JOURNAL OF COMPUTATIONAL PHYSICS 52, 80–104 (1983)

A Simple Inversion Method for Determining Aerosol Size Distributions

THOMAS KAIJSER

National Defence Research Institute (FOA), 581 11 Linköping, Sweden

Received July 30, 1981; revised February 2, 1983

The retrieval of aerosol size distributions from extinction measurements is a problem which has been studied for more than two decades. In this paper a new method for attacking this problem is presented. The method produces a solution representable as a linear combination of a few log-normal functions.

1. INTRODUCTION

The purpose of this paper is to describe a simple method of finding an approximate, but physically realistic, solution to the equations

$$\int_0^{\infty} K(\lambda_i, r) f(r)\, dr = b_i, \qquad i = 1, 2, \ldots, M, \tag{1.1}$$

where (1) the b_i, $i = 1, 2, \ldots, M$, are extinction measurements at M different wavelengths for radiation passing through an aerosol, (2) the function f represents the (unknown) aerosol size distribution ($f(r) =$ the number of particles per cubic centimeter in the radius range $(r, r + dr)$), and (3)

$$K(\lambda, r) = \pi r^2 Q_{\text{ext}}(m_\lambda, 2\pi r/\lambda) \tag{1.2}$$

where $Q_{\text{ext}}(m_\lambda, 2\pi r/\lambda)$ is the extinction for a plane wave of wavelength λ passing through a spherical, homogeneous particle of radius r and refractive index m_λ.

The method is based on the assumption that an aerosol size distribution can be described by a linear combination of a few log-normal functions. However, the restriction to log-normal functions is not crucial; other families of functions—for example gamma functions or "rectangular" functions—can also be used. Nor does the method depend explicitly on the special analytic form of the kernel $K(\lambda, r)$. It is therefore possible that the method can be used for other problems as well.

There exist several inversion methods which have been used for retrieving the aerosol size distribution from extinction measurements. Wolfson *et al.* [15] have analysed and compared four of these, namely, the Chahine-method; the regularization

0021-9991/83 $3.00

method by Phillips, Twomey, and Tikhonov; the method by Backus and Gilbert; and the non-linear regression method. One conclusion obtained by Wolfson *et al.* was, roughly speaking, that all four methods have some disadvantages which cause difficulties when applied in practice.

In two recent and very interesting papers, the first by Rizzi *et al.* [10], the second by King [4], the regularization method is analysed in great detail. Both papers demonstrate the sensitivity of this inversion method to the selection of the so-called Lagrange multiplier needed in the method. These papers therefore also confirm the assertion that it is not easy to use the regularization method for determining aerosol size distributions from extinction measurements.

One of the main advantages of the new method we describe in this paper is that it works very well in practice.

The plan of the paper is as follows. In Section 2 we give a brief description of the kernel K. In Section 3 we introduce some notations and in Section 4 we discuss positive solutions to (1.1). In Section 5 we introduce some classes of log-normal functions, and discuss the relationship between log-normal functions and aerosol size distributions. In Sections 6 and 7 we present the method, and in Section 8 we make some remarks about the method. In Section 9 we apply the method to three cases of real data that can be found in the literature, and in section 10 we do some sensitivity tests. In Section 11 we make a few comments regarding computer-times, and in the last section we summarize the main advantages of the method.

2. A Few Words about the Kernel

As we saw from (1.2) the kernel $K(\lambda, r)$ is the product of πr^2 and the extinction coefficient $Q_{\text{ext}}(m_\lambda, 2\pi r/\lambda)$. The extinction coefficient was determined analytically by G. Mie in 1908. It can be expressed as an infinite sum

$$Q_{\text{ext}} = 2x^{-2} \sum_{0}^{\infty} (2n+1)\,\text{Re}(a_n + b_n)$$

where a_n and b_n are linear fractions of Bessel-functions and where x is equal to $2\pi r/\lambda$. For the precise definitions of a_n and b_n see, e.g., Twomey [12 p. 206].

In Fig. 1, we have depicted the function $Q_{\text{ext}}(m_\lambda, \cdot)$ when m_λ is equal to 1.33. The shape of the function $Q_{\text{ext}}(m_\lambda, \cdot)$ is similar for other values of m_λ. Since $Q_{\text{ext}}(m_\lambda, \cdot)$ oscillates around the line $y = 2$, and since $K(\lambda, r)$ is the product of πr^2 and $Q_{\text{ext}}(m_\lambda, 2\pi r/\lambda)$, the shape of $K(\lambda, r)$ for varying r and fixed λ is like a sequence of terraces along the curve $y = 2\pi r^2$, $r > 0$.

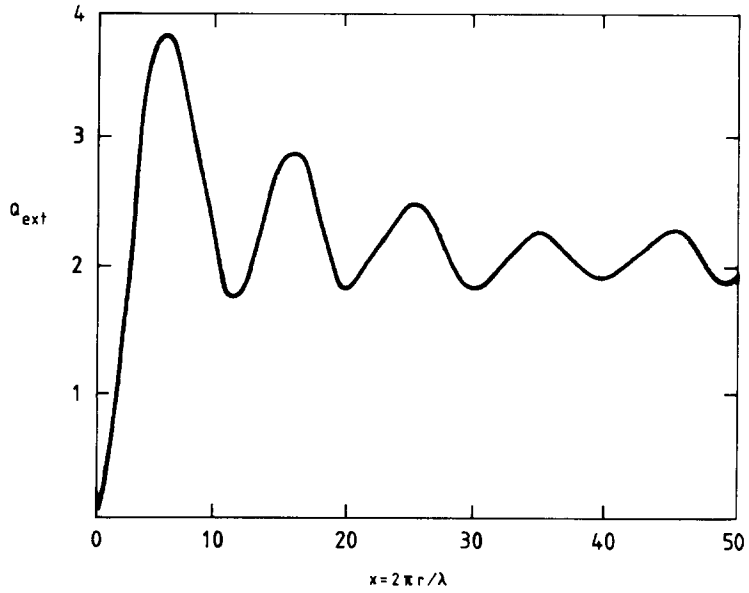

FIG. 1. The graph of the function $Q_{ext}(m_\lambda, \cdot)$ for $m_\lambda = 1.33$.

3. PRELIMINARY NOTATIONS

Let $\mathscr{L}$ denote the set of functions for which the integral

$$\int_0^\infty K(\lambda_i, r) f(r)\, dr$$

exists for $i = 1, 2, \ldots, M$. (From now on, when the integration limits are omitted, it is implicit that they are 0 and ∞.) For each f in $\mathscr{L}$ we set

$$(Kf)_i = \int K(\lambda_i, r) f(r)\, dr \tag{3.1}$$

and define

$$Kf = ((Kf)_1, (Kf)_2, \ldots, (Kf)_M)^* \tag{3.2}$$

where * denotes transposition. We shall call $(Kf)_i$ a *theoretical observation* at λ_i induced by f, and we shall call Kf the *theoretical observation vector* induced by f.

Next set

$$b = (b_1, b_2, \ldots, b_M)^*$$

where the b_i's, $i = 1, 2, \ldots, M$, are the measurements in (1.1). We call b the *measurement vector*.

For a vector x in $\mathbf{R}^N$ $(N \geqslant 2)$ we shall use either the notation x_i or the notation $(x)_i$ to denote the ith component of x. We let $\mathbf{R}_+^N$ denote those vectors in $\mathbf{R}^N$ which have non-negative components, and we let $\mathscr{L}_+$ denote those functions in $\mathscr{L}$ which are non-negative.

By dist$(\cdot, \cdot)$ we shall mean a distance-function on $\mathbf{R}^M$. Examples of distance-functions are

$$\text{dist}(x, y) = \sum_1^M |x_i - y_i| \tag{3.3}$$

$$\text{dist}(x, y) = \left(\sum_1^M |x_i - y_i|^2 \right)^{1/2} \tag{3.4}$$

$$\text{dist}(x, y) = \sum_1^M |x_i - y_i| \, w_i \tag{3.5}$$

$$\text{dist}(x, y) = \left(\sum_1^M |x_i - y_i|^2 \, w_i \right)^{1/2}. \tag{3.6}$$

In (3.5) and (3.6) the w_i's are weights (which may depend on y).

Note that for each f in $\mathscr{L}$, Kf is a vector in $\mathbf{R}^M$ and therefore dist(Kf, b) is well-defined for any distance-function on $\mathbf{R}^M$. For the sake of convenience, we shall call dist(Kf, b) f's *error*, in spite of the fact that dist(Kf, b) is not a measure of f's deviation from the true solution.

4. Non-negative Solutions

When attacking an integral equation of type (1.1), one is almost always faced with one fundamental difficulty, namely, that "for any given function there exist many other functions, usually infinitely many, and quite different from f, with the same theoretical observation vector (see (3.1) and (3.2)) as the one induced by f." This implies that even if the experimental situation is perfect and there are no errors, neither in measurements nor in the model, one cannot expect to retrieve the true solution—or a solution close to the true solution—from Eq. (1.1) alone; and when there are measurement and/or modelling errors things become even worse. In order to overcome this ambiguity it is necessary to impose one or more additional conditions. (The reader can, for example, consult Twomey [13] or Rust and Burrus [11] for further information about integral equations of type (1.1).)

In many problems one knows that the true solution must be non-negative. This, for example, is the case in our problem. In such situations it is natural to start by restricting the set of feasible solutions to the set of positive functions. When one makes this restriction, due to errors in measurements as well as in the model, it often happens that one cannot find *any* function which satisfies (1.1) exactly. Therefore one is lead to look at the following minimization-problem:

(4.1) *Find a function $\hat{f}$ in $\mathscr{L}_+$ such that*

(4.2) $$\operatorname{dist}(K\hat{f}, b) \leqslant \operatorname{dist}(Kf, b)$$

for every function f in $\mathscr{L}_+$.

Unfortunately, this is not a tractable problem since in general there does not exist a solution to the problem. This mathematical nuisance can be resolved by enlarging the set of positive functions to the set of positive measures. However, this would be of little help, since one does not know of any method for finding the minimum. A more practical procedure is to approximate the set of positive functions from "inside" in the following way.

Let I denote an arbitrary but fixed finite partition $0 < r_0 < r_1 < \cdots < r_N < \infty$ of the r-axis and define $\mathscr{L}_+(I)$ as those non-negative (left-continuous) functions which are constant on each interval $(r_{i-1}, r_i]$, $i = 1, 2, \ldots, N$, and which are zero otherwise. Now consider the following minimization-problem:

(4.3) *Find a function f in $\mathscr{L}_+(I)$ such that the inequality* (4.2) *holds for all f in $\mathscr{L}_+(I)$.*

Contrary to problem (4.1) this problem is quite easy to solve, at least if the distance function is given by any of the formulas (3.3) to (3.6). Since the new method we shall describe in Sections 6 and 7 has parts in common with the solution-procedure to problem (4.3) we shall now describe how problem (4.3) is solved.

Set

$$x_j = f(r_j), \qquad j = 1, 2, \ldots, N,$$

$$x = (x_1, x_2, \ldots, x_N)^*$$

$$a_{ij} = \int_{r_{j-1}}^{r_j} K(\lambda_i, r)\, dr, \qquad 1 \leqslant i \leqslant M,\ 1 \leqslant j \leqslant N,$$

and let $A = (a_{ij})$ be the $M \times N$ matrix which has a_{ij} as its elements. For a function f in $\mathscr{L}_+(I)$ it now follows that

$$\int K(\lambda_i, r) f(r)\, dr = \sum_{j=1}^{N} f(r_j) \left(\int_{r_{j-1}}^{r_j} K(\lambda_i, r)\, dr \right)$$

$$= \sum_{j=1}^{N} x_j a_{ij} = (Ax)_i$$

which implies that

(4.4) $$Kf = Ax$$

when f belongs to $\mathscr{L}_+(I)$. Now, suppose we have solved the following minimization problem:

(4.5) *Find a vector* $\hat{x}$ *in* $\mathbf{R}_+^N$ *such that*

$$\operatorname{dist}(A\hat{x}, b) \leqslant \operatorname{dist}(Ax, b)$$

for all vectors x *in* $\mathbf{R}_+^N$.

From (4.4) it then immediately follows that a solution to problem (4.3) is obtained by defining

$$\begin{aligned} \hat{f}(r) &= \hat{x}_j \qquad \text{if } r_{j-1} < r \leqslant r_j, \qquad j = 1, 2, \ldots, N, \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{4.6}$$

Thus to solve problem (4.3) it suffices to solve problem (4.5). But this is quite easy to do if the distance-function is defined by any of the formulas (3.3) to (3.6). For, if $\operatorname{dist}(\cdot, \cdot)$ is defined by (3.3) or (3.5), problem (4.5) can easily be transformed into a standard LP-problem (LP = linear programming), and if $\operatorname{dist}(\cdot, \cdot)$ is defined by (3.4) or (3.6) it can be transformed into a QP-problem (QP = quadratic programming). Since there exist good algorithms for solving LP-problems as well as QP-problems, we conclude that if $\operatorname{dist}(\cdot, \cdot)$ is defined by any of the formulas (3.3) to (3.6) we can solve problem (4.5), and hence also problem (4.3).

The question now arises whether the solution (4.6) is close to the true solution. If the partition is very fine and the number of division-points is much larger than the number of measurements then the solution (4.6) is in general *not close* to the true solution. For, in this case the solution vector $\hat{x}$ to problem (4.5) will have many of its components equal to zero, and this will imply that the solution (4.6) will be physically unrealistic. ($\hat{x}$ can never have more non-zero components than there are measurements.) In order to obtain a solution which is acceptable as a first approximation to the true solution, it is usually necessary to let the number of division-points in the partition be substantially smaller than the number of measurements. The disadvantage however, with a coarse partition is that the solution obtained will not satisfy Eq. (1.1) very well. One can usually find "better-looking" solutions with a smaller error. However, since problem (4.3) is easy to solve, it is convenient to use solutions to problem (4.3) for comparison with solutions obtained by other methods.

5. On Log-Normal Functions and Aerosol Size Distributions

A log-normal function is defined by

$$f(r) = (\sqrt{2\pi})^{-1} r^{-1} (\log \sigma)^{-1} \exp\left(-\frac{1}{2}\left(\frac{\log r - \log \mu}{\log \sigma}\right)^2\right)$$

for $r > 0$ and by $f(r) = 0$ otherwise. A log-normal function is thus defined by two positive parameters μ and σ; μ is called the *geometric mean*; σ is called the *geometric deviation*. We shall also call μ the *localization* parameter and σ the *dispersion*

parameter. We shall call the vector (μ, σ) the *parameter-vector* (determining the log-normal function). A log-normal function with parameter-vector (μ, σ) will be denoted $\mathrm{LN}(\mu, \sigma)$.

We shall next introduce some sets of functions. We let $\mathscr{H}$ denote the set of all log-normal functions. For every set P in the first quadrant for which the second component is larger than one, we let $\mathscr{H}(P)$ denote the set of all log-normal function with parameter-vector in P. For $N = 1, 2,...,$ we let $\mathscr{F}_N$ denote the set of functions which consists of all positive linear combinations of (at most) N log-normal functions and for every parameter-set P we let $\mathscr{F}_N(P)$ denote the set of all positive linear combinations of (at most) N log-normal functions with parameter vectors in P.

According to K. Whitby [14] there is now a considerable body of evidence to suggest that most atmospheric aerosol size distributions are basically trimodal. The three modes are called the nuclei mode, the accumulation mode, and the coarse paticle mode. By using (1) the fact that aerosol size distributions are basically trimodal, (2) the central limit theorem, and (3) the fact that if a stochastic variable X has a log-normal distribution the same is true for any power of X, Whitby concludes [14, pp. 138–143] that aerosol size distributions are best described as a positive linear combination of three log-normal functions, that is, a function in $\mathscr{F}_3$. Similar arguments and conclusions can be found in Patterson and Gillette [9].

Let us now introduce the following minimization-problem which we shall call the $\mathscr{F}_N(P)$-problem:

(5.1) *Find a function $\hat{f}$ in $\mathscr{F}_N(P)$ such that*

$$\mathrm{dist}(K\hat{f}, b) \leqslant \mathrm{dist}(Kf, b)$$

for all f in $\mathscr{F}_N(P)$.

What we would like to do is to solve this problem with a small value on N and with a parameter-set P chosen in such a way that the solution can be used as a description of an aerosol size distribution. Before we show how one can achieve this, we shall first show how the $\mathscr{F}_N(P)$-problem can be solved, when the parameter-set P consists of *exactly* N parameter-vectors.

6. On the $\mathscr{F}_N(P)$-Problem When P Consists of Exactly N Elements

Let P be a parameter-set with exactly N elements. Then the set $\mathscr{H}(P)$ consists of exactly N log-normal functions which we shall denote $h_j, j = 1, 2,..., N$. We call $\{h_j\}_1^N$ the set of *base-functions* (associated with the parameter-set P).

Next, let us determine the vector Kf for an arbitrary f in $\mathscr{F}_N(P)$. Since $\mathscr{H}(P)$ consists of exactly N functions, an arbitrary function in $\mathscr{F}_N(P)$ can—in this case—be written

$$f = \sum_1^N \alpha_j h_j \tag{6.1}$$

where $\alpha = (\alpha_1, \alpha_2, ..., \alpha_N)^*$ belongs to $\mathbf{R}_+^N$. Therefore the ith component of the vector Kf satisfies

$$(Kf)_i = \int K(\lambda_i, r) f(r)\, dr = \int K(\lambda_i, r) \left(\sum_1^N \alpha_j h_j(r) \right) dr \tag{6.2}$$

$$= \sum_1^N \alpha_j \int K(\lambda_i, r)\, h_j(r)\, dr = \sum_1^N \alpha_j d_{ij} = (D\alpha)_i$$

where

$$d_{ij} = \int K(\lambda_i, r)\, h_j(r)\, dr \tag{6.3}$$

and $D = (d_{ij})$ is the $M \times N$ matrix with d_{ij} as its elements. We call D the *base-matrix* associated with the parameter-set.

From (6.2) it follows that if f satisfies (6.1) then

$$Kf = D\alpha$$

where D is the base-matrix associated with the parameter-set P. Hence, if P consists of exactly N parameter-vectors then the $\mathcal{F}_N(P)$-problem is equivalent to the following minimization problem:

(6.4) *Find a vector $\hat{\alpha}$ such that*

$$\operatorname{dist}(D\hat{\alpha}, b) \leqslant \operatorname{dist}(D\alpha, b), \qquad \forall \alpha \in \mathbf{R}_+^N.$$

Once problem (6.4) is solved the solution to problem (5.1) is given by

$$\hat{f} = \sum_1^N \hat{\alpha}_j h_j.$$

Now, looking at problem (6.4) we note that it has *exactly* the same structure as problem (4.5). Hence, if the distance-function in problem (6.4) is defined by (3.3) or (3.5), problem (6.4) can be transformed into an LP-problem, whereas if the distance-function is defined by (3.4) or (3.6), it can be transformed into a QP-problem. Since there exist efficient algorithms for solving LP-problems and QP-problems, we conclude that if the parameter-set P consists of exactly N elements then it is easy to solve the $\mathcal{F}_N(P)$-problem. In the next section we shall see how this fact can be utilized in order to obtain approximate solutions to problem (5.1) (the $\mathcal{F}_N(P)$-problem) for small values of N.

7. The Method

Our aim is now to describe a simple procedure of finding an approximate solution to the $\mathcal{F}_N(P)$-problem when N is small and the parameter-set P satisfies the following condition:

(7.1) *Every parameter-vector $p = (\mu, \sigma)$ in P has a dispersion coefficient σ such that $\sigma \geqslant \sigma_0$ where σ_0 is not too small (preferably $\geqslant 1.5\ \mu m$).*

The motive for condition (7.1) is that it is a guarantee that the solution is physically realistic.

The procedure consists of two phases.

PHASE I. *Step* 1. Specify the parameter-set P.

Step 2. Determine a *net* $\hat{P}$ in P; i.e., determine a finite subset of P such that each element in P is close to some element in $\hat{P}$.

Step 3. Determine the base-functions $\{h_j\}_1^N$ associated with the net $\hat{P}$. ($N =$ the number of elements in $\hat{P}$.)

Step 4. Determine the base-matrix D defined by (6.3). (A quadrature formula is needed in order to determine the elements of D.)

Step 5. Specify the distance-function dist$(\cdot, \cdot)$.

Step 6. Solve problem (6.4).

Step 7. Set

$$\hat{f} = \sum_1^N \hat{\alpha}_j h_j$$

where $\hat{\alpha} = (\hat{\alpha}_1, \alpha_2, ..., \hat{\alpha}_N)^*$ is the solution to problem (6.4).

This ends Phase I. What we have done so far is find a function $\hat{f}$ in $\mathcal{F}_N(\hat{P})$ which solves the $\mathcal{F}_N(\hat{P})$-problem. Moreover, if the net $\hat{P}$ is not too badly chosen, the function $\hat{f}$ will also give an *approximate* solution to the $\mathcal{F}_N(P)$-problem. Note also that we have not imposed any restriction on the number of elements in the net $\hat{P}$.

Before we describe Phase II we shall introduce some further concepts. Two components $\hat{\alpha}_i$ and $\hat{\alpha}_j$ of the solution vector $\hat{\alpha}$ are called *close*, if the corresponding parameter-vectors (μ_i, σ_i) and (μ_j, σ_j) are close. A component $\hat{\alpha}_i$ of $\hat{\alpha}$ is called *active* if it is non-zero. The vector $\hat{\alpha}$ is called *singular* if merely a small portion of the components are active. An active component is called *isolated* if there is no other active component close to it. Two active components are said to belong to the same *cluster* if they are close. Every cluster is called a *mode*. Every isolated component is also called a *mode*.

We shall now state the fact that makes the method work so well.

EXPERIMENTAL FACT. *In all cases with real data so far encountered, the vector $\hat{\alpha}$ has been singular as soon as the number of elements in the net $\hat{P}$ has been chosen*

larger than 12. *Moreover if the lower bound for the dispersion coefficients in the net* $\hat{P}$ *is not chosen unrealistically small* ($\geqslant 1.5$), *then the vector* $\hat{\alpha}$ *has had at most three modes.*

We shall now present the second phase of the procedure.

PHASE II. *Step* 1. Check whether the vector $\hat{\alpha}$ has at most three modes. If not, go back to Phase I and start over with a larger value on the lower bound σ_0.

Step 2. Check whether all non-negative components of $\hat{\alpha}$ are isolated. If this is the case stop the procedure and take $\hat{f}$ as the solution to the problem.

Step 3. (At this point we know that $\hat{\alpha}$ has at most three modes and that at least one mode is a cluster.) Determine a new parameter-set $\tilde{P}$ with the same number of elements as $\hat{\alpha}$ has modes. Each parameter-vector in this new parameter-set shall be a weighted average of parameter-vectors corresponding to components belonging to the same mode.

Step 4. Determine the base-functions $\{\tilde{h}_j\}_1^{N'}$ associated with the parameter-set $\tilde{P}$. Here N' denotes the number of parameter-vectors in $\tilde{P}$ (which is equal to the number of modes of $\hat{\alpha}$).

Step 5. Determine the base-matrix $\tilde{D}$ associated with $\tilde{P}$. (See (6.3).)

Step 6. Find a vector $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_{N'})^*$ which solves problem (6.4) with D replaced by $\tilde{D}$.

Step 7. Set

$$\tilde{f} = \sum_1^{N'} \tilde{\alpha}_j \tilde{h}_j.$$

The procedure is completed!

8. SOME REMARKS

In this section we shall make some remarks about the method. Further comments will be made in the next section where we present some examples.

Remark 8.1. The basic idea in the procedure is taken from the paper [1] by Gustafson. In this paper the author describes a method of fitting a sum of exponentials to measured data.

Remark 8.2. At step 4 of Phase I one has to specify the distance-function dist($\cdot, \cdot$). As has been pointed out there are several possibilities. However, it seems that distance-functions depending on the absolute values of the errors are better than those that depend on the squares of the errors. (That is, distance-functions defined by (3.3) or (3.5) are better than distance-functions defined by (3.4) or (3.6).) There are two reasons for this. The first reason is that distance-functions depending on absolute

values lead to LP-problems, whereas distance-functions depending on squares lead to QP-problems; and LP-problems are almost always easier to solve than QP-problems. The second reason is that distance-functions depending on the absolute values of the errors are more *robust* in the sence that they are less sensitive to systematic errors. For more information concerning the difference between distance-functions depending on absolute values and distance-functions depending on squares see, for example, the paper [7] by Narula and Wellington and papers quoted therein.

Remark 8.3. One difficulty in the procedure is finding the average in step 3 of Phase II. The simplest case to handle is the one where all components belonging to the same mode are associated with base-functions with the same value on the dispersion-parameter. In this case we can do as follows. Let $i_{1,k}$, $i_{2,k}$,..., $i_{m,k}$ denote the indices of the components belonging to the kth mode. (Usually the number of components belonging to a mode is less than or equal to 2.) Let $\mu_{i_1}^{(k)}$, $\mu_{i_2}^{(k)}$,..., $\mu_{i_m}^{(k)}$ denote the localization parameters for the base-functions with indices $i_{j,k}$, $j = 1, 2,..., m$, and let $\sigma^{(k)}$ denote the common value for the dispersion-parameter in the kth mode. Now define $\tilde{\mu}^{(k)}$ by

$$\log(\tilde{\mu}^{(k)}) = \left(\sum_{j=1}^{m} \hat{\alpha}_{i_{j,k}} \log(\mu_{i_j}^{(k)})\right) \Big/ \left(\sum_{j=1}^{m} \hat{\alpha}_{i_{j,k}}\right) \tag{8.1}$$

and define $\tilde{P}$ by

$$\tilde{P} = \{(\tilde{\mu}^{(k)}, \sigma^{(k)}) : k = 1, 2,..., N'\}$$

where N' is the number of modes of $\hat{\alpha}$.

Remark 8.4. The solution $\tilde{f}$ obtained after Phase II usually gives a slightly larger error than the solution obtained after Phase I. The advantage with $\tilde{f}$ compared to $\hat{f}$ is that it is determined by a smaller number of parameters and hence has a simpler analytic form. If one is primarily interested in drawing a graph representing the aerosol size distribution one can use $\hat{f}$ just as well as $\tilde{f}$.

Remark 8.5. In step 2 of Phase I one has to determine a net $\hat{P}$ within P. As a general rule one should take a fairly *coarse* net. The calculations become much quicker and one loses surprisingly little in accuracy.

Remark 8.6. As has been pointed out, it is not always very easy to determine the parameter set $\tilde{P}$ in Phase II. In order to facilitate the path from the parameter set $\hat{P}$ to the parameter set $\tilde{P}$ one can insert an intermediate phase. In this phase one introduces a new parameter set P^* built up by (1) the *active* parameter-vectors in $\hat{P}$, i.e., those parameter-vectors in $\hat{P}$ which correspond to active components of $\hat{\alpha}$; and (2) some new parameter-vectors in the neighbourhood of these. One then uses P^* to determine a new set of base-functions and a new base-matrix after which one solves problem (6.4) with this new base-matrix. Since all the active parameters of $\hat{P}$ also belong to P^*, the solution obtained in this intermediate step will of course have an error which is smaller than or equal to the error of the solution obtained after Phase I.

Remark 8.7. In principle the procedure works just as well whatever parametrizable family of non-negative functions one uses for base-functions; hence instead of having log-normal functions as base-functions, one could use, e.g., "rectangular" functions or gamma-functions. However, for the real data so far encountered, it seems that log-normal functions give the best fit with the least number of functions.

Remark 8.8. Instead of looking for a function describing the number distribution one might be interested in finding functions which describe the surface area distribution or the mass distribution of an aerosol. Now, if one already has a function $f(r)$ describing the number distribution, it is easy to find functions describing the surface area distribution and the mass distribution. However, these functions will in general not be linear combinations of log-normal functions even if $f(r)$ is. If one wants to express the surface area distribution and the mass distribution as a linear combination of log-normal functions one can use the method described in Section 7 with the kernel K divided by πr^2 and $4\pi r^3/3$, respectively.

9. Some Examples with Real Data

data all of which can be found in the literature.

Our first example is taken from a paper by King *et al.* [5], in which the authors present measurements taken on different occasions. We shall use the data collected November 20, 1975. By estimating the coordinates of the black dots on the left-hand side of Fig. 6 in their paper, we obtain the input data given in Table I.

The first step in Phase I of the procedure is to determine a parameter-set. The most important part of this step is to determine a lower bound σ_0 for the dispersion parameter. In this first example we shall take a fairly large value on this parameter, namely, $\sigma_0 = 2.0\,\mu$m. We also have to determine lower and upper bounds for the

TABLE I

Measurements Obtained by King *et al.* [5], November 20, 1975

	λ-Value							
	0.44	0.52	0.61	0.69	0.71	0.79	0.87	1.07
Measurement	0.064	0.044	0.048	0.046	0.045	0.046	0.045	0.048
Refractive index	1.45 – 0.0i							

localization parameter. In this example we let these values be equal to 0.003 and 1.2 μm. We now define the set P by

$$P = \{(\mu, \sigma) : 0.003 \leqslant \mu \leqslant 1.2,\ \sigma = 2.0\}.$$

Note that we have only allowed one value for the dispersion parameter σ. The reason for this will be given later.

The next step in the procedure is to determine a net $\hat{P}$ within P. Before we do this we shall introduce a notation. For any two positive real numbers a and b, $0 < a < b < \infty$, any natural number n, and any real number s, the set defined by

$$\{(\mu_i, \sigma) : \mu_i = a(b/a)^{1/n},\ i = 0, 1, ..., n, \sigma = s\}$$

will be denoted

$$G\{[a, b], N = n, \sigma = s\}.$$

Observe that $\mu_0 = a$, $\mu_n = b$ and that the μ_i's form a geometric series. Therefore $\{\mu_0, \mu_1, ..., \mu_n\}$ determines a partition of the interval $[a, b]$ with $n + 1$ division-points equi-distributed along a logarithmic scale.

By using this notation, we now define the set $\hat{P}$ by

$$\hat{P} = G\{[0.003, 1.2], N = 32, \sigma = 2.0\}. \tag{9.1}$$

The ratio between two consecutive μ-values in $\hat{P}$ will in this case be equal to

$$(1.2/0.003)^{1/32} = (400)^{1/32} \approx 1.206.$$

The third step in the procedure is simply writing down the set of base-functions associated with the parameter-set $\hat{P}$. In the fourth step we determine the base-matrix D. In order to do this, we have to specify the refractive index for the different wavelengths, since they are needed when specifying the kernel $K(\lambda, r)$. Following King *et al.* [5] we choose $m_\lambda = 1.45$ for all eight λ-values. The dimension of the matrix D will this time be equal to 8×33, since the number of observations is equal to 8 and the number of base-functions is equal to 33.

In the fifth step we shall determine a distance-function. In this example, as well as in all subsequent ones, we shall use the *mean of the absolute values of the relative differences* defined by

$$\text{dist}(x, y) = M^{-1} \left(\sum_{i=1}^{M} \left| \frac{x_i - y_i}{y_i} \right| \right) \tag{9.2}$$

as our distance-function. (This is a special case of (3.5).) The advantage of this choice of distance-function is that it can be used as a measure of the relative error between the theoretical observation vector and the measurement vector. For this reason we shall call dist(Kf, b) the *relative error* when dist($\cdot\,, \cdot$) is defined by (9.2).

In the sixth step we shall solve problem (6.4). Since we have defined the distance-function by (9.2), problem (6.4) can easily be transformed into an LP-problem. (Compare, e.g., Murty [6, p. 15].) This LP-problem is then solved by using the IMSL-routine ZX3LP. (We shall always use this routine when solving an LP-problem.) In this case it turns out that only two out of 33 components are different from zero, namely, the 6th and the 29th. The values for these two components are:

$$\hat{a}_6 = 1{,}170{,}000, \qquad \hat{a}_{29} = 6.29. \tag{9.3}$$

In the last step of Phase I we form the sum

$$\hat{f} = \sum_{i=0}^{32} \hat{a}_i \mathrm{LN}(\mu_i, \sigma_0)$$

which in this case reduces to

$$\hat{f} = 1{,}170{,}000 \cdot \mathrm{LN}(0.0077,2.0) + 6.29 \cdot \mathrm{LN}(0.567,2.0) \tag{9.4}$$

since $\mu_6 = 0.0077$, $\mu_{29} = 0.567$, $\sigma = 2.0$, and $\hat{a}_6$ and $\hat{a}_{29}$ satisfy (9.3).

This completes Phase I. Now since the vector $\hat{a}$ has only two active components and both these are isolated, we do not make any further calculations but take (9.4) as our solution.

Before we write down the relative error and the theoretical observation vector induced by the function (9.4) we shall present a solution corresponding to a smaller value for the dispersion parameter. Thus, let

$$\hat{P} = G\{[0.003,1.2], N = 32, \sigma = 1.5\}$$

be the net in step 2 of the procedure. By using the same distance-function as before we now obtain a solution

$$\tilde{f} = 84{,}600 \cdot \mathrm{LN}(0.028,1.5) + 1.55 \cdot LN(0.390,1.5) + 5.07 \cdot \mathrm{LN}(0.844,1.5). \tag{9.5}$$

In this case the solution vector $\hat{a}$ obtained at step 6 of Phase I had three modes, one of which was a cluster. We therefore had to carry out all the steps in Phase II. In particular, at step 3 of Phase II we had to choose a formula when determining the parameter-set $\tilde{P}$. The formula we used here, and will use for all other examples in this paper, is formula (8.1).

In Fig. 2, we have depicted the graphs of functions (9.4) and (9.5). There is a substantial difference between the two graphs for small values of r, but this is not contradictory, since the effect on the extinction at wavelengths larger than 0.44 μm is very small for particles less than 0.02 μm.

In Table II, we have presented the theoretical observations and the relative errors for functions (9.4) and (9.5). In the first row we have repeated the measurements

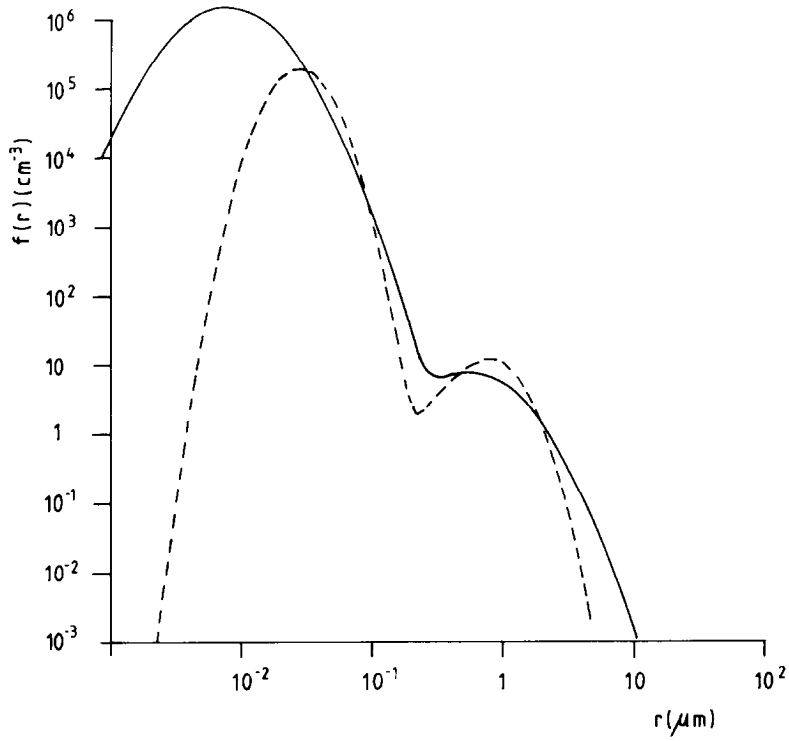

FIG. 2. Calculated aerosol size distributions for the data presented in Table I. The solid curve corresponds to a dispersion-coefficient equal to 2.0; the dashed curve corresponds to a dispersion-coefficient equal to 1.5.

given in Table I. As could be expected the relative error is smaller for the function with the dispersion-coefficient = 1.5.

Before we continue with our next example, we shall compare the solutions obtained

TABLE II

Theoretical Observations and Relative Errors for the Solutions (9.4) and (9.5)

	Measurement								
	0.064	0.044	0.048	0.046	0.045	0.046	0.045	0.048	Relative error
Theoretical observation									
$\sigma = 2.0$	0.0618	0.0526	0.0480	0.0459	0.0456	0.0451	0.0450	0.0455	0.0393
$\sigma = 1.5$	0.0640	0.0532	0.0480	0.0460	0.0450	0.0455	0.0459	0.0480	0.0321

above with some solutions obtained when solving problem (4.3) of Section 4. We shall consider three partitions, namely,

$$I_1 = G\{[0.1, 4], N = 100\}$$
$$I_2 = G\{[0.1, 4], N = 4\}$$
$$I_3 = G\{[0.1, 4], N = 3\}.$$

Here $G\{[a, b], N = n\}$ denotes the partition

$$r_i = a(b/a)^{i/n}, \qquad i = 0, 1, \ldots, n.$$

We use the same distance-function as before, namely, (9.2).

In Fig. 3, we have depicted the solutions corresponding to these three choices of partitions. In Table III we present the theoretical observations and the relative errors for these solutions.

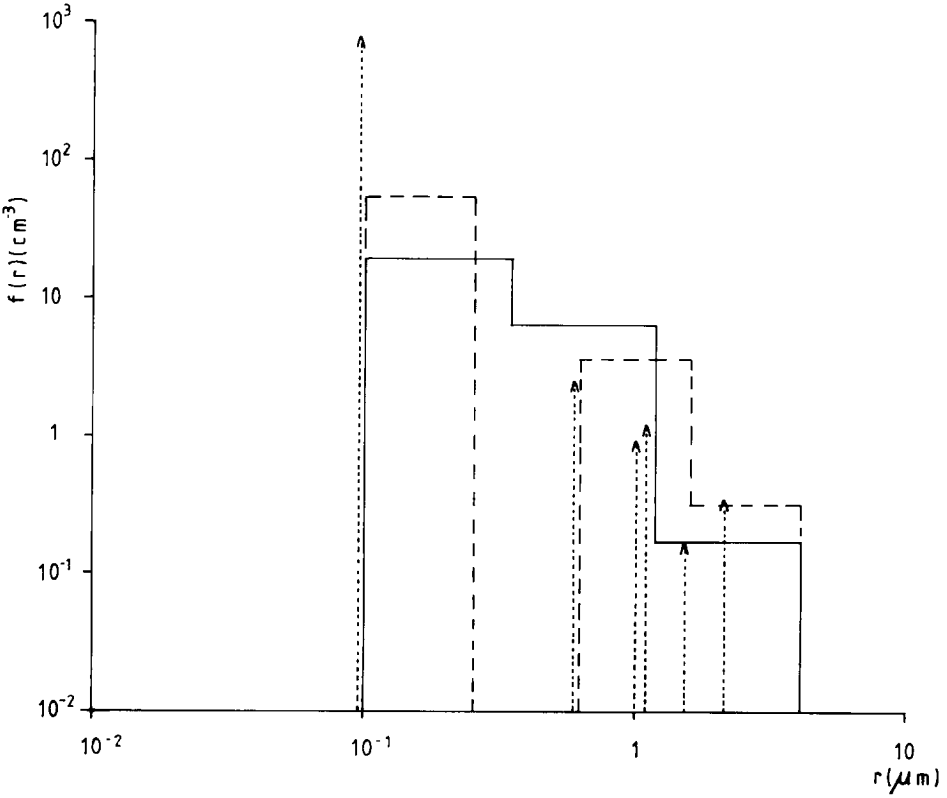

FIG. 3. Aerosol size distributions for the data in Table I obtained by solving problem (4.3). The solution corresponding to partition I_3 is represented by the solid line, the solution corresponding to partition I_2 is represented by the dashed lines, and the solution corresponding to partition I_1 is represented by the dotted lines. (Each dotted line shall be interpreted as a very thin rectangle.)

TABLE III

Theoretical Observations and Relative Errors for the Solutions Depicted in Fig. 3[a]

	Measurement								
	0.064	0.044	0.048	0.046	0.045	0.046	0.045	0.048	Relative error
Theoretical observation									
(I_1)	0.0640	0.0468	0.0480	0.0460	0.0458	0.0460	0.0450	0.0480	0.010
(I_2)	0.0540	0.0512	0.0480	0.0468	0.0458	0.0460	0.0461	0.0455	0.052
(I_3)	0.0457	0.0457	0.0460	0.0460	0.0458	0.0450	0.0450	0.0480	0.050

[a] Row 2 corresponds to partition I_1, row 3 to partition I_2, and row 4 to partition I_3.

Of all the five solutions we have obtained, the one with the smallest relative error is the one obtained when solving problem (4.3) with partition I_1. However, as we can see from Fig. 3, this solution is physically completely unrealistic. If we compare the relative errors for the solutions obtained when solving problem (4.3) using partitions I_2 and I_3 with the solutions obtained when using our method, we observe that the relative errors for the latter are substantially smaller. Note, also that in order to avoid a solution with gaps of zero particles, we need to choose a very coarse partition when using the method described in Section 4.

Our next example is taken from a paper by von Hoyningen–Huene [3]. By estimating the coordinates for 20 of the small circles in Fig. 2a of his paper, we obtain the input data given in Table IV. Now, by assuming a refractive index equal to 1.50 for all wavelengths and choosing the parameter-set

$$\hat{P} = G\{[0.01, 2], N = 33, \sigma = 1.8\}$$

we obtain a solution

$$\tilde{f} = \tilde{\alpha}_1 \mathrm{LN}(\tilde{\mu}_1, \sigma) + \alpha_2 \mathrm{LN}(\tilde{\mu}_2, \sigma)$$

where $\tilde{\mu}_1 = 0.06375$, $\tilde{\mu}_2 = 0.7406$, $\sigma = 1.8$, $\tilde{\alpha}_1 = 31{,}485$ and $\tilde{\alpha}_2 = 44.116$. The relative error is this time equal to 0.038. We omit a figure showing the graph of the solution as well as a table with the theoretical observations. We only mention the fact that the solution we obtain looks very similar to the solution obtained by von Hoyningen–Huene. (See Fig. 2b in [3].)

Our third example is taken from the paper [2] by Hågård *et al.* By estimating the components of the circles of the D-curve in Fig. 8 of their paper we obtain the input data given in Table V. The refractive indices for the different wavelengths are calculated by using a formula developed by B. Nilsson. This formula is based on

TABLE IV

Measurements Obtained by von Hoyningen–Huene [3]

	λ-Value									
	0.34	0.36	0.39	0.43	0.46	0.50	0.53	0.56	0.62	0.65
Measurement	1.84	2.00	2.01	1.87	1.75	1.51	1.47	1.40	1.20	1.12
Refractive index	1.50 – 0.0i									
	λ-Value									
	0.68	0.71	0.75	0.80	0.88	0.93	0.97	1.02	1.08	1.11
Measurement	1.04	1.02	0.99	0.93	0.80	0.75	0.78	0.66	0.67	0.69
Refractive index	1.50 – 0.0i									

results presented in [8]; its main meteorological variable is the relative humidity. For the data obtained by Hågård *et al.* the formula yields the indices given in Table VI.

If one compares the data in this example with the data from the two previous examples the main difference is that in this example the measurements are taken from a much broader spectrum of wavelengths.

By choosing the net $\hat{P}$ in the second step of the procedure equal to

$$G\{[0.01,10], N = 33, \sigma = 2.0\}$$

and using the same distance-function as before, we obtain a solution

$$\tilde{f} = 255{,}000 \cdot \mathrm{LN}(0.024, 2.0) + 376 \cdot \mathrm{LN}(0.26, 2.0) + 1.93 \cdot \mathrm{LN}(4.22, 2.0). \tag{9.6}$$

TABLE V

Measurements Obtained by Hågård *et al.* [2]

	λ-Values										
	0.55	0.58	0.70	0.80	0.90	1.06	1.7	2.2	3.8	4.7	10.8
Measurement	1.5	1.4	1.35	1.25	1.2	1.2	1.0	0.85	0.75	0.80	0.75
Relative humidity	70%										

TABLE VI

Refractive Indices for Different Wavelengths Determined by a Formula du to B. Nilsson[a]

	λ-Value						
	0.55	0.58	0.70	0.80	0.90	1.06	1.70
Refractive index	1.44	1.44	1.44	1.44	1.43	1.43	1.39
	λ-Value						
	2.2	3.8	4.7	10.7			
Refractive index	1.33	1.38	1.37	1.45-0.1i			

[a] See [8].

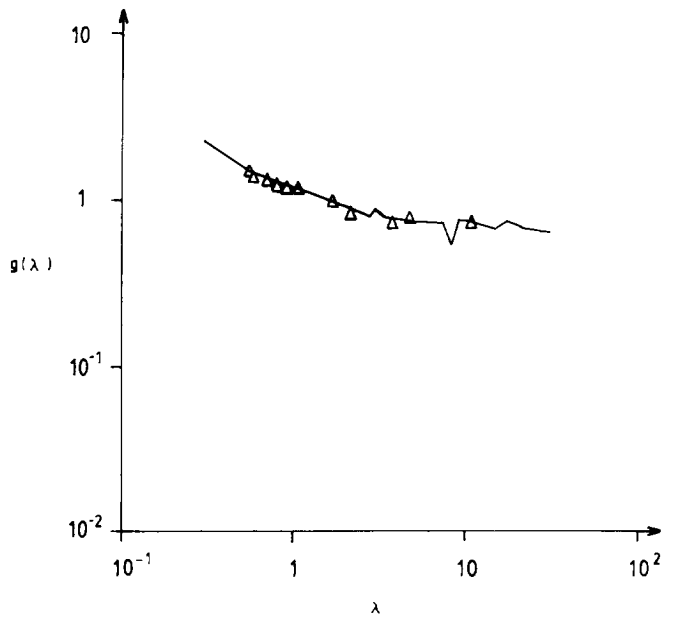

FIG. 4. Extinction-curve for the aerosol size distribution given by (9.6). The triangles represent the measurements given in Table V.

The relative error is this time equal to 0.0213, which is somewhat better than the relative errors for the two previous examples.

In Fig. 4, we have depicted the function $g(\lambda)$ defined by

$$g(\lambda) = \int K(\lambda, r)\tilde{f}(r)\, dr$$

where $\tilde{f}(r)$ is the function defined by (9.6). In this figure the triangles represent the measurements given in Table V.

10. Error and Sensitivity Analysis

In all examples presented in the previous section we have chosen a net $\hat{P}$ of the following form:

$$\hat{P} = G\{[a, b],\ N = n, \sigma = s\}.$$

This means that we have not let the dispersion parameter σ vary, but only the localization parameter μ. The reason for this is simply that experiments have shown that the method always picks a solution with dispersion coefficients as small as possible.

What value one shall choose for the dispersion parameter depends on one's purpose and on how much information one has about the aerosol. A value equal to 2.0 seems to be a reasonable initial choice in most cases, although one obtains a better fit the smaller the value one chooses. But the increase in fitness is often very small. For example, if we compare the theoretical observations induced by the function (9.4) (corresponding to a dispersion coefficient equal to 2.0) with the theoretical observations induced by the function (9.5) (corresponding to a dispersion coefficient equal to 1.5) we find that the distance between these two theoretical observations is equal to 0.017 when the distance is calculated by formula (9.2). This means that the mean value of the absolute values of the relative differences between the components of the two theoretical observation vectors is less than 2%.

Remark 10.1. One way of obtaining solutions, which do not necessarily only use functions with dispersion-coefficients equal to the smallest one allowed, is to introduce an extra term

$$\varepsilon \sum_{1}^{N} \gamma_i \alpha_i$$

into the object function of the minimization problem. First choosing ε very small and then choosing the γ's in such a way that γ_i becomes larger the smaller the corresponding dispersion coefficient is, one can obtain good solutions with dispersion coefficients larger than the smallest one allowed.

In order to illustrate the sensitivity of the solution due to the measurements, we

shall give two examples showing what can happen when one measurement is omitted. Let us first omit the first measurement in the data obtained by King *et al.* presented in the previous section. Our λ-values will then be equal to 0.52, 0.61, 0.69, 0.71, 0.78, 0.87, 1.07 and our measurements will be equal to 0.044, 0.048, 0.046, 0.045, 0.046, 0.045, 0.048. If we take

$$\hat{P} = G\{[0.1, 2.0], N = 20, \sigma = 2.0\},$$

again take $m_\lambda = 1.45 - 0.0i$, and define the distance-function by (9.2), we obtain as our solution

$$\tilde{f} = 4.0\ \mathrm{LN}(0.582, 2.0) + 0.39\ \mathrm{LN}(1.71, 2.0). \tag{10.1}$$

In this case the relative error becomes equal to 0.0183, which is less than half the value 0.0393 which we obtained when the first measurement was not omitted.

The difference between the function (9.4) and the function (10.1) can be observed from Fig. 5, in which both functions are depicted.

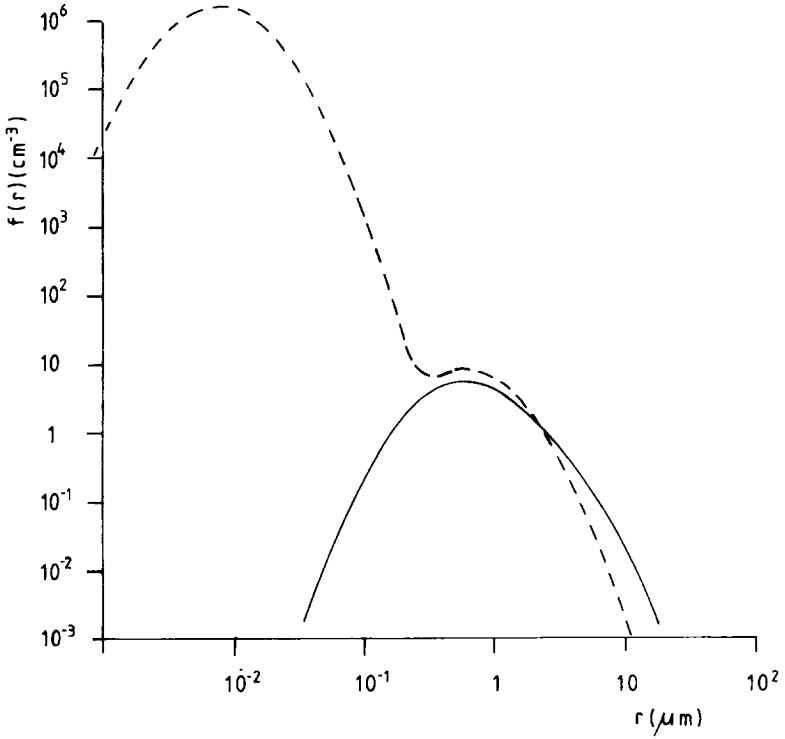

Fig. 5. Calculated aerosol size distributions for the data presented in Table I. The dashed curve corresponds to a solution obtained when all measurements are taken into account, the solid curve to a solution obtained when the first measurement is omitted. Both curves have dispersion-coefficients equal to 2.0.

If instead of omitting the first measurement we had omitted any other measurement there would have been very little change in the solution. Nor would there have been any substantial change in the relative error. Since the relative error *did* decrease substantially when the first measurement was left out, there is a possibility that the first measurement is less precise than the others.

Let us also see what happens if we omit the *last* measurement in the data obtained by Hågård *et al.* presented in the previous section. Since this measurement is taken at wavelength $\lambda = 10.7$ whereas the previous one is taken at wavelength $\lambda = 4.7$, it is not very surprising if the solution depends on this measurement.

The result is as follows. If we take

$$\hat{P} = G\{[0.01,\ 10.0],\ N = 33,\ \sigma = 2.0\},$$

use the same refractive indices as before and the same distance-function, we obtain a solution

$$\tilde{f} = 256{,}000 \cdot \mathrm{LN}(0.024, 2.0) + 362 \cdot \mathrm{LN}(0.27, 2.0) + 1.86 \cdot \mathrm{LN}(4.22, 2.0). \tag{10.2}$$

If one draws a graph of the function (10.2) and compares it with a graph of function (9.6) one will notice that function (10.2) is located more to the left.

Regarding the relative error, it is this time equal to 0.0234, which is even slightly larger than 0.0213, which was the relative error obtained when all measurements were taken into account. This means that there is no reason to believe that the last measurement is in any way less precise than the others.

We shall end our illustration of the method by inverting a set of synthetic data. The data we shall use will essentially be the same as used by Wolfson *et al.* in [15].

We first assume that the aerosol size distribution is given by

$$f(r) = Cr^6 \exp(-1.5r)$$

where the constant C is such that the integral of f is equal to one. We then calculate

$$\int K(\lambda, r) f(r)\, dr$$

for eight wavelengths, namely, $\lambda = 0.45,\ 1.19,\ 1.65,\ 2.25,\ 3.90,\ 6.05,\ 10.0$, and 16.5 using a refractive index equal to $1.33 - i0.0$; we obtain the following "measurements": 160, 166, 170, 176, 210, 250, 185, 86.

We shall now apply our method to this set of data. By taking

$$\hat{P} = G\{[0.8,\ 10.0],\ N = 20,\ \sigma = 2.0\},$$

choosing a refractive index $m_\lambda = 1.33 - i0.0$, defining the distance-function by (9.2), and using the IMSL-routine ZX3LP we obtain a solution

$$\tilde{f} = \tilde{\alpha}_1 \mathrm{LN}(\mu_1, \sigma)$$

where $\tilde{\alpha}_1 = 2.516$, $\mu_1 = 1.92$, and $\sigma = 2.0$. The relative error becomes equal to 0.0784 and the theoretical observations become equal to 157, 168, 176, 186, 199, 189, 144, 86. Note the relatively large difference between the given measurements and the theoretical observations at wavelengths 6.05 and 10.0.

If we choose a parameter-set $\hat{P}$ with smaller values on the dispersion-coefficient we will get a better fit. For example, if we take

$$\hat{P} = G\{[2.0,\ 10.0], N = 20, \sigma = 1.5\}$$

we obtain a solution

$$\tilde{f} = \tilde{\alpha}_1 \mathrm{LN}(\mu_1, \sigma)$$

where $\tilde{\alpha}_1 = 1.1277$, $\mu_1 = 3.98$, and $\sigma = 1.5$. The relative error is this time equal to 0.018 (which must be considered as a small relative error) and the theoretical observations are equal to 160, 166, 170, 177, 216, 236, 175, 86.

11. Computer Times

The method described above has been implemented on a DEC–10 computer. The calculation times are of course dependent on how well we want to represent the functions involved in the procedure. Below, we shall give some examples of computer times when all the functions involved are represented by a vector with 100 components.

(1) *Calculation of the set of base-functions.* The CPU-time to calculate a set of N base-functions is approximately equal to 1 sec when $N = 20$, and 4 sec when $N = 130$.

(2) *Calculation of the kernel.* The CPU-time to determine the kernel $K(\lambda, r)$ for 8 λ-values and 100 r-values is approximately 2.5 sec.

(3) *Calculation of the base-matrix.* The CPU-time to determine a base-matrix of size $M \times N$ is approximately 3 sec when $M = 8$ and $N = 130$, and approximately 1 sec when $M = 8$ and $N = 20$.

(4) *The minimization.* The CPU-time to solve the minimization-problem (6.4) depends mainly on the size of the base-matrix D but also on the choice of distance-function and which algorithm is used. If we choose a distance-function of the form (3.1) or (3.3) we are led to an LP-problem. If we use the IMSL-routine ZX3LP then the CPU-time depends mainly on the number of rows in the base-matrix. For example, if $M = 8$ and $N = 20$ then the CPU-time is approximately equal to 0.35 sec; if $M = 8$ and $N = 100$ the CPU-time is approximately 0.6 sec; if $M = 20$ and $N = 33$

the CPU-time is approximately 2.0 sec; and if $M = 20$ and $N = 100$ then the CPU-time is approximately equal to 3.5 secs. The IMSL-routine ZX4LP is even faster, particularly for larger problems. However, there is one important disadvantage with the ZX4LP-routine, namely, that it is less reliable then the ZX3LP-routine, which hardly ever fails. Especially when the base-matrix consists of columns which are almost identical, the ZX4LP-routine might not work. This situation occurs when we take a very fine net $\hat{P}$ at step 2 of Phase I in the procedure.

12. Conclusions

There are essentially three main advantages with the method described in this paper compared to most other methods.

(i) The method does not require a good initial guess.

(ii) The method is quite robust in the sense that there are seldom any numerical instabilities.

(iii) The method is quite fast and therefore can be used for large series of data.

Acknowledgments

This paper would not have been written had it not been for the support and help from my two colleagues Mr. B. Nilsson and Mr. A. Hågård. Their knowledge of aerosols and of the literature on inversion methods has been invaluable. I am also grateful to Dr S.-Å. Gustafson for drawing my attention to the basic idea in [1], on which the method of this paper is founded. Furthermore I want to thank Mrs. G. Fallgren and Mr. B. Nilsson (again) for helping me with the figures. Finally I want to thank Dr. T. Elfving and Mrs. K. Kocken for valuable advice on a preliminary version of this paper.

References

1. S.-Å. Gustafson, *Z. Angew. Math. Mech.* **61** (1981), T 284.
2. A. Hågård, B. Nilsson, H. Ottersten, and O. Steinvall, *Radio Sci.* **13** (1978), 277.
3. W. von Hoyningen-Huene, *Izv. Acad. Sci. USSR Atmospher. Ocean. Phys.* **2** (1979), 842.
4. M. King, *J. Atmospheric Sci.* **39** (1982), 1356.
5. M. King, D. Byrne, B. Herman, and J. Reagan, *J. Atmospheric Sci.* **35** (1978), 2153.
6. K. Murty, "Linear and Combinatorial Programming," Wiley, New York, 1976.
7. S. Narula and J. Wellington, Linear regression using multiple-criteria, *in* "Multiple Criteria Decision Making Theory and Application" (G. Fandel and T. Gal, Eds.), Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1980.
8. B. Nilsson, *Appl. Opt.* **18** (1979), 3457.
9. E. M. Patterson and D. A. Gillette, *J. Geophys. R.* **82** (1977), 2074.
10. R. Rizzi, R. Guzzi, and R. Legnani, *Appl. Opt.* **21** (1982), 1578.
11. B. Rust and W. Burrus, "Mathematical Programming and the Numerical Solution of Linear Equations," Americal Elsevier, New York, 1972.

12. S. TWOMEY, "Atmospheric Aerosols," Elsevier, Amsterdam, 1977.
13. S. TWOMEY, "Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements," Elsevier, Amsterdam, 1977.
14. K. WHITBY, *Atmospheric Environ.* **12** (1978), 135.
15. N. WOLFSON, J. H. JOSEPH, AND Y. MEKLER, *J. Appl. Meteor.* **18** (1979), 543.