem of atmospheric correction of MODIS spectral bands, carefully selected in the atmospheric windows.

The LUT$_W$ stores path reflectance $R_2^{\text{D}} (\mu_0, \mu, \varphi, W, \tau^a)$ for three bands (B17-B19), and two transmittance ratios $T_{18}^{\text{tot}} / T_{19}^{\text{tot}}$, $T_{19}^{\text{tot}} / T_{17}^{\text{tot}}$, which depend on 4 parameters ($\mu_0, \mu, W, \tau^a$). The LUT$_W$ was calculated with steps $\Delta \mu_0 = \Delta \mu = 0.02$ for the range $\mu=1 - 0.4 (0^o - 66.4^o)$, $\mu_0=1 - 0.34 (0^o - 70.1^o)$, CWV=0 - 7.5 with $\Delta W=0.3$, and $\tau^a_{0.94} = \{0.01, 0.2, 0.5, 1.0\}$. The step in azimuthal angle for the path reflectance is selected $10^o$. The angular resolution of LUT$_W$ is high enough to use nearest neighbor method, avoid bi- or three-linear interpolation in angles.

5.3 AERONET Validation
To test developed algorithm, we processed 1 year (2003) of MODIS TERRA data subsetted for 156 AERONET stations globally. The 50×50 km² subsets of data were provided by the MODAPS (courtesy of Dr. Saleos).

The new AERONET ver. 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). The new algorithm was recently validated against GPS retrievals. The comparison study showed an excellent agreement with zero intercept and a slope of 0.984 in the range of CWV 0 – 5 cm (Smirnov et al., 2004).

For the current study, we used AERONET CWV data within ±30 min of the MODIS TERRA overpass for validation. The retrieved data (CWV₁) were averaged over 9×9 km² region. Because CWV is retrieved before CM algorithm, the cloud contamination was filtered first using a very simple variance criterion: the point was considered potentially cloudy if the difference between the maximum and the average CWV₁ over the 9×9 km² 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 scattershots of the retrieved water vapor against AERONET results are shown in Figure 5.1 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 within 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 5.1 show that a fraction of values CWV₁ 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 CWV₁. Unless the surface is very bright, the aerosol scattering usually increases the measured signal and the band ratio, also causing lower retrieved CWV. Figure 5.2 shows the time series of the retrieved CWV, AERONET CWV, and AERONET aerosol optical thickness at λ=1.06 µm. The correlation between the CWV₁ and AERONET data 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 the retrieved CWV was significantly lower than the AERONET value. These data show the need for aerosol correction of the CWV retrievals on hazy days. The operational MODIS NIR CWV algorithm MOD03 did not implement this correction (Gao and Kaufman, 2003).

5.4 Alternative Algorithms
We have also studied two additional algorithms - an operational MODIS near-IR water vapor algorithm (MOD03) (Gao and Kaufman, 2003), 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 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’s results have 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 CWV retrieval method used in MOD09 uses an empirical relation:

$$CWV = (a_1 \ln R + a_2 \ln^2 R) / m,$$  \hspace{1cm} (5.6)

where $\ln R = \ln(R_{18}/R_{10})$, $m$ is airmass, and $a_1$, $a_2$ are band-dependent coefficients. This relation was used earlier for CWV retrievals from space (Bennartz and Fisher, 2001). Another empirical relation commonly used in CWV retrievals from ground-based sunphotometer measurements (Michalsky et al., 1995) has a form:

$$T_{WV} / T_{Window} = \exp(-a[m_{WV}CWV]^b),$$  \hspace{1cm} (5.7)

where the left-hand side is the direct transmittance ratio in the WV 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 (5.6) and (5.7) for the CWV retrievals after determining the model coefficients. 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 5.3, where the ratio of LUT transmittance is called $ETratio$, and Model 1 and Model 2 represent equations (5.7) and (5.6), respectively. The CWV increases along the x-axes from 0.05 to 7.5 with step 0.3. At each CWV value, the zenith sun and view angles change from $0^\circ$ to $60^\circ$. The transmittance ratio and the model error resemble the comb-like structure. Each period of this structure corresponds to a fixed value of CWV and all combinations of SZA and VZA, with maximal error at SZA=VZA=60°.

Figure 5.3 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. Our comparison of the algorithm (Eq. 5.6) with AERONET data showed that it works remarkably well, providing similar results with our selected approach, only with slightly lower correlation with AERONET data.
Figure 5.1. Comparison of retrieved column water vapor with AERONET water vapor data.

Figure 5.2. Time series of retrieved CWV, AERONET CWV, and AERONET optical thickness at 1.06 μm. The arrows mark the days with high AOT when the retrieved CWV is significantly lower than the AERONET data.
Figure 5.3. 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 CWV 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°.
6. MAIAC CLOUD MASK ALGORITHM

6.1 Introduction

The quality of cloud clearing significantly impacts precision and accuracy of aerosol and surface reflectance records and consistency of their time series. The heritage cloud mask (CM) algorithms for the low-orbit sensors, including the AVHRR CLAVR (McClain, 1993) and MODIS algorithm (Ackerman et al., 2006), use the current measurements of spectral reflectance and brightness temperature and perform processing at the pixel level. The ISCCP CM algorithm (Rossow and Garder, 1993) developed for geostationary platforms, builds the clear-skies composite map from the previous measurements and infers CM for every pixel by comparing current measurement with the clear-skies reference value. The internal CM algorithm used in the MODIS atmospheric correction algorithm (MOD09) is a pixel-level empirical algorithm whose performance has not been documented.

The distinct feature of the new algorithm is utilizing both the previous measurements and an image-based method of analysis combined with pixel-level processing. The clear-skies images of the same surface area have a common textural pattern, defined by the surface topography, boundaries of rivers and lakes, distribution of soils and vegetation etc. This pattern changes slowly as compared to the daily rate of global Earth observations. Clouds disturb this pattern, which can be easily detected by a 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,

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

A high covariance of two images usually implies cloud-free conditions in both images, whereas low covariance usually indicates presence of clouds at least in one of the images. A rapid surface change or significant variation of aerosol density in the area may also reduce covariance. Because covariance removes the average values of signals, this metric is equally successful over the dark and bright surfaces. It works in both clear and hazy conditions if the surface variability is still detectable from space. The MAIAC approach, based on the time series of gridded MODIS measurements, inherently benefits from using temporal dimension of measurements and image texture.

As discussed earlier, MAIAC algorithm first grids MODIS data and splits them into 600 km Tiles. The further processing uses either individual pixels or fixed-size (25×25 km$^2$) areas (blocks) for the covariance analysis. For each block, spectral tests are performed first to detect clear vegetated, snow or water pixels. If successful, these pixels are used to calculate an average brightness temperature of the ground and of the inland water of the block ($BT_G, BT_W$). The core of the new CM algorithm is initialization and regular update of the reference clear-skies image ($refcm$) for every block. The reference image is initially built from the pair of images for which covariance is high, and caution is exercised to exclude correlated cloudy fields. After initialization, the algorithm uses $refcm$ to compute covariance with the latest measurements. If covariance is high, $refcm$ is updated. If covariance is low, then for each pixel the algorithm compares i) measured reflectance with $refcm$ value, and ii) measured brightness temperature with $BT_G$ to separate clouds which are usually somewhat brighter and colder. In the world regions with sparse or no vegetation, the ground brightness temperature comes from the neighbor blocks declared clear by the covariance analysis.
MAIAC CM algorithm implements a relatively simple and straightforward logic. Because of covariance component, the algorithm works successfully over both dark and bright surfaces, including deserts and snow. Dynamic update of the reference clear skies image allows CM algorithm to smoothly adjust to the seasonal surface variability. On the other hand, rapid surface change events (fire burns, snow fall/ablation, floods etc) are handled through repetitive re-initialization of refcm. Keeping memory of the clear-skies image and of the essential statistical properties of reflectance and brightness temperature for every land surface block strongly enhances probability of the correct cloud detection in difficult cases. The details of the algorithm are given below.

6.2. Algorithm Technical Detail

1. The covariance analysis is currently performed (refcm is maintained) for MODIS band 1 (0.67 µm). We have extensively studied use of B6 (1.64 µm) and B5 (1.21 µm). B6 was initially our band of choice because 1) it has a very low molecular absorption and usually low aerosol effect, and 2) the Landsat extensive experience shows that the spectral region of 1.6 µm has the most land surface variability in the visible-SWIR spectrum. We have conducted an independent covariance analysis for 1 year of MODIS data and 160 AERONET locations globally. From considered MODIS bands B2 (0.87 µm), B5, B6 and B7 (2.13 µm), band 6 provided by about 10% more of the high covariance cases. On the other hand, this band has not been working properly on MODIS AQUA, and it has developed occasional problems since 2006 on MODIS TERRA. We have also found that bands 5 and 6 often cannot detect significant spatial aerosol variability or variable semi-transparent clouds. Spatial uniformity of aerosols within the block area is one of the major requirements of MAIAC inversion method, and B1 was found to provide a better overall performance.

2. The size of block is currently selected as 25×25 pixels (km²) for two reasons. First, this size is large enough to capture a variety of spatial variability scales (geological, topographic, ecological etc.) required for covariance analysis. Second, it is sufficiently large to capture surface variability at the edge of scan where the MODIS pixel size grows to ≈2×4 km² for 1 km² nadir pixels. On the other hand, success rate of covariance algorithm to select clear blocks in conditions of broken cloudiness is higher for smaller blocks. The MISR CM algorithm, which also uses covariance analysis, uses the block size of 17.6 km. We plan to evaluate the global performance of cloud mask for the variable block size 25-15 km and select the optimal one for operational application.

3. In order to account for the effects related to the scan angle variation, e.g. pixel size growth, surface BRF or reduction of contrast at higher VZA, two reference clear-skies images are maintained by the algorithm, refcm1 for µ>0.7 (≈0-45°) and refcm2 for µ≤0.7 (45°-60°).

4. In addition to the clear-skies image in band 1 (refcm), which is stored at the Q-level, the queue also keeps the maximal reflectance r1 max and a brightness temperature contrast for a given block, ΔBT=BT max−BT min. We found from MODIS data that the BT contrast is a rather stable metric of a given land block-area in clear conditions. For the pure land blocks (no water or snow pixels), the BT contrast is usually low (3-6 K) for flat terrains. It may increase, sometimes significantly (20-25 K) when the block has a mixture of land and water or snow pixels. In partially clear/cloudy conditions the BT contrast increases because BT min is usually lower over clouds. Use of these two parameters, q.r1 max and q.ΔBT, allows an effective block-level control of processing.
The brightness temperature is calculated using MODIS band 31 (11 µm).

5. The algorithm keeps a two-level cloud mask, the standard mask at the grid (1 km) resolution (CM), and another one at the block resolution (CM_COV). The CM_COV mask is used to efficiently control the algorithm flow.

6. The allowed values of cloud mask are clear (CM_CLEAR), possibly clear (CM_PCLEAR), possibly cloudy (CM_PCLOUD), and cloudy (CM_CLOUD). Two more values of cloud mask are CM_SHADOW for pixels defined as a cloud shadow, and CM_GREY representing 1 dilated pixel on cloud edges.

### 6.3 Algorithm Flowcharts

The general flowchart of CM algorithm is shown in Figure 6.1. Here, rectangles represent separate functions, diamond shapes stand for the separate subroutines (algorithms), and round-corner rectangles indicate decision (branching) points. The thick arrows show the points of exit.

1. As a first step, a Cold cloud test is performed for each pixel of the Tile:
   
   \[ \text{IF } BT(i,j) < 243K + dT(h) \rightarrow \text{CM_CLOUD}, \]
   
   where \( dT(h)=0.0045*h \) is an altitude correction factor for an average lapse rate, \( h \) is surface height above the sea level in meters. The threshold value for this test will be adjusted as a function of latitude in operational application for the northern regions of North America and Eurasia, Arctic and Antarctic.

Next, spectral tests are performed for each block to detect clear vegetated, water and snow pixels (Sec. 2.1). These tests are used to initialize a pixel-level land-water-snow mask, which is used in the cloud mask algorithm and in the aerosol-surface retrieval algorithm. The clear pixels detected in these tests are used to calculate an average brightness temperature of the ground and of the water for the block \( (BT_G, BT_W) \), which is required by the covariance part of CM algorithm.

2. Further processing path depends on whether \( refcm \) has been initialized. If not, then the algorithm tries to initialize \( refcm \) (module \( \text{initRefcm}() \)). During initialization, the algorithm consecutively computes covariance of the last measurements with each of the earlier images of the queue, until the high covariance is found and clear conditions in both images are confirmed. The initialization takes 2-4 clear-skies images, and needs to be done once at the beginning of processing/re-processing, or after a rapid surface change event (e.g. snowfall). If initialization is unsuccessful because of clouds, the algorithm runs a backup pixel-level CM algorithm \( \text{cloudMask1}() \).

If \( refcm \) has been initialized earlier, then the algorithm calculates covariance of the new measurements with \( refcm \), and performs further processing depending on the value of covariance (modules \( \text{CM_highCov}() \) and \( \text{CM_lowCov}() \)). The algorithm pursues a conservative strategy admitting that sometimes partially or completely cloudy images can leak through filters of \( \text{initRefcm}() \), in which case they are used to update \( refcm \). After that, a new clear image will not correlate well with \( refcm \). To reduce possible errors, our algorithm attempts to re-initialize \( refcm \) with latest measurements each time when the calculated covariance with existing \( refcm \) is found low. This approach also helps to re-initialize \( refcm \) when the surface reflectance rapidly changed (e.g. snowfall).
Each of modules $CM_{\text{highCov}}$, $CM_{\text{lowCov}}$, $cloudMask1$ is producing CM for the new tile. Module $initRefcm$ produces CM only if it finds high covariance. The $refcm$ and the necessary statistical parameters of B1 reflectance and BT contrast for every block are updated in the modules $initRefcm$ and $CM_{\text{highCov}}$. The algorithms implemented in the modules $initRefcm$, $CM_{\text{highCov}}$, $CM_{\text{lowCov}}$, $cloudMask1$ are described in sec. 6.3.2-6.3.5.

![Flowchart](image.png)

**Figure 6.1.** The general flowchart of CM algorithm. After spectral tests, the algorithm first tries to initialize the reference clear-skies image ($refcm$), and if fails then it uses a backup cloud mask algorithm (module $cloudMask1$). If $refcm$ is available, the CM algorithm calculates covariance between the latest image and $refcm$, and carries on further analysis depending on whether covariance is high or low.

### 6.3.1 Land-Water-Snow Mask

The MALAC algorithm maintains a dynamic internal Q-level land-water-snow mask ($LWS\_mask$) which governs cloud clearing algorithm and controls selection of the surface BRF model during aerosol-surface reflectance retrievals. The $LWS\_mask$ also helps processing algorithms to adjust to surface changes, such as snow fall/ablation, flooding etc. It has three stable values (MASK\_LAND, MASK\_WATER, MASK\_SNOW) and three transitional values used when surface change is detected (MASK\_TO\_LAND, MASK\_TO\_WATER, MASK\_TO\_SNOW). For example, value MASK\_TO\_LAND represents transition from snow or water to land. In order to handle surface change, the algorithm uses another mask indicating stability of state ($mask\_Change$) which has values of MASK\_STABLE and MASK\_CHANGE.
When the new Tile is received, three spectral tests are used to detect clear pixels and validate or change the LWS_mask:

1. High NDVI test: \[ \text{IF } NDVI = \frac{(r2 - r1)}{(r2 + r1)} > 0.6 \rightarrow LY\_CM = CM\_CLEAR. \]

   If the q-level previous Mask value was MASK\_LAND or MASK\_TO\_LAND, then the value of mask is validated (q.LWS\_mask = MASK\_LAND, q.mask\_Change = MASK\_STABLE). Otherwise, change is detected: q.LWS\_mask = MASK\_TO\_LAND, q.mask\_Change = MASK\_CHANGE.

2. Water test: \[ \text{IF } r2 < 0.07 \text{ AND } r5 < 0.02 \text{ AND } r7 < 0.020 \text{ AND } NDVI \leq 0.2 \rightarrow LY\_CM = CM\_CLEAR. \]

   Test 2 is conducted only for the off-glint geometries, which are defined according to a condition \( r_{CM} < 0.02 \), where \( r_{CM} \) is a Cox-Munk glint reflectance for the wind-ruffled water surface calculated at wind speed of 7 m/s. Theoretical reflectance is pre-calculated using Nakajima and Tanaka (1983) model with the mutual shadowing of waves, and stored in the LUT. The algorithm can use a real time wind speed, if it becomes known operationally from other sources.

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

   In the sun-glint conditions (\( r_{CM} \geq 0.02 \)) the water test is not used. In this case, the algorithm assumes that the state of water pixels (q.LWS\_mask = MASK\_TO\_WATER or MASK\_WATER) has not changed, and cloud mask is established as follows:

   \[ \text{IF } r5 < r_{CM} + 0.03 \rightarrow LY\_CM = CM\_CLEAR, \text{ ELSE } LY\_CM = CM\_CLOUD. \]

   The margin of 0.03 may be increased in the future to allow full-range aerosol retrievals over the in-land water and coastal regions.

3. Snow test: \[ \text{IF } NDSI = \frac{(r4 - r6)}{(r4 + r6)} \geq 0.5 \text{ AND } r1 > 0.20 \text{ AND } BT < 280K \rightarrow \text{ snow or cloud detected.} \]

   This test usually successfully detects pixels fully or partially covered by snow. Infrequently, clouds also pass this test. Assuming that this happens in a random fashion, the value of cloud mask is set to CM\_CLEAR and MASK\_SNOW is validated only if snow had already been detected for this pixel (q.LWS\_mask = MASK\_TO\_SNOW or MASK\_SNOW). If the snow signature was detected for a given pixel for the first time, the q.LWS\_mask value is changed to reflect detected surface change (q.LWS\_mask = MASK\_TO\_SNOW, mask\_Change = MASK\_CHANGE), but the value of cloud mask is set to CM\_CLOUD to avoid possibility of cloud leaks.

   This test needs further development to better account for clouds with spectral signature of snow.

6.3.2 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 (Figure 6.2) calculates a block-level covariance between the new Tile and the previous Tiles stored 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. Prior to cov-analysis, several tests are used in order to exclude possibly cloudy scenes (days) from cov-analysis:

1) On the initialization stage, the algorithm skips the Tile if it was acquired within less than 3h from the new Tile. This restriction is designed to exclude high correlation of cloudy or partially cloudy scenes on stagnant (windless) days.

2) On the re-initialization stage, when the clear-skies parameters (q.R1,max, q.ΔBT) of the block are available, the following filters are used:

   IF ΔBT>q.ΔBT+10K OR r1max<q.R1max+0.1 → skip day.

Next, the algorithm calculates covariance between the new Tile and Tile(day). If HIGH covariance (>0.68) is found between the new day and day, then the algorithm checks the brightness temperature contrast for this block for both days. This test has two thresholds:

1) The high threshold Δ2=7 K for pure water blocks, Δ2=25+dT(h) K for pure land blocks, and currently Δ2=25+dT(h) K for mixed land-water, land-snow, or land-water-snow blocks. If the measured BT contrast for a given day exceeds Δ2, then this day is no longer used, and the algorithm moves to another day in the queue.

2) The low threshold Δ1=q.ΔBT+5K or 15K if q.ΔBT is not initialized. If measured BT contrast for the day or the new day is lower than Δ1, then:

---

**Figure 6.2.** Module initRefcm() of MAIAC CM algorithm.
a) The block is declared CM_CLEAR at the block and at the pixel level;

b) refcm is updated (the B1 image and parameters q,R1max, q,ΔBT). If both day and new day are found cloud-free, then the update preference is given to the latest measurement (new day).

The update for image and maximal reflectance is performed with relaxation, for example q,R1max = (R1max + q,R1max)/2. This softens the effect of possible errors due to residual clouds.

The lowest value of BT contrast for the block is limited to 3.5K. In clear-skies conditions on some days this value is often lower for flat terrains, 1-2K, however the long-term variability is usually higher, 3-5K. Without this restriction, CM algorithm may choose incorrect path of processing on the following clear-skies days with higher BT contrast.

c) The ground brightness temperature is calculated for the new day as the minimum value for the 95% hottest pixels. If BTG was initialized earlier during Spectral tests, then the higher value of the two is selected as a final BTG for a given block.

d) The algorithm exits the day-loop (over the queue).

If measured BT contrast for the day or the new day exceeds Δ1 but it is below Δ2 and covariance is HIGH, then we assume that a small fraction of pixels in the block is cloudy. In this case the algorithm performs a histogram-based analysis for the B1 reflectance and BT, which is a heuristic cloud search, looking for pixels that are simultaneously brighter and colder than the clear surface. It calculates separately the maximal B1 reflectance and minimal BT for 90% of land pixels of the block (r1max, BTmax) and assigns CM_CLOUD values to the pixels which are simultaneously brighter and colder than the found thresholds. The rest of pixels are assigned CM_CLEAR values. The block-level LY_CM_COV value is set to a possibly clear value (CM_PCLEAR).

6.3.3 Module CM_highCov()

This module (Figure 6.3) is called when refcm is initialized and covariance between refcm and new Tile for a given block is HIGH (cov > 0.68). High covariance automatically means clear-skies conditions although a few pixels may still be cloudy.

The algorithm first compares the measured BT contrast for the block with the threshold. If the BT contrast is lower than threshold then the block is declared CM_CLEAR, the ground BT is calculated, and refcm is updated. Otherwise, the algorithm performs a pixel-based analysis comparing measured B1 reflectance with refcm reflectance and measured BT with the ground BT (see Fig.3). If BTG is undefined, it is first evaluated as a minimal value for 90% of the hottest pixels of the block.

6.3.4 Module CM_lowCov()

This module is called when covariance is low (cov≤0.68). Initially, it evaluates BTG for a given block if it is undefined. BTG is computed as an average value of BTG of neighbor blocks for which it was defined either by Spectral Tests (clear vegetated pixels) or by the module CM_highCov(). Next, the algorithm performs a pixel-based analysis comparing measured B1
reflectance with $ref\text{cm}$ reflectance and measured BT with the ground BT if it is available (Fig. 6.4).

Refcm has been initialized and $\text{cov}(ref\text{cm}, \text{last Tile})$ is HIGH.
The assumption below is that this block of \text{last Tile} is mostly clear with possibly few cloudy pixels.

**Figure 6.3.** Module $CM_{\text{highCov}}()$. If the brightness temperature contrast is below expected threshold then the block is declared CM\_CLEAR, the ground BT is calculated, and $ref\text{cm}$ is updated. Otherwise, the algorithm performs a pixel-based analysis comparing measured B1 reflectance with $ref\text{cm}$ reflectance and measured BT with the ground BT.

<table>
<thead>
<tr>
<th>$\Delta\text{BT}&gt;q$, $\Delta\text{BT}+3K$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Figure 6.4.** Module $CM_{\text{lowCov}}()$. The algorithm filters cold pixels and performs a pixel-based analysis comparing measured B1 reflectance with $ref\text{cm}$ reflectance and measured BT with the ground BT if it is available.

**6.3.5 Module cloud\text{Mask1}()**

This module is used at the very beginning of processing/re-processing when $ref\text{cm}$ has not yet been initialized, and $\text{initRefcm}()$ failed because of persistent cloud cover.
The processing (Figure 6.5) consists of two passes. In the first pass, the algorithm uses spectral flatness (SF) test (1) to find optically thick clouds (2) in the block.

Refcm has not been initialized, but B1 image could be updated by this module.

**Figure 6.5. Module cloudMask1() of CM algorithm.**

The SF test is formulated using 4 visible bands, B1, B3, B4, B8. It requires 1) that the reflectance in these bands for a given pixel is higher than 0.15 and is close to the spectrally-average value:

\[
|R1 - R_{av}| < d \quad \text{AND} \quad |R3 - R_{av}| < d \quad \text{AND} \quad |R4 - R_{av}| < d \quad \text{AND} \quad |R8 - R_{av}| < d,
\]

where \( R_{av} = (R1 + R3 + R4 + R8) / 4 \). If band 8 is saturated, only 3 bands are used. The threshold deviation is currently selected as \( d=0.025 \); 2) that the BT contrast between SF pixels (cloud BT) and the rest of pixels (ground BT) in the block exceeded a threshold of 4K:

\[ \text{BT}_G - \text{BT}_G > 4 \]

The last test checks if SF pixels are really clouds, because certain types of soils/sands also pass the SF test.

The SF test captures relatively thick clouds which can be low and warm and would not be detected by the brightness temperature test.

On the second pass, the algorithm follows the logic described earlier. First, if only a few pixels were detected as cloudy and the brightness temperature contrast in the block is low, which indicates clear conditions with possibly a few cloudy pixels, then the algorithm performs a joint histogram analysis of B1 reflectance and BT (3), and updates refcm for clear pixels (4). Second, if ground BT is available from either spectral tests or from CM_CLEAR neighbor blocks, then the algorithm performs a threshold comparison of BT and B1 reflectance with BTG and refcm B1 values (5). If no a priori information is available about the surface, then the algorithm checks B1 reflectance and BT against fixed thresholds (6). Finally, block-level cloud mask (LY_CM_COV) is defined depending on the total number of found CM_CLOUD pixels in the block.
6.4. Performance of CM Algorithm

Performance of the cloud mask algorithm has been extensively tested using 1 year of MODIS TERRA data subsetted for 156 AERONET locations worldwide by MODAPS. The data were received in the swath format with resolution aggregated to 1 km in all bands for the subset area 50×50 km². Before processing, we re-projected and gridded MODIS data to 1 km grid using the nearest neighbor method.

To gain insight on the large scale algorithm performance, we also used the 2005 MODIS TERRA data for the north-east USA, Arabian peninsular and Zambia, each covering the area of 1200×1200 km². The testing was done for at least half a year of continuous data in each case.

The analysis has being conducted by a visual comparison of generated cloud mask against RGB image of TOA reflectance and, if required, the image of brightness temperature. Several examples of cloud mask results after initialization are shown in Figures 6a-6b. Figure 6a shows results for the area centered at the Goddard Space Flight Center (Greenbelt, Maryland, USA) for five 16-days intervals for different seasons of the year. Each column contains 20 images because on some days the area was observed repeatedly from different orbits. The Julian day for the last image is shown on the top of Figure. Each image was divided in 4 blocks during processing. The reproducible spatial pattern of surface on the cloud-free days, clearly visible in these image sequences, is a key to the success of the CM algorithm.

The first column and beginning of the second column of images in Fig. 6a show the initialization stage of covariance algorithm. To make this demonstration, the CM algorithm was run without module CloudMask1(). The blocks 2-3 with largest contrast are initialized first, and block number 4 with lowest contrast is initialized last. A careful analysis of these images shows that the new algorithm at the current basic level of development already has a very high accuracy of cloud discrimination.

The cloud-free regions (blocks) are accurately identified by the covariance analysis, and performance of algorithm is robust globally regardless of the surface brightness. For example, the first and second columns of Fig. 6a show correct identification of snow. Other examples in Figure 6b (Railroad Valley, Arizona, USA and Solar Village, Saudi Arabia) show performance of CM over bright deserts. Bondville, ILL (USA), and Mongu (Zambia) are examples of vegetated or partly vegetated surfaces.

Several examples of CM performance over large areas are shown in Figures 6.7 (north-east USA), 6.8 (Zambia, Africa), and 6.9 (Arabian Peninsula). All examples show a very robust performance of CM algorithm in very different conditions, including bright surfaces (snow and deserts) and heavy aerosol loading (biomass burning and dust storms). One should bear in mind that the current algorithm has been developed for the land. Cloud detection over water at this stage is rudimentary, and presented examples contain a number of artifacts over seas and coastal ocean waters. MAIAC algorithm will be enhanced for these conditions in the near future.

These figures also show comparison with reprojected MODIS Collection 5 cloud mask MOD35. One can see that cloud masks from MAIAC and MOD35 are generally similar. MAIAC has a higher confidence in clear conditions (Fig. 6.8), which is important for land applications. It may also provide a better performance in difficult cases, as over snow (Fig. 6.7a) or in high aerosol conditions (Fig. 6.9). We plan to conduct a global comparison of the developed CM with MOD35 in cooperation with group of S. Ackerman, who developed MODIS cloud mask product.
If covariance analysis helps to select cloud-free conditions, then maintenance of dynamically updated thresholds for every pixel and block of the surface helps to successfully detect clouds in cloudy conditions. Indeed, keeping track of surface reflectance and brightness temperature contrast makes the cloud mask a rather well-defined problem when the new data arrive, as opposed to conventional pixel-level approach lacking an a priory information on the surface properties.

Our analysis of the cloud mask on global subsets of AERONET data and large-scale target areas reveals no major problems. Some of the problems, which haunt all CM algorithms including this, are the following: 1) cases when clouds systematically exhibit spectral signatures of snow, 2) 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 both problems which we plan to address in the very near future.

There are avenues for a further development of MAIAC CM algorithm. One approach would be using BRF calculated from the latest results of atmospheric correction instead of refcm image. This would substitute a current rough division of VZA range in two intervals, 0-45° and 45°-62°, by a smooth angular dependence. Functionally, our software architecture allows very easy implementation of such enhancements and following testing.

6.5. Memory Requirements

For every gridded pixel, our algorithm keeps memory of:
q.LWSmask (1 Byte), q.MaskChange (1 Byte);
B1 reflectance for two ranges of VZA (refcm1 and refcm2) (float);

For every block, the algorithm keeps memory of:
Brightness temperature contrast, q.BT_DIF (float);
Maximal B1 reflectance for two ranges of VZA (refcm1 and refcm2) (float);

6.6. Near-Future Tasks

- Develop algorithm for coastal seas and oceans, and large in-land lakes.
- Improve CM performance on the land-water boundaries.
- Improve discrimination of clouds with spectral signatures of snow.
- Perform global testing of algorithm, and extensive inter-comparison with MODIS CM product (MOD35) in collaboration with University of Wisconsin, Madison (S. Ackerman, R. Frey).
Day 29, GSFC

Day 54, GSFC

Day 192, GSFC

Day 261, GSFC

Day 311, GSFC

(a)

snow

clear

clouds

missing data

Close to edge-of-scan, high VZA

Under-estimated clouds?

Near-nadir image, Low VZA

missing data
Figure 6.6. Examples of performance of MAIAC CM algorithm for GSFC, USA (a) and several locations with different surface brightness worldwide (b). RGB TOA MODIS gridded images are shown on the left, and cloud mask is shown on the right. CM legend: clear – blue, cloudy – red, possibly cloudy – yellow.
Figure 6.7a. Example of MAIAC (top) and MOD35 (bottom) cloud mask over snow from MODIS TERRA data for days 36 (left) and 37 (right) of 2005. The image shows 1 Tile (600×600 km²) for the north-east USA.
Figure 6.7b. Example of MAIAC and MOD35 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 7a. Performance is similar.
Figure 6.8a. Example of MAIAC and MOD35 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 km²). MAIAC CM is more certain over the cloud-free areas.
Figure 6.8b. Example of MAIAC and MOD35 cloud mask in the second half of dry season with intensive biomass burning for days 222 (left) and 269 (right) of 2005. The image area is the same as in Figure 8a.
Figure 6.9. Example of MAIAC and MOD35 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 km²). Performance is similar for day 145. MOD35 overestimates CM over the dust storm areas on day 207.
7. Software Architecture

To be added.
REFERENCES


