Automated cloud screening algorithm for MFRSR data

Mikhail D. Alexandrov
Department of Applied Physics and Applied Mathematics, Columbia University and NASA Goddard Institute for Space Studies, New York, USA

Alexander Marshak
NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Brian Cairns
Department of Applied Physics and Applied Mathematics, Columbia University and NASA Goddard Institute for Space Studies, New York, USA

Andrew A. Lacis and Barbara E. Carlson
NASA Goddard Institute for Space Studies, New York, USA

Received 18 November 2003; revised 22 December 2003; accepted 22 January 2004; published 28 February 2004.

[1] A new automated cloud screening algorithm for ground-based sun-photometric measurements is described and illustrated on examples of real (MFRSR) and simulated data. The algorithm uses single channel direct beam measurements and is based on variability analysis of retrieved optical thickness. To quantify this variability the inhomogeneity parameter ε is used. This parameter is commonly used for cloud remote sensing and modeling, but not for cloud screening. In addition to this an adjustable enveloping technique is applied to control strictness of the selection method. The key advantages of this technique are its objectivity, computational efficiency and the ability to detect short clear-sky intervals under broken cloud cover conditions. Moreover, it does not require any knowledge of the instrument calibration. The performance of the method has been compared with that of AERONET cloud screening algorithm. *INDEX TERMS:* 0305 Atmospheric Composition and Structure: Aerosols and particles (0345, 4801); 0320 Atmospheric Composition and Structure: Cloud physics and chemistry; 0360 Atmospheric Composition and Structure: Transmission and scattering of radiation; 0394 Atmospheric Composition and Structure: Instruments and techniques. **Citation:** Alexandrov, M. D., A. Marshak, B. Cairns, A. A. Lacis, and B. E. Carlson (2004), Automated cloud screening algorithm for MFRSR data, *Geophys. Res. Lett.*, *31*, L04118, doi:10.1029/2003GL019105.

1. Introduction

[2] The absence of an effective cloud screening algorithm has been always a problem for ground-based clear-sky retrievals. Some of the methods currently used for cloud screening of ground-based measurements extract the clear sky portions of their datasets by imposing thresholds on standard deviation of the measured values [*Smirnov et al.*, 2000; *Michalsky et al.*, 2001] or on the ratio between the direct and diffuse measurements [*Long and Ackerman*, 2000; *Augustine et al.*, 2003]. A sunphotometer calibration often needs to be determined from the data (using techniques such as Langley regression), that should be cloud-screened first. This means that a cloud screening algorithm applicable in such a case should work on raw un-calibrated data. The calibration independence requirement makes ineffective the use of the measurement's spectral signature (characterized e.g., by Ångstrom exponent) for cloud screening, because it can be affected by initial calibration uncertainties. We should note, that even if the instrument's calibration is accurately known, the wide range of retrieved Ångstrom exponents (especially for dust) [cf. *Michalsky et al.*, 2001] makes it very difficult, if at all possible, to choose the cloud-aerosol threshold.

[3] In this study we expand the cloud screening tool kit by adding a simple one-channel cloud screening algorithm, that is primarily designed for Multi-Filter Rotating Shadowband Radiometer (MFRSR) data. However its principles are sufficiently general to be broadly applicable. The MFRSR makes precise simultaneous measurements of the solar irradiances at six wavelengths (nominally 415, 500, 615, 670, 870, and 940 nm) at short intervals (20 sec in our case) throughout the day [cf. *Harrison et al.*, 1994]. Time series of direct solar beam extinctions and horizontal diffuse fluxes are derived from these measurements. The proposed cloud screening algorithm is based on local variability of optical thickness derived from the direct beam measurements in 870 nm channel. Provided that this channel is sufficiently narrow, measurements at this wavelength only contain aerosol and cloud contributions, i.e., the channel is situated away from spectral regions with trace gas and water vapor absorption. The 670 nm channel that is slightly impacted by ozone absorption can be used equally well. Thus, we refer to the clear sky optical thickness as aerosol optical thickness (AOT) as distinct from cloud optical thickness (COT). We should note, that the MFRSR-derived COT values may have a bias caused by the shadowbanding technique [*Joseph and Min*, 2003], however this does not affect our cloud screening. Our retrievals from MFRSR data [*Alexandrov et al.*, 2002] are based only on direct beam information (the diffuse flux measurements are used for calibration only). Thus, we declare a measurement "clear sky" whenever the line of sight between the sun and the instrument is free of cloud. This is different from the direct-

Copyright 2004 by the American Geophysical Union.
0094-8276/04/2003GL019105

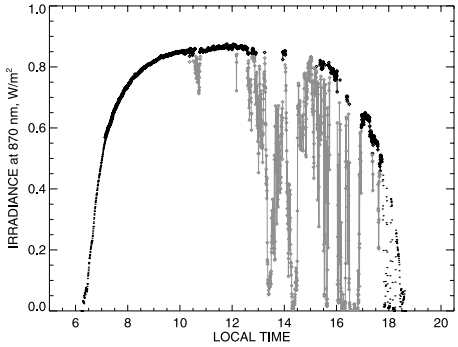

Figure 1. Direct normal irradiances in 870 nm MFRSR channel measured on 11 September 2000 at E13 Extended Facility. Cloudy parts of the data are shown in grey, clear-sky parts - in black. Dotted line depicts the part of the day before 7:00 and after 19:00 local time with airmass greater than 5, which is not analyzed.

diffuse cloud screening definition, where cloud presence anywhere in the sky affects the direct-diffuse ratio, and the "clear sky" flag is restricted only to the measurements with cloud-free hemispheric field of view.

[4] The proposed algorithm is based on the (scale dependent) parameter [*Cahalan*, 1994; *Cairns et al.*, 2000]

$$\varepsilon = 1 - \frac{\exp\left(\overline{\ln \tau}\right)}{\overline{\tau}}, \tag{1}$$

which characterizes the degree of horizontal inhomogeneity of an atmospheric field. Here τ denotes the optical thickness, the overbar indicates a moving average (over 15 data points = 5 min in our case). As a measure of atmospheric variability this parameter is complementary to the ratio of the standard deviation and the mean. The value of ε can vary between 0 for homogeneous and 1 for extremely inhomogeneous datasets and is a useful measure for evaluating the effect of inhomogeneity on the radiative properties of clouds [cf. *Rossow et al.*, 2002]. Note, that this parameter is invariant with respect to multiplication of τ by a constant factor. This means ε characterizes variability regardless of the magnitude of optical thickness. Some types of clouds (thin cirrus, marine boundary layer clouds) may have values of ε similar to those of aerosols. To separate clear-sky intervals from such clouds we apply the ε-test after a renormalization of the optical thickness (described in the next section).

[5] In our previous work [*Alexandrov et al.*, 2002] the cloud screening was done by visual examination of the "roughness" of the optical thickness curve followed by manual selection of clear intervals through an interactive computer program interface. This procedure was time consuming and its results depended on the skill of the examiner. The algorithm described below has several adjustable parameters (such as the moving average window length) appropriate setting of which can make the automatic cloud screening work as well as manual screening, while being faster and more objective. This algorithm facilitates automatic processing of MFRSR data.

[6] We illustrate the method on the data from a partially clear day 11 September, 2000. These data were obtained at the Extended Facility E13 (colocated with the Central Facility) of the DOE Atmospheric Radiation Measurement (ARM) Program (http://www.arm.gov) MFRSR network located at Southern Great Plains (SGP) in northern Oklahoma and southern Kansas. Figure 1 shows the direct-normal radiance in 870 nm channel for the entire day. The clear and cloudy parts designated by our algorithm are shown respectively in black and grey. The same cloud separation is shown on the optical thicknesses plot in Figure 2.

2. Description of the Cloud Screening Method

[7] To prepare the data for cloud screening, we convert the direct normal irradiances I measured in 870 nm channel during one day into (uncalibrated) optical thickness

$$\tau = -\ln\left(\frac{I}{I_0}\right) \cdot \mu - \tau_R, \tag{2}$$

where I_0 is the nominal TOA irradiance at 870 nm, μ is the inverse of the airmass (essentially equal to the cosine of the solar zenith angle), and τ_R is the Rayleigh scattering contribution [*Hansen and Travis*, 1974]. The cloudy intervals with high COT ($I \approx 0$) are removed from the data before Equation (2) is applied. The resulting optical thickness may contain calibration uncertainty of the form const $\cdot$ μ, that is a smooth function of time. We do not include data for airmass greater than 5 due to uncertainties in the MFRSR angular response and shadowband correction at large zenith angles.

[8] As we noted above, ε-test should not be used for cloud screening directly. Our tests show, that aerosols and thin (cirrus) clouds appear to have similar values of ε, thus they cannot be separated by a threshold in this parameter. This means that the optical thickness of cirrus clouds resembles AOT multiplied by a constant factor, while ε is invariant under such transformation. Thus, to separate clear sky from thin clouds we need to modify the original optical thickness by bringing AOT and COT variability to the same mean, while retaining the structure and size of their fluctuations. This is done by subtracting from the optical thickness time series τ its moving average $\overline{\tau}$ (with 15 datapoints = 5 min window) and adding back a constant $\tau_{\text{const}} > 0$ which is more typical as AOT value than COT value. We use $\tau_{\text{const}} = 0.2$.

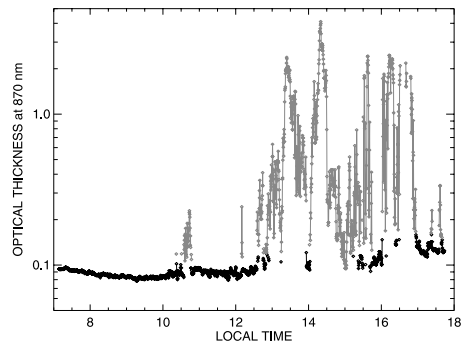

Figure 2. Cloud screening results on the optical thickness plot (in log scale), cloudy parts of the data are shown in grey, clear-sky parts - in black.

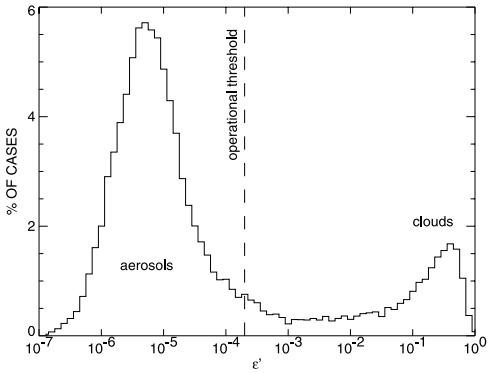

Figure 3. Statistical distribution of values of the inhomogeneity parameter ε' for the 26-day dataset obtained at E13 in September 2000. The threshold value $\varepsilon' = 2 \cdot 10^{-4}$ for initial separation of clear sky from clouds is shown by a dashed line.

The results of cloud screening, however, show almost no dependence on the exact value provided it is of the same order of magnitude. Thus, we obtain a renormalized

$$\tau' = \tau - \overline{\tau} + \tau_{\text{const}} \tag{3}$$

that has a mean value typical for aerosols, but different fluctuations in aerosol and cloud parts. Since the logarithm used in the definition of ε is not defined for non-positive arguments, we a priori designate the points with $\tau' \leq 0$ as cloudy. This approach can misidentify a very short (less than 5 min) clear sky interval between clouds as being cloudy, however this situation is statistically insignificant, since such short breaks in cloud are unlikely to truly be clear-sky. We compute ε' as a function of time for the rest of the data by substituting τ' instead of τ into (1) and using the same 5 min window in the moving average. Figure 3 shows statistical distribution of ε' values derived from 26 days of data (47,296 data points) obtained at E13 in September 2000. The two maxima correspond to the aerosol and cloud modes respectively. We have selected $\varepsilon' = 2 \cdot 10^{-4}$ as the operational threshold value for separating clear sky from clouds (this value is shown in Figure 3 by dashed line). The ε' can be expressed in terms of the nth central moments $m_n = \overline{(\tau - \overline{\tau})^n}$, $n = 1, 2, \ldots$ as

$$\varepsilon' = 1 - \frac{\exp\left(\overline{\ln \tau'}\right)}{\overline{\tau'}} = 1 - \exp\left[\sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} \frac{m_n}{\tau_{\text{const}}^n}\right]. \tag{4}$$

It follows from (4), that in the small variability limit $m_2 \approx 2\varepsilon' \cdot \tau_{\text{const}}^2$, and the threshold value of ε' translates to the threshold value of 0.004 of the standard deviation of τ. This is comparable to values used for MODIS cloud screening over ocean [*Martins et al.*, 2002], where the threshold standard deviation of 0.0025 is used on the spatial scale of 1500 m (equivalent to our 5 min if 5 m/sec conversion wind speed is used).

[9] The selected threshold value of ε' leads to rather conservative cloud screening. However, visual analysis of a set of partially clear days suggested that optimal results can be achieved if another method is used for selection of additional clear-sky points instead of an increase of the ε' threshold. To include data points between those selected as clear sky by ε'-test and exhibiting similar optical thickness values we use a procedure based on an enveloping technique similar to that of the empirical mode decomposition method [cf. *Huang et al.*, 1998]. First, we construct a min-max envelope consisting of the two curves: the min (max) curve is created by taking local minima (maxima) of the time series of the originally selected points and then linearly interpolating between these minima (maxima) to the whole dataset. We found that for optimal results the envelope should be slightly enlarged (the max-curve multiplied by 1.2, and the min-curve divided by the same factor). Then, additional data points which (a) are located within 30 min from an originally selected data point, and (b) have their optical thickness values within the constructed min-max envelope, are classified as clear-sky. This procedure is illustrated in Figure 4.

[10] Our cloud screening algorithm has no significant dependence on the original calibration of the data, except for the case of strong negative calibration error, when significant portion of clear sky data can be designated as cloudy. To avoid this problem we (a) roughly pre-calibrate the data using robust Langley regression before cloud screening, and (b) repeat cloud screening after analysis and calibration of the initially selected clear-sky dataset.

[11] The described algorithm also screens out portions of the data with rapid oscillations due to instrument misalignment, when the shadowband does not completely block the sun during the measurement.

3. Discussion

[12] We have described an automated cloud screening algorithm and demonstrated it on real MFRSR measurements. As with any cloud screening method, our algorithm has several parameters which can be user-optimized based on experience with a particular dataset. These parameters include the width of the moving average window (we have chosen 5 min which is of the order of the characteristic time for boundary layer variability); the threshold for ε' (we use $2 \cdot 10^{-4}$); the width of the min-max envelope (we selected an additional 20%); and the maximal distance from the initially selected data points within the min-max envelope

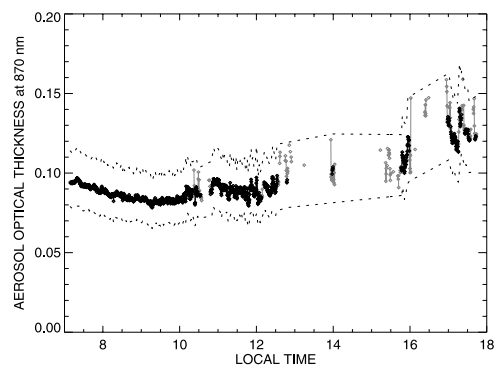

Figure 4. An illustration of the enveloping procedure: originally selected data points (with $\varepsilon' \leq 2 \cdot 10^{-4}$) are shown in black, the additional points - in grey, the min-max envelope used is depicted by dotted curves.

(we chose 30 min). Since the optimal values of these parameters are related to each other we suggest that these values be adjusted in the order in which they are listed above. We should note also, that changing these parameters may impact the result in different ways. For example, while raising of the threshold for ε' on one hand increases the risk of including cloudy points into clear-sky dataset, on the other hand it increases the density of the ε'-test selected clear-sky points and therefore makes the construction of the min-max envelope more reliable.

[13] Unfortunately, objective criteria are not readily available for validation of a cloud screening algorithm, since it is impossible to independently verify if the sunphotometer's field of view is obscured by a thin cloud or aerosol. Thus, we evaluated the performance of our new method by comparison with existing techniques and by conducting tests on simulated data with prescribed cloud positions (assuming that our simulations are sufficiently realistic). We have compared clear-sky selections made according to the proposed method with those chosen for a colocated CIMEL sun-photometer according to the cloud screening algorithm [*Smirnov et al.*, 2000] adopted by Aerosol Robotic Network (AERONET, http://aeronet.gsfc.nasa.gov). Figure 5 shows the comparison for the data obtained at SGP Central Facility on two days: 11 September, 2000 and 20 May, 1998. The latter measurements were made during the Central American fire smoke transport period [*Peppler et al.*, 2000] and show higher and more variable optical thicknesses. While temporal resolution and measurement techniques are quite different for MFRSR and CIMEL, Figure 5 shows certain similarity in clear sky selection in the corresponding datasets. Differences are mainly seen in the smoke case, where the AERONET cloud screening is more conservative than ours.

[14] We also applied our algorithm to a simulated dataset (2048 data points) with 28% cloud fraction (575 points) shown in Figure 6. The distribution of clouds with mean COT of 0.3 was simulated using a bounded cascade model [*Marshak et al.*, 1994]. A fractional Brownian motion [cf. *Mandelbrot*, 1982] was used to generate the AOT time series with the mean of 0.2. Both models had the same scaling parameter $H = 0.25$. Application of the proposed method with the parameter values specified above to this dataset shows good results. Only 71 data points (3.4%) were designated as clear, while actually being cloudy (however with small COT: 0.03 on the mean). On the other hand, only 83 data points (4%) were designated as cloudy, while actually being clear.

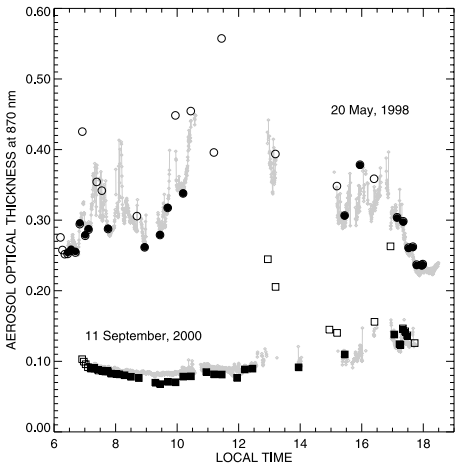

Figure 5. Comparison with AERONET cloud screening. The cloud-screened MFRSR data for 11 September, 2000 and 20 May, 1998 (smoke event) are shown in grey. The corresponding level 1.0 AERONET data points (after rough initial cloud screening) are depicted by circles for 20 May, 1998, and by squares for 11 September, 2000. The data points selected for level 1.5 cloud-screened dataset are filled with black.

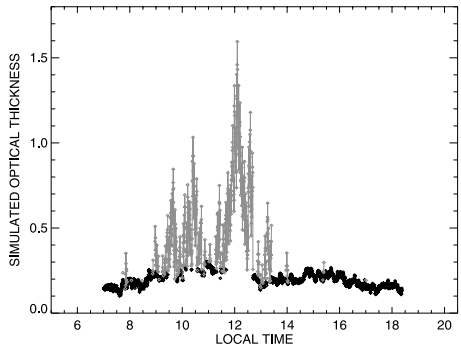

Figure 6. Simulated optical thickness time series. Cloudy parts are shown in grey.

[15] **Acknowledgments.** We would like to thank W. Rossow for useful discussions, and M. J. Bartholomew for her effort in maintaining the AERONET site at SGP. This research was supported by the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Environmental Sciences Division (Interagency Agreements No. DE-AI02-93ER61744 and DE-AI02-95ER61961), and by NASA's Radiation Science Program (RTOP No. 622-46-05-30).

References

Alexandrov, M., A. Lacis, B. Carlson, and B. Cairns (2002), Remote sensing of atmospheric aerosols and trace gases by means of Multi-Filter Rotating Shadowband Radiometer. Part I: Retrieval algorithm, *J. Atmos. Sci.*, *59*, 524–543.

Augustine, J. A., C. R. Cornwall, G. B. Hodges, C. N. Long, C. I. Medina, and J. J. De Luisi (2003), An automated method of MFRSR calibration for aerosol optical depth analysis with application to an Asian dust outbreak over the United States, *J. App. Meteor.*, *42*, 266–278.

Cahalan, R. F. (1994), Bounded cascade clouds: Albedo and effective thickness, *Nonlinear Proc. Geophys.*, *1*, 156–167.

Cairns, B., A. A. Lacis, and B. E. Carlson (2000), Absorption within inhomogeneous clouds and its parameterization in general circulation models, *J. Atmos. Sci.*, *57*, 700–714.

Hansen, J. E., and L. D. Travis (1974), Light scattering in planetary atmospheres, *Space Sci. Rev.*, *16*, 527–610.

Harrison, L., J. Michalsky, and J. Berndt (1994), Automated multifilter shadow-band radiometer: Instrument for optical depth and radiation measurement, *Appl. Opt.*, *33*, 5118–5125.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu (1998), The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. A*, *454*, 903–995.

Joseph, E., and Q. Min (2003), Assessment of multiple scattering and horizontal inhomogeneity in IR radiative transfer calculations of observed thin cirrus clouds, *J. Geophys. Res.*, *108*(D13), 4380, doi:10.1029/2002JD002831.

Long, C. N., and T. P. Ackerman (2000), Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects, *J. Geophys. Res.*, *105*(D12), 15,609–15,626.

Mandelbrot, B. B. (1982), *The Fractal Geometry of Nature*, 460 pp., W. H. Freeman, San Fransisco.

Marshak, A., A. Davis, R. Cahalan, and W. Wiscombe (1994), Bounded cascade models as nonstationary multifractals, *Phys. Rev. E*, *49*, 55–69.

Martins, J. V., D. Tanre, L. Remer, Y. Kaufman, S. Mattoo, and R. Levy (2002), MODIS cloud screening for remote sensing of aerosols over oceans using spatial variability, *Geophys. Res. Lett.*, *29*(12), 8009, doi:10.1029/2001GL013252.

Michalsky, J. J., F. A. Schlemmer, W. E. Berkheiser, J. L. Berndt, L. C. Harrison, N. S. Laulainen, N. R. Larson, and J. C. Barnard (2001), Multiyear measurements of aerosol optical depth in the Atmospheric Radiation Measurement and Quantitative Links programs, *J. Geophys. Res.*, *106*(D11), 12,099–12,107.

Peppler, R. A., et al. (2000), ARM Southern Great Plains site observations of the smoke pall associated with the 1998 Central American fires, *B. Am. Meteorol. Soc.*, *81*, 2563–2591.

Rossow, W. B., C. Delo, and B. Cairns (2002), Implications of the observed mesoscale variations of clouds for the Earth's radiation budget, *J. Clim.*, *15*, 557–585.

Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker (2000), Cloud-screening and quality control algorithms for the AERONET database, *Remote Sens. Environ.*, *73*, 337–349.

M. D. Alexandrov and B. Cairns, Department of Applied Physics and Applied Mathematics, Columbia University and NASA Goddard Institute for Space Studies, New York, USA. (malexandrov@giss.nasa.gov)

A. Marshak, NASA Goddard Space Flight Center, Greenbelt, MD, USA.

A. A. Lacis and B. E. Carlson, NASA Goddard Institute for Space Studies, New York, USA.