A NEW APPROACH TO INVERTING AND DE-NOISING BACKSCATTER FROM LIDAR OBSERVATIONS

Willem Marais[1][*], Yu Hen Hu[1], Robert Holz[2], Edwin Eloranta[3]

[1] *Department of Electrical and Computer Engineering, University of Wisconsin-Madison, USA*
[2] *Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, USA*
[3] *Space Science Engineering Center, University of Wisconsin-Madison, USA*
[*] *Email:* `Willem.Marais@ssec.wisc.edu`

ABSTRACT

Atmospheric lidar observations provide a unique capability to directly observe the vertical profile of cloud and aerosol scattering properties and have proven to be an important capability for the atmospheric science community. For this reason NASA and ESA have put a major emphasis on developing both space and ground based lidar instruments. Measurement noise (solar background and detector noise) has proven to be a significant limitation and is typically reduced by temporal and vertical averaging. This approach has significant limitations as it results in significant reduction in the spatial information and can introduce biases due to the non-linear relationship between the signal and the retrieved scattering properties. This paper investigates a new approach to de-noising and retrieving cloud and aerosol backscatter properties from lidar observations that leverages a technique developed for medical imaging to de-blur and de-noise images; the accuracy is defined as the error between the true and inverted photon rates. Hence non-linear bias errors can be mitigated and spatial information can be preserved.

1. INTRODUCTION

The scatter properties of cloud and aerosol particulates can be inverted through an atmospheric lidar system, and are of great use to the atmospheric science community. NASA has invested in several lidar instruments, eg. CALIPSO, CPL and CATS. And in the near future ESA will launch EarthCare, which will have an onboard HSRL (High Spectral Resolution Lidar) instrument. This paper mainly focusses on ground based photon counting HSRL observations, with the intent of extending the pre-

sented technique for both single and multi channel ground and space based lidar instruments.

The inversion of the desired scatter properties are hindered by solar background and detector noise. Hence accurate backscatter estimates are important when these are used to constrain the inversion of the scattering volume coefficient and the backscatter phase function. The backscatter volume coefficient - or backscatter in short, is one of the scatter properties that can be inverted from HSRL photon rate observations, without making a priori assumptions about the backscatter phase function. Photon counting systems inadvertently introduce noise in the measured photon rates. This is due to the stochastic properties of photon counting detectors, such as PMTs and APDs, and the noise can be described by a Poisson point process.

The most common method to de-noise lidar observations is to simply average, which reduces the noise variance. Averaging is well suited for scenes where aerosol and cloud particulates are homogenous over horizontal layers. But there are two significant drawbacks to this approach. First, nonlinear bias errors are introduced into the retrievals whenever averaging is performed over a region of a cloud which is non-homogenous. This is especially true when the scattering volume coefficient change several orders in magnitude over a short period of time. The second drawback is the loss of spatial information. In this paper we investigate a new approach, a PMLE (Penalized Maximum Likelihood Estimator) technique, which has being used in de-noising and de-blurring medical images.

The image to the left of figure 1 shows a noisy (1 profile) and averaged (40 profiles) retrieved backscatter of a cloudy scene. Between ranges 5 and 6 km the averaged retrieval overestimate the backscatter, and again at 7.2 km. The image to the right of figure 1 shows the molecular and combined
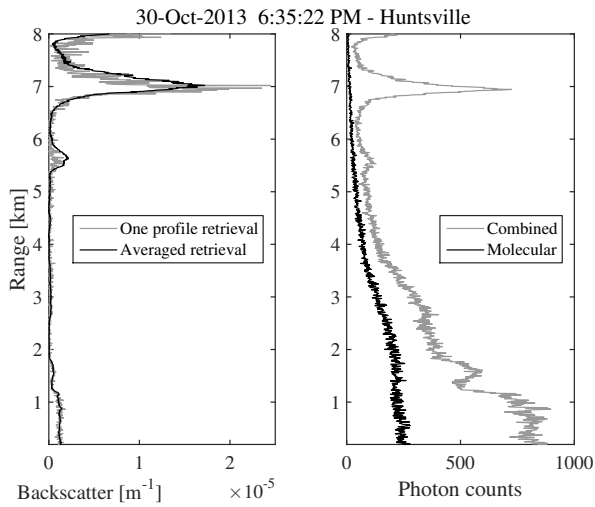
Fig. 1: The leftmost image demonstrates potential biases in the averaged backscatter volume coefficient retrieval, by comparing it to a single profile retrieval. The rightmost image shows the photon rates of the combined and molecular channels that were used for the one profile retrieval.

channel photon counts that were used to retrieve the single profile backscatter; refer to §2 for more information about these two channels. The observations were made with the Madison-WI HSRL system [2].

The PMLE technique has been used in medical imaging systems, such as confocal microscopy imaging, to de-blur and de-noise Poisson-noisy medical images [3]. As with photon counting lidar systems, some medical imaging systems also use PMTs or APDs to measure incoming photon rates. Statistical inversion techniques developed for medical imaging can also be employed in the context of photon counting lidar systems.

This paper makes a contribution by demonstrating that an accurate backscatter profile, relative to the standard retrieval, can be inverted from HSRL measured photon rates. The error criteria which is used to claim the improved accuracy is defined in §3. Both qualitative and quantitative validation approaches are used in this paper. For the qualitative approach a homogenous scene is used to estimate the backscatter, and it is compared against an averaged retrieval of the backscatter. And for the quantitative approach a simulation is used to quantify the performance of the backscatter estimates.

2. PROBLEM FORMULATION

A HSRL measure the photon rates with two distinct channels, which can be used to estimate the unknown backscatter [1]. With the molecular channel a HSRL takes advantage of molecular doppler broadening by attenuating the particulate contribution, and in effect measure the photon rate due to the molecular backscatter, with the molecular and particulate transmittances. Let $\beta \in \mathbb{R}^n_+$ denote the particulate scatter volume coefficient vector, where each vector component corresponds to a range bin; the symbol $\mathbb{R}_+$ denotes non-negative real numbers. Let $\alpha \in [0, 1]^n$ denote the backscatter phase function vector. The discretized single scatter lidar equation for the molecular channel is

$$S_m = x \cdot (\alpha \cdot \beta \cdot \theta_m + \phi_m) \cdot \exp(-2Q\beta) + b_m,$$

where $b_m \in \mathbb{R}^n_+$ is the background photon rate, $Q \in \mathbb{R}^{n \times n}_+$ is a discrete integral operator, $\phi_m \in \mathbb{R}^n_+$ represent both the molecular backscatter and its calibration term. The calibration term is due to - for example - the iodine filter spectral shape and other optical devices [1]. The factor $\theta_m \in [0, 1]$ represent the attenuation due to the iodine filter. The vector $x \in \mathbb{R}^n_+$ includes all the other parameters such as the reciprocal of the squared range, the molecular transmittance, the area of the telescope, the number of photons transmitted, etc. The symbol $\cdot$ is the point-wise multiplication of the vectors.

The photon rate of the second channel, called the combined channel, is where the particulate backscatter is not attenuated. Its single scatter lidar equation similar to that of the molecular channel.

To estimate the backscatter we can rewrite the single scatter lidar equation in a different form. Let $\iota \in \{c, m\}$ be an index holder for the molecular and combined channels. An alternative form of the single scatter lidar equation is

$$S_\iota = x \cdot (\exp(\chi) \cdot \theta_\iota + \exp(\psi) \cdot \phi_\iota) + b_\iota,$$

where $\chi \in \mathbb{R}^n$ and $\psi \in \mathbb{R}^n$. We have that

$$\exp(\chi) = \alpha \cdot \beta \cdot \exp(-2Q\beta) \text{ and } \psi = -2Q\beta.$$

The backscatter can be recovered by $\exp(\chi - \psi) = \alpha \cdot \beta$.

2.1 THE PMLE TECHNIQUE

The noise can be modeled by a Poisson PMF (Probability Mass Function), and with this we can use the MLE technique in conjunction with the single scatter lidar equations to infer the particulate backscatter $(\alpha \cdot \beta)$. This amounts to minimizing a loss function, where the minimizer of the loss function represent the estimate which best "fits" the observe photon rates. The MLE technique - in our case - is statistically ill-posed and the estimation has to be constrained; this is achieved by introducing a penalty term. For example with "optimal estimation" the loss function is the negative logarithm of the Gaussian probability density function, and the penalty term is the Euclidean norm of the parameter being estimated.

Using the rearranged single scatter lidar equation, the estimation of χ and ψ is formulated as

$$(1) \quad (\hat{\chi}_\tau, \hat{\psi}_\tau) = \underset{(\chi,\psi)\in\mathcal{B}}{\mathrm{argmin}}\, l(\chi, \psi) + \tau \mathrm{TV}(\chi, \psi),$$

where $\mathcal{B} \subset \mathbb{R}^n \times \mathbb{R}^n$ is a closed set. The function $l(\chi, \psi)$ is the MLE loss function; it is a function of the photon counts (not shown) and the parameters we want to estimate. The function TV (Total Variation) is the penalty function and the parameter $\tau \in \mathbb{R}_+$ is the regularizer parameter; these are discussed in the next paragraph. The estimates of χ and ψ are expressed as $\hat{\chi}_\tau$ and $\hat{\psi}_\tau$, where each one is function of regularizer parameter τ.

A reasonable penalty function, that has been used in medical image de-blurring and de-noising, is the so called TV penalty [3]. It is the absolute value norm of the discrete derivative. With a large regularizer parameter ($\tau \gg 1$) the discrete derivative is constrained to be small, hence the parameter being penalized is constrained to be smooth. If the regularizer parameter is close to being zero, the discrete derivative is not constrained to be small and penalized parameter is therefore not necessarily smooth. The advantages of using the TV penalty, is 1) that allows discontinuities in the parameter being estimated and 2) it is convex.

To thus far we found - empirically - that the loss function with the rearranged model has one solution, with both simulated data and actual HSRL observations.

3. METHODOLOGY

A variant of the SpaRSA algorithm called SPIRAL is used to find $(\hat{\chi}_\tau, \hat{\psi}_\tau)$ via (1) [3]. The algorithm uses sequential quadratic approximations of the loss function to find a minimizer. In each sequence of the quadratic approximation a sub-optimization problem is solved. The sub-optimization problem is the linear combination of the 2nd order Taylor series approximate of the loss function and the TV penalty function. The parameters are constrained on closed set, thus the algorithm will converge to a minimizer.

A model selection technique is be used to choose the optimum regularizer parameter τ [3, p. 162]. This is achieved by splitting the photon counts profile into two profiles using a technique called "Poisson thinning". The first profile is used to find $(\hat{\chi}_\tau, \hat{\psi}_\tau)$ for every τ, and the second profile is used to find the τ by minimizing a test loss function. It can be shown experimentally that with the described technique minimizes the error criterion

$$(2) \quad \left\|\sqrt{\hat{S}_c} - \sqrt{S_c^*}\right\|_2^2 + \left\|\sqrt{\hat{S}_m} - \sqrt{S_m^*}\right\|_2^2,$$

where S_c^* and S_m^* are the true photon rates of the corresponding channels. The error criteria (2) gives an indication of how "close" the inverted photon rates are relative to the true photon rates.

4. RESULTS AND CONCLUSION

Both qualitative and quantitative approaches are used to validate backscatter volume coefficient estimates. The qualitative approach is done using HSRL observation of the Madison-WI HSRL system. The quantitative approach is based on simulated HSRL observations.

For the qualitative validation approach two cases are considered, where a homogenous daytime scene is used. In the first case a single profile is used to compare the backscatter retrieval and inversion. And in the second case 100 accumulated profiles are used to compare the backscatter retrieval and inversion. For the second case intuitively the backscatter retrieval should be accurate, since the noise variance decrease relatively to the number of samples using for averaging.

The leftmost image of figure 2 shows the inverted and retrieve backscatter of the first case. The chosen scene has fairly homogenous horizontal layers between ranges 0 and 3 km. The rightmost image of figure 2 shows the inverted and retrieved backscatter of the second case. In this case both
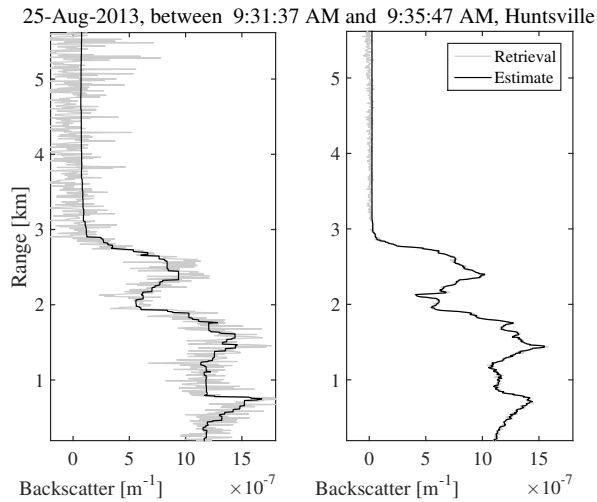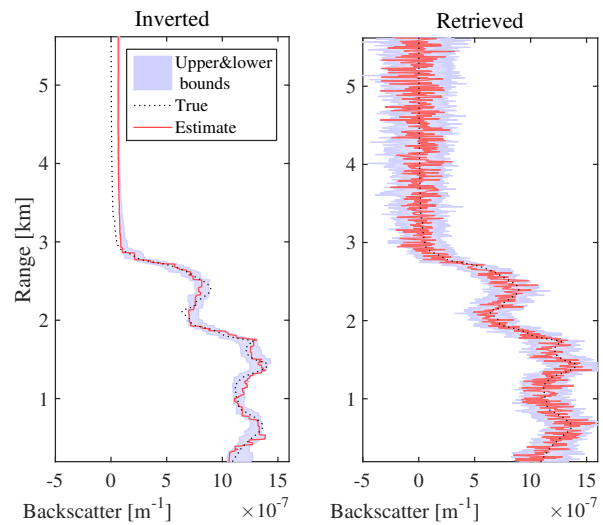

Fig. 2: The leftmost image compares the backscatter retrieval and estimate of one profile. In the rightmost image shows the retrieval and estimate of 100 accumulated profiles.

estimates of the backscatter are nearly in full agreement between ranges 0 and 3 km, while the inverted backscatter is overestimating the backscatter between ranges 3 and 5.6 km.

For the quantitative validation approach the simulated HSRL profile has an aerosol layer similar to figure 2. The dotted curves in figure 3 shows the true backscatter that is used in the simulated HSRL observations. The continuous curve of the left images shows an example of inverted backscatter profile, and the continuous curve of the right image is a retrieved backscatter profile. In order to evaluate the bias and variance of the backscatter estimates, the simulated HSRL profile is recreated several times and of each the backscatter is inverted and retrieved. The light filled curve in the left most image of figure 3 shows the upper and lower bounds of the inverted backscatter; 18 recreated HSRL simulations are used in this case. The image to the right of figure 3 shows the corresponding upper and lower bounds of the retrieved backscatter.

From figure 3 we deduce that the inverted

Fig. 3: The leftmost image shows inverted backscatter profile, in conjunction with the true backscatter profile. The rightmost image shows correpsonding retrieved backscatter profile.

backscatter profile is somewhat biased in comparison to the retrieved backscatter profile. But variance of inverted backscatter profile is significantly smaller than retrieved backscatter profile variance. The mean error in the simulations, according to the definition (2), of the proposed estimation procedure is 54.48. And mean error using standard averaging approach is 769.45. The results indicate that the potential exists for inverting the backscatter volume coefficient reliably, using one or very few profiles, with HSRL observations. Further research is required to 1) derive confidence bounds of the backscatter estimate and 2) determine how to extend the technique for single channel lidar systems.

REFERENCES

[1] E. Eloranta, "High Spectral Resolution Lidar Measurements of Atmospheric Extinction: Progress and Challenges," in *Aerospace Conference, 2014 IEEE*. IEEE, 2014, pp. 1–6.

[2] ——. (2014, 10) http://hsrl.ssec.wisc.edu. [Online]. Available: http://hsrl.ssec.wisc.edu

[3] Z. T. Harmany, "Computational optical imaging systems: Sensing strategies, optimization methods, and performance bounds," Ph.D. dissertation, Duke University, 2012.