Atmos. Meas. Tech. Discuss., 6, 10571–10615, 2013
www.atmos-meas-tech-discuss.net/6/10571/2013/
doi:10.5194/amtd-6-10571-2013

Atmospheric Measurement Techniques
Discussions
Open Access

Multi-modal analysis of aerosol robotic network size distributions for remote sensing applications: dominant aerosol type cases

M. Taylor, S. Kazadzis, and E. Gerasopoulos

Institute for Environmental Research and Sustainable Development (IERSD), National Observatory of Athens (NOA), Metaxa & Vas Pavlou, Penteli, 15236 Athens, Greece

Received: 3 November 2013 – Accepted: 27 November 2013 – Published: 9 December 2013

Correspondence to: M. Taylor (patternizer@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

To date, size distributions obtained from the aerosol robotic network have been fit with bi-lognormals defined by six secondary microphysical parameters: the volume concentration, effective radius, and the variance of fine and coarse particle modes. However, since the total integrated volume concentration is easily calculated and can be used as an accurate constraint, the problem of fitting the size distribution can be reduced to that of deducing a single free parameter – the mode separation point. We present a method for determining the mode separation point for equivalent-volume bi-lognormal distributions based on optimisation of the root mean squared error and the coefficient of determination. The extracted secondary parameters are compared with those provided by AERONET's Level 2.0 Version 2 inversion algorithm for a set of benchmark dominant aerosol types including: desert dust, biomass burning aerosol, urban sulphate and sea salt. The total volume concentration constraint is then also lifted by performing multimodal fits to the size distribution using nested Gaussian mixture models and a method is presented for automating the selection of the optimal number of modes using a stopping condition based on Fisher statistics and via the application of statistical hypothesis testing. It is found that the method for optimizing the location of the mode separation point is independent of the shape of the AVSD, does not require the existence of a local minimum in the size interval 0.439 µm $\leq r \leq$ 0.992 µm, and shows some potential for optimizing the bi-lognormal fitting procedure used by AERONET particularly in the case of desert dust aerosol. The AVSD of impure marine aerosol is found to require 3 modes. In this particular case, bi-lognormals fail to recover key features of the AVSD. Fitting the AVSD more generally with multi-modal models allows automatic detection of a statistically-significant number of aerosol modes, is applicable to a very diverse range of aerosol types, and gives access to the secondary microphysical parameters of additional modes currently not available from bi-lognormal fitting methods.

1 Introduction

Highlights

- A method for optimizing bi-lognormal fits to the size distribution with a single free parameter – the mode separation point
- Sensitivity analysis of the dependence of secondary microphysical parameters on the mode separation point and on aerosol optical depth
- A method for multi-modal analysis of the size distribution using Gaussian mixture models and access to the microphysical parameters of higher modes
- A test of the feasibility of the methods for fitting size distributions of dominant aerosol types of diverse morphology

Satellite retrievals of aerosol optical depth (AOD) and related parameters typically require the use of prescribed models of aerosol size and composition. In particular, the aerosol volume size distribution (AVSD) and the spectral complex refractive index are needed to compute properties such as the scattering phase function, the single scattering albedo and the extinction coefficient, which are in turn used to calculate quantities such as the total AOD from the columnar abundance. In general, the information content of measurements from current satellite radiometers is insufficient to unambiguously retrieve all these parameters particularly when the (spectral and directional) behavior of the surface reflectance is unknown (Hasekamp and Landgraf, 2007). For this reason, aerosol retrieval algorithms employed by most of these sensors are required to make assumptions about microphysical properties. The consequence is that these assumptions then contribute to differences in retrieved AOD – even in the idealized case of a black (non-reflecting) surface (Kokhanovsky et al., 2010).

So, while the ability of satellite retrieval algorithms to represent the radiative-behaviour of real aerosols is still in question (most recently raised in the context of

AMTD
6, 10571–10615, 2013

pure marine aerosol models by Sayer et al., 2012), satellite retrievals are usually validated (e.g. Remer et al., 2005) against co-located and synchronous retrievals provided by ground-based sun photometer and sky radiometer systems like those in the aerosol robotic network (AERONET). The main reason for this is that remote sensing of the AVSD in particular is exceedingly difficult and no one sensor system is capable yet of providing totally unambiguous information (King et al., 2009). Moreover, the aerosol model inter-comparison initiative – AeroCom has carried out analysis of aerosol simulations from various global chemical transport and climate models and found that, even on the scale of yearly averages, aerosol life cycles and particle sizes span a large range of values (Textor et al., 2006, 2007) and the total number of aerosol modes, mode composition and control parameters vary considerably both between models and their final simulation results (Zhang et al., 2010). As a result, there is heavy reliance on the AVSDs provided by AERONET.

The advanced mathematical inversion algorithm developed to provide AERONET retrievals (Dubovik and King, 2000) from direct (sun) and diffuse (sky) measurements of radiation, returns aerosol optical parameters such as the spectral AOD, single scattering albedo and phase function, as well as important microphysical parameters like the AVSD, the complex refractive index, and the percentage of spherical particles (see Dubovik et al., 2002, 2006). Since the AVSD plays a pivotal role in the relation of the radiation field to the microphysics of aerosol particles (Hansen and Travis, 1974) as well as for the determination of aerosol type and composition (Dubovik et al., 2011), this paper focuses on the development of a method for assessing whether or not the AVSD for a couple of characteristic cases can be fit using multiple aerosol modes with a procedure that can be automated.

The AERONET Level 2.0 Version 2 inversion code inverts sky radiances simultaneously at wavelengths available in the CIMEL instrument (most frequently at 440, 675, 870 and 1020 nm) for the complete solar almucantar scenario or principal plane scenario together with measurements of AOD at the operational wavelengths. In particular, the algorithm returns the AVSD $dV(r)/d\ln r$ in 22 equidistant logarithmic radial size

AMTD
6, 10571–10615, 2013

bins spanning the range of particle radii $0.05 \le r \le 15\,\mu m$, normalized to the value of the total volume concentration of aerosol in $\mu m^3\,\mu m^{-2}$. The AERONET inversion code approximates the AVSD using trapezium rule integration (Dubovik and King, 2000) and, while the option of allowing the use of lognormal-shaped bins was included in Dubovik et al. (2006), it has only recently been found that sufficiently accurate modeling of POLDER/PARASOL observations requires such an optimization of the shape of each radial size bin (Dubovik et al., 2011). In particular, lognormal-shaped bins were found to provide notable improvements over the trapezoidal approximation in terms of smoothness and suggested some advantage in modeling the AVSD as a superposition of n-lognormals with the modal volume concentrations V_i, geometric mean radii r_i and standard deviations σ_i as fixed parameters (Dubovik et al., 2011),

$$\frac{dV(r)}{d\ln r} = \sum_{i=1}^{n} \frac{V_i}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln r - \ln r_i}{\sigma_i}\right)^2} \tag{1}$$

This finding is the main motivation for this paper. For several decades now, the AVSD of tropospheric aerosols has been known to contain several distinct modes, each most commonly being modeled by a lognormal function (Whitby, 1978; Remer et al., 1997, 1998; Remer and Kaufman, 1998; O'Neill et al., 2000). The statistical properties of the lognormal and bi-lognormal distribution are well-known (O'Neill et al., 2000) and are applied in this paper to test the feasibility of modeling the AVSD of distinct aerosol cases with super-positions of several ($n \ge 2$) lognormals. The multi-modal method presented here, it is hoped, will add a new layer of detail to existing studies without the need for too much additional mathematical complexity. Furthermore, since many available radiative-transfer codes are now able to take as input lognormal distribution parameters (Sayer et al., 2012), the results of this paper can be readily applied and implemented. In Appendix A, the equations used to calculate secondary microphysical parameters such as V_i, r_i and σ_i are presented.

The paper is organized as follows. The approach adopted for aerosol typing and the selection of sites impacted by dominant aerosol types is presented in Sect. 2.

AMTD
6, 10571–10615, 2013

Section 3 then briefly outlines two new methods for optimizing bi-lognormal fits and for fitting the AVSD with multiple modes. In Sect. 4, the results of applying the two new methods to a cohort of AVSDs representative of different aerosol types are presented and compared with AERONET, and the major impacts and feasibility of these new approaches are noted and analyzed. Finally, we conclude in Sect. 5 with a summary of our findings and an assessment of the potential offered by these new methods for analyzing AVSDs provided by AERONET or other remote sensing instruments.

2 Data selection

In this paper we apply new methodologies (developed in Sect. 3) to a set of dominant aerosol type AVSDs. While portals like NASA's AERONET Data Synergy Tool (http://aeronet.gsfc.nasa.gov/cgi-bin/bamgomas_interactive) and the Multi-sensor Aerosol Products Sampling System (MAPSS: http://giovanni.gsfc.nasa.gov/mapss) provide a framework for multi-sensor aerosol validation, inter-comparison, and joint analysis, a search for dominant aerosol type cases and high load aerosol events must still be done manually or with reference to field campaigns published in the literature. Here, we describe the approach we adopted to isolate candidate AERONET sites as well as those days which are most dominated by desert dust, biomass burning products, urban sulphate, and marine sea salt. We will see below that "dominant" sea salt is the most problematic case for bi-lognormal fitting methods owing to the fact that the data is drawn from an island site where the marine aerosol is mixed to a high degree (in the proportion 60 % : 40 %) with other aerosol species.

The Georgia Institute of Technology–Goddard Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model (Chin et al., 2000, 2002 and Ginoux et al., 2001) used by NASA's GEOS-5, simulates the AOD for major types of tropospheric aerosols. In particular, it provides 3 hourly measurements of the total extinction AOD as well as the contribution to the total extinction AOD of sulphate (SU), black carbon (BC), organic carbon (OC), desert (mineral) dust (DU) and sea salt (SS). It therefore

AMTD
6, 10571–10615, 2013

provides a model-driven aerosol classification. This is complementary to the way the AERONET's *Spectral Deconvolution Algorithm Product* (O'Neill et al., 2003) uses a more generalized set of microphysical assumptions to estimate the contributions of fine and coarse particles to the AOD at visible wavelengths. GOCART data spanning the years 2001–2005 (inclusive) is obtainable from the AERONET data synergy portal and was downloaded for the first 155 AERONET sites (75 % of all *Level 2.0 Version 2 Inversion Product* records N) ranked by the number of daily-averaged data available. Since GOCART provides eight 3 hourly measurements per day, these were averaged to produce daily-averages. The ratio of the contribution of individual aerosol types to the total extinction AOD was then calculated (as percentages) and used to sort the ranked sites by dominant ("nearly-pure") aerosol type. This approach provides a simple and straight-forward method for site selection. During this process, an additional column was added to the data provided by GOCART – so as to monitor a combination of aerosol: the combined percentage of organic and black carbon (OC + BC). The reason for this is that, although SU accompanies the burning of biomass products, the combination OC + BC was found to better distinguish biomass burning sites from urban sites (which can also have high levels of SU). While no site of course has 100 % "pure" aerosol of a single type, this approach enables one to rank sites by *dominant* aerosol type. For example, for the biomass burning products OC + BC, Mongu was selected since: (i) it has the longest AERONET data record (N = 1573 days) during the period 2001–2005, and (ii) it has a very high OC + BC presence (71.3 %), second only to Alta Floresta (77.78 %). Analogous criteria were used to select sites dominated by dust, urban-industrial SU and sea salt aerosol. As a result, the following sites that are representative of the dominant aerosol types were selected:

- Urban-industrial pollution: GSFC-Washington, US (76.840° W, 38.992° N, elevation = 87 m)
- Biomass burning: Mongu, Zambia (23.151° E, 15.254° S, elevation = 1107 m)
- Dust: Banizoumbou, Niger (2.665° E, 13.541° N, elevation = 250 m)

AMTD
6, 10571–10615, 2013

– Marine: Lanai, Hawaii (156.922° W, 20.735° N, elevation = 20 m)

Regarding the selection of temporal data at these sites, daily-averaged AERONET Level 2.0 Version 2 quality-assured and cloud-screened data for each site was aligned (i.e. filtered for synchronous values) with the daily-averaged GOCART AOD extinction data. The day corresponding to the peak % SU at GSFC-Washington, peak % (BC + OC) at Mongu, peak % DU at Banizoumbou and peak % SU at Lanai, was then identified. This allowed for selection of the AERONET daily-averaged Level 2.0 Version 2 AVSDs used in this study. The aerosol composition ranked by dominant aerosol type for each of the 4 sites is shown in Table 1.

3 Methodology

In Sect. 3.1 we briefly describe some anomalies associated with the fitting of key aerosol types using the bi-lognormal fitting procedure used by AERONET. Section 3.2 presents a potential improvement of this approach by lifting constraints on the mode separation point and performs a sensitivity analysis. Finally, in Sect. 3.3 we provide a general multi-modal fitting approach based on Gaussian mixtures. The different approaches are illustrated by using the case of "dominant" marine (sea salt) aerosol at Lanai to provide a context.

3.1 AERONET bi-lognormal fitting

The AVSD is usually assumed to be bi-modal so as to distinguish between a "fine" (accumulation) mode containing small particles ($<\approx 0.6\,\mu m$) and a "coarse" mode containing larger aerosol particles (Omar et al., 2005). With reference to Eq. (1), this partition of the AVSD into $n = 2$ lognormal modes – a fine (f) mode and a coarse (c) mode therefore requires the calculation of 6 secondary parameters: V_f, V_c, r_f, r_c, σ_f and σ_c. The estimation of these parameters by the AERONET retrieval algorithm proceeds as

AMTD
6, 10571–10615, 2013

follows. Firstly, the total volume size distribution $dV/d\ln r$ is divided into two parts at a radial mode separation point r_s. Fine particles of radii $r < r_s$ contribute to the fine mode volume concentration V_f while coarse particles with radii $r > r_s$ contribute to the coarse mode volume concentration V_c. Values of other secondary (derived) parameters such as the volume geometric mean radii r_f and r_c, volume geometric standard deviations σ_f and σ_c and the fine mode fraction η (the fraction of the total volume concentration due to the fine mode), are then calculated from the AVSD by integrating to and from the separation point r_s. The equations necessary for their calculation is presented in Appendix A. To calculate r_f one sets $r_1 = 0.05\,\mu$m and $r_2 = r_s$ and then exponentiates the value of of $\ln r_V$ obtained. Conversely, to calculate r_c one sets $r_1 = r_s$ and $r_2 = 15\,\mu$m and then exponentiates the value of $\ln r_V$ obtained. Similarly, to calculate the geometric fine mode standard deviation σ_f one sets $r_1 = 0.05\,\mu$m and $r_2 = r_s$ in Eq. (A4) and to calculate σ_c one sets $r_1 = r_s$ and $r_2 = 15\,\mu$m. The AERONET Level 2.0 Version 2 inversion algorithm provides the values of all of these parameters including the value of the separation point r_s upon which they all depend. At present, the code estimates r_s by finding the local minimum within the size interval $0.439 \leq r \leq 0.992\,\mu$m (Dubovik et al., 2000). For comparison, the CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm (Omar et al., 2009) refers to r_s as the "fine cut-off radius" and sets it to 1.0 µm for all aerosol types classified ("dust", "smoke", "clean continental", "polluted continental" and "polluted dust"), apart from "clean marine" aerosol which has r_s set at 0.6 µm. The CALIPSO separation point (with the exception of marine aerosol) is therefore fixed at the upper limit of the AERONET operational range.

In the context of AERONET Level 2.0 Version 2 retrievals, the integrals are discrete since $dV/d\ln r$ is provided in 22 logarithmic bins. In order to investigate the accuracy of performing numerical integration with lognormal radial bins, we calculated the integral by first of all fitting $dV(r)/d\ln r$ with a piecewise interpolation between the 22 equidistant logarithmically-spaced points provided in the radial size range $r_1 = 0.05\,\mu$m and $r_2 = 15\,\mu$m in Eq. (A1). We then doubled the number of interpolation points and

AMTD
6, 10571–10615, 2013

re-calculated the integral. This was then repeated until the integral converged. The rationale for interpolating the AVSD is twofold:

1. in order to decrease the radial step size and hence improve the validity of the sensitivity analysis applied to the position of the mode separation point r_s described in Sect. 3.2

2. in order to avoid spikes in the errors propagated using the Gaussian mixture method (GMM) described in Sect. 3.3 that are caused by jaggedness resulting from straight line connections across 22 discrete bins.

For the case of peak marine (sea salt) AVSD at Lanai, Fig. 1 shows how successive doubling of the number of interpolation points leads to a suitable smoothing of the AVSD without introducing spurious features.

Regarding the accuracy of the AERONET retrieval (the grey band in Fig. 1), the overall uncertainty in AOD data (under cloud-free conditions) is ±0.01 for wavelengths longer than 440 nm, and ±0.02 for shorter wavelengths (Dubovik et al., 2000), and the error in aerosol AVSD is estimated to be $< 10\,\%$ for particle radii between 0.1 µm and 7 µm (see Dubovik et al., 2000). While this is true near the maxima of the distribution, the errors can be as large as 35 % for the lowest AVSD values in this particle range (Dubovik et al., 2002). Furthermore, at the edges of the AVSD ($r < 0.1$ µm and $r > 7$ µm) the accuracy of the retrieval drops significantly because of the low sensitivity of aerosol scattering at 440, 670, 870 and 1020 nm to particles of these sizes (Dubovik et al., 2002). Correspondingly, the retrieval errors rise sharply to 80–100 % at the edges but do not significantly affect the derivation of the secondary microphysical parameters because typically the value of the AVSD is very low there (Dubovik et al., 2002). To provide a conservative uncertainty context for the results presented here, AERONET AVSDs are overlaid with an error band that is 10 % at peaks in the interval $1 \leq r \leq 7$ µm, 35 % at local minima in this range, and 100 % in edge regions when $r < 0.1$ µm and $r > 7$ µm. Between these thresholds, the error is interpolated on an equidistant, logarithmically-spaced grid to ensure a smooth transition. Note that the

AMTD
6, 10571–10615, 2013

overall trend with increasing number of points is as follows: V, V_f and V_c were all found to converge with increasing N but did not converge asymptotically to the values given by the AERONET Level 2.0 Version 2 inversion retrieval. In particular when comparing with the quoted AERONET values, the interpolated value of V is slightly lower (by 1.19 %), V_f is moderately lower (16.20 %) and V_c is slightly higher (1.81 %) – but all well within the prescribed AERONET uncertainty band.

In Fig. 3 (in Sect. 3.2), the AERONET-retrieved AVSD together with the associated bi-lognormal fit reconstructed from the AERONET-quoted values of the secondary microphysical parameters is shown. Note that 3 goodness of fit statistics are also provided: the mean of the residuals b, the standard error of the fit s, and the degrees of freedom-adjusted R^2 which measure the difference between the interpolated AERONET AVSD and the bi-lognormal re-constructed using the secondary parameters in Eq. (1) for $n = 2$ modes. All data-model comparisons in this work are accompanied by these measures of bias, location and spread (b, s and R^2 respectively). Figure 3 clearly shows that in the case of maritime (sea salt) aerosol, the fit to the fine mode is good but the fit to the coarse region is inappropriate – both from the perspective of fitting the "double hump" with a single broader peak, and from the perspective of the overall amplitude in this region. Given that the double hump occurs in the coarse region, it is possible that it is caused by the existence of a mixture of 2 coarse populations arising perhaps from: (a) 2 different aerosol types, (b) fresh aerosol with an aged component, or (c) some combination of these. In any case, it appears that the mode separation point r_s for this AVSD leads to a bi-lognormal fit that is good for the fine mode but poor for the coarse mode. The requirement for the total integrated volume concentration to remain constant means that, while it is feasible that the double-hump coarse mode region can be better fit by changing the mode separation point, the fit to the fine mode peak will necessarily have to worsen. This is suggestive of a need for handling the problem in a different way. With this in mind, in the next section, we present a method for unambiguously and automatically calculating the optimal location of r_s by optimizing the statistical measures s and R^2.

AMTD
6, 10571–10615, 2013

3.2 Optimised Equivalent-Volume (OEV) bi-lognormal fitting

Since the AERONET code uses r_s to separate fine and coarse modes and then obtains the spectral AOD extinction and absorption and asymmetry factors for these modes, the location of r_s is central. Furthermore, the fact that such spectral parameters are being used to validate satellite retrievals of fine and coarse modes means that the role played by r_s is becoming more prominent. It is therefore important to test the assumption that bi-lognormals provide the best fit to the AVSD and, if so, to assess the impact of uncertainty in the value of r_s on derived microphysical parameters. For this purpose, we stepped r_s through the set of 2200 (interpolated) equidistant logarithmically-spaced radial bins (excluding the end points). Then, using the interpolated volume concentration as a constraint, in each step, a bi-lognormal was fit to the AVSD and the secondary (derived) microphysical parameters: V_f, V_c, r_f, r_c, σ_f and σ_c, and goodness of fit measures: s and R^2, were calculated and tabulated. While the 2198 fits obtained (minus the end-points) fill a continuum, we show in Fig. 2 the result of applying this procedure to the AERONET AVSD radial bins in the range 0.1–7 µm (the "10 % error range").

While Fig. 2 shows no discernable bias $b = 0.000$ (to 3 decimal places) and a small and stable standard error $s \approx 0.001$ µm, R^2 is much more sensitive to changes in r_s and reveals a peak value of 0.893 at $r_s = 0.286$ µm. This suggests a method for automating the detection of the optical value of r_s related to max(R^2). Despite appearing constant in the legend of Fig. 2, a unique trough was found to exist in the curve of s – suggesting an optimal separation related to min(s) when $r_s = 0.290$ µm ($\approx$ 9 % smaller than that obtained with R^2). These two estimates of r_s are significantly lower than the mode separation point $r_s = 0.439$ µm quoted by AERONET at this site on this day. Since we could not find a clear reason for favouring the max(R^2) method over the min(s) method, we decided to define the optimal r_s as the mean of the values obtained from the min(s) method and the max(R^2) method, i.e. $r_s = (0.290\,\mu m + 0.315\,\mu m)/2 = 0.303\,\mu m$. Figure 3 compares the results of the standard AERONET bi-lognormal fit with that obtained by optimizing the separation point using the method described above.

AMTD
6, 10571–10615, 2013

The OEV method developed above, based on optimisation of statistical measures of goodness of fit, obtains a new mode separation point that marginally improves the bi-lognormal fit to the AVSD (R^2 = 0.894 as compared with R^2 = 0.885). However, while the improvement is only of the order of 1 % in terms of R^2, there is a significant *qualitative* improvement. Visually, the fit to the fine mode with the new OEV method is much better, particularly on the smaller radius side of this marine aerosol AVSD. Its peak, while within the AERONET error in this region, appears to be slightly over-estimated in amplitude but is well-located. The fit to the coarse mode is also better in terms of the peak amplitude. It is also nearer to the raw data on the rising edge of the coarse mode peak region at smaller radii. However, neither the AERONET bi-lognormal fit nor the new OEV fit are able to fit the double-peak in the coarse mode region. In Sect. 3.3, we present a second fitting method based on Gaussian mixture models to investigate whether or not the inclusion of additional modes can account for such features in the AVSD.

With regard to the dependence of the secondary microphysical parameters on r_s, we also performed a sensitivity analysis over the range $0.1 \leq r_s \leq 7$ µm for the AVSD at Lanai, Hawaii on the 21 January 2002 interpolated with 2200 points as described in Sect. S1.1 of the Supplement. In the range of mode separation points 0.286 µm $\leq r_s \leq$ 0.567 µm (i.e. up to the edge of the coarse region ≈ 0.6 µm) the magnitude of the relative errors of the geometric mean radii r_f and r_c, geometric standard deviations σ_f and σ_c and mode volumes V_f and V_c were found not to exceed 30 % as can be seen in Table S1. However, steep gradients were observed outside this range. For example, in the range of mode separation points $0.439 \leq r_s \leq 0.992$ µm used by the AERONET inversion code, the fine mode parameters r_f, σ_f and V_f reached large negative relative errors especially at the higher radius end where $r \approx 1$ µm (see Table S1). This suggests that the AERONET bi-lognormal fit is strongly under-predicting their values. The sensitivity analysis shows that apparently small differences in the value of the deduced separation point r_s, can be seen to translate into large differences in the deduction of secondary microphysical parameters – and hence the shape of the reconstructed

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD

6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

AVSD as can be seen visually in Fig. 3. The results of applying the sensitivity analysis to the 3 other dominant aerosol types are also included in Table S1 for reference and are discussed in Sect. 4. We also performed a sensitivity analysis of the effect of aerosol load as measured by the AOD at 1020 nm on the calculated secondary microphysical parameters. The results for a time window of 20 days around the marine aerosol peak (21 January 2002) at Lanai, Hawaii are presented in Sect. S1.2. All microphysical parameters (except for σ_c) show an increase with increasing AOD (1020 nm). As expected, this proportionality is strongest in the case of the modal volume concentrations V_f and V_c. The large spread of points at values of AOD (1020 nm) ≤ 0.04 in Fig. S2 is due to the fact that Level 2.0 Version 2 retrievals are less reliable at these low loads. Nonetheless, there is a substantial difference in the values obtained by the OEV fit and AERONET for loads AOD (1020 nm) > 0.04 suggesting the need to reassess criteria on a case by case basis.

3.3 Gaussian Mixture Model (GMM) multi-modal fitting

While physical processes in the atmosphere often result in a bi-modal AVSD structure (Remer and Kaufman, 1998), it is known that the appearance of a third mode is also both realistic and likely (Dubovik and King, 2000). For example, a volcanic eruption may produce optically thick stratospheric aerosol, which adds a stable third compositional mode to the commonly appearing fine (accumulation) mode ($r < 0.6$ µm) and coarse mode ($r > 0.6$ µm) composing tropospheric aerosol (Kaufman and Holben, 1996). The standard deviation of the AVSD is also known to vary substantially and to depend on the type of aerosol as well as prevailing atmospheric conditions (Dubovik and King, 2000). Furthermore, in the context of marine aerosol for example, it has been recently noted that the coarse mode is skewed and has a long tail at the lower-radius end (Sayer et al., 2012). The case of dominant marine sea salt at Lanai above presents an even more extreme situation where the coarse mode is double-humped and the fit with a bi-lognormal is problematic. Despite such observations, the division of the AVSD into a single fine mode and a single coarse mode is still the norm. Here, we wish

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD

6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◂ ▸
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

to assess whether or not there are statistically-significant exceptions where additional modes should be included in the analysis. Of course, such discrepancies should be assessed carefully taking into account the uncertainty in AVSD measurements (particularly outside of the interval 0.1–7 µm), and also the fact that the AERONET inversion code retrieves the AVSD from radiation measurements that have their own uncertainties. While the AERONET retrieval does not provide confidence intervals on the microphysical parameters (this is the reason bi-lognormal fits of Sects. 3.1 and 3.2 are presented without stated errors), the use of GMMs however, allows for calculation of 95 % confidence bands by propagating errors (see below). As a result, in what follows, errors will be placed both also on the GMM model fit.

We wish to point out that, while the AVSD can also be fit *directly* with lognormals in the r-domain ($\mathrm{d}V(r)/\mathrm{d}r$), we *chose* instead to fit the AVSD in the $\ln(r)$-domain (i.e. $\mathrm{d}V(r)/\mathrm{d}\ln r$) with normal distributions. The reason is not just because there is mathematical equivalence between a lognormal distribution in r and the analogous Gaussian distribution in $\ln(r)$, but because the lognormal distribution is skewed. Peaks in the AVSD in the r-domain are therefore *always* skewed. On the contrary, in the $\ln(r)$-domain, Gaussians are symmetrical and so an observed skew in a peak of the AVSD is due instead to a loss in goodness of fit arising from the inappropriateness of fitting it with a single (normal) mode, rather than due to any instrinsic asymmetry associated with the distribution itself. In the GMM method then, we fit the AVSD with a mixture of Gaussians under the *proviso* that the independent variable is $\ln(r)$. The GMM fit in this space has the generic form,

$$\frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} = \sum_{i=1}^{n} a_i e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} \qquad (2)$$

Where, in this notation, a_i is the "amplitude", b_i is the "centroid" (or modal value), and c_i is related to the width of each mode i for a mixture containing up to n discrete and independent modes. Comparing with the general n-lognormal equation of Eq. (1), the model coefficients (a_i, b_i and c_i) are related to the secondary (derived) microphysical

AMTD
6, 10571–10615, 2013

parameters (V_i, r_i and σ_i) via the set of relations:

$$a_i = \frac{V_i}{\sigma_i\sqrt{2\pi}} \leftrightarrow V_i = \sqrt{\pi}(a_i c_i) \tag{3}$$

$$b_i = \ln r_i \leftrightarrow r_i = e^{b_i} \tag{4}$$

$$c_i = \sigma_i\sqrt{2} \leftrightarrow \sigma_i = \frac{1}{\sqrt{2}}c_i \tag{5}$$

Our task then was to analyze an array of mixtures (containing 1–6 different modes), obtain the coefficients a_i, b_i and c_i for each mode, and then identify which mixture best fits the AVSD. We begin by noting that while the GMM is a linear sum of modes, it is nonlinear in the coefficients – hence a nonlinear least squares fitting method was adopted. The fitting algorithm used to adjust the coefficients, was a variant of the conjugate gradient method called the "trust region" method. More details on the application of this method are presented in Appendix C. To estimate the error of the fit, we followed the standard procedure of propagating errors (also described in Appendix C). As a result, the nonlinear least squares fitting algorithm provides values for a_i, b_i and c_i plus their upper and lower bounds at the 95 % level of confidence ($p = 0.05$). Then, by making the assumption that the upper and lower confidence intervals are symmetrical about a_i, b_i and c_i (i.e. "2-tailed"), their standard errors SE (a_i), SE (b_i) and SE (c_i) can be calculated – equal to the confidence interval divided by 1.96 (the value of the z score at the 95 % level). These standard errors then allow for an estimate of the upper and lower error bounds on the AVSD fit obtained for each mixture of n modes. Figure 4 shows the results of applying the GMM method to the AVSD of peak marine aerosol at Lanai for the first $n = 1 \rightarrow 6$ modes.

Clearly, 3 modes appear to be sufficient to fit the AVSD for this particular case. The addition of more modes does not contribute much improvement in the goodness of fit measures. For example, for the 6-modal fit the value of $R^2 = 1.000$ is only marginally better than the value $R^2 = 0.998$ for the 3-modal fit. Hence, an important question to

AMTD
6, 10571–10615, 2013

answer at this point is: how can detection of the 3-modal "optimum" be automated? For example, a detection algorithm based on seeking the maximum value of R^2 is likely to flag up the 6-modal fit as the optimal fit to the above AVSD. In particular, care should be taken to ensure that additional modes are *physical* and not just artefacts of the fitting procedure. What is needed therefore is a stopping condition to find the optimal mixture. One way to do this is to define a statistic and then to perform a hypothesis test to assess whether or not adding an extra mode leads to a statistically-significant improvement in the fit. Here, we adopt the protocol outlined by Harel (2009) and work with the square root of the degrees of freedom-adjusted R^2 as a proxy for the Pearson product-moment correlation coefficient ρ. This is based on the assumption that the nesting procedure (i.e. adding more modes and therefore model parameters) does not cause much divergence to occur between the coefficient of determination R_d^2 and the degrees of freedom-adjusted R^2 (defined in Eqs. B4 and B6 in Appendix B). This was verified for all of the dominant aerosol type cases studied here and the results of the calculations are presented in Sect. S2 of the Supplement. The percentage relative error (RE) between R^2 and R_d^2 was found to be very small for GMMs containing 1–6 modes – reaching a maximum value of RE = 0.060 %. Propagating this error into the square root of R^2 (the proxy for Pearson's ρ), we found that this had an effect only on the 4th decimal place and did not impinge on the results of the hypothesis testing procedure (see below) at the 95 % level. Having justified the use of the square root of R^2 as a proxy for Pearson's ρ, we then proceeded to construct confidence intervals on ρ using the Fisher transform (Fisher, 1921):

$$F(\rho) = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \tag{6}$$

where $F(\rho)$ is a transformed value of ρ that follows approximately a normal distribution with standard error $SE = 1/\sqrt{N-3}$ for a sample of N points (Fisher, 1921). If we note that the 0.975 quantile of the normal distribution has a z-score of 1.96, then the upper and lower 95 % confidence limits are simply: $F(\rho) \pm 1.96/\sqrt{N-3}$. We calculated these

AMTD
6, 10571–10615, 2013
Dominant aerosol type cases
M. Taylor et al.

limits for the value of ρ obtained for each GMM (1–6 modes). Two values of $F(\rho)$ (and hence R^2) show a significant statistical difference when the lower confidence limit of the larger $F(\rho)$ value does not overlap the upper confidence limit of the smaller $F(\rho)$ value. In the event of an overlap, the Welch t-statistic for unequal variances (Welch, 1947),

$$t = \left| \frac{F(\rho_1) - F(\rho_2)}{\sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}} \right| \tag{7}$$

reports a significant statistical difference when $t > 1.96$. In this way, a test was performed as modes were successively added to the Gaussian mixture. The optimal GMM fit occurs when adding a new mode does not lead to a significant statistical difference in $F(\rho)$ (or t in the case of over-lapping values) at the 95 % level of confidence. The calculation is presented in Table 2 for the automatic identification of the optimal (3-mode) mixture pertinent to the case of maximum marine sea salt illustrating this section.

Figure 5 shows the resulting 3-modal GMM fit ($b = 0.00$, $s = 0.000$, $R^2 = 0.998$) to the AVSD. Comparing with the AERONET bi-lognormal fit ($b = 0.00$, $s = 0.001$, $R^2 = 0.885$) and the OEV fit ($b = 0.00$, $s = 0.001$, $R^2 = 0.894$) shown in Fig. 3, the result of fitting with the GMM is clearly both quantitatively and qualitatively better.

4 Results

The case of fitting the AVSD of dominant marine aerosol at Lanai with a bi-lognormal shows that things are not so simple but that fitting problems could be overcome with the GMM method. This motivates a study of other geo-locations where the aerosol composition is also clearly defined so as to assess under what conditions, bi-lognormal fits are appropriate or not. With this in mind, in Table S3 in the Supplement accompanying this paper, we collect together the results of fitting the AVSD for each of the 4 dominant aerosol type cases with the methods introduced in Sects. 3.2 and 3.3.

AMTD
6, 10571–10615, 2013

Concerning the comparison of the goodness of fit obtained with the AERONET approach and the OEV method, in relation to the mode separation point, Table S3 shows that there are important differences. For peak dust, the value of $r_s = 0.439\,\mu m$ deduced by the AERONET inversion code is substantially smaller (−94.1 %) than that obtained by the OEV method ($r_s = 0.852\,\mu m$). For biomass burning, urban SU and marine sea salt, the mode separation point r_s deduced by the AERONET inversion code is larger than that obtained by the OEV method: $r_s = 0.648\,\mu m$ vs. $r_s = 0.546\,\mu m$, $r_s = 0.756\,\mu m$ vs. $r_s = 0.528\,\mu m$, and $r_s = 0.439\,\mu m$ vs. $r_s = 0.303\,\mu m$ respectively with relative errors of +18.7 %, +30.2 % and +31.0 %. These differences translate into a direct impact on the goodness of fit as measured by R^2. In the case of dust, $R^2 = 0.913$ for the AERONET bi-lognormal fit and $R^2 = 0.978$ using the OEV method – where the deduced value of r_s is almost double that of AERONET. Nevertheless, the AVSD is clearly well fit with 2 modes (i.e. a bi-lognormal) but application of the OEV method is necessary to ensure a good fit in the "shoulder" region. For biomass burning and urban SU, the value of R^2 obtained with both AERONET and the OEV method are higher ($R^2 > 0.982$) – suggesting that their AVSDs are very satisfactorily fit with a bi-lognormal distribution in both cases. As we have seen, the situation is different for peak marine (sea salt) where, despite a small improvement in the fit provided by the OEV method, the value of R^2 does not exceed 0.894. As was seen in Sect. 3, the bi-lognormal fit is not appropriate for this case – which involves a substantial mixture of aerosol types (only 60 % is sea salt).

With regard to the secondary microphysical parameters, the case of peak dust reveals another stark feature resulting from the existence of a "shoulder" in the distribution in the region: $0.4 \le r \le 1\,\mu m$. There is a large difference in the reported fine mode radius of dust obtained by AERONET ($r_f = 0.195\,\mu m$) and the OEV method ($r_f = 0.361\,\mu m$). While the optimal separation point r_s obtained with the OEV method produces a slightly better fit to the AERONET AVSD (as measured by R^2), its impact on the values of the secondary microphysical parameters is dramatic. With reference to Table S1, the tabulated entry closest to the optimal value obtained with the

AMTD
6, 10571–10615, 2013
Dominant aerosol type cases
M. Taylor et al.

OEV method is $r_s = 0.858\,\mu m$. With this mode separation point, Table S1 shows that the size of the relative error (using AERONET values as a reference) is very high for the fine volume (−115.9 %) and the fine radius (−87.0 %) – i.e. the value quoted by AERONET ($r_f = 0.195\,\mu m$) is 87.0 % *lower* than that found using the optimized fit bi-lognormal fit ($r_f = 0.365\,\mu m$). In contrast, in the case of peak biomass burning at Mongu, the tabulated entry closest to the optimal OEV value is $r_s = 0.567\,\mu m$. With this mode separation point, the relative errors of the microphysical parameters are in good agreement with those derived by AERONET and are in the narrow range: −3.3 % (for σ_c) to +2.6 % (for σ_f). The same is true for peak urban SU at GFSC-Washington, where the tabulated entry closest to the OEV optimum is $r_s = 0.528\,\mu m$. The similarity of this value to that quoted by AERONET for this day ($r_s = 0.756\,\mu m$) also translates into small relative errors, spanning the narrow range: −9.0 % (for σ_c) to +5.1 % (for r_c). The situation takes a turn for the worse in the case of peak marine aerosol. While both the AERONET and OEV methods point to similar mode separation points ($r_s = 0.885\,\mu m$ and $r_s = 0.894\,\mu m$ respectively), referencing the closest tabulated entry of $r_s = 0.885\,\mu m$ in Table S1, shows that the fine mode parameters: r_f, σ_f and V_f are strong under-estimated (−53.69 % to −68.73 %) and the coarse mode parameters r_c, σ_c and V_c are being over-estimated with a magnitude of ≈ 15 %. Dust and marine aerosol microphysical parameters are highly sensitive to the location of the mode separation point r_s.

Turning to the results of the GMM method, the fits to the AVSD for each of the 4 dominant aerosol type cases show a significant improvement over both those obtained by AERONET and using the OEV method. In Fig. 6 below, the interpolated AERONET AVSD retrieval is overlaid with the re-constructed AERONET bi-lognormal fit, the OEV bi-lognormal fit and the GMM optimal fit obtained by hypothesis testing.

Figure 6a shows that the GMM method flags up only peak dust aerosol as an $n = 2$ bi-lognormal. In contrast, it fits peak biomass burning, urban SU and marine (sea salt) AVSDs in Fig. 6b–d with $n = 3$ modes. While the qualitative appropriateness of these detections may be uncertain (see below) the results are statistically-significant in the

AMTD
6, 10571–10615, 2013

context of both the quoted error on the AERONET data and the deduced error on the GMM model. For example, in the case of peak dust aerosol (Fig. 6a), the GMM method does not fit the smallest fine mode which is marginally visible at $r \approx 0.1\,\mu m$. However, the GMM fit recovers the behavior of the AVSD in the central "shoulder" region $\approx 0.4 \leq r \leq 1\,\mu m$ where there is no local minimum and does not need to refer to a mode separation point. In the case of peak biomass burning aerosol (Fig. 6b), all three fitting methods agree quite well for the fine mode despite the possible existence of a slight skew towards smaller radius particles. However, only the GMM method is able to properly fit the coarse mode – the AERONET bi-lognormal fit and the OEV method fit both slightly under-estimate the location and amplitude of the coarse mode peak. This is due to a skew in the coarse mode peak. The GMM fit is able to correct for this effect by invoking the existence of a broad mode that peaks around $r \approx 0.2\,\mu m$. However, this peak is of low total volume concentration and spans too broad a range of particle radii – suggesting that it is physically perhaps less significant than the other two modes, and that arguably such an AVSD is *approximately* bi-lognormal. Note that the GMM method did not add a fourth mode to account for the possible skew in the fine mode peak. This is likely to be due to the fact that the errors on the AERONET data are relatively large in the region of the skew. For peak urban SU aerosol (Fig. 6c), all three methods agree excellently on the shape of the coarse mode ($r > 1\,\mu m$). In the fine mode region, AERONET and the OEV method slightly over-predict the location of the fine mode peak but in general are approximating its amplitude well. However, there is a clear skew in the fine region peak toward smaller radii. Here, despite large errors on the AERONET data at the low-radius end, the GMM method models the peak with a superposition of two modes – so as to fit this feature with two urban aerosol spikes. While this is physically plausible for urban pollution, we would argue that, just as for the case of biomass burning in Fig. 6b, the AVSD here is again *approximately* bi-lognormal. Finally, the case of peak marine (sea salt) aerosol (Fig. 6d) reveals a unique problem – that of fitting a double-hump in the coarse mode region. Only the GMM 3-mode model is able to successfully reproduce the shape of the AVSD in this difficult to fit region. The

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD

6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

fit with AERONET and the OEV method are unable to follow even small portions of the AVSD in the coarse mode region. For small particles below $r \approx 0.15$–$0.25\,\mu$m the OEV fit is very good although marginally over-estimates the amplitude of the fine mode peak. The AERONET approach is under-performing for this type of aerosol where mixtures are present (particularly at inhabited island sites and coastal regions). It is perhaps for this reason that only recently have models of "pure" marine aerosol been developed (Sayer et al., 2012).

5 Discussion

Our study of peak aerosol type cases has revealed that AERONET bi-lognormal fits to the AVSD are not valid for the dominant dust and marine (sea salt) aerosol cases considered here. They are appropriate for the dominant biomass burning and urban SU aerosol cases but cannot explain the (small) effect of the skew of the coarse mode towards larger particles in the case of biomass burning or the skew of the fine mode peak toward smaller particles in the case of urban SU. Relaxing the restriction on the location of the mode separation point used by AERONET with the OEV method slightly-improved the traditional bi-lognormal fit in these latter cases – but had a more prominent impact in the cases of dominant dust and marine aerosol. In the case of dominant dust aerosol, while the AERONET bi-lognormal fit to the coarse mode is good on the high radius side, the fit on the low radius side of the coarse peak is poor and especially poor over nearly the entire fine mode region. This was attributed to a strong under-estimation of the mode separation point by AERONET. Correcting for this with the OEV method led to a significant improvement in the goodness of fit with the fit to the coarse peak being almost indistinguishable for the raw AVSD. However, the constant total volume concentration constraint meant that this improvement was at the expense of the fit to fine mode features, which worsened. In the case of marine (sea salt) aerosol, the goodness of fit using bi-lognormals was the lowest observed for all of the dominant aerosol type cases. Furthermore, the problem of fitting both the fine mode peak and

AMTD
6, 10571–10615, 2013

simultaneously a double-peak in the coarse mode region could not be attributed to miscalculation of the position of the mode separation point – suggesting strongly that 2 modes are not sufficient to fit this type of aerosol. Overall, the OEV method produced an improvement over the AERONET method in all 4 dominant aerosol type cases studied. It also provided a way to make explicit the sensitivity of secondary microphysical parameters to the location of the mode separation point r_s. It was found that displacing r_s by even a single bin could lead to relative errors of the order of several tens of percent on the derived parameters. While this is precisely what gives the OEV method its capacity for improving bi-lognormal fits, it highlights that some care should be taken when using the values of microphysical parameters provided by AERONET – particularly in the case of dust and marine aerosol.

In relation to the GMM method, in all cases, it was possible to fit dominant aerosol type cases with a very high goodness of fit which is almost indistinguishable from the raw AERONET AVSD. For peak dust, the best fit was found to be bi-modal whereas for peak biomass burning (BC + OC), urban SU and marine (sea salt), the best fit was tri-modal. The use of iterated nonlinear least squares to obtain the microphysical coefficients was very efficient – although it was necessary to interpolate the AVSD with a 100-fold increase in the number of points (from 22 bins to 2200 bins) so as to avoid numerical instability (i.e. so that the propagated errors of the fit were stable at the 95 % level of confidence). Using the square root of R^2 as a proxy for the Pearson product-moment correlation coefficient ρ and applying the Fisher z transform allowed us to perform a test for a statistically-significant improvement in the fit with the addition of each additional mode. The null hypothesis at the 95 % confidence level provided a stopping condition that enabled automatic identification of the optimal number of modes in the mixture. The GMM method, in addition to providing a better overall fit, provides important details concerning the amplitude, location and width of each mode contributing to the mixture and hence allows for determination of secondary microphysical parameters in the case of $n > 2$ modes – something not currently possible with methods based on a single mode separation point. As such, the multi-modal content of the fit to the

AMTD
6, 10571–10615, 2013

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

AERONET AVSD is obtained directly. We would like to point out that the retrieval algorithm used by AERONET obtains AVSDs that would have the same *optical impact* on radiances as the "real" one expected from chemical composition analyses. However, as we have seen, bi-lognormals, reconstructed from the microphysics (geometric radii, standard deviations and volume concentrations) quoted by AERONET, provide poor fits to the parent AVSDs from which they are calculated in the case of peak dust and marine sea salt size distributions. As such, the total volume concentration (analogous to the "weight" of the measurement in mass terms) is erroneously distributed between the two modes in these cases. What the OEV method does is to better simulate the location, spread and amplitude of the fine and coarse modes so that the optical analogue is also better. The GMM method then generalizes this approach to a greater number of modes.

6 Conclusions

In this work, it was found that the results of the GOCART model could be used to rank and filter AERONET sites by aerosol type and used to select sites and individual daily-averaged records of the AERONET retrieved AVSD for dominant types. As a result, a dataset was produced that comprised of 4 daily-averaged AVSDs representative of 4 dominant aerosol types: dust, biomass burning (BC + OC), urban SU and marine SS. The AVSDs display a broad range of morphologies and provided a good test-bed for comparisons of AERONET microphysical parameter fits with bi-lognormals against the two new fitting methods that form the crux of this paper. Firstly, a method (the OEV method) was developed based on applying sensitivity analysis to the mode separation point r_s and led to improvements in the bi-lognormal fitting procedure used by AERONET particularly in the case of dust. Furthermore, it is able to have potentially wider applicability since it is independent of the *shape* of the AVSD (and therefore the aerosol type). Secondly, a method was developed for fitting the AVSD more generally with multiple ($n > 2$) modes (the GMM method) which produces consistently high

AMTD
6, 10571–10615, 2013
Dominant aerosol type cases
M. Taylor et al.

goodness of fit for all of the dominant aerosol types studied, could be automated to detect an optimal number of physically-plausible modes, and lifts restraints on both the total volume concentration and the need for additional mode separation points that would be required for example in a tri-lognormal equivalent of the OEV method. Furthermore, the GMM method gives access to the microphysical parameters associated with "higher" modes ($n > 2$) – not currently accessible by bi-lognormal fitting methods. In particular, this study showed that:

1. the constraint of constant total integrated volume concentration allows bi-lognormal AVSDs (and hence all secondary microphysical parameters) to be calculated from knowledge of a single free parameter – the mode separation point r_s,

2. deduction of r_s by detection of a local minimum in the size interval $0.439\,\mu m \leq r \leq 0.992\,\mu m$ is not always feasible – particularly in the case of desert dust where constraining r_s to this interval prevents the AERONET algorithm from obtaining the optimal bi-lognormal fit,

3. the OEV method is able to detect the optimum separation point instead by maximizing the goodness of fit (as expressed through R^2) independent of the shape of the AVSD and independent of the possible existence of a local minimum in the size interval $0.439\,\mu m \leq r \leq 0.992\,\mu m$,

4. it is possible to perform sensitivity analysis of the dependence of secondary microphysical parameters on (a) r_s and (b) the aerosol load (as measured by the AOD as a proxy),

5. bi-lognormal distributions do not necessarily provide the best fit to AERONET AVSDs – especially in the case of desert dust and marine sea salt AVSDs,

6. the GMM method allows automatic detection of a statistically-significant number of aerosol modes required to fit the AVSD on any given day and for a very diverse range of aerosol types.

AMTD
6, 10571–10615, 2013

With regard to estimation of the accuracy of the methods developed here, the mean bias b, standard error of the fit s, and the degrees of freedom-adjusted coefficient of determination R^2, were found to be very useful statistics for assessing the goodness of fit of the OEV method and the GMM method with respect to AERONET AVSDs. Consideration of the rate of change of s and R^2 with respect to changes in r_s was what gave the OEV method its capacity to automate the detection of the optimal separation point. In the case of the GMM method, the calculation of R^2 for consecutive mixtures (n modes vs. $n+1$ modes) in conjunction with Fisher statistics, allowed for the development of a stopping condition to automatically detect the optimal aerosol mixture that best fits the AVSD. Note that, while the estimated errors on AERONET-retrieved AVSD are modeled, they serve only as a visual point of reference since they are still yet to be verified. Having said this, application of nonlinear least squares fitting and standard error propagation allowed 95 % confidence bounds to be placed on the multi-modal fits to the interpolated AERONET AVSD in the GMM method.

It is hoped that the methods presented here will help contribute to the vast body of knowledge already provided by AERONET. AERONET retrievals are now being used for the accurate calculation of atmospheric broadband fluxes and aerosol radiative-forcing, and have been shown to agree very reasonably with available coincident ground-based flux observations in desert regions (Derimian et al., 2008) and also globally (Garcia et al., 2008). Furthermore, new retrieval algorithms are being developed to extend the capability of AERONET and to transfer knowledge to new remote sensing domains. For example, Dubovik et al. (2011) developed an inversion procedure for spectral multi-angle, polarimetric, satellite observations from POLDER/PARASOL, and it is hoped that the new methods introduced here will help contribute additional information content as this exciting field evolves.

AMTD
6, 10571–10615, 2013

Appendix A

Calculation of secondary microphysical parameters

From retrieved AERONET AVSDs, the volume concentration V occupied by particles spanning the range of sizes $r = [r_1, r_2]$ is easily calculated by integrating the $\mathrm{d}V/\mathrm{d}\ln r$ over the complete range of values of $\ln r$,

$$C_\mathrm{V} = \int_{r_1}^{r_2} \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r \quad \text{(A1)}$$

In principle, the aerosol number size distribution (ANSD) $\mathrm{d}N(r)/\mathrm{d}\ln r$ or $\mathrm{d}N(r)/\mathrm{d}r$ could equally well be used instead of the AVSD (e.g. see King et al., 1978). The conversion between the volume and number distribution parameters is also straight-forward (see for example Appendix A of Sayer et al., 2012). In particular, the spread σ remains the same for both AVSD and ANSD (King et al., 1978). However, it has been found that the AVSD is preferable to the ANSD as it is more accurate when inverting optical data that is highly sensitive to aerosol particle size (Dubovik et al., 2011). The AERONET inversion code approximates the AVSD using trapezium rule integration (Dubovik and King, 2000) and, while the option of allowing the use of lognormal-shaped bins was included in the calculations of Dubovik et al. (2006), it has only recently been found that accurate modeling of POLDER/PARASOL observations can only be achieved by optimizing the shape of each radial size bin in this way (Dubovik et al., 2011). For an overview of the properties of lognormal distributions in the physical sciences we refer the reader to Limpert et al. (2001). By using a mode separation point $r = r_\mathrm{s}$, the fine mode fraction η – a key parameter in aerosol forcing estimates (Kaufman et al., 2002) – can then be calculated as follows:

AMTD
6, 10571–10615, 2013

$$\eta = \frac{V_\mathrm{f}}{V_\mathrm{f} + V_\mathrm{c}} = \frac{\int_{r_1}^{r_s} \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r}{\int_{r_1}^{r_s} \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r + \int_{r_s}^{r_2} \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r} \tag{A2}$$

The fine mode fraction reflects the contribution of the fine mode to the total volume concentration. For desert (mineral) dust in the Sahara and the Arabian Peninsula it is low ($\eta \approx 25\,\%$), for multi-year averages of biomass burning in Africa and South America, and regional pollution in the eastern US, south-east Asian and Europe, the fine mode contribution is high and spans the range 92–95 %, while for maritime aerosol over the Pacific it is more moderate $\approx 67\,\%$ (Kaufman et al., 2002). Other important secondary microphysical parameters are statistical measures of central location and dispersion used to characterize individual aerosol modes in the AVSD. The logarithmic volume geometric mean radius (mean logarithm of radius) is a measure of the typical size of aerosol particles and is given by,

$$\ln r_\mathrm{V} = \frac{\int_{r_1}^{r_2} \ln r \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r}{\int_{r_1}^{r_2} \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r} \tag{A3}$$

The geometric mean radius is obtained by exponentiating the result. In addition, the geometric standard deviation is a measure of the spread (“width”) of the particle mode(s) and is given by,

$$\sigma_\mathrm{V} = \sqrt{\frac{\int_{r_1}^{r_2} (\ln r - \ln r_\mathrm{V})^2 \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r}{\int_{r_1}^{r_2} \frac{\mathrm{d}V(r)}{\mathrm{d}\ln r} \mathrm{d}\ln r}} \tag{A4}$$

AMTD
6, 10571–10615, 2013

Appendix B

Comparative statistics measures

Comparative statistics approaches necessarily involve the calculation of the differences (or residuals) between the model data ($\hat{y}_i$) and the target AERONET AVSD data (y_i). An initial picture is presented by the bias, b (or mean of the residuals) which, for N pairs of data points is given by,

$$b = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{y}_i \tag{B1}$$

This statistic is used to assess whether or not models systematically under-predict or over-predict. As a measure of the average difference, we avoid statistics such as the mean relative error and the χ^2 statistic since, while dependent on residuals, they are fractional quantities and contain y_i in the denominator. This can inflate their values in the tails of the AVSD where the values of y_i are extremely small compared to regions, for example, dominated by modal ("fine" and "coarse") peaks. Instead, statistics involving sums of squares of the residuals were selected that are strongly sensitive to outliers – and hence better able to discriminate between good and bad fits. In particular, we calculated the sum of squares of the residual errors (SSE),

$$\mathrm{SSE} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{B2}$$

From this, the standard error of the fit, s was calculated,

$$s = \sqrt{\frac{\mathrm{SSE}}{N-p-1}} \tag{B3}$$

AMTD
6, 10571–10615, 2013

where N is the number of points and p is the number of independent model parameters. This is our choice of location (average) measure and is the unbiased sample estimator version of the traditional root mean square of the errors (RMSE). This measure is sensitive to outliers (due to its dependence on SSE) and is also an interval scale quantity – i.e. it has the same measurement units as y_i. In order to asses the dispersion (or spread) in the residuals of the fits, we decided to use a regression statistic known as the coefficient of determination (R_d^2),

$$R_\mathrm{d}^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \tag{B4}$$

where SST is the total sum of squares of the difference between the target AVSD data and its *mean* value $\bar{y}_i$:

$$\mathrm{SST} = \sum_{i=1}^{N} (y_i - \bar{y}_i)^2 \tag{B5}$$

R_d^2 measures how well a model reproduces data in terms of the amount of the total variance it explains (Steel and Torrie, 1960) and ranges from 0 to 1 such that $R_\mathrm{d}^2 = 0.95$ is taken to mean, for example, that the model fit explains 95 % of the total variance in the data. Care must be taken when using this statistic since models involving more modes have more model parameters p and their improved fit is reflected in a smaller value of SSE. This results in a correspondingly gradual increase in the value of R_d^2 with the number of parameters. To compensate for this effect, we therefore use, in this paper, the *degrees of freedom-adjusted* R^2 statistic which penalizes the value of R_d^2 as extra parameters are included in the model:

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \left(\frac{N-1}{N-p-1} \right) \tag{B6}$$

AMTD
6, 10571–10615, 2013

This is known to be a good indicator of fit quality when comparing nested models – i.e. a series of models each of which adds additional coefficients to the previous model (Harel, 2009) as is the case of the GMM method outlined in Sect. 3.3.

Appendix C

The GMM fitting procedure and propagation of errors

The identification of the best mixture in the GMM method of Sect. 3.3 is an optimization problem involving the minimization of the residuals between the observed AVSD data and the GMM model:

$$\varepsilon = y - [f(x, \beta)] \tag{C1}$$

where ε is an $N \times 1$ vector of "residuals", y is an $N \times 1$ vector of "target" values (the AVSD), x is an $N \times p$ "design matrix" for the model (in our case an n-Gaussian), β is a $p \times 1$ vector of "parameters" (the coefficients – i.e. the mode volumes, radii and standard deviations), and f is a function of x and β (i.e. the linear sum of independent Gaussians expressed by Eq. (1) in Sect. 1 of this paper. Such problems cannot be solved using simple matrix techniques due to the nonlinear dependence of ε on β in the Gaussians. However, an approximate solution can be obtained by iteratively applying the nonlinear least squares method as follows:

- Step 1: provide an initial estimate for each coefficient
- Step 2: produce the fit curve for the current set of coefficients
- Step 3: adjust the coefficients using a conjugate-gradient method
- Step 4: iterate the process, returning to step 2 until the fit converges

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD
6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

This method was scripted in MATLAB using its in-built object-oriented scripting language, and required initial constraints to be placed on the values of the coefficients. For reproducibility of results obtained here, we provide the interested reader with the parameters used in the fitting procedure. The lower bounds were set to 0.0005 since the amplitudes, locations and spreads must be non-zero and positive for each GMM. The upper bounds were set to 3, ln(15) and 3 for the amplitudes, log-locations and spreads respectively. The fit at each step was then obtained by minimizing the least absolute (total) residual (LAR). This "Trust region" method also calculates the Jacobian of $f(x,\beta)$ to determine whether or not the fit is improving (based on the direction and magnitude of the previous adjustment). The minimum and maximum change in the coefficients for this finite difference Jacobian was set to 10^{-8}. Regarding convergence criteria, we set the maximum number of model evaluations in each iteration to the default value of 600 and the maximum number of overall iterations to its default value of 400. The stopping condition on the minimum value of the LAR was set to 10^{-6} which is 1/100th of the minimum volume concentration in our dataset. This entire recipe was then repeated 8 times – being applied to GMMs containing 1 to 8 modes in succession.

Confidence bounds for the optimal GMM (obtained with the fitting procedure that uses the stopping condition outlined in Sect. 3.3) was calculated by the standard approach of propagating errors. This could be achieved because a closed form exists for the optimal GMM as given by Eq. (1) for n-modes. For each independent mode,

$$f_i = a_i e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} \tag{C2}$$

the standard error SE_i is given by,

$$\mathrm{SE}_i = \sqrt{\left(\frac{\partial f_i}{\partial a_i} \times \mathrm{SE}(a_i)\right)^2 + \left(\frac{\partial f_i}{\partial b_i} \times \mathrm{SE}(b_i)\right)^2 + \left(\frac{\partial f_i}{\partial c_i} \times \mathrm{SE}(c_i)\right)^2} \tag{C3}$$

AMTD
6, 10571–10615, 2013
Dominant aerosol type cases
M. Taylor et al.

in terms of the partial derivatives,

$$\frac{\partial f_i}{\partial a_i} = a_i e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} \tag{C4}$$

$$\frac{\partial f_i}{\partial b_i} = \frac{2a_i(\ln r - b_i)}{c_i^2} e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} \tag{C5}$$

$$\frac{\partial f_i}{\partial c_i} = \frac{2a_i(\ln r - b_i)^2}{c_i^3} e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} \tag{C6}$$

The upper and lower 95 % confidence bounds for the overall GMM fit are then obtained by noting that the standard errors of the modes also combine as root mean squares and are centred on the sum of the modes, i.e.:

$$\text{upper bound} = \sum_{i=1}^{n} a_i e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} + 1.96\sqrt{\sum_{i=1.n} (\text{SE}_i)^2} \tag{C7}$$

$$\text{lower bound} = \sum_{i=1}^{n} a_i e^{-\left(\frac{\ln r - b_i}{c_i}\right)^2} - 1.96\sqrt{\sum_{i=1.n} (\text{SE}_i)^2} \tag{C8}$$

Supplementary material related to this article is available online at http://www.atmos-meas-tech-discuss.net/6/10571/2013/amtd-6-10571-2013-supplement.pdf.

Acknowledgements. The authors would like to thank all the principal investigators of AERONET for their free provision of timely and high quality aerosol retrieval data, Mian Chin and Tom Kucsera for making GOCART model data available during the years 2001–2005 for all AERONET sites at the AERONET data synergy portal, the MODIS rapid response team for making available true colour images with fire sources from AQUA and TERRA, and David Giles and Brent Holben at NASA for curating the AERONET data synergy tool at http://aeronet.gsfc.nasa.gov/cgi-bin/bamgomas_interactive. MT was supported by a Marie-Curie IEF funded project "AEROMAP: Global mapping of aerosol properties using neural network inversions of ground and satellite based data", CN: 300515, and would like to thank the members of IERSD-NOA for excellent training in the field, and in particular, Spiros Lykoudis for useful discussions on statistical hypothesis testing of nested models.

References

Chin, M., Rood, R. B., Lin, S. J., Müller, J. F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties, J. Geophys. Res., 105, 24671–24687, 2000.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, J. Atmos. Sci., 59, 461–483, 2002.

Derimian, Y., Leon, J.-F., Dubovik, O., Chiapello, I., Tanré, D.,Sinyuk, A., Auriol, F., Podvin, T., Brogniez, G., and Holben, B. N.: Radiative properties of aerosol mixture observed during the dry season 2006 over M'Bour, Senegal (African Monsoon Multidisciplinary Analysis campaign), J. Geophys. Res.-Atmos., 113, D00C09, doi:10.1029/2008JD009904, 2008.

Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, J. Geophys. Res., 105, 20673–20696, 2000.

Dubovik, O., Smirnov, A., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F., and Slutsker, I.: Accuracy assessment of aerosol optical properties retrieval from AERONET sun and sky radiance measurements, J. Geophys. Res., 105, 9791–9806, 2000.

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci., 59, 590–608, 2002.

Dubovik, O., Sinyuk, A., Lapyonok, T., Sinyuk, A., Mishchenko, M. I., Yang, P., Eck, T. F., Volten, H., Munoz, O., Veihelmann, B., van der Zander, W. J., Sorokin, M., and Slutsker, I.:

AMTD
6, 10571–10615, 2013

Application of light scattering by spheroids for accounting for particle non-sphericity in remote sensing of desert dust, J. Geophys. Res., 111, D11208, doi:10.1029/2005JD006619, 2006.
Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, Atmos. Meas. Tech., 4, 975–1018, doi:10.5194/amt-4-975-2011, 2011.
Fisher, R. A.: On the "probable error" of a coefficient of correlation deduced from a small sample, Metron, 1, 3–32, 1921.
Garcia, O. E., Diaz, A. M., Exposito, F. J., Diaz, J. P., Dubovik, O., Dubuisson, P., Roger, J.-C., Eck, T. F., Sinyuk, A., Derimian, Y., Dutton, E. G., Schafer, J. S., Holben, B. N., and Garcia, C. A.: Validation of AERONET estimates of atmospheric solar fluxes and aerosol radiative forcing by ground-based broadband measurements, J. Geophys. Res., 113, D21207, doi:10.1029/2008JD010211, 2008.
Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., and Lin, S. J.: Sources and distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106, 20255–20273, 2001.
Hansen, J. E. and Travis, L. D.: Light scattering in planetary atmospheres, Space Sci. Rev., 16, 527–610, 1974.
Harel, O.: The estimation of R^2 and adjusted R^2 in incomplete data sets using multiple imputation, J. Appl. Stat., 36, 1109–1118, 2009.
Hasekamp, O. P. and Landgraf, J.: Retrieval of aerosol properties over land surfaces: capabilities of multi-viewing-angle intensity and polarization measurements, Appl. Optics, 46, 3332–3344, 2007.
Kaufman, Y. J. and Holben, B. N.: Hemispherical backscattering by biomass burning and sulfate particles derived from sky measurements, J. Geophys. Res., 101, 19433–19445, 1996.
Kaufman, Y. J., Tanré, D., Boucher, O.: A satellite view of aerosols in the climate system, Nature, 419, 215–223, 2002.
King, M. D., Byrne, D. M., Herman, B. M., and Reagan, J. A.: Aerosol size distributions obtained by inversion of spectral optical depth measurements, J. Atmos. Sci., 21, 2153–2167, 1978.
King, M. D., Kaufman, Y. J., Tanré, D., and Nakajima, T.: Remote sensing of tropospheric aerosols from space: past, present, and future, B. Am. Meteorol. Soc., 2229–2259, 2009.

Kokhanovsky, A. A., Deuzé, J. L., Diner, D. J., Dubovik, O., Ducos, F., Emde, C., Garay, M. J., Grainger, R. G., Heckel, A., Herman, M., Katsev, I. L., Keller, J., Levy, R., North, P. R. J., Prikhach, A. S., Rozanov, V. V., Sayer, A. M., Ota, Y., Tanré, D., Thomas, G. E., and Zege, E. P.: The inter-comparison of major satellite aerosol retrieval algorithms using simulated intensity and polarization characteristics of reflected light, Atmos. Meas. Tech., 3, 909–932, doi:10.5194/amt-3-909-2010, 2010.

Limpert, E., Stahel, W. A., and Abbt, M.: Log-normal distributions across the sciences: keys and clues, J. Bioscience, 51, 341–352, 2001.

Omar, A. H., Won, J. G., Winker, D. M., Yoon, S. C., Dubovik, O., and McCormick, M. P.: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, J. Geophys. Res.-Atmos., 110, D10S14, doi:10.1029/2004JD004874, 2005.

Omar, A. H., Winker, D. M., Kittaka, C., Vaughan, M. A., Liu, Z., Hu, Y., Trepte, C. R., Rogers, R. R., Ferrare, R. A., Lee, K. P., Kuehn, R. E., and Hostetler, C. A.: The CALIPSO automated aerosol classification and Lidar ratio selection algorithm, J. Atmos. Ocean. Tech., 26, 1994–2014, 2009.

O'Neill, N. T., Ignatov, A., Holben, B. N., and Eck, T. F.: The lognormal distribution as a reference for reporting aerosol optical depth statistics; Empirical tests using multi-year, multi-site AERONET sunphotometer data, Geophys. Res. Lett., 27, 3333–3336, 2000.

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., and Thulasiraman, S.: Spectral discrimination of coarse and fine mode optical depth, J. Geophys. Res., 108, 4559–4573, 2003.

Remer, L. A. and Kaufman, Y. J.: Dynamic aerosol model: Urban industrial aerosol, J. Geophys. Res., 203, 13859–13871, 1998.

Remer, L. A., Gasso, S., Hegg, D. A., Kaufman, Y. J., and Holben, B. N.: Urban/industrial aerosol: ground-based Sun/sky radiometer and in situ measurements, J. Geophys. Res., 102, 16849–16859, 1997.

Remer, L. A., Kaufman, Y. J., Holben, B. N., Thompson, A. M., and McNamara, D. P.: Biomass burning aerosol size distribution and modeled optical properties, J. Geophys. Res., 103, 31879–31891, 1998.

Remer, L. A., Kaufman, Y. J., Tanre, D., Mattoo, S., Chu, D. A., Martins, J., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, J. Atmos. Sci., 62, 947–973, 2005.

Sayer, A. M., Smirnov, A., Hsu, N. C., and Holben, B. N.: A pure marine aerosol model, for use in remote sensing applications, J. Geophys. Res., 117, 1–25, 2012.

AMTD
6, 10571–10615, 2013

Steel, R. G. D. and Torrie, J. H.: Principles and Procedures of Statistics with Special Reference to the Biological Sciences, McGraw Hill, New York, 187–287, 1960.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I., Iversen, I., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: Analysis and quantification of the diversities of aerosol life cycles within AeroCom, Atmos. Chem. Phys., 6, 1777–1813, doi:10.5194/acp-6-1777-2006, 2006.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Feichter, J., Fillmore, D., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I. S. A., Iversen, T., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, M. S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: The effect of harmonized emissions on aerosol properties in global models – an AeroCom experiment, Atmos. Chem. Phys., 7, 4489–4501, doi:10.5194/acp-7-4489-2007, 2007.

Welch, B. L.: The generalization of "Student's" problem when several different population variances are involved, Biometrika, 34, 28–35, 1947.

Whitby, K. T.: The physical characteristics of sulfur aerosols, Atmos. Environ., 12, 135–159, 1978.

Zhang, K., Wan, H., Wang, B., Zhang, M., Feichter, J., and Liu, X.: Tropospheric aerosol size distributions simulated by three online global aerosol models using the M7 microphysics module, Atmos. Chem. Phys., 10, 6409–6434, doi:10.5194/acp-10-6409-2010, 2010.

Table 1. Dataset comprising the 4 dominant aerosol type cases studied in this work. Peak percentages are highlighted. Note that SS at Lanai is the least "dominant" aerosol type case with marine aerosol being mixed with other aerosols in the proportion ≈ 60 % : 40 %.

AERONET site	Peak date	SU	OC	BC	DU	SS	OC + BC
Banizoumbou	16 Mar 2005	1.02 %	0.74 %	0.31 %	97.91 %	0.03 %	1.04 %
Mongu	14 Aug 2003	5.61 %	77.36 %	16.76 %	0.22 %	0.05 %	94.12 %
GSFC-Washington	17 Aug 2005	87.53 %	8.31 %	2.72 %	1.38 %	0.05 %	11.04 %
Lanai	21 Feb 2002	28.92 %	5.31 %	2.31 %	3.32 %	60.14 %	7.61 %

AMTD
6, 10571–10615, 2013

Table 2. Statistical testing of R^2 during application of the GMM fit to the interpolated AVSD of dominant marine (sea salt) aerosol at Lanai, Hawaii on the 21 January, 2002. In the case of a single mode, statistical testing is not performed. In the case of both $n = 2$ and $n = 3$ modes, the lower confidence limit $\mathrm{CI}_2(l)$ of the larger-valued $F(\rho_2)$ is less than the upper confidence limit $\mathrm{CI}_1(u)$ of the lower-valued $F(\rho_1)$, and the t-Welch statistic being > 1.96 shows that there is a statistically-significant improvement in R^2. In the case of $n = 4$ modes, two things should be noted. Firstly, that $F(\rho_2) < F(\rho_1)$ (i.e. a reduction in the improvement in the goodness of fit). Secondly, since the lower confidence limit $\mathrm{CI}_1(l)$ of the larger-valued $F(\rho_1)$ is greater than the upper confidence limit $\mathrm{CI}_2(u)$ of the lower-valued $F(\rho_2)$ then this reduction is statistically-significant, i.e. the addition of the 4th mode worsens the fit and the optimal number of modes is therefore $n = 3$.

n Modes	$R^2(n)$	$R^2(n+1)$	$F(\rho_1)$	$F(\rho_2)$	$\mathrm{CI}_1(l)$	$\mathrm{CI}_1(u)$	$\mathrm{CI}_2(l)$	$\mathrm{CI}_2(u)$	t Welch
1	0.777								
2	0.777	0.819	1.38	1.50	1.34	1.54	1.46	1.54	3.87
3	0.819	0.998	1.50	3.80	1.46	3.84	3.76	3.84	76.26
4	0.998	0.993	3.80	3.17	3.76	3.21	3.13	3.21	20.80

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD
6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

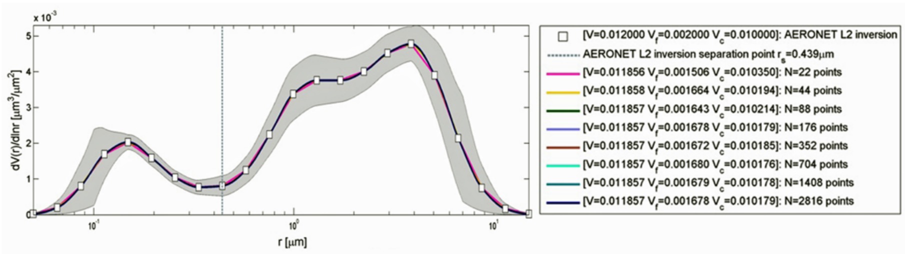

Fig. 1. Interpolation of the AVSD with between 22 and 2816 equidistant logarithmically-spaced points (7 doublings) for dominant marine (sea salt) aerosol at Lanai on the 21 January 2002. Only the interpolations with 22 points (pink) and 44 points (yellow) show visible deviations from the interpolation having the maximum 2816 points. Also shown is the value of the mode separation point $r_s = 0.439\,\mu m$ provided by the AERONET retrieval on this day. The grey band is the uncertainty on the AERONET AVSD obtained by following the approach described in the next paragraph.

AMTD
6, 10571–10615, 2013

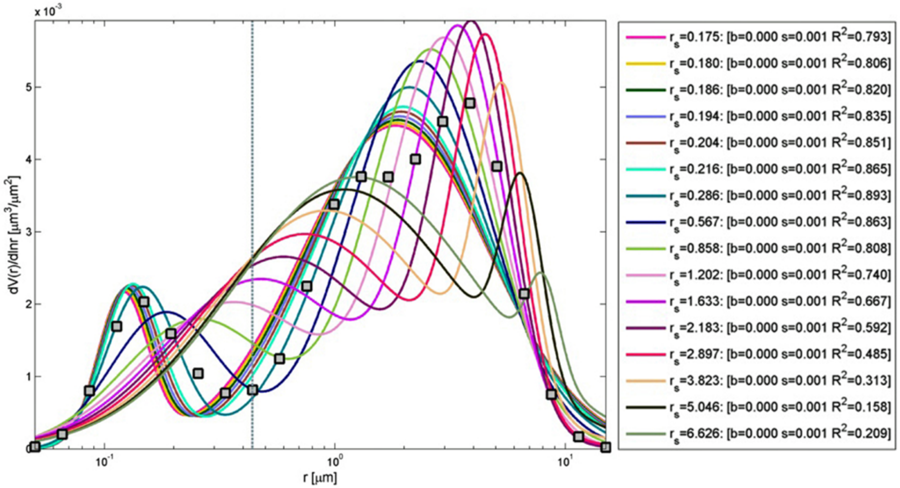

Fig. 2. Sensitivity analysis of the equivalent volume bi-lognormal fit to the AERONET AVSD data with varying mode separation point r_s for dominant marine (sea salt) aerosol at Lanai, Hawaii on the 21 January 2002. The grey squares are the values of the AERONET AVSD.

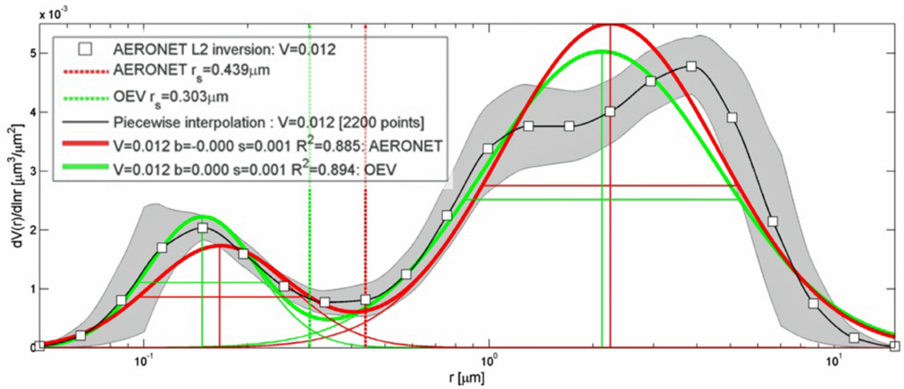

Fig. 3. Comparison of the interpolated AVSD for dominant marine (sea salt) aerosol at Lanai, Hawaii on the 21 January 2002 with the AERONET bi-lognormal fit and the optimized equivalent volume (OEV) fit. The grey band is the uncertainty on the AERONET AVSD.

AMTD
6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

AMTD
6, 10571–10615, 2013

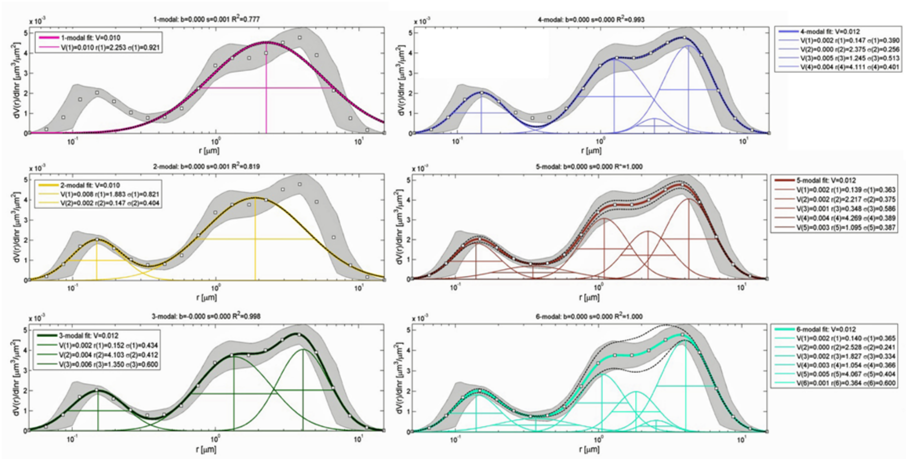

Fig. 4. Gaussian mixture model (GMM) 1–6 modal fits of the interpolated AVSD for dominant marine (sea salt) aerosol at Lanai, Hawaii on the 21 January 2002. The grey shaded region is the error on the AERONET data and the black dotted lines (most visible in the 5 and 6-modal plots) are the 95 % confidence level curves on the fit.

AMTD
6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.

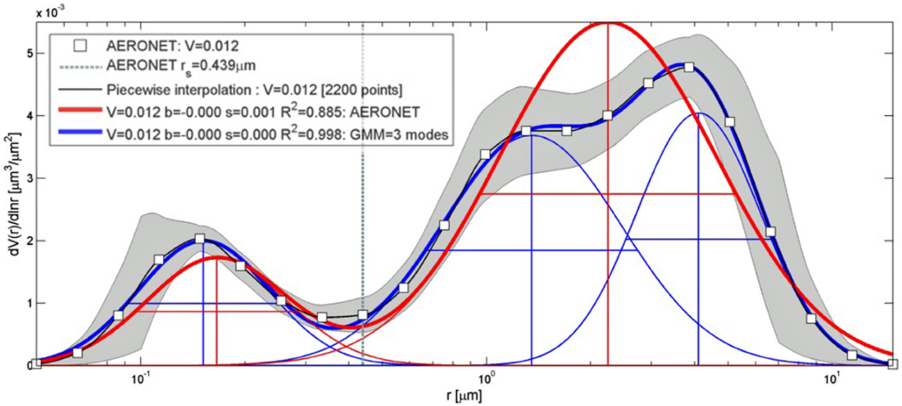

Fig. 5. Comparison of the interpolated AVSD for dominant marine (sea salt) aerosol at Lanai, Hawaii on the 21 January 2002 with the AERONET bi-lognormal fit and the optimal tri-modal Gaussian mixture model (GMM) fit. The grey band is the uncertainty on the AERONET AVSD.

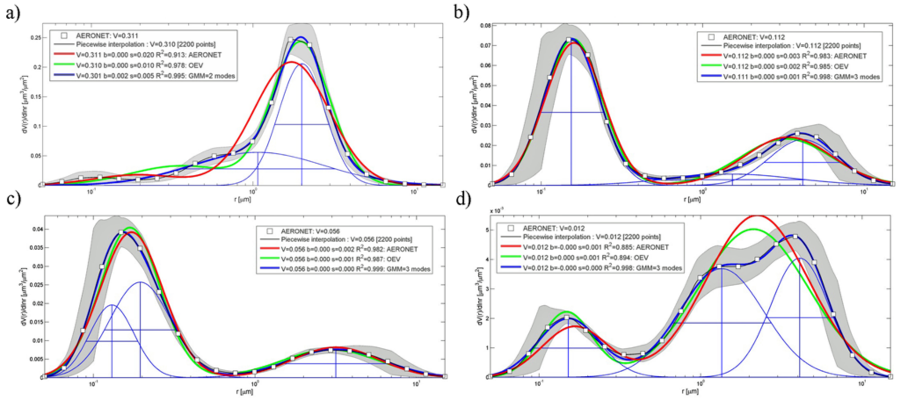

Fig. 6. Comparison of the interpolated AVSD, the AERONET bi-lognormal fit, the OEV bi-lognormal fit and the GMM optimal fit for **(a)** dominant dust aerosol at Banizoumbou, Niger on the 16 March 2005, **(b)** dominant biomass burning aerosol at Mongu, Zambia on the 14 August 2003, **(c)** dominant urban SU aerosol at Washington-GSFC, US on the 17 August 2005, and **(d)** dominant marine (sea salt) aerosol at Lanai, Hawaii on the 21 February 2002. The grey band is the uncertainty on the AERONET AVSD.

AMTD
6, 10571–10615, 2013

Dominant aerosol type cases

M. Taylor et al.