Simple algorithm to determine the near-edge smoke boundaries with scanning lidar

Vladimir A. Kovalev, Jenny Newton, Cyle Wold, and Wei Min Hao

We propose a modified algorithm for the gradient method to determine the near-edge smoke plume boundaries using backscatter signals of a scanning lidar. The running derivative of the ratio of the signal standard deviation (STD) to the accumulated sum of the STD is calculated, and the location of the global maximum of this function is found. No empirical criteria are required to determine smoke boundaries; thus the algorithm can be used without a priori selection of threshold values. The modified gradient method is not sensitive to the signal random noise at the far end of the lidar measurement range. Experimental data obtained with the Fire Sciences Laboratory lidar during routine prescribed fires in Montana were used to test the algorithm. Analysis results are presented that demonstrate the robustness of this algorithm. © 2005 Optical Society of America

OCIS codes: 280.3640, 280.0280, 280.1100, 280.1120.

1. Introduction

Wildland fires are a major contributor of particulate matter and other pollutants to the atmosphere.1–3 High concentrations of particulate matter emitted by wildland fires often violate air quality standards. Meanwhile, under the provisions of the Clean Air Act of the U.S. Environmental Protection Agency, all states by 2005 must institute special programs to reduce the emissions of particles smaller than 2.5 μm (PM2.5) in nonattainment areas. These regulations may challenge the planned increase in the Forest Service’s and other federal agencies’ use of prescribed fire to reduce hazardous fuels.

Operative monitoring of smoke particulate dynamics and concentrations in forest fire areas would allow critical, time-sensitive information on smoke distribution and concentrations to be obtained and documented. This in turn would be helpful for prompt estimation of the scale and intensity of fires and for assessment of the visibility, air quality, and public health effects from the wildfires and prescribed fires. Lidar is an instrument that is potentially capable of measuring smoke particulate characteristics remotely, over an extended area, and in real time. To achieve these objectives, the Fire Sciences Laboratory initiated the development of a ground-based two-wavelength mobile lidar instrument. In April 2004, lidar measurements were performed during a prescribed fire in the vicinity of Dillon, Montana, in which two-dimensional distributions of smoke particulates were measured.

Similar to clouds, smoke plumes are generally marked by sharp temporal and spatial changes of the particulate concentration at the smoke plume boundaries. To process lidar data, regions with high levels of smoke backscattering must be separated from regions of clear atmosphere, and the distance from the lidar to the nearest boundary of the smoke plume should be established.4,5 In principle, lidar can easily detect the boundaries between different atmospheric layers so that there is no problem visualizing the location and boundaries of heterogeneous areas, for example, the location of the atmospheric turbid layer or clouds from lidar scans. However, use of an automatic method to select these boundaries is always an issue. The exact position of any heterogeneous layer is not well specified, and a large amount of interpretation is often involved in the selection of a value for the boundary layer or the cloud height. Different methodologies to process such data have been proposed that make it possible to discriminate the layering with increased particulate loading from clear-air areas.6–13

In most cases, the location of the boundaries of the
enhanced scattering is found by use of empirical criteria. For example, in the threshold method, the height of the atmospheric boundary layer is determined as a point where the backscatter intensity exceeds that of the free atmosphere or the Rayleigh signal by an established percent difference. This value needs to be empirically established. Different methods that analyze the gradient change at the transition zone from clear air to the turbid layer were developed that analyze the first- or second-order derivative of the range-corrected lidar signal with respect to the altitude.

In some recent studies of tropospheric structures, including investigations of the boundary-layer characteristics, the wavelet technique is used. This technique has many attractive specifics. Particularly, it can retrieve structures at a variety of spatial scales and determine multiple boundaries from lidar profiles, segregating these according to their strength and sign. However, the technique requires the selection of a particular wavelet, and this may impede its use in an automated mode. Additional issues are identifying atmospheric structures with little backscatter gradients in the transition zone and the signal noise.

In all the methods listed above, the shape of the lidar signal is analyzed, and a sharp increase or decrease in the backscatter signal intensity is considered to be a boundary of the aerosol plumes or the boundary layer. The principal problem in determining turbid layer boundaries is that, due to the large variability of atmospheric situations, the shape and intensity of the backscattered signals may be significantly different. With these methods, it is quite difficult to establish reliable criteria to discriminate a boundary between clear air and a turbid zone in an automated mode.

There is an alternative method in which the variance in the lidar signal intensity is used that allows localization of the boundary layer without use of empirical thresholds. To obtain convective boundary-layer depths and associated characteristics, the backscatter signals from each altitude range are mapped on horizontal planes, and the variance of the signal on each horizontal plane is calculated. The lowest-altitude local maximum peak of the variance profile is considered to be the convective boundary-layer mean depth. To prevent selection of a random local maximum, the behavior of the horizontal variance at five consecutive altitudes was specified. Obviously, the results obtained with this algorithm depend on both the behavior of the investigated aerosol inhomogeneity and the selected altitude resolution. A change of height resolution for the analyzed data points may produce different measurement results.

As follows from the above discussion, two functions are generally used in these methods to determine the boundaries of the aerosol structures with lidar: either the range-corrected signal or the signal variance (or the standard deviation) versus range. The problem that we met when analyzing our experimental smoke plume data was the strong diffusion of the smoke plume at distant ranges; this dramatically reduced the intensity and gradients of the backscatter signal from the smoke, impeding the reliable determination of the near-edge smoke boundary. Moreover, analysis showed that false spikes, originated by the noise at the far end of the range-corrected signal, can mask the slight increase of the backscattering at the near edge, making it impossible to discriminate the near-edge smoke boundary. This forced us into looking for alternative functions that would have an increased gradient at the near edge of the smoke plume.

In this study a special ratio function is implemented to facilitate the automatic determination of the near-edge boundaries of smoke plumes with a gradient method. As compared with the above lidar signal or variance profile versus range method, the ratio function has increased positive gradients at the near edge of the smoke and strongly suppressed false noise spikes over the far end of the measurement range. Two variants of the algorithm used to determine smoke plume boundaries with a scanning lidar are discussed. These algorithms are based on the determination of the location of the ratio derivative maximum and do not require use of numerical criteria or threshold values to determine the location of the near-edge boundaries of the smoke plume under investigation. Both variants of this algorithm were examined, and results of this examination are discussed.

2. Algorithm and Examples of Its Application for Lidar Experimental Data

A. Variant 1

The idea behind the proposed algorithm is as follows. Consider a range-corrected backscatter lidar signal $P(r)r^2$ from a synthetic clear atmosphere, which incorporates a distant turbid area over some range $r$ from the lidar, for example, from 2000 to 2600 m (Fig. 1, curve 1). Because of the sharp increase of aerosol backscattering in this area, the lidar signal has a sharp increase at the boundary of the turbid zone $r_b = 2000$ m. The integral of the signal (curve 2) also increases in the zone of the turbid air; however, the latter increase is delayed as compared with the sharp increase in the signal at $r_b$. Accordingly, the ratio of the range-corrected signal $P(r)r^2$ to its integral from some range $r_{\text{min}}$ to $r$ has a sharp spike at the boundary range $r_b$, which in turn can be additionally increased by differentiation of the ratio. Thus the function to determine the location of the boundary between the clear air and turbid (smoke) area can be written as

$$D_{\text{sign}}(r) = \frac{d}{dr} \left[ \frac{P(r)r^2}{\int_{r_{\text{min}}}^{r} P(r')r'^2dr'} \right],$$

where $r_{\text{min}}$ is the minimal range of the integration; this range can initially be selected to exclude the
nearest incomplete overlap zone over which the signal $P(r)r^2$ sharply increases with range. The shape of the function $D_{\text{sign}}(r)$ for the above example is shown in Fig. 1 (curve 3). One can see that the location of the maximum of this function coincides well with the near boundary at $r_b = 2000$ m; thus, by determining the location of the maximum, one determines the unknown boundary $r_b$. Note also that the integral in the denominator has increased values at the far end of the measurement range as compared with that close to $r_b$. Therefore the signal noise component in real range-corrected signals does not create strong local maxima in $D_{\text{sign}}(r)$ over the distant ranges that would be comparable with the spikes at the near-edge $r_b$. This specific of the ratio function is illustrated in Fig. 2, in which the same range-corrected signal $P(r)r^2$ as that in Fig. 1 but now corrupted with artificial random noise (curve 1) is given. The function $D_{\text{sign}}(r)$ for this signal is shown as curve 2. One can see that, in spite of extremely large signal noise over the far end of the measurement range, the local maxima of the function $D_{\text{sign}}(r)$ over the distant ranges remain less than the global maximum at $r_b$ and do not prevent the determination of the boundary $r_b$ in an automatic mode.

To examine the value of this method, real lidar experimental data obtained during two prescribed fires performed by the Forest Service were used. The results shown in this paper were obtained from lidar signals measured during a prescribed fire near Dillon, Montana, on 23 April 2004. The lidar, developed at the University of Iowa, was operated at the wavelengths of 355 and 1064 nm in both the horizontal and the vertical scanning mode. For an illustration of how the algorithm works, an azimuthal scan at 1064 nm was selected from a set of the lidar scans. The scan was made along the fixed elevation angle $\phi$ deg and over a wide azimuthal range $\phi$, from $\phi = 80$ deg to $\phi = 170$ deg with 1-deg steps; the signal for every line of sight was an average of 30 shots. A 12-bit digitizer was used to sample the signals with a 2.4-m range resolution. For the analysis, a simplified form of Eq. (1) was used, where the scaling factor, that is, the range resolution $\Delta r$ of the digitized signal in the denominator, was omitted:

$$D_{\text{sign}}(r) = \frac{\frac{d}{dr}\left\{\frac{P(r)r^2}{\sum_{r_{\text{min}}} |P(r)r^2|}\right\}}{P(r)r^2}.$$  

In Fig. 3 the top panel shows the range-corrected signal for the azimuth $\phi = 90$ deg from the smoke plume, in which the near-end edge is located at the range $\sim 1400$ m. Because only the increase in the range-corrected signal $P(r)r^2$ is the subject of interest when we are determining the near-edge boundary of the turbid structure, the negative values of $D_{\text{sign}}(r)$...
are omitted. The scale of the signal magnitude in arbitrary units is shown on the left side of the panel. The signal has three bulges, that is, there are three range-separated smoke plumes along this azimuthal direction. The positive values of the function $D_{\text{sign}}(r)$, which are the subject of interest, are shown in the bottom panel. Here and in Figs. 4–12, the running derivative is determined for five adjacent points, that is, over the range interval of 9.6 m. One can see three separated bulges of the function $D_{\text{sign}}(r)$ where maxima, located at the ranges 1460, 1550, and 1680 m, coincide with the near-edge boundaries of three smoke plumes. Note that the location of the global maximum of the function $D_{\text{sign}}(r)$ ($\sim 1460$ m) coincides with the range where the initial increase of the range-corrected signal takes place, so that it can be taken as the nearest boundary of the smoke $r_b$. The magnitudes of the function $D_{\text{sign}}(r)$ over the range 1200–1400 m are small as compared with that at the nearest boundary of the smoke at $r_b = 1460$ m and do not prevent the determination of the actual $r_b$ in the automatic mode. Note that actually a large number of positive spikes of the function $D_{\text{sign}}(r)$ over distant ranges (1700–2000 m) exist; however, their magnitudes are so small that they cannot be visualized unless a log scale is used.

The analysis of this variant revealed that the function $D_{\text{sign}}(r)$ nicely discriminates the individual layers in multilayer atmospheres. It also works well when we are determining the boundary $r_b$ between clear air and a dense smoke area with intense backscattering and a well-defined boundary between them. However, when working in highly inhomogeneous smoke plumes with high variations of backscattering, a number of intense local maxima may appear within the smoke-polluted area, such as shown in Fig. 3. In some cases this can impede use of the automatic mode to determine $r_b$. The intensity of the maxima depends on the level of spatial variations in the range-corrected signal, and the most intense spike will be obtained from the aerosol structures with the largest positive gradient of its magnitude. The global maximum of $D_{\text{sign}}(r)$ may not be located at the nearest boundary of interest; it can be located somewhere within the smoke plume zone. This situation is often met when the intensity of smoke backscattering at the near edge is relatively low and comparable to the backscattering from adjacent zones of clear air. Thus, to use this variant to determine $r_b$ in an automated mode, some additional numerical criteria have to be selected and applied to separate cases when the global maximum is located within the smoke-polluted areas, that is, at the ranges $r > r_b$.

B. Variant 2

Another variant of this algorithm is based on the evaluation of the signal variance at the same ranges, similar to that in the studies of Refs. 19 and 20. However, instead of using the variance profile, we apply a ratio function similar to that in Eq. (2). In particular, the function in the form

$$D_{\text{STD}}(r) = \frac{d}{dr} \left[ \frac{\text{STD}(r)}{\sum_{r_{\text{min}}} \text{STD}(r)} \right]$$

is used, where $\text{STD}(r)$ is the standard deviation of the three to five range-corrected signals at the same range $r$ for adjacent azimuths. After an analysis of the obtained results, we concluded that the function $D_{\text{STD}}(r)$ is more appropriate for the determination of the near-edge boundary of the smoke area than the function $D_{\text{sign}}(r)$. Figure 4(a) shows the same range-corrected signal for the azimuth $\phi = 90^\circ$ as that in Fig. 3(a). In Fig. 4(b) the standard deviation of the signal as a function of the range, $\text{STD}(r)$, is shown. Here a stepped standard deviation is calculated for every five adjacent lines of sight and is assigned to the central line. In Fig. 4(c) the positive values of the function $D_{\text{STD}}(r)$, which are subjects of interest, are shown. A local running derivative is determined for the five adjacent points, similar to that in variant 1. The location of the maximal positive value of the derivative ($\sim 1430$ m) where the initial increase of the $\text{STD}(r)$ takes place is insignificantly shifted relative to that determined with variant 1 (1460 m). One can also see small positive, nonzero values of $D_{\text{STD}}(r)$ around the ranges 1600 and 1750 m, where the derivative of the standard deviation in Fig. 4(b) changes its sign from negative to positive. However, here the magnitudes of the function $D_{\text{STD}}(r)$ at the ranges $r > r_b$ are much smaller than that at the nearest boundary of the smoke at $r_b = 1430$ m. In some cases, additional spikes (local maxima) of $D_{\text{STD}}(r)$ with magnitudes comparable with that at $r_b$ may appear at $r > r_b$, close to $r_b$ (Figs. 5–7). Fortunately, unlike that of $D_{\text{sign}}(r)$ obtained with Eq. (2), their magnitudes in most cases are less than that at the range $r_b$ even for sparse smoke with decreased backscattering. The local maxima that appear at $r < r_b$ are discussed below.

The most important specific of the algorithm in Eq. (3) is that it is relatively insensitive to significant changes in the intensity of the lidar signals at the smoke boundaries. In other words, the algorithm in
Eq. (3) functions properly even if the lidar return from smoke areas varies significantly during the scanning. In Figs. 4 and 5, the range-corrected signal intensity from the smoke plume areas ranges from 600 to 1150 arbitrarily selected units. In Fig. 6 the maximal signal intensity is approximately 100 arbitrary units; in Figs. 7 and 8 the intensity is 35 and 20–25 arbitrary units, respectively. These units are calculated with the same scale factor for all signals, so that the relative changes in the signals given in

Figs. 3–9 represent real changes in atmospheric backscattering. Thus the signal intensities differ here by as much as 30–50 times; however, no noticeable worsening for the determination of the boundary \( r_b \) by the \( D_{\text{STD}}(r) \) global maximum occurs.

The data analysis showed that use of the function \( D_{\text{STD}}(r) \) instead of \( D_{\text{sign}}(r) \) to determine \( r_b \) significantly reduces the likelihood of our obtaining false maxima within the area of smoke plumes at \( r > r_b \). However, in some cases of extremely inhomogeneous smoke plumes, especially when combined with a significantly reduced signal intensity at the nearest boundaries of the smoke plume, the global maximum of the function \( D_{\text{STD}}(r) \) may be located inside the plume, rather than at the boundary \( r_b \). Such a case is shown in Fig. 9. Here, because of the sharp increase in the standard deviation within the smoke plume area, the global maximum of the function takes place at the range 1474 m, whereas the actual location as determined with the first large spike would be 1320 m. In this particular case, the automated method would overestimate the boundary \( r_b \) by as much as 154 m. Fortunately, when we use the function \( D_{\text{STD}}(r) \), this generally occurs only in individual cases, and is much rarer than when we use \( D_{\text{sign}}(r) \). To correct or remove these outliers, a conventional technique of removing outliers can be applied to the retrieved function \( r_b(\phi) \) (see below).
Another issue of this method is related to the presence of the noise in the function $D_{\text{STD}}(r)$ over the clear-air zone at $r < r_b$ (Figs. 5–9). This is because in the areas of the lidar measurement range, close to the selected $r_{\min}$, the accumulated sum $\sum \text{STD}(r)$ in the denominator of Eq. (3) is small, and accordingly the function $D_{\text{STD}}(r)$ is extremely sensitive even to minor local fluctuations of the standard deviation STD($r$). In some cases, local maxima in $D_{\text{STD}}(r)$ appear close to $r_{\min}$, whose magnitudes are larger than that at the smoke boundary $r_b$. To avoid obtaining false global maxima in the nearest zone, an additional dead zone of $\Delta r = 100–500$ m, adjacent to $r_{\min}$, should be established when one is determining the location of the global maximum in $D_{\text{STD}}(r)$. In other words, the summation in the denominator of Eq. (3) should start at $r_{\min}$, but the global maximum of the function $D_{\text{STD}}(r)$ should be determined over the range starting from $r_b = r_{\min} + \Delta r$ (rather than from $r_{\min}$). To clarify the principles of the selection of the dead zone $\Delta r$, let us consider the real two-dimensional image of the smoke plume obtained with our lidar (Fig. 10).

The relative amount of backscattering not corrected for the two-way atmospheric transmission, that is, the intensity of the range-corrected signal in the arbitrary units, is defined here in gray scale given on the right side Fig. 10. The dark structures are the signal intensity from smoke plume areas with enhanced backscattering. The initial data points of $r_b$ retrieved with the selected $r_{\min} = 50$ m and $\Delta r = 360$ m are shown as white squares. One can see that, on the right side of the plot, most of the adjacent data points $r_b$ are close to each other and their location changes monotonically with the change of the azimuth. On the left side of the plot, within the azimuthal range $\phi = 127–167$ deg, some data points are highly scattered over a wide range from 350 to 800 m. Our analysis revealed that only some of these data points originated from clearly visualized local smoke tatters. Most of these points originated from slight signal fluctuations, which were difficult to interpret. To solve the problem of the bad points, some principle of separation of the weak smoke-originated variations from noise fluctuations should be established. In particular, the following procedure can be used. Unlike the good points, the adjacent noise spike data points are poorly correlated to each other; therefore the running standard deviation of $r_b$ versus azimuthal range $\phi$ will have an increased value in areas of scattered data points. This allows one to separate and remove these dubious quantities. Because these points are located at near distances, it is more effective to use the running fractional standard deviation to distinguish such outliers. In Fig. 11 such a dependence for the three-point running fractional standard deviation for the function $r_b(\phi)$ is shown as the bold curve at the bottom of the panel. Note the significant increase of this value over the range where the function $r_b(\phi)$ (dotted curve) has large deviations. Following the conventional principles of removing outliers, data points were removed where the local fractional standard deviation was three or more times larger than the mean standard deviation of $r_b(\phi)$ over the range of intense smoke backscattering (excluding only the boundary points between these areas). In Fig. 11, the data points of $r_b$ that remain after the outliers are removed are shown as black squares. Now one can find more accurate locations of $r_b$ after removal of these bad points. This can be achieved by determining global maxima of $D_{\text{STD}}(r)$ by use of an increased range $\Delta r$. In particular, the newly selected $\Delta r$ should be extended enough to put all established outliers in an enlarged dead zone. In our case this can be achieved by selecting $\Delta r = 800–900$ m. The boundaries $r_b$ for the same smoke plume as in Fig. 10 but now determined with $\Delta r = 850$ m are shown in Fig. 12.
3. Summary

In this study we propose a modified algorithm for the gradient method, in which we use a special ratio function, to determine near-edge smoke plume boundaries using backscatter signals of a scanning lidar. Two variants of the algorithm are examined to determine the location of such boundaries. In the first variant we determined the derivative of the ratio of the range-corrected signal to the integral of the signal and the location of the global maximum of this derivative. In the second variant we determined the derivative of the ratio of the standard deviation of the range-corrected signal to the integral of the signal and the location of the global maximum of this derivative. We tested both variants with experimental data obtained during prescribed fires to establish whether the algorithms can be used to automatically determine the smoke boundary without applying a priori selected criteria.

The analysis of the behavior of the function $D_{\text{sign}}(r)$ [variant 1, Eq. (2)] revealed that the algorithm nicely discriminates separate zones in highly heterogeneous, multilayering smoke plumes, yielding intense well-defined local maxima within the smoke area. However, the global maximum may be located somewhere within the smoke plume zone, rather than at the smoke near-edge boundary $r_b$, so that use of this variant to determine $r_b$ in an automated mode might be an issue, especially for relatively weak intensity backscattering. The alternative function $D_{\text{STD}}(r)$ [Eq. (3)] used in variant 2, in which such a situation is rarely met, is more appropriate for automatic determination of the near boundary of the smoke than $D_{\text{sign}}(r)$.

Analyzing variant 2, we determined the running derivative, varying the least-squares (regression) range from 4.8 to 14.4 m, and calculated standard deviations, varying the number of lines of view from three to seven. In most cases, these variations either do not influence the established locations of the global maximum of $D_{\text{STD}}(r)$ or resulted in insignificant changes in these locations; however, use of five lines of view instead three when we determined STD($r$) resulted in some reduction in the number of false maxima cases, particularly when we determined $r_b$ for relatively thin smoke plumes.

The method is relatively insensitive to the lidar signal random noise at the far end of the measurement range, enhanced by the range-squared correction, and works reliably enough even when the signal gradients in the transition zone between clear-air and smoke plume areas are reduced; however, special precautions should be taken to avoid false global maxima in the areas close to the lower limit of the integration of the function STD($r$).

This method can be useful to determine the boundary-layer top with airborne or spacecraft lidars.

References

14. C. Kiemle, M. Kästner, and G. G. Ehret, “The convective boundary layer structure from lidar and radiosonde measure-