

Interactive comment on “Non-parametric and least squares Langley plot methods” by P. W. Kiedron and J. J. Michalsky

P. W. Kiedron and J. J. Michalsky

joseph.michalsky@noaa.gov

Received and published: 19 November 2015

Comments and Responses to Reviewer #3

Reviewer #3: From my point of view this work should be considered as a review of several methods for obtaining Langley plot calibration for Sun photometers. This also because, as expected and recognized by the authors, is not possible to give an answer on which method is the most appropriate.

Response: The word “review” in the above statement implies existing and known methods. Except for LSF and OA methods the remaining techniques that we compared are non-parametric methods that to our knowledge have not been used in Langley plots. The non-parametric methods like the ones by Thiel (1950) and Siegel (1982) are de-

C3959

scribed in the literature. They provide slope and intercept in every case, but they do not define outliers. Without identified outliers one cannot apply metrics such as RMS to determine the quality of the Langley plot. So we had to supplement Thiel (1950) and Siegel (1982) methods with procedures of outlier identification. We proposed two novel methods: Sequential Removal of Outliers (SRO) from dispersion of slopes and Outlier Sorting Method (OSM). Both identify outliers and permit one to calculate RMS of residuals of the points defining the Langley plot. We also introduced a new histogram method that is very efficient for extracting Langley plots in cases when the number of outliers is very large. For these reasons we believe this paper should not be considered as a review.

Reviewer #3: In some cases it's difficult to follow the different steps of a specific method, but appropriate references are provided to whom who want to deeply understand and/or apply it. Maybe, more examples of the application of the different methods to specific cases, as in Figure 1, could help the reader to quickly have an idea on the usefulness of them.

Response: Fig. 1 illustrates the proposed histogram method and the potential that histograms can have in analysis of Langley plots. However we cannot think of equivalent illustrations for other methods. For all methods (except for OA and LSF) formulas are given with explanations that, we believe, are sufficient to replicate and implement methods demonstrated in our work.

Reviewer #3: Specific Comments

Response: In all cases indicated by Reviewer #3 we agree that the difficult to follow parts in our text needed a rewrite to make our intended meaning clearer and more intelligible. Some of these parts are difficult because we tried to convey in the text mathematical operations w/o invoking complex formulas that would further complicate and muddle the text. We hope that our rewrites (see below) make the text better.

Reviewer #3: Page 4201, lines 13-16: this is an example of what I said in the general

C3960

comment. I can't follow the steps.

Response: Lines 13-19 were replaced with the following paragraph:

For a given α and β we sort all points according to the ascending value of their residuals ($r_j \leq r_{j+1}$). Then we calculate the root mean square of residuals (rms_J) of the first $j = 0, \dots, J-1$ points. We find J for which $rms_J \leq rms_{max}$ and $rms_{J+1} > rms_{max}$. All points with indices $j \geq J$ are considered outliers. Points with indices $j < J$ are retained. They define the Langley plot, which is considered to be successful when $J \geq n/3$. Residuals of retained points can be de-biased by performing a LSF on them. This reduces the value of rms_J and slightly changes α and β .

Reviewer #3: Page 4203, lines 6-20: another series of steps difficult to follow.

Response: Lines 6-22 were replaced with this paragraph:

In this method we analyze a histogram of slopes to identify the subset of nearly collinear points. The histogram of slopes is constructed from elements of the matrix $\{b_i, j\}$. We locate the cell $[b, b+\Delta b]$ at which the histogram's counts are maximum, where Δb denotes the width of histogram cells. Next we identify all pairs of points $[(y_i, x_i), (y_j, x_j)]$ that produce slope b_i, j from within the interval $[b, b+\Delta b]$. We are interested in points (y_i, x_i) that create many such pairs. Let c_i denote the number of such pairs that the i -th point creates. The points for which $c_i = 1$ are rejected, and the median of remaining c_i is calculated. Then the points with c_i less than the median are rejected. LSF is applied to the remaining (y_i, x_i) points and the rms is calculated. This method defines the Langley plot with a small number of points, but is very efficient at detecting subsets of collinear points. For this reason we did not require that the number of retained points had to be larger than $n/3$ for this method. We labeled this method H- β .

Response: We also shortened the text in line 23 p. 4203 to line 4 p. 4204 to these two sentences:

C3961

A similar process was tried with the histogram of intercepts. The results, however, were not as good as with the histogram of slopes.

Reviewer #3: Figure 2 and discussion at page 4206, lines 18-26: again, I can't follow the discussion and the value of such a plot.

Response: Lines 18 to 28 are replaced with the following text:

In Fig. 2 we demonstrate the effect of the number of outliers on the differences among the methods. The average of the standard deviations of $\Delta\alpha$ for all methods is plotted against the percentage of points retained by Langley methods. When Langley plots have no more than 10% outliers, the standard deviations between the methods are an order of magnitude smaller than when the number of outliers is up to 67%. This implies that the main differences between methods are due to different handling of outliers by each method and the outcome is more method dependent when the Langley plot consists of a smaller number of points. Also, we plotted the number of Langley plots vs. the number of points remaining in the Langley plot for the LSF_SRO-x method to demonstrate how strongly the number of available Langleys diminishes with the number of outliers for the site in Oklahoma.

Reviewer #3: Page 4208, lines 1-6: I can't understand the definitions of $acc(d)$ and $a(d_j)$. Why do you add the subscript j ? Is it necessary?

Response: The subscript "j" indicates days for which we have Langleys while d w/o a subscript indicates any day in between the dates within which we try to get calibration constants. So d_j 's are the subset of d 's. Within the lines 1-6 we added explanation. See addition in brackets [] below:

We compare the behavior of time series of derived calibration constants $\alpha_{cc}(d)$, where d indicates each day from 5 October 2003 to 30 March 2006. The $\alpha_{cc}(d)$ are obtained from the time series of $\alpha(d_j)$ independently for each Langley method[, where d_j denotes days at which Langleys were obtained]. The $\alpha_{cc}(d)$ might be considered "the

C3962

best” estimate of the calibration constant for a given day, “the best” in a sense of the method that we use to remove outliers, interpolate and smooth the series $\alpha(d_j)$.

Reviewer #3: Page 4208, line 8: please check "by a moving a boxcar filter".

Response: We replaced “which is followed by a moving a boxcar filter” with “which is followed by a moving boxcar filter”.

Reviewer #3: In the acknowledgements, please check "Centreto"

Response: We changed “Centreto” to “Center”.

Comments and Responses to Reviewer #4

Reviewer #4: The authors describe 11 different mathematical procedures or methods based on the Classical Langley Plot method for radiometer calibration in field sites (or in situ) in order to determine the best Langley plot or the best method (between them) that obtains the best calibration. The general topic of this work is of general interest for radiometer calibration but as the authors mentioned it is not possible to give an answer on which method is the most appropriate because the results demonstrated that there is not a significant difference between them in order to determine the calibration constants when they are smoothed or interpolated.

Response: This assessment is correct when we look at the methods by comparing filtered and smoothed time series of intercepts, where we use a wide averaging window that is necessitated by a large standard deviation of time series of ε at the site in Oklahoma. On the other hand, Table 1 shows relative differences in terms of bias, standard deviation and number of Langleys between methods.

Reviewer #4: Certainly I have not major objection about the paper, as a whole is well structured and written, although some points are difficult to follow for the detailed mathematical description. Also Figure 2 is not easy to understand.

Response: We changed the description of Fig. 2 on page 4206 in response to Re-

C3963

viewer #3 above.

Reviewer #4: What I can not derive of the paper is if the authors do a Langley plot for every day of measurements or they put some criteria to select special days where apparently the AOD might be a constant. This point must be said explicitly.

Response: We applied Langley plot method to every day except for the overcast days. (see section 8 The Data Set). The overcast days were detected by lack of signal in the direct irradiance. No additional information from any ancillary measurements was used. So, obviously we had no knowledge which days or periods of times may have had a constant AOD. When one has such information, the calibration problem becomes a trivial one as it excludes anomalous Langleys. Then the intercept obtained from a single Langley plot exactly equals the calibration constant.

When responding to the issue of constancy of AOD brought up by the reviewer we realized that our introduction should be improved so there are no ambiguities or misunderstandings regarding the objective of our work. In this version we stress the central role of anomalous Langley events in the calibration process.

The following text replaces lines 7-29 on page 4194 and 1-7 on page 4195:

The process of removing outliers from the Langley plot may actually facilitate selecting a straight line from the data that will contain a spurious value of ε . Different methods of removing outliers may cause an inadvertent selection of a different value of ε , some larger and some smaller (positive or negative). We will call a Langley plot with a nonzero ε an anomalous Langley plot. The anomalous Langley plot cannot be identified because data (V_i, m_i) do not contain information on the presence of a nonzero ε . Statistical analysis of the time series of intercepts $\ln(V_0) - \varepsilon$ leads to a better estimate of the calibration constant $\ln(V_0) = \langle \ln(V_0) - \varepsilon \rangle = \ln(V_0) - \langle \varepsilon \rangle$, where $\langle \rangle$ denotes average used in time series analysis that usually is a combination of mean and median moving averaging windows. When the statistics of ε is unbiased $\langle \varepsilon \rangle$ tends to zero as the number of samples in the time series increases. Mountaintops like Mauna Loa in

C3964

Hawaii or Izaña on Tenerife provide environments where the constancy of the optical depth ($\varepsilon = 0$) is frequent. With a small standard deviation $SD(\varepsilon)$ at mountain tops a lower number of Langley's is necessary to achieve the desired precision of calibration. A long history of measurements at sites like Mauna Loa also warrants the belief that the statistics of ε there is considered to be unbiased. Thus, within the range of validity of this belief an accurate calibration is possible.

In most places where sun-photometers are deployed, periods of stable atmospheres are much less common, and they are frequently interrupted by cloud passages, changes in atmospheric conditions like varying humidity that promulgate aerosol size changes, and by aerosol plume incursions. Large numbers of outliers call for special Langley plot analyses going beyond standard straight line fitting procedures. The statistics of ε is likely to be biased with a large standard deviation $SD(\varepsilon)$. This necessitates a larger number of points in the time series to achieve a desired precision while the frequency of Langley events at sites like these is low. There is no guarantee that the statistics of ε is unbiased. It should be emphasized that the difference between easy sites like mountaintops and difficult sites like Billings, Oklahoma, is quantitative not qualitative. The same statistical analysis of time series must be applied in both cases, however precision and accuracy of results will differ. When there are no other available independent measurements, time series analysis is the only option for in situ calibration of sun photometers. The photometers that can also measure aureole radiance simultaneously with direct solar flux can be calibrated when optical depth is not constant (Tanaka et al., 1986; Nieke et al., 1999; Zieger et al., 2007). These methods can identify anomalous Langley events and estimate the value of ε .

The main objective of the paper is to analyze the efficacy of non-parametric and least squares methods of straight line fitting to identify Langley plots useful for calibration. We use time series analysis only to determine the impact of the methods on the estimated calibration constant. While we are not concerned with the identification of anomalous Langley plots, their presence is manifested in outliers of the time series.

C3965

We do not deal with issues of ambiguities related to the definition of the air mass when various air constituents, aerosols, in particular, are present and higher order effects like atmospheric refraction's dependence on wavelength impact air mass. In other words, we presume that the BLB as given by Eq. (1) is valid. Our comparisons use real data from a real sun radiometer that contains data departing from the BLB law chiefly because air masses of Rayleigh scattering and aerosols are different, and the latter is usually known only approximately.

Reviewer #4: My major objection is related with the calibration methodology itself. At the end the authors develop or analyse all these procedures in order to get the best possible calibration method based on the application of the Langley Plot Method. They have chosen the Classical Langley Plot Method (avoiding other related methods that do not meet their approaches) where the main assumption is the constancy of the aerosol optical depth (AOD) in the morning or the afternoon. However, they have forgotten to reference the works of Cachorro et al., 2008a,b, c; and 2004 (and some other references inside), that in my opinion contain one of the most advanced procedure (they named KCICLO method) about this problem based on the Classical Langley Method and where the authors consider other references that analyse the same problem.

The authors insist in the use of statistical-mathematical methods to solve the problem of calibration forgetting in part the physic of the problem, when the exposed procedures are only the auxiliary tools. Precisely, as auxiliary tools these analysed procedures are necessary because always the fitting of the observations need to be done, with a more or less modification of the original equation of the Langley method. In this sense the analysis made in this paper is valid and of interest. The authors mention the parabolic shape of AOD (the authors call anomalous Langley plot) as an external problem when it is consubstantial with the Langley plot procedure itself according to Cachorro et al. papers. Therefore a justification is necessary of the convenience or not of using the method of the KCICLO.

A suggestion to the authors, it would be good the application of the KCICLO method

C3966

extensively described in the four papers of Cachorro et al., to determine the calibration constants of their AOD data series and analyse the results in comparison with these 11 procedures. Otherwise, bear in mind that the KCICLO method can be applied as a correction method of the actual AOD data without any information about the calibration constant (this is one of the major advantages) but if the previous calibration constants are known the new can be determined. Besides the KCICLO can be also used for a direct calibration process. The latter is not used in Cachorro et al., (2008a, b) but the authors describe this possibility as a two-step Langley Plot calibration.

In any case a good statistical or mathematical method is necessary to avoid outliers and do a correct fit and hence all the procedures described in the article are valid in a general context. Hence I found valid this article but it must bear in mind that to assume the best statistical method as the best procedure to solve the calibration problem is not the best solution, as the authors want to reflect in their article.

Response: The referee is correct in that we failed to mention the papers by Cachorro et al. As too often happens, we were not aware of them. Having read these papers, it appears that this technique is equivalent to Langley plot method. Diurnal cycle plot in Fig. 6 in

Cachorro, V. E., C. Toledano, A. Berjón, A. M. de Frutos, B. Torres, M. Sorribas, and N. S. Laulainen (2008), An "in situ" calibration correction procedure (KCICLO) based on AOD diurnal cycle: Application to AERONET–El Arenosillo (Spain) AOD data series, *J. Geophys. Res.*, 113, D12205, doi:10.1029/2007JD009673.

is equivalent to a Langley plot. Both approaches (KCICLO and Langley) hinge on successful identification of regions where the actual optical depth is constant. In both cases the manifestation of constant optical depth are points lying along a straight line. This follows from BLB law. Also from BLB law, it follows that a straight line is only a necessary condition, but not a sufficient condition of constancy of optical depth. Optical depth that has a hyperbolic component ε/m produces straight lines as well. For this

C3967

reason $\ln V_0$ or $\ln K$ are always burdened with an unknown error ε . Therefore in the second stage of the process (both for Langley plot or KCICLO) a time series of $\ln V_0$ or $\ln K$ (Fig. 7 in Cachorro et al., 2008) are analyzed to reduce ε by averaging it out. The presence of unknown ε in data series is the chief limiting factor on the accuracy of Langley or KCICLO methods. This explains why in the introduction we emphasize the concept of anomalous Langleys that helps to account for different uncertainties at different locations.

Our paper is concerned with devising and evaluating various methods of identifying viable Langley plots among data sets that contain outliers and regions where optical depth clearly could not have been constant. We use time series analysis only to evaluate the impact of these different methods on filtered and smoothed calibration constants. While the subject of our paper is not concerned with comparative analysis of various calibration methods, definitively Cachorro et al. papers (2004 and 2008) need to be cited. It is mathematically equivalent to a Langley plot method and thus the fitting methods that we devised and compared can be utilized in KCICLO to locate straight line events in KCICLO plots (like Fig.6). However, instead of finding an intercept as in the Langley plot, a slope would be sought in KCICLO method.

After line 15 on page 4193 we added this paragraph that refers to Cachorro et al.:

The Langley plot method of calibration consists of locating a subset of raw data points ($\ln V_i, m_i$) to which a straight line can be fit. The intercept of the straight line estimates the calibration constant $\ln V_0$. This is a consequence of the BLB law. Often already processed data of retrieved optical depth τ_i require verification and correction if the sun radiometer was poorly calibrated or its calibration has drifted. The BLB law implies that the slope of the straight line fitted to points $(\tau_i, 1/m_i)$ estimates the calibration constant correction factor (Cachorro et al. 2004 and 2008). Finding this slope is a mathematically equivalent approach to finding the intercept in Langley plot method. In both approaches the calibration success hinges on correctly locating the subset when the optical depth of atmosphere is constant. It is often overlooked, that the presence of

C3968

the straight line in the data set $(\ln V_i, m_i)$ or $(\tau_i, 1/m_i)$ does not imply that the actual τ is constant.

Reviewer#4 Specific correction

Reviewer #4: Page 4207 line 17. the sentence "We conclude that the OA, if it errs, it errs on being conservative, i.e., What means errs?"

Response: We consulted English dictionaries. The usage of the verb "to err" is correct.

Reviewer #4: 2. Page 4208, line 8: please check "by a moving a boxcar filter" 3. The word Centrete as Center in the Acknowledgements.

Response: Thank you, we corrected it. See response to Reviewer #3 who also spotted the two mistakes.

Additional changes:

(1) In Acknowledgments we added the following sentence:

The final shape of the paper owes much to constructive input of two anonymous reviewers.

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 4191, 2015.