THE IMPROVEMENT OF LIDAR ANALYSIS THROUGH NON-LINEAR REGRESSION

A. C. Povey[1], R. G. Grainger[1], D. M. Peters[1], J. L. Agnew[2], and D. Rees[3]

[1]*University of Oxford, Atmospheric, Oceanic, and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, United Kingdom, povey@atm.ox.ac.uk*

[2]*STFC Rutherford Appleton Laboratory, Harwell, Oxford, OX11 0QX, United Kingdom*

[3]*Hovemere Ltd., Tonbridge, TN9 1RF, United Kingdom, drees@hovemere.com*

ABSTRACT

Lidars are ideally placed to investigate the effects of aerosol and cloud on the climate system due to their unprecedented vertical and temporal resolution. Dozens of techniques have been developed in recent decades to retrieve the extinction and backscatter of atmospheric particulates in a variety of conditions. These methods, though often very successful, are fairly *ad hoc* in their construction, utilising a wide variety of approximations and assumptions that makes comparing the resulting data products with independent measurements difficult and their implementation in climate modelling virtually impossible.

As with its application to satellite retrievals, the methods of non-linear regression can improve this situation by providing a mathematical framework in which the various approximations, estimates of experimental error, and any additional knowledge of the atmosphere can be clearly defined and included in a mathematically 'optimal' retrieval method, providing rigorously derived error estimates. In addition to making it easier for scientists outside of the lidar field to understand and utilise lidar data, it also simplifies the process of moving beyond extinction and backscatter coefficients and retrieving microphysical properties of aerosols and cloud particles.

Such methods have been applied to a prototype Raman lidar system. A technique to estimate the lidar's overlap function using an analytic model of the optical system and a simple extinction profile has been developed. This is used to calibrate the system such that a retrieval of the profile extinction and backscatter coefficients can be performed using the elastic and nitrogen Raman backscatter signals.

Key words: Optimal estimation, calibration, Raman lidar.

1. INTRODUCTION

The techniques of non-linear regression are applied throughout the scientific community as a means to analyse data [1, 2]. A model is produced to describe the system under investigation in terms of a finite number of parameters, producing a simulated measurement. The set of parameters giving a simulated measurement that is most consistent with the actual measurement is then sought through some iterative scheme (most often minimising the chi-squared). Polynomial regression is a simple form of this, whereby the system is assumed to be described by a polynomial. Though not a perfect process, non-linear regression provides a widely understood means to estimate from indirect measurements a set of variables that will be physically reasonable and consistent, unlike *ad hoc* methods which can produce physically impossible values if not carefully designed and assessed.

As the amount of light observed by a lidar in the elastic channel is a function of both extinction and backscatter, the analysis of lidar data is an underconstrained problem. This is compounded when uncertainties in the calibration of the instrument and state of the atmosphere are also considered. Many means have been proposed to overcome this difficulty [3, 4, 5], but these solve the lidar equation in a deterministic manner and must resort to various smoothing or averaging techniques to minimise noise in the final product. By solving the problem in a statistical sense, non-linear regression fundamentally avoids this issue.

This abstract briefly introduces two applications of non-linear regression to lidar analysis. Firstly, the calibration of a Raman channel of a lidar system in stable conditions through the use of a relatively simple analytic formulation of the overlap function is demonstrated. This work is under review with *Applied Optics* titled 'Estimation of the lidar overlap function by non-linear regression.' Then, the retrieval of the aerosol extinction and backscatter from a Raman and the elastic channel of a lidar is outlined.

2. OPTIMAL ESTIMATION

Optimal estimation is a non-linear regression scheme which fits a set of parameters to a measurement, incorporating any prior information about the measurement system. It solves for $\mathbf{x}$ the inverse problem,

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \varepsilon, \tag{1}$$

where $\mathbf{y}$ describes a set of measurements with noise ε; the state of the observing system and atmosphere are sum-

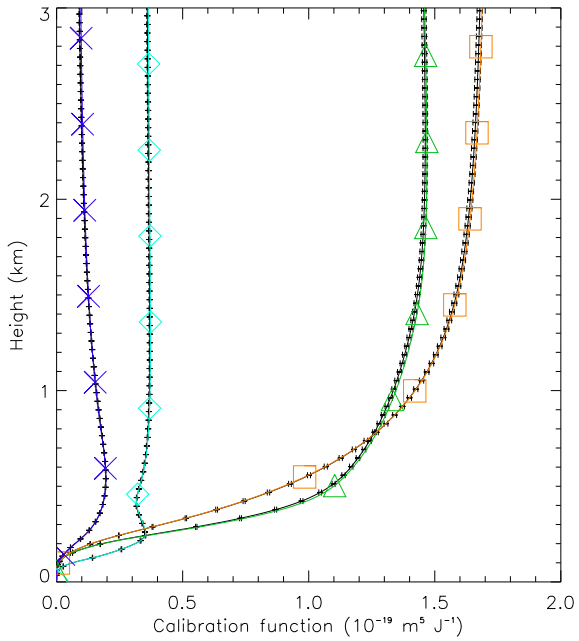

Figure 1. Retrieved $C_{\rm ra}A(r)$ for four different model alignments. True profiles are plotted in colour.

marised by unknown parameters $\mathbf{x}$ and known parameters $\mathbf{b}$; and the forward model $\mathbf{F}$ translates this state into a simulated measurement.

It is shown in [1] that the iteration,

$$\mathbf{x}_{i+1} = \mathbf{x}_i + [(1+\Gamma_i)\mathbf{S}_a^{-1} + \mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}\mathbf{K}_i]^{-1} \{\mathbf{K}_i^T\mathbf{S}_\epsilon^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{x}_i, \mathbf{b})] - \mathbf{S}_a^{-1}(\mathbf{x}_i - \mathbf{x}_a)\}, \quad (2)$$

converges to the state with maximal probability, given the measurement. The *a priori* state $\mathbf{x}_a$ gives the expected state of the system; $\mathbf{S}$ are covariance matrices describing the uncertainty on the measurement or *a priori*; Γ_i is a scale factor that is decreased after an iteration that improves the fit between $\mathbf{y}$ and $\mathbf{F}(\mathbf{x}_i, \mathbf{b})$ and is increased otherwise; and $\mathbf{K}_i = \nabla_{\mathbf{x}}\mathbf{F}(\mathbf{x}_i, \mathbf{b})$. The iteration is considered converged when the change in $\mathbf{x}$ is much less than the predicted error.

3. ESTIMATION OF THE OVERLAP FUNCTION

A common model for the energy observed by a lidar from a range r due to Raman scattering is, [6]

$$E_{\rm ra}(r) = E_0\, C_{\rm ra} \frac{A(r)}{r^2}\, N_{\rm X}(r) \exp\left[-\int_0^r \alpha_m(\lambda_L, r') + \alpha_m(\lambda_{\rm X}, r') + \alpha_a(\lambda_L, r')\left(1 + \frac{\lambda_L}{\lambda_{\rm X}}\right) {\rm d}r'\right], \quad (3)$$

where $\alpha(\lambda, r)$ is the extinction coefficient; the subscripts m and a denote molecular and aerosol scattering; λ_L and $\lambda_{\rm X}$ are the wavelengths of the laser beam and Raman scattered radiation; E_0 is the energy of the laser pulse; $N_{\rm X}(r)$ is the number density of the scattering species; and $C_{\rm ra}$ is a constant.

For estimation of the overlap function with optimal estimation, $\mathbf{y}$ consists of the profile $E_{\rm ra}(r)$ and $\mathbf{x}$ will describe the unknowns $A(r)$, $\alpha_a(r)$, and $C_{\rm ra}$. To further constrain the problem, model forms of $A(r)$ and α_a are introduced such that these profiles can be expressed in terms of a few unknown parameters.

An analytic formulation for $A(r)$ was proposed in [7].

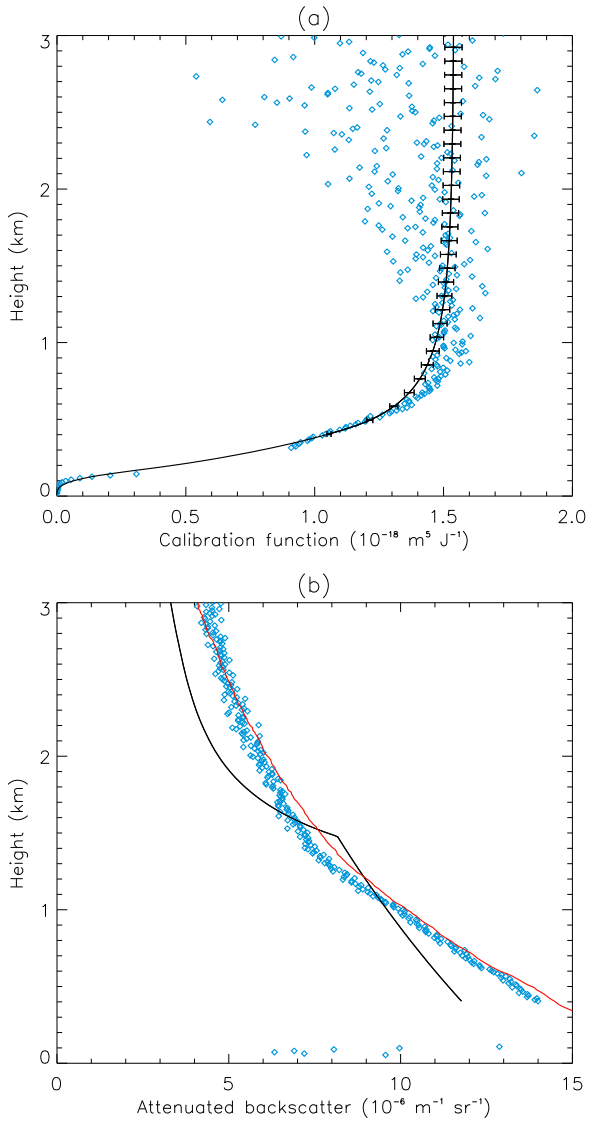

Figure 2. (a) The retrieved calibration function with errors over an arithmetic inversion of eqn. (3) using the a priori extinction profile. (b) The attenuated backscatter coefficient at 355 nm for the retrieved aerosol profile (black), the elastic profile corrected with the retrieved overlap function (blue), and as reported independently by an EZ lidar at the same site (red).

This presented the overlap function as an integral over the overlap of two circles — the assumed circular, continuous beam and the telescope's FOV. It neglects the effects of any components after the telescope and any variations in the beam profile. Some rearrangement of the form originally presented was made to improve accuracy and stability of the integration.

It is then hypothesised that for stable PBLs, extinction can be approximated as constant to some height z_0 and exponentially decaying above that over a scale height H. This is a simplification of the profile presented in [8]. This profile can then be constrained by observations of the atmosphere, such as measurement of the aerosol optical thickness, χ, with a sun photometer.

As shown in fig. 1, with simulated data the scheme successfully retrieves the calibration function (the product $C_{\mathrm{ra}}A(r)$) for a variety of hypothetical systems. Data simulated for a variety of aerosol and laser beam profiles not explicitly included in the models above could also be fit with an error of < 5 % in most cases. Degeneracy in the calibration function, whereby very similar functional forms can be obtained for very different values of $\mathbf{x}$ and $\mathbf{b}$, produces these practically useful fits. However, the retrieval may converge to a solution that is a poor fit to the data, but is a much better fit than other solutions with similar parameter values. As such, it is important that the goodness of fit is carefully assessed.

Observations over two hours of a Spring morning with $\chi = 0.13$ were evaluated. Fig. 2(a) shows the retrieved calibration function. For comparison, a simple arithmetic inversion of eqn. (3) is also shown. The retrieved profile is slightly smaller from 0.5 – 1 km than would be expected from the data without retrieval as the retrieved scale height is greater than initially guessed. The retrieval returns larger errors than observed in the simulated data. The reason for this is clear from the broad scatter of data points above 1 km. The ability to fit a physically consistent function regardless is one of optimal estimation's strengths.

The retrieved calibration function was used to approximately correct the instrument's elastic channel to determine the attenuated backscatter coefficient,

$$\beta^*(r) = \frac{E_{\mathrm{el}}(r)\, r^2}{C_{\mathrm{el}} E_0 A(r)} \exp\left[2\int_0^r \alpha_m(\lambda_L, r')\,\mathrm{d}r'\right] \quad (4)$$

$$\equiv \frac{\beta_m(\lambda_L, r) + \beta_a(\lambda_L, r)}{\exp\left[2\int_0^r \alpha_a(\lambda_L, r')\,\mathrm{d}r'\right]}. \quad (5)$$

This is compared in fig. 2(b) to both the retrieved profile and an independent measurement by a Leosphere EZ lidar operated at the same site. We see that the retrieved α_a is reasonable up to 1.5 km, but then underestimates the scale height. However, the fairly good correspondence between the published β^* and that determined from RACHEL's elastic channel (the RMS deviation between them reduces by a factor of two with the correction) gives confidence that despite the extinction profile, the retrieved correction is useful. The difference in α_a may be due to a change in the lidar ratio between the PBL and free troposphere.

4. RETRIEVAL OF EXTINCTION AND BACKSCATTER

Assuming that the retrieved calibration function does not change significantly over several days to weeks, it can be used to retrieve extinction and backscatter for other measurements. This uses eqn. (3) and, for the elastic channel,

$$E_{\mathrm{el}}(r) = E_0\, C_{\mathrm{el}} \frac{A(r)}{r^2}\, N_{\mathrm{X}}(r)\,[\beta_m(r) + \beta_a(r)] \exp\left[-2\int_0^r \alpha_m(\lambda_L, r') + \alpha_a(\lambda_L, r')\,\mathrm{d}r'\right]. \quad (6)$$

Here, $\mathbf{y}$ contains both $E_{\mathrm{el}}(r)$ and $E_{\mathrm{ra}}(r)$ and $\mathbf{x}$ gives α_a and β_a at a series of points (from which the profiles are interpolated cubically).

Development of this retrieval method is still under way,

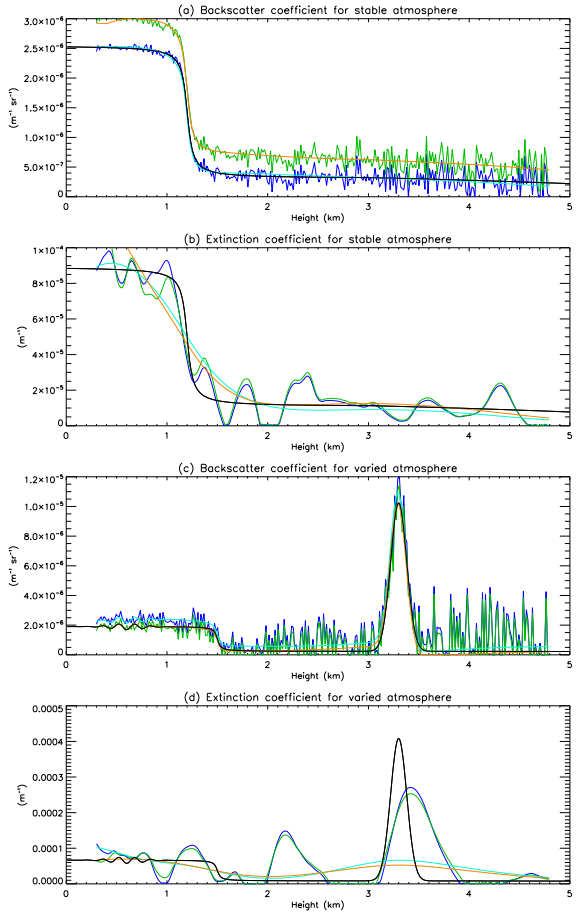

Figure 3. Retrievals of α_a and β_a from simulated data. True profiles in black with retrieval from noiseless simulations in light blue and noisy simulation in dark blue. Retrieval using an incorrect calibration plotted in orange and green.

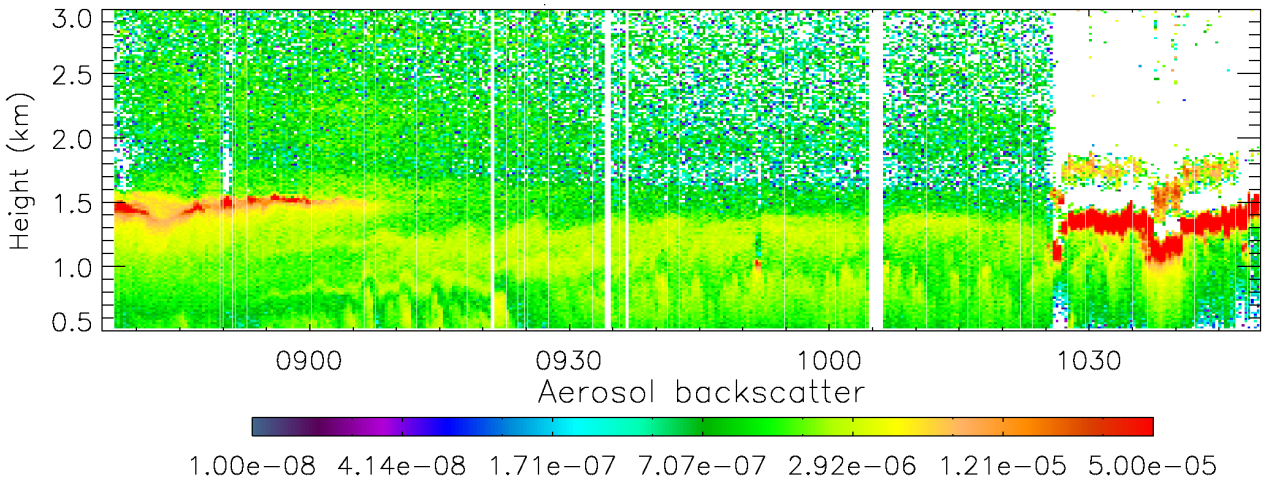

Figure 4. Aerosol backscatter on 29 Aug 2007 as derived by optimal estimation retrieval, at a resolution of 20 s and 17 m (second down).

but some early investigations are be presented here. Fig. 3 show the retrieval of extinction (plots b and d) and backscatter (plots a and c) for two different model atmospheres simulated for 5 minutes of observation. The retrieved backscatter is consistent with the true profile at all levels retrieved (errors not shown) and captures the features of the profile with acceptable noise.

The extinction product is still under development. It is clear that excessively large vertical correlations are assumed for α_a within $\mathbf{S}_a$. This acts to smooth the profile. The impact of the *a priori* upon the retrieval is under investigation.

The retrieved backscatter from observations on a Summer morning is presented in fig. 4 (10 s averages). Values below 500 m are omitted as this plot neglected the overlap function. Aerosol variations through the mixed layer are clearly visible with significantly better noise than for evaluation at this level of averaging for analysis by methods [4, 5].

ACKNOWLEDGEMENTS

Thanks to the staff of the NERC-funded CFARR for their assistance during the field work. Radiosonde data provided by the UK Meteorological Office. AERONET data available at `http://aeronet.gsfc.nasa.gov`.

REFERENCES

1. Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and practice*, volume 2 of *Series on Atmospheric, Oceanic, and Planetary Physics*. World Scientific, Singapore.

2. Gelb, A. (1974). *Applied Optimal Estimation*. MIT Press, Cambridge, Massachusetts.

3. Fernald, F., Herman, B., Reagan, J. (1972). Determination of aerosol height distributions by lidar. *Journal of Applied Meteorology*, **11**(3), pp. 482–489.

4. Klett, J. D. (1981). Stable analytical inversion solution for processing lidar returns. *Applied Optics*, **20**(2), pp. 211–220.

5. Ansmann, A., Riebesell, M., Weitkamp, C. (1990). Measurement of atmospheric aerosol extinction profiles with a Raman lidar. *Optics Letters*, **15**(13), pp. 746–748.

6. Measures, R. M. (1992). *Lidar Remote Sensing: Fundamentals and Applications*. Krieger Publishing Company, Malabar, Florida.

7. Halldórsson, T., Langerholc, J. (1978). Geometrical form factors for the lidar function. *Applied Optics*, **17**(2), pp. 240–244.

8. Steyn, D. G., Baldi, M., Hoff, R. M. (1999). The detection of mixed layer depth and entrainment zone thickness from lidar backscatter profiles. *Journal of Atmospheric and Oceanic Technology*, **16**(7), pp. 953–959.