Soot particle sizing by inverse analysis of multiangle elastic light scattering using Bayesian inference

Burr, D. W.; Daun, K. J.; Link, O.; Thomson, K. A.; Smallwood, G.

National Research Council Canada   Conseil national de recherches Canada

Canada

Soot Particle Sizing by Inverse Analysis of Multiangle Elastic Light Scattering Using Bayesian Inference

D. W. Burr[a*], K. J. Daun[a], O. Link[b], K. A. Thomson[b], and G. Smallwood[b]

[a] *Department of Mechanics and Mechatronics Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1*
[b] *Institute for Chemical Process and Environment Technology, National Research Council, 1200 Montreal Road Ottawa, Ontario K1A 0R6, Canada*

1. Introduction

Spherical soot particles form during combustion and agglomerate into polydisperse fractal aggregates. These aggregates affect human health and the environment in different ways depending on the aggregate size and morphology. For instance, smaller particles can penetrate deeply into the lungs and translocate into the bloodstream, causing cardiopulmonary disease [1]. Larger aggregates act as condensation nuclei for clouds or increase the absorptivity of glaciers causing melting [2]. Thus, it is important to be able to quickly and accurately determine the soot aggregate size distribution of a soot aerosol to determine what impact it will have on human health and the environment.

One way to determine soot aggregate size is physical sampling, where aggregates are removed from the flame and surrounding aerosol and examined using electron microscopy. This method is time consuming and, furthermore, may introduce a bias due to the preferential attraction to the collection slide of some aggregate sizes over others. A more efficient method is to shine collimated light through the aerosol and measure the scattering at multiple angles [3]. This method provides a higher temporal and spatial resolution than physical sampling, and furthermore, it is nonintrusive.

The measured scattering data is related to the soot aggregate size distribution through a Fredholm integral equation of the first kind,

$$g(\theta) = c_1 \int_1^\infty P(N_p) K(N_p, \theta) \mathrm{d}N_p, \tag{1}$$

where $g(\theta)$ is the measured angular scattering data, c_1 is an unknown scaling constant, $P(N_p)$ is the unknown aggregate size distribution with N_p representing the number of primary particles, and $K(\theta, N_p)$ is the kernel function that is derived from light scattering theory, coupled with a model of the system. This integral equation is ill-posed because an infinite number of different aggregate size distributions exist that satisfy Eq. (1) within experimental accuracy. Thus, a small amount of error in the measured data can potentially lead to a large variation in the recovered aggregate size distribution.

Rayleigh-Debye-Gans Polydisperse Fractal Aggregate (RDG-PFA) theory is used to model light scattering by soot aggregates due to its simplicity and relative accuracy for the aggregate sizes and shapes in question [4]. The model provides a phase function, $\Phi(\theta, N_p)$, which gives the relative intensity of the scattered light as a function of scattering angle and aggregate size,

---

[*] Corresponding author: dburr@engmail.uwaterloo.ca

$$\Phi\left(\theta, N_p\right) = 4\pi \frac{\overline{C}_{vv}^{agg}}{\overline{C}_{s}^{agg}}, \tag{2}$$

where $\overline{C}_{vv}^{agg}$ and $\overline{C}_{s}^{agg}$ are the vertically polarized cross section of the aggregates and cross section of the particles, respectively. Substituting this phase function into the radiative transfer equation results, after significant simplification, in

$$K\left(N_p, \theta\right) = \frac{N_p^2 f\left(q(\theta), R_g\left(N_p\right)\right)}{\sin(\theta)}, \tag{3}$$

where $q(\theta)$ is the scattering wave vector [3], $R_g(N_p)$ is the radius of gyration [4], and $f\left(q(\theta), R_g\left(N_p\right)\right)$ is the form factor. Various scattering form factors for soot aggregates are presented in the literature; we chose to use the one presented by Yang and Köylü [5].

Once the kernel function is known, in principle Eq. (1) can be solved for $P(N_p)$. One way to do this is by least-squares fitting a presumed distribution type, such as a lognormal distribution, to the data [6]. This ignores the underlying ill-posedness of the problem, however, and this could lead to a fitted-distribution which would produce the scattering data yet not resemble the actual size distribution in the aerosol [3]. Furthermore, the parameters of the fitted distribution are very sensitive to error in the measured data.

An alternative is to use a regularization technique that augments Eq. (1) with assumed *a priori* information about the system, most of which work on an ill-conditioned matrix derived from that of Eq. (2). There are a number of regularization techniques, such as Twomey's linear [7] and nonlinear techniques [8], and Tikhonov regularization [9], that have been applied to ill-posed problems in the past, each with its own unique advantages and disadvantages.

We first discuss Tikhonov regularization in the context of this problem before moving to the more generalized Bayesian inference technique from which Tikhonov is derived. Next we present a rigorous method for determining the scaling constant which is integral to inverting Eq. (3). Finally, a set of experimental data will be inverted and a reconstructed distribution compared to that found through physical sampling.

2. Regularization Theory

The information provided by the ill-conditioned matrix equation is insufficient to accurately reconstruct the aggregate size distribution by itself; regularization incorporates presumed system characteristics to aid in the reconstruction of a solution. Such information might include smoothness, concavity, or non-negativity, which are known from aggregation mechanics or other similar theories.

Regularization algorithms differ on how this information is incorporated into the analysis. Two very similar techniques, Tikhonov and Twomey linear regularization, solve the least squares objective function,

$$x^* = \operatorname{argmin}\left(\left\|Ax_\lambda - g\right\|_2^2 + \lambda^2 \left\|Lx_\lambda\right\|_2^2\right), \tag{4}$$

where L is a smoothing matrix, λ is the regularization parameter, A is the discrete kernel function (which contains the unknown scaling parameter, c_l), and x is the discrete aggregate size distribution [10]. This is generally done by rewriting Eq. (4) as a matrix equation and solving for the x^* that minimizes the norm over a range of λ values.

$$x^* = \operatorname{argmin}\left\|\begin{bmatrix} A \\ \lambda L \end{bmatrix} x - \begin{bmatrix} g \\ 0 \end{bmatrix}\right\|_2^2 \tag{5}$$

Since the assumed information isn't "exactly" right, it is critical to determine the optimal amount of regularization, which is done by choosing an appropriate value of the regularization parameter, λ. The parameter selection technique used in this paper is the L-Curve Criterion, which is based on the parametric plot of $\log\left(\left\|Ax_\lambda - g\right\|_2\right)$ versus $\log\left(\left\|Lx_\lambda\right\|_2\right)$ in terms off λ, that resembles the letter "L." The vertical section of the curve corresponds to solutions that are under-regularized and the horizontal section corresponds to solutions that are over-

regularized. A good compromise value of λ is found at the corner of the L (point of maximum curvature) which is found by maximizing

$$\kappa(\lambda) = \frac{\rho'\eta'' - \rho''\eta'}{\left(\left(\rho'\right)^2 + \left(\eta'\right)^2\right)^{3/2}}, \tag{7}$$

with $\rho = \log\left(\left\|Ax_\lambda - g\right\|_2\right)$ and $\eta = \log\left(\left\|Lx_\lambda\right\|_2\right)$, $v' = \delta v/\delta\lambda$ and $v'' = \partial^2 v/\partial\lambda^2$.

In many problems Tikhonov regularization is sufficient by itself to stabilize deconvolution of the ill-posed problem. Unfortunately, the extreme ill-posedness of the current problem necessitates the incorporation of additional prior information to obtain a robust solution. Bayesian inference is a more formalized and generic method of incorporating prior information into inverse problems. The basic least-squares objective function follows from Bayes' theorem for several sets of data,

$$P(x \mid g) = \frac{P(g \mid x)}{P(g)} \times P_{model}(x), \tag{8}$$

with $P(x \mid g)$ is the probability of x being correct given that g has been observed, $P(g \mid x)$ is the probability of the model being true given x having been observed (sometimes called the likelihood function), $P(g)$ is the marginal probability of the data (which is assumed to be unity as the data has been measured experimentally and thus is absolutely probable), and $P_{model}(x)$ is the prior probability of the model being correct [11]. If $P_{model}(x)$ is set to unity all models are assumed to be correct, which is known as an uninformed prior.

If the data is assumed: to follow a Gaussian distribution; to be linear such that $Ax = g$; to be scaled properly so that $\sigma_j = \sigma$, then [12]

$$P(g \mid x) \propto \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left\|Ax - g\right\|_2^2\right). \tag{9}$$

Maximizing the likelihood $P(x \mid g)$ is equivalent to minimizing the argument in the exponential function. This is problematic, however, due to the ill-conditioning of A; multiple solutions of x exist that "almost" minimize $\left\|Ax - g\right\|_2^2$, and thus maximize the likelihood function. Consequently, additional information must be added through multiple Bayesian priors; the product of the priors forms $P_{model}(x)$.

The first prior added is the smoothness criterion or Gibb's prior, which promotes a smooth solution [13].

$$P_{pr}(x) \propto \exp\left(-\beta\left\|Lx\right\|_2^2\right) \tag{10}$$

This prior is a discrete difference operator [10] and implies that the variation between two values in x should be small, thereby promoting a smooth (continuous, differentiable) solution.

Nonnegativity is enforced using the indicator function, which equals unity if x_i is within the noted range and zero everywhere else. This guarantees that the probability of a negative value in the reconstructed solution, x, is zero.

$$P_{pr}(x) \propto \prod_{j=1}^{n} I_{[0,\infty)}(x_j) \tag{11}$$

The Heaviside function was not used firstly due to the ambiguity of its value at $x_i = 0$, which by convention is taken to be 0.5, but can also be taken to be either zero or one as needed, and secondly that, although in this case both functions have the same range, in general the indicator function is much more powerful.

Finally, it is possible to promote likely solutions that match a prescribed self-preserving, scaling distribution [3] using

$$P_{pr}(x) = \prod_{k=1}^{n} p_k(\theta)^{m_k}, \tag{12}$$

where $p_k(\theta)$ is the probability of an aggregate being found in the *kth* prescribed size bin in the scaling distribution that is specified by θ, and m_k is the number of aggregates in an experiment that fall within the kth bin [14]. This was

implemented as follows: an initial guess for the distribution was made, and the cumulative density function (CDF) of a self-preserving, scaling distribution was iteratively fitted to the CDF of the initial guess using the Kolmogorov-Smirnov goodness-of-fit test [15]. The fitted distribution was taken to be $p_k(\theta)$ and was converted to probabilities of an aggregate being found in each bin of aggregate sizes. If it is assumed that an experiment was undertaken such that ten thousand aggregates were observed, and the initial guess of the distribution was converted into probabilities, then the number of aggregates, m_k, found in each bin was taken to be ten thousand multiplied by the probability. This was used to calculate $P_{pr}(x)$, which allowed the least-squares objective function to develop a new initial guess through an iteration of Bayesian inference.

The idea behind this prior is that each time an aggregate is experimentally found at a given size the probability for the whole experiment is multiplied by the probability that an aggregate would be that size. If a great number of aggregates are found at a size that has a low probability, the total probability of the experiment happening is low. Likewise, if a great number of aggregates are found of a size that is highly probable, the experiment is much more likely.

Combining these priors with the likelihood function an informed prior for the soot aerosol system is obtained:

$$P(x \mid g) \propto \exp\left(-\frac{1}{2\sigma^2}\|Ax - g\|_2^2\right) \times \exp\left(-\beta\|Lx\|_2^2\right) \times \left(\prod_{j=1}^{n} I_{[0,\infty)}(x_j)\right) \times \left(\prod_{k=1}^{n} p_k(\theta)^{m_k}\right). \tag{13}$$

If we let $\beta = \lambda^2/2\sigma^2$ this can be rewritten as

$$P(x \mid g) \propto \exp\left(-\|Ax - g\|_2^2 - \lambda^2\|Lx\|_2^2\right) \times \left(\prod_{j=1}^{n} I_{[0,\infty)}(x_j)\right) \times \left(\prod_{k=1}^{n} p_k(\theta)^{m_k}\right). \tag{14}$$

Due to the monotonicity of the logarithm function, maximizing Eq. (14) is equivalent to minimizing

$$x^* = \arg\min\left\{\|Ax - g_2^2\| + \lambda^2\|Lx\|_2^2 - \ln\left(\prod_{j=1}^{n} I_{[0,\infty)}(x_j)\right) - \ln\left(\prod_{k=1}^{n} p_k(\theta)^{m_k}\right)\right\}. \tag{15}$$

Note that if only the Gibbs prior, Eq. (10), is implemented, solving Eq. (8) is equivalent to Tikhonov regularization, Eq. (4).

3. Determination of the Scaling Parameter

A complicating factor in the analysis of angular scattering data is the unknown scaling coefficient in Eq. (1), c_1, which depends on the aggregate number density in the aerosol, the intensity of the interrogating light source, and the efficiency of the optical apparatus, among other constants. While some of these terms can be found through calibration, the aggregate number density in the aerosol is difficult to evaluate, so c_1 remains unknown. Equation (4) is therefore a function of both λ and c_1, making this a nonlinear inverse problem.

For the sake of accuracy, the scaling constant was rewritten as $c_1 = 1/c_2$, where c_2 is the value being solved for. A study of soot aggregates carried out using electron microscopy by Tian et al. [16] was used to create a histogram of soot aggregate. A lognormal curve was fit to the histogram with parameters σ_{2g} and N_g found as 3.35 and 23.14, respectively, and a set of data was generated using the lognormal curve through numerically integrating Eq. (1). The norm of the data was compared to that of the experimentally measured data, which found a value of c_2 to be 2.05×10^5. A similar comparison was carried out on a set of data found by numerically integrating a lognormal distribution fitted to the experimental data, with parameters σ_{2g} and N_g being 1.49 and 38.47, respectively. The value of c_2 for the second case was found to be 8.0×10^4.

With those two values as guidelines, the problem was solved using Tikhonov regularization over a range of c_2 values from 4.0×10^4 to 2.05×10^5 and the residual norms were plotted against the value of c_2. As can be seen below in Fig. 1a there is no appreciable difference in the residual norms over the various values of the scaling constant. Next we added the non-negativity prior; the regularization parameter λ, however, is not a smooth function of with respect to the residual norm, and so an optimal value was obtained. To get around this, the value of λ found for Tikhonov regularization was used, which produced a smooth plot of the scaling constant versus the residual norm squared as

shown in Fig. 1b. The minimum c_2 value was found to be 6.0×10^4, which was used to solve the iterative minimization algorithm. The problem finally was then solved using the full Bayesian algorithm, Eq. (15), over a range of scaling constant values determined by looking at the lognormal fitted data, and plotted to find a minimum value, found to be around 8.0×10^4.
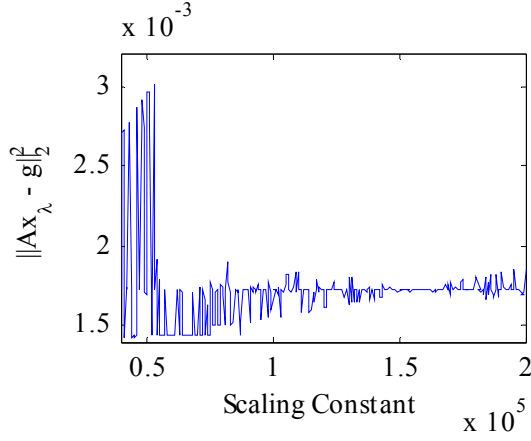

Fig. 1a - Scaling constant versus the residual norm for Tikhonov Regularization
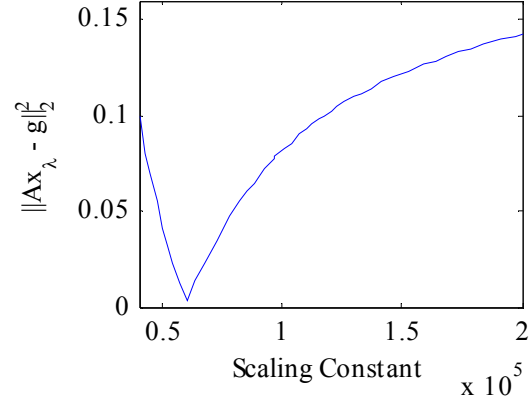
Fig. 1b - Scaling constant versus the residual norm with the non-negativity constraint included

4. Results

Equation (1) was first rewritten as Eq. (5) and solved using Tikhonov regularization, employing the L-Curve Criterion to determine the optimal value of λ and then QR decomposition to reconstruct the distribution. The recovered distribution was found to be negative for some aggregate sizes, which is a non-physical result. The general shape and magnitude of the recovered distribution was as expected, however, and only the extreme ill-conditioning of the matrix A prevented the generally-powerful method from working properly.

To find a physical solution non-negativity was added to the least-squares objective function, as in Eq. (15), in conjunction with the scaling distribution which the soot aggregates should follow [3]. The problem was solved using a built-in iterative minimization method in MATLAB, at the value of the scaling constant found by minimizing the residual norm versus the scaling constant for the solution to Eq. (15). The solution to the Tikhonov regularization and the Bayesian approach are plotted in Fig. 2a along with the lognormal fit to the TEM data. The different distributions were then used to reconstruct the measured data, $g(\theta)$, shown in Fig. 2b.

None of the distributions recovered match exactly the distribution found through TEM analysis, which speaks to the ill-posedness of the problem. While the distributions are different, the reconstructed data set from each is almost identical to that of the measured data, with the difference being well within the experimental error. This shows clearly that the scattering data in and of itself is not sufficient to invert the problem and other knowledge, from sources such as aerosol dynamics, must be added to the problem to develop a solution that matches the expected distribution type.
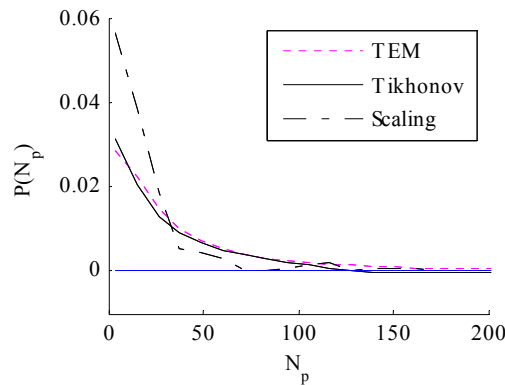
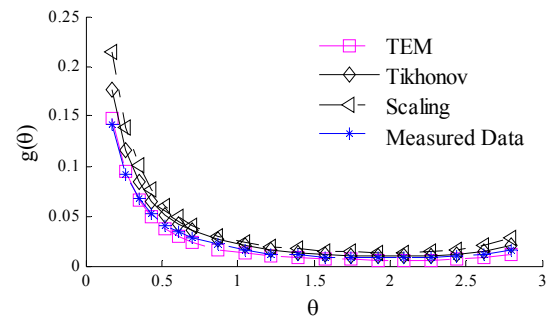Fig. 2a - Recovered distributions found using the different methods mentioned.

Fig. 2b - Reconstructed data from the recovered distributions versus the measured data

5. Conclusion

There is a need to quickly and accurately determine the aggregate size distribution of a soot aerosol system to determine its effect on human health and the environment. This task is challenging as one must solve a Fredholm integral equation of the first kind, which is an extremely ill-posed problem that strongly amplifies any noise present in the system. In this paper this was done by converting the equation into an ill-conditioned matrix equation and applying two regularization techniques to solve for the unknown aggregate size distribution.

Tikhonov regularization was applied to the problem but proved unable to overcome the extreme ill-posedness present to produce physically realistic solutions. Bayesian inference allows us to add assumed information (priors) in a relatively simple and mathematically rigorous manner that, while computationally more expensive, resulted in distributions similar to the expected distributions. The differences between the reconstructed distributions and the experimentally determined distributions highlight the difficulty in solving this problem. Future work will seek to add other priors to more accurately reconstruct the experimental distribution. Finally, a determination of the optimal size for the problem (the dimension of the matrix A and the number of scattering measurements, $g(\theta)$, to make) will hopefully reduce some of the ill-posedness of the problem.

References
[1] C.A. Pope, D.W. Dokery, Journal. Air and Waste Management 56 (2006) 749-752.
[2] T.C. Bond, H. Sun, Environmental Science. Technology 39 (2005) 5921-5926.
[3] C.M. Sorensen, Aerosol Science. Technology 35 (2001) 648-687.
[4] V. Eymet, A.M. Brasil, M. El Hafi, T.L. Farias, P.J. Coelho, J. Quant. Spectrosc. Radiat. Transf. 74 (2002) 697-718.
[5] B. Yang, Ü. Ö. Köylü, J. Quant. Spectrosc. Radiat. Transf. 93 (2005) 89-295.
[6] D.R. Snelling, G.J. Smallwood, F. Liu, Ö. L. Güilder, W.D. Bachalo, Applied Optics 44 (2005) 6773-6785.
[7] S. Twomey, Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurement, Elsevier, Amsterdam, 1977, p. 122-127
[8] S. Twomey, Journal of Computational Physics 18 (1975) 188-200.
[9] A.N. Tikhonov, Y.V. Arsenin, Solutions of Ill-Posed Problems, Winston, Washington, DC (1977) 73-94.
[10] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia (1998) 99-106.
[11] A. Zellner, An Introduction to Bayesian Inference in Econometrics, Wiley and Sons, New York (1971) 10-14.
[12] W. Menke, Geophysical Data Analysis, Academic Press, San Diego (1984) 29-31, 83- 87.
[13] R. Zdunek, Computational Methods in Science Technology 13 (2007) 67-77.
[14] V.E. Johnson, Annals of Statistics 6 (2004) 2361-2384.
[15] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes, Cambridge University Press, Cambridge (2007) 736-738.
[16] K. Tian, K. Thomson, F. Liu, D.R. Snelling, G.J. Smallwood, D. Wang, Combust. Flame 144 (2006) 782-791.