

# Iterative algorithms for a non-linear inverse problem in atmospheric lidar

Giulia Denevi<sup>1,3</sup>, Sara Garbarino<sup>2</sup> and Alberto Sorrentino<sup>3,4</sup>

<sup>1</sup> Istituto Italiano di Tecnologia, Genova

<sup>2</sup> Centre for Medical Image Computing, Department of Computer Science, University College London, London

<sup>3</sup> Dipartimento di Matematica, Università di Genova, Genova

<sup>4</sup> CNR-SPIN, Genova

**Abstract.** We consider the inverse problem of retrieving aerosol extinction coefficients from Raman lidar measurements. In this problem the unknown and the data are related through the exponential of a linear operator, the unknown is non-negative and the data follow the Poisson distribution. Standard methods work on the log-transformed data and solve the resulting linear inverse problem, but neglect to take into account the noise statistics. In this study we show that proper modelling of the noise distribution can improve substantially the quality of the reconstructed extinction profiles. To achieve this goal, we consider the non-linear inverse problem with non-negativity constraint, and propose two iterative algorithms derived using the Karush-Kuhn-Tucker conditions. We validate the algorithms with synthetic and experimental data. As expected, the proposed algorithms out-perform standard methods in terms of sensitivity to noise and reliability of the estimated profile.

## 1. Introduction

Lidars are widely used instruments that collect information on remote objects by illuminating them with a laser and then measuring the amount of backscattered light and the corresponding traveling time. Lidars have many different applications, including ecosystem mapping [1], navigation in urban environments [2] and stratigraphic modelling in geology [3]. In this study we are concerned with the atmospheric application of lidar: here the laser is headed towards the sky, and the backscattered light carries information about the optical parameters of the atmosphere, namely extinction and backscattering coefficient, as functions of the height [4]. These, in turn, can be used to infer the presence and composition of clouds and aerosols at various altitudes, in order to characterize pollution or the atmospheric consequences of disruptive events [5, 6, 7].

More specifically, in this study we focus on the estimation of the extinction coefficient from so-called *Raman* lidar data [8], where part of the backscattered light arrives at a different wavelength than that emitted by the laser, as a consequence of *inelastic* scattering; this allows us to assume knowledge of the backscattering coefficient,

which is unknown for *elastic* scattering. As a consequence, the noiseless lidar equation can be written as

$$P(z) = d(z)e^{-\int_0^z \alpha(x)dx}, \quad (1)$$

where  $P(z)$  is the power of backscattered light from altitude  $z$ ,  $d(z)$  is a known function and  $\alpha(x)$  is the unknown extinction coefficient, which can be assumed to be non-negative.

In the last decade, several methods have been proposed for the solution of this problem, including Tikhonov regularization [9, 10], Levenberg–Marquardt [11] and Richardson–Lucy [12]. These approaches have represented a substantial improvement with respect to previously used restoration techniques, which consisted in computing the numerical derivative of the smoothed data [13]: first, because they explicitly cast the problem as an inverse problem and use explicit regularization (Tikhonov) or an early stopping technique (Levenberg–Marquardt and Richardson–Lucy) to avoid excessive backprojection of noise on the estimated extinction profile; second – this only holds true for Richardson–Lucy, although it could be easily introduced in Levenberg–Marquardt – because they impose a non-negativity constraint, which further helps to reduce instability. On the other hand, to the best of our knowledge, there has been no attempt so far to properly account for noise statistics in the inversion procedure. Indeed, all the proposed methods work on the log-transformed problem

$$y(z) := \log\left(\frac{d(z)}{P(z)}\right) = \int_0^z \alpha(x)dx \quad (2)$$

and noise on  $y(z)$  is treated either as Gaussian with constant variance (in plain Tikhonov and Levenberg–Marquardt), or as Poisson (in Richardson–Lucy). However, neither of these is a realistic assumption: in general, it is reasonable to assume that  $P(z)$  is Poisson, and therefore  $y(z)$  is not; neither it is reasonable to assume that the variance of  $y(z)$  is the same at all altitudes.

In order to account for the noise statistics, in this study we solve the inverse problem in its non-linear form. We consider two frameworks, one being early stopped maximum likelihood, the other one with an explicit  $\ell^2$  regularization term. Specifically, we derive two iterative algorithms by using the Karush-Kuhn-Tucker conditions [14], so as to include a non-negativity constraint for the solution. We use synthetic and experimental data to show that the extinction profiles recovered by the proposed algorithms are substantially better than those recovered by state-of-the-art methods. In order to further investigate the effects of modelling noise statistics, we also apply a weighted Tikhonov algorithm to the log-transformed problem and show that this small modification (as opposed to plain Tikhonov) does reduce the instability, even though is not as effective as the newly proposed algorithms.

The plan of the paper is as follows: in Section 2 we provide a brief introduction to the lidar inverse problem; in Section 3 the proposed algorithms are derived and their properties are described; in Section 4 we present the numerical experiments on synthetic

data and in Section 5 an example of application to experimental data. Our conclusions are offered in Section 6.

## 2. The Raman lidar equation

In the atmospheric application of lidar technology, a laser emitted at wavelength  $\lambda$  is headed towards the sky; when the backscattered light is measured at the same wavelength, as a consequence of *elastic* scattering, the signal  $P_\lambda(z)$  from altitude  $z$  can be expressed as

$$P_\lambda(z) = \frac{C_\lambda}{z^2} \beta_\lambda(z) e^{-2 \int_0^z \alpha_\lambda(x) dx} . \quad (3)$$

Here  $C_\lambda$  is a known constant accounting for the laser optical power and the overall detection efficiency, the term  $\frac{1}{z^2}$  is the signature of standard signal decay from the backscattering altitude  $z$  and  $\beta_\lambda(z)$  and  $\alpha_\lambda(z)$  are the (non-negative) backscattering and extinction coefficient, respectively. In this equation both  $\alpha_\lambda(z)$  and  $\beta_\lambda(z)$  have to be regarded as unknown. However, the backscattered light can also be measured at a different wavelength  $\mu$ , as a consequence of inelastic (Raman) scattering. In this case, the backscattering can be entirely addressed to the *molecular* contribution [8], which can be assumed to be proportional to the molecular density profile of the atmosphere  $\rho(z)$ , and the signal takes the form

$$P_{\mu|\lambda}(z) = \frac{C_\mu}{z^2} \rho(z) e^{-\int_0^z (\alpha_\mu(x) + \alpha_\lambda(x)) dx} , \quad (4)$$

where  $C_\mu$  includes laser optical power, detection efficiency and Raman scattering cross-section and  $\alpha_\mu(z)$  is the extinction coefficient at the backscatter wavelength  $\mu$ . The advantage of Raman equation is evident: since  $\rho(z)$  can be determined from the Rayleigh theory of scattering, the total extinction coefficient  $\alpha(z) := \alpha_\lambda(z) + \alpha_\mu(z)$  can be estimated. Furthermore, the relation between the particles' contribution at the two different wavelengths can be derived from Mie scattering theory [4], allowing one to disentangle the two different contributions, just by their sum. Most modern lidar systems have both elastic and vibrational Raman channels; therefore, in this study we focus on estimating  $\alpha(z)$  from Raman measurements (the estimate of  $\beta_\lambda(z)$  from elastic measurements, once  $\alpha_\lambda(z)$  is known, is straightforward).

As a first step, we conveniently rewrite equation (4) as

$$P(z) = d(z) e^{-\int_0^z \alpha(x) dx} , \quad (5)$$

where subscripts  $\lambda, \mu$  are omitted but implied, and the function  $d(z)$  is defined as

$$d(z) := \frac{C_\mu}{z^2} \rho(z) > 0. \quad (6)$$

We now discretize equation (5) by introducing two (a priori independent) sets of points  $\{z_i\}_{i=1,\dots,N}$  and  $\{x_j\}_{j=1,\dots,M}$ , and defining:

- the noise-free measurement vector  $\mathbf{P}^*$  such that  $P_i^* := P(z_i)$ ,  $i = 1, \dots, N$ ;
- the known system function vector  $\mathbf{d}$  such that  $d_i := d(z_i)$ ,  $i = 1, \dots, N$ ;
- the unknown exact solution vector  $\boldsymbol{\alpha}^*$  such that  $\alpha_j^* := \alpha(x_j)$ ,  $j = 1, \dots, M$ ;
- the linear operator  $\mathbf{L}$ , discretizing the integral.

Therefore, we rewrite (5) in the following discrete compact form

$$\mathbf{P}^* = \mathbf{d} \odot e^{-\mathbf{L}\boldsymbol{\alpha}^*}, \quad (7)$$

where  $\odot$  is the Hadamard product. From here on, we assume that the noisy data  $\mathbf{P}$  have a Poisson distribution, whose mean value is given by the noise-free signal  $\mathbf{P}^*$  (7). In addition, we assume that noise values at different altitudes are independent, so that the likelihood of a given (permissible) extinction profile  $\boldsymbol{\alpha}$  is given by

$$p(\mathbf{P}|\boldsymbol{\alpha}) = \prod_{i=1}^N p(P_i|\boldsymbol{\alpha}) = \prod_{i=1}^N e^{-\left(d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i}\right)} \frac{\left(d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i}\right)^{P_i}}{P_i!} \quad (8)$$

and its domain of definition, i.e. the set of the permissible objects, is the non-negative orthant. Therefore, our inverse problem is the one of retrieving a non-negative estimate of the exact solution  $\boldsymbol{\alpha}^*$ , given the recorded signal  $\mathbf{P}$ .

### 3. Iterative algorithms for the inverse problem

In this Section we introduce two inversion algorithms to solve the lidar inverse problem, by taking into account the Poisson statistics of the recorded signal  $\mathbf{P}$ . The first one belongs to the class of the early stopped maximum likelihood methods; in this context, it is often more convenient to maximize the log-likelihood instead of the likelihood itself, i.e.:

$$l(\boldsymbol{\alpha}) := \log(p(\mathbf{P}|\boldsymbol{\alpha})) = \sum_{i=1}^N \left( P_i \log(d_i) - (\mathbf{L}\boldsymbol{\alpha})_i P_i - d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} - \log(P_i!) \right), \quad (9)$$

where the algorithm is stopped at an appropriate iteration, before reaching convergence, in order to prevent excessive backprojection of noise onto the reconstruction. A second strategy to solve the inverse problem is to supplement the log-likelihood function with a proper penalty term (*prior* in Bayesian terms), accounting for the regularization; the most natural choice in our application is to add an  $\ell^2$  penalty term weighed by a regularization parameter  $\gamma > 0$ , and then devise an algorithm to maximize the penalized log-likelihood

$$S(\boldsymbol{\alpha}) := l(\boldsymbol{\alpha}) - \gamma \|\boldsymbol{\alpha}\|_2^2. \quad (10)$$

In the following, we use the Karush-Kuhn-Tucker (KKT) conditions to derive two iterative algorithms that maximize the log-likelihood (9) and the penalized log-likelihood (10), respectively.

### 3.1. KKT: Maximum Likelihood for Poisson data

The first iterative algorithm we propose aims at maximizing (9). Before proceeding with the derivation of the algorithm, we first prove that function (9) has a unique maximum in its domain of definition.

*Existence and uniqueness of the maximum.* The maximum of the log-likelihood exists and is unique, provided that  $N \geq M$ ,  $\mathbf{L}$  is full rank and  $P_N > 0$ . By elementary computations we obtain that the gradient and the Hessian matrix of  $l(\boldsymbol{\alpha})$  are given by

$$(\nabla l)_j(\boldsymbol{\alpha}) = \frac{\partial l(\boldsymbol{\alpha})}{\partial \alpha_j} = \sum_{i=1}^N \left( L_{i,j} (d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} - P_i) \right), \quad (11)$$

and

$$[\mathbf{H}(\boldsymbol{\alpha})]_{k,j} = \frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \alpha_k \partial \alpha_j} = - \sum_{i=1}^N \left( L_{i,j} L_{i,k} d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} \right), \quad (12)$$

for  $j, k = 1, \dots, M$ .

From equation (12), it follows that, for any vector  $\mathbf{v} \in \mathbb{R}^N$  the relation

$$\mathbf{v}^T \mathbf{H}(\boldsymbol{\alpha}) \mathbf{v} = \sum_{k,j=1}^M \left( v_k [\mathbf{H}(\boldsymbol{\alpha})]_{k,j} v_j \right) = - \sum_{i=1}^N \left( d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} (\mathbf{L}\mathbf{v})_i^2 \right) \quad (13)$$

holds. Since  $d_i > 0 \forall i = 1, \dots, N$ , assuming that  $N \geq M$  and  $\mathbf{L}$  is full rank, we have  $\mathbf{v}^T \mathbf{H}(\boldsymbol{\alpha}) \mathbf{v} < 0 \forall \mathbf{v} \neq \mathbf{0}$ ; therefore the Hessian matrix is negative definite, implying  $l(\boldsymbol{\alpha})$  is strictly concave. We now observe that the likelihood can be written in the following way, up to constant terms:

$$\begin{aligned} l(\boldsymbol{\alpha}) &\sim - \sum_{i=1}^N \left( (\mathbf{L}\boldsymbol{\alpha})_i P_i + d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} \right) \\ &= - \left( \sum_{j=1}^M \left( (\mathbf{L}^T \mathbf{P})_j \alpha_j \right) + \sum_{i=1}^N \left( d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} \right) \right) \end{aligned} \quad (14)$$

and it is defined on the non-negative orthant ( $\boldsymbol{\alpha} \geq \mathbf{0}$ ).

From here on, we assume that the last component  $P_N$  of the data  $\mathbf{P}$  is non-zero (otherwise, it is sufficient to cut the data at the last non-zero component). In addition,  $\mathbf{L}$  discretizes the cumulative integral from the first height, therefore  $\mathbf{L}^T$  can be thought of as a cumulative, weighted sum, from the last height. These two facts guarantee that  $(\mathbf{L}^T \mathbf{P})_j$  never vanishes (it is actually always positive). We now observe that

$$0 < \sum_{i=1}^N \left( d_i e^{-(\mathbf{L}\boldsymbol{\alpha})_i} \right) \leq \sum_{i=1}^N d_i; \quad (15)$$

moreover, since  $(\mathbf{L}^T \mathbf{P})_j \neq 0 \forall j = 1, \dots, M$ , if we take a vector  $\boldsymbol{\alpha}$  such that  $\alpha_j \rightarrow +\infty$ , for any  $j$ , we have  $\sum_{j=1}^M \left( (\mathbf{L}^T \mathbf{P})_j \alpha_j \right) \rightarrow +\infty$ . Therefore, thanks to the first inequality in (15), we have

$$\lim_{\|\boldsymbol{\alpha}\| \rightarrow +\infty} l(\boldsymbol{\alpha}) = -\infty. \quad (16)$$

This fact, together with the strict concavity, guarantees that the likelihood admits a unique global maximum point (see Lemma 2 in [15]).

*Derivation of the algorithm.* In general, the KKT conditions are first-order necessary conditions for optimality under given constraints. In our context, for the maximum  $\tilde{\boldsymbol{\alpha}}$  of  $l(\boldsymbol{\alpha})$ , they can be expressed as

$$\begin{cases} \alpha_j \frac{\partial l(\boldsymbol{\alpha})}{\partial \alpha_j} \Big|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}} = 0 & j = 1, \dots, M \\ \frac{\partial l(\boldsymbol{\alpha})}{\partial \alpha_j} \Big|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}} \leq 0 & \text{if } \tilde{\alpha}_j = 0. \end{cases} \quad (17)$$

Importantly, since  $l(\boldsymbol{\alpha})$  is concave, in our case the KKT conditions are not only necessary but also sufficient conditions for an object  $\tilde{\boldsymbol{\alpha}}$  to be the maximum of  $l(\boldsymbol{\alpha})$ .

By combining equation (11) with equations (17) we obtain that the maximum of  $l(\boldsymbol{\alpha})$  must be solution of the following non-linear equation

$$\tilde{\alpha}_j = \frac{(\mathbf{L}^T(\mathbf{d} \odot e^{-\mathbf{L}\tilde{\boldsymbol{\alpha}}}))_j}{(\mathbf{L}^T\mathbf{P})_j} \tilde{\alpha}_j \quad j = 1, \dots, M, \quad (18)$$

where, as pointed out above, the denominator is always positive. Let us now introduce the non-linear operator  $\mathbf{T} : \mathbb{R}^N \rightarrow \mathbb{R}^N$

$$\mathbf{T}(\boldsymbol{\alpha}) := \frac{\mathbf{L}^T(\mathbf{d} \odot e^{-\mathbf{L}\boldsymbol{\alpha}})}{\mathbf{L}^T\mathbf{P}} \odot \boldsymbol{\alpha}, \quad (19)$$

where the division has to be intended as the inverse of the Hadamard product, replaced with the usual division symbol for the sake of readability. Therefore, it follows that the maximum of  $l(\boldsymbol{\alpha})$  must be fixed point of  $\mathbf{T}$ , i.e.

$$\tilde{\boldsymbol{\alpha}} = \mathbf{T}(\tilde{\boldsymbol{\alpha}}). \quad (20)$$

By applying the method of successive approximations (as done, for instance, in [16]), to this fixed point equation, we obtain the following iterative algorithm for the maximization of  $l(\boldsymbol{\alpha})$

$$\boldsymbol{\alpha}^{(n+1)} := \mathbf{T}(\boldsymbol{\alpha}^{(n)}) = \frac{\mathbf{L}^T(\mathbf{d} \odot e^{-\mathbf{L}\boldsymbol{\alpha}^{(n)}})}{\mathbf{L}^T\mathbf{P}} \odot \boldsymbol{\alpha}^{(n)}. \quad (21)$$

We notice that there are no obvious guarantees that algorithm (21) will converge; however, it is possible to interpret (21) as follows

$$\boldsymbol{\alpha}^{(n+1)} = \boldsymbol{\alpha}^{(n)} + \mathbf{D}(\boldsymbol{\alpha}^{(n)})\nabla l(\boldsymbol{\alpha}^{(n)}). \quad (22)$$

Now this is a scaled gradient algorithm with a constant and unitary step size, and a diagonal scaling matrix  $\mathbf{D}(\boldsymbol{\alpha}) := \text{diag}\left(\frac{\boldsymbol{\alpha}}{\mathbf{L}^T\mathbf{P}}\right)$ . For this class of algorithms convergence is guaranteed by introducing a line search procedure [17, 18]; hence, in the numerical simulations below we have utilized the Armijo rule [19], designed to respect the positivity constraint.

Even though the numerical results discussed below have been obtained with the scaled

gradient algorithm (22), we provide here some additional comments on the multiplicative algorithm (21).

**Remark 1.** If all the components of the initial guess  $\boldsymbol{\alpha}^{(0)}$  are (strictly) positive, then such property is preserved by each  $\boldsymbol{\alpha}^{(n)}$ ,  $n > 0$ . As a consequence,  $\boldsymbol{\alpha}^{(n)}$  is a fixed point of the operator  $\mathbf{T}$  only if it is the maximum of the log-likelihood function.

**Remark 2.** If the iterations of the multiplicative algorithm (21) converge, they converge to the maximum (i.e. the limit point satisfies both the KKT conditions (17)); this is a consequence of the concavity of  $l(\boldsymbol{\alpha})$ . In fact, in all the numerical experiments performed for this study we did observe numerical convergence for the multiplicative algorithm (21); these results are not shown in this article.

**Remark 3.** Algorithm (21) has been described in [15] for an imaging problem in Computerized Tomography,  $\mathbf{L}$  being the Radon Transform rather than the cumulative integral. The authors discuss the convergence properties of the algorithm. By assuming the existence and the uniqueness of an optimal point  $\tilde{\boldsymbol{\alpha}}$  satisfying the strict constraint  $\tilde{\alpha}_j > 0 \forall j = 1, \dots, M$ , they show that local convergence is guaranteed if  $(\mathbf{L}\tilde{\boldsymbol{\alpha}})_i < 1$ ,  $\forall i = 1, \dots, N$ .

As a final comment, we recall that in this study the iterations will not be performed until convergence. Indeed, due to the ill-posedness of the problem, the maximum likelihood estimate would not provide a sensible reconstruction of the unknown. Even though the theoretical properties of early stopping are still matter of debate, heuristics suggest that early stopping can have a regularizing effect on the solution, and our numerical simulations below confirm this conjecture.

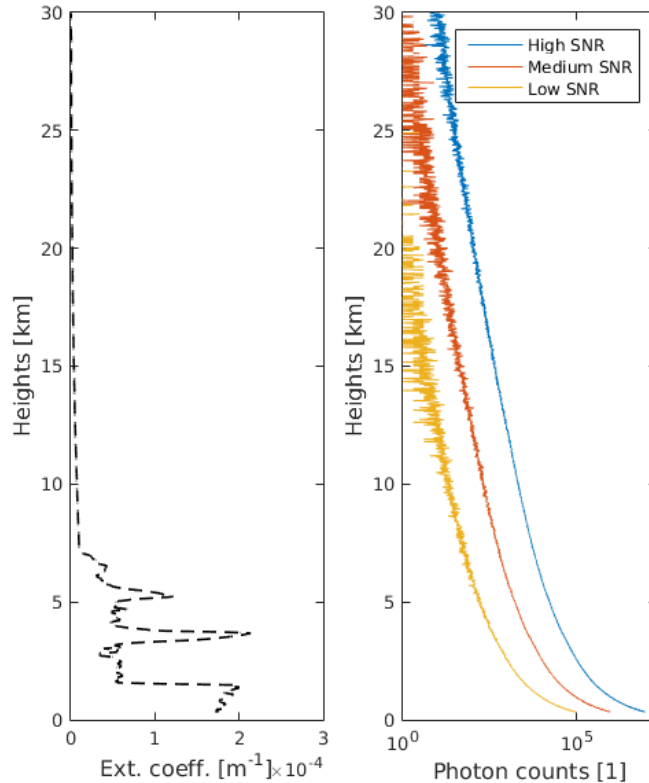
### 3.2. KKT-L2: Penalized maximum likelihood for Poisson data

We now consider the maximization of the penalized log-likelihood (10). Since we adopt the very same strategy used for the maximization of (9) we are not going to repeat the derivation here. In fact, thanks to the regularizing term, function (10) is strictly concave even without the requirements  $N \geq M$  and  $\mathbf{L}$  full rank; in addition, it has the same limiting behavior (16) of the log-likelihood  $l(\boldsymbol{\alpha})$ ; consequently, it admits a unique global maximum in its domain. By writing down the KKT conditions with the non-negativity constraint, a similar iterative algorithm is derived, which reads as follows:

$$\boldsymbol{\alpha}^{(n+1)} = \frac{\mathbf{L}^T(\mathbf{d} \odot e^{-\mathbf{L}\boldsymbol{\alpha}^{(n)}})}{\mathbf{L}^T\mathbf{P} + 2\gamma\boldsymbol{\alpha}^{(n)}} \odot \boldsymbol{\alpha}^{(n)}, \quad (23)$$

where  $\gamma$  is the regularization parameter. Again, we interpret the algorithm as a scaled gradient one and apply the line search procedure to ensure convergence. We notice that the considerations expressed in Remarks 1 and 2 hold also for the multiplicative version of this algorithm. In addition, we remark the following facts.

**Remark 4.** Since regularization is obtained here by introduction of a regularizing term in the functional, there is no need to stop the iterations in advance. In the numerical sections, we will perform 200 iterations, and always observe convergence of the algorithm.



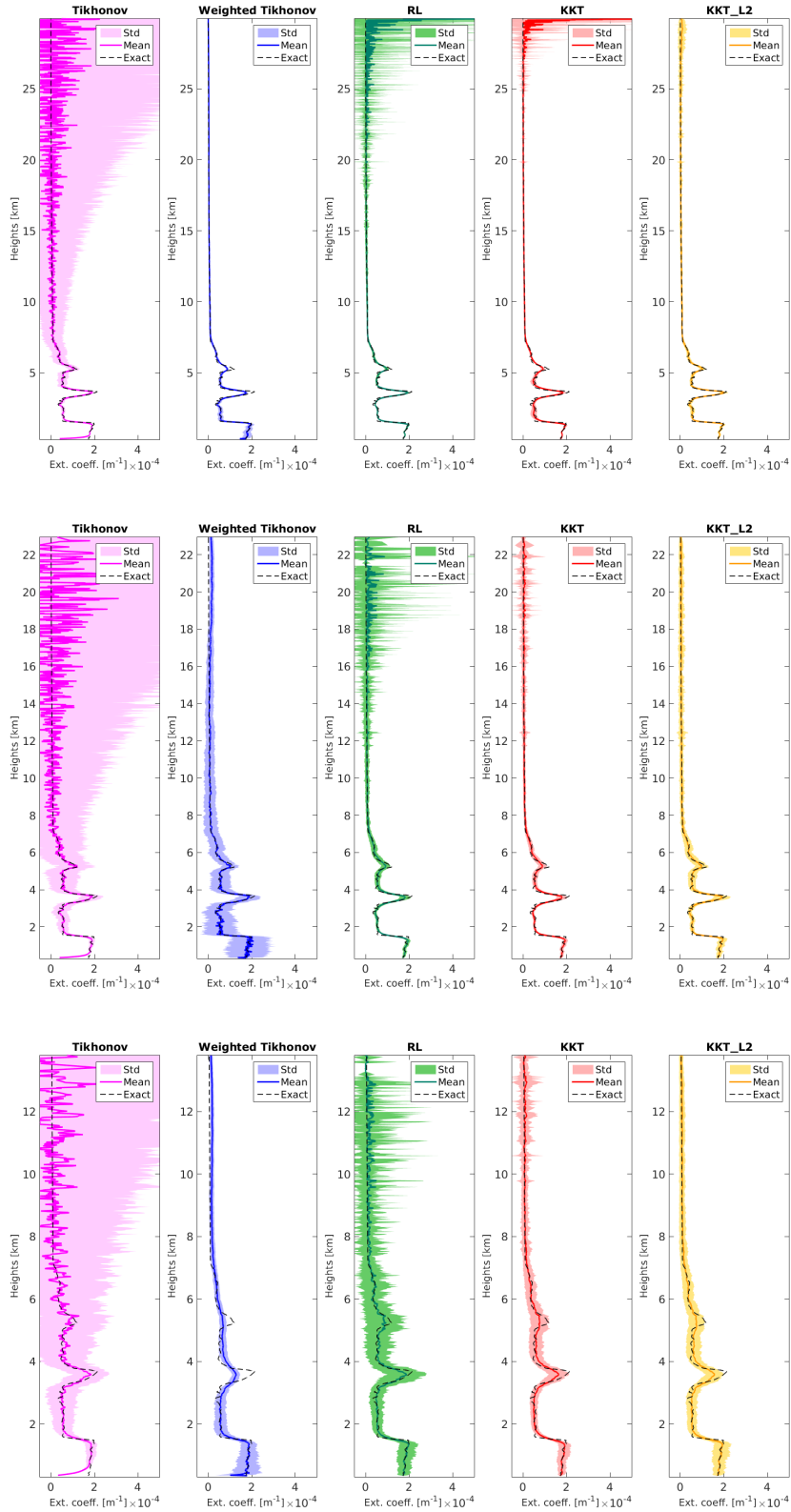
**Figure 1.** Left panel: the true extinction profile for Simulation 1. Right panel: synthetic data with Poisson noise at three different SNRs (logarithmic scale on the  $x$  axis).

**Remark 5.** In standard Tikhonov regularization, the regularization parameter can be interpreted as the ratio between the variance of the prior and the variance of the likelihood function. In our model, since the likelihood is Poisson, the regularization parameter is interpretable as the inverse variance of the prior distribution only. Therefore the optimal regularization parameter should not depend on the signal-to-noise ratio.

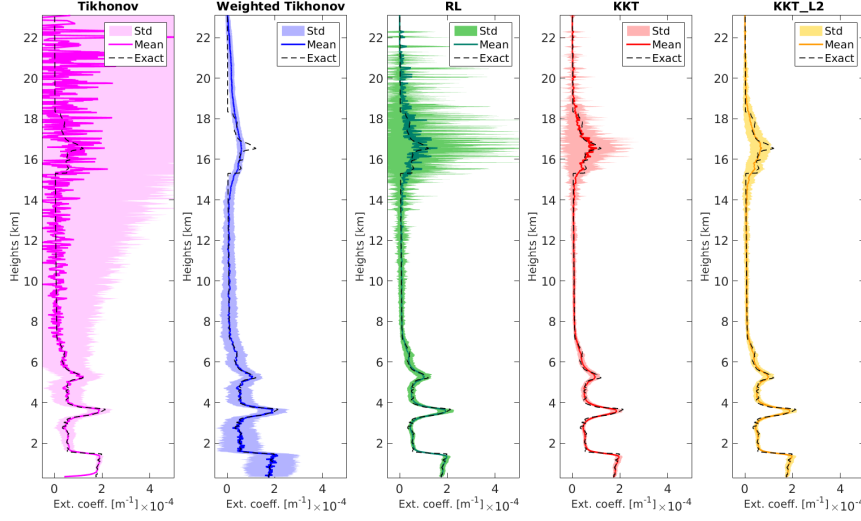
#### 4. Numerical experiments

In this Section we evaluate the proposed algorithms against two sets of synthetic data and compare them to two state-of-the-art methods in this field, namely plain Tikhonov [9, 11] and Richardson–Lucy (RL) [12]. We omit direct comparison against Levenberg–Marquardt (LM), which was shown in [12] to be outperformed by RL, in the same experimental set-up used in this paper (i.e. Raman lidar data). Since plain Tikhonov assumes that the noise variance is the same at every altitude, but we know that this assumption is incorrect for lidar data, we add to the comparison also a weighted





**Figure 2.** Simulation 1. Mean and standard deviation of the estimated profiles over 100 noise realizations for high (top), medium (centre) and low (bottom) SNR. The parameters are as follows:  $\gamma_{Tikhonov} = 5 \times 10^3$ ,  $\gamma_{w.Tikhonov} = 2 \times 10^7$ ,  $n_{RL} = 6000$ ,  $n_{KKT} = 100$ ,  $\gamma_{KKT_2} = 2 \times 10^6$  (top);  $\gamma_{Tikhonov} = 5 \times 10^3$ ,  $\gamma_{w.Tikhonov} = 2 \times 10^7$ ,  $n_{RL} = 4000$ ,  $n_{KKT} = 120$ ,  $\gamma_{KKT_2} = 2 \times 10^6$  (centre);  $\gamma_{Tikhonov} = 10^4$ ,  $\gamma_{w.Tikhonov} = 3 \times 10^7$ ,  $n_{RL} = 3500$ ,  $n_{KKT} = 140$ ,  $\gamma_{KKT_2} = 4 \times 10^6$  (bottom).



**Figure 3.** Simulation 2. Mean and standard deviation of the estimated profiles over 100 realizations for medium SNR. Results are obtained with the following parameters:  $\gamma_{Tikhonov} = 10^5$ ,  $\gamma_{w.Tikhonov} = 10^7$ ,  $n_{RL} = 6000$ ,  $n_{KKT} = 170$ ,  $\gamma_{KKT_2} = 2 * 10^6$ .

Tikhonov algorithm. We recall that Tikhonov and RL work on the log-transformed problem

$$\mathbf{y} = \mathbf{L}\boldsymbol{\alpha} , \quad (24)$$

where  $\mathbf{y} := \log(\mathbf{d}/\mathbf{P})$ . The weighted Tikhonov solution is then given by

$$\tilde{\boldsymbol{\alpha}} = (\mathbf{L}^T \mathbf{W} \mathbf{L} + \gamma \mathbf{I})^{-1} (\mathbf{L}^T \mathbf{W} \mathbf{y}) , \quad (25)$$

where  $\gamma > 0$  is a regularization parameter and  $\mathbf{W}$  is a diagonal weight matrix, accounting for altitude-dependent noise variance: the element  $W_{ii}$  represents the inverse of the noise variance at altitude  $z_i$ . In order to obtain an estimate of these values, we adopt the following strategy: we generate 100 Poisson realizations of  $\mathbf{P}$ , compute the corresponding  $\mathbf{y}$  and then compute the empirical variance of these 100 realizations of  $\mathbf{y}$ . For the sake of completeness, we also recall that RL, when applied to the log-transformed data as in [12], is the iterative algorithm whose iterations are described by

$$\boldsymbol{\alpha}^{(n+1)} = \frac{\boldsymbol{\alpha}^{(n)}}{\mathbf{L}^T \mathbf{1}} \odot \mathbf{L}^T \frac{\mathbf{y}}{\mathbf{L} \boldsymbol{\alpha}^{(n)}} . \quad (26)$$

We now proceed with the description of the two simulated experiments. In each of the two simulations, we start from a true extinction profile and compute the corresponding exact signal through eq. (4). To assess the sensitivity to noise, we then generate 100 Poisson realizations with the same mean, we compute the solutions with all the algorithms, and then show mean and standard deviation of the retrieved extinction coefficients. In order to assess the impact of the signal-to-noise ratio (SNR), we repeat this procedure for three different values of the SNR (high, medium, low). Clearly, as the

data are Poisson, the only way to change the SNR is by changing the signal strength; this is done by modifying the constant  $C_\mu$  in (6).

*Simulation 1.* The first simulation we consider is based on the extinction profile previously used to assess the performances of various inversion algorithms [20] in the framework of the European Aerosol Research Lidar Network (EARLINET). Here  $M = N = 2000$ .

In Figure 1 we show the true extinction profile and three examples of synthetic data, for three different noise levels. The extinction profile and the data are plotted according to the standard vertical visualization of lidar quantities: altitudes are on the  $y$  axis, the signals and the extinction coefficient on the  $x$  axis. In Figure 2 we show the average estimated profiles, with standard deviation, for the five methods and three different values of the SNR. In order to point out the differences between the algorithms, the number of iterations and the regularization parameters have been selected as those that minimize the discrepancy between the mean estimated profile and the true profile, at the lower altitudes.

The results indicate that plain Tikhonov (magenta in the Figure) is extremely affected by noise: when the SNR is high, top row, the profile at low altitudes is correctly reconstructed with almost zero variance; however, the variability of the reconstruction at higher altitudes is large, so that the reconstruction from a single realization (as in a real case) is useless above, say, 10 km; for lower values of the SNR, even the low altitudes present high variability. Of course one could regularize more, but then the estimates at the lower altitudes would be oversmoothed.

Weighted Tikhonov (blue in the Figure) is better: in the high SNR case, the reconstruction is very good, with almost no variance; when the SNR goes down, however, more variance appears at the lower altitudes, and the higher altitudes are over-smoothed. This behaviour is opposite with respect to plain Tikhonov, and suggests that either the Gaussian approximation is not good, or the estimated weights are not optimal.

As discussed in [12], Richardson–Lucy (green in the Figure) improves over Tikhonov: in the high and medium SNR case, the estimated profile is affected by noise only in the higher part of the range; for low SNR, the reconstructions have more variability, as expected, but the profile remains discernible. We observe that the standard deviation in the figure exhibits an irregular behaviour: this is due to the presence of large noise-induced spikes in the single reconstructions, which change from one noise realization to the other one.

The KKT algorithm (red in the Figure) does not seem to depart qualitatively from RL, but it is quantitatively better: while retaining approximately the same mean value, it has lower variance, systematically, at all altitudes and for all three SNR values. This is pleasant and expected, and confirms that the KKT algorithm is better suited to treat the Poisson noise on the signal  $\mathbf{P}$ .

Finally, KKT\_L2 appears to exhibit very good robustness to Poisson noise: all the spikes of RL and KKT disappear here. The reconstructions at high SNR have almost

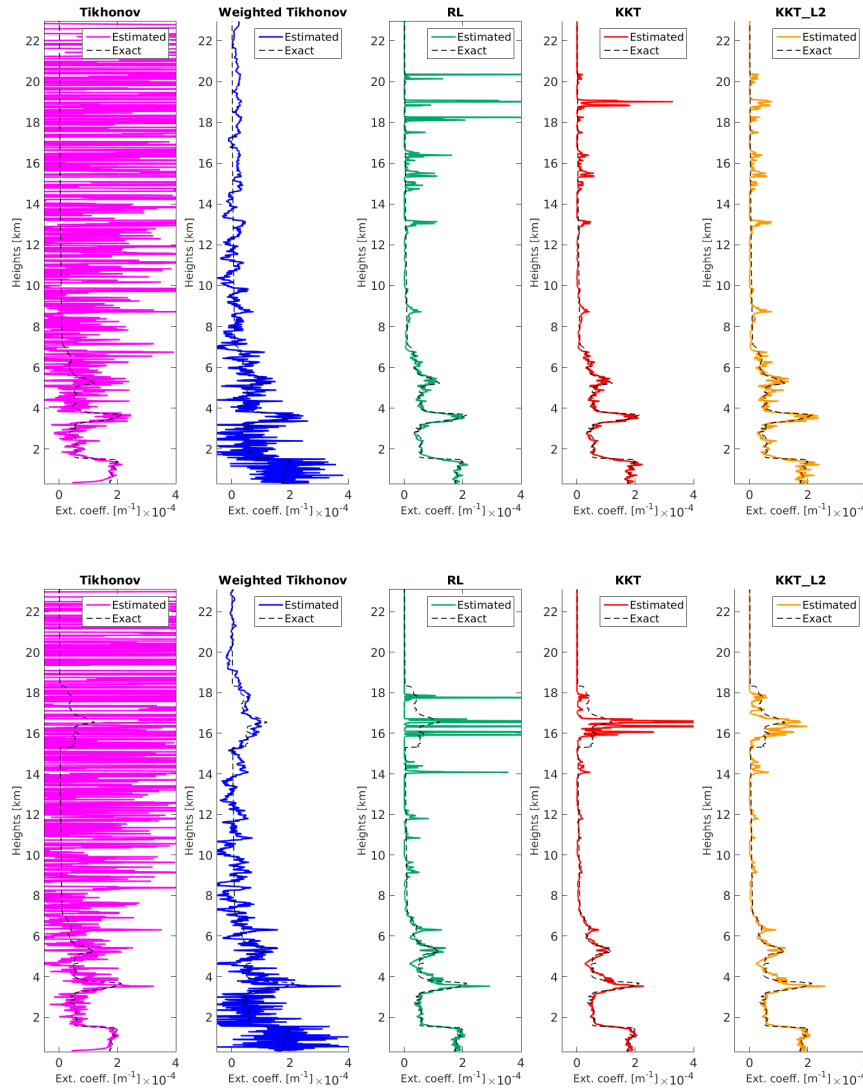
negligible variance and almost no noise, and appear to outperform those of weighted Tikhonov. For lower values of the SNR, we witness some more variance and a little oversmoothing at the lower altitudes, which is expected due to the presence of the regularizing term.

*Simulation 2.* Simulation 1 clearly shows that the KKT algorithm improves slightly over RL, but KKT\_L2 improves substantially when it comes to the stability of the solution. However, it may be objected that KKT\_L2 has a considerable smoothing effect at high altitudes, and therefore it may be questioned whether KKT\_L2 would be able to recover something at high altitudes. To answer this question, in Simulation 2 we take the same synthetic extinction profile of Simulation 1 and add a structure in the higher part of the range.

In Figure 3 we show the estimated profiles obtained by the five inverse methods, in the case of medium SNR; similarly to the first simulation, the parameters have been chosen in order to minimize the discrepancy between the mean estimated profile and the exact profile, at the lower heights. As expected, KKT\_L2 has a smoothing effect at high altitudes, but its reconstructions remain relatively stable. As far as RL and KKT are concerned, they may seem to provide better results, because their mean values follow the true profile more accurately. However, this is due to the fact that they are maximum likelihood methods, and therefore more unbiased than KKT\_L2 (not completely unbiased because they are stopped early); but the variance of their reconstructions is so large (for RL especially) that the reconstruction from a single realization is almost useless in the higher part of the range.

To highlight this point, in Figure 4 we present the estimated profiles from a single realization: in the top row we plot the results from Simulation 1, in the bottom row from Simulation 2. Clearly, it would be difficult to tell which is which by looking either at plain Tikhonov, where oscillations completely destroy the information content of the signal, or Richardson–Lucy, where similar spiky structures appear at around the same altitudes. Weighted Tikhonov manages to recognize the presence of the structure on the top, but the retrieved profile at the lower altitudes is extremely noisy. KKT is better than RL, as it produces less spikes when the true underlying profile is flat and seems to recognize the presence of the peak, although over-estimating its intensity and, finally, KKT\_L2 manages to provide a substantially different reconstruction, with a clear indication of the presence and of the intensity of a structure in the higher part of the range.

It is worth to observe that a better estimate of the structure at high altitudes can be obtained also by RL and KKT and, to a lesser degree, plain Tikhonov, just modifying the stop iterations and the regularization parameter. As shown in Figure 5, the profiles estimated at high altitudes by RL and KKT are much closer to the true solution, and very similar to the one reconstructed by KKT\_L2; however, this improvement at high altitudes comes at the price of a strong oversmoothing in the lower part of the range, particularly for RL.



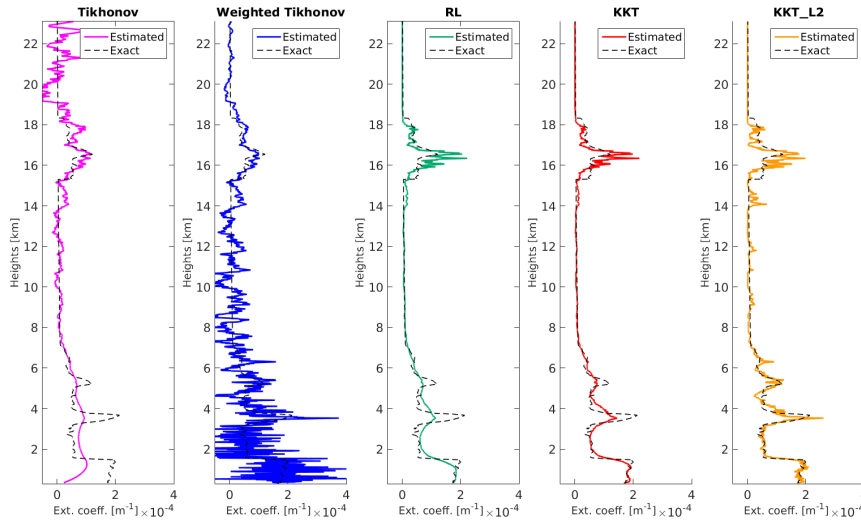
**Figure 4.** Reconstruction from a single realization for medium SNR: top, Simulation 1; bottom, Simulation 2.

## 5. Application to experimental data

In this section we compare the performances of all the algorithms using a set of experimental data.

The signal was acquired on June 17th 2013 by the MALIA (Multiwavelength Aerosol Lidar Apparatus) lidar [21] installed in Naples. The laser wavelength was 355 nm, with a pulse energy of 100 mJ, 20 Hz repetition rate and a 30 cm aperture telescope. The recordings took place during a Saharan dust event, whose traces can be recognized in the estimated profile. Here  $M = N = 1000$ .

In Figure 6 we present the reconstructed extinction coefficients. Following the results from Simulation 2 above, we present two different estimated profiles, for two

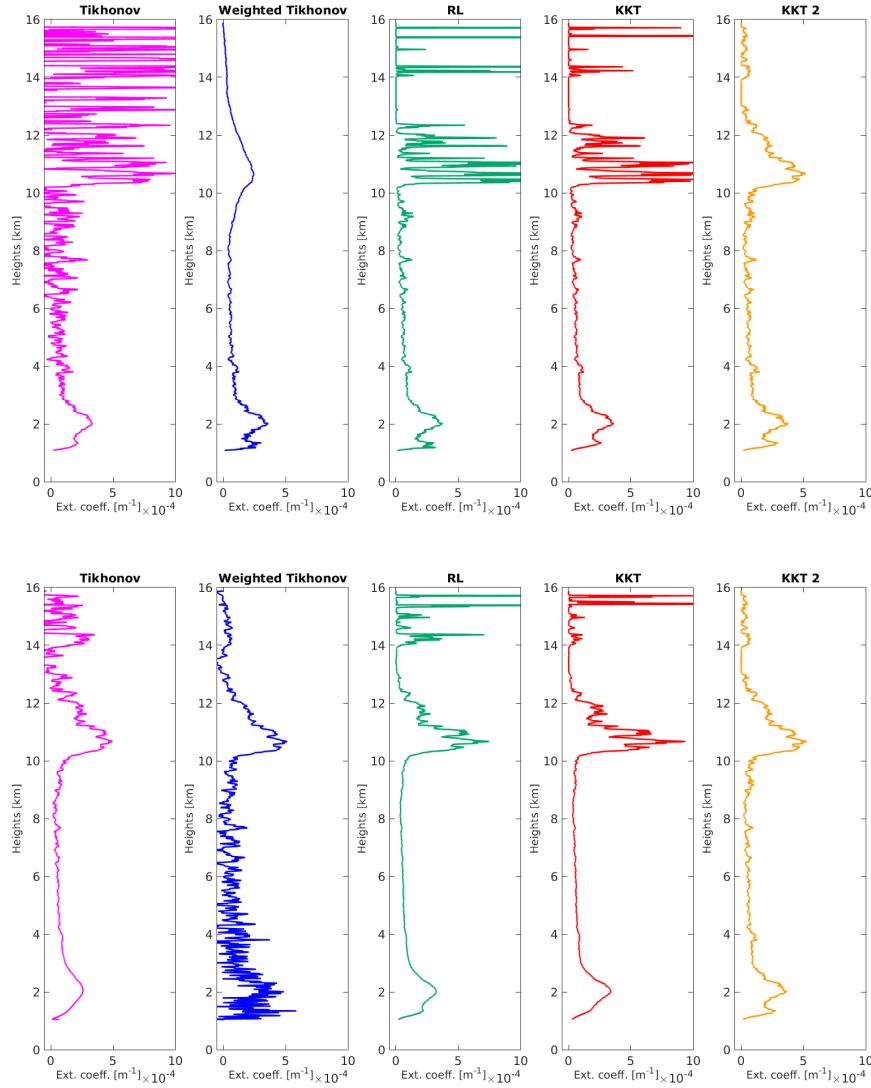


**Figure 5.** Reconstruction from a single realization for Simulation 2, where the regularization parameters have been set in order to provide a better reconstruction of the structure at high altitudes. Explicitly, the parameters are as follows:  $\gamma_{Tikhonov} = 5 * 10^5$ ,  $\gamma_{w.Tikhonov} = 10^7$ ,  $n_{RL} = 350$ ,  $n_{KKT} = 70$ ,  $\gamma_{KKT_2} = 2 * 10^6$ .

different choices of the regularization parameters: one (top row) where the parameters have been selected to obtain similar results (across methods) at low altitudes, one (bottom row) where the parameters have been selected to obtain similar results at high altitudes.

The results are consistent with those obtained on synthetic data. In the top row, the algorithms obtain very similar reconstructions at the lower altitudes, indicating that the information content of the data is high, and the different priors do not play a major role. In particular, plain Tikhonov provides a good reconstruction at the low heights but wide oscillations at high altitudes; weighted Tikhonov has the opposite behaviour, providing a very smooth reconstruction at high altitudes; the profile estimated by KKT appears to remain reliable up to around 10 kilometres, but presents high peaks from there on; RL is similar to KKT; KKT\_L2 seems to provide a reliable reconstruction across the whole range.

In the second row, the estimated profiles present similar features but are not identical at high altitudes, indicating that the signal here is less informative, and the prior becomes more relevant. With this choice, the profiles estimated by plain Tikhonov, RL and KKT are overly smoothed at low altitudes: the general structure is preserved, but the fine details are lost. The profile reconstructed by weighted Tikhonov has, again, the opposite behaviour, becoming too oscillating at low altitudes. Again KKT\_L2 provides the best solution: a good balance between fine resolution and smoothness is achieved uniformly at all the altitudes.



**Figure 6.** Real data. In the first row the results obtained with the following parameters:  $\gamma_{Tikhonov} = 10^4$ ,  $\gamma_{w.Tikhonov} = 10^8$ ,  $n_{RL} = 2000$ ,  $n_{KKT} = 170$ ,  $\gamma_{KKT_2} = 5 * 10^6$ . In the second row the results obtained with the following parameters:  $\gamma_{Tikhonov} = 10^5$ ,  $\gamma_{w.Tikhonov} = 10^7$ ,  $n_{RL} = 300$ ,  $n_{KKT} = 40$ ,  $\gamma_{KKT_2} = 5 * 10^6$ . We point out that the regularization parameter used for KKT\_L2 is the same in both cases.

## 6. Discussion and conclusion

In this article we have introduced two methods, KKT and KKT\_L2, for estimating the atmospheric extinction coefficient from Raman lidar data. The inverse problem under investigation is non-linear, and we have used the Karush–Kuhn–Tucker conditions to incorporate a non-negativity constraint and derive the optimization algorithms. The main novelty of the proposed methods is that they have been devised to model the Poisson statistics of the data, differently from all methods available in the literature on

this topic.

We have compared the proposed algorithms with the state of the art in the field, namely Richardson–Lucy and Tikhonov. The first one assumes that noise is Poisson on the log-transformed data, which is incorrect. The second one, which also works with the log-transformed data, assumes that noise is Gaussian and with constant variance; in order to account for an altitude-variable noise variance, we have also introduced a weighted Tikhonov approach, where the weights have been obtained from the empirical variance of the signals.

The numerical results on synthetic and experimental data appear consistent and indicate that the proposed methods outperform the state of the art. In particular, KKT is better than its most direct competitor RL, inasmuch the estimated profiles are either less affected by noise or less over-smoothed. This improvement, clearly visible but not striking, was expected due to the incorrect assumption on the noise distribution in RL. The comparison of KKT\_L2 with Tikhonov is more striking: while the improvement over plain Tikhonov is obvious, because the noise variance changes strongly across the altitudes and plain Tikhonov fails to take this into account, the fairly substantial improvement over weighted Tikhonov comes a bit more unexpected. We speculate that the relatively unsatisfactory performances of weighted Tikhonov might be due either to a non-optimal choice of the weighting matrix, or quite possibly to the fact that the Gaussian distribution does not model correctly the log-transformed data, which are likely to be more skewed than a Gaussian can manage. However, more work is needed to better address this point.

Our results seem also to point out that the early stopping approach is, in this case, less effective than regularization through an explicit  $\ell^2$  term. Indeed, while the profiles estimated by KKT are still affected by noise-induced spikes, or, alternatively, by some degree of oversmoothing, those obtained by KKT\_L2 present a homogeneous behaviour across altitudes. In practice, when using KKT it is almost necessary to tune the stopping iteration differently, depending on whether one is interested in the lower or higher part of the range. Instead, KKT\_L2 provides a reliable reconstruction at both low and high altitudes with the same value of the regularization parameter.

The results in this study further add to the field of inverse problems with Poisson noise, a lively field, attracting increasing interest recently [22]. We believe there are several interesting directions for future work. One is to consider an altitude-dependent regularization parameter by means of multi-parameter approaches, such as the one in [23]. Alternatively, it would be interesting to investigate the use of more sophisticated regularizers, such as Total Variation [24] or sparsity-inducing norms on a frame expansion, like has been done in [25] for a problem with the same structure. Of course, due to the presence of non-smooth terms, in this case more sophisticated algorithms, such as proximal and primal–dual methods [26, 27, 28], should be used.



## Acknowledgments

The authors acknowledge EARLINET for providing aerosol synthetic lidar profiles. The authors also acknowledge Advanced Lidar Applications srl for providing real lidar measurements. AS has been partially supported by Gruppo Nazionale per il Calcolo Scientifico.

## References

- [1] Michael A Lefsky, Warren B Cohen, Geoffrey G Parker, and David J Harding. Lidar remote sensing for ecosystem studies. *BioScience*, 52(1):19–30, 2002.
- [2] Jesse Levinson, Michael Montemerlo, and Sebastian Thrun. Map-based precision vehicle localization in urban environments. In *Robotics: Science and Systems*, volume 4, page 1. Citeseer, 2007.
- [3] Jerome A Bellian, Charles Kerans, and David C Jennette. Digital outcrop models: applications of terrestrial scanning lidar technology in stratigraphic modeling. *Journal of sedimentary research*, 75(2):166–176, 2005.
- [4] Albert Ansmann, Ulla Wandinger, Maren Riebesell, Claus Weitkamp, and Walfried Michaelis. Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined raman elastic-backscatter lidar. *Applied Optics*, 31(33):7113–7131, 1992.
- [5] V Mohan Karyampudi, Stephen P Palm, John A Reagen, Hui Fang, et al. Validation of the saharan dust plume conceptual model using lidar, meteosat, and ecmwf data. *Bulletin of the American Meteorological Society*, 80(6):1045, 1999.
- [6] A Ansmann, Matthias Tesche, S Groß, V Freudenthaler, P Seifert, A Hiebsch, J Schmidt, U Wandinger, I Mattis, D Müller, et al. The 16 april 2010 major volcanic ash plume over central europe: Earlinet lidar and aeronet photometer observations at leipzig and munich, germany. *Geophysical Research Letters*, 37(13), 2010.
- [7] Paolo Di Girolamo, Donato Summa, Rohini Bhawar, Tatiana Di Iorio, Marco Cacciani, Igor Veselovskii, Oleg Dubovik, and Alexey Kolgotin. Raman lidar observations of a saharan dust outbreak event: Characterization of the dust optical properties and determination of particle size and microphysical parameters. *Atmospheric Environment*, 50:66–78, 2012.
- [8] Albert Ansmann, Maren Riebesell, and Claus Weitkamp. Measurement of atmospheric aerosol extinction profiles with a raman lidar. *Optics letters*, 15(13):746–748, 1990.
- [9] Valery Shcherbakov. Regularized algorithm for raman lidar data processing. *Applied optics*, 46(22):4879–4889, 2007.
- [10] Pornsarp Pornsawad, Christine Böckmann, Christoph Ritter, and Mathias Rafter. Ill-posed retrieval of aerosol extinction coefficient profiles from raman lidar data by regularization. *Applied optics*, 47(10):1649–1661, 2008.
- [11] Pornsarp Pornsawad, Giuseppe D’Amico, Christine Böckmann, Aldo Amodeo, and Gelsomina Pappalardo. Retrieval of aerosol extinction coefficient profiles from raman lidar data by inversion method. *Applied optics*, 51(12):2035–2044, 2012.
- [12] Sara Garbarino, Alberto Sorrentino, Anna Maria Massone, Alessia Sannino, Antonella Boselli, Xuan Wang, Nicola Spinelli, and Michele Piana. Expectation maximization and the retrieval of the atmospheric extinction coefficients by inversion of raman lidar data. *Optics Express*, 24(19):21497–21511, 2016.
- [13] James D Klett. Stable analytical inversion solution for processing lidar returns. *Applied Optics*, 20(2):211–220, 1981.
- [14] Willard I Zangwill. *Nonlinear programming: a unified approach*. 1969.
- [15] Kenneth Lange and Jeffrey A Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. *IEEE Transactions on Image Processing*, 4(10):1430–1438, 1995.

- [16] Henri Lanteri, Muriel Roche, and Claude Aime. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. Inverse problems, 18(5):1397, 2002.
- [17] Henri Lanteri, Muriel Roche, Olga Cuevas, and Claude Aime. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints. Signal Processing, 81(5):945–974, 2001.
- [18] Silvia Bonettini, Riccardo Zanella, and Luca Zanni. A scaled gradient projection method for constrained image deblurring. Inverse problems, 25(1):015002, 2008.
- [19] Larry Armijo. Minimization of functions having lipschitz continuous first partial derivatives. Pacific Journal of mathematics, 16(1):1–3, 1966.
- [20] Gelsomina Pappalardo, Aldo Amodeo, Marco Pandolfi, Ulla Wandinger, Albert Ansmann, Jens Bösenberg, Volker Matthias, Vassilis Amiridis, Ferdinando De Tomasi, Max Frioud, 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(28):5370–5385, 2004.
- [21] Antonella Boselli, Mario Armenante, Loredana D’Avino, Massimo D’Isidoro, Gianluca Pisani, Nicola Spinelli, and Xuan Wang. Atmospheric aerosol characterization over Naples during 2000–2003 earlinet project: Planetary boundary-layer evolution and layering. Boundary-layer meteorology, 132(1):151–165, 2009.
- [22] Thorsten Hohage and Frank Werner. Inverse problems with poisson data: statistical regularization theory, applications and algorithms. Inverse Problems, 32(9):093001, 2016.
- [23] V Bortolotti, RJS Brown, P Fantazzini, G Landi, and F Zama. Uniform penalty inversion of two-dimensional nmr relaxation data. Inverse Problems, 33(1):015003, 2016.
- [24] Michael K Ng, Pierre Weiss, and Xiaoming Yuan. Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. SIAM journal on Scientific Computing, 32(5):2710–2736, 2010.
- [25] Clothilde Melot, Yannick Boursier, Jean-François Aujol, and Sandrine Anthoine. Some proximal methods for poisson intensity cbct and pet. Inverse problems and Imaging, 6(4):565–598, 2012.
- [26] Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In Fixed-point algorithms for inverse problems in science and engineering, pages 185–212. Springer, 2011.
- [27] Emilie Chouzenoux, Jean-Christophe Pesquet, and Audrey Repetti. Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function. Journal of Optimization Theory and Applications, 162(1):107–132, 2014.
- [28] Laurent Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. Journal of Optimization Theory and Applications, 158(2):460–479, 2013.