Parameterization of cloud lidar backscattering profiles by means of asymmetrical Gaussians

Massimo Del Guasta, Marco Morandi, Leopoldo Stefanutti

A fitting procedure for cloud lidar data processing is shown that is based on the computation of the first three moments of the vertical-backscattering (or -extinction) profile. Single-peak clouds or single cloud layers are approximated to asymmetrical Gaussians. The algorithm is particularly stable with respect to noise and processing errors, and it is much faster than the equivalent least-squares approach. Multilayer clouds can easily be treated as a sum of single asymmetrical Gaussian peaks. The method is suitable for cloud-shape parametrization in noisy lidar signatures (like those expected from satellite lidars). It also permits an improvement of cloud radiative-property computations that are based on huge lidar data sets for which storage and careful examination of single lidar profiles can't be carried out.

Key words: Lidar, signal processing, cloud

1. Introduction
A backscattering lidar is able to provide the vertical profile of a cloud with a resolution of the order of a few meters. Such a high resolution is useful for the study of cloud evolution and microphysics, but at the same time it is too high for purposes that require a simple parametrization of the profile. In such cases the magnificent lidar resolution is often rendered superfluous because the cloud's backscattering profile is approximated to a rectangular one that is limited by a sharp base and a top. Base- and top-height determinations can be carried out automatically in multilayer clouds as well (e.g., Pal et al.1).

Many clouds show an asymmetrical vertical profile with a smooth base because of water or ice precipitation. Such situations have been often observed in midlevel antarctic clouds by Del Guasta et al.,2 and they are expected from mixed-phase midlevel cloud simulations (Sassen3,4). Warm cirrus clouds have been observed by Platt et al.5 to show a typical asymmetrical backscattering profile as well. In such cases determination of the cloud base and the top level is still possible, but the meaning of such information is not so clear and the modeling of such clouds as rectangular ones seems inadequate.

In the case of good lidar profiles, a vertical-extinction profile can in principle be derived that permits the computation of some radiative properties of the cloud (e.g., the IR radiance) when the temperature profile and some basic information about the cloud composition are known. For these purposes the complete lidar extinction profile should be used in a radiative-transfer model, but this results in a waste of computer memory when huge lidar data sets are used for statistical purposes. To avoid this problem, a rectangular parametrization of clouds is often used instead of the complete profile. In thick low clouds the base and top heights are very important because radiatively they can be considered to be black bodies that are irradiating downward and upward at their characteristic temperatures, regardless of the vertical-density profile of the cloud, and a rectangular cloud-shaping parameterization is adequate. However, in optically thinner midlevel and high clouds that often extend vertically for more than 1000 m,2 the lidar-derived base and top heights do not have the same importance because the downward and upward IR fluxes are produced by the whole cloud layer. In such cases the vertical profile of the cloud becomes more important for the radiative computations, and the parametrization of the lidar profile should roughly preserve the cloud shape. As an example, a typical 2000-m-thick altocumulus cloud, as often observed in Antarctica by Del Guasta et al.,2 that has a −30 °C midcloud temperature, an averaged 532-nm extinction of 10−3 m−1, and an asymmetrical shape (because of ice precipitation) has been used by the author to compute the downward IR radiance by means of a
simplified radiative-transfer model. The computed radiance was approximately 15% larger when the real cloud profile was used than when an equivalent rectangular model of the same cloud was used.

Rather than using a rough rectangular model that is defined by three values (base, top, and averaged extinction or backscattering), we can parameterize each cloud layer in a lidar signature by a truncated asymmetrical Gaussian that is expressed through six figures (base and top heights, peak extinction or backscattering, peak position, and two standard deviations); multilayer clouds can be expressed as a sum of such functions (see below).

When a large amount of noisy lidar data is acquired (e.g., from future satellite-borne lidars) with no hope of retrieval of optical information other than the backscattering profile (i.e., there is too much noise to permit an optical-extinction retrieval) a fast and memory-saving way to store the data is required. Also for such a case, the algorithm presented here can be useful because of its stability with respect to noise and signal distortion.

2. Asymmetrical-Gaussian Method

The asymmetrical Gaussian adopted for this study is composed of two half-Gaussians with different standard deviations, $\sigma_1$ and $\sigma_2$, that are joined at the cloud's peak height, where continuity and derivability are respected (Fig. 1). It can be expressed by

$$G(z) = (S_1/\sigma_1\sqrt{2\pi})\exp\left[-\frac{(z-z_p)^2}{2\sigma_1^2}\right], \quad z < z_p,$$

$$G(z) = (S_2/\sigma_2\sqrt{2\pi})\exp\left[-\frac{(z-z_p)^2}{2\sigma_2^2}\right], \quad z \geq z_p,$$

with $S_1$ equals the area under a complete Gaussian when $\sigma = \sigma_1$, $S_2$ equals the area under a complete Gaussian with $\sigma = \sigma_2$, and the continuity condition at $z = z_p$ is valid: $S_1/\sigma_1 = S_2/\sigma_2$. The moments $M_n$ of

the curve $G(z)$ with respect to $z = 0$ are expressed by

$$M_n = \int_{-\infty}^{\infty} G(z)z^n dz / \int_{-\infty}^{\infty} G(z) dz.$$

This expression can easily be rewritten in terms of finite summations in the case of discrete data points.

The first three moments computed with the center on the cloud peak ($z_p$) are given by

$$M_{np} = \int_{-\infty}^{\infty} G(z)(z - z_p)^n dz / \int_{-\infty}^{\infty} G(z) dz.$$

The first three moments are then

$$M_{1p} = \left(2/\sqrt{2\pi}\right) \frac{S_2\sigma_2 - S_1\sigma_1}{S_2 + S_1},$$

$$M_{2p} = \frac{S_2\sigma_2^2 - S_1\sigma_1^2}{S_2 + S_1},$$

$$M_{3p} = \left(4/\sqrt{2\pi}\right) \frac{S_2\sigma_2^3 - S_1\sigma_1^3}{S_2 + S_1},$$

and they still have the boundary condition

$$S_1/\sigma_1 = S_2/\sigma_2.$$

Solving the system we have

$$M_{1p} = \sqrt{2/\pi}(\sigma_2 - \sigma_1),$$

$$M_{2p} = \sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2,$$

$$M_{3p} = -\sqrt{2/\pi}(\sigma_1 - \sigma_2)(\sigma_1^2 + \sigma_2^2).$$

These are the moments around $z = z_p$. We can convert such moments to those around the midcloud position $M_1$ by using the relations

$$M_{1b} = 0,$$

$$M_{2b} = M_{2p} - M_{1p}^2,$$

$$M_{3b} = M_{3p} + 2M_{1p}M_{2p} - 3M_{1p}^2M_{2p},$$

and when we substitute the previous findings, we have

$$M_{1b} = 0,$$

$$M_{2b} = (\sigma_2^2 + \sigma_1^2)(1 - 2/\pi) - \sigma_1\sigma_2(1 - 4/\pi),$$

$$M_{3b} = \sqrt{2/\pi}[(\sigma_1^3 - \sigma_2^3)(1 - 4/\pi)$$

$$- (\sigma_1\sigma_2^2 - \sigma_2\sigma_1^2)(12/\pi - 4)].$$

When the lidar data are processed, the cloud backscat-
scattering profile \( \beta(z) \) is computed. It is then possible to compute \( M_1, M_2, \) and \( M_3 \) for \( \beta(z) \) [or for extinction \( \sigma(z) \)] with Eq. (2), also automatically. Moments around the midcloud point can immediately be computed with Eqs. (11) and (12), in the forms
\[
M_{2b} = M_2 - M_2^2,
\]
\[
M_{3b} = M_3 + 2M_1^3 - 3M_1M_2.
\]

By inverting Eqs. (13) and (14) we find that it is then possible to retrieve \( \sigma_1 \) and \( \sigma_2 \) as functions of the moments:
\[
\sigma_2 = \frac{-(4/\pi - 1)\sigma_1 + [4/\pi - 1]^2\sigma_1 - 4(1 - 2/\pi)(\sigma_1^2(1 - 2/\pi) - M_{2b})]/2(1 - 2/\pi)}{2/\pi}.
\]

In this way \( \sigma_2 \) is expressed as a function of \( \sigma_1 \) and \( M_{2b} \). \( \sigma_1 \) can't be explicited in terms of the moments, and it remains in the implicit form of Eq. (14).

The solution of the system comprising Eqs. (14) and (15) can easily be obtained numerically if \( \sigma_1 \) is varied by steps, the corresponding \( \sigma_2 \) is computed with Eq. (15), and the residual function is evaluated:
\[
\text{res}(\sigma_1) = \left| M_{3b} - \sqrt{2/\pi} \left[ -(\sigma_1^3 - \sigma_2^2)(1 - 4/\pi) + (\sigma_1\sigma_2^2 - \sigma_2\sigma_1^2)(\frac{12}{\pi} - 4) \right] \right|,
\]
where \( \text{res} \) denotes the residual. Next one searches for the residual function's minimum, which corresponds to the solution of the system.

The maximum range of variation of \( \sigma_1 \) to be used in the procedure is determined by the square root present in Eq. (15), and the reality and positivity of solutions must be verified:
\[
0 < \sigma_1 < \left( \frac{1}{\sqrt{2.75}} M_{2b} \right)^{1/2}.
\]

Once the minimal \( \sigma_1 \) and \( \sigma_2 \) are known, it is possible to compute \( z_p \) with Eq. (7):
\[
z_p = M_1 - \sqrt{2/\pi(\sigma_2 - \sigma_1)}.
\]

To achieve complete knowledge of the fitting function, one can obtain \( S_1, S_2, \) and \( G(z_p) \) from the integrated \( G(z) \) [integrated backscattering in the case of cloud backscattering \( G(z) = \beta(z) \)] by using Eqs. (1) and (6), which leads to
\[
G(z_p) = \frac{\int G(z) dz}{\sqrt{2/\pi(\sigma_1 + \sigma_2)}}.
\]

With this method it is possible to fit any cloud with an asymmetrical Gaussian simply through computation of the first three moments of the lidar backscattering profile.

Of course, the same fitting problem could be solved with the least-squares method. In such a case the minimum of the following function is searched:
\[
\chi(a, b, c, x_0) = \int_{-\infty}^{+\infty} \left[ \log G(z) - c + a(z - x_0)^2 \right] dz + \int_{-\infty}^{+\infty} \left[ \log G(z) - c + a(z - x_0)^2 \right] dz,
\]
(20)
where \( a = 1/(2\sigma_2^2), \) \( b = 1/(2\sigma_1^2), \) and \( c = \log[S_1/(\sigma_1\sqrt{2\pi})]. \)

We search the minimum by changing, in an iterative way, the \( z_p \) positions and solving for \( a, b, \) and \( c \) with the least squares system. At each \( z_p \) step, a new solution of the system must be searched for, which requires computing that involves all the data points; this procedure wastes time when the data records are long. The solution method described in this paper overcomes this problem because an iterative procedure is required only to solve a nonlinear equation of one variable, and no further computing of the data points is necessary.

One of the main advantages of the use of Gaussians in cloud-profile parametrization lies in the scaling properties of such functions. In fact it is possible to elevate a Gaussian to any power \( m \) and still preserve the Gaussian shape. Simple scalings of only the peak value and the standard deviations are involved in this process.

Conservation of the Gaussian shape means that every relation between \( \sigma_1, \sigma_2, M_{2b}, \) and \( M_{3b} \) that was obtained above is still valid when the cloud profile \( G(z) \) (the data being fitted) is elevated to the power \( m \), so that the algorithm can work both on \( G(z) \) profiles and on \( G^m(z) \). When \( G(z) \) is elevated to a power \( m \), the signal-to-noise ratio is strongly enhanced in the transformed cloud layer, which permits better stability in the fitting procedure with respect to noise and to the offset of the backscattering profile. Of course the use of \( m \gg 1 \) leads to a nonlinearly fitted cloud profile, which yields a well-fitted cloud peak and a less-well-fitted weak and noisy part of the cloud.

When \( G^m(z) \) is used instead of \( G(z) \), two standard deviations, \( \sigma_{z_1} \) and \( \sigma_{z_2} \), are obtained with the algorithm instead of \( \sigma_1 \) and \( \sigma_2 \). The latter can be computed with the simple expressions
\[
\sigma_1 = \sigma_{z_1}/m,
\]
\[
\sigma_2 = \sigma_{z_2}/m,
\]
where \( z_p \) is the same for \( G(z) \) and for \( G^m(z) \).
3. Test of the Procedure

Several tests of the procedure have been carried out on simulated noisy cloud profiles to check the sensitivity of the retrieved $z_p$, $\sigma_1$, and $\sigma_2$ values to noise and offsets that affect the $G(z)$ data. A standard cloud with $z_p = 4000$ m, $G(z) = \beta(z_p) = 10^{-3}$ m$^{-1}$ sr$^{-1}$, $\sigma_1 = 40$ m, and $\sigma_2 = 400$ m has been adopted; the height resolution is 15 m (Fig. 2). The range in which the moments of $m(z)$ were computed was $1000 \div 7000$ m. A random noise with Gaussian distribution has been added, with a specified signal-to-noise ratio (S/N) at the cloud peak that ranges from 5 to 90 (a S/N < 10 at the cloud peak corresponds to noisy lidar data that is generally considered to be unreliable during lidar data processing).

Here, $m = 5$ has been adopted, the odd figure having been preferred because it holds the null-mean value of the noise. An offset that linearly increases with the height (starting from 0 at 1000 m) has been introduced, and the slope is variable. The slope was varied from 0 (null offset) to $2.5 \times 10^{-4}$ at 7000 m, which corresponds to 25% of the cloud peak at 7000 m. When one considers lidar signals, an offset of such magnitude can be present in $\beta(z)$ because of the errors that result from bad data processing.

Processing for simulated clouds without offsets was carried out through computation of $z_p$, $\sigma_1$, and $\sigma_2$ for 100 simulated clouds for each value of S/N (ranging from 5 to 90, with a step interval of 5). At the end, the mean values and standard deviations of such quantities were produced for each S/N value. Results are shown in Figs. 3, 4, and 5. The results show that, for S/N > 15, an error of less than 10 m for the mean $z_p$ is achieved, with a $\pm 20$-m dispersion of the retrieved values (Fig. 3). Similar errors for the mean value and similar data dispersions are present in $\sigma_2$ (Fig. 4). Dispersion of the $\sigma_1$ data is still of the order of $\pm 20$%, but the mean value is more accurate than 10 m (Fig. 5).

Fig. 2. Example of cloud reconstruction by means of the algorithm presented in this study. The original cloud (without noise) is represented by the solid curve, and the reconstruction by the dotted curve. For the original cloud, $z_p = 4000$ m, $\sigma_1 = 40$ m, and $\sigma_2 = 400$ m; for the reconstructed cloud, $z_p = 3963$ m, $\sigma_1 = 27$ m, and $\sigma_2 = 449$ m.

Fig. 3. Mean retrieved $z_p$, and standard deviation (S.D.) of the retrieved values of $z_p$ as functions of the S/N at the cloud peak. The original $z_p$ was 4000 m.

Fig. 4. Mean retrieved $\sigma_2$ and standard deviation (S.D.) of the retrieved values of $\sigma_2$ as functions of the S/N at the cloud peak. The original $\sigma_2$ was 400 m.

Fig. 5. Mean retrieved $\sigma_1$ and standard deviation (S.D.) of the retrieved values of $\sigma_1$ as functions of the S/N at the cloud peak. The original $\sigma_1$ was 40 m.
of cloud fitting for the conditions with a 25% offset at 7000 m is given in Fig. 6. The results are plotted in Figs. 7, 8, and 9. The effects of an offset are obviously changes in the cloud moments, so that drifts in \( z_p \), \( \sigma_1 \), and \( \sigma_2 \) are expected. Such drifts are not present in the case of a 3% offset slope, but they are well visible in the extreme case of a 25% slope. For high S/N ratios the drift is approximately 70 m on \( z_p \) (Fig. 7) and 160 m on \( \sigma_1 \) (Fig. 8). The effect of the offset on \( \sigma_2 \) (Fig. 9) is to cause not only a drifting (nearly 7 m over 40 m), but also a stabilization of the solutions of the algorithm: The retrieved \( \sigma_1 \) and \( \sigma_2 \) values when a weak offset is present are spread around the mean value of 40 m by approximately ±20 m, and this is also the case for high S/N’s. With a steeper offset slope (25% at 7000 m), the spreading is strongly reduced. This phenomenon is visible to a lesser extent in Figs. 7 and 8.

These tests show how stable the algorithm is with respect to noise of different kinds. This stability implies that the method is particularly suited for use with noisy lidar data processing for all cases in which single peaks in a cloud can be isolated. In Fig. 10 an example of fitting for a triangular cloud that was performed with this method is reported; it shows how the cloud is reconstructed in terms of asymmetrical Gaussians.

4. Multilayer Clouds

The presence of multilayer clouds is at least as common as that of single peaks. This is a problem for simple cloud parametrizations. For cases in which the \( n \) peaks are well separated, the single peaks can be manually or automatically split,\(^1\) and the procedure can be applied to the \( n \) single layers. The whole cloud is then parametrized by the sum of \( n \) asymmetrical Gaussian functions, each one parametrized by four figures. When peaks are not fully separated, the shown stability of the procedure with respect to noise and offset (with \( m > 1 \)) is very
clouds often show a more-complex profile. If the cloud profile is, for instance, rectangular or quite ragged, the procedure as shown is useless, and a common rectangular representation of the cloud can be less presumptuous. The problem is to assess the validity of the Gaussian method on a statistical basis. For this purpose, the algorithm has been applied to a subset of a large lidar data set that was collected in 1989 at Dumont d’Urville (DDU), Antarctica, by Del Guasta et al. The test data set contained 300 low-cloud and 200 midlevel and high-cloud profiles. Each cloud profile was averaged over 30 min, which corresponds to a horizontal spatial averaging of up to 30 km when the tropospheric winds observed over DDU are considered. This averaging obviously smooths the final lidar profile of the cloud with respect to shortly averaged profiles.

The test of the asymmetrical Gaussian method was carried out on the data set described above in the following way: For each single-layer cloud, the base and the top heights were determined, and the asymmetrical Gaussian algorithm \( m = 3 \) was applied to the backscattering profile between these two limits. The resulting fitting function was set to 0 below the base and above the top and was then rescaled to obtain an integral equal to the integrated backscattering of the real cloud. At the same time, a rectangular fitting procedure was performed on the cloud between the base and the top. We confined the Gaussian fitting to the area between the base and the top, so it was possible to compare the quality of the fitting procedures that were performed by means of the rectangular and Gaussian functions, and it was also possible to maintain the definitions of the base and the top, as is so popular among lidar-studies people. Multilayer clouds, when possible, were sliced as described in the previous paragraph, with each layer’s having been processed as a monolayer cloud. When the layer separation was not possible, the whole multilayer cloud was processed.

To estimate the quality of the fittings that were obtained with the Gaussian and rectangular functions, we used a standard estimation error (SEE) procedure defined by

\[
\text{SEE} = \left( \frac{1}{2} \sum_{i=1}^{n} [\beta_i(z) - \bar{\beta}(z)]^2 / (n - 2) \right)^{1/2},
\]

where \( \beta(z) \) is the measured lidar profile of \( n \) samples and \( \bar{\beta}(z) \) is a fitting function. The SEE represents the residual variance of the cloud profile after it is fitted. The SEE is high when the cloud profile is ragged or noisy with respect to the fitting function. Better fitting produces a lower SEE on the same cloud profile.

We computed the SEE for both Gaussian and rectangular fitting functions using the test data set. The ratio between the two SEE values was computed for all the clouds, and to the resulting histograms are shown in Figs. 12, 13, and 14. The ratio is close to or higher than 1 if the results of Gaussian fitting

---

Fig. 10. Example of a non-Gaussian cloud (triangles) fitted by means of the algorithm \( m = 5 \).

Fig. 11. Example of the reconstruction of real lidar cloud data by means of two asymmetrical Gaussians \( m = 3 \). The solid curve represents the lidar data whereas the dashed curve represents the fitting curve. Backscattering is \( n \times 10^{-6} \), altitude is \( n \times 10^4 \).

---
Fig. 12. Histogram of the ratio between the standard estimation errors (SEE) obtained with the asymmetrical Gaussian (AG) and the rectangular (rec) functions. The histogram is extended to all the test clouds.

procedure are not better or are worse than those of the rectangular one, but the ratio is lower than 1 if the asymmetrical Gaussian procedure fits the cloud profile (or layer) better. In Fig. 12 the whole data set is represented, whereas in Figs. 13 and 14 it has been split into sets for low clouds (i.e., with a base lower than 2000 m) and for higher ones. It is remarkable that fewer than 5% of the test clouds show a SEE ratio close to 1, which suggests that, for those clouds, the asymmetrical fitting procedure was inadequate. For the rest of the clouds the asymmetrical Gaussian fitting procedure represents an improvement with respect to the rectangular function. A comparison of Figs. 13 and 14 makes it evident that low clouds are better fitted with this method than are higher clouds. That is obvious in DDU clouds because most of low clouds are stratified and have thick precipitation trails, so that the cloud backscattering profile is a thin belly-shaped layer that is shifted toward the cloud top. Such clouds are poorly fitted by means of rectangular functions. Higher DDU clouds are thicker and more ragged than lower clouds. For this reason high clouds are more difficult to fit by means of asymmetrical Gaussians.

6. Conclusions
A method to fit lidar cloud profiles through the use of an asymmetrical Gaussian has been reported. It is relatively stable in the presence of noise and is suitable for use with cloud profiles that are almost belly shaped. It can be used to parametrize monolayer clouds or multilayer clouds if the cloud is split into single layers on which the method can be applied. Each cloud (or cloud layer) is parametrized by means of six figures (base and top heights, peak value, peak position, and two standard deviations). This way, cloud asymmetry is preserved in the fitting function.

Such a method proved valuable when low-cloud lidar profiles were fitted, as was observed with the data from Dumont d'Urville (DDU), and yielded a much better cloud-fitting characterization than the usually used rectangular model. Midlevel and high clouds are more difficult to fit with the new method, but in most cases an improved fit is obtained. The suggested method can be introduced as a quick routine in any lidar-data-processing software to provide an alternative way to parametrize the cloud profiles. Computing of the SEE values for both the Gaussian and the rectangular fitting functions can
assist the data processing, with the lowest value's
being used to choose the fitting function that is most
suitable for each profile.

This work was carried out in the frame of the
following projects and contributions: The Italian
Program for Antarctic Research, the Istitute Polaire
Francaise, and contract E55V CT92-0066 from the
Council of European Community. The authors
thank all the wintering teams that operated the lidar
at Dumont d'Urville, Antarctica.

References
for lidar determination of cloud base height and vertical ex-
tent," Rep. ISTS-APL-TR90-001 (Atmospheric Physics Labora-
tory, Institute for Space and Terrestrial Science, York Univer-
2. M. Del Guasta, M. Morandi, L. Stefanutti, J. Brechet, and J.
Picquard, "One year of cloud lidar data from Dumont D'Urville
(Antarctica). Part I: general overview of geometrical and
3. K. Sassen, H. Zhao, and G. C. Dodd, "Simulated polarization-
diversity lidar returns from water and precipitating mixed-
4. K. Sassen and G. C. Dodd, "Haze particle nucleation simulations in cirrus clouds, and applications for numerical and lidar
5. C. M. R. Platt, J. C. Scott, and A. C. Dilley, "Remote sounding of
high clouds. Part VI: optical properties of midlatitude and