Effect of water vapor on the determination of aerosol direct radiative effect based on the AERONET fluxes

J. Huttunen1,2, A. Arola1, G. Myhre3, A. V. Lindfors1, T. Mielonen1, S. Mikkonen2, J. S. Schafer4, S. N. Tripathi5, M. Wild6, M. Komppula1, and K. E. J. Lehtinen1,2

1Finnish Meteorological Institute, Kuopio, Finland
2Department of Applied Physics, University of Eastern Finland, Kuopio, Finland
3Center for International Climate and Environmental Research – Oslo (CICERO), Oslo, Norway
4NASA/Goddard Space Flight Center (GSFC), Biospheric Sciences Branch, Greenbelt, MD, USA
5Department of Civil Engineering, Indian Institute of Technology, Kanpur, India
6Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

Correspondence to: J. Huttunen (jani.huttunen@fmi.fi)

Received: 24 October 2013 – Published in Atmos. Chem. Phys. Discuss.: 10 January 2014
Revised: 14 May 2014 – Accepted: 16 May 2014 – Published: 20 June 2014

Abstract. The aerosol direct radiative effect (ADRE) is defined as the change in the solar radiation flux, \( F \), due to aerosol scattering and absorption. The difficulty in determining ADRE stems mainly from the need to estimate \( F \) without aerosols, \( F^0 \), with either radiative transfer modeling and knowledge of the atmospheric state, or regression analysis of radiation data down to zero aerosol optical depth (AOD), if only \( F \) and AOD are observed. This paper examines the regression analysis method by using modeled surface data products provided by the Aerosol Robotic Network (AERONET). We extrapolated \( F^0 \) by two functions: a straight linear line and an exponential nonlinear decay. The exponential decay regression is expected to give a better estimation of ADRE with a few percent larger extrapolated \( F^0 \) than the linear regression. We found that, contrary to the expectation, in most cases the linear regression gives better results than the nonlinear. In such cases the extrapolated \( F^0 \) represents an unrealistically low water vapor column (WVC), resulting in underestimation of attenuation caused by the water vapor, and hence too large \( F^0 \) and overestimation of the magnitude of ADRE. The nonlinear ADRE is generally 40–50% larger in magnitude than the linear ADRE due to the extrapolated \( F^0 \) difference. Since for a majority of locations, AOD and WVC have a positive correlation, the extrapolated \( F^0 \) with the nonlinear regression fit represents an unrealistically low WVC, and hence too large \( F^0 \). The systematic underestimation of \( F^0 \) with the linear regression is compensated by the positive correlation between AOD and water vapor, providing the better result.

1 Introduction

Significant uncertainties exist in the current estimates of aerosol effects on climate (IPCC, 2013). This holds also for the aerosol direct radiative effect (ADRE) and aerosol direct radiative forcing (ADR). The ADRE defines the attenuation of the (cloud-free sky) surface solar radiation flux (\( F \)) due to aerosol scattering and absorption. Herein, we consider the solar radiation flux at the surface, although ADRE applies also for the longwave flux and above the atmosphere. In the definitions of ADRE and ADRF, effects relate to both anthropogenic and natural aerosol particles, while forcing refers to the impact of anthropogenic aerosol particles. Although, for example, Myhre (2009) recently showed an increment of the consistency between observation-based and global aerosol model estimates, with a reduction in the uncertainty of this effect, other studies (e.g., Loeb and Su, 2010) highlight that considerable uncertainties are still associated with ADRE, mainly due to the uncertainties in single scattering albedo (SSA). Satheesh and Ramanathan (2000) employed a method in which ADRE is estimated using the aerosol direct effect efficiency (ADREE), which is the ADRE normalized by the aerosol optical depth (AOD), and it is estimated by fitting a
straight line into surface solar flux and AOD observations. A linear dependence between aerosol attenuation and AOD has been commonly assumed when estimating ADRE (e.g., Kaufman et al., 2002; Bush and Valero, 2002, 2003; Dunka et al., 2006; Roger et al., 2006; di Sarra et al., 2008; García et al., 2009; Satheesh et al., 2010). Typical attenuation of radiation intensity, however, implies nonlinear decay, as considered by for example Conant et al. (2003), Markowicz et al. (2008) and Kudo et al. (2010). Thus, a linear fit to \( F \) and AOD data may result in an incorrect extrapolation of \( F^0 \).

The aim of this paper is to examine the uncertainties involved in estimating ADRE, both using the linear fitting method and a nonlinear approach if \( F \) and AOD data are available from surface or satellite measurements. For this, we use Aerosol Robotic Network (AERONET) products (http://aeronet.gsfc.nasa.gov/) from all available AERONET stations, which cover different aerosol types and surface reflectance properties and provide modeled surface solar radiation fluxes also. We conducted our analysis using these modeled fluxes since they represent realistically enough the aerosol-induced relative changes in \( F \) and furthermore give an estimate for \( F^0 \), which is self-consistent within the selected \( F \) (AOD) data set. As AERONET provides an estimation of \( F^0 \), we can compare the estimations immediately with the baseline (AERONET). Special attention is paid to the possible effect of water vapor on estimating ADRE.

2 Methods and data

AERONET is a ground-based remote-sensing global network of Cimel sun photometers (Holben et al., 1998) including the AERONET inversion code with radiative transfer code implementation. The inversion strategy, described in Dubovik and King (2000), provides a group of parameters, e.g., AOD, Ångström exponent (AE) and water vapor column (WVC) from the sun measurements and, for example, SSA, asymmetry parameter (ASYM) and size distribution from the sky measurements. AOD is provided with wavelength channels 340, 380, 440, 500, 670, 870, 1020 and 1640 nm (all or some of these, depending on site of AERONET), WVC from 940 nm and, for example, SSA and ASYM from 440, 670, 870 and 1020 nm. The discrete ordinates method provides broadband fluxes (both at the top of atmosphere and at the surface, with and without aerosols), calculated with the correlated-\( k \) distribution in the Global Atmospheric Model (GAME) code from 200 to 4000 nm. The ozone is based on monthly averaged climatology by the Total Ozone Mapping Spectrometer (TOMS). Moreover, the US standard 1976 atmosphere model sets the atmospheric gaseous profile. The surface reflectivity is approximated by the bidirectional reflectance distribution function (BRDF) and observations from the Moderate-Resolution Imaging Spectroradiometer (MODIS). More details about the AERONET description from, for example, García et al. (2012). The uncertainty of AOD is 0.01–0.02 depending on the wavelength (Eck et al., 1999), the uncertainty in SSA approximately 0.03 (Dubovik et al., 2000), and the uncertainty in WVC of 12% (Holben et al., 1998). We used broadband modeled surface shortwave fluxes from this data set. In this study, level 1.5 sky AERONET data are divided into groups by station, season (December–February, March–May, June–August and September–November) and by solar zenith angle (SZA) (3° steps in the range 0–80°). A data set was included in the analysis if it had at least 20 observations and the data contained AOD 550 nm values above 0.3 and below 0.1. We chose to use level 1.5 data because using level 2.0 would leave out all quality-assured data with AOD 440 nm < 0.4 (including, for example, quality assured SSA and \( F \) calculations).

The drawback of this choice is that, at these low values of AOD, there are significant uncertainties in the optical properties retrieved. This is especially true for SSA, which is an important parameter. Thus, we applied all other level 2 criteria except for AOD (and SZA) limit, in order to enhance the accuracy of the data set selected. Moreover, we have imposed an additional data flagging criterion, removing those SSA points at the AOD 440 nm < 0.4, which are outside the average SSA ± standard deviation, defined for the AOD 440 nm > 0.4.

ADRE at the surface is the difference between the solar flux with and without aerosols: \( \Delta F = F^\text{aer} - F^0 \) (\( F^\text{aer} \) is flux with aerosols). The major challenge obviously is the determination of \( F^0 \). The methodology for its estimation employed in this study is illustrated in Fig. 1, in which \( F^\text{aer} \) (+ symbols) is plotted as a function of AOD (from now on 550 nm) for the AERONET site in Kanpur station (26° N, 80° E) for the spring months March–May with SZA \( \pm 69° \pm 1.5° \) (\( F^\text{aer} \) values were normalized for the average Earth–Sun distance and cosine correction of \( F^\text{aer} \) was done within SZA ranges to its midpoints). \( F^0 \) represents the case \( AOD = 0 \), but with measurements only at AOD above ca. 0.15, we have to extrapolate down to 0. In Fig. 1 we show two such extrapolations: a linear fit (dashed line) and a nonlinear decay fit (solid line) with the data.

We chose this data subset since it represent a case in which the \( F^\text{aer} \) and AOD data exhibit the natural nonlinear behavior of radiation intensity decay. Thus the resulting intercepts of the two curves at \( AOD = 0 \) are quite different: 317 Wm\(^{-2}\) with linear extrapolation and 349 Wm\(^{-2}\) with nonlinear regression, with a difference of 32 Wm\(^{-2}\) when estimating ADRE. Also, for each \( F^\text{aer} \) we show the corresponding AERONET \( F^0 \) (circles), based on the retrieved WVC and surface albedo, and calculated with a radiative transfer model (e.g., García et al., 2008; Derimian et al., 2008). We use the ADRE obtained by averaging these \( F^0 \) (circles) values (bar at \( F = 325 \) Wm\(^{-2}\) on the y axis) as the benchmark against which the extrapolation methods are evaluated.
Thus, $F$ with the constants.

Figure 1. Radiative flux with aerosols $F_{\text{aer}}$ (plusses) and without aerosols $F^0$ (circles) as a function of AOD for the AERONET site in Kanpur in March–May and with SZA = $69^\circ \pm 1.5^\circ$. The bar on the vertical axis represents the mean value of the estimated $F^0$ (all circles). The solid and dashed lines represent the exponential and linear fits to the data, respectively.

Mathematically, our analysis can be summed up as a comparison between the extrapolated ADRE

$$ADRE_{\text{extrapol}} = \frac{1}{n} \sum F_{i}^{\text{aer}} - F_{i}^{0 \text{ extrapol}}$$  \hspace{1cm} (1)$$

and the AERONET ADRE

$$ADRE_{\text{AERONET}} = \frac{1}{n} \sum F_{i}^{\text{aer}} - \frac{1}{n} \sum F_{i}^{0}$$  \hspace{1cm} (2)$$

where $F_{i}^{\text{aer}}$ and $F_{i}^{0}$ are $F^{\text{aer}}$ and $F^{0}$, respectively, with $i$ varying from one to the number of data set, $n$. Notably, the extrapolated $F^{0 \text{ extrapol}}$ (derived with fits) represents a single value for a data set, but in the AERONET, $F^{0}$ is determined side by side with each $F^{\text{aer}}$. $F^{0 \text{ extrapol}}$ is calculated using fits as follows:

$$F_{i}^{0 \text{ extrapol}} = x_1 + x_2 * x_3 \ast \exp (-x_3 \ast \text{AOD}_i);$$

$$F_{i}^{\text{lin} \text{ extrapol}} = x_1 + x_2,$$  \hspace{1cm} (3)

$$F_{i}^{\text{nonlin}} = x'_1 + x'_2 \ast \text{AOD}_i;$$

$$F_{i}^{0 \text{ lin}} = x'_1,$$  \hspace{1cm} (4)

where $F_{i}^{\text{nonlin}}$ and $F_{i}^{\text{lin}}$ are estimated $F^{\text{aer}}$ derived for each AOD with the nonlinear and linear method, respectively. Constants of fits are $x_1, x_2, x_3, x'_1$ and $x'_2$. $F_{i}^{0 \text{ nonlin}}$ and $F_{i}^{0 \text{ lin}}$, thus $F_{i}^{0 \text{ extrapol}}$ of the nonlinear and linear fits, are provided with the constants.

Our decision to use the modeled $F$ from AERONET, instead of pyranometer measurements, was based on two different aspects. First, this allowed us to include a multiple number of sites, with very different and varying aerosol conditions. Second, AERONET data provided interesting ancillary measurements to support and better understand our analysis, WVC being the most crucial one. In addition, the AERONET $F$’s agree with pyranometer measurements with a correlation better than 99%, and the relative difference varies from 0.98 to 1.02 (Garcia et al., 2008). Moreover, we tested the analysis at two sites – Alta Floresta and Goddard Space Flight Center (GSFC) – by using pyranometer measured fluxes $F$ and found no significant difference of the results at these two sites, if compared to the corresponding analysis using the AERONET-modeled fluxes instead (see Supplement Sect. S1).

3 Results

As further examples of determining ADRE using regression analysis, we show $F^{\text{aer}}$ and AOD data from four sites in Fig. 2. In addition, the linear (dashed line) and nonlinear decay (solid line) fits to the data are shown. The bar on the vertical axis represents the average (with SD) value for $F^0$. GSFC (39° N, 77° W) (SZA = $70^\circ$) (Fig. 2a) and Rio Branco (10° S, 68° W) (SZA = $70^\circ$) (Fig. 2b) represent cases in which the data are of sufficient quality for estimating ADRE: AOD values reach close zero with only minor changes in WVC, aerosol optical properties and surface reflectance for a given AOD, resulting in a narrow spread in the data. In these cases, since the nonlinear decay represents a more realistic decay of radiation intensity (based on squared values of residuals), the intersection of the nonlinear fit with the AOD = 0 axis (y axis) is within the SD of the baseline value. Dhadnah (26° N, 56° E) (SZA = $70^\circ$) (Fig. 2c) and GSFC at SZA = $22^\circ$ (Fig. 2d) are examples of more challenging cases: in Fig. 2c only data points with AOD > 0.2 exist so that a more extensive extrapolation is needed, and in Fig. 2d there is significant scatter in the points.

Perhaps the most interesting feature shown in Fig. 2, which also significantly affects the quality of ADRE estimation, is the correlation of $F^0$ with AOD. In Fig. 2a–d there is a negative correlation while in Fig. 2b the correlation is positive. The negative correlation between $F^0$ and AOD is indirectly caused mainly by a positive correlation of AOD with WVC due to humid air masses with large aerosol concentration. Only in some cases, where air masses are dominated by dust aerosols, the correlation is negative. With increasing AOD and WVC, the WVC dims an increasing fraction of the radiation intensity – resulting in a smaller $F^0$. The opposite occurs if AOD and WVC have a negative correlation. Increase in the AOD as a function of WVC is presumably partly due to hygroscopic growth (e.g., Kitamori et al., 2009), although probably a major part of the correlation can be attributed to...
a large variance in atmospheric conditions of aerosol properties and air humidity during seasons.

The intersections of the nonlinear decay fits (solid lines in Fig. 2) with the AOD = 0 axis – 313.5 W m$^{-2}$ (Fig. 2a), 295.9 W m$^{-2}$ (Fig. 2b), 327.4 W m$^{-2}$ (Fig. 2c) and 1008.9 W m$^{-2}$ (Fig. 2d) – approximate the $F_0$ value at AOD = 0. This is clear from the figure, if one imagines straight line fits through the circles and extrapolates fits down to AOD = 0. This approximation is, however, not necessarily a good one for the mean $F_0$, if $F_0$ and AOD correlate (through the AOD–WVC correlation). For the negative correlation cases (Fig. 2a–d), the intersections of the nonlinear decay fits with the AOD = 0 axis tend to therefore overestimate the mean baseline $F_0$ (307.3 W m$^{-2}$ for Fig. 2a, 312.9 W m$^{-2}$ for Fig. 2c, and 972.1 W m$^{-2}$ for Fig. 2d) – as the majority of $F_0$ values are below the extrapolated $F_0$. Typically, for the positive correlation cases (Fig. 2b, mean of $F_0 = 303.4$ W m$^{-2}$) the opposite occurs. As the linear fit obviously results in a lower estimation of $F_0$, the linear regression method can result often in a better estimation of the mean $F_0$, as is clearly the case in Fig. 2c (mean $F_0 = 306.7$ W m$^{-2}$) and Fig. 2d (mean $F_0 = 973.0$ W m$^{-2}$) – even if the nonlinear regression is physically more correct.

The performance of the two different regression methods and, in particular, the WVC and AOD correlation effect on the performance is illustrated as scatterplots in Fig. 3. In Fig. 3a all data are presented in ADRE (nonlinear decay method) and ADRE (AERONET $\Delta F^{\text{average}}$, Eq. 2) form. The color of the single points indicates the correlation of the WVC and AOD. In Fig. 3b the same is shown for the linear regression case. Evidently, a majority of the cases are such that WVC and AOD have a strong positive correlation (red colored points). In addition, it seems that, for most of these cases, the linear regression method (Fig. 3b) results in a better ADRE estimation than the nonlinear decay regression method (Fig. 3a). This means that the inaccuracy inherent in the linear regression cancels out errors caused by the WVC and AOD correlation. For a weak WVC and AOD correlation, the nonlinear decay method appears to be clearly better. Other parameters such as surface albedo, ASYM or SSA do not play as crucial of a role as WVC. We classified the ADRE estimates of the both methods against the baseline in respect of AOD, albedo, ASYM or SSA and WVC. It was evident that only WVC can explain the observed differences of both methods when compared against the baseline (see Supplement Sect. S2). Moreover, we confirmed, by modeling a short wavelength range (310–500 nm), that this WVC effect vanishes, if some other wavelength band as for example the visible range of 400–700 nm containing no significant water vapor absorption is under consideration, instead of the broadband wavelength range of $F^{\text{aer}}$ (see Supplement Sect. S3).

Next we investigated possible geographical features of this correlation. Figure 4 shows the WVC and AOD correlation (in the color scales) at all the sites available from AERONET,
Figure 3. ADRE predicted with exponential decay (a) and linear (b) regression methods (Eq. 1), compared with AERONET values (Eq. 2). The color of the data points represents the correlation coefficient of the AOD and WVC correlation, with red color indicating positive and blue color negative correlation.

in this case for seasons: December–February (DJF, Fig. 4a), March–May (MAM, Fig. 4b), June–August (JJA, Fig. 4c) and September–November (SON, Fig. 4d) (all years available). Most of the points are colored either green or red, indicating an absent or a positive correlation. The strongest positive correlation is for the stations in Europe and eastern USA, presumably due to aerosol hygroscopic growth. This holds especially for the JJA and SON seasons. The DJF and MAM seasons provide weaker positive correlation, indicating that the linear method can then provide there somewhat underestimated ADRE. Interestingly, the strongest negative correlation appears during the JJA season in the west Saharan region and Central America, probably caused by a strong desert dust domination and low WVC in the Saharan outflow region (Marsham et al., 2008). During those particular cases, the linear method can significantly underestimate ADRE, as indicated by the points of largest negative WVC vs. AOD correlation in Fig. 3b, while the nonlinear decay provides then a better estimate.

Finally, the ADRE estimations of all data are grouped together in numerical form in Table 1. As already evident from the figures, the nonlinear decay regression method overestimates (mean = −57.2 Wm$^{-2}$) while the linear method underestimates (mean = −39.4 Wm$^{-2}$) the magnitude of ADRE (AERONET value = −46.1 Wm$^{-2}$). Overall, the linear method yields better results than the nonlinear decay method.

Previous studies have shown that the AERONET WVC agrees well with radiosonde sounding data (e.g., Prasad and Singh, 2009; Bokoye et al., 2007). Also, Smirnov et al. (2004) indicate that the AERONET WVC provides the root mean square difference of 7.0 % in a multiyear comparison with WVC measurements derived from GPS. We also compared AERONET WVC measurements against radiosonde data from five sites (Alta Floresta, Cuiaba-Miranda, Niamey, Thessaloniki and Wallops) and observed similarly high correlations between these two data sources. However, we wanted to assess in particular whether there exists any systematic dependence between WVC from these two data sources as a function of AOD, which could affect our ADRE analysis based on the modeled $F$. We found that while the ratio between the AERONET and radiosonde WVC is essentially constant for AODs (at 500 nm) larger than about 0.1, in many sites WVC can deviate for the cases of smallest AOD (below 0.1). We estimated how our ADRE values (based on the $F$ and AOD relation) would change if we normalized the AERONET-modeled fluxes to incorporate the WVC from the radiosonde measurements instead of AERONET-measured WVC. We found that the increased WVC uncertainty at the lowest AOD values introduces an insignificant change in our ADRE estimates.

4 Conclusions

Determining the ADRE at the earth’s surface from radiative flux, $F$, measurements is not straightforward because it involves the estimation of the flux without aerosols $F^0$. This requires either radiative transfer modeling or an extrapolation of $F$ down to AOD = 0.

We have evaluated two such extrapolation methods: (i) a linear fit and (ii) a nonlinear decay fit to the $F$ and AOD relation. As a reference we used the AERONET ADRE data in which $F^0$ (and $F$) is calculated with radiative transfer modeling. Radiation attenuation due to multiple scattering and absorption results typically in a near-nonlinear decay of the
Table 1. The estimated ADRE \((F_{\text{aer}})\) with standard deviations compared with the AERONET value. MAD – mean absolute deviation. Units are in \(\text{Wm}^{-2}\), except for the correlation coefficient (CC).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>AERONET</th>
<th>Method</th>
<th>Estimate</th>
<th>Est. – AERONET</th>
<th>CC</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADRE</td>
<td>(-46.1 \pm 20.4)</td>
<td>Exp. decay</td>
<td>(-57.2 \pm 23.4)</td>
<td>(-11.1)</td>
<td>0.75</td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear</td>
<td>(-39.4 \pm 16.9)</td>
<td>(+6.7)</td>
<td>0.89</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Figure 4. Geographical distribution of the AOD and WVC correlation, at all AERONET stations considered in this study for (a) December–February, (b) March–May, (c) June–August and (d) September–November (all available years).

intensity, and thus the nonlinear decay regression is expected to give a better estimation of ADRE. This would be the case if the typically positive correlation of WVC and AOD did not affect the dependency. \(F^0\) represents an unrealistically low WVC, resulting in an underestimation of attenuation caused by the WVC, and hence a too large \(F^0\). This leads to an overestimation of the magnitude of ADRE. For stations and data series in which there is no correlation between WVC and AOD, the nonlinear decay fit is superior.

As the WVC effect was found to be of such importance, we also investigated the geographical correlation of WVC and AOD. The positive correlations clearly dominate, and clear negative correlations occur predominantly in desert-dust-dominated data series, such as the regions at the Saharan outflow. The strongest positive correlation was found in stations in Europe and eastern USA. Our results indicate that the regression method, either linear or nonlinear, can readily produce a significant error due to the correlation of WVC and AOD. Since for a majority of locations AOD and water vapor column (WVC) have a positive correlation, the linear method gives somewhat better results in general than the nonlinear approach, for the reasons discussed above. However, there are specific regions of strong negative WVC and AOD correlation, most notably in the Saharan dust outflow region, where the opposite takes place and nonlinear approach results in better estimate for ADRE. Therefore, based on our results we recommend that when the surface ADRE is estimated by using pyranometer and AOD measurements, the site-specific correlation between WVC and AOD should be also estimated to deduce whether linear or nonlinear approach is more suitable. We moreover recommend taking one step forward and additionally attempting to correct for the possible bias due to WVC and AOD correlation. When the data for the WVC become available, then better ADRE accuracy is likely achieved if the flux measurements are normalized to constant WVC amount with simple scaling obtained from RT modeling.

The Supplement related to this article is available online at doi:10.5194/acp-14-6103-2014-supplement.
Acknowledgements. We thank the AERONET team, principal investigators and other participants for their efforts in establishing and maintaining the network. This study is supported by the Academy of Finland Doctoral Programme ACCC and the Maj and Tor Nessling Foundation. We also thank Larry Oolman from Department of Atmospheric Science, University of Wyoming, for providing radiosonde data of atmospheric water vapor column abundance.

Edited by: E. Gerasopoulos

References


