Development of a Two-Year-Long ARM Total Optical Depth Data Set Using the Least-Median-Squares Technique

M.-D. Cheng
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Introduction

A 2-year-long total optical depth (TOD) data set of 30-minute resolution from September 20, 1995, to May 18, 1997, has been constructed for the Southern Great Plains (SGP) site. The TOD values were derived from the surface shortwave irradiance measurements taken by one [E13/Solar Infrared Observing System (SIROS)] of the multifilter rotating shadowband radiometers (MFRSRs). This data set represents the "best" TOD estimate that spans several ARM intensive observation periods (IOPs). The best estimate of TOD value was computed as the composite of TODs that were obtained by employing two fundamentally different computing algorithms. These include the digital data filtering followed by the least-squares Langley analysis and the least-median-squares (LMS) Langley analysis without any filtering followed by outlier detection. The former algorithm was reported by Harrison and Michalsky (1994), while the later scheme was presented in Cheng (1997) and Cheng and Kopetz (1998). An intercomparison of two 30-minute TOD time series was made in this report. The data set is available through anonymous FTP from the Atmospheric Radiation Measurement (ARM) Archive. The aerosol optical depth (AOD) result based on LMS was not reported in this manuscript due to page limitation, but the data set is available from the author.

Description of Methods

TOD was retrieved from the spectral radiometric measurement by MFRSR E13 located at the central facility. The estimation of TOD starts with the Langley analysis that is given:

$$I_{(\lambda,t)} = I_{o\,(\lambda)} \exp[\,-\tau_{(\lambda,t)}\, m_t] \qquad (1)$$

where I is the direct normal narrowband irradiance (in units of W m^{-2} nm^{-1}) measured at the surface, and I_o (in units of W m^{-2} nm^{-1}) is the irradiance at the top of the atmosphere (TOA) to be estimated, τ is the TOD and dimensionless, and m is the dimensionless air mass or thickness of the atmosphere between the surface and TOA. The subscripts, λ and t, denote the physical quantities (i.e., I, I_o, and τ) that are functions of wavelength (λ) and time (t). The usage of subscripts will be dropped in the following text for simplicity. If I and m are known, it is straightforward to compute τ and monochromatic I_o by solving Eq. (1) using the least-square (LS) method (Lawson and Hanson 1974) if it is transformed into a linear model such as:

$$\ln(I) = \ln(I_o) - \tau * m \qquad (2)$$

where ln(•) indicates the natural log transformation of I and I_o at a given wavelength. When AOD is needed, it can be obtained by substituting the estimated I_o back to Eq. (1) to yield the instantaneous TOD. Corrections are needed for the Rayleigh molecular scattering and ozone absorption. Water vapor absorption is required at wavelengths longer than 0.9 μm.

The traditional approach for obtaining the unknown parameters in Eq. (1) that are I_o and τ was to identify "bad" data points (or outliers) and then a LS regression was applied to Eq. (2) after the bad data points were removed. The identification and removal of bad data points bear the same rationale as the "digital filtering" in signal communication and telecommunication. An objective approach was developed by Harrison and Michalsky (1994) to perform the filtering. A number of assumptions and efforts were made to design the filter. Although it has been working reasonably well, it is generally believed that there is no perfect filter.

If one carefully examines the problem of the retrieval of I_o and τ described, two findings emerge: 1) the digital filter was needed because the bad data points "may" harm the subsequent LS estimation of I_o and τ, and 2) the bad data points "cannot" be used when the TOD and AOD are to be inferred based on the I_o. It is intuitive that a perfect filter does not exist. An alternative approach for computing I_o and τ may be devised without the use of an artificial filter. The research-grade computing techniques that belong to a field called "robust statistics" (Huber 1981, Hampel 1975) have been available for more than a decade, but have not been popular because the LS regression software and subroutines was readily available. Moreover, the concept of robust statistics are difficult to assimilate. Here, we employed one of the robust computing techniques based on the concept of breakdown point originally proposed by Hampel (1975) and the algorithm devised mainly by Rousseeuw and his coworkers (Rousseeuw 1984, Rousseeuw and Leroy 1987) to Eq. (2). Cheng (1997) demonstrated this technique to the retrieval problem of AOD. Interested readers are referred to Cheng and Kopetz (1998) for the discussion in detail of this technique.

This robust computing technique is based on the LMS estimator in contrast to the LS estimator. The LS estimator has a zero percent tolerance to bad data points, while the LMS has a 50% tolerance according to the theory of robust statistics. In simple term, a single outlier may completely disrupt the LS estimate leading to an erroneous result. However, it takes 50% of the bad data to corrupt the LMS estimate. The filtering protection the LS required is not critical or required when the LMS is employed, unless more than 50% of the data are bad. If this, however, is the case, then the data set should not be used in the first place.

Two-Year Time Series Record of 30-Minute Total Optical Depth

Due to page limitation, we present the results from the 500-nm MFRSR channel in this manuscript.

Figure 1 shows the time series plot of the 30-minute TOD derived by Michalsky and his coworkers based on the digital-filter assisted LS estimator discussed in Harrison and Michalsky (1994). The percentiles, min, and max of the 2-year TOD values obtained by this method were 0.197 (25^{th} percentile), 0.213 (50^{th}), 0.245 (75^{th}), 0.155 (min.), and 0.446 (max.), respectively. The total number of 30-minute TOD estimates derived by this method for the period was 2121.

Figure 2 shows the time series plot of the TOD derived by using the LMS approach. The percentiles, min, and max of the 2-year TOD values obtained by this method were 0.197 (25^{th} percentile), 0.229 (50^{th}), 0.284 (75^{th}), 0.123 (min.), and 0.501 (max.), respectively. The total number of 30-min TOD estimates derived by the LMS was 3152.

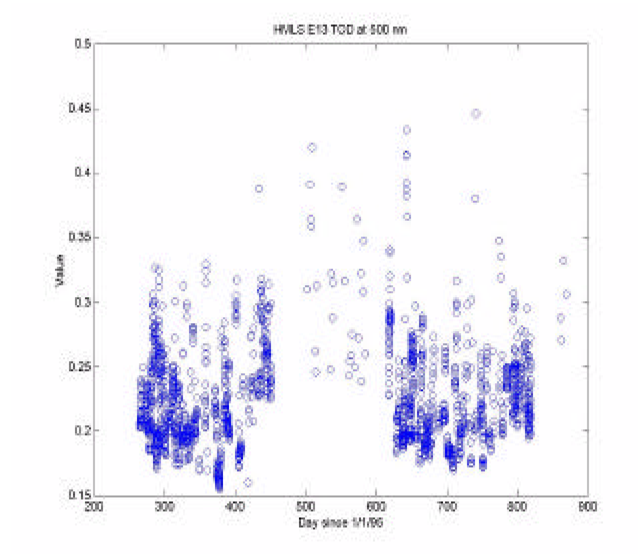

Figure 1. Time series plot of the 30-minute TOD based on the digital-filter assisted LS estimator. (For a color version of this figure, please see *http://www.arm.gov/docs/documents/technical/conf_9803/cheng-98.pdf*.)

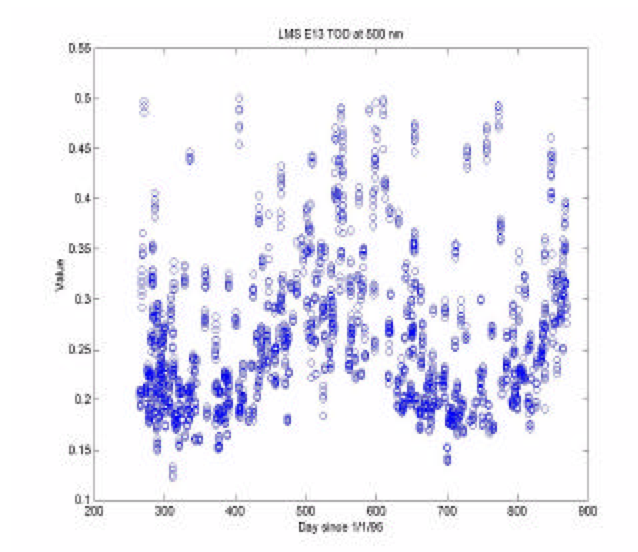

Figure 2. Time series plot of the TOD derived by using the LMS approach. (For a color version of this figure, please see *http://www.arm.gov/docs/documents/technical/conf_9803/cheng-98.pdf*.)

Figure 3 shows two time series on one plot. A cursory comparison of both time series shows that two methods agreed with each other most of the time. The Harrison-Michalsky Least Squares (HMLS) method appeared to impose more stringent constraints on the data points than the LMS approach; thus, it produced less number of TOD values than the LMS method did. This discrepancy of number of TOD values becomes significant during a certain time period (e.g., from day 450 to 600). That was around the end of March 1996 to the end of August 1996 during which HMLS yielded few TOD values. A closer examination is under way to investigate this difference.

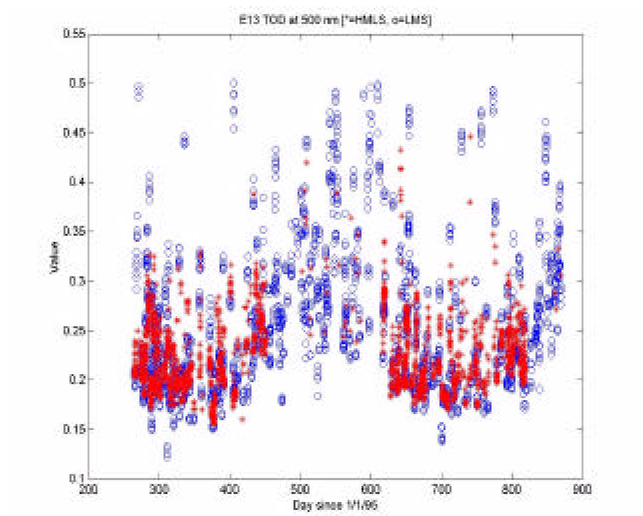

Figure 3. Two time series shown on one plot using both the HMLS and LMS methods. (For a color version of this figure, please see *http://www.arm.gov/docs/documents/technical/conf_9803/cheng-98.pdf*.)

The percentile plot shown in Figure 4 indicates approximately 99% of the TOD values differ by ≤ 0.05. The percentile plot shows that difference greater than 0.05 exists only for about 1% of the data points, when two time series were paired.

In short, TOD values produced by two fundamentally different approaches agreed with each other reasonably well. There were times when LMS appeared to assimilate outliers better and produced reasonable TOD values for these times, but HMLS did not. It is most likely that these data points were removed by the Harrison-Michalsky filter to protect of the LS estimator from generating erroneous results. Thus, no HMLS estimates were obtained for these bad times.

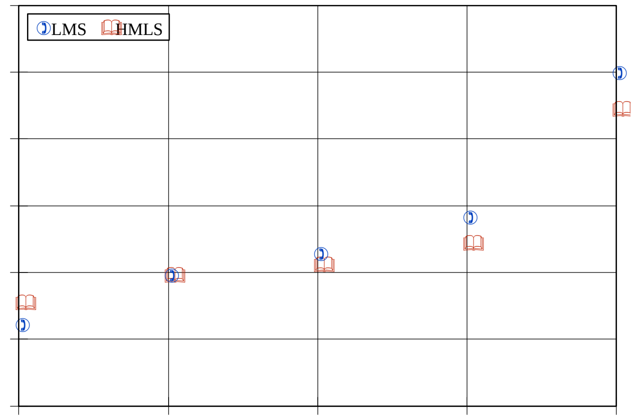

Figure 4. Percentile plot indicating approximately 99% of the TOD values differ by ≤0.05. (For a color version of this figure, please see *http://www.arm.gov/docs/documents/technical/conf_9803/cheng-98.pdf*.)

Availability of the 30-Minute Optical Depths

The 2-year TOD and AOD derived afterwards are available by contacting the author at Oak Ridge National Laboratory. His email address is: chengmd@ornl.gov, and phone and fax number are respectively: 423-241-5918 and 423-576-8646.

Acknowledgments

Jim Schlemmer of the Atmospheric Science Research Center at the State University of New York (SUNY)-Albany is acknowledged for providing the Harrison-Michalsky filter LS-based optical depth. The author thanks Joe Michalsky, also of SUNY-Albany, for discussion on HMLS Langley analysis. Terra Nash of the Environmental Sciences Division at Oak Ridge National Laboratory provided the computational support for the LMS-based optical depth. This work was partially funded by the U.S. Department of Energy's ARM Program. Oak Ridge National Laboratory is managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract number DE-AC05-96OR22464.

References

Cheng, M. D., 1997: Preprints of *The 3rd Conference on Atmospheric Chemistry*, The American Meteorological Society Annual Meeting, in Long Beach, California.

Hampel, F. R., 1975: Beyond Location Parameters: Robust Concepts and Methods. *Bull. Int. Stat. Inst*., **46**, 375.

Harrison, L., and J. Michalsky, 1994: Objective Algorithms for the Retrieval of Optical Depths from Ground-Based Measurements. *Appl Opt*., **33**, 5126-5132.

Huber, P. J., 1981: *Robust Statistics*, John Wiley & Sons, New York.

Lawson, C. L., and R. J. Hanson, 1974: *Solving Least Squares Problems*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 340 pp.

Rousseeuw, P., 1984: Least Median of Squares Regression. *J. Amer. Stat. Assoc*., **79**, 871.

Rousseeuw, P., and A. M. Leroy, 1987: *Robust Regression and Outlier Detection*. John Wiley & Sons, New York.

Other Publications in Progress

Cheng, M. D., and E. S. Kopetz, 1998: Retrieval of aerosol optical thickness by means of the least-median-squares robust algorithm I: Methodology. *J. Aerosol Sci*., submitted.