Assessment of lidar inversion errors for homogeneous atmospheres

Francesc Rocadenbosch, Adolfo Comerón, and Daniel Pineda

The inversion of lidar returns from homogeneous atmospheres has been done customarily through the well-known slope method. The logarithmic operation over the range-corrected and system-normalized received signal used in this method introduces a bias in the statistics of the noise-affected processed signal that can severely distort the estimates of the atmospheric attenuation and backscatter coefficients under measurement. It is shown that a fitting of the theoretically expected exponential signal to the range-corrected received one, using as the initial guess the results provided by the slope method and a least-squares iterative procedure, can yield enhanced accuracy under low signal-to-noise ratios and especially in moderate-to-high extinction conditions. © 1998 Optical Society of America

OCIS codes: 010.0010, 010.1290, 010.3640.

1. Introduction

The basic single-scattering lidar equation may be formally expressed as[1]

$$P(R) = \frac{K}{R^2}\,\beta(R)\exp\left[-2\int_0^R \alpha(r)\mathrm{d}r\right], \qquad (1)$$

where $P(R)$ is the range-received power (W), $\beta(R)$ is the range-dependent volume backscatter coefficient of the atmosphere ($\mathrm{km}^{-1}\ \mathrm{sr}^{-1}$), $\alpha(R)$ is the range-dependent extinction coefficient (km^{-1}), R is the range (km), and K is the system constant defined as

$$K = 10^{-9}(Ec/2)A_r \qquad (\mathrm{W\ km}^3), \qquad (2)$$

where E is the transmitted energy (J), c is the speed of light (m/s), A_r is the effective receiver area (m^2), and 10^{-9} is a conversion constant between the length units used (m^3 to km^3). Note that the term A_r/R^2 in Eq. (1) has units of solid angle (sr).

The goal of lidar inversion is to retrieve the optical parameters $\alpha(R)$ and $\beta(R)$ from the return signal $P(R)$ that is corrupted by noise of different sources, $n(R)$. In a homogeneous atmosphere, the optical parameters are assumed to be constant over the entire lidar range [$\alpha(R) \approx \alpha$, $\beta(R) \approx \beta$]. If Eq. (1) is rewritten in differential form, it takes the form

$$\frac{\mathrm{d}S(R)}{\mathrm{d}R} = \frac{1}{\beta(R)}\frac{\mathrm{d}\beta(R)}{\mathrm{d}R} - 2\alpha(R), \qquad (3)$$

where

$$S(R) \triangleq \ln[R^2P(R)]. \qquad (4)$$

The application of the homogeneous approximation over small range intervals leads to the conjecture

$$\frac{1}{\beta(R)}\left|\frac{\mathrm{d}\beta(R)}{\mathrm{d}R}\right| \ll 2\alpha, \qquad (5)$$

at least over most of the $S(R)$ curve. Unfortunately, assumptions such as this are not well justified in some situations of interest, e.g., under the conditions prevailing in dense clouds or smoke or in any situation in which significant local inhomogeneities occur. However, the utility of the homogeneous approximation often increases with increasing optical depth, such that changes in the fractional gradient of extinction or backscatter cause small variations in the signal[2] [since extinction and backscatter are usually highly correlated, this means that, under these circumstances, the first term of Eq. (3) can be neglected].

We extend the evaluation methodology presented by Kunz and de Leeuw[3] and assess the inversion errors of the slope-method algorithm and of a direct fitting of the range-corrected received signal to the theoretical exponential curve that it should ideally

The authors are with the Department of Signal Theory and Communications, Antennas, Microwaves, Radar, and Optics Group, Polytechnic University of Catalonia, C/Sor Eulalia de Anzizu s/n, 08034 Barcelona, Spain.

Received 26 February 1997; revised manuscript received 5 August 1997.

0003-6935/98/122199-08$15.00/0

conform to. Inversion errors are expressed in terms of α and β rms errors for different signal-to-noise ratios (SNR's) and average atmospheric extinction coefficients. The inversion results presented can be representative of many short-range tropospheric horizontal-pointing lidar systems.

2. Methodology and Simulation Criteria

The considered noise source encompasses signal-induced and dark-current shot noises as well as electronic thermal noise. It is known that shot noise is described by Poisson's statistics; however, if the count numbers are high enough (i.e., the signal strength is more than 50 photons over half of the inverse of the receiving system noise equivalent bandwidth B_N or, equivalently, over the integration time[4]), their discrete statistics may be approximated by continuous, Gaussian ones. For example, at 532-nm wavelength and 10-MHz noise equivalent bandwidth, the approximation is valid for return powers greater than 0.4 nW. Hence noise is modeled by an equivalent Gaussian noise range-dependent spectral density $\sigma_{eq}^2(R)$ that merges into a single body the range-dependent signal-induced and dark-current shot noise spectral densities as well as the thermal noise spectral density [$\sigma_{sh,s}^2(R)$, $\sigma_{sh,d}^2$, and σ_{th}^2, respectively]. Thus the range-dependent SNR is computed according to

$$\mathrm{SNR}(R) = \frac{R_v LP(R)}{\sigma_{eq}(R)\sqrt{B_N}} = \frac{R_v LP(R)}{[\sigma_{sh,s}^2(R)B_N + \sigma_{sh,d}^2 B_N + \sigma_{th}^2 B_N]^{1/2}} \quad \left(\frac{\mathrm{V}}{\mathrm{V}}\right), \tag{6}$$

where R_v is the receiver responsivity (V/W), $P(R)$ is the range-dependent return power (W), B_N is the noise equivalent bandwidth (Hz), and L is the system optical loss, which includes the receiver's spectral transmission factor and either the overlap factor or the range-dependent geometrical form factor. A detailed explanation of the computation of SNR(R) from the system parameters is given in Appendix A.

The noise equivalent bandwidth B_N for a shaping filter in the receiver chain with spectral response $H(f)$ (f is the frequency) is defined as the bandwidth of the equivalent rectangular filter that, given the same white-noise spectral density S_{nn} (W/Hz) at its input, would yield the same noise power (W) as that from the real one. Formally,

$$B_N = \frac{\int_0^\infty S_{nn}|H(f)|^2 \mathrm{d}f}{S_{nn}|H(f)|_{\max}^2} = \frac{\int_0^\infty |H(f)|^2 \,\mathrm{d}f}{|H(f)|_{\max}^2}, \tag{7}$$

where $|H(f)|_{\max}^2$ is the gain at the center frequency [$f = 0$ for a low-pass filter (LPF)].[5] For a practical nth-order Butterworth LPF with 3-dB signal bandwidth equal to B, both concepts are related by

$$B_N = \frac{\pi B}{2n \sin(\pi/2n)}. \tag{8}$$

In this study a fourth-order 10-MHz signal bandwidth Butterworth LPF is used to model the receiver bandwidth, so that both the noise equivalent and the signal bandwidth virtually coincide. This is close to the behavior of an ideal filter, whereas, for example, a first-order Butterworth LPF would have $B_N = 1.57B$.

From Eqs. (1) and (6), it is obvious that under the approximation of a homogeneous atmosphere, the range-dependent SNR is a function of only the volume extinction α, the volume backscatter coefficient β, and the system constant K [Eq. (2)]. Therefore, given the SNR at some minimum range $R_{\min}$, SNR(R) is known for every other distance.

The criteria used with regard to the system ranges are the following:

The minimum system range R_{ovf} is determined by the optical overlap factor (ovf) reaching the 100% (and assuming that the receiver stages are far enough from saturation because of unreasonable high gains in the receiver chain). A value for R_{ovf} of 200 m, which is typical of a short-range tropospheric lidar system, is assumed. However, the minimum inversion range is set to begin at R_{ovf} compensated by the addition of a guard range $\Delta R_{\min} = 2c/B$ (where c is the speed of the light and B is the signal bandwidth of the receiver or, equivalently, the 3-dB cutoff frequency of the filter). This is done on account of the smoothing or smearing effect that is caused by the filter on the output signal in response to the abrupt rising transition of the optical power until the full overlap at $R_{ovf} = 200$ m occurs. The value $2c/B$ corresponds to the spatial interval equivalent to a settle time Δt equal to 8 times the rise time of the filter, which can be roughly approximated by $1/2B$ (after Ref. 5). When we take into account the two-way path of the lidar signal, the time interval $\Delta t = 8 \times 1/2B$ is translated into a spatial interval $\Delta R_{\min} = c\Delta t/2 = 2c/B$. This handy rule of thumb has been extensively tested in the simulation results, showing little influence for settle times greater than or equal to four times the rise time. Therefore the minimum inversion range $R_{\min}$ is computed as

$$R_{\min} = R_{ovf} + \Delta R_{\min}. \tag{9}$$

For example, the 10-MHz filter used yields $R_{\min} = 260$ m, and a 50-MHz filter yields $R_{\min} = 212$ m. The value for $R_{\min}$ of 260 m will be assumed throughout the rest of the paper.

The maximum system range, which coincides with the maximum inversion range $R_{\max}$, is taken as the minimum of the following two ranges:

(1) the range R for which SNR(R) = 1; and

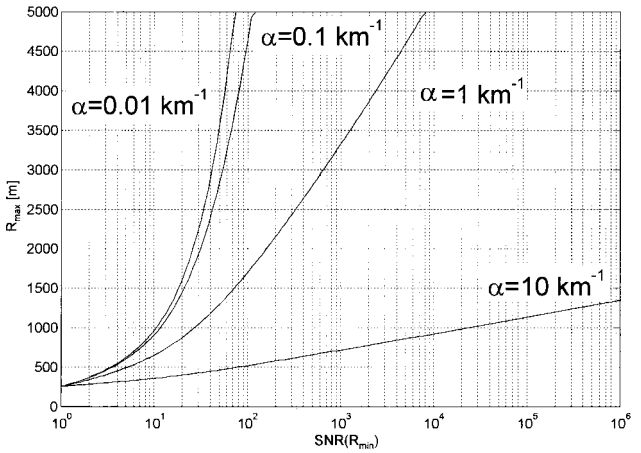

Fig. 1. Maximum range of the lidar, expressed in the required SNR at $R_{\min}$ for different extinction coefficients.

(2) 5 km, which has been considered to be a reasonable typical figure for the maximum range at which the overlap between the illuminated cross section of the atmosphere and the receiving system field of view is still 100% in the case of such a lidar system.

Certainly, the latter value will strongly vary from system to system, since it depends on the actual lidar arrangement (i.e., coaxial or biaxial, bistatic or monostatic) and atmospheric optical parameters. Therefore we have

$$R_{\max} \triangleq \min\begin{cases} R & \text{for which SNR}(R) = 1 \\ 5 \text{ km} \end{cases}. \quad (10)$$

Figure 1 shows the $R_{\max}$ criterion versus SNR($R_{\min}$) for different extinction coefficients and for the system parameters given in Table 1. The computation of $R_{\max}$ follows these steps: First, for each SNR($R_{\min}$) specified, the related optical power P_s is solved from Eq. (6) [or, equivalently, Eq. (A2) below]. This results in a quadratic equation in P_s. Second, the system constant K (which is directly related to the output energy of the system) is computed from Eq. (1), given P_s, the atmospheric extinction α, the $R_{\min}$ specification, and the backscatter coefficient β from Table 2. The range-dependent return power $P(R)$ is then available, and so is SNR(R), the range-dependent SNR. Finally, $R_{\max}$ is solved as indicated in expression (10).

Table 1. Default Values Used in the Simulations

Parameter	Value
APD quantum efficiency	55%
APD current responsivity R_{io}	236 mA/W
APD multiplication gain M	100
APD excess-noise factor F	4.0
APD dark surface current I_{ds}	14 nA
APD dark bulk current I_{db}	6 pA
Preamplifier equivalent transimpedance gain G_T	10^5 Ω
Preamplifier equivalent input noise density $\sigma_{\text{th},i}$	5 pA Hz$^{-1/2}$
Preamplifier 3-dB cutoff frequency B	10 MHz
Optical system losses L	30%
Typical background power P_{back}	1.6 nW

Table 2. Assumed Optical Parameters for Different Visibilities

	Extinction	Backscatter	Atmospheric Condition Based on	
Visibility (V_M)	α (km^{-1})	β (km^{-1} sr^{-1})	Ref. 1	Ref. 6
391.2 m	10	5×10^{-1}	Light water cloud	Moderate fog
3.912 km	1	3×10^{-2}	Moderate/light haze	Haze
39.12 km	10^{-1}	4×10^{-3}	Clear air	Standard/ very clear
391.2 km	10^{-2}	1×10^{-3}	Rayleigh gaseous	Exceptionally clear

As for the backscatter coefficient β, it is assumed to be related to the extinction coefficient, which is the driving parameter in this study, as indicated in Table 2, after the approximate figures proposed by Collis and Russell.[1] For example, at a wavelength of 532 nm and according to the definition of visibility of Koschmieder[7] and Kruse *et al.*,[8] the extinction coefficients indicated in Fig. 1 (α = 10, 1, 0.1, and 0.01 km^{-1}) correspond to simulated backscatter values of 5×10^{-1}, 3×10^{-2}, 4×10^{-3}, and 10^{-3} km^{-1} sr^{-1}, respectively. These atmospheric extinction conditions range from moderate fog (light water cloud) to exceptionally clear air (Rayleigh gaseous).

The result of the computation of the system constant K versus SNR($R_{\min}$) is plotted in Fig. 2 along with a typical system-constant range (to be explained in Section 4). For system constants (energy levels) linked to a SNR($R_{\min}$) specification greater than or equal to approximately 10^2, the system constant is proportional to the square of SNR($R_{\min}$). This is jus-

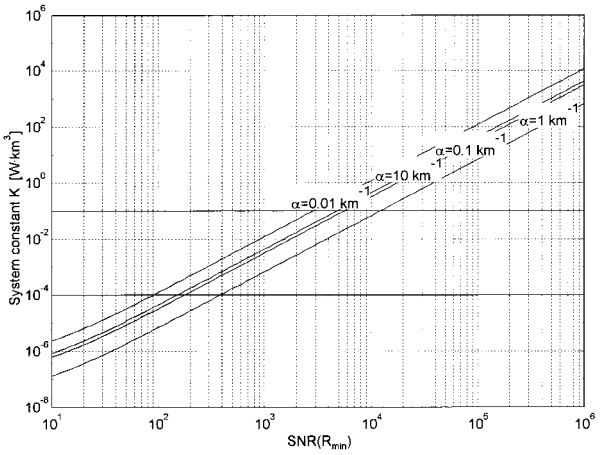

Fig. 2. System constants versus SNR($R_{\min}$) for different atmospheric extinctions (α = 10, 1, 0.1, and 0.01 km^{-1}). Solid horizontal lines indicate a typical system constant range.

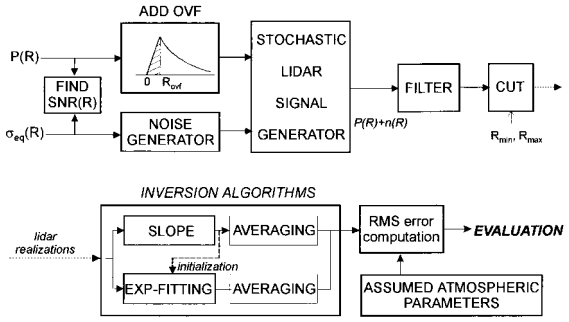

Fig. 3. Block diagram for the error evaluation procedure.

tified because in that range of SNR's and considering the typical system default parameters of Table 2, the system is under a signal-induced shot-limited mode at $R_{\min} = 260$ m. A common result is that the closest observation cells to the system are under a signal-induced shot-limited mode and that they progressively tend to a thermal-limited mode near the maximum range. In a few cases, however, the closest observation cells may be thermal noise limited because the system constant (output energy) is so low that the term ${\sigma_{\text{th}}}^2$ in Eq. (6) begins to dominate ${\sigma_{\text{sh},s}}^2$ at $R_{\min}$. This justifies the lines of Fig. 2 becoming curved for approximately $\text{SNR} < 10^2$. As long as one moves toward lower SNR's, the slope of these curves tends to unity.

The core of the simulation procedure is the synthesis of noise-corrupted lidar return signals by the addition of noise realizations to the simulated range-dependent backscattered return power. The procedure is sketched in Fig. 3. Once $P(R)$ from Eq. (1) is computed, the lidar signal is modeled in a more realistic way by incorporating the effect of the ovf into the signal. This is roughly modeled as a linear increase of the optical power from 0 to $P(R_{\text{ovf}})$ in the overlap range interval $[0, R_{\text{ovf}}]$, where $R_{\text{ovf}} = 200$ m. At the output of the stochastic lidar signal generator, a noise-corrupted lidar signal is available so that, for each range R, a noise stochastic variable $n(R)$ with Gaussian range-dependent standard deviation $\sigma_{\text{eq}}(R)$ is simulated, as discussed in Appendix A. Next, the data burst is filtered and cut according to the $R_{\min}$ and $R_{\max}$ criteria stated in Eq. (9) and expression (10). At this point filtered noise-corrupted lidar realizations are available with a spatial sampling rate of 7.5 m. To compute error inversion statistics, the synthesis and inversion procedures are repeated $M = 10$ times for each particular $\text{SNR}(R_{\min})$ and driving atmospheric extinction coefficient.

The inversion error in an optical parameter of the atmosphere, let us say α, is computed in a rms sense after averaging the M inversion runs. Thus the average extinction inversion error is expressed as

$$e_{r,\alpha} = \left[\frac{1}{M}\sum_{i=1}^{M}\left(e_{r,\alpha}{}^{i}\right)^2\right]^{1/2} \times 100\%, \tag{11}$$

with

$$e_{r,\alpha}{}^{i} = \frac{\alpha_e{}^{i} - \alpha}{\alpha}, \tag{12}$$

where α is the actual extinction coefficient and $\alpha_e{}^{i}$ is the extinction coefficient yielded by the ith inversion. Since the noise-corrupted lidar return signal is stochastic in nature, such an evaluation procedure yields an estimate, itself a random variable, of the mean relative squared error.

3. Slope Method versus Exponential-Curve Fitting

Both the slope method and exponential direct fitting of an exponential curve are, in fact, least-squares algorithms, the former taking advantage of the closed expressions appearing when the fitted function is a linear one. The slope method first proposed by Collis[9] lies in the range-corrected version of expression (4) under the assumption of a homogeneous atmosphere in Eq. (1). If the noise contribution $n(R)$ is considered, the range-corrected function takes the form

$$\begin{aligned} S(R) &= \ln(R^2[P(R) + n(R)]) \\ &= \ln(K\beta) - 2\alpha R + \ln\left[1 + \frac{n(R)}{P(R)}\right], \end{aligned} \tag{13}$$

to which one can adjust a line by choosing values of m and c that minimize

$$\|S(R) - mR - c\|^2 = \sum_{i=1}^{N}[S(R_i) - mR_i - c]^2, \tag{14}$$

where R_i denotes the range of the ith resolution cell; m and c estimate -2α and $\ln(A\beta)$, respectively.

As much as its computational straightforwardness is of advantage, the weak point of the slope method lies in the bias that the unavoidable system noise introduces in any real received signal. In effect, although the parameter estimation will be good when the noise term is small, the estimates of α and $\ln(A\beta)$ tend to increase as the noise term $n(R)$ approaches $-P(R)$, producing large negative peaks in $S(R)$. A sample run with an approximate bias of 12.4% for $\text{SNR} = 50$ at 260 m is illustrated in Fig. 4, where a simulated realization of $S(R)$ is compared with the ideal $S(R)$ (i.e., without noise) and with the best linear fit. The slope and the intercept of the noiseless range-corrected function are -2α and $\ln(A\beta)$, respectively, and the slope and the intercept of the noisy $S(R)$ from which the optical parameters are estimated are m and c, respectively. The figure shows how the negative noise spikes contribute to an overestimation of the sought-after parameters (α and β).

Instead of using a linear function, we can formulate the lidar inversion problem by using an exponential-

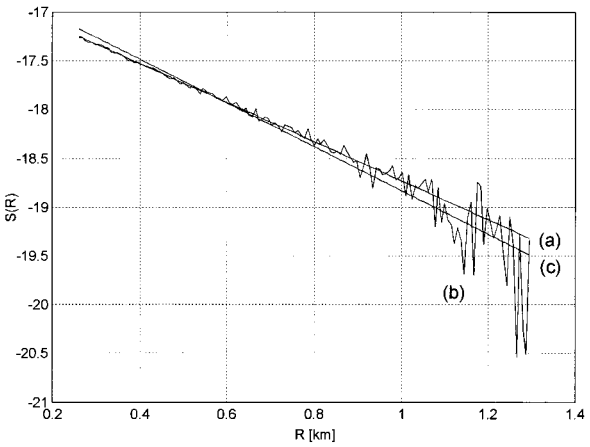

Fig. 4. (a) Ideal range-corrected function $S(R)$, (b) simulated noise-corrupted $S(R)$, (c) $S(R)$ as inverted by the slope method [simulation parameters: $\alpha = 1\ \mathrm{km}^{-1}$, $\beta = 3 \times 10^{-2}\ \mathrm{km}^{-1}\ \mathrm{sr}^{-1}$, $\mathrm{SNR}(R_{\min}) = 50$].

curve fitting. In particular, one chooses the variables a and b that minimize

$$\|R^2P(R) - b\exp(-aR)\|^2 = \sum_{i=1}^{N} [R_i^2P(R_i) - b\exp(-aR_i)]^2, \quad (15)$$

where a and b estimate the terms 2α and $A\beta$, respectively. Unlike the slope method, the nonlinearity of the fitting function with respect to variables a and b makes the derivation of a closed analytical solution impossible. Even though historically this may have been a deterrent, the computational tools currently available render the task of numerically minimizing Eq. (15) straightforward. Direct fitting of the exponential curve has the advantage of introducing no bias in the experimental data to which the estimates (a, b) are to be adjusted. Moreover, the points that correspond to the closer distances, for which the SNR is higher, are given more weight in the minimization of Eq. (15). Yet, a drawback of the algorithm is the need for a good first guess to initialize the minimization procedure[10] so that it converges to an absolute minimum. This can be tackled by using the result of the slope-method inversion as the first guess. With this rationale both algorithms have been implemented and tested by using the simulation procedure described in Section 2.

The exponential-curve fitting iterative algorithm has been implemented by using MATLAB's *leastsq* function, which is based on a cubic Levenberg–Marquardt search algorithm.[11] This method generally requires fewer function evaluations but more gradient evaluations, and for this reason the gradient of the fitting function is supplied analytically. The optimization does not finish until the following two termination criteria are met or a maximum number of 200 iterations is exceeded. The termination criteria are (1) worst-case precision of the independent variables (a, b) equal to 10^{-5} and (2) minimum precision of the objective function [Eq. (15)] at the solution equal to 10^{-8}. A single inversion through the fitting of the exponential curve seldom exceeds 120 kflops, which translates into less than an 800-ms execution time when running on a PC-486 33-MHz platform. As pointed out above, the fitting has been initialized for each lidar realization with the α and β estimates that have resulted from the inversion by means of the slope method.

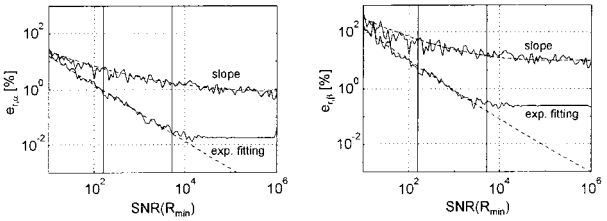

Fig. 5. Comparison between extinction and backscatter inversion errors by use of the slope method and exponential-curve fitting. The final results are shown as solid curves, and the results obtained by filtering only the noise component are shown as dashed curves. Vertical solid lines indicate the $\mathrm{SNR}(R_{\min})$ margin for typical system constants (simulation parameters: $\alpha = 10\ \mathrm{km}^{-1}$, floor $= -23$).

As in the case of the slope method, for each range R the negative noise spikes may drive $1 + n(R)/P(R)$ in Eq. (13) to values that are close to zero or even negative; hence a rule must be implemented to deal with "abnormal" results in taking the logarithm. Note that the lower the SNR at a range R, the higher the likelihood that this happens in a particular run. Concerning the results of the slope-method inversions that are presented in Section 4, the criterion has been to replace the logarithm in question by a floor value equal to -23 whenever its argument (computed in W km^2) becomes equal to or less than $\exp(-23)$ or negative. The -23 figure is a threshold close to the value of $\ln[R^2P(R)]$ (-21.88 with R in kilometers and $P(R)$ in watts) at $R_{\max} = 0.359$ km yielded by Eq. (10) and Fig. 1 when $\alpha = 10\ \mathrm{km}^{-1}$ (worst atmospheric condition) and $\mathrm{SNR}(R_{\min}) = 10$.

4. Simulation Results and Discussion

Figures 5–8 (solid curves) show the estimated rms relative errors in the inverted values of α and β derived by using both the slope method and the exponential-curve fitting algorithm, for several extinction coefficients, representative of what can be expected in conditions ranging from moderate fog ($\alpha = 10\ \mathrm{km}^{-1}$) to exceptionally clear air ($\alpha = 0.01\ \mathrm{km}^{-1}$).

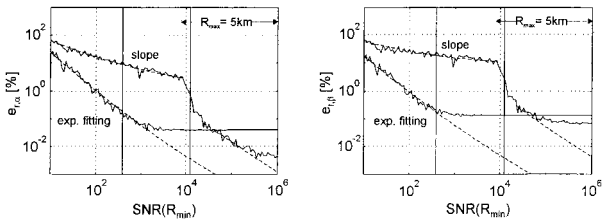

Fig. 6. Same as Fig. 5, but for $\alpha = 1\ \mathrm{km}^{-1}$.

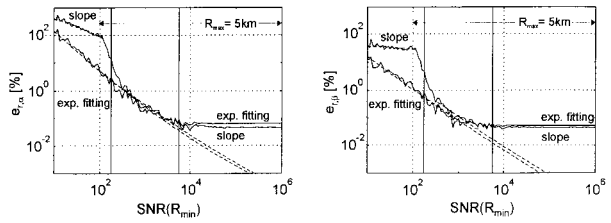

Fig. 7. Same as Fig. 5, but for $\alpha = 0.1$ km^{-1}.

As indicated in Section 3, the backscatter coefficients β are given in Table 2. The abscissas represent SNR($R_{\min}$), and the ordinates indicate the rms relative inversion error according to Eq. (11). The ripple in the curves results from the finite number M (M = 10 in this case) of simulated signals that have been inverted to estimate the relative inversion error for each SNR($R_{\min}$). With M = 10, and 121 points in the abscissas, each error inversion curve involves 1210 inversions. Note that in spite of the visual impression, the logarithmic scaling in the ordinates indicates a larger ripple for lower values of SNR($R_{\min}$).

It is important to stress that for each driving atmospheric extinction and SNR($R_{\min}$), each lidar realization of the sequence of 10 serves as input to the slope-method algorithm for inversion, so that each slope inversion outputs a different initialization (α and β estimates) to the exponential fitting iterative algorithm, which is also fed with the same lidar realizations (see Fig. 3). If we average the inversion errors in the optical parameters inverted by the exponential fitting iterative algorithm, the effect of the different initializations on the final performance (or sensitivity to the initial guess) becomes included in the final inversion error figure. Consequently, each of the error inversion plots presented must be understood as a set of stochastic variables, one for each SNR($R_{\min}$), whose mean represents the average inversion error *in a statistical sense* and whose variance or ripple seems to be a good indicator of the sensitivity of the algorithm to the noise realizations in the case of the slope method and to them and the initialization guess as well in the case of the exponential-curve fitting iterative algorithm.

To compare the performance of both algorithms, it is interesting to define a typical operation range in the abscissas for practical systems. Assuming that the reasonable range of the system constant K for an atmospheric elastic lidar system typically lies between 10^{-4} and 10^{-1} W km^3 [this assumes energies between 40 mJ and 1 J and aperture diameters of the receiving optics between 15 cm and 1 m in Eq. (2)], the reasonable range of SNR($R_{\min}$) falls roughly between 9×10^1 and 3×10^3 V/V when the extinction coefficient α is 10^{-2} km^{-1}, and between 4×10^2 and 10^4 V/V when α is 1 km^{-1}. Figure 2 shows the system constant K versus SNR($R_{\min}$) parameterized for the extinction coefficients used (α = 10, 1, 0.1, and 0.01 km^{-1}) along with the maximum and minimum practical limits.

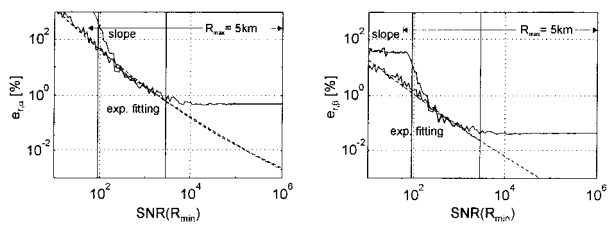

Fig. 8. Same as Fig. 5, but for $\alpha = 0.01$ km^{-1}.

When the inversion algorithm is the exponential-curve fitting and one considers homogeneous atmospheric conditions with medium-to-high extinction coefficients (α = 10 and 1 km^{-1} in Figs. 5 and 6, respectively), there is a clear improvement in the estimation of both the extinction and backscatter coefficients within the practical limits of SNR($R_{\min}$). Under these conditions, only a fraction of the exponential curve is available (the maximum inversion range is indicated in Fig. 1). For the lowest atmospheric extinctions (α = 0.1 and 0.01 km^{-1} in Figs. 7 and 8, respectively), there are no significant differences between the rms inversion error yielded by either method in the SNR($R_{\min}$) range of interest, which corresponds to situations with $R_{\max}$ = 5 km.

The same Figs. 5–8 allow us to compare the effect of the receiver signal bandwidth on the signal component of the lidar realizations. The error plots discussed so far (solid curves) have been computed by filtering both the signal and noise components of the lidar realizations, and the inversion errors obtained after filtering only the noise component are superimposed as dashed curves (these curves have been interpolated in one or two subintervals by using a second-order method for illustrative purposes). If both sets of plots in Figs. 5–8 are compared, the effect of the filter on the signal component, which basically causes overshoot and ringing in response to the abrupt transition of the signal in the overlap range interval, becomes evident through a saturation effect on the error plots in the solid curves from some SNR($R_{\min}$) up. For medium-to-high atmospheric extinctions (α = 10 and 1 km^{-1} in Figs. 5 and 6, respectively), such an effect begins for lower SNR($R_{\min}$) in the case of the exponential fitting algorithm, whereas for low extinctions (α = 0.1 and 0.01 km^{-1} in Figs. 7 and 8, respectively), it is basically the same for both algorithms. When only the noise is filtered (dashed curves), the saturation disappears. Hence it is found that the slope method is the less influenced algorithm. A suitable explanation for this is that all the samples of the range-corrected function $S(R)$ have the same weight in the linear least-squares fitting [Eq. (14)] from which the optical parameters are derived. Yet, in the case of the exponential-curve fitting iterative algorithm, the nonlinear fitting to a decreasing exponential curve [Eq. (15)] gives more weight to the first samples of the inversion interval, where these phenomena are precisely more important. Thus, for high SNR's at $R_{\min}$, the inversion error becomes dominated by the smearing effect of

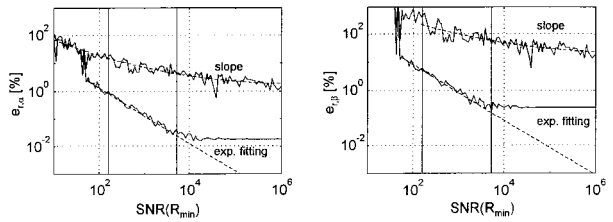

Fig. 9. Same as Fig. 5, but for floor = −30.

the limited bandwidth of the receiver on the signal rather than by the noise itself.

If we ignore the saturation effects imposed by the filter, the change in the slope of the dashed curves of the slope-method inversion, which also appeared in the results presented by Kunz and de Leeuw,[3] arises from the maximum range considered, switching from the condition $\mathrm{SNR}(R_{\max}) = 1$–5 km, in accordance with Eq. (10). Comparing the results shown for the slope method with those from that previous study, we see that they are in close agreement with the error plots shown as dashed curves in Figs. 5–8. The differences with the plots shown as solid curves can possibly be justified by the improvement of the simulation model, which now permits modeling of the filter's response to the abrupt transition of the lidar signal in the overlap interval and the extension of the noise variance to a range-dependent one that merges shot and thermal noise contributions into a single body.

The effect of lowering the limiting floor value has also been investigated. By comparison of Figs. 5 and 9, changing it down to −30 instead of −23 (default setting) results in a slight increase of the rms error in the α and β estimates obtained through the slope method but does not have any noticeable effect on the inversion error when the estimates are obtained by the exponential-curve fitting iterative algorithm, except when the assumed extinction coefficient is 10 km^{-1} and $\mathrm{SNR}(R_{\min})$ is below 50. In the latter case, there are instances in which the iterative algorithm appears to be unable to improve the initial guesses for α and β provided by the already poorly performing slope method. The performance of the slope method in this interval is dominated by large biases and, consequently, by large inversion errors. (These are largely evidenced in the backscatter error plot of Fig. 9.) Another comforting result from Fig. 9 is the reduction in the ripple of the error plots of the exponential-curve fitting iterative algorithm as compared with those of the slope-method plots. Despite the large variance distribution of the initial guesses that come from the slope method, this variance compression leads to more reliable results when the iterative exponential fitting algorithm is used.

Finally, by comparing the inversion errors in the plots of Figs. 5–8 with those of Fig. 9 within the abscissa intervals limited by $\mathrm{SNR}(R_{\max}) = 1$, two additional points arise:

(1) In the situations limited by the criterion $\mathrm{SNR}(R_{\max}) = 1$, the α-inversion error increases as the atmospheric extinction lowers. This can be better understood by considering a homogeneous atmosphere with low extinction; then the noisy range-received power can be approximated by

$$P(R) = \frac{K}{R^2}\,\beta \exp(-2\alpha R) + n(R)$$

$$\approx \frac{K}{R^2}\,\beta(1 - 2\alpha R) + n(R). \tag{16}$$

Qualitatively, as long as the atmospheric extinction decreases, the term $-2\alpha R$ becomes smaller than unity (at least for all the inversion cells until some boundary range) and tends to be masked by the noise term. Thus low-noise spikes added to any power sample must be offset by large deviations in the extinction value during the inversion process. This leads to large inversion errors.

(2) A further effect to be noted of these algorithms is that plots of β error versus SNR tend to coincide for small extinctions. This happens because for extinctions equal to or smaller than 0.1 km^{-1} the effect of the transmittivity term in Eq. (1) becomes less important and, consequently, the $\mathrm{SNR}(R)$ function dependence is approximately proportional to β/R^2 regardless of the atmospheric extinction in that range.

5. Conclusions

The rms inversion errors in the estimates of the extinction and backscatter coefficients obtained from the inversion of lidar returns in homogeneous atmospheres with use of both the slope method and an iterative exponential-curve fitting to the received signal have been compared. The inversion errors have been estimated by repeatedly inverting simulated range-dependent noise-corrupted lidar signals. When working with moderate-to-high extinction coefficients, it has been found that an exponential-curve fitting iterative algorithm initialized with the slope-method estimates exhibits lower inversion errors than those with the stand-alone slope method. In the case of the exponential-curve fitting iterative algorithm, it has been shown that this improvement is due to the absence of the large negative peaks that appear in the range-corrected function of the slope method and that bias the estimation procedure. This effect becomes especially important when low signal-to-noise ratios (SNR's) are found in the exploration range. Another point in favor of the algorithm is the variance compression in the distribution of the inversion estimates yielded by the iterative exponential fitting in spite of the large variance distribution of the initial guesses that result from the slope method in difficult situations.

To compare the inversion error figures of both algorithms, typical elastic backscatter lidar systems ranging from 40-mJ to 1-J energy and from 15-cm to

1-m aperture diameter in the receiving optics and limited to short-range tropospheric exploration have been considered. The inversion errors of the exponential fitting iterative algorithm are usually more than 1 order of magnitude lower than those of the slope method for atmospheric extinctions between 1 and 10 km^{-1} and SNR's (defined at the minimum range) roughly between 10^2 and 10^4. The exact improvement ratio depends on the actual decision rule used to handle the negative noise spikes of the range-corrected function of the slope method and on the guard range left between the minimum system range and the minimum inversion range, the value of $2c/B$ (where c is the speed of the light and B is the receiver signal bandwidth) being a reasonable figure. Otherwise, overshoot and ringing caused by the limited receiver bandwidth on the first samples of the lidar signal may hamper the performance of the exponential fitting algorithm, which gives more weight to the initial part of the inversion interval.

In conclusion, even though the analytical simplicity of the slope method has historically been of advantage, this has been superseded by the computational tools (which enable the practical implementation of the nonlinear exponential algorithm) available today.

Appendix A: Signal-to-Noise Ratio for an Elastic Backscatter Lidar

The SNR expression considers a typical avalanche photodiode (APD)/amplifier combination that is used to detect the return from backscattered pulses in an elastic lidar system. The range-dependent SNR, SNR(R), is defined at the receiver output as the ratio of voltage that is due to the primary photocurrent to the total noise voltage. Based on Refs. 12 and 13, signal-induced and dark-current shot noise and thermal noise spectral densities are computed, respectively, as follows:

$$\sigma_{sh,s}{}^2(R) = 2qG_T{}^2FM^2R_{io}[P(R) + P_{back}]L,$$
$$\sigma_{sh,d}{}^2 = 2qG_T{}^2(I_{ds} + FM^2I_{db}),$$
$$\sigma_{th}{}^2 = \sigma_{th,i}{}^2G_T{}^2, \tag{A1}$$

all in units of $V^2\ Hz^{-1}$, where $P(R)$ is the range-received power (W), P_{back} is the background-radiation power (W), R_{io} is the APD current responsivity without multiplication (i.e., at $M = 1$) (A/W), G_T is the transimpedance receiver gain (V/A), F is the excess-noise factor ( ), M is the APD multiplication factor ( ), I_{ds} is the APD surface dark current (A), I_{db} is the APD bulk dark current (A), $\sigma_{th,i}$ is the amplifier input noise current density ($A\ Hz^{-1/2}$), L are the optical losses ( ), and h, c, and q are physical constants (Planck's constant, speed of light, and electron charge, respectively). [The symbol ( ) indicates that the parameter under consideration has no units.]

From Eqs. (A1) the range-dependent SNR is calculated as

$$\frac{S}{N} = \frac{R_vLP(R)}{[2qG_T{}^2(I_{ds} + FM^2\{I_{db} + R_{io}L[P(R) + P_{back}]\})B_N + \sigma_{th}{}^2B_N]^{1/2}} \quad \left(\frac{V}{V}\right), \tag{A2}$$

where $R_v = R_{io}MG_T$ is the receiver voltage responsivity and B_N is the noise equivalent bandwidth.

In the simulations these parameters have been modeled by assuming typical noise specifications for a lidar receiver based on an APD and a low-noise preamplifier. Table 1 summarizes the default values used in the simulations.

The receiver operation is limited by signal-induced shot noise in the first part of the exploring range and by thermal noise at large ranges.

As far as the background power estimation (P_{back}) is concerned, because it is a function of the receiver effective area, receiver field of view, and optical filter bandwidth, the simulations have been based on a typical estimate of P_{back} rather than on a specific optical configuration.

From the relations above and under the hypothesis of a homogeneous atmosphere, the range R for unity SNR is determined by equating Eq. (A2) to unity.

This study has been carried out under the Spanish Interministry Committee for Science and Technology grants TIC431-93 and AMB1144-C02-02.

References

1. R. T. H. Collis and P. B. Russell, "Lidar measurement of particles and gases by elastic backscattering and differential absorption," in *Laser Monitoring of the Atmosphere*, E. D. Hinkley, ed. (Springer-Verlag, New York, 1976), Chap. 4, pp. 71–102.
2. J. D. Klett, "Stable analytical inversion solution for processing lidar returns," Appl. Opt. **20,** 211–220 (1985).
3. G. J. Kunz and G. de Leeuw, "Inversion of lidar signals with the slope method," Appl. Opt. **32,** 3249–3256 (1993).
4. M. E. Tiuri, "Radio astronomy receivers," IEEE Trans. Antennas Propag. **AP-12,** 930–938 (1964).
5. A. B. Carlson, "Signal transmission and filtering," in *Communication Systems*, 3rd ed. (McGraw-Hill, Singapore, 1986), Chap. 3, pp. 177–178.
6. R. M. Measures, *Laser Remote Sensing: Fundamentals and Applications* (Krieger, Malabar, Fla., 1992), Chap. 4, pp. 138–145.
7. H. Koschmieder, "Theorie der horizontalen Sichtweite," Beitr. Phys. Freien Atmos. **12,** 33–53 (1925).
8. P. W. Kruse, L. D. McGlauchlin, and R. B. McQuiston, *Elements of Infrared Technology: Generation, Transmission and Detection* (Wiley, New York, 1962).
9. R. T. H. Collis, "Lidar: a new atmospheric probe," Q. J. R. Meteorol. Soc. **92,** 220–230 (1966).
10. R. J. Barlow, *Statistics* (Wiley, New York, 1989), Chap. 6.
11. J. J. More, "The Levenberg–Marquardt algorithm: implementation and theory," in *Numerical Analysis*, Lecture Notes in Mathematics 630, G. A. Watson, ed. (Springer-Verlag, New York, 1977), pp. 105–116.
12. W. B. Jones, *Introduction to Optical Fiber Communication Systems*, (Holt, Rinehart & Winston, New York, 1988), Chaps. 7, 8.
13. R. J. McIntyre, "Multiplication noise in uniform avalanche diodes," IEEE Trans. Electron Devices **ED-13,** 164–168 (1966).