A Spatio-Temporal Approach for Global Validation and Analysis of MODIS Aerosol Products

Charles Ichoku¹, D. Allen Chu¹, Shana Mattoo¹, Yoram J. Kaufman², Lorraine A. Remer², Didier Tanré³, Ilya Slutsker¹, and Brent N. Holben⁴

¹ Science Systems and Applications Inc., NASA/GSFC, Greenbelt, MD 20771
² Laboratory for Atmospheres, NASA/GSFC, Greenbelt, MD 20771
³ Laboratoire d'Optique Atmosphérique, Centre National de la Recherche Scientifique et Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France
⁴ Laboratory for Terrestrial Physics, NASA/GSFC, Greenbelt MD 20771

Abstract. With the launch of the MODIS sensor on the Terra spacecraft, new data sets of the global distribution and properties of aerosol are being retrieved, and need to be validated and analyzed. A system has been put in place to generate spatial statistics (mean, standard deviation, direction and rate of spatial variation, and spatial correlation coefficient) of the MODIS aerosol parameters over more than 100 validation sites spread around the globe. Corresponding statistics are also computed from temporal subsets of AERONET-derived aerosol data. The means and standard deviations of identical parameters from MODIS and AERONET are compared. Although, their means compare favorably, their standard deviations reveal some influence of surface effects on the MODIS aerosol retrievals over land, especially at low aerosol loading. The direction and rate of spatial variation from MODIS are used to study the spatial distribution of aerosols at various locations either individually or comparatively. This paper introduces the methodology for generating and analyzing the data sets used by the two MODIS aerosol validation papers in this issue.

1. Introduction

Aerosol physical parameters are among the numerous products operationally retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Earth Observing System's (EOS) Terra satellite launched on 18 December 1999. The MODIS aerosol algorithm routinely retrieves aerosol optical thickness (AOT or \( \tau_a \)) at 0.47 and 0.66 \( \mu m \) wavelengths (and interpolates at 0.55 \( \mu m \)) over vegetated land surfaces, and at seven spectral bands (0.47, 0.55, 0.66, 0.87, 1.2, 1.6, and 2.1 \( \mu m \)) over oceans (Kaufman et al., 1997; Tanré et al., 1997). Other essential parameters retrieved include Angstrom exponent over land and ocean, and effective radius and ratio of small to large size modes over ocean (Chu et al., 2001; Remer et al., 2001).

MODIS aerosol products are validated with ground-based sunphotometer data, particularly those of the Aerosol Robotic Network (AERONET), comprising automatic tracking Sun photometers/sky radiometers located at over 100 sites around the world (Holben et al., 1998). AERONET sunphotometers derive AOT at 0.34, 0.38, 0.44, 0.50, 0.67, 0.87, and 1.02 \( \mu m \) wavelengths from direct solar radiation measurements. These data (http://aeronet.gsfc.nasa.gov/) undergo cloud screening (Level 1.5) and quality assurance (Level 2.0) (Smirnov et al., 2000), although for near real-time data availability, we use the Level 1.5 data for our validation. AERONET data are very
A Spatio-Temporal Approach for Global Validation and Analysis of MODIS Aerosol Products

by

Charles Ichoku¹, D. Allen Chu¹, Shana Mattoo¹, Yoram J. Kaufman², Lorraine A. Remer², Didier Tanré³, Ilya Slutsker¹, and Brent N. Holben⁴

1 Science Systems and Applications Inc., NASA/GSFC, Greenbelt, MD 20771
2 Laboratory for Atmospheres, NASA/GSFC, Greenbelt, MD 20771
3 Laboratoire d’Optique Atmosphérique, Centre National de la Recherche Scientifique et Université des Sciences et Technologies de Lille, Villeneuve d’Ascq, France
4 Laboratory for Terrestrial Physics, NASA/GSFC, Greenbelt MD 20771

Submitted to Geophysical Research Letters

Popular Summary

With the launch of the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor onboard the Earth Observing System’s (EOS) Terra satellite on 18 December 1999, new data sets of the global distribution and properties of aerosols are being retrieved, and need to be validated and analyzed. Aerosols, which are solid or liquid particles generated by natural processes (such as wind and wildfires) or as products or byproducts of human activities (e.g., from automobiles and industries), and suspended in the atmosphere, affect people and the environment in various ways. One of the most prominent parameters for quantifying the concentration of aerosols along the vertical atmospheric column over any given location is the aerosol optical thickness (AOT), which represents the degree of attenuation of solar radiation by aerosols through the processes of scattering and absorption. Algorithms developed by the MODIS aerosol group are used operationally to retrieve AOT both over land and over ocean from MODIS radiance measurements in different visible and infrared wavelengths of the electromagnetic spectrum. Also retrieved are other parameters useful for studying aerosol physical characteristics, such as Ångstrom exponent (representing the wavelength dependence of AOT), retrieved over land and ocean, as well as the effective radius (representing the dominant size mode of the aerosol particles over a given location) and the ratio of small to large size modes, both retrieved over ocean.

The validation of the MODIS aerosol products is essential to establish confidence in their quality in order to maintain high integrity in their use for climate and other environmental studies. MODIS aerosol products are validated with ground-based sun-photometer data, particularly those of the Aerosol Robotic Network (AERONET), comprising automatic tracking Sun photometers/sky radiometers located at over 100 sites around the world. The practical way of achieving this posed a challenge because, whereas MODIS achieves an almost complete global coverage once or twice a day, AERONET sun-photometer retrievals occur several times a day, but only over individual instrumented locations. To bring the two data systems into a compatible form that would facilitate inter-comparison and validation, we developed a software/database system to collocate and match spatial statistics of MODIS data with temporal statistics of AERONET data. This involves extracting subset of MODIS data within a 50 x 50 km box centered over each AERONET station for use in computing spatial statistics. Correspondingly, AERONET data measured within ±30 minutes of MODIS overpass are extracted and used to compute temporal statistics. The MODIS spatial statistics includes mean, standard deviation, direction and rate of spatial variation, and spatial correlation coefficient, while the corresponding AERONET temporal statistics includes mean, standard deviation, rate of temporal variation, and linear correlation coefficient. The software/data system is known as MAPSS (MODIS Aerosol and associated Parameters Spatio-temporal Statistics), and the data can be accessed online at http://modis-atmos.gsfc.nasa.gov/mapss.html.

The means and standard deviations of identical parameters from MODIS and AERONET were compared. Although, their means compare favorably, their standard deviations reveal some influence of surface effects on the MODIS aerosol retrievals over land, especially at low aerosol concentration. The direction and rate of spatial variation from MODIS are used to study the spatial distribution of aerosols at various locations either individually or comparatively. The spatio-temporal approach developed here has enabled the objective and rapid validation of satellite aerosol retrievals from MODIS with ground aerosol retrievals from AERONET. This is in spite of the differences in the characteristics of the two data sources. The methodology will allow the integration of satellite-retrieved aerosol parameters (and even other satellite products, such as water vapor) with data from various other sources to accomplish a wide variety of studies almost seamlessly.
widely used for various aerosol-related studies including satellite retrieval validation (e.g. Zhao et al., 2000).

The objective of this paper is to present the data structure developed for comprehensive and rapid global validation of MODIS aerosol products at near real time. A general assessment of the spatial characteristics of the products is also presented. Specific validations of the MODIS aerosol inversion schemes for retrievals over land and ocean are discussed in separate papers in this issue (Chu et al., 2001; Remer et al., 2001), which also show validation results for different geographical areas where different aerosol types (urban, industrial, marine, and smoke) are dominant.

2. Validation Database Design Concept

Whereas MODIS achieves an almost complete global coverage once or twice a day, sunphotometer retrievals (example, AERONET) occur several times a day, but only over individual instrumented locations. The MODIS Level 2 aerosol products are raster data sets with a pixel size of 10 x 10 Km. AERONET AOT data are acquired at 15-minute intervals on the average. It would be incongruous to compare single MODIS pixel values directly to AERONET point measurements for many reasons. First, the parameter value in an image pixel represents a spatial average over the pixel surface (an area of 10 x 10 Km for MODIS aerosol), and cannot be justifiably equated with a point value measured with a sunphotometer. Secondly, even if the pixel is small enough to represent a point, it is extremely unlikely that it would represent the same conditions as a sunphotometer point data, since their observation axes are different and the atmosphere is constantly in motion. Again, clouds may obscure a MODIS pixel directly over a sunphotometer site, but may not affect nearby pixels. In the same way, times of sunphotometer measurements seldom coincide exactly with MODIS overpass times, the closest pair being often larger than 5 min apart. With cloud contamination and consequent data filtering, the shortest time difference can stretch even longer. Therefore, to achieve a meaningful and balanced validation, we compare spatial statistics from MODIS with corresponding temporal statistics from sunphotometers. The justification is that, since airmasses are constantly in motion, an airmass captured by MODIS across a certain horizontal span over a sunphotometer site, will be sampled by the sunphotometer during a certain time period.

Corresponding MODIS and sunphotometer (currently only AERONET) data subsets are collocated in time and space, and extracted. Spatial and temporal statistics are computed and stored for use in validation and other studies. An automated system known as MAPSS (MODIS Aerosol and associated Parameters Spatio-temporal Statistics) has been developed for this purpose, and the data can be accessed at http://modis-atmos.gsfc.nasa.gov/mapss.html.

3. MODIS Data Subsetting And Statistics

The basis for the procedure is to identify within each MODIS aerosol image each pixel falling over a validation point by its longitude and latitude. Then an N x N Km subset centered on that pixel is extracted and its mean (µ) and standard deviation (σ) are computed. In addition, a linear multiple regression plane is fitted to each subset data and its greatest slope (θ) and azimuth of this slope (αθ) as well as the multiple correlation coefficient (Rs) are computed. The subscript 's' designates spatial distribution. Standard procedures have been used for linear multiple regression plane fitting as well as for computing µ, σ, and Rs (e.g. Fox,
As regards $\theta_0$ and $\alpha_n$, they are derived from the equation of the fitted plane, which can be expressed as:

$$ax + by + cz = k$$

where, for each pixel of the MODIS data subset being processed, $x$ and $y$ represent longitude and latitude values; $z$ is the aerosol parameter value (e.g. AOT); $a$, $b$, and $c$ are equation coefficients derived from the regression fit; and $k$ is a constant. A vector perpendicular to the plane is referred to as the normal vector. The cosines of the angles the normal vector makes with the $x$, $y$, and $z$ axes (termed the $x$, $y$, and $z$ direction cosines) can be used to define the slope and slope azimuth of the plane as follows:

$$\theta_s = \cos^{-1}(d_z)$$

$$\alpha_s = \cos^{-1}\left(\frac{d_y}{\sqrt{(d_x^2 + d_y^2)}}\right)$$

where, $d_x = \frac{a}{r}$, $d_y = \frac{b}{r}$, and $d_z = \frac{c}{r}$ are the $x$, $y$, $z$ direction cosines derived from the equation of the plane coefficients, and $r = \sqrt{(a^2 + b^2 + c^2)}$.

In this work, the computed $\theta_0$ and $\alpha_n$ are expressed in degrees. However, for slope, $\theta_s$, it should be noted that the parameter represented by the $z$ component of coordinate system are not in the same units as the $x$ and $y$ axes. It is necessary to remember the respective units when evaluating the physical gradient of the parameter variation. Thus, for instance, since a slope of 30° arbitrarily represents a vertical change of 0.577 units per unit horizontal distance, in a longitude-latitude coordinate system, a 30° slope in AOT represents a change of 0.577 per degree of geographic coordinates.

All the computed quantities ($\mu_n$, $\sigma_n$, $\theta_n$, $\alpha_n$, and $R_n$) are useful indicators of the local spatial characteristics of the aerosol parameter being analyzed, at a given location and time. Thus, $\mu_n$ represents the average value of the parameter, $\sigma_n$ expresses its local spatial variability, $\theta_n$ would indicate the degree of the spatial trend if any, $\alpha_n$ gives the direction of the trend, and $R_n$ shows how smooth the trend is.

To evaluate the effect of window size on parameter statistics, four window sizes ($30 \times 30$, $50 \times 50$, $70 \times 70$, and $90 \times 90$ Km) were tested. Larger window sizes could introduce undesirable errors due to topographic or aerosol type heterogeneity. Figure 1 shows a random sample of MODIS AOT at 0.55 $\mu$m for April 19, 2000 over different sites (land and ocean). The bar heights represent the means, while the spikes atop them represent the standard deviations. We found that the window-size dependence is generally small and has no specific trend, at least for the window-size range tested. However, we use $50 \times 50$ Km window-size for calculating all validation spatial statistics, because for the 10-Km aerosol products, $30 \times 30$ Km corresponds to only 3 x 3 pixels, which is a small statistical sample, insufficient for plane fitting. Moreover, the average travel speed of an aerosol front is of the order of 50 Km/h. This was visually estimated from animated daily sequences of TOMS aerosol index images (http://jwocky.gsfc.nasa.gov/aerosols/aermovie.html) for July to September 1988, where aerosol fronts are seen crossing the Atlantic from the west coast of Africa to the East coast of America (approximately 6000 Km) in about five or six days. Therefore, the $50 \times 50$ Km window would match a
1-hour sunphotometer data segment. All references to MODIS spatial statistics in the rest of this paper imply those based on the 50 x 50 Km (5 x 5 pixel) subset grid boxes.

4. AERONET Data Subsetting And Statistics

For each location identified in the MODIS data, the AERONET sunphotometer data segment acquired within the 1-hour period centered on the MODIS overpass time are extracted. Statistics of the sunphotometer data subsets, mean \( (\mu_t) \) and standard deviation \( (\sigma_t) \) are computed. However, since the data are not spatially distributed, it is not possible to fit a plane as for MODIS. Instead, the slope \( (\theta_t) \) and the linear correlation coefficient \( (r_t) \) are computed in the temporal domain (as designated by the subscript 't').

5. Result Analysis And Applications

The derived spatio-temporal statistics allow us not only to identify uncertainties in the retrievals, but also to analyze the local spatial behavior of the aerosol parameters at individual stations and between different locations. Because of possible gaps in the data subsets used in deriving the statistics, only those computed from a certain minimum number of values (5 pixels for MODIS and 2 data points for AERONET) will be considered in the following analysis. Examples of the symbol representation in the following analysis are: \( \mu_6(\tau_{660}) \), which stands for 'spatial mean of MODIS AOT at 660 nm'; and \( \sigma_t(\tau_{670}) \) denoting 'temporal standard deviation of AERONET AOT at 670 nm'.

Figure 2 shows scatterplots of \( \mu_6(\tau_{660}) \) against \( \mu_t(\tau_{660}) \) corresponding to MODIS retrievals for October 2000 over (a) land, and (b) ocean. The standard deviations \( \sigma_6(\tau_{660}) \) and \( \sigma_t(\tau_{660}) \) are plotted as error bars. The corresponding 1-1 lines (broken) as well as the least squares lines (solid) show the associated equations and correlation coefficients, \( R_t \), are shown. The correlation and regression coefficients show excellent agreement with AERONET measurements over the ocean (Figure 2b). Over land (Figure 2a), these coefficients reveal a relatively poorer performance, and at low AOT values \( \mu_t(\tau_{670}) < 0.2 \) the standard deviations appear to be larger for MODIS than for AERONET. This is probably due to the effects of surface variability in the MODIS retrievals over land. Figure 3a shows \( \sigma_6(\tau_{670}) \) (AERONET) and \( \sigma_t(\tau_{660}) \) (MODIS land and ocean) plotted against \( \mu_t(\tau_{670}) \) for October 2000. The least squares line and \( R_t \) as well as the root mean square (rms) value for each data set, are shown. Obviously, the values for MODIS over land show the least correlation and the most scatter, especially at low AOT \( (\mu_t(\tau_{670}) < 0.2) \). Reflectance at 2100 nm \( (\rho_{2100}) \) enables a good view of the surface variability since most aerosols are transparent at this wavelength. Figure 3b shows \( \sigma_6(\tau_{670}) \) and \( \sigma_t(\tau_{660}) \) plotted against \( \sigma_t(\rho_{2100}) \) for \( \mu_t(\tau_{670}) < 0.2 \). Amazingly, the correlations, \( R_t \), practically exceed those of Figure 3a, showing indeed that when AOT is low \( (\tau_{670} < 0.2) \), the variability of MODIS AOT over land is highly influenced by the variability of the land surface properties.

The local spatial distribution of the aerosols can be visualized from the spatial slope \( (\theta_t) \) and slope azimuth \( (\alpha_t) \) computed from MODIS. Figure 4a shows serial plots of \( \mu_6(\tau_{670}) \) and \( \mu_t(\tau_{660}) \), depicting \( \theta_t(\tau_{670}) \) and \( \theta_t(\tau_{660}) \) by the circle sizes, and \( \alpha_t(\tau_{670}) \) and \( \alpha_t(\tau_{660}) \) by the arrow directions. They represent MODIS AOT over NASA Goddard Space Flight Center (GSFC) in Greenbelt, Maryland, USA, for the period of September to December 2000. The slope \( (\theta_t) \) is an indicator for the relative aerosol loading gradient, while the
slope azimuth ($\alpha$) points to the direction of lowest aerosol concentration, assumed to point away from the source direction. For most days, the spatial distribution is approximately the same at both wavelengths. Because of the directional ambiguity associated with very shallow slopes, directions are not shown for $\theta < 1^\circ$ (equivalent to an AOT change of less than 0.018 over a distance of 1 degree (or about 110 Km)). Figure 4b shows similar plots only for $\tau_{660}$ comparing two sites: GSFC and the Maryland Science Center (MSC) located at about 50 Km NNE of GSFC. These two locations (GSFC and MSC) almost always seem to have approximately the same direction, suggesting that they share the same aerosol source(s), except perhaps on October 4. However, except for September 17 and October 8, MSC almost always has a relatively higher average aerosol loading $\mu$ ($\tau_{660}$) and steeper gradient $\theta$ ($\tau_{660}$). The low value of $\theta$ ($\tau_{660}$) on September 17 suggests that this could not be caused by any permanent surface artifact. Rather, it is probably due to the impact of the source strength of urban aerosols on most days, since MSC is in the city of Baltimore.

6. Conclusions

The spatio-temporal approach developed here has enabled the objective and rapid validation of satellite aerosol retrievals from MODIS with ground aerosol retrievals from AERONET. This is in spite of the differences in the characteristics of the two data sources. Furthermore, it has been possible to condense the local spatial distribution of each MODIS aerosol parameter for every site into just a few parameters ($\mu$, $\alpha$, $\theta$, $\alpha_s$, and $R_s$). This offers great opportunities to use an easy approach, such as spreadsheet analysis, to study the local spatial distribution of aerosols at various locations from satellite data. The methodology developed here will allow the integration of satellite-retrieved aerosol parameters (and even other satellite products, such as water vapor) with data from various other sources to accomplish a wide variety of studies almost seamlessly.

Acknowledgement. We would like to thank the various MODIS software development and support teams for the production and distribution of the MODIS data, and the AERONET teams for collecting, processing, and making available ground-based aerosol observations around the world. We would also like to thank Eric Moody (of the MODIS Atmosphere Group) for developing a MODIS data subsetting code, some of which routines were used in the MAPSS software.

References


Figure 1. Averaging results of MODIS AOT at 550 nm, for 19 April 2000, over different locations using different window sizes (30x30 Km, 50x50 Km, 70x70 Km, and 90x90 Km). The means are represented by the heights of the wide bars while the standard deviations are plotted as the top spikes.

Figure 2. Scatterplots of means of $\tau_{500}$ (MODIS 50x50 Km subsets) against $\tau_{670}$ (AERONET 1-hour subsets) for all concurrent retrievals around the world during October 2000 (a) over land (b) over ocean. The error bars represent the respective standard deviations. The 1-1 lines are shown as the diagonal broken lines, while the solid lines are the linear regression fits, with their corresponding equations and correlation coefficients, $R$, shown at the lower right corners.
Figure 3. Scatterplots of (a) standard deviations (sdev) of \( \tau_{660} \) over land and ocean (MODIS 50x50 Km subsets) and \( \tau_{670} \) (AERONET 1-hour subsets) against mean \( \tau_{670} \) (AERONET 1-hour subsets) (b) sdev of \( \tau_{470} \) and \( \tau_{660} \) over land against the sdev of surface reflectance (Refl) at 2100 nm wavelength, only for cases where the AERONET mean \( \tau_{670} \) < 0.2. The plots represent all concurrent retrievals around the world during October 2000. In all cases, linear regression lines are fitted and their correlation coefficients, \( R \), are shown.

Figure 4. MODIS mean AOT over 50x50 Km subsets, showing corresponding spatial slopes (circle diameters) and spatial slope directions (arrows). The scale at top right corner of each plot shows the angular relationships expressed in degrees. For slope, a value of \( \theta_{s} \) represents a change of \( \tan \theta_{s} \) per geographical degree (about 110 Km) of horizontal distance. The arrow point designates the principal direction of lower AOT values. Arrows are not shown for very shallow slopes (\( \theta_{s} < 1^\circ \)) because of associated directional ambiguity, but triangles are used to designate such points. The plots represent: (a) \( \tau_{670} \) and \( \tau_{660} \) over the NASA Goddard Space Flight Center (GSFC, Greenbelt, MD) validation site (b) only \( \tau_{660} \) over two validation sites, GSFC and the Maryland Science Center (MSC, Baltimore, MD), which are about 50 km apart.
Figure 1. Averaging results of MODIS AOT at 550 nm, for 19 April 2000, over different locations using different window sizes (30x30 Km, 50x50 Km, 70x70 Km, and 90x90 Km). The means are represented by the heights of the wide bars while the standard deviations are plotted as the top spikes.
Figure 2. Scatterplots of means of $\tau_{a660}$ (MODIS 50x50 Km subsets) against $\tau_{a670}$ (AERONET 1-hour subsets) for all concurrent retrievals around the world during October 2000 (a) over land (b) over ocean. The error bars represent the respective standard deviations. The 1-1 lines are shown as the diagonal broken lines, while the solid lines are the linear regression fits, with their corresponding equations and correlation coefficients, R, shown at the lower right corners.
Figure 3. Scatterplots of (a) standard deviations (sdev) of $\tau_{0660}$ over land and ocean (MODIS 50x50 Km subsets) and $\tau_{0670}$ (AERONET 1-hour subsets) against mean $\tau_{0670}$ (AERONET 1-hour subsets) (b) sdev of $\tau_{0470}$ and $\tau_{0660}$ over land against the sdev of surface reflectance (Refl) at 2100 nm wavelength, only for cases where the AERONET mean $\tau_{0670} < 0.2$. The plots represent all concurrent retrievals around the world during October 2000. In all cases, linear regression lines are fitted and their correlation coefficients, $R$, are shown.
Figure 4. MODIS mean AOT over 50x50 Km subsets, showing corresponding spatial slopes (circle diameters) and spatial slope directions (arrows). The scale at top right corner of each plot shows the angular relationships expressed in degrees. For slope, a value of $\theta_s$ represents a change of $\tan \theta_s$ per geographical degree (about 110 Km) of horizontal distance. The arrow point designates the principal direction of lower AOT values. Arrows are not shown for very shallow slopes ($\theta_s < 1^\circ$) because of associated directional ambiguity, but triangles are used to designate such points. The plots represent: (a) $\tau_{470}$ and $\tau_{660}$ over the NASA Goddard Space Flight Center (GSFC, Greenbelt, Md) validation site (b) only $\tau_{660}$ over two validation sites, GSFC and the Maryland Science Center (MSC, Baltimore, Md), which are about 50 km apart.