

Using global optimization for a microparticle identification problem with noisy data

M. C. Bartholomew-Biggs

(M.Bartholomew-Biggs@herts.ac.uk)

Numerical Optimization Centre, University of Hertfordshire

Z.J. Ulanowski

(Z.Ulanowski@herts.ac.uk)

STRC, University of Hertfordshire

S. Zakovic

(S.Zakovic@ic.ac.uk)

Department of Computing, Imperial College, London

Abstract

We report some experience with optimization methods applied to an inverse light scattering problem for spherical, homogeneous particles. Such particles can be identified from experimental data using a least squares global optimization method. However, if there is significant noise in the data, the “best” solution may not correspond well to the “actual” particle. We suggest a way in which the original least squares solution may be improved by using a constrained optimization calculation which considers the position of peaks in the data. This approach is applied first to multi-angle data with varying amounts of artificially introduced noise and then to examples of single-particle experimental data patterns characterized by high noise levels.

1 Introduction

Substantial advances have been made over the last hundred or so years in the development of the theory of light scattering from particulate matter. Computation of the properties of scattered electromagnetic fields - the direct scattering problem - is now possible in many situations. Rigorous solutions exist for particle types such as homogeneous and inhomogeneous spheres, ellipsoids, cylinders with various cross-sections and generalized axisymmetric particles. Of far greater practical importance, however, is the determination of properties of scatterers from the knowledge of scattered fields - the inverse scattering problem. This arises in applications, ranging from astronomy and remote sensing, through aerosol and emulsion characterization, to non-destructive analysis of single particles and living cells [1, 2, 3, 4, 5, 6, 7, 8].

The inverse problem has proved to be much harder to solve, even for the simplest particle shapes. Lack of rigorous solutions has motivated the development of methods based on approximate models of scattering, for example, assuming that the particles are weak (Rayleigh-Debye) scatterers or that diffraction or even geometrical optics can adequately describe the interaction process [3, 5, 9, 10, 11, 12]. However, when such methods are inappropriate, empirical procedures have to be used which are based on generating solutions to the direct problem (after making assumptions concerning the shape, internal structure of the particle etc.) and matching these solutions to experimental data [1, 2, 4, 13, 14]. Combining empirical and analytical (eigenfunction) methods can provide particle size distributions as well as complex refractive indices under some constraining conditions [12]. Such procedures can be slow, difficult to implement rigorously and often require substantial computing resources. More rapid solutions can be obtained using neural network methods which take advantage of the capability of radial basis function neural networks to approximate multidimensional functions [15] and use one or two hidden-layer networks trained by back-propagation [7, 8]. An analytical method which directly yields particle radius but may eventually lead to a full inverse solution is based on expanding scattering data in terms of Legendre or Gegenbauer polynomials [16, 17]. Several approaches based on numerical optimization have also been reported [18, 19, 20].

Experimental data are inevitably distorted by the presence of noise and numerous sources of error, for example optical aberrations, nonlinearity of the detection system, multiple scattering or particle nonsphericity. All existing inversion algorithms are sen-

sitive to such distortion to a greater or lesser extent, which results in error [3, 11]. This problem is especially acute in, but not limited to, measurements on single particles. Therefore, the sensitivity of the inversion to error in input data should be properly considered for all new techniques. The present paper focuses on an approach to inverse light-scattering based on global optimization previously reported in [18, 19] and introduces a refinement procedure which can sometimes be used to compensate for the influence of noise in input data. This new approach is then applied both to multi-angle data with varying amounts of artificially introduced noise and also to representative examples of single-particle experimental data sets some of which are characterized by high noise levels.

In choosing to use a global optimization approach for particle identification, we are assuming that the inverse light scattering problem for homogeneous spheres has a unique solution. If this were not so and the error function had several global minima then there would still be ambiguity about the true properties of the scattering particle. Mireles [21] has shown that a unique solution does exist for the very closely related problem of infinite circular cylinders; and this, together with other more intuitive reasons, leads to a general consensus in the field of particle characterization that multi-angle light scattering data can give unique solutions to the homogeneous sphere problem as long as a sufficient number of measurements is available. Recent investigations confirm that this is the case [22]. On the other hand, the theoretical question of uniqueness of inverse scattering may not be very relevant to practical problems of particle identification in which there is the additional complication of accuracy and the discrete character of experimental measurements. These complications may arise, for instance, due to finite sampling density and aliasing, incompleteness of data, as well as the presence of noise. Further discussion of these issues is available elsewhere [15, 18, 19].

A good starting point for the development of methods for solving the inverse light scattering problem for small particles is the case of a homogeneous, isotropic, non-absorbing sphere. If a plane incident wave of known wavelength and state of polarization and a known medium surrounding the particle are assumed, the particle can be completely described using its radius r and refractive index n . In the scattering geometry considered in the present study (see Figure 1), the intensity of the light scattered by the particle is measured in one plane only and can, therefore, be described by a function of the scattering angle

$$I_1(\theta) = \varphi(\theta, r, n) \quad (1)$$

where θ is defined as the angle between the direction of propagation of the incident wave and the direction of observation. This arrangement leads to a one-dimensional scattering pattern which is representative of the properties of the particle and has been used as a basis for characterization of single particles and particle distributions in both routine and research applications [1, 2, 3, 4, 5, 7, 8, 12].

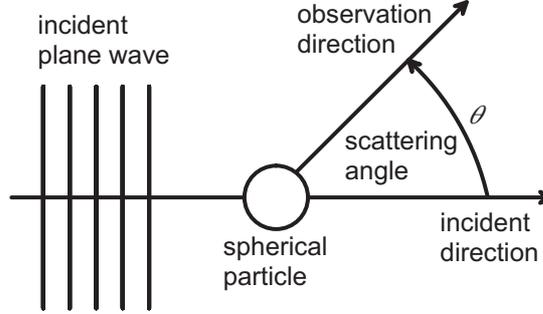


Figure 1: Scattering geometry

More practical detail about such experimental measurements can be found in [23]. A typical scattering pattern is shown in Figure 2.

The direct problem of computing a scattering pattern for a spherical particle can be solved using the series expansions of Lorenz-Mie theory [10]. We now give a brief outline of the Lorenz-Mie model of intensity to indicate the work involved in a typical function evaluation in the global optimization calculations discussed in subsequent sections. Suppose the incident light has intensity I_0 and wavelength *in vacuo* λ . Suppose also that the refractive index of the scattering medium is n_0 . Then the intensity of scattered light is given by

$$I_1 = \frac{I_0}{k^2 R^2} |S_1|^2$$

where R is the radial coordinate, $k = 2\pi n_0/\lambda$ and S_1 is defined by

$$S_1 = \sum_j \frac{2j+1}{j(j+1)} (a_j \pi_j + b_j \tau_j). \quad (2)$$

The values of π_j and τ_j depend on the scattering angle θ and are obtained from recurrence relations involving Legendre polynomials P_j . Specifically

$$\pi_j = (2j-1)P_j + \pi_{j-2} \quad \tau_j = j(j+1)P_j - \pi_j \cos \theta$$

with initial conditions $\pi_0 = 0$, $\pi_1 = 1$, $\tau_0 = 0$, $\tau_1 = \cos \theta$. The values of a_j , b_j depend on Bessel-Ricatti functions

$$\psi_j(x) = \sqrt{x} J_{j+1/2}(x) \quad \chi_j(x) = \sqrt{x} Y_{j+1/2}(x)$$

where J, Y respectively denote half-order Bessel functions of the first and second kinds.

If we now define the relative refractive index

$$n_r = \frac{n}{n_0}$$

and let $\xi_j(x)$ denote the complex function $\Psi_j(x) + i\chi_j(x)$ then

$$a_j = \frac{n_r \Psi_j(n_r x) \Psi_j'(x) - \Psi_j(x) \Psi_j'(n_r x)}{n_r \Psi_j(n_r x) \xi_j'(x) - \xi_j(x) \Psi_j'(n_r x)}$$

and

$$b_j = \frac{\Psi_j(n_r x) \Psi_j'(x) - n_r \Psi_j(x) \Psi_j'(n_r x)}{\Psi_j(n_r x) \xi_j'(x) - n_r \xi_j(x) \Psi_j'(n_r x)}.$$

The summation in (2) continues until the imaginary parts of a_j and b_j become sufficiently small.

2 Identification using least-squares data fitting

Suppose we have experimental measures of scattered light intensity

$$I_1(\theta_1), I_1(\theta_2), \dots, I_1(\theta_m)$$

and we wish to determine corresponding values for particle radius r and refractive index n . A standard approach involves finding r and n to minimize

$$E_1 = \sum_{i=1}^m (I_1(\theta_i) - k \cdot \varphi(\theta_i, r, n))^2 \quad (3)$$

where φ is the function appearing in the Lorenz-Mie model (1). The additional variable k appears because experimental measurements usually determine *relative* intensities at each θ_i and so a scaling is necessary to match the model values.

In practice, because the intensities vary widely in magnitude over the range $0^\circ \leq \theta_i \leq 180^\circ$, it may be advisable to consider an objective function of the form

$$E_2 = \sum_{i=1}^m (i(\theta_i) - \psi(\theta_i, r, n) - c)^2, \quad (4)$$

where i denotes $\log I_1$, ψ denotes $\log \varphi$ and $c = \log k$. Experimental results are often presented in this form in order to give increased weighting to scattering at large angles.

There are a number of optimization techniques which can be used to minimize (3) or (4). An obvious possibility is the Gauss-Newton method which is designed to deal with functions of the form

$$F(x) = \sum_{i=1}^m f_i(x)^2.$$

It uses the fact that

$$\nabla F(x) = 2J(x)^T f(x) \quad \text{and} \quad \nabla^2 F(x) = 2J(x)^T J(x) + \sum_{i=1}^m f_i(x) \nabla^2 f_i(x)$$

where $f(x) = (f_1(x), \dots, f_m(x))^T$ and $J(x)$ is the Jacobian matrix whose (i, j) -th element is $\partial f_i(x)/\partial x_j$. If the minimum value of F is near-zero and/or the subfunctions f_i are near-linear the second term in $\nabla^2 F(x)$ may be neglected and hence the iteration

$$x^{(k+1)} = x^{(k)} + d^{(k)} = x^{(k)} - J(x^{(k)})^T J(x^{(k)})^{-1} J(x^{(k)})^T f(x^{(k)}) \quad (5)$$

can be regarded as an approximate form of the Newton method for minimizing $F(x)$. Practical implementations of the Gauss-Newton algorithm include a line search so that $x^{(k+1)} = x^{(k)} + \alpha d^{(k)}$, where α is a scalar chosen to ensure $F(x^{(k+1)}) < F(x^{(k)})$.

The minimization of either E_1 (or E_2) can be regarded as a three-variable or a two-variable problem. In the case of E_1 we can write

$$\partial E_1 / \partial k = -2 \sum_{i=1}^m \varphi(\theta_i, r, n) \cdot (I_1(\theta_i) - k \cdot \varphi(\theta_i, r, n)), \quad (6)$$

and since $\partial E_1 / \partial k = 0$ at the minimum, we can obtain the optimal value of k in terms of the other two variables as

$$k = \frac{\sum_{i=1}^m (I_1(\theta_i) \cdot \varphi(\theta_i, r, n))}{\sum_{i=1}^m \varphi(\theta_i, r, n)^2}. \quad (7)$$

Similarly, the optimal value for c in (4) is

$$c = \frac{\sum_{i=1}^m (i(\theta_i) - \psi(\theta_i, r, n))}{m}. \quad (8)$$

We have discussed elsewhere [18, 19] the application of a Gauss-Newton method to the minimization of (3) and (4) in order to identify r and n from *perfect* scattering data – i.e. values of $i(\theta)$ which have been generated from the Lorenz-Mie model. In this relatively simple case (3) and (4) both have a global minimum of zero (and hence the Gauss-Newton method can be expected to work well). However, it turns out to be difficult to obtain the correct values of r and n because (3) and (4) have many local minima. For instance, if we use perfect data derived from the Lorenz-Mie model with $n = 1.525$, $r = 1.475$ then the function (4) has *six* local minima within the region $1.475 \leq n \leq 1.575$ and $1.375 \leq r \leq 1.575$. In such a case, the region of convergence for the Gauss-Newton method about the global solution is relatively small. In practice we may not have good estimates of the optimum values of n and r and so we will need to use a global rather than a local optimization technique.

2.1 Global optimization of (3) and (4)

One possibility, considered in [19], is to use the Gauss-Newton method within the framework of a multi-start approach due to Rinooy Kan & Timmer [24, 25]. Essentially, this uses many *local* optimizations from different starting points in order to seek *all* the minima within the region of interest. The Rinooy Kan & Timmer algorithm includes tests (based on cluster analysis) designed to avoid redundant local searches by rejecting starting points which are too close to each other or are near to areas where a local minimum has already been found. Termination occurs when the number of minima actually found is sufficiently close to a Bayesian estimate of the total number of minima.

In this paper we shall also use an alternative global optimization method DIRECT [26]. DIRECT (which is an acronym for DIviding RECTangles) is a non-gradient method which minimizes a function $F(x)$ in some chosen hyperbox $l_i \leq x_i \leq u_i$. It proceeds by systematic subdivision of this region into smaller hyperboxes. This is done quite efficiently by only subdividing boxes which pass a test for *potential optimality*.

DIRECT begins with a given hyperbox defined by its centre point, c_0 , the value of the objective function, $F_0 = F(c_0)$, and the n vector of displacements s_0 . These displacements are such that $l_i = c_{0i} - s_{0i}$ and $u_i = c_{0i} + s_{0i}$ for $i = 1, \dots, n$. This initial hyperbox is then systematically split into smaller ones, using the procedure **subdivide** described below. For each hyperbox, j , ($= 1, \dots, J$) we have a centre c_j (where the function value is F_j) and a vector of semi-sides s_j . Hyperboxes are grouped according to a size parameter δ_j , which is the distance from centre to any corner. We shall suppose that among the J hyperboxes there are only $K_J \leq J$ different size values.

When the procedure **subdivide** is applied to an existing hyperbox characterized by $(c_j, F_j, s_j, \delta_j)$ it only shrinks the *longest* edges. If there is a unique longest edge then DIRECT replaces the existing box j by three new ones, constructed by trisecting the appropriate side. If several edges of hyperbox j all have the same “longest” length, then the trisection process is repeated for each of them. (It should be noted, however, that the boxes created by establishing new centres parallel to the second and subsequent sides will be smaller than the boxes created by division along the first one. It is suggested in [26] that the *order* in which long edges are dealt with should be based on some exploratory function evaluations, with a view to enclosing the smallest new function values in the largest of the new hyperboxes.)

At each iteration of DIRECT, some of the current hyperboxes $j = 1, \dots, J$ are selected for further subdivision. The aim is to explore the whole region efficiently by only computing extra function values in regions which can be termed “potentially optimal”. Potentially optimal hyperboxes are chosen via the procedure **identify** given below. We note first of all, however, that we need only examine K_J of the hyperboxes – i.e. for each of the different δ_j -sized candidates we need only consider the one whose centre has the least function value.

We now explain how the procedure **identify** selects from the current set of hyperboxes those which are worth further exploration. Suppose first that Ω is a Lipschitz constant for the function F – i.e. that $\|\nabla F\| < \Omega$. Then a lower bound for F inside the hyperbox j is given by $\underline{F}_j = F_j - \Omega\delta_j$. Hence the most promising box would be the one for which \underline{F}_j is smallest. This argument, of course, assumes that a valid Lipschitz constant Ω is known – which will not usually be the case. The basis of DIRECT is therefore to consider whether there exists any Lipschitz constant such that box j *could* contain a lower function value than any other box. Thus, box j is more promising than box k if there exists a positive Ω such that

$$F_j - \Omega\delta_j < F_k - \Omega\delta_k.$$

We note that no such Ω exists if $\delta_j = \delta_k$ and $F_j \geq F_k$. Hence, as mentioned above, we only need to test the potential optimality of the box with size δ_j the smallest F value. If $\delta_j > \delta_k$ then box j can be potentially optimal only if

$$\Omega > \Omega_{min_k} = \frac{F_j - F_k}{\delta_j - \delta_k};$$

while if $\delta_j < \delta_k$ then box j the corresponding condition is

$$\Omega < \Omega_{max_k} = \frac{f_j - f_k}{\delta_j - \delta_k}.$$

We can calculate Ω_{min_k} or Ω_{max_k} for the smallest-valued hyperbox for each size δ_k ($\neq \delta_j$) and then set $\Omega_{min} = \max\{\Omega_{min_k}\}$; $\Omega_{max} = \min\{\Omega_{max_k}\}$. Box j can then only be potentially optimal if $\Omega_{max} > 0$ and $\Omega_{min} < \Omega_{max}$.

Even if there is a valid range $[\Omega_{min}, \Omega_{max}]$ we can apply a further filter to try and reduce the number of boxes to be subdivided. We only treat box j as potentially optimal if it might produce a worthwhile decrease in F_{min} , the least function value found so far. Thus we test

$$F_j - \Omega_{max}\delta_j < F_{min} - \varepsilon|F_{min}|$$

where ϵ is a user-specified parameter. If this inequality fails then box j is judged to be not worth further subdivision at the present time.

Successful experience with DIRECT on practical problems is reported in [20, 27].

2.2 The effect of noise in experimental data

We now consider how the global solution of (4) can be affected by the presence of noise in real-life experimental data. One obvious point is that the global solution for noisy data will not now be characterized by $E_2 = 0$. A consequence of this may be to cast doubt upon the connection between the global minimum of (4) and a true identification of the particle. Consider for instance the real-life scattering dataset **py12log** (Figure 2) which is discussed more fully in a later section.

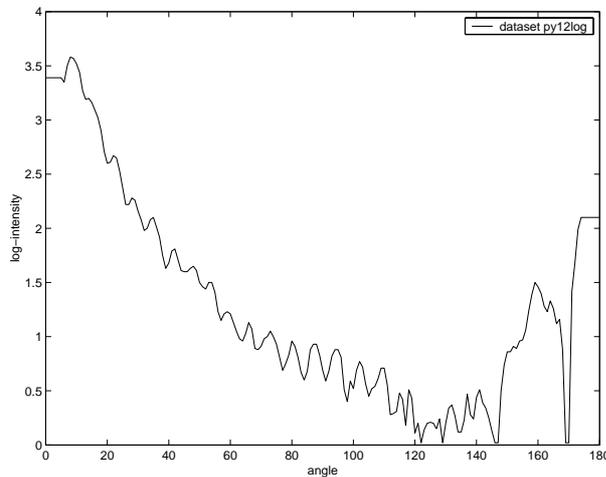


Figure 2: Scattering pattern **py12log**

Because this scattering pattern shows some truncation or flattening of the intensity measurements in the region of $\theta = 0^\circ$ and $\theta = 180^\circ$, the identification is based only on readings in the range $20^\circ \leq \theta \leq 160^\circ$. With this data, the function (4) turns out to have (at least) three minima all of which correspond quite well to prior knowledge about the particle which produced the pattern. These minima are:

$$n \approx 1.5032, r \approx 1.8014 \text{ giving } E_2 \approx 19.07$$

$$n \approx 1.4737, r \approx 1.8210 \text{ giving } E_2 \approx 19.62$$

$$n \approx 1.4909, r \approx 1.8219 \text{ giving } E_2 \approx 19.88$$

Here the global solution is not very much better than its two nearby competitors. In fact, the differences in the values of E_2 are probably small in relation to the experimental errors

which cause the minimum of (4) to be so much greater than zero. Therefore we might not be sure that the least value of E_2 defines the particle very precisely.

The error function (4) for the dataset **py12log** is visualised in Figures 3 and 4 below.

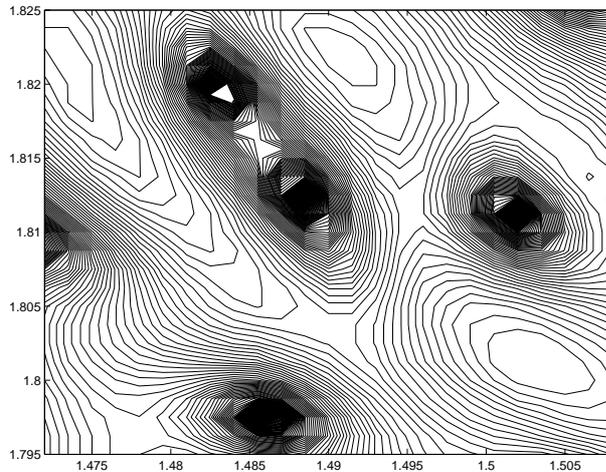


Figure 3: Contour plot of error function (4) for **py12log**

In the contour plot – which shows the local minima quite well – the dark areas represent sharp peaks. These can be seen more clearly in the surface plot in Figure 4.

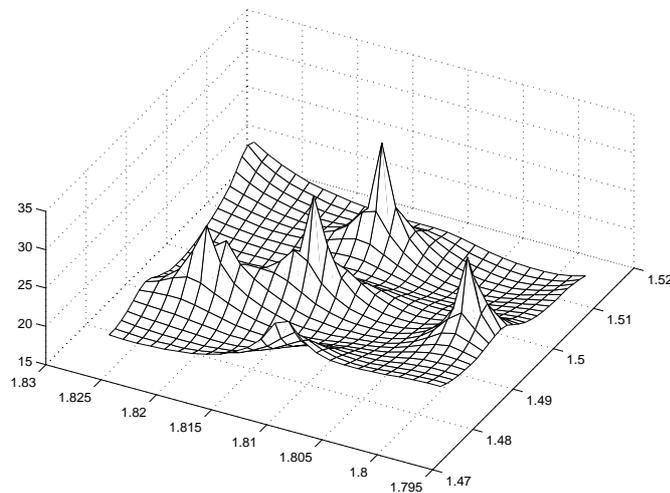


Figure 4: Surface plot of error function (4) for **py12log**

The dataset **py12log** provides an even more damaging piece of evidence about the dangers of blindly associating the global minimum of (4) with an identification of the “actual” particle. By enlarging the search region for the global optimization methods we find that there are two more optima of E_2 which are “better” than the ones quoted in the

previous paragraph. These are:

$$n \approx 1.6604, r \approx 0.5328 \text{ giving } E_2 \approx 16.1$$

$$n \approx 0.8951, r \approx 0.4352 \text{ giving } E_2 \approx 12.2.$$

However, neither of these “solutions” is physically acceptable. The first is inconsistent with what is known about the radius of the particle; and the second is completely outside the bounds of possibility set by the experimental conditions.

To look at the question of experimental noise in a more general way, suppose that the particle has radius and refractive index r^* and n^* but that there are errors ε_i (with zero mean and standard deviation σ_m) in the log-intensity measurements $i_i (= \psi_i + \varepsilon_i)$. If we calculate ψ_i using the correct value of $x^* (= (n^*, r^*, c))$, the expected value of E_2 is

$$E_2(x^*) = \sum_i \varepsilon_i^2 = m \cdot \sigma_m^2 \quad (9)$$

where m is the number of measurements. If we had an estimate of σ_m then we might use it to put bounds on the value of E_2 rather than trying to minimize it. (Indeed, if we make E_2 “too small” then we may be modelling the noise rather than the underlying intensity pattern.) We can deduce a plausible value for σ_m for a particular data set once we have found \hat{x} as the global optimum of E_2 . If we assume experimental errors are normally distributed with zero mean then, by (9), we can approximate the standard deviation of the errors by the root-mean-square value

$$\sigma_m \approx E_2^{rms}(\hat{x}) = \sqrt{(E_2(\hat{x})/m)}.$$

Hence we may wish to consider solutions $x = (n, r)$ such that

$$E_2^{rms}(\hat{x}) \leq E_2^{rms}(x) \leq (1 + \tau)E_2^{rms}(\hat{x}) \quad (10)$$

for some small positive τ . If we have *a priori* estimates \bar{x}_1 and \bar{x}_2 for the refractive index and radius then, as suggested by Dixon [28], we could consider minimizing

$$e_1(x) = \sum_{i=1}^2 (x_i - \bar{x}_i)^2 \quad (11)$$

subject to the constraints (10). This can be thought of as trying to find a better match to our expectations about the solution without causing too much of an increase in the least-squares error function. For certain particles a “good estimate” \bar{x} can sometimes be made – examples include microbial cells or microparticles produced in an industrial process. However it seems inadvisable to rely overmuch on the availability of prior information and in the next section we shall present a more general approach.

2.3 Using peak-matching in identification

If we do not have reliable initial estimates of n and r then we can use ideas which mimic a visual approach to particle identification. There are occasions when visual comparison of experimental and theoretical data – i.e., exploiting intrinsic feature-selection and data-processing capabilities of the human brain – can produce satisfactory results, even where least squares methods fail [2]. In particular, it has been found that *positions* of intensity peaks can be useful in particle identification procedures [2, 4, 14, 29]. As an example of this, consider Figures 5 and 6.

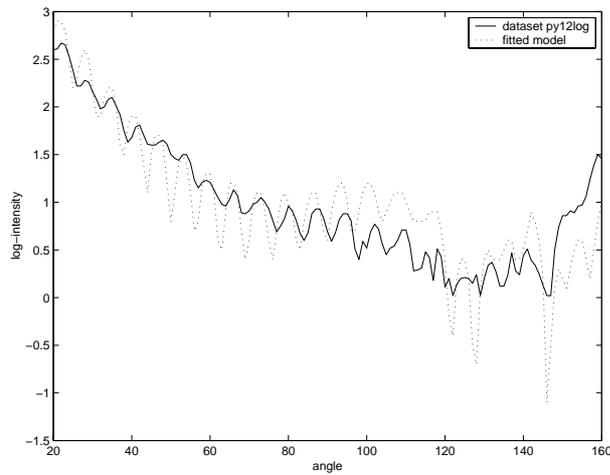


Figure 5: Theoretical and experimental scattering patterns for **py12log**

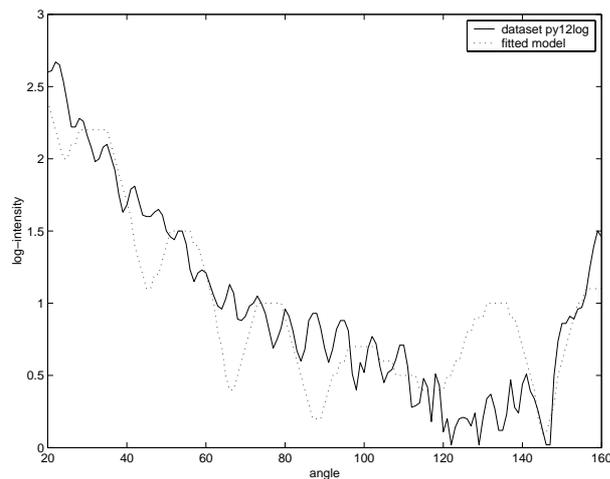


Figure 6: Theoretical and experimental scattering patterns for **py12log**

Figure 5 shows the scattering pattern **py12log** (solid line) together with the pattern

produced by the Lorenz-Mie model with $n = 1.5032, r = 1.8014$. Figure 6 gives a similar comparison between the dataset and theoretical model when $n = 1.6604, r = 0.5328$. Although the dotted line in Figure 6 is a better fit to the data in the simple least-squares sense, it is clear that it does not correspond at all well to the essential *shape* of the given pattern. Hence we now consider a way of including *peak-matching* in the solution of the inverse light-scattering problem.

For brevity we let i_k denote the data value $i(\theta_k) = \log I_1(\theta_k)$. We also let K be the set of indices of reference angles θ_k at which the given log-intensities satisfy

$$i_{k-2} < i_{k-1} < i_k \quad \text{and} \quad i_k > i_{k+1} > i_{k+2}.$$

These conditions suggest the data has a peak near θ_k . (In practice, for noisy data, this is more reliable than deducing peak positions simply on the basis of $i_k > i_{k-1}$ and $i_k > i_{k+1}$.) If $\delta\theta$ is the spacing between the θ -values in the data then we can estimate first and second derivatives by the second order formulae

$$i'_k = \frac{i_{k+1} - i_{k-1}}{2\delta\theta} \quad \text{and} \quad i''_k = \frac{i_{k+1} - 2i_k + i_{k-1}}{\delta\theta^2}.$$

More accurately we can use fourth-order formulae

$$i'_k = \frac{8(i_{k+1} - i_{k-1}) - (i_{k+2} - i_{k-2})}{12\delta\theta}$$

$$i''_k = \frac{16(i_{k+1} + i_{k-1} - 2i_k) - (i_{k+2} + i_{k-2} - 2i_k)}{4\delta\theta^2}.$$

These have proved more effective in the numerical examples quoted later. By Newton's method we can then deduce the peak position as occurs at

$$\hat{\theta}_k \approx \theta_k - \frac{i'_k}{i''_k}.$$

For all $k \in K$, corresponding peaks in the model data will be at

$$\bar{\theta}_k(n, r) \approx \theta_k - \frac{\Psi'(n, r, \theta_k)}{\Psi''(n, r, \theta_k)}$$

where the expressions for Ψ' and Ψ'' are similar to those for i' and i'' . We can now define an error function

$$E_3 = \sum_{k \in K} (\bar{\theta}_k(n, r) - \hat{\theta}_k)^2. \quad (12)$$

As an illustration we consider the dataset **py12log** again. At the least-squares solution $\hat{x} \approx (1.5032, 1.8014)$ the value of (12) is $E_3 \approx 70.7$. This is large compared with the

corresponding $E_2 = 19.07$ and suggests that a better identification might be obtained by seeking $x = (n, r)$ to solve a problem of the form

$$\text{Minimize } E_3(x) \quad \text{s.t. } E_2^{rms}(x) \leq (1 + \tau)E_2^{rms}(\hat{x}). \quad (13)$$

For small values of τ , problem (13) can be seen as a way of seeking values for radius and refractive index to give good agreement about the peak positions while limiting the amount by which the overall error E_2 can exceed its minimum value.

Bearing in mind the fact that E_2 (and perhaps E_3) can be highly non-convex we shall want to seek a global solution of (13). One way to do this is by applying an unconstrained global optimizer to an exact penalty function. For (13) we can use

$$F_1(x) = E_3(x) + \rho |\max(0, E_2^{rms}(x) - (1 + \tau)E_2^{rms}(\hat{x}))| \quad (14)$$

where ρ is a positive penalty parameter. Provided ρ is “sufficiently large”, the function (14) has a global minimum at the global solution of (13). (In practice we minimize (14) for an increasing sequence of ρ values until we get the same feasible solution for two successive values of ρ .)

Since F_1 is non-smooth we must use a non-derivative global optimization method and a suitable candidate is DIRECT [26], already outlined in an earlier section.

3 Tests with artificial data

In order to consider how problem (13) might be used in practice, we carry out some preliminary tests using artificially noisy data. In the following examples we have taken perfect data from the Lorenz-Mie model (with $n = 1.525$, $r = 1.475$) and then superimposed different noise distributions to create pseudo-experimental data. The noise distributions all have mean zero but with different standard deviations. The names **noise2***, **noise3*** distinguish two basic noise patterns; and the second subscript indicates the standard deviation – i.e. **noise*k** has $\sigma_m = k \times 0.15$.

In Table 1, the first five columns give information about the solution obtained by (global) minimization of the error function (4) – i.e. the root-mean-square values of E_2 and E_3 , the particle parameters \hat{n} and \hat{r} and the identification error defined as

$$D = \sqrt{(\hat{n} - 1.525)^2 + (\hat{r} - 1.475)^2}.$$

The last two columns give the root-mean-square values of E_2 and E_3 corresponding to $n = 1.525, r = 1.475$, the parameters of the “true” particle.

Problem	$E_2^{rms}(\hat{x})$	$E_3^{rms}(\hat{x})$	\hat{n}	\hat{r}	D	$E_2^{rms}(x^*)$	$E_3^{rms}(x^*)$
noise21	0.1496	0.294	1.5244	1.4750	6×10^{-4}	0.150	0.295
noise31	0.1479	0.237	1.5237	1.4746	1.4×10^{-3}	0.150	0.244
noise22	0.2993	1.166	1.5238	1.4750	1.2×10^{-3}	0.300	0.863
noise32	0.2963	0.354	1.5220	1.4748	3×10^{-3}	0.300	0.358
noise23	0.4491	1.366	1.5233	1.4750	1.7×10^{-3}	0.450	0.876
noise33	0.4441	0.551	1.5193	1.4754	5.7×10^{-3}	0.450	0.473
noise24	0.5988	1.680	1.5229	1.4751	2.1×10^{-3}	0.600	1.179
noise34	0.5911	0.742	1.5161	1.4761	9×10^{-3}	0.600	0.624

Table 1: Errors at the least-squares and exact solutions of artificial problems

We comment first of all that the approach based on global minimization of (4) seems fairly robust. Even at high levels of noise we get quite good identification of the actual particle. The worst results are for datasets **noise34**, **noise33**, **noise23** and **noise24**; and there are particularly large values for E_3^{rms} at the computed least-squares solutions for **noise22**, **noise23** and **noise24**. For these three cases the solution of (13) might be expected to provide a better identification than that given by the solution of (4). On the other hand, if we compare the values of E_3^{rms} in the final column with those in column two then we see that, in several cases, the peak-matching errors are *smaller* at the least-squares solution than at the true solution. In these instances it seems unlikely that the use of (13) will yield a better identification.

If we consider **noise22** and solve (13) with the constraint $E_2^{rms}(x) \leq 0.3$ we get $n \approx 1.5247$, $r \approx 1.4755$ with $E_3^{rms} \approx 0.827$.

Since we know the “true” solution we can see that this is an appreciable improvement in the identification of the particle compared with that given by minimizing (4). However, even without such prior knowledge we can *infer* that the solution of (13) is better if we look at the relative changes in E_3 and E_2 . We let \hat{x} denote the global minimum of (4) and \tilde{x} the solution of (13) and consider the ratio

$$\kappa = \frac{E_3^{rms}(\tilde{x})}{E_3^{rms}(\hat{x})}. \quad (15)$$

For this problem we find that $\kappa \approx 0.71$, indicating that there has been about a 30% improvement in peak-matching. This has been obtained at the expense of a small (less than 0.3%) increase in the least-squares error function. Therefore it seems reasonable to regard \tilde{x} as a better solution to the identification problem.

Of course, in a real problem, we would not know in advance that $E_2^{rms}(x) \leq 0.3$ is an appropriate limit to use in (13). A more general way of looking at the situation is suggested by the remark at the end of the previous paragraph. The improvement in peak-matching due to going from \hat{x} to \tilde{x} is given by

$$\Delta_3(\tilde{x}) = 1 - \frac{E_3^{rms}(\tilde{x})}{E_3^{rms}(\hat{x})}$$

while the loss of accuracy in overall data-matching is given by

$$\Delta_2(\tilde{x}) = \frac{E_2^{rms}(\tilde{x})}{E_2^{rms}(\hat{x})} - 1.$$

Since we want the change in E_2 to be small *relative to* the change in E_3 we can replace the constraint in (13) by the requirement that $\Delta_2(\tilde{x}) < \varepsilon \Delta_3(\tilde{x})$ where ε is a small positive constant. Thus we obtain a more general problem instead of (13), namely

$$\text{Minimize } E_3(x) \quad \text{s.t.} \quad \frac{E_2^{rms}(\tilde{x})}{E_2^{rms}(\hat{x})} - 1 \leq \varepsilon \left(1 - \frac{E_3^{rms}(\tilde{x})}{E_3^{rms}(\hat{x})}\right). \quad (16)$$

The global solution of this constrained optimization problem can be obtained if we apply DIRECT to an exact penalty function of similar form to (14).

Table 2 shows the results of solving (16) with $\varepsilon = 0.01$ for some of the noisy datasets considered above. (For the other datasets the refinement procedure makes no significant change to the original least-squares solution \hat{x} .)

Problem	$E_2^{rms}(\hat{x})$	$E_3^{rms}(\hat{x})$	\hat{n}	\hat{r}	D
noise22	0.3	0.793	1.5249	1.4756	6×10^{-4}
noise23	0.45	0.704	1.5254	1.4759	9.8×10^{-4}
noise33	0.444	0.52	1.5200	1.4757	5×10^{-3}
noise24	0.601	1.061	1.5256	1.4741	1.1×10^{-3}
noise34	0.592	0.689	1.5173	1.4765	7.8×10^{-3}

Table 2: Errors at solutions of (16) for artificial problems

From our knowledge of the “true” solution to these particular examples we can see that the refinement process gives a worthwhile improvement in the computed estimates of

n and r . It is worth noting that, on these examples at least, it is the estimate of refractive index more than the estimate of radius that is improved by the use of (16). Specifically, for **noise23** both the peak match and the accuracy of the estimated (n, r) improve by about 40% while for **noise24** a peak-match improvement of about 37% yields a 48% reduction in the errors in (n, r) . On the other hand, however, the refinement makes, at best, only small improvements to n when the peak-matching errors at \hat{x} are relatively small. We can get slightly better results from (16) in the cases involving **noise33** and **noise34** if we change ε to 0.02. We then obtain for **noise33** the results:

$$n \approx 1.5207, r \approx 1.4760 \text{ with } E_2^{rms} \approx 0.445, E_3^{rms} \approx 0.501, D \approx 4.4 \times 10^{-3}$$

while for **noise34** we get

$$n \approx 1.5183, r \approx 1.4767 \text{ with } E_2^{rms} \approx 0.592, E_3^{rms} \approx 0.661, D \approx 6.9 \times 10^{-3}$$

In both cases the peak-match is improved by about 10% and the error in the computed (n, r) is reduced by about 40%. This indicates that the choice of ε in (16) may be somewhat problem dependent. If, for instance, we use $\varepsilon = 0.02$ for the cases **noise23** and **noise24** then the refinement process *over-corrects* and produces worse estimates of both n and r . Notwithstanding these remarks, however, the experience reported in this section seems sufficiently encouraging that we shall now turn our attention to problems involving real-life experimental data.

4 Experimental data results

We now consider four sets of experimental data. The first two sets (**py12log**, **lp29log**) are scattering patterns for fungal spores and the second two (**n1log**, **p1log**) are measurements from polystyrene microspheres. All patterns were obtained from single particles suspended in water [2, 4], leading to the presence of distortion and/or noise at quite high levels in some cases. In these realistic situations we have, at best, only a rough estimate of the “true” solution.

Figure 2 shows the scattering pattern **py12log**. Knowledge of the original experiment suggests that this particle has refractive index $n \approx 1.5$ and radius $r \approx 1.8 \mu\text{m}$. As we observed earlier, the identification is based only on data points in the range $20^\circ - 160^\circ$ because the intensity measurements appear to have been truncated at low and high scattering angles.

Figure 7 shows the scattering pattern **lp29log** in the range $20^\circ - 160^\circ$. Prior knowledge

in this case suggests a particle with $n \approx 1.5$ and $r \approx 1.5 \mu\text{m}$.

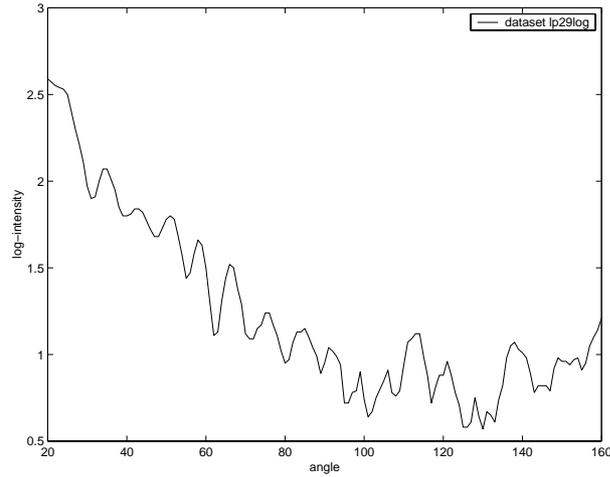


Figure 7: Scattering pattern **lp29log**

The third data set, **n1log**, for which the expected values are $n \approx 1.6$ and $r \approx 1.0 \mu\text{m}$, is shown in Figure 8. Here we only have reliable measurements in the range $20^\circ - 140^\circ$.

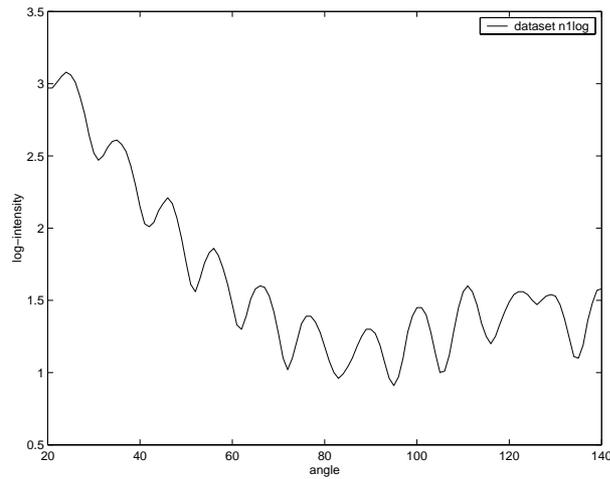


Figure 8: Scattering pattern **n1log**

The last data set, **p1log** is shown in Figure 9 where measurements are confined to the range $10^\circ - 120^\circ$. The expected values of the refractive index and radius in this case are $n \approx 1.6$ and $r \approx 0.6 \mu\text{m}$.

The last two patterns came from particles which were known to have good sphericity and homogeneity of refractive index; the refractive index was larger than for the first two particles, leading to a relatively stronger scattering pattern with respect to noise. In

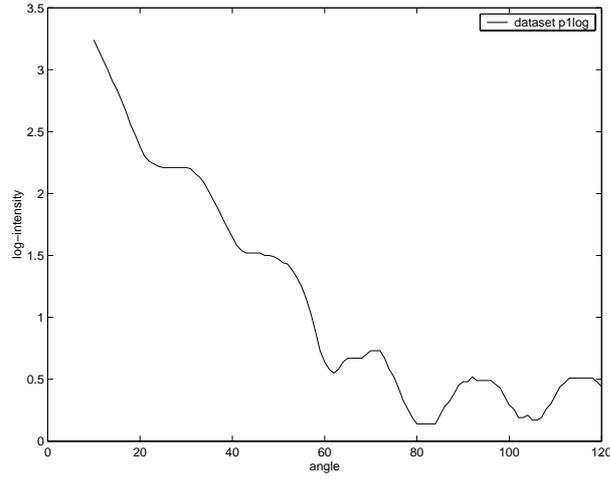


Figure 9: Scattering pattern **p1log**

addition, **n1log** came from a larger particle than **p1log**, giving a further improvement in the signal-to-noise ratio.

4.1 The dataset **py12log**

Since we believe that the data comes from a particle with $n \approx 1.5$ and $r \approx 1.8$, the ranges $1.4 \leq n \leq 1.6$, $1.7 \leq r \leq 1.9$ were chosen for the global optimization of (4). It turns out that there are 19 local optima in this region. As already mentioned, the best three solutions are:

$$n \approx 1.5032, r \approx 1.8014 \text{ giving } E_2^{rms} \approx 0.368, E_3^{rms} \approx 1.979$$

$$n \approx 1.4737, r \approx 1.8210 \text{ giving } E_2^{rms} \approx 0.373, E_3^{rms} \approx 2.002$$

$$n \approx 1.4909, r \approx 1.8219 \text{ giving } E_2^{rms} \approx 0.376, E_3^{rms} \approx 3.679$$

While these candidate solutions are relatively close in terms of E_2 , none of them are very satisfactory in terms of the peak-matching error E_3 . Therefore we solve problem (16), taking \hat{x} as the best of the three quoted minima of E_2 and using $\varepsilon = 0.01$. We get

$$n \approx 1.5007, r \approx 1.8002 \text{ giving } E_2^{rms} \approx 0.371, E_3^{rms} \approx 0.522$$

Here the values of n and r have moved nearer to the expected value and, as a more objective comment, the peak match error E_3^{rms} is reduced by over 70% compared with what it was at the least-squares solution.

If we solve (16) with \hat{x} taken as the second of the local optima we obtain

$$n \approx 1.4993, r \approx 1.7989 \text{ giving } E_2^{rms} \approx 0.376, E_3^{rms} \approx 0.372$$

In other words, peak-matching considerations have moved the solution substantially away

from the second-best local minimum of (4) and we can now have more confidence that, to two decimal places at least, the refractive index and radius of the particle can be taken as 1.50 and 1.80.

4.2 The dataset lp29log

Because of our expectations about this particle, the search ranges $1.4 \leq n \leq 1.6$ and $1.3 \leq r \leq 1.7$ were used for the global optimization of (4). At least thirteen local minima exist in this region, of which the best two are

$$n \approx 1.5428, r \approx 1.6218 \text{ giving } E_2^{rms} \approx 0.3555, E_3^{rms} \approx 3.079$$

$$n \approx 1.5373, r \approx 1.4411 \text{ giving } E_2^{rms} \approx 0.3562, E_3^{rms} \approx 0.884$$

These two solutions are very close in terms of E_2 , although the first is inferior in terms of peak-matching. Both solutions are plausible in terms of our limited prior knowledge about the particle.

When we solve (16), using $\varepsilon = 0.01$ and taking \hat{x} as the better of the local solutions above we obtain

$$n \approx 1.5448, r \approx 1.6240 \text{ giving } E_2^{rms} \approx 0.359, E_3^{rms} \approx 0.431.$$

This gives about an 85% reduction in E_3^{rms} . Moreover, if we base (16) on the second-best candidate solution it produces virtually the same result. Once again, we have been able to use peak-matching to distinguish between two quite similar solutions to the identification problem and hence to deduce that the refractive index and radius can be taken as about 1.54 and 1.62, respectively.

4.3 The dataset n1log

On the basis of *a priori* information, a search region $1.5 \leq n \leq 1.7$, $0.8 \leq r \leq 1.5$ was used for the global minimization of (4). There are 29 local minima in this box, but in this case the global solution is clearly distinct from the rest. It is

$$n \approx 1.5869, r \approx 1.1082 \text{ giving } E_2^{rms} \approx 0.1165, E_3^{rms} \approx 0.242.$$

If we take this point as \hat{x} and solve (16) with $\varepsilon = 0.01$ we get no significant change in the computed values of n and r which indicates that peak-matching cannot improve this already quite good solution.

4.4 The dataset p1log

For this problem, the range for global optimization of (4) was $1.5 \leq n \leq 1.7$, $0.4 \leq r \leq 0.7$.

The result obtained is

$$n \approx 1.6537, r \approx 0.5811 \text{ giving } E_2^{rms} \approx 0.17, E_3^{rms} \approx 0.818$$

If we base problem (16) on this point and use $\varepsilon = 0.01$ we get the result

$$n \approx 1.6538, r \approx 0.5805 \text{ giving } E_2^{rms} \approx 0.17, E_3^{rms} \approx 0.77.$$

This small shift in n and r represents only about a 5% improvement in peak matching compared with the original least-squares solution. Hence we have a reasonable indication that the particle has already been quite well identified by the basic least-squares approach.

5 Discussion

We have considered a standard least squares approach to the solution of an inverse light scattering problem and noted that, for noisy data, it may not always give satisfactory results. Hence we have considered an alternative technique using a constrained optimization problem (16) which is based on the idea of minimizing a “peak-matching” function subject to constraints on the least-squares error function (4). Some initial tests of (16) have been performed using data sets which include artificial noise. In cases where the solution obtained by global optimization of (4) does not agree very well with the position of intensity peaks in the data, (16) has been shown to be quite successful in producing better estimates of refractive index and radius which give better correspondence to the “shape” of the scattering data.

Based on these ideas, we have developed a composite approach to the identification of particles from ‘real-world’ experimental data, which may be strongly distorted and/or contain high levels of noise. The first phase finds the global optimum of (4) to get a trial solution (\hat{n}, \hat{r}) for refractive index and radius. The constrained problem (16) is then solved to find (\tilde{n}, \tilde{r}) to minimize a peak matching error function subject to a restriction on the permitted increase in the overall error function (4). This increase is in turn dependent on the amount of improvement in the peak-matching error function.

The new composite approach performs quite well in compensating for noise that is artificially imposed on “perfect” scattering data. However, it is clear that fitting real experimental data is a more difficult problem. Figure 2, for example, shows that real data

can be subject to distortions – such as truncation of extreme values – which may not correspond very well to the normally-distributed artificial noise applied to the perfect data in section 3. In practice, detailed information concerning noise levels and/or filtration may be either unavailable or unreliable – as, for example, when adaptive noise reduction procedures are used to allow dealing with variable noise levels in raw data. The effects of the distortion may be only partly alleviated by our strategy of basing the identification on a restricted range of the data.

In using our refinement procedure on real-life examples we have found that it can sometimes produce significant improvements to the least-squares estimates of the particle parameters; but in other cases it makes little or no difference to the results. The difficulty in assessing the effectiveness of the changes that are (or are not) made is that we only know approximately what the “right answer” should be! Comparison between the Lorenz-Mie scattering patterns generated at the solutions of (4) and (16) shows that the differences are quite subtle. Re-positioning of peaks at the solution of (16) seems mainly to be restricted to the angular range from about 80° to about 110° . Hence it is partly on the basis of measurably useful performance on the artificial data that we are inclined to believe that the refined solutions from (16) are in fact better than the ones from the original least-squares approach. Conversely, if refinement based on (16) does not make a significant change to the solution then we take it as some degree of confirmation that the particle has been identified reasonably well. We can say that the use of (16) has helped to resolve ambiguities in the solutions obtained using (4) alone.

Of course, it is not only the presence of noise which adversely affects solutions to the inverse light-scattering problem. Some difficulties may arise due to the common practice of fitting intensity data in logarithmic form, as this type of scaling can have the undesirable effect of emphasizing noise. Since evidence has recently been provided that logarithmic scaling can be inferior to linear scaling, future studies should examine the benefits of using linearly scaled data with various angular weighting functions, such as $\sin(\theta)^4$. [15]

Another source of difficulty in identifying real particles from experimental data is of course that our scattering model is strictly appropriate for homogeneous, spherical particles. We are not alone in making the working assumption that the Lorenz-Mie model will be adequate for practical purposes [1, 8, 13, 29]. Other models exist, however, such as the one describing a coated sphere - i.e. a particle which has one refractive index in

an outer "shell" and a different refractive index from the center out to a radius r_2 . Such a particle is therefore described by four parameters and it is possible that this somewhat more complex model could be adjusted by an optimization process to match physical data more closely, whatever form of error criterion is being used. It would not add significantly to the difficulty of the problem to increase the number of variables by two. Although the number of shapes for which scattering solutions can be obtained is limited, there are other forms for which rigorous solutions exist, such as ellipsoids. The extension of all the techniques used in this paper to deal with coated spherical or ellipsoidal particles would be a worthwhile investigation.

Finally, the study confirms the intuitive assertion that while a small number of data points can be used for accurate inversion of angular scattering data in the case of low-noise input – about 10 points for a 2-parameter solution with an average accuracy of 0.1 % [15] – a much greater amount of data may be needed when significant levels of noise are present.

References

- [1] P. J. Wyatt, Some chemical, physical and optical properties of fly ash particles, *Appl. Opt.* **7**, 975 (1980).
- [2] Z. J. Ulanowski, *Investigations of microbial physiology and cell structure using laser diffractometry*, PhD thesis, Hatfield Polytechnic (1988).
- [3] G. Gousbet and G. Grehan, *Optical particle sizing* (Plenum, New York, 1988).
- [4] Z. J. Ulanowski and I. K. Ludlow, Water distribution, size and wall thickness in *Lycoperdon pyriforme* spores. *Mycolog. Res.* **93**, 28 (1989).
- [5] H. G. Barth and S. T. Sun, Particle-size analysis, *Anal. Chem.* **63**, R1 (1991).
- [6] L. A. de Pieri, I. K. Ludlow, and W. M. Waites, The application of laser diffractometry to study the water content of spores of *Bacillus sphaericus* with different heat resistances, *J. Appl. Bacteriol.* **74**, 578 (1993).
- [7] C. A. O. Nascimento, R. Guardani and M. Giulietti, Use of neural networks in the analysis of particle size distributions by laser diffraction, *Powd. Technol.* **90**, 89 (1997).

- [8] P. G. Hull and M. Quinby-Hunt, A neural-network to extract size parameter from light-scattering data, *SPIE Proc.* **2963**, 448 (1997).
- [9] L. P. Bayvel and A. R. Jones, *Electromagnetic scattering and its applications* (Applied Science Publishers, London, 1981).
- [10] C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles* (Wiley, New York, 1983).
- [11] K. Shimizu and A. Ishimaru, Differential Fourier-transform technique for the inverse scattering problem, *Appl. Opt.* **29**, 3428 (1990).
- [12] M. R. Jones, B. P. Curry, M. Q. Brewster and K. H. Leong, Inversion of light - scattering measurements for particle size and optical constants: theoretical study, *Appl. Opt.* **33**, 4025 (1994).
- [13] G. M. Quist and P. J. Wyatt, Empirical solution to the inverse light scattering problem by the optical strip - map technique, *J. Opt. Soc. Am. A* **2**, 1979 (1985).
- [14] V. P. Maltsev and V. N. Lopatin, Parametric solution of the inverse light-scattering problem for individual spherical particles, *Appl. Opt.* **36**, 6102 (1997).
- [15] Z. J. Ulanowski, Z. Wang, P. H. Kaye and I. K. Ludlow, Application of neural networks to the inverse light scattering problem for spheres, *Appl. Opt.* **37**, 4027 (1998).
- [16] I. K. Ludlow and J. Everitt, The application of Gegenbauer analysis to light scattering from spheres, theory, *Phys. Rev. E* **51**, 2516 (1995).
- [17] I. K. Ludlow and J. Everitt, Systematic behavior of the Mie scattering coefficients of spheres as a function of order, *Phys. Rev. E* **53**, 2909 (1996).
- [18] S. Zakovic, *Global optimization applied to an inverse light scattering problem*, PhD thesis, University of Hertfordshire (1997).
- [19] S. Zakovic, Z. J. Ulanowski, and M. C. Bartholomew-Biggs, Application of global optimization to particle identification using light scattering, *Inverse Problems* **14**, 1053 (1998).

- [20] S. Zakovic, Z.J. Ulanowski and M.C. Bartholomew-Biggs Particle identification using light scattering: a global optimization problem Technical Report, Numerical Optimisation Centre, University of Hertfordshire, 2002
- [21] R. Mireles, The inverse problem of electromagnetic theory. I. Uniqueness theorem for cylinders, *J. Mathematics Phys. (MIT)* **45**, 179 (1966).
- [22] I. K. Ludlow and J. Everitt The inverse Mie problem, *J. Opt. Soc. Am. A* (submitted)
- [23] Z. Ulanowski, R.S. Greenaway, P.H. Kaye & I.K. Ludlow, Laser diffractometer for single-particle scattering measurements, *Measurement Science and Technology* **13**, 292-296 (2002)
- [24] A. Rinnooy Kan and G. T. Timmer, Stochastic global optimization methods. Part I: Clustering methods, *Math. Program.* **39**, 27 (1987).
- [25] A. Rinnooy Kan and G. T. Timmer, Stochastic global optimization methods. Part II: Multilevel methods, *Math. Program.* **39**, 57 (1987).
- [26] D.R.Jones, C.D. Perttunen and B.E.Stuckman, Lipschitzian optimization without the Lipschitz constant, *Journal of Optimization Theory and Applications*, vol 79, pp157-181, 1993
- [27] M.C. Bartholomew-Biggs, S.C. Parkhurst and S.P. Wilson Using DIRECT to solve an aircraft routing problem, *Computational Optimization and Applications*, Vol 21, pp 311-323, 2002.
- [28] L. C. W. Dixon, personal communication (1996).
- [29] F. Robillard and A. J. Patitsas, Determination of size, size distribution and refractive index of Dow latexes EP-1358-38 by the Mie scattering method, *Can. J. Phys.* **52**, 1571 (1974).