APPLICATION OF THE CHI-SQUARED TECHNIQUE TO QUANTIFY THE AEROSOL EXTINCTION WITH A RAMAN LIDAR

Felicita Russo[1], David N. Whiteman[2], Belay Demoz[2], Raymond M. Hoff[1]
[1] *University of Maryland-Baltimore County, Baltimore, MD 21250(USA)*
[2] *NASA Goddard Space Flight Center, Greenbelt, MD 20771(USA)*
Tel: +1-301-614-6716, Email: rfeli@umbc.edu

ABSTRACT

We report on the improvement of a Raman lidar algorithm for calculating aerosol extinction. In order to calculate aerosol extinction from Raman lidar data it is necessary to evaluate the derivative of a molecular Raman signal with respect to range. The typical approach taken in the lidar community is to make an *a priori* assumption about the functional behavior of the data in order to calculate the derivative. It has previously been shown that the *a priori* assumption is not necessary and goes against the rules of statistics. Here the chi-squared test is used to eliminate the need for such assumption and is then applied to both simulated and experimental data. The detailed validation of the technique has been submitted to Applied Optics.

1. INTRODUCTION

One of the important capabilities of Raman lidar systems is the ability to retrieve range-resolved profiles of aerosol extinction [1]. Networks of lidars like EARLINET (European Aerosol Research LIdar NETwork)[2] provide datasets particularly useful for studying the influence of aerosols on climate. The Raman lidar stations in EARLINET use various different techniques to calculate aerosol extinction [3]. We will show that in general different techniques can result in significantly different uncertainties and that the technique used should be supported by statistical considerations, eliminating in this way the need for making *a priori* assumptions.

The aerosol extinction coefficient is usually calculated from the Raman lidar nitrogen signal with the equation [1]:

$$\alpha_{aer}(\lambda_L, z) = \frac{\frac{d}{dz}\ln\left(\frac{N_N(z)}{z^2P(\lambda_N, z)}\right) - \alpha_{mol}(\lambda_N, z) - \alpha_{mol}(\lambda_L, z)}{1 + \left(\frac{\lambda_L}{\lambda_N}\right)} \quad (1)$$

where $\alpha_{aer}(\lambda_L,z)$ is the aerosol extinction at the wavelength of the laser λ_L, λ_N is the wavelength of the molecular nitrogen channel, $\alpha_{mol}(\lambda_L,z)$ and $\alpha_{mol}(\lambda_N,z)$ are the molecular extinction respectively, at the laser wavelength (λ_L) and at the nitrogen Raman wavelength (λ_N), $N_N(z)$ is the number density of molecular nitrogen and $z^2P(\lambda_N,z)$ is the range-squared-corrected Raman nitrogen lidar signal. Here the Angstrom coefficient is considered equal to one.

One of the major challenges in the calculation of the aerosol extinction with Eq. 1 is to calculate the derivative term. The derivative is defined only for continuous functions but the argument of the derivative here is not a continuous function. Therefore what is traditionally done in the lidar community is to choose *a priori* a polynomial to fit to the data and calculate the derivative from the derived parameters. It has been shown [4] that the *a priori* choice of the model is not necessary and that statistical tools, such as the chi-squared test, can be used to find the function that is most likely to fit the data.

The chi-squared test is defined starting from the assumption that the data possess a Gaussian distribution. We consider the lidar signal in each range bin to be determined by a counting process according to Poisson statistics. Since the number of accumulated counts required in a given range cell in order for the Raman lidar signal to be useful for deriving aerosol extinction is much greater than 10, we can assume that the statistic pertaining to each range cell is essentially Gaussian. However, as has been described before [4], the argument of the derivative in Eq.(1) does not follow a Gaussian distribution. In order to illustrate this, a chi-squared hypothesis test has been performed on simulated data. A set of 10000 values of $N(z)$ and $z^2P(z)$ are randomly extracted from a Gaussian distribution and a Poisson distribution respectively. The hypothesis that these distributions are Gaussian is tested using the chi square test for a distribution. In both cases the hypothesis is verified with 5% confidence level. This demonstrates that $z^2P(z)$, that is extracted from a Poisson distribution with 5000 counts, is essentially Gaussian. The argument of the derivative in Eq. 1 is the ratio of $N_N(z)$ and $z^2P(z)$. The ratio of two Gaussian variables in general does not possess Gaussian statistics and the distribution of the ratio $N_N(z)/\ z^2P(\lambda_N,z)$ is at best approximately Gaussian [5]. The hypothesis that this distribution is Gaussian has been tested and was rejected at the 5% confidence level. The same test is executed on the term $Ln(N_N(z)/\ z^2P(\lambda_N,z)$. The test is rejecting the hypothesis that the distribution is Gaussian to the 5% level of confidence.

Since the form of Eq. (1) distorts the statistical distribution of the quantities to be regressed, the term containing the derivative in Eq. (1) has to be re-written in a form that preserves the statistics of both $N_N(z)$ and $z^2P(\lambda_N,z)$ [4]:

$$\frac{d}{dz}\ln\left(\frac{N_N(z)}{z^2P(\lambda_N,z)}\right)=\frac{1}{N_N(z)}\frac{dN_N(z)}{dz}-\frac{1}{z^2P(z)}\frac{d}{dz}\left[z^2P(z)\right] \quad (2)$$

If the transformation outlined in Eq. 2 is applied to Eq. 1 the terms on which the derivative is applied are $N_N(z)$ and $z^2P(\lambda_N,z)$ that follow in general a Gaussian distribution, allowing the chi-squared test to be applied for the choice of the most likely model to fit the lidar data.

2. THE CHI-SQUARED TECHNIQUE

We describe here the chi-squared technique for the calculation of the aerosol extinction from Raman nitrogen lidar data. Since the transformation described in Eq. 2 has to be used in the calculation of the aerosol extinction with this technique, the terms $N_N(z)$ and $z^2P(z)$ are regressed separately. In this illustration of the technique the random error in $N_N(z)$ is considered constant. For this reason the chi-squared test is applied to the term $z^2P(z)$. Tests not shown here indicate that the model used to regress the term $N_N(z)$ has to be of the same order as the model used to regress $z^2P(z)$. To do otherwise can increase the variability of the results and introduce small biases in the retrieved extinction. If realistic random errors for $N_N(z)$ are considered, the chi-squared test should be applied to this regression to determine which model is most likely to fit the measurements of $N_N(z)$.
In order to perform the derivatives in Eq. 2 a sliding window of 5 points was chosen. As the window slides along the profiles a linear, quadratic and cubic regression are performed on the 5 data points. For each of the three regressions a value of the chi-squared and the corresponding cumulative probability are calculated. The expected value of the chi-squared in a regression is close to the degrees of freedom of the regression. The value of the chi-square in a regression is a measure of how well the function fits the experimental data points, according with the random error in the data. In fact when the deviation between the fit polynomial and the experimental points is large with respect to the random error in the data the value of the chi squared will be larger that the expected value. On the other hand if these deviations are smaller than the random error in the data then the chi-squared will be smaller than the expected value. The values of the chi-squared cumulative probability corresponding to these situations are 1, and 0 respectively. In this application of the chi-squared test a polynomial is considered to be the most likely to fit the data if the corresponding chi-squared cumulative probability is the closest to 0.5[6]. The application of the chi-squared test on the entire profile will generate a composite extinction profile (that will be indicated by *chosen*) in which every value of the extinction is resulting from the regression over 5 points of the most likely model among the three that are tested.

3. APPLICATIONS OF THE CHI-SQUARED TECHNIQUE TO SIMULATED DATA

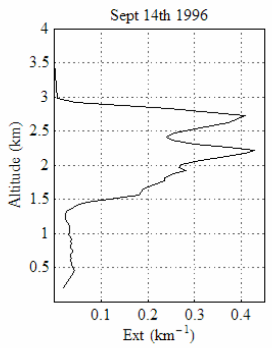

Fig. 1: Extinction profile used to simulate the Raman nitrogen signals analyzed here.

In order to examine the consequence of the use of the chi-squared technique of the calculation of the aerosol extinction and the corresponding uncertainty some simulations were performed.
A previously validated numerical model[7] was used to simulate Raman nitrogen signals as would be measured by a lidar system with the same operational characteristics as the NASA/GSFC Scanning Raman Lidar (SRL) during 1997-2000[8][9]. In this way the results obtained in the simulations could be validated by applying the chi-squared technique to the large body of experimental data that was collected in that period. Information regarding the experimental configuration of the SRL during the time of these measurements may be found in Whiteman et al. 1999[8] and Whiteman et al. 2001 [9].
For a given aerosol extinction profile in the atmosphere, a particular lidar measurement of that profile will be governed by the statistics of the measurement process. Therefore a statistically significant ensemble of simulations of a single atmospheric profile needs to be used if general conclusions are to be inferred from the results of such simulations. Previous studies[10] have been performed on idealized aerosol extinction profiles. Here we use an actual Raman lidar extinction retrieval from Sept 14, 1996, shown in Fig. 1, as input to the model. This extinction profile was used in order to generate a set of Raman nitrogen signals on which to apply the chi-squared technique. The extinction profiles retrieved with different orders of polynomials will then be compared with the profile in Fig. 1. In this illustration a set of 200 lidar profiles corresponding to an averaging time of 600 s were simulated. The retrieved extinction profiles obtained with different orders of polynomials are shown in Fig. 2, together with the corresponding extinction uncertainties and the frequency of the choice by the chi-squared test of each of the three models as most likely to fit the Raman nitrogen signal. In Fig. 2-a the input extinction profile and the extinction profiles retrieved using linear, quadratic, cubic models and the chi-squared technique

are shown. There is general agreement among the different techniques except in the altitude range between

study submitted to Applied Optics showed that for an idealized extinction profile decreasing with altitude

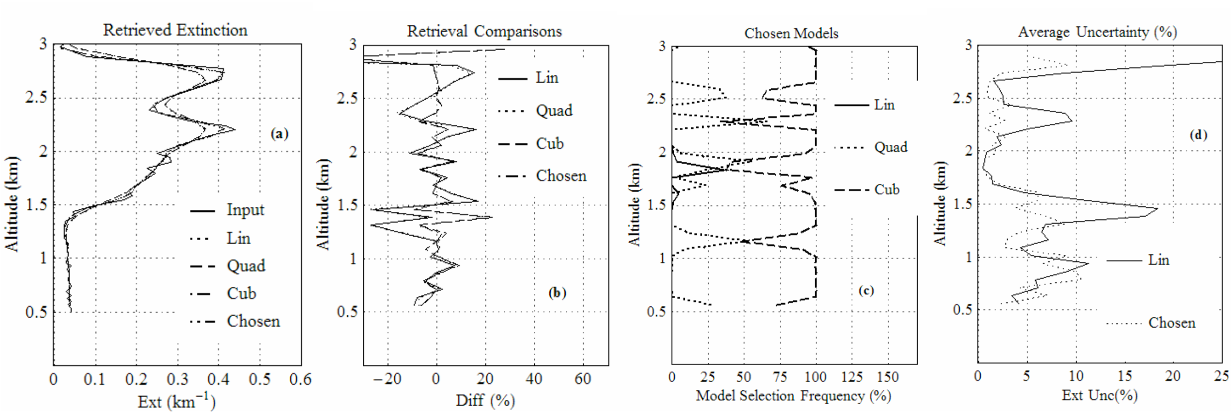

Fig.2: Results of the regression of the simulated signals. Panel (a) shows the simulated and the average retrieved extinction profiles with the different orders of polynomials. Panel (b) shows the average percentage differences between the retrieved and the simulated extinctions. Panel (c) shows the frequency of selection of the three models and panel (d) shows the extinction uncertainty obtained with the linear model and with the chi-squared technique.

2.1 and 3 km. Here in fact the linear and quadratic regressions result in extinction profiles that are not able to reproduce the rapid changes that there are in the input extinction. The cubic regression and the chi-squared technique on the other hand reproduce these features more accurately. This is due to the fact that the extinction retrievals performed with different orders of polynomials possess in general different effective spatial resolutions. Tests performed elsewhere[10] indicate that the retrieval done using a sliding window of 5 points on a Raman nitrogen signal with a spatial resolution of 75m and a linear or quadratic polynomial for the derivative in Eq. 2 possess effective spatial resolution of ~260 m. The retrieval obtained with a cubic polynomial has an effective spatial resolution of ~160 m. The effective spatial resolution of the retrieval obtained with the chi-squared technique depends on the model chosen and in general is a function of altitude, ranging from ~160 m to ~260 m. The panel 2-b shows that in general the cubic and the chosen model more accurately reproduce the input extinction (especially between 1.5 km and 2.7 km). Panel 2-c shows the frequency of choice of each model by the chi-squared test as a function of altitude in the 200 profiles dataset. Notice that the linear polynomial is the model chosen the least frequently as the most probable. Panel 2-d shows the extinction uncertainty corresponding to the retrieval with the linear least squares fit and the retrieval obtained using the chi-squared technique. The uncertainties are similar except for altitudes of around 1.4 km and around 2.3 km where the linear regression results in an uncertainty about a factor of two larger than the uncertainty of the chi-squared retrieval. The average extinction uncertainty calculated with the linear model and shown in panel 2-d is 6% while the average uncertainty of the chosen model is 4.2%. Results of the

within the boundary layer the extinction uncertainty obtained with the linear model was a factor of four larger than that obtained with the chi -the most probable depends on the shape of the retrieved extinction profile and the extinction uncertainty calculated with the chi-squared technique is on average smaller than the uncertainty obtained with the traditional technique.

4. APPLICATIONS OF THE CHI-SQUARED TECHNIQUE TO EXPERIMENTAL DATA

The chi-squared technique was applied to a large body of experimental data. The data considered are from two campaigns at DOE/ARM site in Lamont, Oklahoma held in fall 1997 (WVIOP2) and 2000 (WVIOP3)[11]. These campaigns focused on quantifying accuracies and determining limitations of atmospheric water vapor measurement technologies. Besides water vapor measurements, aerosol backscatter and extinction measurements were also acquired by the SRL but were not previously analyzed. Approximately 100 hours of extinction profiles were analyzed using the chi-squared technique. Extinction retrievals were performed on nighttime lidar measurements that were acquired over 300 seconds. From these retrieved extinctions, data with extinction uncertainties lower than 50% were selected for use in the statistics presented here. The results considered here are a subset of those studied earlier [10] in order to illustrate that the same conclusions are drawn with smaller portions of the dataset.

With the purpose of investigating the relationship between the most probable model, the aerosol extinction and the signal uncertainty, the extinction and extinction uncertainty points were combined. In Fig. 3-a the points in the dataset are displayed in 5×5 cells and among the points in one cell the most frequently chosen model is indicated. As for the case of simulated data, the linear model is chosen the least frequently in these experimental data. Moreover there is a strong preference

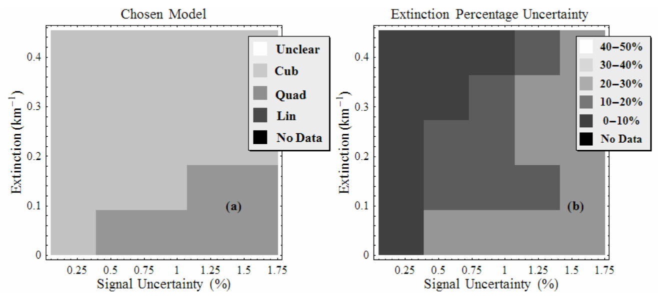

Fig. 3: Panels showing the results for data acquired with SRL during fall 1997 and 2000. Panel (a) shows the model chosen with the chi-squared technique as a function of extinction and signal uncertainty. Panel (b) shows the average extinction uncertainty as a function of the extinction and signal uncertainty.

for the cubic model (76% likelihood).

In Fig. 3-b the average extinction uncertainty is shown as a function of aerosol extinction and signal uncertainty. We can see that progressing toward the upper left, corresponding to higher extinction and smaller signal random errors, the average extinction uncertainty is always below 20%, while in general it increases progressing toward the lower right. Characterizations of the extinction uncertainty by means of these diagrams could allow evaluating the magnitude of the extinction uncertainty expected for different combinations of aerosol extinction and signal uncertainty.

5. CONCLUSIONS AND FUTURE WORK

The chi-squared technique for the calculation of the aerosol extinction with a Raman lidar has been validated. This algorithm uses the chi-squared test to choose the most probable least squares fit model as opposed to choosing one *a priori*, which is the standard method of evaluating aerosol extinction from Raman lidar data. In order to correctly apply the chi-squared test the data need to follow a Gaussian distribution. Here the distribution of the data on which the derivative is performed in the traditional form of the aerosol extinction equation was tested. To guarantee that the derivative is performed on data that follow a Gaussian distribution, the traditional equation of the aerosol extinction must be reformulated. Here simulated and experimental data were presented. However they do not include all the possible combinations of aerosol types and atmospheric conditions. Nonetheless they point toward the conclusion that the *a priori* selection of model is likely to both significantly increase the uncertainty and lower the vertical resolution of the retrieved aerosol extinction. A more detailed validation of this technique has been submitted to Applied Optics and is in the review process. We are currently applying these techniques to a comparison of the extinction retrievals from the SRL and the DOE/ ARM Raman Lidar in Northern Oklahoma.

References

1. A. Ansmann, et al, Measurements of atmospheric aerosol extinction profiles with a Raman Lidar, *Optics Letters*, 15, 746-748 (1990).

2. J. Bösenberg et al., EARLINET: a European aerosol research lidar network, in *Advances in Laser Remote Sensing*, A. Dabas, C. Loth, and J.Pelon, eds. (Editions de L'Ecole polytechnique, Palaiseau Cedex, France, 2001), pp. 155-158.

3. G. Pappalardo, et al., Aerosol lidar intercomparison in the framework of the EARLINET project. 3. Raman lidar algorithm for aerosol extinction, backscatter, and lidar ratio, *Applied Optics*, 43, 5370-5385, (2004).

4. D. N. Whiteman, Application of statistical methods to the determination of slope in lidar data, *Applied Optics* 38, 3360-3369 (1999).

5. G. Marsaglia, Ratios of normal variables and ratios of sums of uniform variables, *Journal of American Statistical Association* 60, 193-204, (1965).

6. P. R. Bevington, D.K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* 2nd Ed. (McGraw-Hill, 1992).

7. D.N. Whiteman, et al., Performance modeling of an airborne Raman water vapor lidar, *Applied Optics*,40, 375-390 (2001).

8. D. N. Whiteman, S.H. Melfi, Cloud liquid water, mean droplet radius and number density measurements using a Raman lidar, *Journal of Geophysical Research*, 104, 31,411-31,419 (1999).

9. D. N. Whiteman, et al., Raman lidar measurements of water vapor and cirrus clouds during the passage of Hurricane Bonnie, *Journal of Geophysical Research*, 106, 5211-5225 (2001).

10. F. Russo, et al., Validation of Raman lidar algorithm for quantifying aerosol extinction *Applied Optics* (submitted).

11. H.E. Revercomb, et al. The ARM program's water vapor intensive observation periods - Overview, initial accomplishments, and future challenges, *Bulletin of American Meteorological Society*, 84, 217-236 (2003).