Modified version of the Chahine algorithm to invert spectral extinction data for particle sizing

Fabio Ferri, Alessandra Bassini, and Enrico Paganini

A modified version of the nonlinear iterative Chahine algorithm is presented and applied to the inversion of spectral extinction data for particle sizing. Simulated data were generated in a λ range of 0.2–2 μm, and particle-size distributions were recovered with radii in the range of 0.14–1.4 μm. Our results show that distributions and sample concentrations can be recovered to a high degree of accuracy when the indices of refraction of the sample and of the solvent are known. The inversion method needs no *a priori* assumptions and no constraints on the particle distributions. Compared with the algorithm originally proposed by Chahine, our method is much more stable with respect to random noise, permits a better quality of the retrieved distributions, and improves the overall reliability of the fitting. The accuracy and resolution of the method as functions of noise were investigated and showed that the retrieved distributions are quite reliable up to noise levels of several rms percent in the data. The sensitivity to errors in the real and imaginary parts of the refraction index of the particles was also examined.

Key words: Particle sizing, inversion algorithm, spectral extinction.

1. Introduction

Optical techniques are well-established methods for characterizing particle-size distributions in a variety of application fields.[1–4] Indeed they have several advantages over other methods, such as electron microscopy and sieving, because high-quality measurements can be performed *in situ* and in real time. Optical techniques can provide either direct information on particle distributions when measurements are carried out on single particles one at a time or indirect information when the techniques are used to study samples containing many particles. The last case is the most convenient because measurements can be performed simultaneously on a large number of particles and because of simpler experimental setups. Indirect optical characterization is primarily achieved by measuring (1) the angular distribution of the scattered light of a fixed wavelength, (2) the light scattered at several wavelengths for a fixed angle, (3) the intensity autocorrelation function of the light scattered at a fixed angle for a fixed wavelength, or (4) the spectral extinction of light through the sample.

In principle, measurements of scattered light are superior because higher sensitivities can be achieved and independent information can be provided by the two orthogonal polarizations of the incident light. However, in certain experimental configurations, typically encountered when dealing with natural aerosols, the large illuminated volumes and the large collection solid angles do not permit the separation of the two polarizations.[5] On the other hand, the multispectral extinction technique is very simple in terms of measurement principle and very convenient with regard to the optical arrangement and can be easily performed by the adaptation of commercial spectrophotometers. This technique is suitable for remote sensing too and has indeed been applied to the characterization of the atmospheric aerosol size distribution for a long time.[6,7]

The main difficulty in retrieving the particle-size distribution from measurements of light attenuation at several wavelengths lies in the solution of a Fredholm integral equation of the first kind. Various inversion methods have been developed to solve this classic ill-posed problem, whose characteristics are common to many other indirect sensing experiments. In practice, ill-posed problems invariably

F. Ferri is with the Institute of Mathematical, Physical and Chemical Sciences, University of Milan at Como, via Lucini 3, 22100 Como, Italy. A. Bassini and E. Paganini are with the Centro Informazioni Studi Esperienze, P.O. Box 12081, 20134 Milan, Italy.

Received 20 September 1994; revised manuscript received 16 December 1995.

0003-6935/95/255829-11$06.00/0.

lead to highly unstable solutions because even arbitrarily small noise components in the measured quantities can give rise to extremely large spurious oscillations in the solution.[8,9]

The simplest approach to tackling the inversion problem consists of making an *a priori* assumption about the shape of the size distribution; the parameters defining the distribution are then determined by a best-fit procedure on the experimental data. Particle-size information based on the assumed shapes of the distributions was recovered for air suspensions,[10] liquid suspensions,[11] and atmospheric aerosols.[12,13] However, these methods are very restrictive because, in many circumstances, the forms of the distributions are unknown, and consequently more general methods are preferred.

Two of the most popular inversion techniques are the methods of Philips,[14] Twomey,[15] and Backus and Gilbert,[16] which are linear constrained algorithms and are described in detail in Ref. 17; a quantitative comparison of them is also reported in Refs. 18 and 19. Linear inversion methods and their successive modifications were applied to a variety of experimental situations including atmospheric aerosols[20,21] and condensation of water droplets in supersonic steam flows.[22] However, in spite of their popularity and wide applicability, the linear constrained methods have some evident disadvantages in that their success depends on the accurate tuning of certain fitting parameters, which vary from case to case and are user dependent. Moreover some constraints, such as smoothness and positiveness, are to be imposed on the solutions.

In more recent years, McWhirter and Pike[23] introduced a new approach, known as the analytic eigenfunction theory, which is based on the Mellin transform of the kernel function. This technique, which is particularly suited to problems in which the kernel is a relatively simple analytical function, was applied by Viera and Box[24] and by Bertero and co-workers[25,26] to the inversion of spectral extinction data in the anomalous diffraction approximation. Recently Box *et al.*[27] applied the eigenfunction theory, using the exact Mie extinction kernel, and obtained satisfactory computer simulations, although with a tendency to overestimate the content in small particles.

A different approach was suggested by Chahine[28,29] who proposed a nonlinear iterative method to determine temperature profiles of the atmosphere from measurements of its emerging radiance as a function of frequency. The Chahine method was first applied by Grassl[30] to invert spectral transmission data and later on was used for the inversion of light-scattering data.[31–33] The main advantages of the Chahine method rely on the fact that no tuning of external parameters is needed, no constraints are imposed on the solutions that are always positive, and large amounts of data can be reduced efficiently. On the other hand, the technique is sensitive to the presence of noise, so that highly unstable and noisy solutions may occur. To increase the stability, a modification of this inversion procedure, the Twomey–Chahine algorithm, was proposed[8,17] and applied successfully to the inversion of spectral extinction data.[34] However, the method must be handled with care because, as pointed out in Ref. 35, it may lack convergence if the estimated initial distribution is too different from the actual distribution.

In this paper we propose a new inversion algorithm that is based on the classical Chahine scheme. We have applied the method to the inversion of spectral extinction data and tested it by computer simulations. An exhaustive analysis of its accuracy, resolution, and reliability as a function of noise has been performed. Our results show that the method permits a large variety of particle distributions to be accurately reconstructed, greatly improves the stability of the solutions against noise, and is able to suppress the typical indented and spiky-shaped appearance of the distributions obtained with the classical Chahine method. Moreover, when the kernel function exhibits secondary maxima and has high-frequency oscillations superimposed, as in the current case of spectral extinction data, our technique seems to be particularly suitable. We also present a quantitative analysis of the sensitivity of the method to errors in the refraction index of the particles and discuss the possibility of recovering it by means of spectral extinction measurements.

2. Theory

When a beam of monochromatic radiation impinges on a sample containing particles with an index of refraction that is different from that of the medium, scattering and absorption lead to an attenuation of the transmitted beam. If P_0 and P_T denote the powers of the incident and transmitted beams, respectively, when multiple scattering can be neglected, it results that[36]

$$P_T = P_0 \exp[-\alpha(\lambda)L], \tag{1}$$

where $\alpha(\lambda)$ is the extinction coefficient and L is the length of the sample. If the sample is a diluted homogeneous suspension of noninteracting polydisperse spheres, the extinction coefficient $\alpha(\lambda)$ is given by

$$\alpha(\lambda) = \int \pi r^2 Q_{\text{ext}}(r, \lambda, m) N(r) \mathrm{d}r, \tag{2}$$

where $N(r)\mathrm{d}r$ is the concentration of particles (cm^{-3}) with radii between r and $r + \mathrm{d}r$ and Q_{ext} is a known function of the Mie theory.[37] Here λ is the wavelength of the radiation in the medium, and m is the refraction index of the particles relative to the medium. Equation (2) is a first-kind Fredholm integral equation where $\alpha(\lambda)$ is provided by the experiment, $\pi r^2 Q_{\text{ext}}$ is the kernel, and $N(r)$ is the unknown function. Q_{ext} is an adimensional quantity called an extinction efficiency factor that actually depends on r and λ only through the ratio $x = 2\pi r/\lambda$. The function Q_{ext} is shown in Fig. 1 as a function of x for different values

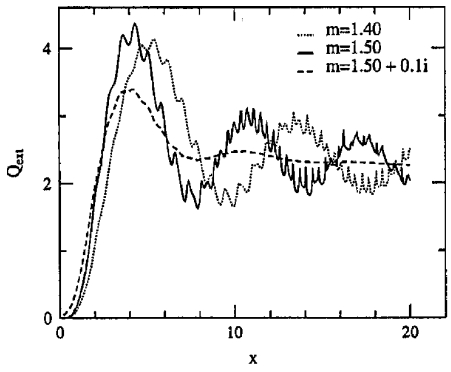

Fig. 1. Behavior of the extinction efficiency $Q_{ext}(x, m)$ as a function of the adimensional parameter $x = 2\pi r/\lambda$ for different values of the relative refraction index of the particle m.

of m. For small x, when m is real, Q_{ext} increases as x^4, which corresponds to Rayleigh scattering. Then it reaches a maximum for r that is comparable with λ and, for large x, tends, with oscillatory behavior, toward a value of 2, which corresponds to the diffraction limit. Figure 1 shows that the real part of the refraction index of the particle affects the positions of the principal and secondary peaks of Q_{ext}, whereas the presence of an imaginary part smears out its fine structure and depresses the amplitude of its oscillations. We can deduce from Fig. 1 that particles of different radii have maxima located at different λ values, which is fundamental in the use of spectral extinction data for particle sizing.

To invert Eq. (2), we should consider that only a finite number of λ values can be accessed experimentally and within a limited range $(\lambda_{min}, \lambda_{max})$. As a consequence, $N(r)$ can be recovered only over a limited range of radii (r_{min}, r_{max}). Let us call $\lambda_1, \lambda_2, \ldots, \lambda_q$ the q measured wavelengths and divide the radius range in q intervals or classes. Assuming that inside each class $N(r)$ is a constant equal to N_j, we can transform Eq. (2) into

$$\alpha(\lambda_i) = \sum_j N_j \mathbf{A}_{ij}, \qquad j = 1, 2, \ldots, q, \tag{3}$$

where $\mathbf{A}_{ij}$ is a $q \times q$ matrix given by

$$\mathbf{A}_{ij} = \int_{r_{j-1}}^{r_j} \pi r^2 Q_{ext}(r, \lambda_i, m)\mathrm{d}r, \tag{4}$$

where (r_{j-1}, r_j) is the interval that defines the jth class of width $\delta_j = r_j - r_{j-1}$. If the relative refraction index m is known for all wavelengths, the elements of the matrix $\mathbf{A}_{ij}$ are known and depend on the jth class and the ith wavelength only. Equation (3) is then a complete set of q linear algebraic equations.

Let us now consider how to subdivide the interval (r_{min}, r_{max}) into classes. Although this is somewhat arbitrary, it is fairly convenient to define the classes so that their relative width is constant over the entire range (r_{min}, r_{max}). To do that, we should scale r_k as a geometrical progression given by

$$r_k = r_{min} a^k, \qquad k = 0, 1, \ldots, q, \tag{5}$$

where a is the ratio of the progression given by $a = (r_{max}/r_{min})^{1/q}$. If we denote with R_j the average radius of the jth class, it results from Eqs. (5) that the relative width δ_j/R_j is given by $\delta_j/R_j \sim a - 1$. We have now to find how the two ranges, (r_{min}, r_{max}) and $(\lambda_{min}, \lambda_{max})$, are related to each other. Following the idea originally proposed by Chahine, we associate for each λ_j a class of particles that gives a signal sequence $\alpha(\lambda)$ with the maximum located as close as possible to λ_j. This is equivalent to, if the classes are narrow enough, their average radius R_j being given by

$$R_j = \frac{x_{peak}}{2\pi}\lambda_j, \qquad j = 1, 2, \ldots, q, \tag{6}$$

where x_{peak} is the value of x for which the function Q_{ext} has its maximum (see Fig. 1). Equation (6) provides the recipe for selecting the λ_j values when a distribution with assigned R_j values is wanted, and vice versa it gives the values of R_j when a set of λ_j is available for the experiment. From Eqs. (5) and (6) it follows that the wavelengths λ_j scale as a geometrical progression:

$$\lambda_j = \lambda_{min} a^{j-1}. \tag{7}$$

We now briefly recall how the classical Chahine algorithm works. As we know, this inversion method is an iterative nonlinear scheme that tells you how to obtain the next iteration distribution once the previous distribution is known. Let us suppose that a distribution ${N_j}^p$ has been recovered after p iterations. This distribution gives rise to a sequence of signals ${\alpha_{calc}}^p(\lambda_j)$ [calculated according to Eq. (3)], which will be different from the sequence of the measured signals $\alpha_{meas}(\lambda_j)$. To find a better distribution at the $p + 1$ iteration, the Chahine scheme corrects the population of the jth class ${N_j}^p$ in the following way:

$${N_j}^{p+1} = {N_j}^p \frac{\alpha_{meas}(\lambda_j)}{{\alpha_{calc}}^p(\lambda_j)}. \tag{8}$$

Note that one performs the correction by comparing the measured and calculated signals only for the λ_j associated with the class being corrected. Indeed the largest relative contribution to the extinction at λ_j is given by particles with radius R_j. By use of Eq. (8) a good match between the calculated and measured signals can be achieved fairly quickly. However, although this match remains stable as the number of iterations rises, an instability mechanism can arise and make the distributions noisier and noisier. This effect is particularly disturbing in the presence of experimental noise and in the case of kernels with secondary maxima and high-frequency oscillations superimposed, as those used in the inversion of spectral extinction data and shown in Fig. 1. The final result is a spikylike distribution where only a limited number of classes are used to reconstruct

the desired distribution. Examples of this effect can be found in Refs. 19 and 32 and will be shown below.

Our improvement of the classical Chahine method consists of using the whole signal sequence $\alpha_{\text{meas}}(\lambda_i)$ to correct the population of each class N_j. This is carried out in the following way:

$$N_j{}^{p+1} = N_j{}^p \sum_{i=1}^{q} W_{ij} \frac{\alpha_{\text{meas}}(\lambda_i)}{\alpha_{\text{calc}}{}^p(\lambda_i)}, \tag{9}$$

where W_{ij} is a normalized weight function given by

$$W_{ij} = \frac{A_{ij}}{\sum_i A_{ij}}. \tag{10}$$

Note that, for each term of the sum of Eq. (9), the correction factors $[\alpha_{\text{meas}}(\lambda_i)/\alpha_{\text{calc}}{}^p(\lambda_i)]$ are of the same type as those used in the classical Chahine scheme. Equation (9) is therefore a sort of smooth version of Eq. (8), which is fundamental in making the algorithm much more stable and reliable in the presence of noise. Note also that, because of Eq. (4), in the limit of infinitesimal classes, W_{ij} has the same shape as Q_{ext}. As a result the highest weight to the correction factors is for $i = j$, i.e., for that wavelength associated with the class being corrected. Note that the choice of W_{ij} defined by Eq. (10) is somehow arbitrary. Actually we used different weight functions (for example, Gaussian functions with different widths) and obtained distributions that are always much smoother and more reliable with respect to noise than those retrieved with the classical Chahine algorithm. However, these distributions are still characterized by a slight tendency to present some spurious oscillations that depend on the particular weight function that was adopted. We think that this is due to the presence of secondary maxima in the kernel function Q_{ext}, and, in this respect, a weight function equal to the kernel is the one that optimizes the relative contributions of the different channels. Finally, note that the idea of using Q_{ext} as a weight function was suggested several years ago by Twomey and was applied to another version of the Chahine method known as the Twomey–Chahine algorithm.[8,17] However, we found that the Twomey–Chahine scheme may be highly unstable when the supposed initial distribution is too different from the distribution to be retrieved, and we think that this is due to a nonadequate normalization in the weight function used in that method. On the contrary, our technique seems to be completely reliable in this respect, and we never noticed a dependence of the results on different starting distributions.

We estimated the convergence and the reliability of our method by monitoring the behavior, as a function of the number of iterations, of the following parameter:

$$\text{rme} = \left\{ \frac{1}{q} \sum_{i=1}^{q} \frac{[\alpha_{\text{meas}}(\lambda_i) - \alpha_{\text{calc}}(\lambda_i)]^2}{[\alpha_{\text{calc}}(\lambda_i)]^2} \right\}^{1/2}. \tag{11}$$

The root mean error (rme) describes the average rms relative deviation of the retrieved signals, $\alpha_{\text{calc}}(\lambda_i)$, from the measured signals, $\alpha_{\text{meas}}(\lambda_i)$.

In all our tests we observed that the rme decreases with the number of iterations, and no evidence of a minimum was ever observed. We found that, for noiseless signals, the inversion procedure is very stable with respect to the number of iterations. As the procedure goes on, the retrieved distributions match the expected ones better, and no sign of instability is observed whatsoever. We thus stopped the inversion procedure when the retrieved distributions had attained their steady-state shapes.

The case of noisy signals is different. Although the rme decreases with the number of iterations, some instabilities in the retrieved distributions might grow if the inversion procedure is pushed too far away. This occurs when the rme reaches a value comparable with the rms amount of noise present on the data, which is a fair indication that the signal sequence has been reconstructed as well as possible. We noted also that broad distributions permit a much faster convergence than narrow distributions but exhibit a higher tendency to produce instabilities. A stopping criterion should then be optimized, depending on both the particular distribution being tested and the level of noise present in the signals. However, because these data are not readily available in an experiment, some general criteria must be adopted. We found that a good method is to stop the procedure when the relative variation of the rme parameter between one iteration and the next one goes below a fixed threshold. We fixed this value equal to 10^{-3} and determined that this is a reasonable trade-off to permit both narrow and broad distributions to be retrieved fairly well.

3. Computer Simulations

We have tested our inversion method by computer simulations. Each simulation consists of two steps. First, simulated extinction signals (input signals) are generated by computer according to Eq. (2) for a given number distribution $N_{\text{inp}}(r)$ (input distribution) and for a given relative refraction index m_{inp}. Second, the refraction index m_{out} is supposed to be known, and then the input signals are processed through our algorithm. The retrieved number distribution $N_{\text{out}}(r)$ (output distribution) is compared with the input distribution on the basis of the three parameters: the average radius, the standard deviation, and the overall number concentration. The retrieved signals (output signals) associated with $N_{\text{out}}(r)$ are obtained with Eq. (3) and are compared with the input distributions on the basis of the rme parameter defined in Eq. (11). In most cases random noise is added to the input signals, and for each level of noise the procedure is repeated with 100 different samples of noise. Consequently the output distributions and the output signals are characterized by means of the average and standard deviations of the above parameters. In the following we use the symbols $\mathbf{r}_{\text{inp}}$, σ_{inp},

and c_{inp} to indicate, respectively, the average radius, the standard deviation, and the concentration of the input distribution; the symbols $\langle \mathbf{r}_{out} \rangle$, $\langle \sigma_{out} \rangle$, and $\langle c_{out} \rangle$ indicate the average values of the corresponding parameters for the output distributions, whereas the symbol $\langle \mathrm{rme} \rangle$ will represent the average relative root-mean deviation of the output to input signals. Here $\langle \cdots \rangle$ stands for an average over 100 samples of noise. In all the tests described below, both the input and output distributions are number distributions and the associated parameters are number average parameters.

In our simulations we selected $q = 50$ wavelengths, scaled as in Eq. (7) and spanning a range of a decade (0.2–2 μm). Therefore the ratio of the geometrical progression was $a = 1.047$, and the relative width of the classes $\delta_j / R_j = 4.7\%$. According to Eq. (6), we studied distributions covering a range of radii R_j of a decade with $R_{min} = 0.14$ μm and $R_{max} = 1.4$ μm. The index of refraction of the particles relative to the medium was $m_{inp} = 1.50$ at all wavelengths. The zero-iteration distribution was chosen to be constant ($N_j = \text{const.}$), and we verified that the final results were independent of this initial guess.

A. Results for the $m_{out} = m_{inp}$ Case

In this series of tests, we supposed the refraction index of the particles to be known exactly, so that all the inversions were carried out by imposition of $m_{out} = m_{inp}$.

In the first test we show how the method works with monodisperse distributions. We chose three input distributions with $\mathbf{r}_{inp}$ equal to 0.25, 0.50, and 1.0 μm, and we added a 3% rms noise to the input signals. As reported in Table 1, the parameters of the input distributions are recovered fairly well for all three cases with an accuracy of better than 2% for the concentration and the average radius. The full widths of the three distributions are ~20% of the respective average radii, and this is a reasonable estimate of the resolution of our method. In Fig. 2(a) we show the output distributions for the case of 0.50 μm corresponding to the noiseless case and to the first 10 samples of noise. Note that the noiseless distribution is somewhat narrower than the noisy distributions. This is due to the input signals for a monodisperse distribution always being very structured with deep oscillations superimposed [see Fig. 2(b)]. As a consequence the convergence of the method is somewhat slowed, and, for the noisy cases, the inversion procedure is stopped when the rme is still higher than the rms level of the noise added to the data. Vice versa in the noiseless case the inversion procedure is stopped when convergence has been attained, and consequently the distribution is much narrower.

Table 1. Output Parameters Corresponding to Three Different Monodisperse Input Distributions[a]

$\mathbf{r}_{inp}$ (10^{-3} μm)	$\langle \mathbf{r}_{out} \rangle$ (10^{-3} μm)	$\langle \sigma_{out} \rangle$ (10^{-3} μm)	$\langle c_{out} \rangle$ (10^{6} cm^{-3})	$\langle \mathrm{rme} \rangle$ (%)
250	246 ± 1	23 ± 1	1.02 ± 0.01	5.4 ± 0.4
500	493 ± 1	49 ± 3	1.02 ± 0.01	6.0 ± 0.4
1000	992 ± 5	101 ± 9	1.01 ± 0.01	5.9 ± 0.3

[a]For all three cases $c_{inp} = 1.00 \times 10^6$ cm^{-3}, and the noise added to the input data was 3% rms.

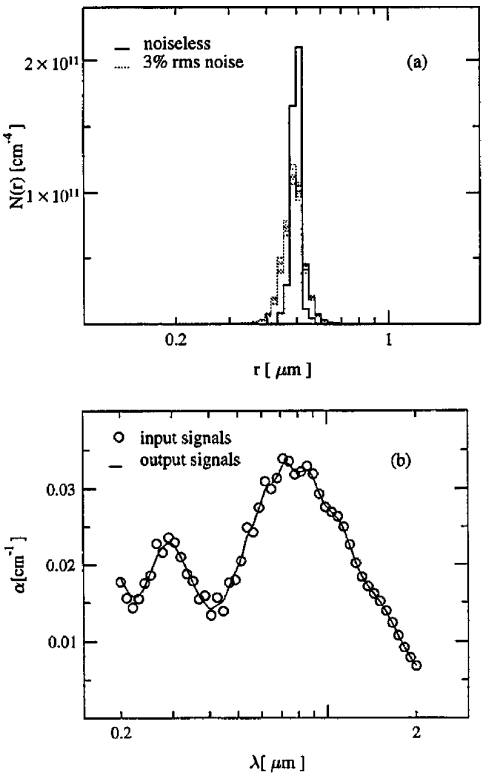

Fig. 2. Results of our inversion method in the case of a monodisperse input distribution with $\mathbf{r}_{inp} = 0.5$ μm: (a) Output distributions for the noiseless case (solid lines) and for 10 different samples of 3% rms noise (dotted lines). (b) Input signals (circles) and output signals corresponding to the noiseless case (solid curve).

The second test was devised to study how the algorithm works with broader distributions, and the results are shown in Fig. 3. A Gaussian-shaped input distribution characterized by $\mathbf{r}_{inp} = 0.5$ μm, $\sigma_{inp} = 0.1$ μm, $c_{inp} = 10^6$ cm^{-3} was used, and a 3% rms noise was added to the input signals. As a result of the inversion we obtain $\langle \mathbf{r}_{out} \rangle = 0.496 \pm 0.004$ μm, $\langle \sigma_{out} \rangle = 0.114 \pm 0.005$ μm, and $\langle c_{out} \rangle = (1.005 \pm 0.001) \times 10^{-6}$ cm^{-3}, which are in excellent agreement with the input parameters. Figure 3(a) shows the input distribution, the output distribution for the noiseless case, and the output distributions corresponding to the first 10 samples of noise. Without noise the output distribution is reconstructed in an excellent way, whereas with noise there is a slight tendency for tails to grow on the left side of the range. However, note that the distributions reported in Fig. 3(a) are number distributions, which are much more sensitive to the presence of small particles than volume or weight distributions. Had we plotted weight distributions, the match between input and output would have been much better, with no tails on small particles. In Fig. 3(b) the output signals corresponding to the

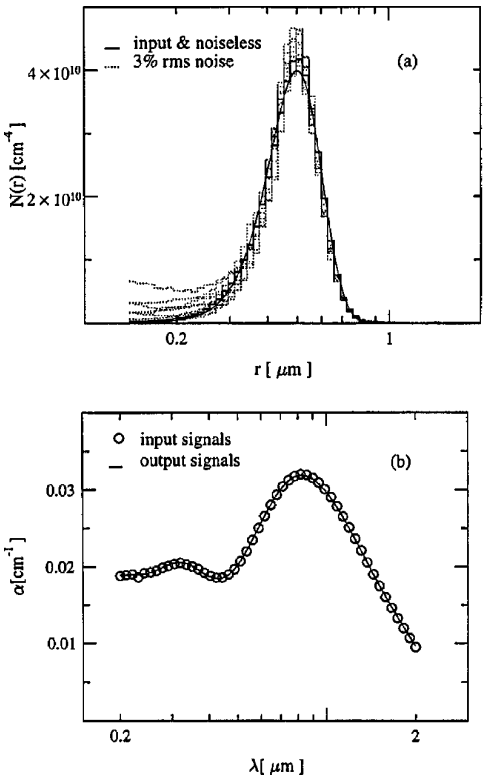

Fig. 3. Results of our inversion method in the case of a Gaussian input distribution characterized by $\mathbf{r}_{inp} = 0.5$ μm and $\sigma_{inp} = 0.1$ μm, $c_{inp} = 10^6$ cm^{-3}: (a) Input distribution (solid curve), output distribution for the noiseless case (solid lines), and output distributions for 10 different samples of 3% rms noise (dotted lines). (b) Input signals (circles) and output signals corresponding to the noiseless case (solid curve).

noiseless output distribution are compared with the input signals. These data are fitted much better than in the case of monodisperse distributions as we expect because of the smoothness of the input signals. Correspondingly the average stopping value of the rme parameter was $\langle \text{rme} \rangle = (3.1 \pm 0.2)\%$, which is very close to the 3% rms noise added to the input signals.

In the next figure we show an example of how our method works compared with the classical Chahine method. We have repeated the test shown in Fig. 3, inverting the same input signals (with the same 100 samples of 3% rms noise) but using the classical Chahine method. Results are shown in Fig. 4 where we report the input distribution (solid curve), the output distribution for the noiseless case (solid lines), and, for clarity, only one output distribution corresponding to a noisy case (dashed lines), all the others being very similar. The distributions reported in Fig. 4 are typical of the classical Chahine method. With noiseless data the distribution is likely to be characterized by the presence of spurious and deep oscillations; when noise is present in the data, this effect is strongly enhanced and the distribution may become dramatically noisy and spiked. However, noisy distributions do not always mean inaccurate results. Indeed, coarse grain averaging gives the

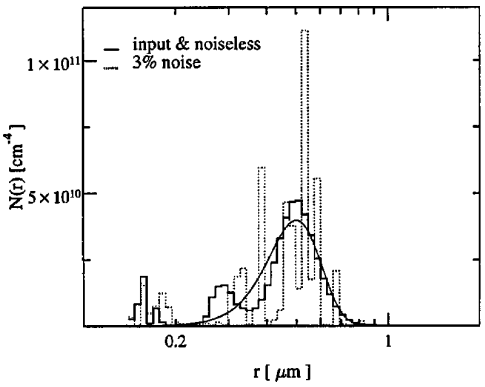

Fig. 4. Results of the classical Chahine inversion method for the same input distribution of Fig. 3: input distribution (solid curve), output distribution for the noiseless case (solid lines), and a typical output distribution for a 3% rms noise (dotted lines).

following results: $\langle \mathbf{r}_{out} \rangle = 0.495 \pm 0.008$ μm, $\langle \sigma_{out} \rangle = 0.117 \pm 0.009$ μm, and $\langle c_{out} \rangle = (1.04 \pm 0.03) \times 10^{-6}$ cm^{-3}, which are fairly satisfactory. Nevertheless note that, although the average values of the above parameters are somewhat comparable with those obtained with our method, the error bars associated with them are at least twice as large (for c_{out} there is even more than 1 order of magnitude), implying that our method is much more immune to noise than the classical Chahine method. This is also confirmed by the fact that the average stopping value of the rme parameter is $\langle \text{rme} \rangle = (4.8 \pm 0.7)\%$, meaning that the input signals have been more poorly reconstructed. A summary of the results of the comparison between our method and the classical Chahine method is reported in Table 2.

To investigate the effect of noise on our method, we used the same Gaussian distribution of Fig. 3 and added different levels of noise to the input signals. The results for the parameters of interest are in Fig. 5(a) where the ratios $\langle \mathbf{r}_{out}/\mathbf{r}_{inp} \rangle$, $\langle \sigma_{out}/\sigma_{inp} \rangle$, $\langle c_{out}/c_{inp} \rangle$, and respective error bars are plotted as a function of the noise level. Figure 5(a) shows that there are two effects on increasing the noise level. First, the mismatching between input and output parameters grows larger, implying that the accuracy of the method deteriorates at high noise levels; the systematic trend to reduce the average radius and increase both the standard deviation and the concentration is consistent with that shown in Fig. 3(a), i.e., with the presence of spurious tails on the left side of the particle range. Second, the error bars associated with the output parameters become larger, indicating

Table 2. Comparison between the Classical Chahine Method and our Method for a Gaussian Input Distribution Characterized by $r_{inp} = 0.5$ μm, $\sigma_{inp} = 0.1$ μm, and $c_{inp} = 1.00 \times 10^6$ cm^{-3}[a]

Method	$\langle \mathbf{r}_{out} \rangle$ (10^{-3} μm)	$\langle \sigma_{out} \rangle$ (10^{-3} μm)	$\langle c_{out} \rangle$ (10^6 cm^{-3})	$\langle \text{rme} \rangle$ (%)
Chahine	495 ± 8	117 ± 9	1.04 ± 0.03	4.8 ± 0.7
Ours	496 ± 4	114 ± 5	1.005 ± 0.001	3.1 ± 0.2

[a]Three percent rms noise was added to the input signals.

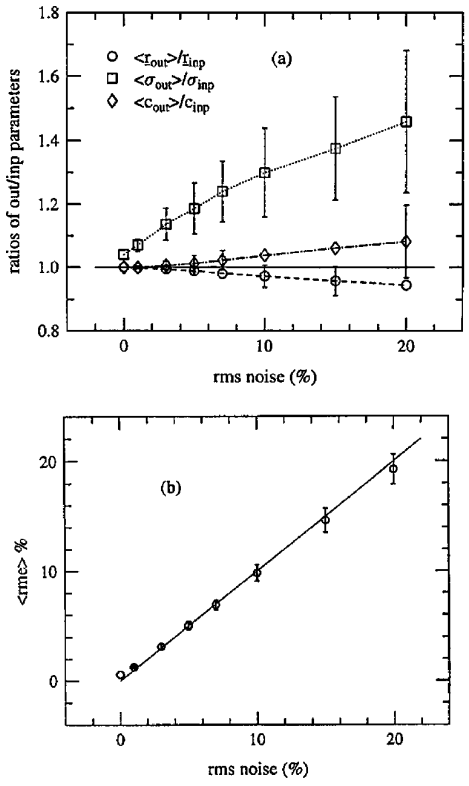

Fig. 5. Comparison between output and input parameters characterizing the same Gaussian distribution shown in Fig. 3: (a) Behavior of the ratios $\langle \mathbf{r}_{out} \rangle / \mathbf{r}_{inp}$, $\langle \sigma_{out} \rangle / \sigma_{inp}$, and $\langle c_{out} \rangle / c_{inp}$ as a function of the noise level. The error bars show the spread of the results obtained over 100 different samples of noise. The lines through the symbols are guides to the eye. (b) Behavior of the stopping value of the $\langle$rme$\rangle$ as a function of the noise level. The line corresponds to $\langle$rme$\rangle$ = rms noise.

that the method is becoming more and more sensitive to noise, with σ_{out} being the most critical parameter. This implies that, depending on the particular sample of noise being considered, the output distributions will be significantly different from the input distribution and their number will increase, statistically increasing the noise level. (We observed that, for noise levels near 10–20% rms, this number is of the order of 10%.) Figure 5(a) shows also that the above effects, i.e., the systematic and statistical deviations between input and output parameters, are of the same order of magnitude. Therefore they are both to be taken into account when the performances of the method are to be ascertained on noisy data. In particular, the parameters $\langle \mathbf{r}_{out} \rangle$ and $\langle c_{out} \rangle$ can be recovered quite well, and, even with 10% rms noise, their accuracy can be ~3% and ~4%, respectively. However, for the parameter $\langle \sigma_{out} \rangle$ a noise level of 3% rms is sufficient to reduce its accuracy to ~14%. As a final comment, to improve the results of Fig. 5(a), a better or different stopping criterion could be selected, and we are still working on this. However, the criterion adopted here is fairly satisfactory, and this is shown in Fig. 5(b) where we report the stopping value of the $\langle$rme$\rangle$ parameter as a function of the noise level. It is remarkable that, within the error bars, $\langle$rme$\rangle$ is equal to the rms noise for the entire range of noise.

From the above test we notice that σ is the parameter most sensitive to noise. To investigate further the ability of our inversion procedure to recover distributions with different σ values, we have generated input signals corresponding to Gaussian distributions with $\mathbf{r}_{inp}$ = 0.5 μm and σ varying between 0 and 0.2 μm. The dependence of the output ratio $\langle \sigma_{out} \rangle / \langle \mathbf{r}_{out} \rangle$ on the input ratio $\sigma_{inp} / \mathbf{r}_{inp}$ for different levels of noise is reported in Fig. 6. As shown, the output distributions tend to be always larger than the input ones, with the mismatching increasing with the noise level. For broad distributions the accuracy of the results is fairly satisfactory, although for narrow distributions the discrepancy is obviously a result of the finite resolution of the inversion procedure.

As a further test for investigating the resolving power of our method, we considered bi-Gaussian input distributions and looked for the minimum distance between the two peaks, which can be resolved. We performed the tests by adding 3% rms noise to the input data. We show in Fig. 7 the results when the two Gaussians are located near the minimum resolved distance. The two input Gaussians are characterized by $\mathbf{r}_{inp}{}^{a}$ = 0.40 μm, $\sigma_{inp}{}^{a}$ = 0.02 μm, $c_{inp}{}^{a}$ = 0.5×10^6 cm^{-3} and by $\mathbf{r}_{inp}{}^{b}$ = 0.60 μm, $\sigma_{inp}{}^{b}$ = 0.03 μm, $c_{inp}{}^{b}$ = 0.5×10^6 cm^{-3}. Figure 7 is consistent with the estimated resolution of the method reported in Table 1 where we show that, for 3% rms noise, the full width of a monodisperse distribution is of the order of 20% of the average radius, corresponding to approximately five classes of particles.

Finally, we investigated the ability of the method to reconstruct multipeaked distributions. We chose a tri-Gaussian input distribution characterized by the three average radii equal to 0.30, 0.60, and 1.20 μm. Their standard deviations were equal to 10% of their average radii, the concentrations were all equal to 0.33×10^6 cm^{-3}, and 1% rms noise was added to the input signals. The results are in Fig. 8 where it can be seen that the three peaks are reconstructed fairly

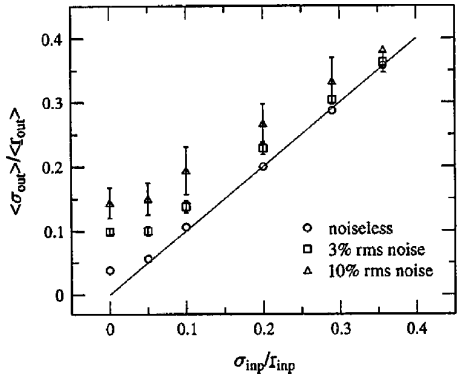

Fig. 6. Behavior of the output ratio $\langle \sigma_{out} \rangle / \langle \mathbf{r}_{out} \rangle$ as a function of the input ratio σ_{inp} / r_{inp} for different levels of noise. Input distributions were Gaussians with $\mathbf{r}_{inp}$ = 0.5 μm and σ_{inp} varying between 0 and 0.2 μm. The error bars indicate the spread of the different values obtained over 100 different samples of noise.

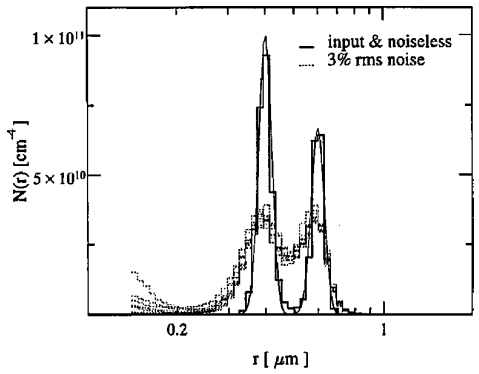

Fig. 7. Results of our inversion method in the case of a bi-Gaussian input distribution with the two peaks located near the minimum resolved distance. The two Gaussians are characterized by $\mathbf{r}_{inp}{}^{a} = 0.4$ μm, $\sigma_{inp}{}^{a} = 0.02$ μm, $c_{inp}{}^{a} = 0.5 \times 10^{6}$ cm^{-3} and by $\mathbf{r}_{inp}{}^{b} = 0.6$ μm, $\sigma_{inp}{}^{b} = 0.03$ μm, $c_{inp}{}^{b} = 0.5 \times 10^{6}$ cm^{-3}. Three percent rms noise was added to the input signals. The input distribution (solid curve), output distribution for the noiseless case (solid lines), and output distributions for 10 different samples of noise (dotted lines) are shown.

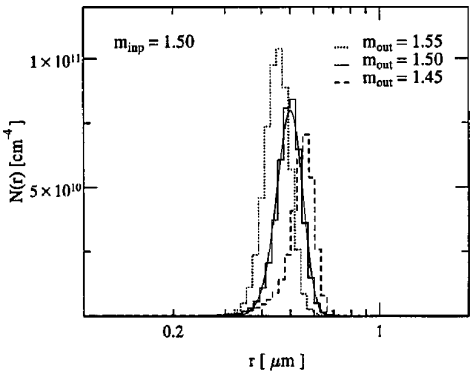

Fig. 9. Results of our inversion method when $\mathrm{Re}\{m_{out}\} \neq \mathrm{Re}\{m_{inp}\}$. The input distribution is a Gaussian with $\mathbf{r}_{inp} = 0.5$ μm, $\sigma_{inp} = 0.05$ μm, $c_{inp} = 10^{6}$ cm^{-3}, and $m_{inp} = 1.50$ (solid curve). We retrieve the output distributions supposing that $m_{out} = 1.45$ (dashed lines), $m_{out} = 1.50$ (solid lines), and $m_{out} = 1.55$ (dotted lines). All the tests were done with no noise added to the input signals.

well, with the first being less accurate. We have not investigated how much further we can push our method, but from Fig. 8 we can conclude that, with 1% rms noise present on the data, at least 3 peaks/decade can be resolved.

B. Results for the $m_{out} \neq m_{inp}$ Case

This last series of tests is aimed at the investigation of the sensitivity of the inversion method to errors in the estimate of the index of refraction of the particles. These tests were done with no noise added to the input signals. In the first test we studied the dependence on the real part of the refraction index for a Gaussian distribution with $\mathbf{r}_{inp} = 0.5$ μm, $\sigma_{inp} = 0.05$ μm, $c_{inp} = 1.00 \times 10^{6}$ cm^{-3}, and $m_{inp} = 1.50$. The data were inverted by use of different values of m_{out}, and the results are in Fig. 9 for the $m_{out} = 1.45$, $m_{out} = 1.50$, and $m_{out} = 1.55$ cases. It is consistent with Fig. 1 that when $m_{inp} < m_{out}$ the output distribution is shifted toward small particles and is enhanced, whereas if $m_{inp} > m_{out}$ it is shifted toward large particles and is depressed. In both cases the errors in the output average radii and concentrations were of the order of 10% and 16%, respectively. Figure 9 shows also that the widths of the distributions were approximately the same for the three retrieved distributions.

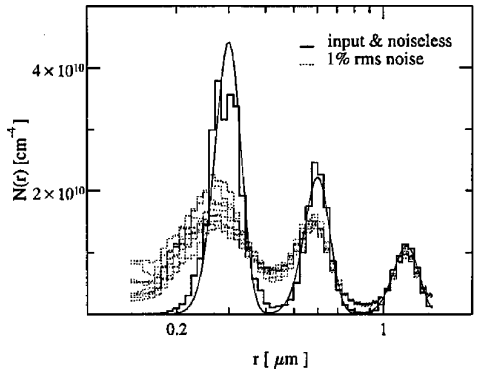

Fig. 8. Results of our inversion method in the case of a tri-Gaussian input distribution characterized by $\mathbf{r}_{inp}{}^{a} = 0.3$ μm and $\sigma_{inp} = 0.03$ μm, $\mathbf{r}_{inp}{}^{b} = 0.6$ μm and $\sigma_{inp}{}^{b} = 0.06$ μm, and $\mathbf{r}_{inp}{}^{c} = 1.2$ μm and $\sigma_{inp}{}^{c} = 0.12$ μm. The concentrations were all the same equal to $c_{inp} = 0.33 \times 10^{6}$ cm^{-3}. One percent rms noise was added to the input data. The input distribution (solid curve), output distribution for the noiseless case (solid lines), and output distributions for 10 different samples of noise (dotted lines) are shown.

In the second test we inverted the same data, supposing that m_{out} has an imaginary part. The results are shown in Fig. 10(a) for the $m_{out} = 1.50 + 0.05i$ case. The average radius and concentration of the input distribution are recovered fairly well, with an accuracy of ≈1% and ≈5%, respectively, whereas the width of the output distribution is approximately one half that of the input one. This can be understood by consideration of the large mismatch between the input and output signals shown in Fig. 10(b). Indeed, it is consistent with Fig. 1 that the two signal sequences have maxima located at the same λ, but the output signals are much flatter than the input ones. As a consequence the output distribution is centered correctly but its width is artificially narrowed.

To study the behavior of the parameters of the output distributions as a function of m_{out}, we have repeated the same tests as in Figs. 9 and 10 over a range of values of $\mathrm{Re}\{m_{out}\}$ and $\mathrm{Im}\{m_{out}\}$, and results are reported in Figs. 11 and 12. Figure 11 shows that $\mathbf{r}_{out}$ and c_{out} are very sensitive to errors in $\mathrm{Re}\{m_{out}\}$, and, if they are desired with an accuracy of 1%, $\mathrm{Re}\{m_{out}\}$ has to be known with a precision of nearly 0.4% and 0.2%, respectively. Figure 11 shows also that σ_{out} is slightly less sensitive to $\mathrm{Re}\{m_{out}\}$, because the corresponding curve presents a minimum when $\mathrm{Re}\{m_{out}\} = \mathrm{Re}\{m_{inp}\}$. In fact, as long as errors in $\mathrm{Re}\{m_{out}\}$ are less than a few percent, the artificial broadening of the output distribution is always smaller than its intrinsic width because of the finite resolution of the inversion procedure. Figure 12 shows the behavior of the same parameters of Fig.

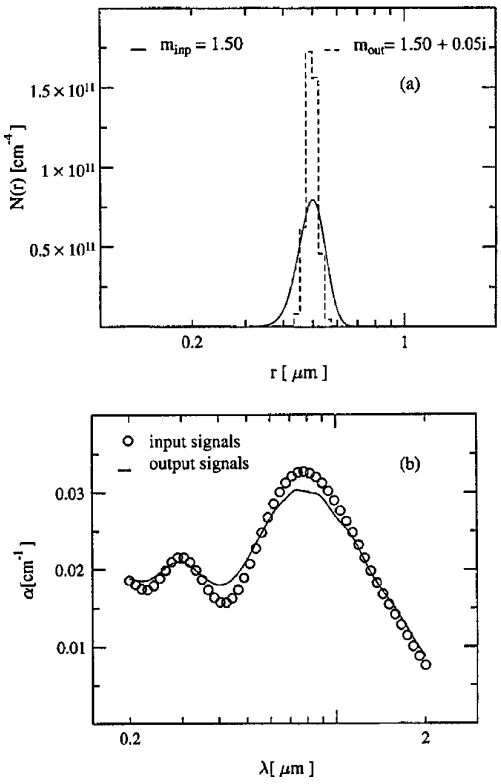

Fig. 10. Results of our inversion method when $\mathrm{Im}\{m_{out}\} \neq \mathrm{Im}\{m_{inp}\}$. The input distribution is a Gaussian with $\mathbf{r}_{inp} = 0.5$ μm, $\sigma_{inp} = 0.05$ μm, $c_{inp} = 10^6$ cm^{-3}, and $m_{inp} = 1.50$. (a) Input distribution (solid curve) and output distribution retrieved if $m_{out} = 1.50 + 0.05i$ is supposed. (b) Input signals (circles) and output signals (solid curve). The test was done with no noise added to the input signals.

11 as a function of $\mathrm{Im}\{m_{out}\}$. In this case $\mathbf{r}_{out}$ and c_{out} are the most insensitive parameters, and they can be recovered with accuracies of 1% and 5%, respectively, even if particles with a purely real refraction index are supposed to be fairly absorbing with $\mathrm{Im}\{m_{out}\} = 0.05$. On the contrary the parameter σ_{out} is extremely sensitive to $\mathrm{Im}\{m_{out}\}$, and a value of $\mathrm{Im}\{m_{out}\} = 0.01$ is enough to lower the accuracy of σ_{out} to ≈25%.

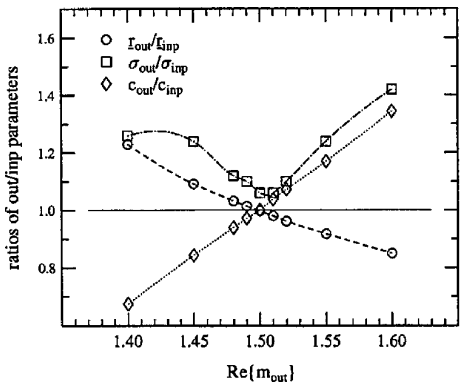

Fig. 11. Behavior of the ratios $\mathbf{r}_{out}/\mathbf{r}_{inp}$, $\sigma_{out}/\sigma_{inp}$, and c_{out}/c_{inp} as a function of $\mathrm{Re}\{m_{out}\}$. The input distribution was the same as in Fig. 9 with $m_{inp} = 1.50$. All the tests were done with no noise added to the input data. The lines through the symbols are guides to the eye.

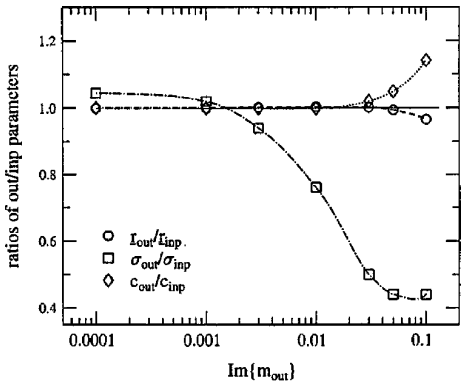

Fig. 12. Behavior of the ratios $\mathbf{r}_{out}/\mathbf{r}_{inp}$, $\sigma_{out}/\sigma_{inp}$, and c_{out}/c_{inp} as a function of $\mathrm{Im}\{m_{out}\}$. The input distribution was the same as in Fig. 9 with $m_{inp} = 1.50$. All the tests were done with no noise added to the input data. The lines through the symbols are guides to the eye.

As a final comment, note that the curves in Figs. 9–12 and the relative results refer to a specific case in which a particular distribution has been considered; so we should take care when generalizing to other cases. Nevertheless we think that the above results indicate the accuracies attainable when samples with an unknown index of refraction are studied. Moreover the strong dependence of the parameters shown in Figs. 11 and 12 could be exploited to measure the index of refraction of particles. Indeed, if samples of particles with known distributions but unknown refraction index are considered, we could recover their refraction index by finding that value of m_{out} that permits the best fit between input and output distributions and signals.

4. Conclusions

In this work we have tested the accuracy, reliability, and sensitivity of an innovative inversion algorithm applied to the inversion of spectral extinction data for particle sizing.

The technique proposed is a modified version of the algorithm originally developed by Chahine in 1968, but, compared with it, our technique is much more stable, reliable, and accurate with respect to noise. Moreover, in contrast to the original scheme, which works well when the kernel is a smooth function with a single maximum, our method is able to cope with kernels that present oscillations and secondary maxima, as in the case of spectral extinction data.

The results of our computer simulations show that the method can accurately recover a large variety of different distributions, including monodisperse, narrow, and broad bell-shaped distributions and multipeaked distributions as well. The accuracy and resolution of the inversion scheme have been studied as a function of the noise level present in the data and showed that the first three moments of the particles' distributions, i.e., concentration, average radius, and standard deviation, can be recovered fairly well. In

particular, we have shown that, for monodisperse distributions with up to 3% rms noise, the accuracies on radius and concentration are better than 2%; in these conditions the resolution of the method is of the order of 20%. For broader distributions the performances of the method are even better and accuracies of a few percent on the average radius and concentration can be reached even in the presence of 10% rms noise. The overall shape of the distributions is also reconstructed fairly well, with no sign of spurious oscillations or noisy peaks. In particular, the noisy and indented appearance typical of the distributions recovered with the classical Chahine method has been removed. Distributions retrieved with our method are always very smooth, and there is only a slight tendency for tails to grow toward the small particle side of the range. However, this discrepancy would have been more negligible if we had expressed our results in terms of volume or weight distributions.

We have also studied the influence of errors in the estimated refractive index of the particles on the retrieved distributions. Our results show that there is a strong sensitivity of the parameters characterizing the output distributions on both the real and imaginary parts of the particle's refractive index. This suggests the possibility of exploiting spectral extinction data to measure the refractive index of samples with known size distributions.

Finally, we have applied our method to the inversion of real extinction data obtained by adapting a commercial spectrophotometer. Our results are quite satisfactory and will be published in another paper.

After we submitted this paper, M. Bertero drew to our attention that the inversion method proposed here for particle sizing has also been used in the field of medical imaging and emission tomography. In this area the method is known as the expectation maximization method and was originally proposed by Richardson[38] and Lucy.[39]

We thank A. Andreoni, Institute of Mathematical, Physical and Chemical Sciences–Como, for helpful discussions and M. Bertero, Department of Physics, University of Genova–Genova, for drawing our attention to the existence of the expectation maximization method.

References

1. M. Kerker, *The Scattering of Light and Other Electromagnetic Radiation* (Academic, New York, 1969), Chap. 7.
2. G. Gouesbet and G. Gréhan, eds., *Proceedings of an International Symposium on Optical Particle Sizing: Theory and Practice* (Plenum, New York, 1988).
3. E. D. Hirleman, ed., *Proceedings of the Second International Congress on Optical Particle Sizing* (Arizona State University Printing Services, Tempe, Ariz., 1990), pp. 169–435.
4. M. Maeda, S. Nakae, and M. Ikegami, eds., *Proceedings of the Third International Congress on Optical Particle Sizing* (n.p., 1993).
5. S. Twomey and H. B. Howell, "Some aspects of the optical estimation of microstructure in fog and cloud," Appl. Opt. **6,** 2125–2131 (1967).
6. A. Ångström, "On the atmospheric transmission of sun radiation and on dust in the air," Geogr. Ann. **11,** 156–166 (1929).
7. J. A. Curcio, "Evaluation of atmospheric aerosol particle size distribution from scattering measurements in the visible and infrared," J. Opt. Soc. Am. **51,** 548–551 (1961).
8. S. Twomey, "Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions," J. Comput. Phys. **18,** 188–200 (1975).
9. S. Twomey, "The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements," J. Franklin Inst. **279,** No. 2, 95–109 (1965).
10. E. E. Uthe, "Particle-size evaluations using multiwavelength extinction measurements," Appl. Opt. **21,** 454–459 (1982).
11. R. L. Zollars, "Turbidimetric method for online determination of latex particle number and particle-size distribution," J. Colloid Interface Sci. **74,** 163–172 (1980).
12. H. Quenzel, "Determination of size distribution of atmospheric aerosol particles from spectral solar radiation measurements," J. Geophys. Res. **75,** 2915–2921 (1970).
13. G. E. Shaw, "Intercomparison of equatorial and polar multiwavelength atmospheric optical depths and sky radiance," Bull. Am. Meteorol. Soc. **54,** 1073–1080 (1973).
14. D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," J. Assoc. Comput. Mach. **9,** 84–97 (1962).
15. S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," J. Assoc. Comput. Mach. **10,** 97–101 (1963).
16. G. E. Backus and J. F. Gilbert, "Numerical applications of a formalism for geophysical inverse problems," Geophys. J. R. Astron. Soc. **13,** 247–276 (1967).
17. S. Twomey, *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements* (Elsevier, Amsterdam, 1977), Chap. 7, p. 179.
18. N. Wolfson, J. H. Joseph, and Y. Mekler, "Comparative study of inversion techniques. Part I: Accuracy and stability," J. Appl. Meteorol. **18,** 543–555 (1979).
19. N. Wolfson, Y. Mekler, and J. H. Joseph, "Comparative study of inversion techniques. Part II: Resolving power, conservation of normalization and superposition principles," J. Appl. Meteorol. **18,** 556–561 (1979).
20. G. Yamamoto and M. Tanaka, "Determination of aerosol size distribution by spectral attenuation measurements," Appl. Opt. **8,** 447–453 (1969).
21. 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).
22. P. T. Walters, "Practical applications of inverting spectral turbidity data to provide aerosol size distributions," Appl. Opt. **19,** 2353–2365 (1980).
23. 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).
24. 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).
25. M. Bertero, C. De Mol, and E. R. Pike, "Particle size distributions from spectral turbidity: a singular-system analysis," Inverse Problems **2,** 247–258 (1986).
26. M. Bertero, C. De Mol, and E. R. Pike, "Particle sizing by inversion of extinction data," in Ref. 2, pp. 55–61.
27. 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).

28. M. T. Chahine, "Determination of the temperature profile in an atmosphere from its outgoing radiance," J. Opt. Soc. Am. **58,** 1634–1637 (1968).
29. M. T. Chahine, "Inverse problems in radiative transfer: determination of atmospheric parameters," J. Atmos. Sci. **27,** 960–967 (1970).
30. H. Grassl, "Determination of aerosol size distributions from spectral attenuation measurements," Appl. Opt. **10,** 2534–2538 (1971).
31. R. Santer and M. Herman, "Particle size distribution from forward scattered light using the Chahine inversion scheme," Appl. Opt. **22,** 2294–2301 (1983).
32. F. Ferri, M. Giglio, and U. Perini, "Inversion of light scattered data from fractals by means of the Chahine iterative algorithm," Appl. Opt. **28,** 3074–3082 (1989).
33. A. Bassini, S. Musazzi, E. Paganini, U. Perini, F. Ferri, and M. Giglio, "Optical particle sizer based on the Chahine inversion scheme," Opt. Eng. **31,**1112–1117 (1992).
34. E. Trakhovsky, S. G. Lipson, and A. D. Devir, "Atmospheric aerosols investigated by inversion of experimental transmittance data," Appl. Opt. **21,** 3005–3010 (1982).
35. J. G. Crump and J. H. Seinfeld, "A new algorithm for inversion of aerosol size distribution data," Aerosol Sci. Technol. **1,** 15–34 (1982).
36. C. F. Bohren and D. R. Hufman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983), p. 77.
37. H. C. Van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1981), Chap. 9, p. 127.
38. W. H. Richardson, "Bayesian based iterative method of image restoration," J. Opt. Soc. Am. **62,** 55–59 (1972).
39. L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron. J. **79,** 745–754 (1974).