

Inclusion of *a priori* information in an analytic eigenfunction inversion of Mie extinction measurements

Gail Patricia Box

The analytic eigenfunction inversion technique to retrieve aerosol columnar size distributions from Mie extinction measurements has been extended to include *a priori* information, specifically surface area. The earlier (standard) and new (subtracted) techniques are compared by use of synthetic data that cover typical aerosol size distributions. Two different measurement wavelength ranges are considered. It is shown that the most appropriate inversion technique depends on the particle sizes, with the standard technique being more appropriate for small particles and the subtracted being more appropriate for large particles. Also presented is a simple method to determine whether a particular inversion technique is likely to produce meaningful results with a particular data set. © 2005 Optical Society of America

OCIS codes: 010.1320, 280.1100, 010.1110, 280.1310, 290.4020.

1. Introduction

One of the techniques used to remotely sense the size distribution of atmospheric aerosols is based on the measurement of aerosol extinction as a function of wavelength and the inversion of the resulting Fredholm integral equation:

$$\tau(\lambda) = \int_0^{\infty} \pi r^2 Q(r, \lambda, m) n(r) dr. \quad (1)$$

Here $\tau(\lambda)$ is the aerosol extinction (optical depth) at wavelength λ , Q is the Mie extinction efficiency factor,^{1,2} m is the complex refractive index, and $n(r)$ is the aerosol size distribution.

The inversion of Eq. (1) is a classic ill-posed problem. A number of different techniques have been employed, including constrained linear,^{3–5} iterative,^{6,7} analytic,^{8,9} kernel expansion,¹⁰ singular function theory,¹¹ and empirical orthogonal functions.^{12,13}

McWhirter and Pike¹⁴ introduced a new approach, suitable for product kernels such as the Laplace transform and based around the Mellin transform of the kernel. Viera and Box¹⁵ applied this technique to

the anomalous diffraction approximation to the Mie extinction¹ and later to two diffraction approximations to forward scattering.¹⁶ These applications used an approximate kernel rather than exact Mie theory and assumed there were no restrictions on the measurement range. Box *et al.*¹⁷ applied this technique with a full Mie kernel and for a finite, realistic measurement range.

In their earlier study, Viera and Box¹⁵ showed that more accurate inversions for size distribution could be obtained if certain *a priori* information was included. In this paper I extend that research to the case of full Mie theory and a realistic measurement range. The basic theory is outlined in Section 2, and in Section 3 I compare the earlier standard inversion technique with the new subtracted inversion technique. Section 4 contains the discussion and conclusions.

2. Analytic Eigenfunction Theory

Since the details of the analytic eigenfunction theory have been given elsewhere,^{14–16} only a brief outline is given here. The generic Fredholm integral equation can be written as

$$g(x) = \int_0^{\infty} K(x, y) f(y) dy, \quad (2)$$

where $g(x)$ is the data function, $K(x, y)$ is the kernel, and $f(y)$ is the function to be retrieved. Analytic eigenfunction theory applies only to product kernels, that

G. P. Box (gpb@newt.phys.unsw.edu.au) is with the School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia.

Received 16 March 2004; revised manuscript received 20 October 2004; accepted 22 October 2004.

0003-6935/05/071288-08\$15.00/0

© 2005 Optical Society of America

is, kernels of the form $K(x, y) = K(xy)$. The Mie theory extinction kernel is a product kernel if we write it in the form $Q(r, \lambda, m) = Q(kr, m)$, where $k = 2\pi/\lambda$ and the refractive index m is assumed constant.

The eigenfunctions $\phi_\omega(y)$ and eigenvalues $\lambda(\omega)$ of such a product kernel are defined by

$$\int_0^\infty K(xy)\phi_\omega(x)dx = \lambda(\omega)\phi_\omega(y). \quad (3)$$

McWhirter and Pike¹⁴ were able to construct $\lambda(\omega)$ and $\phi_\omega(y)$ by taking the Mellin transform of Eq. (3). They found

$$\lambda(\omega) = |\widetilde{K}(1/2 + i\omega)|,$$

$$\phi_\omega(y) = y^{-1/2-i\omega}\chi(\omega),$$

where

$$\chi(\omega) = [\widetilde{K}(1/2 + i\omega)/\pi\lambda(\omega)]^{1/2},$$

$$\widetilde{K}(1/2 + i\omega) = \int_0^\infty t^{-1/2+i\omega}K(t)dt \quad (4)$$

is the Mellin transform of the kernel K . Convergence of this integral places some constraints on $K(t)$. As it stands, the Mellin transform of the Mie extinction kernel clearly diverges at the large t end. This is easily overcome if we factor the integral equation and multiply K by a factor such as t^{-1} (for full details, see Viera and Box¹⁵).

The retrieval of $f(y)$ requires us to effectively expand both it and $g(x)$ in terms of the eigenfunctions ϕ . First we define

$$G(\omega) = \chi(\omega) \int_0^\infty g(x)x^{-1/2-i\omega}dx,$$

and then $f(y)$ is obtained from the truncated summation

$$f(y) = y^{-1/2}\text{Re} \int_0^{\omega_m} G(\omega)y^{-i\omega}\chi(\omega)/\lambda(\omega)d\omega. \quad (5)$$

The cutoff for this integral is determined by the decay of the eigenvalue spectrum $\lambda(\omega)$, relative to the experimental error level. In general ω_m will be the value for which $\lambda(\omega)$, normalized relative to the largest eigenvalue, is equal to the relative error in the measurements.

A. Standard Inversion

In the case considered by Box *et al.*,¹⁷ which is called the standard kernel, to satisfy convergence requirements on Eqs. (4), we selected as our kernel $K(kr)$

$= \pi Q(kr)/kr$. Hence $f(r) = r^3n(r)$ and $g(k) = k^{-1}\tau(k)$. Thus, in this case, Eq. (2) becomes

$$k^{-1}\tau(k) = \int [\pi Q(kr)/kr][r^3n(r)]dr.$$

The next step is the construction of the Mellin transform of $K(t)$ by numerical integration of Eqs. (4) for values of ω from 0 to 10 by use of the same approach as in Viera and Box.¹⁵

In practice, measurements are available only for a finite range of wavelengths, and thus $g(k)$ is defined only for a corresponding range $k_1 < k < k_2$. Assumptions have to be made about its behavior outside of this range in the final retrieval process. The assumptions made by Box *et al.*¹⁷ were

$$g(k) \approx g(k_1)(k/k_1)^2 \quad k < k_1,$$

$$g(k) \approx g(k_2)(k_2/k) \quad k > k_2.$$

B. Subtracted Inversion

In their study on the application of analytic eigenfunction theory to the anomalous diffraction approximation, Viera and Box¹⁵ considered three distinct cases corresponding to three distinct kernels. These were the standard kernel outlined above and two cases where it was assumed there was some *a priori* information available, leading to what is called the subtracted kernels. The advantage of the subtracted kernels is that the eigenvalue spectrum decays much more slowly, permitting a higher value of ω_m to be used in Eq. (5) and leading to more accurate retrievals.¹⁷ (See Box *et al.*¹⁷ for plots illustrating the decay of the eigenvalue spectrum for the standard and subtracted kernels.)

The case considered here is that of where the total cross-sectional area (second moment) of the size distribution is known. The cross-sectional area A is related to the size distribution by $A = \int \pi r^2 n(r)dr$. If we define the subtracted kernel $K_s(kr)$ as $K_s(kr) = \pi[Q(kr) - 2]$, then from Eq. (2) we can write the subtracted data function as

$$g_s(k) = \int_0^\infty \pi[Q(kr) - 2]f_s(r)dr,$$

where $f_s(r) = r^2n(r)$. There are no convergence problems with this kernel because the leading term in the asymptotic expansion of the integrand now behaves like $(kr)^{-7/6}$ (see the Appendix of Box *et al.*¹⁷).

Relating this to the size distribution, optical thickness, and cross-sectional area, we can see that

$$\begin{aligned} g_s(k) &= \tau(k) - 2 \int_0^\infty \pi r^2 n(r)dr \\ &= \tau(k) - 2A. \end{aligned}$$

The Mellin transform of $K_s(kr)$ is constructed by numerical integration of Eqs. (4) for values of ω from 0 to 10. The subtracted inversion procedure follows the same steps as for the standard inversion, but with the subtracted forms of the data function and Mellin transform.

Again, when faced with a finite measurement range, and thus $g_s(k)$ defined only for a finite range of values $k_1 < k < k_2$, we need to make some assumptions about the behavior of $g_s(k)$ outside this range. To define the behavior of $g_s(k)$ outside the measurement range, we need to consider the behavior of $g_s(k)$ at the extremes.

For small values of k , $\tau(k) \rightarrow 0$ as $k \rightarrow 0$ so that $g_s(k) = -2A$ when $k = 0$. If we assume that the data function is linear between the lower limit and $k = 0$, we can approximate the data function behavior in this region by

$$g_s(k) \approx [g_s(k_1) + 2A]k/k_1 - 2A \quad k < k_1.$$

At the other extreme, it has been shown that, for large values of k (Refs. 18 and 19),

$$\tau(k) \sim 2A + 1.9923861\pi M_{4/3} k^{-2/3} + O(k^{-1}),$$

where the i th moment of the size distribution is given by

$$M_i = \int_0^\infty r^i n(r) dr. \quad (6)$$

Thus, for the upper limit, we can say that

$$g_s(k) \approx g_s(k_2)(k_2/k)^{2/3} \quad k > k_2.$$

In practice we do not normally have any estimates for the total cross-sectional area of the size distribution, so the inversions are done as a two-step process. The first step is to invert the data by use of the standard inversion technique and, from the size distribution obtained, calculate the second moment of the distribution and hence a first estimate A_{integ} for the cross-sectional area. The second stage uses the subtracted inversion technique with A_{integ} as the *a priori* information.

When the subtracted inversion is carried out and the resulting size distribution is used to calculate $\tau(k)$ for comparison with the input data, the errors are often larger than with the standard inversion, particularly if A_{integ} is not close to the true cross-sectional area. Thus an iterative approach was adopted. Starting with A_{integ} as the first estimate for A , at each successive iteration in the subtracted inversion process a new estimate for A is calculated by increasing or decreasing the previous value by a small fraction. This process continues until the value A_{iter} , which gives the smallest error when the calculated optical thicknesses is compared with the measured values, is

Table 1. Size Distribution Characteristics

Parameter	Value					
r_{eff}	0.2	0.2	0.4	0.4	0.8	0.8
σ	1.6	2.0	1.6	2.0	1.6	2.0
r_g	0.115	0.060	0.230	0.120	0.461	0.241
r_a	0.179	0.157	0.358	0.315	0.716	0.629
α	48.50	105.66	12.13	26.42	3.0	6.60

found. The direction in which A needs to be adjusted is easily established after the first few iterations, and in most cases only five or six iterations are needed. (This procedure converges faster than when the value of A calculated from the second moment of the new size distribution is used and also avoids problems that can arise due to the oscillatory nature of the eigenfunctions.)

3. Comparison of Standard and Subtracted Inversions

A. Simulated Data Sets

The two inversion procedures were compared by use of six sets of simulated optical thickness measurements derived from log-normal size distributions. Although size distributions are often multimodal, in this paper I consider only single-mode distributions representative of the accumulation mode that is optically active size range at these wavelengths. A log-normal distribution was chosen because it allows both the geometric mean radius r_g and the spread of the distribution σ to be varied. For easier comparison, data were normalized to give the size distribution a value of 1 at the mode radius r_m of the volume-weighted distribution. The log-normal distribution is given by

$$n(r) = \frac{N}{\sqrt{2\pi r} \ln \sigma} \exp\left[-1/2\left(\frac{\ln r - \ln r_g}{\ln \sigma}\right)^2\right], \quad (7)$$

where N is the total number of particles. The volume-weighted distribution thus takes the form

$$f(r) \equiv r^3 n(r) \equiv ar^2 \exp\left[-1/2\left(\frac{\ln r - \ln r_g}{\ln \sigma}\right)^2\right],$$

where $a = \exp[-2(\ln \sigma)^2]/r_g^2$ is the normalization factor and $r_m = r_g \exp[2(\ln \sigma)^2]$ is the mode radius of the volume-weighted distribution.

Six different size distributions were used, comprising three different values of the effective radius (0.2, 0.4, and 0.8 μm) and two values of the spread parameter ($\sigma = 1.6$ and 2.0). The effective radius $r_{\text{eff}} = M_3/M_2 = r_g \exp[2.5(\ln \sigma)^2]$, where M_2 and M_3 are the second and third moments of the distribution, respectively. A summary of the parameters is given in Table 1. These parameters were used to calculate the optical thicknesses at a number of discrete wavelengths, similar to those commonly used in radiometers. (The normalization of the size distributions was

chosen arbitrarily and results in optical thickness values that are equally arbitrary; this is a linear problem.) The wavelengths used in this study are 0.352, 0.414, 0.501, 0.670, 0.862, 1.020, 1.250, and 1.55 μm . Two sets of inversions were performed, one with the full set of wavelengths and the other with the subset of wavelengths from 0.414 to 0.862 μm . The latter set, called the restricted wavelength range, corresponds to the multifilter rotating shadow-band radiometer²⁰ measurement wavelengths, whereas the full set are the Stratospheric Aerosol and Gas Experiment III (SAGE III)²¹ wavelengths and cover the wavelength range of other instruments, such as the NASA Ames Airborne Tracking Sunphotometer-14.²²

B. Data Function as a Predictor of Inversion Accuracy

In earlier research on analytic eigenfunction inversions,¹⁷ it was found with the standard inversion technique that, if the peak of the $g(k)$ versus the k curve did not lie within the measurement range, then the quality of the retrievals was adversely affected because a significant portion of the information was missing. Consequently the upper and lower tails used in the integrations will not be good approximations to the behavior of the curve in this region. This is obviously also a consideration with the subtracted inversion technique because of the necessity to make assumptions about the behavior of $g_s(k)$ outside the measurement range.

It is possible to obtain some indication of whether a satisfactory retrieval will be possible by looking at the data function $g(k)$. If there is no peak within the measurement range, then retrievals may be less than satisfactory—although if an extremity of the data range is on the peak, then reasonable results can be obtained.

Figures 1(a) and 1(b) are plots of $g(k)$ versus k for the standard and subtracted cases, respectively. Also plotted are vertical lines representing the measurement ranges; the inner pair corresponds to the restricted set and the outer pair to the full set of wavelengths. Examination of these plots shows that for the subtracted inversion technique the peak of the $g(k)$ versus the k curve is shifted toward larger k values, which suggests that this method may be more appropriate for larger particles whereas the standard inversion technique may be more appropriate for smaller particles. It is also evident that the wavelength range has a bearing on the viability of an inversion technique.

On the basis of these plots, we would expect that the standard inversion technique would produce accurate retrievals for $r_{\text{eff}} = 0.2$ with both the full and the restricted ranges, although the slow drop-off at large k could cause some problems. For the subtracted inversion technique, however, we would expect the retrievals to be poor.

For $r_{\text{eff}} = 0.4$, we would expect that both the standard and the subtracted inversion techniques should work well for the full wavelength range, although the subtracted inversion technique could be considered

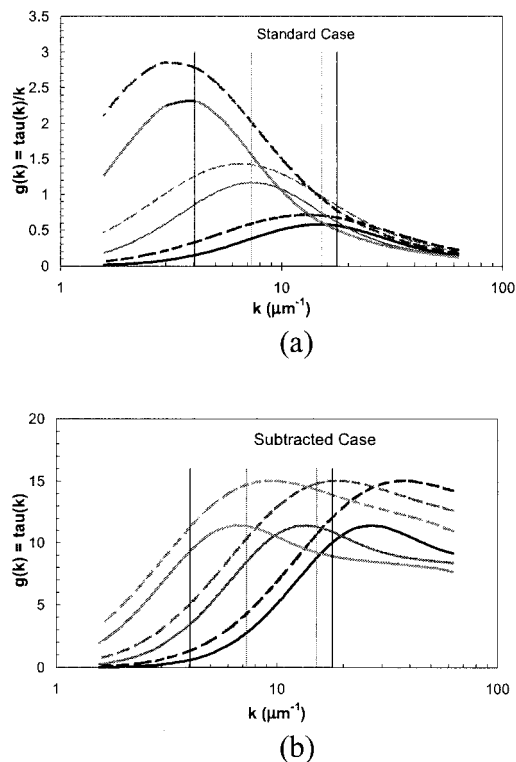


Fig. 1. Plot of data function $g(k)$ versus k , where $k = 2\pi/\lambda$. (a) Standard case [$g(k) = \tau(\lambda)/k$] and (b) subtracted case [$g(k) = \tau(\lambda)$]. Shown are $r_{\text{eff}} = 0.2$ (black curves), $r_{\text{eff}} = 0.4$ (dark grey curves), $r_{\text{eff}} = 0.8$ (light grey curves) (leftmost peaks). Solid curves are $\sigma = 1.6$ and dashed curves are $\sigma = 2.0$. The vertical lines indicate the measurement ranges.

marginal for $\sigma = 2.0$. For the restricted wavelength range, however, both the standard and the subtracted inversion techniques are marginal as the peak of the $g(k)$ curve is close to the upper k edge of the range for the subtracted inversion and the lower edge of the range for the standard inversion.

For $r_{\text{eff}} = 0.8$, the standard inversion technique is, at best, only marginal for the full range, with the peak at the low k edge. For the subtracted inversion technique, however, reasonable inversions can be expected for both wavelength ranges. The restricted range retrievals could be marginal, especially for $\sigma = 1.6$, where the peak is at the low k end of the range.

C. Comparison of True and Retrieved Distributions

Use of simulated data allows comparison of retrieved size distributions with the true distributions and also provides a way to check the validity of predictions based on the $g(k)$ versus the k curves. In Figs. 2(a)–2(c), the true distribution (thin curves), and the size distributions retrieved by both the standard inversion (solid curves) and the subtracted inversion (dashed curves) techniques are plotted for the $\sigma = 2.0$ cases when there were no errors in the data and for integration cutoff $\omega_m = 6$. (The effect of errors in the data is discussed in Subsections 3.D and 3.E.) The plots for $\sigma = 1.6$ and for $\omega_m = 7$ –10 are qualitatively similar. These plots suggest that the subtracted in-

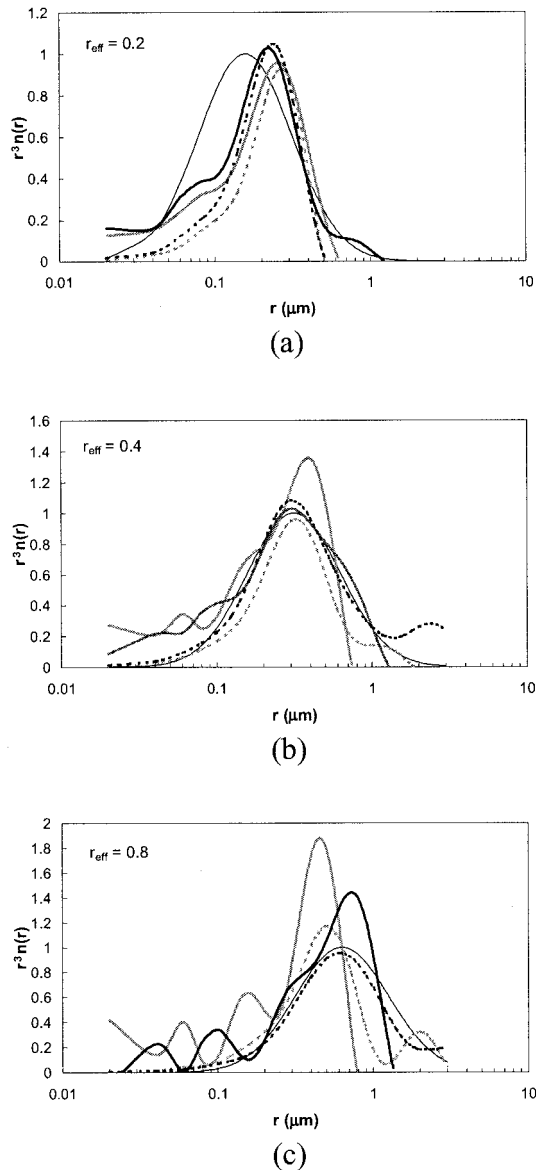


Fig. 2. Plots of normalized volume-weighted size distributions for $\sigma = 2.0$. (a) $r_{\text{eff}} = 0.2$, (b) $r_{\text{eff}} = 0.4$, (c) $r_{\text{eff}} = 0.8$. Each plot shows the true distribution (thin curves) and retrieved distributions from both standard (solid curves) and subtracted (dashed curves) techniques by use of both the full (dark grey curves) and the restricted (light grey curves) measurement wavelength ranges.

version technique does not always yield a significantly better retrieval than the standard inversion technique. As was pointed out above, this is to be expected.

The expansion nature of the retrieval process used here means that each distribution is the sum of a number of oscillatory terms, which may cancel or enhance one another. Mathematically, the retrieved size distribution can be defined for all values of radius, and this may include negative values at some radii. To obtain meaningful results it is necessary to set upper and lower radius cutoff criteria for the retrieved size distributions. Obvious bounds are the radii on either side of the volume-weighted size dis-

tribution peak, at which the value first goes negative; this is what was chosen here. However, there were cases in which the size distribution did not go negative, so in those cases the cutoffs were taken to be the lowest and highest values of r for which the size distributions were evaluated, 0.02 and 3.0 μm , respectively.

Several points need to be made about the subtracted inversions, the most notable of which is that they do more accurately reflect the behavior of the size distribution for small radii. This is a significant improvement over the standard inversion technique, which tends to produce extra modes at the small radius end of the distribution. It can also be seen from the plots that the standard inversion technique is more likely to give a high value at the mode radius r_m and to drop off sharply at the large radius end of the distribution. The subtracted inversions were generally closer to the true distribution.

Another feature of the retrieved size distributions is the tendency of the inversions to suggest an extra mode at larger radii. This is particularly evident with the subtracted inversions, which are less likely to go negative. The extra modes are more evident with the restricted wavelength range than the full wavelength range. Several test inversions with wavelength ranges of 0.352–2.50 and 0.20–2.50 μm confirmed that the magnitude of the extra modes is reduced in both the standard and the subtracted inversion techniques as the wavelength range is expanded. The oscillatory behavior can be more pronounced as more terms are added to the expansion, but for the inversions presented here it was found that there was little difference in the basic features of the retrievals for values of ω ranging between 6 and 10.

The effect of wavelength range on the quality of the standard inversion was considered in an earlier study.²³ For all the inversions, the results with the full wavelength set are better than with the restricted set. In the case of $r_{\text{eff}} = 0.2$, for both the standard and the subtracted inversions, r_m is too high and the distribution too narrow for both wavelength ranges. This is what would be expected for the subtracted inversion technique but not for the standard inversion technique. The slow drop-off of $g(k)$ at large k noted above is a likely reason: The upper tail might be a poor approximation to the true behavior.

For $r_{\text{eff}} = 0.4$, the subtracted inversion is clearly better than the standard inversion at the short radius end since the extra modes present in the standard inversion are no longer evident, although there is an extra mode at the large radius end. The restricted range inversions, which were expected to be marginal, retrieve r_m reasonably accurately and give some idea of the spread of the distribution. The additional measurements in the full wavelength set, mainly at the long-wavelength end, result in much improved inversions with both the standard and the subtracted inversion techniques, as was expected, although the subtracted inversion also suggests an extra large r mode.

In the case of $r_{\text{eff}} = 0.8$, as expected, the subtracted

Table 2. Percentage Error in Calculated Optical Thickness

Size Distribution	Standard Inversion		Subtracted Inversion	
	No Error	Random Error	No Error	Random Error
Range: 0.414–0.862 μm				
0.2, 1.6	2.11	2.11 \pm 0.18	3.65	3.66 \pm 0.18
0.2, 2.0	1.46	1.56 \pm 0.29	2.28	3.82 \pm 4.77
0.4, 1.6	2.63	2.69 \pm 0.07	0.98	1.03 \pm 0.07
0.4, 2.0	2.51	2.57 \pm 0.12	0.99	1.11 \pm 0.15
0.8, 1.6	5.95	5.95 \pm 0.10	1.56	1.58 \pm 0.12
0.8, 2.0	4.76	4.80 \pm 0.10	1.05	1.09 \pm 0.11
Range: 0.352–1.55 μm				
0.2, 1.6	2.80	2.42 \pm 1.02	8.06	8.44 \pm 2.75
0.2, 2.0	0.57	0.73 \pm 0.10	3.05	2.80 \pm 0.54
0.4, 1.6	0.51	0.62 \pm 0.09	1.53	1.55 \pm 0.17
0.4, 2.0	0.72	0.78 \pm 0.05	1.28	1.32 \pm 0.16
0.8, 1.6	2.40	2.44 \pm 0.08	0.27	0.48 \pm 0.07
0.8, 2.0	2.21	2.23 \pm 0.06	0.25	0.50 \pm 0.20

inversions are clearly better than the standard inversions, which suggests several extra modes at the short radius end and high values at the mode radius. These features are not evident in the subtracted inversions, although they do have an extra mode at the large radius end. The effect of wavelength range is more pronounced for these cases. For both the standard and the subtracted inversion techniques, the restricted set indicated a mode radius that was too low and a narrow distribution, whereas the full wavelength set retrieved the mode radius and spread reasonably well for both techniques. These results show that the $g(k)$ versus the k curve does give a good guide to the likely quality of the inversion process.

D. Calculating Back to the Input Data

With simulated data it is possible to compare the retrieved size distribution with the true size distribution and also to examine how well the various moments of the distribution can be calculated. In practice, however, the only measure we have of how good an inversion is, is how well it calculates back to the input data.

One measure of this is the mean of the square root of the sum of squares of percentage errors in calculated τ , defined as

$$\text{error} = \frac{1}{n} \left\{ \sum_{i=1}^n \left[\frac{\tau(\lambda_i) - \tau_{\text{calc}}(\lambda_i)}{\tau(\lambda_i)} \right]^2 \right\}^{1/2} \quad (8)$$

Here τ is the measured value, τ_{calc} is the predicted value by use of the retrieved size distribution, and n is the number of measurements.

Table 2 lists these values for both the standard and the subtracted retrievals for all the distributions considered, including the means and standard deviations of errors for sets of ten retrievals with 2.5%

random error. The random error sets provide a means of gauging the reliability of retrievals in the presence of measurement error.

Examination of the values in Table 2 shows that for the $r_{\text{eff}} = 0.2$ distributions, percentage error is better with the standard inversion technique than with the subtracted inversion technique. This is what would be expected based on the discussion above. For all the other distributions, the percentage error was lower (or similar) for the subtracted inversion technique. The percentage errors for each distribution are lower for the full wavelength range, as would be expected given the extra information available. The values for the random error cases generally agree well with those when there is no error in the measurements, indicating that the inversion process is stable in the presence of random error.

E. Predicting Moments of the Distributions

The moments of a distribution yield useful and important information. Here we consider the zeroth moment that gives the total number of particles N ; the second and third moments that relate to area and volume, respectively; and the effective radius $r_{\text{eff}} = M_3/M_2$. The moments for the standard and subtracted inversions were calculated numerically from the retrieved size distribution by Eq. (6). The true values were calculated analytically from²⁴

$$M_i = N r_g^i \exp[0.5(\ln \sigma)^2 i^2].$$

The ratios of the calculated moments to the analytic moments for both wavelength ranges are given in Table 3. (Moments calculated from the random error cases are not given, but agreed well with the values in Table 3.) As a general rule, the moments derived from the full wavelength set inversions were more accurate, and the subtracted inversion technique was better than the standard inversion technique for the larger sizes.

Neither the standard nor the subtracted inversion technique was able to recover the zeroth moment particularly well. For the standard inversion the values were too high, particularly with the restricted range, and there was no pattern for the subtracted inversion. The second and third moments, along with the effective radius, were in many cases retrieved reasonably well. For the standard inversion, the second moment was within 15% of the true value for $r_{\text{eff}} = 0.2$ for both wavelength ranges, but overestimated the value for the other size distributions. The third moment and hence r_{eff} were retrieved well for $r_{\text{eff}} = 0.2$, as would be expected. For the other size distributions, the third moment and hence r_{eff} were underestimated, with the level of error increasing with particle size. The patterns were the same for both the restricted and the full wavelength range, although the accuracy is higher with the full range.

For the subtracted inversions the results are mixed; the second and third moments are mostly underestimated for the restricted wavelength range, re-

Table 3. Ratio of Calculated to True Moments of Distributions

Parameter	Value					
r_{eff}	0.2	0.2	0.4	0.4	0.8	0.8
σ	1.6	2.0	1.6	2.0	1.6	2.0
0.414–0.862 μm : standard inversion						
M_0	5.29	2.31	43.89	17.99	4.14	100.88
M_2	1.00	0.84	1.49	1.29	1.04	1.40
M_3	0.93	0.83	0.95	0.81	0.53	0.50
r_{eff}	0.93	0.99	0.64	0.62	0.51	0.36
0.414–0.862 μm : subtracted inversion						
M_0	1.11	0.36	3.47	1.40	6.65	4.18
M_2	0.78	0.56	1.26	0.76	0.75	0.92
M_3	0.89	0.66	2.89	0.70	0.48	0.73
r_{eff}	1.14	1.14	2.28	0.93	0.64	0.79
A_{iter}	0.73	0.68	1.08	1.05	1.03	1.02
0.352–1.55 μm : standard inversion						
M_0	8.11	2.90	17.86	7.81	1.27	8.86
M_2	1.15	0.94	1.28	1.19	1.04	1.16
M_3	0.98	0.95	1.04	0.96	0.86	0.76
r_{eff}	0.85	1.01	0.81	0.81	0.83	0.65
0.352–1.55 μm : subtracted inversion						
M_0	1.72	0.60	3.88	2.01	0.82	2.58
M_2	0.87	0.68	1.36	1.22	0.97	0.97
M_3	0.89	0.72	3.39	2.99	1.40	1.12
r_{eff}	1.03	1.06	2.50	2.45	1.45	1.16
A_{iter}	0.93	0.77	1.04	0.96	1.03	1.04

sulting in values of r_{eff} mostly within 20% of the true value. The $r_{\text{eff}} = 0.4, \sigma = 1.6$ distribution gives high values for the second and third moments. For the extended range, the second moment is too low for $r_{\text{eff}} = 0.2$ and quite accurate for $r_{\text{eff}} = 0.8$, but too high for the other cases, with the errors being most pronounced for $r_{\text{eff}} = 0.4$. The third moment is also low for $r_{\text{eff}} = 0.2$, but is too high for the other cases, with the errors being most pronounced for $r_{\text{eff}} = 0.4$. The main explanation for the overestimation of the second and third moments with the subtracted inversions is the presence of extra modes at the large radius end.

The iterative procedure used by the subtracted inversion technique to determine the value A_{iter} , which gives the best fit, provides an alternative way of finding the second moment. To evaluate the accuracy of these determinations, the cross-sectional area was determined analytically for each distribution. The ratios of the retrieved values to the analytical values are given in Table 3.

In the majority of cases the values of A_{integ} (i.e., M_2) and A_{iter} agree with the analytic value to within 30%, and in many cases the discrepancy is considerably less. As a general rule, except for the $r_{\text{eff}} = 0.2$ case, A_{iter} is the most accurate value of the cross-sectional area. One interesting feature to note from Table 3 is that in most cases there was not much difference

between the values of A_{iter} for the restricted and full wavelength sets. This suggests that, even though the moments derived from the subtracted inversions may have large errors due to the presence of extra modes, the iterative procedure does produce an accurate estimate for the cross-sectional area of the distribution. This can be used to gauge the accuracy of the moments by a comparison of A_{iter} with πM_2 ; if A_{iter} is less than πM_2 , then the second and third moments are likely to be too high and vice versa.

4. Summary and Conclusions

In this paper we have presented an inversion method that incorporates *a priori* information into the inversion process. The method is based on analytic eigenfunction theory and makes use of the surface area of the size distribution. Six simulated data sets were generated, and two wavelength ranges were used in the inversion process. Two inversion processes are compared: the standard inversion, which makes no assumptions about the underlying size distribution, and the subtracted inversion, which assumes that *a priori* information, in the form of surface area, is available.

The inversion method used here is in fact a two-step process incorporating both the standard and the subtracted inversion techniques. This was done because independent measurements of the surface area

of an aerosol size distribution are not usually available, so the standard inversion is used to determine a first estimate as input to the subtracted inversion.

The quality of an inversion depends on the information content of the measurements. Here we present a simple method to determine whether a particular measurement set is likely to provide reasonable results and also whether the standard inversion technique or the subtracted inversion technique is better. The method, which examines the behavior of the data function (derived from the measurements) as a function of $k(= 2\pi/\lambda)$, indicates that the viability of an inversion depends on both the particle sizes and the measurement wavelength range. It was found that the standard inversion technique is most appropriate for size distributions where the particles are small, and the subtracted inversion technique is more suited to larger particle sizes. Also, as a general rule, the greater the wavelength range, the better the inversion was likely to be. The inversion results confirmed these predictions.

One weakness of the analytic eigenfunction inversion process is the tendency to produce extra modes at either or both the large and the small radius ends of the distribution. This arises from the expansion nature of the inversion process and the oscillatory nature of the terms being added. The subtracted inversion tends to minimize this problem at short radii, and the addition of measurements at longer wavelengths helps at the large radius end. These artifacts of the inversion process mean that moments calculated from the retrieved size distributions are not always accurate. However, the iterative procedure used to determine the value of A during the subtracted inversion does produce a quite accurate value for A . Comparison of this with that derived from the second moment of the final distribution gives an alternative way of gauging the accuracy of the retrieved distribution.

In conclusion, it has been shown that the inclusion of *a priori* information in the analytic eigenfunction inversion technique can lead to improved retrievals of aerosol size distribution. The two-step process used here can be used when no measured values of a surface are available, and the iterative procedure used by the subtracted inversion technique provides a good estimate for the surface area. It was also shown that the reliability of a particular inversion can be gauged quite simply from an examination of the data function.

References

- H. C. van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1981).
- C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).
- S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by inversion of the linear system produced by quadrature," *J. Assoc. Comput. Mach.* **10**, 97–101 (1963).
- M. D. King, D. M. Byrne, B. M. Herman, and J. A. Reagan, "Aerosol size distributions obtained by inversion of spectral optical depth measurements," *J. Atmos. Sci.* **35**, 2153–2167 (1978).
- G. Yamamoto and M. Tanaka, "Determination of aerosol size distribution by spectral attenuation measurements," *Appl. Opt.* **8**, 447–453 (1969).
- J. Heintzenberg, H. Muller, H. Quenzel, and E. Thomalla, "Information content of optical data with respect to aerosol properties: numerical studies with a randomized minimization-search-technique inversion algorithm," *Appl. Opt.* **20**, 1308–1315 (1981).
- E. Trakhovskiy, S. G. Lipson, and A. D. Devin, "Atmospheric aerosols investigated by inversion of experimental transmittance data," *Appl. Opt.* **21**, 3005–3010 (1982).
- A. Ya. Perelman and K. S. Shifrin, "Improvements to the spectral transparency method for determining particle size distribution," *Appl. Opt.* **19**, 1787–1793 (1980).
- M. A. Box and B. H. J. McKellar, "Analytic inversion of multispectral extinction data in the anomalous diffraction approximation," *Opt. Lett.* **3**, 91–93 (1978).
- C. D. Capps, R. L. Henning, and G. M. Hess, "Analytic inversion of remote sensing data," *Appl. Opt.* **21**, 3581–3587 (1982).
- G. Viera and M. A. Box, "Information content analysis of aerosol remote sensing experiments using singular function theory. 1. Extinction measurements," *Appl. Opt.* **26**, 1312–1327 (1987).
- A. Ben-David, B. M. Herman, and J. A. Reagan, "Inverse problem and the pseudoempirical orthogonal function method of solution 1. Theory," *Appl. Opt.* **27**, 1235–1242 (1988).
- A. Ben-David, B. M. Herman, and J. A. Reagan, "Inverse problem and the pseudoempirical orthogonal function method of solution. 2. Use," *Appl. Opt.* **27**, 1243–1254 (1988).
- J. G. McWhirter and E. R. Pike, "On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind," *J. Phys. A*, **11**, 1729–1745 (1978).
- G. Viera and M. A. Box, "Information content analysis of aerosol remote-sensing experiments using an analytic eigenfunction theory: anomalous diffraction approximation," *Appl. Opt.* **24**, 4525–4533 (1985).
- M. A. Box and G. Viera, "Information content analysis of aureole inversion methods: differential kernel versus normal," *J. Opt. Soc. Am. A* **7**, 1015–1018 (1990).
- G. P. Box, K. M. Sealey, and M. A. Box, "Inversion of Mie extinction measurements using analytic eigenfunction theory," *J. Atmos. Sci.* **49**, 2074–2081 (1992).
- H. M. Nussenzweig and W. J. Wiscombe, "Efficiency factors in Mie scattering," *Phys. Rev. Lett.* **45**, 1490–1494 (1980).
- P. Attard, M. A. Box, G. Bryant, and B. H. J. McKellar, "Asymptotic behavior of the Mie-scattering amplitude," *J. Opt. Soc. Am. A* **3**, 256–258 (1986).
- L. Harrison, J. Michalsky, and J. Berndt, "Automated multi-filter rotating shadow-band radiometer: an instrument for optical depth and radiation measurements," *Appl. Opt.* **33**, 5118–5125 (1994).
- P.-H. Wang, G. S. Kent, M. P. McCormick, L. W. Thomason, and G. K. Yue, "Retrieval analysis of aerosol-size distribution with simulated extinction measurements at SAGE III wavelengths," *Appl. Opt.* **35**, 433–440 (1996).
- P. B. Russell, J. M. Livingston, O. Dubovik, S. A. Ramirez, J. Wang, J. Redemann, B. Schmid, M. A. Box, and B. N. Holben, "Sunlight transmission through desert dust and marine aerosols: diffuse light corrections to Sun photometry and pyrhelimetry," *J. Geophys. Res.* **109D**, doi:10.1029/2003JD004292 (2004).
- G. P. Box, "Effects of smoothing and measurement-wavelength range on the accuracy of analytic eigenfunction inversions," *Appl. Opt.* **34**, 7787–7792 (1995).
- A. Deepak and G. P. Box, "Representation of aerosol size distribution data by analytic models," in *Atmospheric Aerosols: The Formation, Optical Properties and Effects*, A. Deepak, ed. (Spectrum, Hampton, Va., 1982), pp. 79–109.