Improving Aerosol Retrieval Accuracy by Integrating AERONET, MISR and MODIS Data

Qifang Xu¹, Zoran Obradovic¹, Bo Han¹, Yong Li¹, Amy Braverman², Slobodan Vucetic¹

¹Center for Information Science and Technology
Temple University,
1805 N. Broad Street,
Philadelphia, PA 19122
USA
{qfxu, zoran, hanbo, liyong, vucetic@ist.temple.edu}

²California Institute of Technology,
4800 Oak Grove Drive,
Pasadena, CA 91109-8099
USA
Amy.J.Braverman@jpl.nasa.gov

Abstract – Retrieval of Aerosol Optical Thickness (AOT) by ground- and satellite-based remote sensing provides different accuracy, coverage and resolution. An important challenge is how to best utilize information from multiple instruments to further improve the quality of retrievals. In this study we explored whether the accuracy of AOT retrievals could be improved by fusion of ground- and satellite-based data using neural network techniques. MISR and MODIS satellite data were obtained for several 16-day periods during 2002 and 2003 covering the continental USA. These data are joined spatially and temporally with AOT measurements from 34 AERONET ground-based stations over the continental USA. The R² accuracies of MODIS and MISR retrievals were estimated at 0.57 and 0.66, when AERONET AOT is used as the ground truth. When radiance and geometric attributes are used together with MISR and MODIS AOT as attributes for prediction of AERONET AOT, the R² accuracy was increased up to 10%.

Keywords: data fusion, aerosols, MISR, MODIS, AERONET.

1 Introduction

Satellite remote sensing of the Earth and its atmosphere is a staple of modern geoscience. For many years now a network of geostationary (e.g. GOES-I series, METEOSAT) and polar orbiting (e.g. NOAA-14–NOAA-16) satellites have been providing data used in meteorology and atmospheric science research. More recently, NASA launched three low-altitude satellites (Terra, Aqua and Aura) as a part of the Earth Observing System (EOS) that provides steady, massive streams of data from multiple instruments. These instruments provide an unprecedented opportunity for long-term global observation of the land surface, biosphere, solid Earth, atmosphere, and oceans. The information is used to characterize how the Earth system is changing, as well as to identify and understand the primary causes of that variability [5].

As a basic principle of remote sensing, satellite instruments measure radiances reflected from Earth. The primary source of this reflected energy can be natural (e.g. solar radiation) or man-made (e.g. as for radar instruments). Depending on the instrument purpose and various operational constraints, different Earth observation satellite instruments have different spectral, spatial, and temporal resolutions. As an example, MODIS (the Moderate Resolution Imaging Spectrometer) instruments, of which there is one each aboard Terra and Aqua, observe the Earth in 36 spectral bands with spatial resolution of 250 m to 1 km, and at temporal resolution of about 1 or 2 days [12]. This results in about 1.5 GB of radiance observations daily, and an order of magnitude larger amount of processed data. These data are used to estimate underlying geophysical parameters such as atmospheric temperature profiles, cloud/aerosol properties, snow/ice cover, and vegetation characteristics. These derived, or retrieved, in the parlance of remote sensing, parameters are then used in various applications ranging from natural resource monitoring to development of general circulation models of the atmosphere used in studies of climate change.

This paper addresses data fusion challenges related to retrieval of aerosols. Aerosols are small particles produced by natural and man-made sources that both reflect and absorb incoming solar radiation. One of the biggest challenges of current climate research is to characterize and quantify the effect of aerosols on Earth’s radiation budget [11, 20]. There are two major sources of aerosol data: (1) satellite instruments, such as AVHRR-2, GOME, TOMS, SeaWiFS, POLDER, MODIS, and MISR; and (2) ground-based instruments, such as those used by AERONET [6], a federated international network of about 180 sun/sky radiometers collecting data since 1993. Satellite instruments provide global coverage with high spatial resolution, relatively low temporal resolution and allow moderately accurate retrievals. Ground-based instruments have limited spatial coverage, relatively high temporal resolution (many measurements per day), and are generally regarded as more accurate retrievals. As a result, AERONET is often used for validation of satellite-based retrievals.

It is evident that proper integration of data from
different instruments, platforms, and retrieval algorithms could lead to improvements in retrieval quality. However, in most cases retrievals are derived separately for each specific instrument. Two main reasons are (1) the daunting technical issue of joining and managing huge volumes of data stored in distributed data warehouses, and whose size increases by tens or hundreds of TB each year, and (2) the basic research issue concerning how to utilize all data sources in an optimal way through data fusion techniques. Current research efforts are only beginning to address this. Much of the recent work in aerosol retrieval improvement is concentrated on validation of retrieval quality from satellite instruments. Often, satellite derived AOT values are validated by comparing them with AERONET data, which are considered ground truth [17]. For example, such investigations have determined that MISR and MODIS AOT retrievals are more accurate than those of previous generation instruments such as AVHRR-2, TOMS, SeaWiFS, and POLDER [15], and that retrievals are generally most accurate over ocean and least accurate over bright areas such as deserts and snow/ice-covered terrain [1, 8]. Based on validation study results, domain scientists often manually adjust parameters of retrieval algorithms to improve their accuracy.

In this paper, we report on the results of a pilot study that explores a different approach to the construction of AOT data sets from multiple data sources. The approach is based on integrating multisource data into a single data table consisting of attributes derived from satellite-based observations (i.e. by MISR and MODIS instruments) and of a target attribute representing accurate ground-based retrievals (i.e. based on AERONET). Given such a data table, the problem of deriving an optimal estimate of AOT can be treated as a nonlinear regression. We refer to this approach as the multisource statistical retrieval. In Section 4 we compare accuracy of the multisource statistical algorithm with that of operational, satellite-based, single-instrument retrievals.

Since data produced from single-instrument operational retrieval algorithms are readily available along with the corresponding satellite radiance observations, we also explore potential benefits of using single instrument data as attributes in the multisource statistical retrieval. Our assumption is that the resulting retrieval algorithm should be at least as accurate as the best single-instrument retrieval. If so, the multisource statistical retrieval approach would represent a promising new method for efficient retrieval of other quantities from multiple sources in a number of different remote sensing applications.

## 2 Data Sets

### 2.1 MISR Data

MISR (Multi-angle Imaging SpectroRadiometer [3, 4]) is a key instrument aboard the Terra satellite for collection of aerosol and cloud related information. Terra was launched into polar orbit on December 19, 1999 [9] and started collecting data on February 24, 2000. MISR measures reflected solar radiation from nine view angles along the direction of flight (along-track), and in four spectral bands at each angle. Its spatial resolution is 275 m or 1.1 km depending on band and angle. On each Terra orbit, MISR sweeps out a 360 km wide swath of data from north to south while in daylight (Figure 1). Since MISR does not collect data at night, consecutive swaths are separated geographically resulting in 14 or 15 evenly spaced swaths per day. MISR’s ground footprint repeats nearly exactly every 16 days, which is the time it takes Terra to fly 233 distinct orbital paths. We obtained more than 100GB of radiance-related MISR data covering the continental US over five 16-day cycles from 07/01/2002 – 09/02/2002 and 12/24/2002 – 01/08/2003. From these data, we extracted all 36 radiance values at resolution 1.1 km.

![Figure 1. MISR and MODIS swaths](http://delenn.gsfc.nasa.gov/~imswww/pub/imswelcome/)

Low-quality radiance data (MISR radiometric data quality index (RDQI ≥ 2) were flagged and not used in our study.

We also obtained aerosol data for the study area and time periods These were retrieved from MISR radiances by complex, state-of-the-art, deterministic forward simulation models [14]. The main derived quantities are Aerosol Optical Thickness (AOT) at wavelength 558nm, and aerosol type, representing the most likely of 24 candidate types for each scene. For reasons related to computational complexity the resolution of MISR AOT is 17.6km. Thus, each data point represents the average AOT over a region of size 17.6×17.6km² obtained using the radiance information within the scene (Figure 2a). Aerosol information is retrieved only for regions not covered by clouds.

MISR data contains AOT derived in several different ways. Here we used « Regional Mean AOT » as the MISR retrieved AOT. We also extracted 37 geometric attributes describing MISR viewing geometry (solar zenith angle, 9 scattering angles, 9 glittering angles, 9 camera view angles, zenith angle, and 9 camera relative view azimuth values). These provide important information about camera angles and solar position. The angle attributes are provided at 17.6km resolution and are also important parameters of the operational retrieval algorithm [14].

1 obtained from Atmospheric Sciences Data Center at NASA Langley Research Center (http://delenn.gsfc.nasa.gov/~imswww/pub/imswelcome/).
2.2 MODIS Data

The MODIS instrument aboard Terra is an important instrument for understanding global dynamics and processes occurring on land, in the oceans, and in the lower atmosphere. Unlike MISR, MODIS has a single camera observing radiances over 36 spectral bands between 0.41μm and 14μm at three different spatial resolutions (250m, 500m, 1km) [18]. MODIS has a swath width of 2330km (Figure 1). Due to its larger orbital swath, MODIS achieves the global coverage every 1 – 2 days. MODIS also collects data during both day and night portions of Terra’s orbit.

We obtained MODIS radiance data covering the continental US over eight 16-day cycles: 07/01/02 – 08/01/02, 10/01/02 – 11/01/02, 12/24/02 – 01/23/03, 03/14/03 – 04/14/03. The collected data had spatial resolution of 5km and had the total size of about 80GB. We then extracted radiances in the 7 lowest wavelengths (0.47 – 2.1μm) since only those are used in the MODIS operational aerosol retrieval algorithm [10]. We also extracted the uncertainty indices of these radiances as potentially useful information.

We also obtained AOT at 550nm from the MODIS Aerosol Data Product. As with MISR computational complexity of the retrieval algorithm dictates, MODIS AOT is provided at reduced resolution, in this case 10km. Each data point thus represents the average AOT over a region of size 10×10km² (Figure 2b). We also extracted 5 geometric attributes (solar azimuth angle, solar zenith angle, sensor azimuth angle, sensor zenith angle, and scattering angle) from the MODIS data. These are important parameters in the MODIS Aerosol Retrieval Algorithm for Land. Finally, as with MISR, aerosol information is not retrieved for regions containing clouds.

2.3 AERONET Data

The Aerosol Robotic Network (AERONET) is a global, ground-based, remote sensing network supported by NASA and numerous international institutions that provides aerosol information from the ground [6]. AERONET radiometers measure AOT in 10 spectral bands between 340nm and 1640nm. AERONET has relatively high accuracy and precision [6] and is widely used in validation of satellite-based AOT retrievals [2, 7, 13, 16].

We obtained Level 2.0 cloud-screened and quality-assured AERONET data for 34 sites (Figure 3) in the continental US that overlapped with MISR and MODIS observations during the study time periods. Daily average AOT was extracted for each AERONET site. Since AOT is a function of wavelength, and since MISR and MODIS AOT retrievals are both available at wavelength of 558nm, AERONET AOT at 558nm was calculated by logarithmic interpolation of daily AOT estimates at 440nm and 675nm using the Angstrom exponent $\alpha_{440-675}$ [13]. If $\alpha_{440-675}$ was not provided in the AERONET data file, $\alpha_{440-870}$ was used. We used this interpolated AOT in our experiments as the AERONET AOT.

![Figure 2. a) MISR region (17.6km) and its subregions (1.1km); b) MODIS region (10km) and its subregions (1km).](http://aeronet.gsfc.nasa.gov/index.html)

![Figure 3. The 34 Aeronet sites](http://aeronet.gsfc.nasa.gov/index.html)
3 Methodology

3.1 Integration of aerosol data

From the raw data described in Section 2, we constructed the following 3 integrated data tables: (1) MisrAeronet, by spatially and temporally joining MISR and AERONET data, (2) ModisAeronet, by joining MODIS and AERONET data, and (3) MisrModisAeronet, by joining MISR, MODIS, and AERONET data. To join AERONET and satellite data, we apply the following condition: MISR/MODIS AOT retrievals and AERONET AOT retrievals must overlap both spatially and temporally. More precisely, the join is performed when satellite AOT is available for at least one region within 30km of the AERONET location (Figure 4) during the day for which daily AERONET AOT is available.

Figure 4. MISR/MODIS co-location with AERONET

**MisrAeronet Data.** AOT from 28 AERONET sites are joined with MISR data from the 5 orbital cycles. The resulting MisrAeronet data set contains 231 data points. Each data point consists of 72 radiance attributes, 37 angle attributes, 1 MISR AOT attribute, and 1 AERONET AOT attribute. The first 36 radiance attributes are obtained by calculating the average radiances over non-cloudy regions within a 30km radius from the AERONET location. The second 36 attributes are obtained as the corresponding standard deviations. The 37 angle attributes and the MISR AOT attributes are calculated by averaging of geometric attributes over the non-cloudy regions.

**ModisAeronet Data.** AOT from 34 AERONET sites are joined with MODIS data from the 8 orbital cycles. The resulting ModisAeronet data set contains 1465 data points. Each data point consists of 14 radiance attributes, 7 uncertainty attributes, 5 angle attributes, 1 MODIS AOT attribute, and 1 AERONET AOT estimate. The 14 radiance attributes are the averages and standard deviations of the 7 MODIS radiances over the non-cloudy regions within a 30km radius of the corresponding AERONET location. The 7 uncertainty attributes, 5 angle attributes, and 1 MODIS AOT attribute are calculated by similar averaging over non-cloudy regions.

**MisrModisAeronet Data.** This data set is obtained by joining MisrAeronet and ModisAeronet data tables. The available MISR and MODIS data overlapped during only three 16-day cycles, so the number of MisrModisAeronet data points was just 118. Each data point contains all attributes present in MisrAeronet and ModisAeronet data sets.

3.2 Multisource statistical retrieval

In the multisource statistical retrieval problem we learn a regression model that is able to accurately predict an underlying geophysical parameter such as AOT using attributes derived from satellite observations of multiple sensors. Attributes could include observed radiances, ancillary attributes (e.g. cloud or snow cover, topographical properties, vegetation index), or any quantity derived from these variables (e.g. by spatial or temporal aggregation, or nonlinear combination of attributes). In this study, we use attributes listed in the previous subsection, including AOT retrieved by MISR and MODIS operational algorithms. Since AERONET AOT retrievals are assumed to be significantly more accurate than satellite-based retrievals, AERONET AOT was used as the target attribute to be predicted by our regression model. We used a feedforward neural network with one hidden layer, 10 neurons in the hidden layer, and a backpropagation learning algorithm.

An important challenge in evaluating multisource statistical retrievals is how to estimate their accuracy. A standard cross-validation in which data are randomly divided into training and cross-validation samples is likely to produce overly optimistic accuracy estimates because of significant spatial correlation in AOT [19]. To avoid this we used leave-one-site-out cross-validation instead. In other words AERONET data was first grouped per sites. Then, a neural network was trained on the data from all the AERONET sites except the one left out. The neural network is then tested on the data from the excluded site. The leave-one-site-out process is repeated for all AERONET sites and the average accuracy is reported as an estimate of retrieval quality. We report accuracy as R-squared, calculated as Eq. (1), where MSE is the Mean Squared Error (MSE) of prediction, and Var(AOT) is the variance of the AERONET AOT target attribute.

\[
R^2 = 1 - \frac{MSE}{Var(AOT)}
\]

4 Experimental Results

4.1 MISR-AERONET fusion

In this experiment we used MisrAeronet data described in Section 3. The scatter-plot in the left panel of Figure 5 compares MISR and AERONET AOT retrievals for 231 MisrAeronet points. There is a strong correlation of 0.83 between MISR AOT and AERONET AOT. The R-squared accuracy of MISR operational retrieval algorithm was estimated as 0.66 based on the MisrAeronet data (note that by removing the five marked data points from the left panel of Figure 5 as outliers, the R-squared accuracy improves to 0.83). This accuracy serves as the basis for comparison between the operational MISR
retrieval algorithm and the proposed multisource statistical retrieval algorithm.

In Table 1 we summarize accuracies for different neural network regression models derived from the MisrAeronet data set. It can be observed that inclusion of the 36 radiance attributes already resulted in quite high accuracy, comparable to the state-of-the-art operational MISR algorithm. By adding the 36 radiance standard deviation attributes the accuracy was improved by 6%, while the inclusion of the 37 angle attributes improved the accuracy by additional 2%. The resulting R-squared accuracy of 0.708 is significantly higher than 0.662 of the MISR algorithm. This indicates that the statistical retrieval algorithm shows promise as a means of utilizing accurate ground-based retrievals to improve satellite-based retrievals. It is interesting to note that the addition of MISR AOT as an attribute improves the accuracy (from 0.65 to 0.68) when used in addition to the 36 radiance attributes, but not with the larger set of the 72 radiance and 37 angle attributes.

### Table 1. Accuracy comparison over MisrAeronet

<table>
<thead>
<tr>
<th>Attributes</th>
<th>With MISR AOT</th>
<th>Without MISR AOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MISR AOT (only)</td>
<td>0.662</td>
<td>-</td>
</tr>
<tr>
<td>36 Radiances (avg)</td>
<td>0.682±0.033</td>
<td>0.655±0.021</td>
</tr>
<tr>
<td>72 Radiances (avg, std)</td>
<td>0.701±0.055</td>
<td>0.694±0.026</td>
</tr>
<tr>
<td>72 Radiances, 37 Angles</td>
<td>0.705±0.044</td>
<td>0.708±0.043</td>
</tr>
</tbody>
</table>

Nevertheless, this result was obtained over a relatively small time period and spatial area and so should be treated as somewhat tentative. A more extensive study is needed before more definite conclusions could be made.

### 4.2 MODIS-AERONET fusion

In this experiment we used ModisAeronet data described in Section 3. The scatter-plot in the right panel of Figure 5 compares MODIS and AERONET AOT retrievals for 1,463 ModisAeronet points. Compared to MISR there is a somewhat weaker correlation of 0.74 between MODIS AOT and AERONET AOT. The R-squared accuracy of the MODIS operational retrieval algorithm was estimated as 0.57 based on the ModisAeronet data.

As seen from Table 2, the R-squared accuracy of statistical algorithms that use radiance, uncertainty index, and angle attributes was very small (with range between 0.23-0.28) and significantly below the MODIS operational algorithm. However, when MODIS AOT was added as additional attribute, the R-squared accuracy was significantly higher (with range between 0.59 and 0.64) than that of MODIS operational algorithm. It is interesting to note that the neural network with the smaller number of attributes (only 7 radiances and 7 uncertainty indices) appeared to have higher accuracy than the neural network with the complete set of attributes. Understanding the reasons for such behavior is a topic of our ongoing study.

### Table 2. Accuracy comparison over ModisAeronet

<table>
<thead>
<tr>
<th>Attributes</th>
<th>With MODIS AOT</th>
<th>Without MODIS AOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODIS AOT (only)</td>
<td>0.566</td>
<td>N/A</td>
</tr>
<tr>
<td>7 Radiances (avg)</td>
<td>0.606±0.022</td>
<td>N/A</td>
</tr>
<tr>
<td>7 Radiances, 7 UncInd (avg)</td>
<td>0.640±0.026</td>
<td>0.278±0.070</td>
</tr>
<tr>
<td>7 Radiances, 7 UncInd (avg, std)</td>
<td>0.610±0.040</td>
<td>0.235±0.062</td>
</tr>
<tr>
<td>7 Radiances, 7 UncInd (avg, std), 5 Angles</td>
<td>0.591±0.046</td>
<td>0.263±0.06</td>
</tr>
</tbody>
</table>

### 4.3 MISR-MODIS-AERONET fusion

In this experiment we used MisrModisAeronet data described in Section 3. Consistent with the previous results, for 118 MisrModisAeronet points MISR retrievals achieve higher correlation and R-squared accuracies (0.77 and 0.59, respectively) with AERONET retrievals than MODIS retrievals (0.62 and 0.39, respectively).

The first 3 rows of Table 3 are consistent with observations from Table 1, while the second 3 rows are consistent with observations from Table 2. It should be observed that the uncertainty in R-squared estimation is significantly larger than in Tables 1 and 2 due to the significantly smaller data set of 118 data points. The last 3
rows of Table 3 correspond to jointly using attributes from both MISR and MODIS. The results suggest that the integration is highly beneficial by being able to increase the R-squared accuracy to above 0.70. This is a strong indication that fusion of data from multiple sources could maximize retrieval quality compared to traditional single-source approaches. It also shows that the statistical retrieval algorithm is both a simple and robust approach for utilizing multisource information.

Table 3. Accuracy comparison over MisrModis/Aeronet. MISR AOT (MODIS AOT) attribute. MisrAeronet (ModisAeronet) data excluding the MISR AOT (MODIS AOT) data.

<table>
<thead>
<tr>
<th>MISR AOT Attr</th>
<th>MODIS AOT Attr</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>0.5935</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>0.632±0.065</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>0.651±0.056</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>0.3901</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>0.545±0.076</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>0.312±0.061</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>0.552±0.046</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>0.700±0.065</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>0.684±0.047</td>
</tr>
</tbody>
</table>

5 Conclusions and Future Work

In this paper we propose a statistical approach for exploiting multiple sources of remote sensing data to improve the quality of AOT data. The utility of the proposed approach is illustrated on an important problem in remote sensing: the retrieval of aerosol optical thickness satellite measurements. Specifically aerosol-related data from the MISR and MODIS instruments aboard the Terra satellite are merged spatially and temporally with data from ground-based AERONET instruments. Results show that a significant improvement in retrieval accuracy can be achieved. While further study is required, we are encouraged by the results presented here and believe that multisource statistical retrievals promise to provide a robust and relatively straightforward framework for integration of data from multiple sources in remote sensing.

We plan to pursue a more in-depth analysis of this methodology using more data with better spatial and temporal coverage. We also need to explore whether a more effective set of attributes could be constructed that would lead to further improvements in the accuracy.

Acknowledgments

We thank Ralph A. Kahn and John Martonchik, the MISR Science Team members at the Jet Propulsion Laboratory, for their generous help with understanding of the domain. We thank Atmospheric Sciences Data Center at NASA Langley Research Center for their support in collection of MISR and MODIS data. We also thank the AERONET PIs for providing the ground-based aerosol data.

References

[13] Yang Liu, Jeremy A. Sarnat, Brent A. Coull, Petrok Kouttrakis, and Daniel J. Jacob. Validation of Multiangle Imaging Spectroradiometer (MISR) aerosol optical thickness measurements using Aerosol Robotic Network (AERONET) observations over the contiguous United


