A New Cloud Screening Algorithm for Ground-Based Direct-Beam Solar Radiation

MAOSI CHEN AND JOHN DAVIS
UV-B Monitoring and Research Program, Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado

WEI GAO
UV-B Monitoring and Research Program, Natural Resource Ecology Laboratory, and Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, Colorado

(Manuscript received 8 April 2014, in final form 23 July 2014)

ABSTRACT

Cloud screening of direct-beam solar radiation is an essential step for in situ calibration and atmospheric properties retrieval. The internal cloud screening module of a Langley analysis program (LA) used by the U.S. Department of Agriculture (USDA) UV-B Monitoring and Research Program (UVMRP) is used for screening the uncalibrated direct-beam measurements and for deriving Langley offset voltages $V_{LO}$ for calibration of the UV version of the Multifilter Rotating Shadowband Radiometer (UV-MFRSR). The current LA cloud screening module utilizes data from extended clear-sky periods and tends to ignore shorter periods that typify periods of broken cloudiness, and as a result, fewer $V_{LO}$ values are generated for sites with higher frequencies of cloudy days (cloudy sites). A new cloud screening algorithm is presented that calculates the total optical depth (TOD) difference between a target point and pairs of points, and identifies the target as cloudy if the mean TOD difference exceeds a certain threshold. The screening is an iterative process that finishes when no new cloudy points are found. The result at a typical clear/sunny site shows that $V_{LO}$ values from partly cloudy days are consistent with those from cloud-free days, when the new method is employed. The new cloud screening algorithm picks up significantly more $V_{LO}$ values at cloudy sites. The larger decrease of the annual mean value of $V_{LO}$ at cloudy sites than at relatively clear sites suggests the potential for improving calibration accuracy at cloudy sites. The results also show that the new cloud screening method is capable of detecting clear points in short clear windows and in transitional regions.

1. Introduction

Ground-based solar radiation measurement systems are generally considered simple, reliable, and necessary to validate satellite measurements and retrievals (Smirnov et al. 2000; Krotkov et al. 2005). Separation of clear-sky and cloudy portions in measurements is an essential requirement that all these ground systems have to fulfill during in situ calibration and while producing retrievals of atmospheric properties [e.g., aerosol optical depth (AOD)]. One of the most common in situ calibration methods for direct-beam-measuring instruments [e.g., sunphotometer and Multifilter Rotating Shadowband Radiometer (MFRSR)] is Langley regression (Stephens 1994), which can only be applied on cloud-screened data.

The U.S. Department of Agriculture (USDA) UV-B Monitoring and Research Program (UVMRP) has been observing solar UV radiation at 37 sites across United States for over a decade. The primary instrument it uses is the ultraviolet version of MFRSR (UV-MFRSR). The UV-MFRSR receives the direct normal, diffuse horizontal, and total horizontal solar radiation at seven UV channels characterized by 2-nm full width at half maximum (FWHM) bandpasses and are nominally centered at wavelengths of 300, 305, 311, 317, 325, 332, and 368 nm (Slusser et al. 2000).

Chen et al. (2013) reviewed ground-based cloud screening methods published over the last two decades. There are four common types of cloud screening. In the
first type, cloud screening is performed on uncalibrated voltage data—the standard Langley analysis or its variants fall into this category. In the second type, cloud screening is performed on calibrated irradiance data. In the third type, cloud screening is performed on derived AOD data. In the fourth type, cloud screening is performed using colocated auxiliary equipment/data, such as a Total Sky Imager (TSI). Here we briefly review some common algorithms.

If the cloud screening method is based on ratios of measured voltages or irradiances rather than their absolute quantities, then this cloud screening method can apply to both the first and second types because the uncalibrated voltage data and the calibrated irradiance data have a constant ratio. Most type 2 cloud screening algorithms are not designed for the purpose of calibration. It is desirable to have a cloud screening algorithm suitable for both applications.

The current cloud screening module used by the UVMRP is the Langley Analyzer (LA), which was developed by the Atmospheric Solar Radiation Group at the University at Albany, State University of New York, Albany, New York. This technique, which uses the methodology described in Harrison and Michalsky (1994), is a two-step filter on a series of log transformed voltage (lnV) and airmass points. In the first step, points are sorted by air mass in ascending order and segments beginning at the point where lnV starts to increase and ending at the point where lnV starts to decrease are classified as cloudy (Chen et al. 2013). In the second step, the concavity test, the points with the slope of lnV exceeding a given threshold are cloudy. One of the LA variants developed by Lee et al. (2010) uses the maximum value composite (MVC) technique to acquire the largest voltage values in narrow airmass intervals. Those composite values are considered to be close to voltages measured under clear-sky conditions.

Long and Ackerman (2000) developed a cloud screening algorithm for irradiance data from pyranometers. The downwelling irradiances—the total and the diffuse components—are normalized by air mass. Applying thresholds on these two airmass-normalized components eliminates scenes containing optically thick clouds or haze, as well as high thin clouds, such as cirrus. Applying thresholds on irradiance variation over time and in comparison to a normalized diffuse ratio may eliminate some other cloudy points. When new cloudy points are detected in these processes, threshold values mentioned above will be adjusted and a new iteration of screening is triggered. Otherwise, the surviving points are the final clear-sky points.

Alexandrov et al. (2004) proposed an automated cloud screening algorithm for the time series of AOD derived from a single MFRSR channel. The method renormalizes the AOD time series by removing the local AOD average and calculates the corresponding inhomogeneity index (e') for each point. The statistical distribution of e' over a long period will show two distinctive maxima that correspond to the aerosol and cloud modes. Applying a threshold between the two maxima, the method separates the clear-sky points and cloudy points.

Smirnov et al. (2000) developed an automatic cloud screening algorithm on the time series of AOD derived from calibrated sunphotometers of the Aerosol Robotic Network (AERONET). Essentially, this algorithm requires that clear-sky points 1) are within a certain AOD range, 2) have stable and smooth AOD in nearby points, and 3) do not exceed a certain standard deviation of AODs from the daily AOD average.

The limitations of the existing cloud screening algorithms include

1) being able to perform only on the time series of optical depth (e.g., the index of inhomogeneity algorithm; Alexandrov et al. 2004), which requires accurate calibration that may not be available at the current step;
2) requiring consideration of the variation of nearby points or points in a relatively short window (e.g., in the index of inhomogeneity algorithm, the measurements taken within 5 min (or 17 points) are used to determine the cloudiness of the center point); or
3) missing some clear intervals due to their short duration or contamination by slight noise. For example, in LA, the slope between nearby points cannot exceed a certain threshold.

This paper proposes a new method that examines the surface-measured direct-beam irradiance (voltage) with the purpose of identifying the clear-sky points in the data. The main improvements of the proposed method include the following: 1) cloudy points can be compared to all points within a time (air mass) window of any length; 2) clear-sky transitional points and short clear-sky segments can survive the screening; and 3) the method does not require the Langley offset voltage and provides more clear points, which may be missed by the other methods.

The following sections will describe the new TOD pairing cloud screening algorithm, show its advantages mentioned above compared to other methods (especially the LA method), and its ability to obtain more Langley offsets on the cloud-screened data.

2. The algorithm description

a. The basis

Beer’s law tells us that

\[ V_t = V_{LO} \exp(-\tau_{Total,t} m_t). \]  

(1)
where \( m_t \) is the air mass at time \( t \); \( V_t \) is the cosine-corrected voltage of the direct beam at \( t \) at one channel; \( V_{LO} \) is the corresponding voltage as if measured at the top of the atmosphere (air mass \( 5.0 \); also known as the Langley voltage offset); \( \tau_{Total} \) (TOD\(_t\)) is the total optical depth (TOD in the nadir direction) at \( t \).

Taking the natural logarithm and rearranging, Eq. (1) can be recast as

\[
\text{TOD}_t = m_t^{-1} \ln V_{LO} - m_t^{-1} \ln V_t. \tag{2}
\]

It is noted that the Beer’s law is based on the single scattering assumption, which cloudy measurements do not obey. However, the purpose of this work is to distinguish cloudy points from clear points rather than to give an accurate optical depth at each point. The value of cloud plus aerosol optical depth value can be derived from the signal of a well-calibrated radiometer after subtraction of the Rayleigh scattering and gaseous absorption. By definition, cloudy points have higher total optical depth values than clear points. Therefore, clear points tend to distribute around the upper envelope in the \((m_t, \ln V_t)\) coordinate system, while the cloudy points are scattered lower than the envelope. This is one of the few assumptions on which the current cloud screening algorithm depends.

\section*{b. The transformed coordinate system}

Instead of the original measurement pair \((m_t, V_t)\), we use the transformed pair \((m_t^{-1}, m_t^{-1} \ln V_t)\) in the following discussion. The benefit of this transformation is that the TOD\(_t\) can be examined at any \( t \) directly because the slant paths have already been corrected or normalized by moving the \( m_t \) term to the right-hand side of the Eq. (2).

Figure 1 illustrates the transformed coordinate system. The mathematical basis for the whole figure is given by Eq. (2). In the normal Langley calibration method, \( V_{LO} \) is used with the extraterrestrial irradiance to determine the responsivity of the instrument, and it is assumed here that \( V_{LO} \) or its natural logarithm \( \ln V_{LO} \)
is constant during a short time range (e.g., one day). If the TODs were zeroes, then the ground measurements \( (V_i) \) would be close to the extraterrestrial counterpart \( (V_{LO}) \), which is constant by our assumption. The pair of \( (m_i^{-1}, m_i^{-1} \ln V_{LO}) \) would be on a straight line (the blue dashed line) approaching the origin. According to Eq. (2), any measurement with TOD greater than zero would have a \( V_i \) less than \( V_{LO} \). Therefore, the blue dashed line is the upper limit for ground measurements because the real TODs are greater than zero, because of the Rayleigh scattering and gaseous absorptions. When measurements are taken on a clear day, the TODs would be stable and Eq. (2) applied to these measurements could be treated as a linear regression problem, where \( m_i^{-1} \) is the independent variable, \( m_i^{-1} \ln V_i \) is the dependent variable, \( -\text{TOD}_{LR} \) is the intercept of the regression line, and \( \ln V_{LO} \) is the slope of the regression line. The green dashed line (the linear regression line) on black solid circles (clear-day measurements) gives an example of such a regression. Note that the TOD at any time is given as the vertical difference between the extraterrestrial line and the clear-sky line.

The coordinate transformed measurements are sorted by \( m_i^{-1} \) in ascending order. When applying the algorithm to more than half of a day, it is possible that multiple points have the same \( m_i^{-1} \). In our procedure, the first occurrence of such points is kept, while the rest are removed. Since more than one-half day may be considered, the usage of the word “time” refers to the corresponding \( m_i^{-1} \).

c. The outline of the algorithm

At the beginning, the algorithm treats every measurement as an indeterminate point, which means the cloudiness of this measurement has not been determined. The methodology of the procedure is to start by comparing the TOD of this point with other points to determine whether it is cloudy; and if it is not definitely cloudy, then to use it in determining whether other points are cloudy. The point whose cloudiness is to be determined is called the target point. All indeterminate points within a certain time range surrounding but excluding the target point constitute the local window for that target point.

The procedure pairs an indeterminate point in the local window with all other indeterminate points in the same local window. For a pair of such points, a particular type of weighted average TOD is calculated. If the two points of a pair represent nearly clear-sky points, then the TOD difference between the target and the weighted average of such a pair is an indicator of the target point’s cloudiness. The sections below mathematically explain why the algorithm uses the weighted TOD average rather than the standard average in comparing with the target’s TOD. Then a description is given of how to calculate the TOD difference between the target and a paired points’ weighted averages without knowing the value of \( V_{LO} \).

Figure 2 gives an example of a target (the purple pentagon) and one pair of points in its local window (green triangles).

In practice, comparing the target’s TOD to the average of only one pair may not be determinate because of the possibility of including cloudy points in the pair. Therefore, an assumption is made that there are many pairs of clear-sky points in the local window and the differences between the target’s TOD and those pairs’ weighted TOD averages would cluster at one value. Then, we calculate the TOD differences between the target and all pairs’ weighted averages in the local window. With the assumption above, outliers of the TOD differences are removed and the mean TOD difference of the remaining pairs is considered as a more robust indicator of the target’s cloudiness. If the mean TOD difference is positive and greater than a reasonable threshold, then the target is definitely cloudy.

When applying the screening routine described above on every indeterminate point, the examining order does not matter and doing so completes one iteration. Cloudy points will retain the cloudy status and be excluded in the future operations. A new iteration is triggered if any new cloudy points are found in the last iteration. If there were no new cloudy points found in the last iteration, then the cloud screening finishes. The surviving indeterminate points are considered clear points.

Following the algorithm design, the points with higher TODs are more likely to be labeled cloudy in the early iterations, while the points with TODs close to the baseline are more likely to survive the screening as would be anticipated.

d. Algorithm implementation

1) Pairing and TOD differences

The explicit expressions for the total optical depth at each point in the local window are

\[
\text{TOD}_0 = m_0^{-1} \ln V_{LO} - m_0^{-1} \ln V_0 \\
\text{TOD}_1 = m_1^{-1} \ln V_{LO} - m_1^{-1} \ln V_1 \\
\vdots \\
\text{TOD}_k = m_k^{-1} \ln V_{LO} - m_k^{-1} \ln V_k \\
\vdots \\
\text{TOD}_{Np-1} = m_{Np-1}^{-1} \ln V_{LO} - m_{Np-1}^{-1} \ln V_{Np-1},
\]

Equation (3)
where \( N_p \) is the number of indeterminate points in the local window of the target (Tgt). Since there should be at least one pair in the local window, the minimum value for \( N_p \) is 2. There is no maximum limitation for \( N_p \). In practice, the window size for UVMRP data is set to 256, which is large enough to include all points in a day (UVMRP data are measured every 20 s and the average recorded every 3 min).

The explicit expression for the total optical depth at the target Tgt is

\[
\text{TOD}_{\text{Tgt}} = m_{\text{Tgt}}^{-1} \ln V_{\text{LO}} - m_{\text{Tgt}}^{-1} \ln V_{\text{Tgt}}. \tag{4}
\]

Defining \( \Delta m_k^{-1} \) as the difference between the reciprocal of air mass at the \( k \)th point \( (m_k^{-1}) \) and that at the target \( (m_{\text{Tgt}}^{-1}) \),

\[
\Delta m_k^{-1} = m_k^{-1} - m_{\text{Tgt}}^{-1}. \tag{5}
\]

With Eqs. (5) and (4), Eq. (3) can be rewritten as

\[
\text{TOD}_0 = (m_{\text{Tgt}}^{-1} + \Delta m_0^{-1}) \ln V_{\text{LO}} - (m_{\text{Tgt}}^{-1} + \Delta m_0^{-1}) \ln V_0
\]

\[
\text{TOD}_k = (m_{\text{Tgt}}^{-1} + \Delta m_k^{-1}) \ln V_{\text{LO}} - (m_{\text{Tgt}}^{-1} + \Delta m_k^{-1}) \ln V_k
\]

\[
\text{TOD}_{N_p-1} = (m_{\text{Tgt}}^{-1} + \Delta m_{N_p-1}^{-1}) \ln V_{\text{LO}} - (m_{\text{Tgt}}^{-1} + \Delta m_{N_p-1}^{-1}) \ln V_{N_p-1}.
\]

Since no two points have the same air mass, direct TOD comparison between the target and any local window point is impossible when \( V_{\text{LO}} \) is unknown (see appendix A). For the same reason, it is also almost impossible to compare a target’s TOD with a standard TOD average of two local window points (see appendix B).

The standard average is a special case of a linear combination. Fortunately, for our purpose—comparing the target’s TOD to a pair of local window points (A and B)—any linear combination of TOD\(_A\) and TOD\(_B\) fulfills the requirement. The derivation of Eq. (7) to Eq. (11) shows how to calculate the TOD difference between the target.
and a pair points’ weighted average without knowing the
calibration term.

Defining \( wTOD_{AB} \) as the weighted TOD average of
points A and B,

\[
wTOD_{AB} = \frac{(M_A TOD_A + M_B TOD_B)}{M_T},
\]

(7)

where \( M_A, M_B, \) and \( M_T \) are three nonzero real value
multipliers. We are free to define the relationship be-
tween the three multipliers as

\[
M_T = M_A + M_B.
\]

(8)

Since \( M_A \) and \( M_B \) can be any nonzero real values, we can
set \( M_A = \Delta m_A^{-1} \) and \( M_B = -\Delta m_A^{-1} \), and get the following
equation:

\[
M_A \Delta m_A^{-1} + M_B \Delta m_B^{-1} = 0.
\]

(9)

Replacing TOD\(_A\) and TOD\(_B\) with Eq. (6) and applying
Eq. (8) and (9), Eq. (7) can be recast as

\[
wTOD_{AB} = m_T^{-1} \ln V_{LO} - \frac{M_A m_A^{-1} \ln V_A + M_B m_B^{-1} \ln V_B}{M_T}.
\]

(10)

Using Eqs. (4) and (10), we define \( \Delta TOD_P \) as the
difference between the target’s TOD (TOD\(_{Tgt}\)) and
the weighted TOD average of points A and B
(wTOD\(_{AB}\)):

\[
\Delta TOD_P = TOD_{Tgt} - wTOD_{AB}
\]

\[
= m_T^{-1} \ln V_{Tgt} + \frac{M_A m_A^{-1} \ln V_A + M_B m_B^{-1} \ln V_B}{M_T}.
\]

(11)

It is noted that all \( V_{LO} \) terms cancel out in Eq. (11),
which means that the calculation of \( \Delta TOD_P \) does not depend
on the Langley voltage offset. Since points A and B represent any two local window points, this
advantage applies for all other pairs.

2) OUTLIERS

In practice, examining the difference between the
target’s TOD and a single pair’s average TOD in the window
may not be conclusive because of the possibility of in-
cluding cloudy points in the pair. Figure 3 shows an
example of the histogram of \( \Delta TOD_P \). The x axis is \( \Delta TOD_P \)
with the bin size of 0.05. The y axis is the frequency
(number of cases) of the target’s pairs’ \( \Delta TOD_P \) falling
into the bins. It is seen that the histogram of \( \Delta TOD_P \) has
an obvious cluster center and long tails on both sides. Since there are ample clear-sky-point pairs in this
example, the cluster center represents clear-sky-point
pairs’ \( \Delta TOD_P \) and the tails represent pairs containing
cloudy points. Figure 4 illustrates the effect of pair points’
cloudiness in calculating \( \Delta TOD_P \). Figure 4a shows the
case when both pair points (A and B) are clear. In this
case, no matter where the target’s location—left, right,
or between—relative to the A–B pair, all \( \Delta TOD_P \) values
are the same. Figure 4b shows the case when pair point
A (left) is cloudy and pair point B (right) is clear. In this
case, targets on the left side of point B will have \( \Delta TOD_P \) lower than \( \Delta TOD_P \) compared to Fig. 4a, while
the target on the right side of point B will have \( \Delta TOD_P \) higher than \( \Delta TOD_P \) compared to Fig. 4a. Figure 4c shows the
case when pair point A (left) is clear and pair point B
(right) is cloudy. In this case, targets on the left side of
point A will have \( \Delta TOD_P \) higher than \( \Delta TOD_P \) compared to
Fig. 4a, while the target on the right side of point A will
have \( \Delta TOD_P \) lower than \( \Delta TOD_P \) compared to Fig. 4a.
Since the assumption has been made that there are many
pairs of clear-sky points in the local window, it will result
in \( \Delta TOD_P \) clustering to one value—the peak in Fig. 3—
and the tails represent pairs containing cloudy points. The process is
repeated two to five times varying by sites.

In Eq. (12) \( P_R \) are the sets of remaining pairs (near the
peak in Fig. 3), which are also considered as clear-sky pairs
and \( \Delta TOD_{Tgt} \) is the mean difference between the TOD of
the target and the weighted TOD of the remaining clear-
sky points, which is the simple average of \( \Delta TOD_P \) on the remaining pairs \( (P_R) \):

\[
\Delta TOD_{Tgt} = \frac{\sum_{AB \in P_R} \Delta TOD_P}{\text{size}(P_R)}.
\]

(12)

3) THRESHOLD

Since cloudy points have higher total optical depth
values than clear points, we can set a threshold value,
\( T_{\Delta TOD} \), above which the target’s TOD is too high com-
pared to the weighted TODs of clear-sky pairs in the local
window. Therefore, the condition of \( \Delta TOD_{Tgt} > T_{\Delta TOD} \)
suffices to determine the target to be cloudy.

If the purpose of using this cloud screening algorithm
is to identify the clearest points in a period and to obtain
the Langley offset from those points, a lower value of
\( T_{\Delta TOD} \) (e.g., 0.008) is preferred. If the purpose is to
identify points that are not contaminated by thick clouds
and to study the other constituents in the atmosphere (e.g., aerosols and trace gases), a higher value of $T_{\Delta TOD}$ is preferred.

The flowchart of the whole cloud screening algorithm is presented in Fig. 5.

e. Accuracy assessment

Although it has been assumed that there are adequate clear-sky points to obtain a cluster of AODs near the clear-sky value, nevertheless it is useful to confirm that the points identified as clear actually are. As an example of how this can be achieved, the measured diffuse-to-direct ratio (DDR) can be compared to the modeled value for a clear sky. The UV-MFRSR provides both the direct normal and diffuse horizontal irradiance measurements simultaneously. Before calibration, these irradiances are measured in units of millivolts. However, the direct normal and diffuse horizontal irradiance ratio

---

**FIG. 3.** An example of the histogram of $\Delta TOD_p$ for a certain target point at a certain iteration step. The x axis is $\Delta TOD_p$ with the bin size of 0.05. The y axis is the frequency (number of cases) of the target’s pairs’ $\Delta TOD_p$ falling into the bins.

**FIG. 4.** Illustration of the effect of pair points’ cloudiness in calculating $\Delta TOD_p$. (a) The case when both pair points (A and B) are clear. (b) The case when pair point A (left) is cloudy and pair point B (right) is clear. (c) The case when pair point A (left) is clear and pair point B (right) is cloudy.
(DDR) of the calibrated irradiances and uncalibrated voltages are the same because the same responsivity is used to convert both voltage measurements to their respective irradiance values. If the AOD, the Rayleigh optical depth (ROD), the ozone optical depth, solar geometry, and site location are known, then a radiative transfer routine such as the moderate resolution atmospheric transmission (MODTRAN) for UV and visible channels or the tropospheric UV model (Madronich 1993) for UV channels is capable of simulating DDR. Solar geometry and site location are measured and known properties. ROD is a function of ground pressure.

Fig. 5. The flowchart of the new TOD pairing cloud screening algorithm. The details of the process “Examine Tgt Cloudiness” are presented separately in the box below with a dashed outline. Vcc stands for the cosine-corrected voltage, asc.d. stands for ascending, and RecAM stands for the reciprocal of air mass.
and site location (Bodhaine et al. 1999) and is relatively stable over a day. Ozone optical depths are negligible at the 368-nm channel. When it is important, the total column ozone amount data are available from satellite. By elimination, AOD is the main unknown source that affects the DDR simulation. Therefore, AODs can be estimated by matching the MODTRAN-modeled and UV-MFRSR-measured DDRs on points indicated as clear. Using the retrieved mean AOD value, the direct normal irradiances at the clear-sky points are simulated by MODTRAN. The Langley regression can be applied to the measurements indicated as clear and the $V_{LO}$ obtained, which is used to calibrate the direct normal voltage data. The closer the direct normal irradiances from the Langley calibration and MODTRAN simulation, the more accurate the $V_{LO}$. An accurate $V_{LO}$ suggests that the points selected by the cloud screening algorithm are clear-sky points. Figure 6 summarizes the accuracy assessment procedure described above.

**Fig. 6. Schematic diagram of validation of cloud screening performance via comparing direct normal irradiances from Langley analysis and the radiative transfer model (MODTRAN).**
3. Results

The HI02 site at Mauna Loa, Hawaii, is a climatology site operated by the USDA UV-B Monitoring and Research Program (UVMRP). Because of its high altitude (3409 m) and great distance from any continents, its atmosphere is less affected by varying aerosol loading and is relatively stable, which makes HI02 a good site for applying Langley regression. The original Langley analysis at HI02 gives more \( V_{LO} \) hits with lower fluctuation compared to other UVMRP sites. To demonstrate that the new cloud screening method has the ability of selecting good points for calibration, the values of \( V_{LO} \) from linear regression on clear points given by the new cloud screening method and those from the original Langley analysis in 2013 are compared.

Figure 7 shows a comparison of the relative differences of \( V_{LO} \) (sun–Earth distance factor normalized) between values derived from linear regression of Eq. (2) on the points passing the new cloud screening algorithm and those from the original Langley analysis at HI02 in 2013. The circles represent the sunny-day cases when almost all points are measured with no cloud contamination and the squares represent the partly cloudy cases. Except for some sporadic outliers, the relative difference is about \(-2\%\) to 0%. On sunny days, the points remaining after applying the new cloud screening algorithm are usually the same as those passing the internal cloud screening module in the Langley analysis. The relative difference of \( V_{LO} \) at around \(-1\%\) in a such case is believed to result from the fact that the coordinate system of \((m_i^{-1}, m_i^{-1} \ln V_i)\) used in this paper gives relatively balanced weights to both points with high and low air mass in the linear regression. It is also evident that the relative differences of \( V_{LO} \) between the sunny-day cases and the partly cloudy cases compare well, which suggests that the new cloud screening algorithm gives points that generate unbiased \( V_{LO} \) in partly cloudy cases as compared to the sunny-day cases.

The stable atmosphere at the HI02 site is optimal for Langley calibration; however, for many sites where air pollution, clouds, and the combination of both are frequent, the Langley method is not as reliable as a means of calibration. In contrast, the FL02 site at Homestead, Florida, is characterized by frequent and fast-moving stratocumulus clouds. The internal cloud screening module of the original Langley analysis often misses short periods of clear points. As a result, there may not be sufficient clear points to calculate \( V_{LO} \). Figure 8 shows the FL02 site’s 368-nm channel’s values of \( V_{LO} \) (sun–earth distance factor normalized) in 2013 from the original Langley analysis and...
those from the linear regression of the points passing the new cloud screening algorithm. The latter gives about 56% more values of $V_{LO}$ than the original LA in 2013. The mean value of $V_{LO}$ using the new cloud screening algorithm is about 2.8% lower than that of the original LA, and the standard deviation is about 4% higher.

Figure 9 shows an example of the cloud screening performance of the original Langley analysis (top) and the new method (bottom) at the FL02 site on 26 September 2013. The point sets for the two methods are the same: morning points with an airmass range between 1.5 and 3.0. For the Langley analysis 12 points survived from its cloud screening procedure and 5 clear points (3 in transition between clear and cloudy and 2 in short clear periods) are missed in comparison to the new method. After the removal of regression outliers, the Langley analysis allows 9 points, which is less than its minimum requirement (12 points), and therefore the original Langley analysis does not generate $V_{LO}$ on that day. The new cloud screening method picks up all 5 clear points missed by the original Langley analysis and has 17 points for the further linear regression process. Although 2 of them are the regression outliers, there are still 15 points left to generate $V_{LO}$ on that day. The $V_{LO}$ (raw, without sun–earth distance normalization) generated using the clear points that survived the new cloud screening algorithm is 1576.40. One can calculate the corresponding $V_{LO}$ (norm, sun–earth distance normalized) as 1584.29, which is in the middle of all $V_{LO}$ (norm) in 2013 at the FL02 site. This example shows the new cloud screening algorithm’s ability of picking up clear points in transitional regions and in short clear-sky periods. It also shows that the calibration based on Langley regression can benefit from including those clear points.

To demonstrate the performance of $V_{LO}$ generated in the last example (FL02, 368 nm, 26 September 2013), the simulation of direct normal irradiance using the radiative transfer model MODTRAN is conducted. The AOD is determined by matching the MODTRAN-modeled and UV-MFRSR-measured DDRs on clear-sky points. The mean AOD value for this case is 0.103 at 550 nm (or 0.154 at the 368-nm channel). The values of important input parameters to the MODTRAN model are listed in appendix C. Figure 10 displays the MODTRAN-simulated and the new method-calibrated direct normal irradiance at the FL02 site on 26 September 2013. The clear-sky points’ irradiances are plotted with dark blue circles and the cloudy points’ irradiances are plotted with light blue circles. The MODTRAN simulations are the small red points. It is seen that all clear-sky points’ irradiances are close to the MODTRAN
results, including the five points saved by the new cloud screening method but missed by the LA cloud screening module. The five points are indicated by the arrows in Fig. 10. The mean square error (MSE) for this case is $2.87 \times 10^{-5}$, while the MSE value for a day inappropriate for Langley analysis may be two to three magnitudes higher. It suggests that the clear-sky points have nearly equal AOD values. It is also evident that cloudy points
determined by the new cloud screening algorithm have much lower irradiances than clear-sky points (Fig. 10).

Table 1 lists the statistics of $V_{LO}$ before and after applying the new cloud screening algorithm to measurements of 368-nm irradiances at five UVMRP sites in 2013. For the relatively clearer sites, HI02 and NM02, the number of $V_{LO}$ increased slightly and the annual mean value of $V_{LO}$ decreased by 1% after substituting the cloud screening module in the original Langley analysis. For sites where cloudy measurements are more frequent—that is FL02, OK02, and CO02—the number of $V_{LOS}$ increased between 33.8% and 56.7% and the annual mean value of $V_{LOS}$ decreased between 2.51% and 3.41%. The increased number of $V_{LOS}$ suggests that the new TOD pairing cloud screening algorithm picks up significantly more $V_{LOS}$ at cloudy sites. The larger decrease in the annual mean value of $V_{LO}$ at cloudy sites suggests the potential for improving calibration accuracy at those sites.

4. Summary

A new cloud screening algorithm for narrowband direct-beam measurements is developed. The mathematical basis of this algorithm is Beer’s law. Measurements are reorganized to a converted coordinate system that emphasizes the relative magnitude of measurements’ total optical depth (TOD). Instead of examining the fluctuation of a target measurement with nearby points, this algorithm calculates the TOD difference between a target and pairs of all indeterminate points and considers the target a cloudy point if the TOD difference exceeds a certain threshold value. All points are of indeterminate status at the beginning of cloud screening. Each point is examined with all other indeterminate points. If new cloudy points are found, then a new

<table>
<thead>
<tr>
<th>$V_{LO}$ source</th>
<th>oLA</th>
<th>CSLA</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI02</td>
<td>274</td>
<td>1748.64</td>
</tr>
<tr>
<td></td>
<td>289</td>
<td>1729.28</td>
</tr>
<tr>
<td>FL02</td>
<td>83</td>
<td>1631.45</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>1585.60</td>
</tr>
<tr>
<td>NM02</td>
<td>242</td>
<td>1388.49</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1374.13</td>
</tr>
<tr>
<td>OK02</td>
<td>106</td>
<td>1565.01</td>
</tr>
<tr>
<td></td>
<td>152</td>
<td>1511.72</td>
</tr>
<tr>
<td>CO02</td>
<td>148</td>
<td>1946.82</td>
</tr>
<tr>
<td></td>
<td>198</td>
<td>1897.92</td>
</tr>
</tbody>
</table>
iteration of examination is triggered. The cloud screening finishes when no new cloudy points are found in the last iteration. The surviving indeterminate points are considered clear points. The new cloud screening method is verified by comparing the Langley voltage offsets ($V_{LO}$) determined from the clear-sky intervals to values generated by linear regression in the original calibration program LA. The results at the relatively clear site at Mauna Loa Observatory, Hawaii, shows that values of $V_{LO}$ from partly cloudy days are not biased in comparison to those from sunny days. The results at the more cloudy site at Homestead, Florida, shows that 56% more $V_{LO}$ at cloudy sites. The larger decrease of the annual mean value of $V_{LO}$ at relatively cloudy sites than at relatively clear sites suggests the potential for improving calibration accuracy at cloudy sites. The result also shows that the new cloud screening method is capable of picking up clear points in short clear windows and in transitional regions.

Acknowledgments. This work is supported by the USDA UV-B Monitoring and Research Program under Grant USDA NIFA projects (2012-34263-19736 and 2013-34263-20931).

APPENDIX A

**TOD Comparison (Direct)**

The quantity TOD$_{D1}$ is defined as the difference between the target’s TOD (TOD$_{Tgt}$) and one local point’s TOD:

$$\text{TOD}_{D1} = \text{TOD}_{Tgt} - \text{TOD}_k,$$

where $k$ lies within the target’s local window. Using Eqs. (4) and (6) on point $k$, we get

$$\text{TOD}_{D1} = -\Delta m_k^{-1} \ln V_{LO} - m_{Tgt}^{-1} \ln V_{Tgt} + m_k^{-1} \ln V_k.$$

(A2)

In Eq. (A2), $\Delta m_k^{-1}$ is a nonzero value because every point (both the target and the local window points) has a unique air mass and $V_{LO}$ is the unknown calibration parameter. Therefore, it is impossible to calculate directly. Using a coarsely estimated $V_{LO}$ to do the TOD comparison was found to be impractical. First, it is unknown how the responsivity of the instrument may change based on its previous behavior. It could be stable for years, but it could also drop or increase quickly in a short period.

Second, the TOD$_{D}$ calculation using Beer’s law [Eq. (2)] shows that the bias of the daily $V_{LO}$ approximation is constant, while the airmass varies a lot in a day. As a result, the bias in TOD$_{D}$ varies a lot in a day. The comparison of such TOD$_{D}$ with varying magnitudes of bias will make it extremely complicated if not impossible to set a threshold for excluding cloudy points.

**APPENDIX B**

**TOD Comparison (Standard Average)**

The quantity TOD$_{D2}$ is defined as the difference between the target’s TOD (TOD$_{Tgt}$) and the standard average of any two local points’ (A and B) TODs:

$$\text{TOD}_{D2} = \text{TOD}_{Tgt} - (\text{TOD}_{A} + \text{TOD}_{B})/2,$$  

(B1)

where the points A and B lie within the target’s local window. Using Eqs. (4) and (6) on points A and B, we get

$$\text{TOD}_{D2} = -\Delta m_A^{-1} + \Delta m_B^{-1} - 2 \ln V_{LO} - m_{Tgt}^{-1} \ln V_{Tgt} + m_A^{-1} \ln V_A + m_B^{-1} \ln V_B.$$  

(B2)

Equation (B2) contains the unknown calibration parameter ($V_{LO}$), which makes it impossible to calculate TOD$_{D2}$ directly. Although it is possible that $\Delta m_A^{-1} + \Delta m_B^{-1}$ is or close to zero for some situations (when $\Delta m_A^{-1}$ and $\Delta m_B^{-1}$ have opposite signs and their absolute values are the same or close), normally it is not true (especially not true for the target in the early morning, late afternoon, and near solar noon, when most $\Delta m_A^{-1}$ and $\Delta m_B^{-1}$ have the same sign).

**APPENDIX C**

**Important MODTRAN Parameters**

To simulate the 368-nm-channel direct normal and the diffuse horizontal solar irradiance at the FL02 site on 26 September 2013, the following MODTRAN parameters are used. The parameters are for MODTRAN, version 5.3.

**Card I**

- **MODEL = 1** Tropical atmosphere
- **ITYPE = 3** Vertical or slant path to ground
- **IEMSCT = 4** Execute in spectral solar radiance mode with no thermal scatter
- **IMULT = 1** Execute with multiple scattering

**Card IA**

- **DIS = T** Use Discrete Ordinate Radiative Transfer (DISORT) discrete ordinate multiple scattering algorithm
DISALB = T
Calculate spectral albedo and diffuse transmittance

NSTR = 8
Number of streams to be used by DISORT

O3STR ‘a0.2784’: Column ozone amount (ATM-cm), data source: Earth Observing System (EOS) Aura Ozone Monitoring Instrument (OMI) daily level 3 global 0.25° gridded data (http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=omi)

LSUNFL = T
Read a user-specified top of the atmosphere (TOA) solar irradiance data

LBMNAM = T
Read the root name of the band model parameter data

Card 1A1
DATA/SUN01-SA02010.dat
Data provided by Chance and Kurucz (2010)

Card 1A2
01_2009
The root name of 1.0 band model

Card 2
IHAZE = 1
Rural extinction
VIS = -0.103
Negative of the 550-nm vertical aerosol optical depth
GNDALT = 0.000
Altitude of surface relative to sea level (km)

Card 3
H1ALT = 0.000
Initial altitude (km)
OBSZEN = 180.000
Initial zenith angle (°) as measured from H1ALT

Card 3A1
IPARM = 2
Method of specifying solar geometry on Card 3A2
IPH = 2
Select Mie-generated database of aerosol phase functions
IDAY
Day of year 269 (26 September 2013)
ISOURC = 0
Extraterrestrial source is the sun

Card 3A2
PARM1
Azimuth angle, which varies at each observation
PARM2
Solar zenith angle, which varies at each observation

Card 4
V1 = 26789
Initial frequency in wavenumber (cm⁻¹)
V2 = 27526
Final frequency
DV = 1
Frequency increment
FWHM = 2
Slit function full width at half maximum
FLAGS (7:7) = F
Write a spectral flux (.flx) file
MLFLX = 1
Number of atmospheric levels for which spectral fluxes are output

REFERENCES


