OPEN ACCESS Asian Journal of Scientific Research

ISSN 1992-1454
DOI: 10.3923/ajsr.2018.

Research Article
Development of a Quality Control and Cloud-screening Algorithm for Sun Photometry Langley Calibration

[1]Chang Hian Wui Jackson, [2]Chee Fuei Pien and [2]Dayou Jedol

[1]Preparatory Center for Science and Technology, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
[2]Energy, Vibration and Sound Research Group (e-VIBS), Faculty science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia

Abstract

Background and Objective: This study presents the development of a quality control and cloud-screening algorithm for sun photometer data to improve the state-of-the-art well-known Langley calibration method. The quality control algorithm accounts for removing obvious cloudy days and selecting the ideal Langley plot within a restricted air mass range to guarantee the reliability and reproducibility results. The cloud-screening algorithm is developed for sun photometer data collected from less ideal observatory sites such as low and mid-altitude sites. The algorithm includes calculating the moving averages of the Perez index under different windows time combined with a procedure for detecting negative derivatives. **Materials and Methods:** The algorithms were evaluated using sun photometer measurements from Laban Rata, (6.05°N, 116.56°E, 3,270 m a.s.l.), Kinabalu Park (6.0°N, 116.0°E, 1,574 m a.s.l.) and Sepanggar in Kota Kinabalu (6.03°N, 116.12°E, 18 m a.s.l.). A total of 60 Langley plots at four distinct wavelengths, 470, 500, 670 and 870 nm, were collected using an LED-based sun photometer. **Results:** The analysis of the results indicates that both algorithms provide comparable results despite the varying altitude measurements. On a point-by-point observational basis, the calibration constant of the sun photometer maintains a difference of less than 0.09, depending on wavelength. **Conclusion:** This study concludes that both algorithms are useful for improving the state-of-the-art well-known Langley calibration method in terms of its reliability and reproducibility.

Key words: Langley calibration, sun photometer, cloud-screening, solar calibration, high altitude, Perez model

Received: Accepted: Published:

Citation: Chang Hian Wui Jackson, Chee Fuei Pien and Dayou Jedol, 2018. Development of a quality control and cloud-screening algorithm for sun photometry langley calibration. Asian J. Sci. Res., CC: CC-CC.

Corresponding Author: Chang Hian Wui Jackson, Preparatory Center for Science and Technology, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia Tel: +60-088-320 000, Fax: +60-088-320 049

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Atmospheric aerosols are small particles suspended in the air that affect the Earth's radiation budget by scattering and/or absorbing solar and infrared radiation[1,2]. They also indirectly affect weather circulation patterns as cloud condensation and ice nuclei[3,4]. The monitoring of atmospheric aerosols is commonly done by measuring the aerosol optical depth (AOD), which is a measure of the total extinction of the radiation that enters and passes through the atmosphere and is of critical importance to the understanding of how the climate is forced radiatively. Sun photometers are a common instrument used for measuring AODs due to the simplicity and portability of the instrument. However, the key to the accurate determination of AODs is the calibration of the sun photometer[4,5]. The most common and economical method of calibrating a sun photometer is the Langley method, which is a method of determining the sun's radiance at the top of the atmosphere with ground-based instrumentation. Of critical importance to any sun photometer application is the routine calibration of the instrument under stable atmospheric conditions[6,7]. An ideal site for these aerosol-stable conditions and little to insignificant cloud contamination is at a high altitudes. Nevertheless, while the impacts of aerosols are often pronounced downwind of mid-altitudes, high-altitude sites are also sensitive to variations of aerosol concentrations[8]. This is especially true in the tropics, where the airborne pollutants, dust and smoke from transboundary areas are transported poleward during the monsoon season and/or haze period[9,10]. Therefore, the need for a quality control algorithm for sun photometer calibration is essential to guaranteeing suitable Langley days to be used for calibrations, even for data collected from high-altitude site measurements[11].

Despite the aerosol-stable conditions, cloud contamination is another major challenge in Langley calibrations, especially in the tropics[12]. Many studies have shown that Langley calibrations at low altitudes are feasible provided that sufficiently strict algorithms are implemented to account for accuracy[13]. The traditional problem in the tropics is the abundant cloud cover and inconsistent short-interval thin cirrus clouds[2,14]. These features are difficult to detect and are often removed subjectively by observing the outlier trends of the best-fit lines. This method is unscientific, heavily experience-dependent and always entails erroneous extrapolations[15]. Therefore, an objective and robust cloud-screening algorithm is of crucial importance for Langley calibrations, especially for measurements collected anywhere but high altitudes.

In this study, two important aspects of Langley calibration are discussed. The first is the proposed quality control (QC) algorithm for sun photometer data at high altitudes and the second is the cloud-screening (CS) algorithm for sun photometer data at low and mid altitudes. The QC algorithm is used to select suitable Langley days for data collected at high altitudes.

It is assumed that cloud contamination over high-altitude observatories is relatively low and therefore, the QC algorithm emphasizes the suitable air mass range to use[16]. Conversely, the CS algorithm removes the random thin cloud occurrences of the data collected from low or mid altitudes. The filtration is done in a completely automated and objective manner. The proposed algorithms are both investigated in this study.

MATERIALS AND METHODS

A total of 60 Langley plots was collected from three different observatory sites at varying altitudes using a portable radiometer: An ASEQ LR-1 spectrometer (Vancouver, Canada). This study was conducted from August, 2015-March, 2016. Each observatory represents a different altitude: Low (<10 m a.s.l.), mid (<2,000 m a.s.l.) and high altitudes (>3,000 m a.s.l). Measurements were made on visibly clear forenoons starting at sunrise, between 600 and 900 local time at periodic intervals of 3 min. Measurements from afternoon data were avoided due to the abundant cloud cover that was always prevailing during the sunset hours, especially for the tropical climates, where thick fog and rainfall is regularly expected over the study area[17].

The instrument has a 3648-element CCD-array silicon photodiode detector from Toshiba that enables optical resolutions as precise as 1 nm (FWHM). Each measurement series consists of global and diffuse irradiance components. The direct irradiance component was determined by subtracting the diffuse irradiance scans from the corresponding global irradiance scans as[18]:

$$I_{direct}^{\lambda,t} = I_{glocal}^{\lambda,t} - I_{diffuse}^{\lambda,t} \quad (1)$$

where, λ is the wavelength of a particular spectral light and t represents the time of the measurement. The LR-1 spectrometer was not equipped with a shadowing band, the diffuse irradiance component was measured using a manual shading disk diffuser after each global irradiance measurement. The diffuse component irradiance was measured for each scan of the global component irradiance using a shading disc to overfill the image of solar disc on

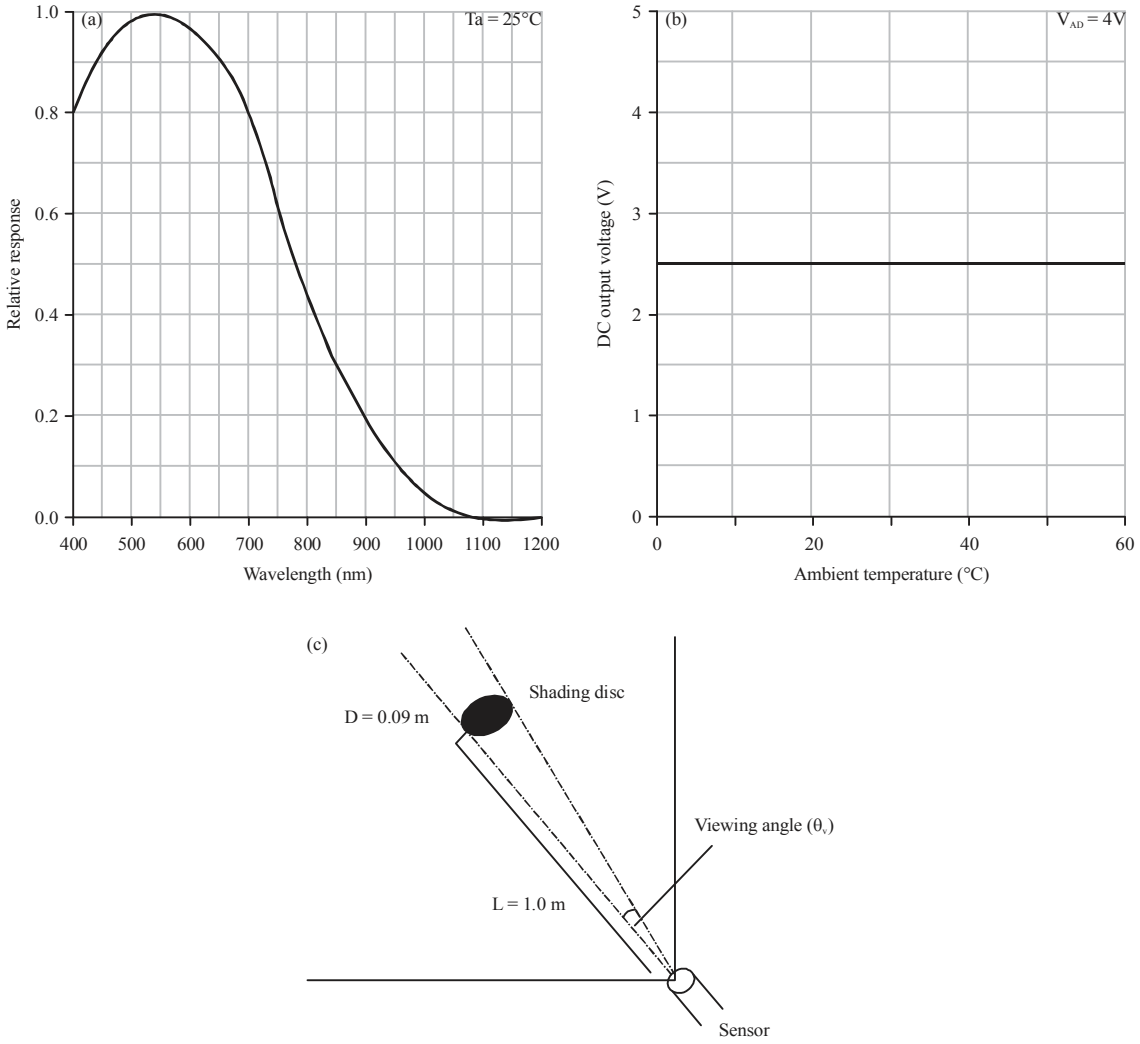

Fig. 1(a-c): Typical performance curves of a Toshiba TCD1304 linear silicon CCD array

(a) Spectral response, (b) Ambient temperature provided in the datasheet and (c) Diffuse flux measurement using a shading disc with a 0.09 m diameter held at 1 m parallel to the sensor to ensure the shading angle θ_s to the sensor is the same as the viewing angle θ_v of the sensor

parallel axis from direct viewed by the sensor. This shading disc has 0.09 m diameter (D = 0.09 m) and was held 1.0 m from the sensor (Fig. 1c). The dimension of the shading disc is determined by following the condition that the shading angle θ_s of the shading disc to the sensor should be same as the viewing angle θ_v of the sensor. Here, the viewing angle is defined as the maximum angle at which the senor can detect light radiation with an acceptable accuracy. To reduce the uncertainties incurred in Eq. 1, a cosine corrector was used and attached to the sensor head. Given that the viewing angle of the cosine corrected sensor was 5.0°, the ratio of the shading disc radius R to the distance from the shading disc L to the sensor should meet the following condition:

$$\tan\theta_v = \frac{D}{L} \quad (2)$$

where, D represents the diameter of the shading disc. In this way, the shade of the shading disc over the shaded spectrometer covers at least the whole of the sensor head of the spectrometer, but the margin area is kept to a minimum. The data acquisition of the radiometer was averaged at 10 ms for the temporal stability of the spectrometer. The typical performance curves of a Toshiba TCD1304 linear silicon CCD array for (a) Spectral response and (b) Ambient temperature measurements are as shown in Fig. 1a-c. Details of the Langley measurement protocol used with this portable spectrometer are discussed elsewhere[13,16].

RESULTS AND DISCUSSION

Quality control algorithm for sun photometer data at high altitudes: A conventional Langley plot can be performed at high-altitude sites for clean and aerosol-stable conditions[5,6] but not all Langley days are suitable for sun photometer calibration. This is especially true for sites with tropical climates, where abundant cloud clover occurs regularly[2]. To objectively select ideal Langley plots from a pool of datasets, an objective QC algorithm for sun photometer data from high altitudes is necessary. The flow chart of the QC algorithm for sun photometer data proposed in this study is shown in Fig. 2. The first step of Fig. 2 removes obvious cloudy days that are not suitable for Langley plots. This was evaluated by examining the correlation R^2 of the untreated Langley plots where only the Langley plots with $R^2>0.80$ for $\lambda<870$ nm and $R^2>0.70$ for $\lambda>870$ nm were selected. For more accurate results, Langley cases with higher R^2 are recommended, but our results were capped at $R^2>0.80$ ($\lambda<870$ nm) and 0.70 ($\lambda>870$ nm). The results of the untreated Langley plots collected at high altitudes are shown in Table 1. The QC algorithm following the first step eventually selected Day 1 and Day 5 for further analysis.

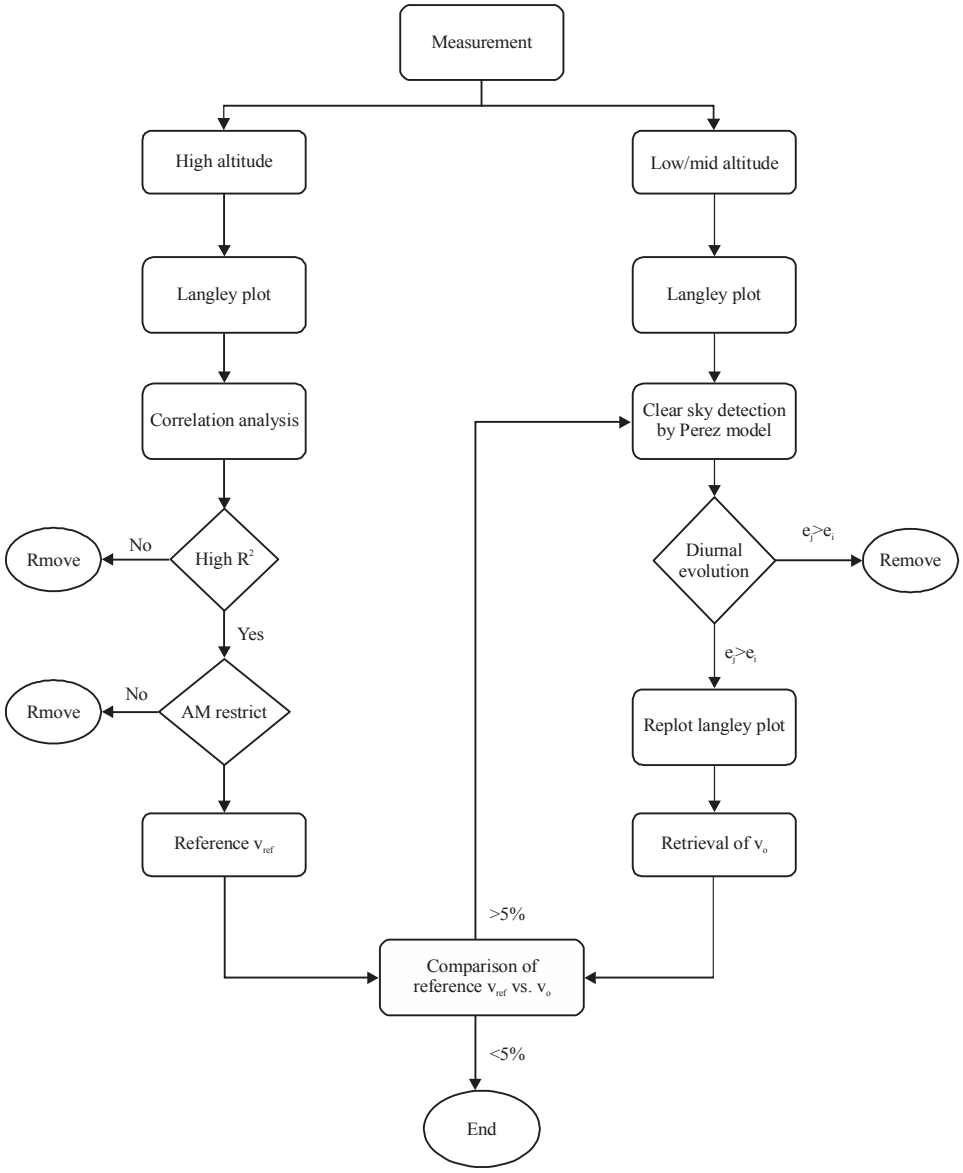

Fig. 2: Flowchart of the quality control (QC) algorithm for the sun photometer data with the high-altitude measurements (left) and cloud-screening (CS) algorithm applied to low/mid-altitude measurements (right)

Table 1: Untreated Langley plot obtained from the measurement of sun photometers at high altitudes

	Day 1	Day 2	Day 3	Day 4	Day 5
440 nm					
Regression line	y = -0.47x+10.62	y = -1.26+11.61	y = -0.37x+10.01	y = -0.54x+9.49	y = -0.43x+10.51
Corr. R^2	0.8665	0.5800	0.6605	0.3463	0.8446
AOD, δ_a	0.4730	1.2592	0.3717	0.5350	0.4311
Cal. constant, V_o	10.618	11.613	10.098	9.4851	10.509
No. of data, n	47	26	46	44	48
QC algorithm	Y/clear sky	N/cloudy	N/cloudy	N/cloudy	Y/clear sky
500 nm					
Regression line	y = -0.29x+10.47	y = -0.93x+11.18	y = -0.19x+9.97	y = -0.57x+9.88	y = -0.26x+10.38
Corr. R^2	0.7898	0.5514	0.4371	0.3426	0.7575
AOD, δ_a	0.2889	0.9347	0.1899	0.5671	0.259
Cal. constant, V_o	10.471	11.178	9.9723	9.8769	10.382
No. of data, n	47	26	46	44	48
QC algorithm	Y/clear sky	N/cloudy	N/cloudy	N/cloudy	Y/clear sky
670 nm					
Regression line	y = -0.14x+9.50	y = -1.36x+11.39	y = -0.04x+8.96	y = -0.63x+9.31	y = -0.12x+9.43
Corr. R^2	0.7954	0.5906	0.0532	0.3415	0.8063
AOD, δ_a	0.1444	1.3617	0.0436	0.6273	0.1214
Cal. constant, V_o	9.4964	11.387	8.9623	9.3089	9.4307
No. of data, n	47	26	46	44	48
QC algorithm	Y/clear sky	N/cloudy	N/cloudy	N/cloudy	Y/clear sky
870 nm					
Regression line	y = -0.13x+7.17	y = -1.55x+9.48	y = -0.01x+6.61	y = -0.64x+6.96	y = -0.1x+7.17
Corr. R^2	0.7333	0.6165	0.0036	0.3141	0.7432
AOD, δ_a	0.1296	1.552	0.0109	0.6359	0.103
Cal. constant, V_o	7.1677	9.4769	6.6056	6.9616	7.1676
No. of data, n	47	26	46	44	48
QC algorithm	Y/clear sky	N/cloudy	N/cloudy	N/cloudy	Y/clear sky

The second step of the algorithm addresses the saturation and parabolic effects due to unsuitable air mass ranges. Measurements acquired at higher $m>m_{i,j}$ require long exposure times, particularly for filters with low quantum efficiencies and thus, the signal collected can experience significant changes during the collection time. Thus, the signals collected during high air mass occurrences tend to curve the Langley regression into a parabolic arc line that is herein called the parabolic effect. Conversely, the saturation effect causes erroneous signals to be collected during low air mass occurrences, especially for low quantum efficiency filters. Similar findings were also reported in previous studies[5,18]. Day 1 and Day 5 are selected to illustrate the saturation and parabolic effects visually. The untreated Langley plots of Day 1 (upper pane) and Day 5 (lower pane) where the measurements were acquired in the air mass range $2<m_{i,j}<8$ are shown in Fig. 3. In the low air mass region where the intensity of the sun light is high, the measured signal ln V began to be saturated and gives values of almost the same readings. This shows the so-called saturation effect, which is intended to highlight our results[16]. This effect severely inhibits the Langley extrapolation of a robust regression of high R^2 values. The saturation effect is greater in low air mass regions because the changes of the air mass are too small to capture the insignificant increases in solar signal intensities, particularly with low quantum efficiency filters[7]. The preceding effect from the signal saturation effect is a catalyst of the parabolic effect that bends the Langley plot into a parabola-like curve. An even more erroneous parabolic curve is yielded when the data from high air mass occurrences further bend the curve into a quadratic like form. Instead of following a liner regression line, the resultant best-fit line becomes parabola like (Fig. 3).

The third step of the algorithm computes the uncertainty of $V_o \pm \Delta V$ using the post-filtered dataset with the highest R^2 and lowest δ_a. The final product of the algorithm estimates the Langley calibrated constant at a nominal wavelength $V_\lambda \pm \Delta V_\lambda$. The outcome of the AM restriction and final product of the algorithm for $m<8$ to $m<2$ is summarized in Table 2. It was When more stringent air mass ranges [m_i, $m_{i,j}$] were used, V_o tends to be gradually reduced for $\lambda<870$ nm. In addition, higher R^2 and lower δ_a are also expected for stricter $m_{i,j}$, although the pattern is not perfectly consistent for all wavelengths. The results suggest that there are no definite air mass ranges that are fixed for a particular wavelength but that the two indices vary independently. In fact, the dispersity of the calibration parameters R^2 and δ_a are heavily dependent on the manipulation of the Langley dataset on a daily basis. Nevertheless, the variability of V_o is not discretely large until a fictitious extrapolation is suspected. The estimated uncertainty of the accuracy of the Langley calibration at high altitudes ranges from $\pm$0.4% at 670 nm to $\pm$2.3% at 500 nm.

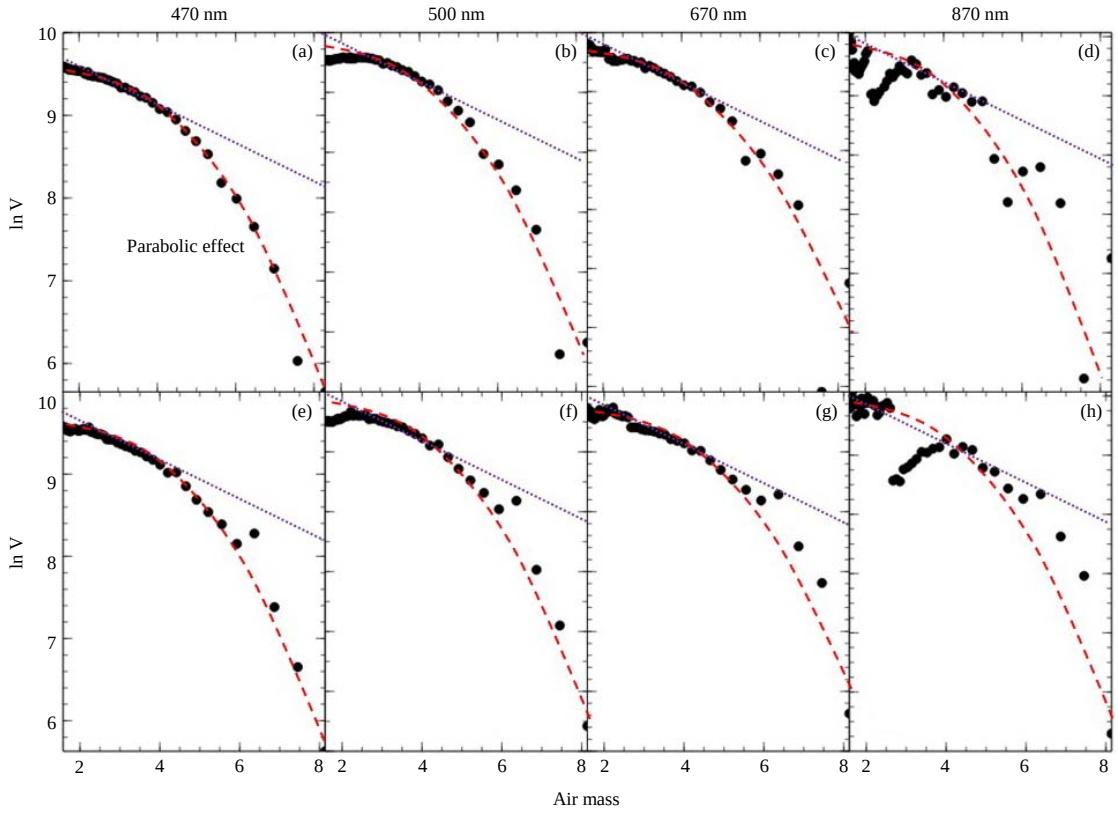

Fig. 3(a-h): Untreated Langley plot on Day 1: (a-d) Upper pane and Day 5: (e-h) Lower pane, where, the measurements were acquired in the air mass range of $2<m_{i,j}<8$. The parabolic effect is represented by the red curving dotted line

Cloud-screening algorithm for sun photometer data at middle and low altitudes: In this section, a new CS algorithm for sun photometer data collected at middle and low altitudes is proposed for use in identifying cloud data to be removed from calibrations to allow for accurate Langley results. The accuracy of the aerosol optical properties heavily depends on the impacts of cloud contamination in the field of view of the instrument and remains one the biggest sources of error in inversion retrievals[4,15]. The details of the Langley sun photometer data at each observatory site in this study are summarized in Table 3. All the sun photometer data were collected using the same units for consistent results. The main difference lies between the observatory sites and the algorithms implemented for these datasets. For data from high altitudes, the QC algorithm was used to select ideal Langley days for the calibration constant computation, which will be used as a reference value (RV). For data from mid or low altitudes, the CS algorithm was used to select the suitable Langley days for calibration. The flow chart of the CS algorithm for sun photometer data from mid/low altitudes proposed in this work is shown in Fig. 2.

The first step of the CS algorithm calculates the clearness index $\varepsilon_{i,j}$ for each data point using Perez's clear-sky classification model. The Perez model defines the discrete sky clearness based on eight categories bounded by a lower limit of 1.0, for completely overcast days and an upper limit of 6.2, for completely clear days. The index is calculated using the relationship between the diffuse and global components of solar irradiance as follows[13,19]:

$$\varepsilon_{i,j} = \frac{\left(\left(\int_{\lambda_i}^{\lambda_j} I_{ed}\,d\lambda + \int_{\lambda_i}^{\lambda_j} I_{dir}\,d\lambda\right) \Big/ \int_{\lambda_i}^{\lambda_j} I_{ed}\,d\lambda\right) + 1.041\varphi_H^{\,3}}{1 + 1.041\varphi_H^{\,3}} \quad (3)$$

where, I_{ed} is the diffuse irradiance component, I_{dir} is the direct irradiance component, $\lambda_{i,j}$ is the spectral range and φ_H is the solar zenith angle in radians.

The second step of the algorithm plots the calculated Perez index as a function of air mass and its diurnal evolution. The Perez index can be interpreted as an indicator of the sky clearness index, where a higher index values represents clearer sky conditions and a lower index exhibits the opposite.

Table 2: Outcome of the QC algorithm for sun photometer data from high altitudes

	Day 1			Day 5			Evaluation*			
	R^2	δ_a	V_o	R^2	δ_a	V_o	R^2	δ_a	V_o	$V_o \pm \Delta V$
440 nm										
$m_{i,j}<8$	0.8665	0.4730	10.168	0.8446	0.4311	10.509	N	N	N	9.82±0.12
$m_{i,j}<7$	0.8949	0.3604	10.321	0.8993	0.3162	10.210	N	N	N	
$m_{i,j}<6$	0.9138	0.2702	10.097	0.9124	0.2757	10.109	N	N	N	
$m_{i,j}<5$	0.9471	0.2266	9.9946	0.9170	0.2178	9.9738	Y	N	Y	
$m_{i,j}<4$	0.9753	0.1746	9.8794	0.9146	0.1605	9.848	N	N	N	
$m_{i,j}<3$	0.9806	0.1403	9.8100	0.7770	0.1094	9.7457	Y	Y	Y	
$m_{i,j}<2$	0.9342	0.1629	9.8021	0.3791	0.0812	9.6869	N	Y	Y	
500 nm										
$m_{i,j}<8$	0.7898	0.2889	10.471	0.7575	0.2550	10.382	N	N	N	9.97±0.23
$m_{i,j}<7$	0.8041	0.2002	10.238	0.7930	0.1674	10.154	Y	N	Y	
$m_{i,j}<6$	0.7469	0.1318	10.068	0.7655	0.1338	10.070	N	N	N	
$m_{i,j}<5$	0.7382	0.0937	9.9789	0.6591	0.0858	9.9578	N	N	N	
$m_{i,j}<4$	0.5232	0.0399	9.8599	0.2651	0.0301	9.8359	N	N	N	
$m_{i,j}<3$	0.2128	0.0103	9.7583	0.2630	0.0288	9.7177	N	Y	Y	
$m_{i,j}<2$	0.7140	0.0372	9.7189	0.6975	0.0633	9.6493	N	N	N	
670 nm										
$m_{i,j}<8$	0.7954	0.1444	9.4964	0.8063	0.1214	9.4307	N	N	N	9.28±0.04
$m_{i,j}<7$	0.9105	0.1033	9.3892	0.9225	0.0856	9.3374	N	N	N	
$m_{i,j}<6$	0.8885	0.0835	9.3401	0.9279	0.0774	9.3169	Y	N	Y	
$m_{i,j}<5$	0.9561	0.0679	9.3032	0.9060	0.0652	9.2884	Y	N	Y	
$m_{i,j}<4$	0.9330	0.0576	9.2803	0.8008	0.0545	9.2650	N	N	N	
$m_{i,j}<3$	0.8226	0.0559	9.2768	0.4446	0.0407	9.2375	N	Y	Y	
$m_{i,j}<2$	0.8728	0.1053	9.3366	0.1061	0.0339	9.2184	N	Y	Y	
870 nm										
$m_{i,j}<8$	0.7333	0.1296	7.1677	0.7432	0.1030	7.1676	Y	N	Y	7.09±0.09
$m_{i,j}<7$	0.7189	0.0958	7.0796	0.7217	0.0746	7.0933	N	N	N	
$m_{i,j}<6$	0.5306	0.0807	7.0422	0.6394	0.0737	7.0910	N	N	N	
$m_{i,j}<5$	0.3437	0.0463	6.9614	0.5304	0.0768	7.0981	N	N	N	
$m_{i,j}<4$	0.1650	0.0432	6.9547	0.5875	0.1157	7.1822	N	Y	Y	
$m_{i,j}<3$	0.2485	0.0897	7.0464	0.5087	0.1564	7.2656	N	N	N	
$m_{i,j}<2$	0.6552	0.3312	7.3963	0.1113	0.0589	7.0754	N	Y	Y	

*Evaluation is based on the highest R2 and lowest δa on a daily basis to select the Vo for the final product of $V_o \pm \Delta V$. Bold figures represent the selected V_o values

Table 3: Details of each observatory site for Langley sun photometer measurements

Site	Site A (High)	Site B (Mid)	Site C (Low)
Altitude	3,270 m a.s.l.	1,574 m a.s.l.	18 m a.s.l.
Location	Laban rata	Kinabalu park	Sepanggar, UMS
Coordinate	6.05°N, 116.56 °E	6.00°N, 116.54°E	6.03°N, 116.12°E
Measurement date (dd-mm-yy)	22/01/16-27/01/16	26/08/15-30/08/15	13/01/16-02/03/16
Measurement period (h)	2-3	2-3	2-3
Measurement Interval (min)	3	3	3
Air mass range	1.0-8.0	1.0-6.0	1.0-4.0
Solar zenith angle (deg)	46-86	42-80	40-77
QC algorithm	Y	N/A	N/A
CS algorithm	N/A	Y	Y

Therefore, plotting the index as a function of airmass renders a picture of the stability of the atmospheric conditions during the Langley measurements. The daily diurnal evolution of the Perez index plotted against the air mass variations for sun photometer data at low (upper pane) and mid (lower pane) altitudes is shown in Fig. 4. In general, two types of evolution patterns are observed: (a) A gradual evolution pattern or a (b) Punctuated evolution pattern. In the context of this study, the increment of the Perez index against the air mass was the determining factor that characterizes whether the pattern was gradual or punctuated. The gradual evolution pattern was characterized as a steady evolution where the increment of change is consistent with time. For example, this type of pattern was best illustrated on Day 2, when the diurnal change of the Perez index as a function of air mass was consistent and regular. Conversely, a punctuated evolution pattern is a

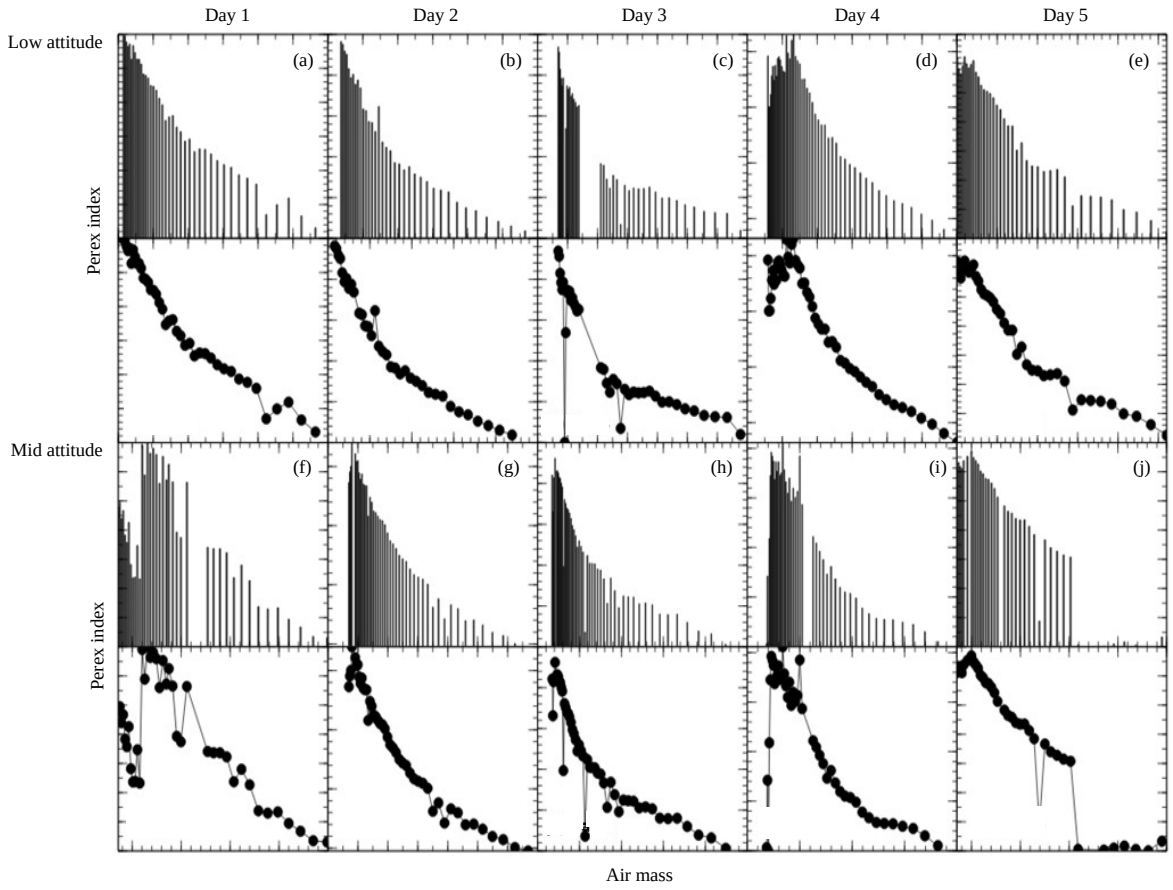

Fig. 4(a-j): Daily diurnal evolutions of the Perez index plotted against the air mass in the box chart and scatter plot for sun photometer data taken at low altitudes (a-e) Upper pane and mid altitudes (f-j) Lower pane

non-steady evolution where the increment of change is not consistent with time and often there is virtually no change at all (Fig. 5). The characterization of the Perez index pattern provides an objective way to identify and filter potential contaminated data from Langley plots. The identification was based on the reasoning that a perfect Langley plot should exhibit an ideally gradual evolution pattern that has no negative derivatives at any time within the measurement period. This would mean than any instances of data that have negative derivatives are likely to be contaminated by cloud cover, aerosol loading, or unstable atmospheric turbidity. Another similar finding from previous studies also suggested that the cause and necessity of AOD reprocessing in Langley plots were due to the existence of an observed fictitious diurnal cycle (including negative values)[20]. Using this rule, several potentially contaminated data points have been identified on each Langley plot.

To visualize this filtration procedure, Day 5 of the mid-altitude measurements was selected as an example. (a) The original Langley plot and (b) The improved Langley plot at 500 nm are shown in Fig. 5a-b. Data P1 was the initial point, so its derivative Perez index was unable to be determine. Data P2 was the second data point and its derivative Perez index (-0.14) (Fig. 6b) was calculated with respect to its preceding point, which is P1. For the next data points (P3-P6), the respective derivative values were still calculated with respect to P1. This sequence was continued until a positive derivative was obtained, in this case, the sequence was stopped at P7. Thereafter, the calculation of the derivative value for P8 follows the normal sequence with respect to the preceding point. A similar practice was followed whenever a negative derivative is obtained. For example, the next negative derivative value lies at P12, hence, the derivative value for the data point P13 was calculated with respect to that of P11 instead of its preceding point. Finally, all data with negative derivative values were identified and filtered.

The final step of the CS algorithm is to retrieve the V_o using the improved Langley plot from Step 2, which is done by extrapolating the regression to a zero air mass and finding the intercept. Table 4 summarizes the final product of the CS

Table 4: Final product of the cloud-screening algorithm for sun photometer data from low- and mid-altitude measurement sites

		Original langley			Improved langley			CS algorithm analysis		
		R^2	δ_a	V_o	R^2	δ_a	V_o	n	m	f
Low altitude										
Day 1	440 nm	0.9139	0.6312	10.472	0.9274	0.5519	10.315	41	32	0.78
	500 nm	0.8717	0.4106	10.403	0.8733	0.3502	10.285			
	670 nm	0.8708	0.2588	9.5197	0.8790	0.2275	9.4654			
	870 nm	0.7799	0.3364	7.4898	0.7749	0.3122	7.4603			
Day 2	440 nm	0.9750	0.5315	10.198	0.9750	0.5216	10.178	36	27	0.75
	500 nm	0.9605	0.3443	10.215	0.9564	0.3371	10.200			
	670 nm	0.9809	0.2128	9.3712	0.9811	0.2167	9.3836			
	870 nm	0.7411	0.1558	6.8996	0.8361	0.1801	6.9723			
Day 3	440 nm	0.8814	0.9091	10.633	0.9886	0.8342	10.541	40	24	0.60
	500 nm	0.8232	0.6684	10.580	0.9857	0.6079	10.520			
	670 nm	0.6861	0.4716	9.6598	0.9768	0.4201	9.6282			
	870 nm	0.6038	0.4041	7.3743	0.9216	0.3542	7.3376			
Day 4	440 nm	0.8916	0.6529	10.280	0.9642	0.7606	10.627	51	29	0.57
	500 nm	0.8360	0.4380	10.279	0.9557	0.5365	10.593			
	670 nm	0.8573	0.2789	9.428	0.9627	0.3320	9.598			
	870 nm	0.7162	0.2125	7.1194	0.8498	0.2731	7.3019			
Day 5	440 nm	0.9495	0.7935	10.609	0.9817	0.8132	10.712	39	27	0.69
	500 nm	0.9351	0.5655	10.575	0.9794	0.5857	10.671			
	670 nm	0.9149	0.3611	9.6051	0.9886	0.3778	9.6813			
	870 nm	0.8186	0.3363	7.3965	0.9180	0.3531	7.4812			
Mid altitude										
Day 1	440 nm	0.6629	1.1584	10.226	0.9232	1.7413	11.890	40	18	0.45
	500 nm	0.6204	1.0175	10.367	0.9143	1.6046	12.027			
	670 nm	0.5750	0.8816	9.4625	0.9112	1.4927	11.145			
	870 nm	0.5924	0.9299	7.0681	0.8964	1.5497	8.7742			
Day 2	440 nm	0.9525	0.4052	10.181	0.9673	0.4029	10.201	46	36	0.78
	500 nm	0.8810	0.2454	10.198	0.9191	0.2459	10.224			
	670 nm	0.9026	0.1677	9.4070	0.9716	0.1634	9.4109			
	870 nm	0.6407	0.1384	6.9829	0.6717	0.1146	6.9189			
Day 3	440 nm	0.7549	0.6283	10.546	0.9170	0.5338	10.404	50	37	0.74
	500 nm	0.6719	0.4530	10.533	0.8726	0.3735	10.423			
	670 nm	0.5351	0.3279	9.5963	0.8445	0.2611	9.5244			
	870 nm	0.3242	0.2999	7.1715	0.7585	0.2508	7.1596			
Day 4	440 nm	0.5212	0.7995	10.315	0.9777	1.0730	11.275	45	21	0.47
	500 nm	0.4151	0.6648	10.424	0.9769	0.9168	11.321			
	670 nm	0.3011	0.5310	9.4848	0.9678	0.7585	10.295			
	870 nm	0.3925	0.5936	7.3792	0.9241	0.7849	8.0969			
Day 5	440 nm	0.6023	2.4266	13.446	0.7350	0.4343	10.050	38	29	0.76
	500 nm	0.5635	2.3596	13.666	0.5153	0.2814	10.214			
	670 nm	0.5734	2.1850	12.707	0.3911	0.2086	9.3421			
	870 nm	0.5701	1.7510	9.5403	0.0249	0.0563	6.6669			

n is the number of data points in the original Langley, m is the number of data points in the improved Langley, f is the ratio of m and n

algorithm for sun photometer data at low and mid-altitude measurement sites. Note that after the CS algorithm, the improved Langley showed a better R^2 and lower δ_a for all wavelengths regardless of altitude. The CS algorithm objectively removes cloudy data, assuming that the ever-rising calculated Perez index is expected for clean and clear-sky condition. The resulting regression thereby showed a better correlation and reduced slope of the incline. However, for cases when the opposite pattern is observed, with poorer R^2 and higher δ_a after the CS algorithm, the resultant improved Langley plot was unrealistic. For instance, a fictitious regression was observed on Day 5 of the mid-altitude measurement at 870 nm where, $R^2 = 0.5701$ was observed in the original Langley analysis but an even lower $R^2 = 0.0249$ was observed in the improved Langley. In this case, that dataset should be discarded to avoid inaccurate results.

Assessment with high-altitude measurement: In this section, the capability of the CS algorithm for sun photometer data collected at middle and low altitudes to reproduce the V_o that was obtained at a higher altitude. This section presents the calibration constant obtained from the original Langley

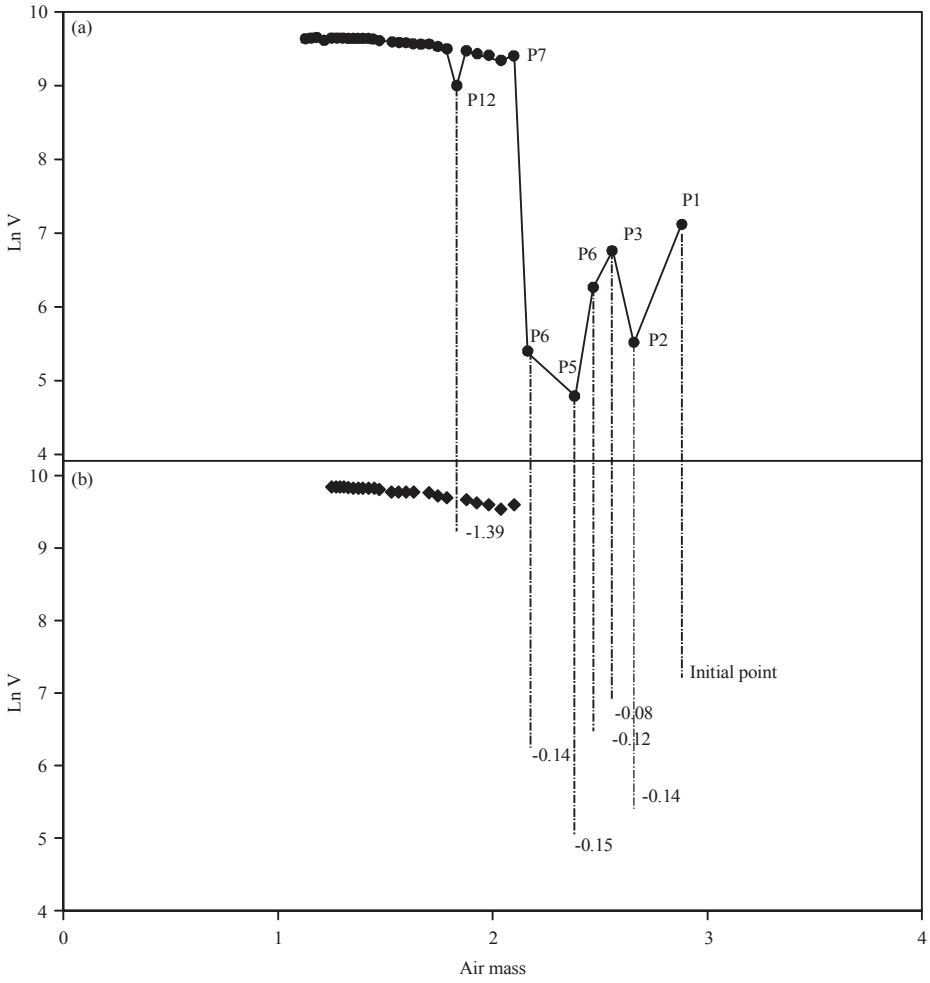

Fig. 5: Langley plot at 500 nm on Day 5 of the mid-altitude measurements (a) Before filtration and (b) After filtration
Values presented in the lower figure are the derivatives of the Perez index at distinct airmasses, Data with negative derivatives are filtered

method, $V_{o,ori}$ and the improved Langley method, $V_{o,imp}$ and compares them with the reference calibration constant, $V_{o,ref}$, obtained from high-altitude measurements. The diurnal variations of $V_{o,ori}$ and $V_{o,imp}$ and their comparisons with $V_{o,ref}$ at all wavelengths for the low and mid-altitude measurements are represented in Fig. 6a. Major discrepancies between $V_{o,ori}$ and $V_{o,imp}$ were observed on days 1, 4 and 5 and mid and minor discrepancies were observed on days 4 and 5 of the low-altitude measurements. This pattern was more visually evident when one examines the clustered column in Fig. 6b, which was calculated by finding the ratio $|V_{o,ori}-V_{o,imp}|$ against the reference calibration constant, $V_{o,ref}$. This was noted that the ratio, ψ, can be used as an objective indicator for determining suitable Langley days, where higher indexes indicate poor datasets and lower indexes demonstrate the opposite. When $V_{o,ori}$ and $V_{o,imp}$ are nearly identical and ψ is close to zero, the sky conditions were presumed to be ideally pristine and cloud-free. For instance, Day 2 of both the low and mid-altitude measurements have ratios of ψ close to zero, indicating that the datasets have little to insignificant effects from cloud contamination and very stable atmospheric conditions throughout the Langley measurements. Under pristine and clean environmental conditions, the calibration constants of the instruments estimated from the original Langley method agreed well with those of the improved Langley method. In addition, both types of values were close to the reference calibration constant, with errors as low as ~0.37 (3.8%) at 470 nm, ~0.24 (2.4%) at 500 nm, ~0.12 (1.3%) at 670 nm and ~0.14 (2.0%) at 870 nm.

Although the proposed CS algorithm was useful for identifying suitable Langley days for calibration, its application for reproducing the Langley results for a given dataset that is originally poor is implausible. The mid-altitude measurements on days 1, 4 and 5 are perfect examples to illustrate this

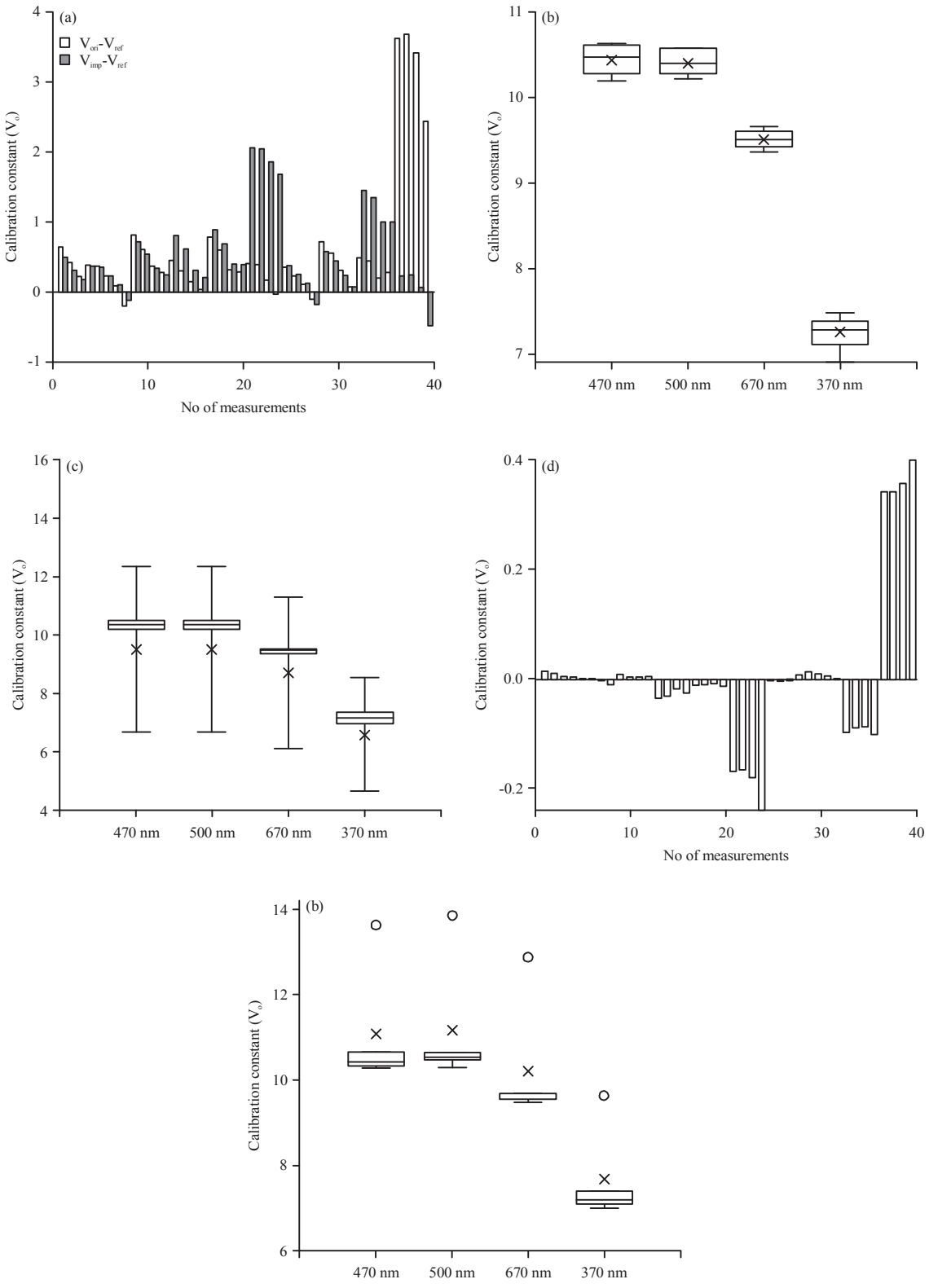

Fig. 6: Comparison between the calibration constants obtained from the original Langley method, $V_{o,ori}$ and the improved Langley method, $V_{o,imp}$, against the reference calibration method, $V_{o,ref}$.

limitation. Despite using the improved Langley plot, the obtained calibration constant was still far beyond the RV and the errors were as high as ~1.25 (12.7%) at 470 nm, ~1.22 (12.2%) at 500 nm, ~0.98 (10.57%) at 670 nm and ~0.76 (10.66%) at 870 nm. Note that Day 1 of the mid-altitude measurements actually exhibited remarkably high R^2 values after the CS algorithm is applied to all wavelengths, but the calibration constants were highly erroneous. In other words, these results tell us that the high correlations in Langley plots do not guarantee clean and clear-sky conditions at high altitudes[14]. In fact, the combined conditions of high R^2 and low δ_a values were the preferred conditions for an ideal Langley plot (Table 4). The diurnal variations of the aerosol optical depths and Angstom's exponent can be significant in low AOD conditions, while those of other AODs are negligible[6]. The calibration constant obtained on Day 2 of both the low and mid-altitude measurements, for instance, measured remarkably high R^2>0.84 and low δ_a<0.50 values for all wavelengths and agreed well with the reference calibration values with errors of less than <4.0%.

The variation of V_o obtained from the data from low and mid altitudes are shown in Fig. 6c and d, respectively. It was observed that V_o varies significantly at both the 470 and 500 nm wavelengths, which could be due to the quantum efficiencies of the filters used. The box plot also indicates that a comparison between low and mid altitudes shows that the latter demonstrates a poor consistency with $V_o$and significant outliers are detectable at all wavelengths. The result agrees well with the ψ interpretation that states that the higher index indicates a poor dataset where mid altitude sun photometer data exhibits higher ψ values than those from low altitudes. In tropical climate regions, bad Langley days are unavoidable due to the regular cloud loading that is abundant but can be objectively identified and removed from the calibrations. To account for this discrepancy, the proposed CS algorithm was implemented and the output revealed that days 1, 2 and 3 of the low-altitude measurements and days 2 and 3 of the mid-altitude measurements have the potential to reproduce the V_o that was obtained at high altitudes. Figure 6e shows the box plot of the final product of V_o after the CS algorithm was applied. The uncertainty analysis shows that the variations of $\Delta V_o$are considerably small at all wavelengths, suggesting that the algorithm is capable of producing consistent results, provided that enough data are collected. The lowest ΔV_o was measured at the wavelength 670 nm (0.10), followed by those of the 500 nm (0.14), 470 nm (0.17) and 870 nm (0.22) wavelengths. Under pristine and clean atmospheric conditions, the proposed algorithm was able to reproduce V_o values compatible with those of high-altitude measurements, with acceptable errors ranging from 1.27% at 870 nm to 5.60% at 470 nm (Fig. 6e).

CONCLUSION

This analysis of the 60 Langley plots after implementing the two proposed QC and CS algorithms indicates that the algorithms provide comparable results despite varying altitude measurements. In the QC algorithm protocol, there is no fixed air mass range defined for any particular wavelength. The CS algorithm objectively removes cloudy data by examining on the evolution pattern of the calculated Perez index. It is assumed that an increasing pattern of the calculated Perez index was expected for clean and clear-sky conditions. Although, the proposed CS algorithm was useful for identifying suitable Langley days for calibrations, its ability to reproduce ideal Langley results for a given dataset that was originally poor is implausible. The results highlight that high correlations in Langley plots do not guarantee clean and clear-sky conditions at high altitudes. In fact, the combined conditions of high R^2 and low δa values are the preferred conditions for an ideal Langley plot. On a point-by-point observational basis, the calibration constant of the sun photometer maintains a difference of less than 0.09, depending on the wavelength. This study concludes that both algorithms can be useful for improving the state-of-the-art well-known Langley calibration method in terms of both reliability and reproducibility.

SIGNIFICANCE STATEMENTS

In this study, a novel quality control and cloud-screening algorithm for sun photometer data was developed to improve the state-of-the-art well-known Langley calibration method. A Langley plot is a conventional calibration method that requires stable atmospheric conditions and clear sky condition. For this reason, it is typically performed at high altitudes. This work shows that the Langley calibration at less ideal sites is feasible provided that an objective quality control and cloud-screening algorithm is implemented to select suitable solar calibration data. The outcome of this study will help researchers recommend alternative approaches to calibrate sun photometers when frequent calibrations at high altitudes are unfeasible. Thus, a new theory on sun photometry calibration may be developed.

ACKNOWLEDGMENT

This study was supported by the Malaysian Ministry of Education under the research grant number RAG0071-SG-2015 and this help is gratefully acknowledged.

REFERENCES

1. Guleria, R.P. and J.C. Kuniyal, 2016. Characteristics of atmospheric aerosol particles and their role in aerosol radiative forcing over the Northwestern Indian Himalaya in particular and over India in general. Air Qual. Atmos. Health, 9: 795-808.
2. Tan, F., H.S. Lim, K. Abdullah, T.L. Yoon and B. Holben, 2015. Monsoonal variations in aerosol optical properties and estimation of aerosol optical depth using ground-based meteorological and air quality data in Peninsular Malaysia. Atmos. Chem. Phys., 15: 3755-3771.
3. Nirmalkar, J. and M.K. Deb, 2016. Impact of intense field burning episode on aerosol mass loading and its possible health implications in rural area of Eastern Central India. Air Qual. Atmos. Health, 9: 241-249.
4. Li, X., L. Zhang, X. Cao, J. Quan, T. Wang, J. Liang and J. Shi, 2016. Retrieval of precipitable water vapor using MFRSR and comparison with other multisensors over the semi-arid area of Northwest China. Atmos. Res., 172-173: 83-94.
5. Nieke, J., B. Pflug and G. Zimmermann, 1999. An aureole-corrected Langley-plot method developed for the calibration of HiRES grating spectrometers. J. Atmos. Solar-Terr. Phys., 61: 739-744.
6. Kreuter, A., S. Wuttke and M. Blumthaler, 2013. Improving Langley calibrations by reducing diurnal variations of aerosol Angstrom parameters. Atmos. Meas. Tech., 6: 99-103.
7. Li, Z., P. Goloub, O. Dubovik, L. Blarel and W. Zhang *et al*., 2009. Improvements for ground-based remote sensing of atmospheric aerosol properties by additional polarimetric measurements. J. Quant. Spectrosc. Radiat. Transfer, 110: 1954-1961.
8. Ul-Saufie, A.Z., A.S. Yahaya, N.A. Ramli, N. Rosaida and H.A. Hamid, 2013. Future daily PM_{10} concentrations prediction by combining regression models and feedforward backpropagation models with Principle Component Analysis (PCA). Atmos. Environ., 77: 621-630.
9. Latif, M.T., D. Dominick, F. Ahamad, M.F. Khan, L. Juneng, F.M. Hamzah and M.S.M. Nadzir, 2014. Long term assessment of air quality from a background station on the Malaysian Peninsula. Sci. Total Environ., 482-483: 336-348.
10. Aouizerats, B., G.R. van der Werf, R. Balasubramanian and R. Betha, 2015. Importance of transboundary transport of biomass burning emissions to regional air quality in Southeast Asia during a high fire event. Atmos. Chem. Phys., 15: 363-373.
11. Kanniah, K.D., D.G. Kaskaoutis, H. San Lim, M.T. Latif, N.A.F.K. Zaman and J. Liew, 2016. Overview of atmospheric aerosol studies in Malaysia: Known and unknown. Atmos. Res., 182: 302-318.
12. Azmi, S.Z., M.T. Latif, A.S. Ismail, L. Juneng and A.A. Jemain, 2010. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. Air Qual. Atmos. Health, 3: 53-64.
13. Chang, J.H.W., J. Dayou and J. Sentian, 2014. Development of a Near-Sea-Level calibration algorithm for aerosol optical depth measurement using a ground-based spectrometer. Aerosol Air Qual. Res., 14: 386-395.
14. Guerrero-Rascado, J.L., M.J. Costa, A.M. Silva and F.J. Olmo, 2013. Retrieval and variability analysis of optically thin cloud optical depths from a Cimel sun-photometer. Atmos. Res., 127: 210-220.
15. Holben, B.N., T.F. Eck, I. Slutsker, T. Tanre and J.P. Buis *et al*., 1998. AERONET-A federated instrument network and data archive for aerosol characterization. Remote Sens. Environ., 66: 1-16.
16. Chang, H.W.J., N.H.N. Maizan, F.P. Chee, J. Sentian, S.S.K. Kong and J. Dayou, 2016. Saturation and parabolic effects of langley calibration at different altitude levels. ARPN J. Eng. Applied Sci., 11: 12681-12689.
17. Rahman, S.R.A., S.N.S. Ismail, M.F. Raml, M.T. Latif, E.Z. Abidin and S.M. Praveena, 2015. The assessment of ambient air pollution trend in Klang valley, Malaysia. World Environ., 5: 1-11.
18. Reynolds, R.M., M.A. Miller and M.J. Bartholomew, 2001. Design, operation and calibration of a shipboard fast-rotating shadowband spectral radiometer. J. Atmos. Oceanic Technol., 18: 200-214.
19. Perez, R., R. Ineichen, R. Seals, J. Michalsky and R. Stewart, 1990. Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy, 44: 271-289.
20. Cachorro, V.E., C. Toledano, M. Sorribas, A. Berjon, A.M. de Frutos and N. Laulainen, 2008. An "*in situ*' calibration-correction procedure (KCICLO) based on AOD diurnal cycle: Comparative results between AERONET and reprocessed (KCICLO method) AOD-alpha data series at El Arenosillo, Spain. J. Geophys. Res.: Atmos., Vol. 113. 10.1029/2007JD009001.