CALIOP Algorithm Theoretical Basis Document

Part 2: Feature Detection and Layer Properties Algorithms

Primary Authors:

Mark A. Vaughan, Science Applications International Corp. (SAIC), Hampton, Virginia, USA
David M. Winker, NASA Langley Research Center, Hampton, Virginia, USA
Kathleen A. Powell, Science Applications International Corp. (SAIC), Hampton, Virginia, USA

Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations

CALIOP Algorithm Theoretical Basis Document

Document No: PC-SCI-202 Part 2

Prepared By:

_____

Mark A. Vaughan   Date

_____

Kathleen A. Powell   Date

Approved By:

_____

David M. Winker   Date
CALIPSO Principal Investigator

Table of Contents

1. Introduction

During its three-year mission, the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite will acquire a global suite of measurements from which the first observationally based estimates of aerosol direct climate forcing will be made. A key component of these measurements will be made by a two-wavelength, polarization sensitive backscatter lidar know as CALIOP (i.e., Cloud-Aerosol Lidar with Orthogonal Polarization). The global extent of CALIOP's vertically resolved aerosol and cloud measurements will enable new insights into the roles of aerosols and clouds in the Earth's climate system. The most basic parameters derived from CALIOP are the base and top altitudes of cloud and aerosol layers. In this document we describe the theoretical bases and practical implementations of the algorithms used to retrieve these important data products.

The CALIOP "feature finder" algorithm described in this document detects aerosol and cloud layers and determines layer base and top heights. In addition to locating aerosol and cloud layers (referred to generically here as "features"), the feature finder algorithm also computes a set of simple descriptive statistics, referred to as "layer descriptors", from the lidar profile data. These layer boundaries and their associated layer descriptors are the primary inputs to the scene classification algorithms described in the CALIPSO Scene Classification ATBD (PC-SCI-202 Part 3), and provide the requisite roadmap for the retrievals of range-resolved optical properties described in the CALIPSO Extinction ATBD (PC-SCI-202 Part 4).

1.1. Terminology

As the CALIPSO lidar orbits the Earth, backscatter profiles are acquired which contain signals returned from several different classes of geophysical objects. Of particular interest are those portions of the profiles backscattered from clouds, aerosol layers, and/or Earth's surface. However, in its initial scan of the profiles, our detection scheme does not attempt to distinguish among these different classes of objects. Before we attempt to classify a given object as a cloud, an aerosol, or a surface spike, we first must determine its spatial boundaries. That is, before we can confidently say *what* a thing is, we first must know *where* it is…and we cannot use specifics of the "*what*" to help us in our search for the "*where*".[1] We have therefore adopted the generic term "*feature*" to refer to all legitimate non-molecular targets that can be detected in a lidar profile. It is only after determining a feature's boundaries that we can begin to quantify the optical properties that will allow us to classify it as either cloud or aerosol. The classification process is discussed in detail in the Scene Classification ATBD (PC-SCI-202 Part 3).

For the purposes of our detection scheme we define a feature as any extended and contiguous region of enhanced backscatter signal that rises significantly above the expected "clear air" value. Clearly this definition encompasses all of our targets of interest: that is, clouds, aerosol layers, and surface returns. Because the CALIOP signal-to-noise ratio (SNR) is often low, noise excursions may have magnitudes similar to weak features. The task of the feature finding algorithm is to separate the genuine features from the pseudo-features (i.e., noise excursions) and to determine their spatial boundaries.

---

[1] Depending on the terrain and our pointing knowledge, returns from the Earth's surface can represent an exception to this dictum.

The fundamental measurement made by a backscatter lidar is a range-resolved *profile* of laser backscatter intensity. The term "profile" is also somewhat generic: it is used to refer both to the data from a single laser pulse and to the data obtained by averaging an arbitrary number of consecutive pulses. However, a profile is always a one-dimensional array of range-resolved lidar data.

As shown in Figure 1.1, multiple features of different types can be present in a single lidar profile. This profile, representing data from a single laser pulse, was acquired during the LITE mission (Winker et al., 1996) in a pass over south central Africa. In it we see clouds of varying backscatter intensities centered at 17.3 km, 14.0 km, 12.2 km, and 9.3 km, aerosol layers centered at 8.3 and 3.4 km, and a surface return at 1.4 km above mean sea level.
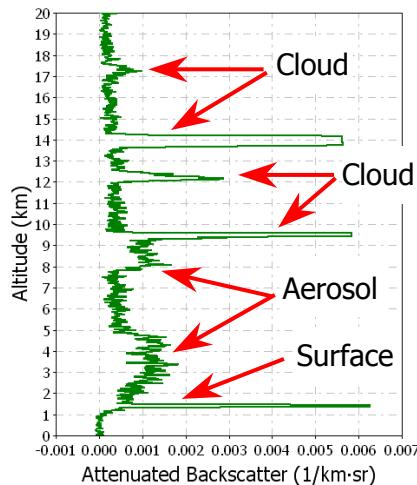
Figure 1.1: Space-based Lidar Profile acquired by LITE over South Central Africa (11.9S, 19.2E)

Individual lidar profiles contain information on the vertical distribution of features within a single atmospheric column. To augment this vertical knowledge with the corresponding horizontal information we display a large number of consecutive profiles in a single *scene*. Figure 1.2 shows a complete scene that provides the three-dimensional context necessary for evaluating the radiative significance of the features displayed in the profile shown in Figure 1.1. The vertical coordinates of each feature are given by the Y-axis values (i.e., altitude). The horizontal coordinates are given by the upper X-axis values (latitude and longitude of the lidar footprint; the lower X-axis provides a temporal coordinate given in mission elapsed time (MET)). The uncalibrated, absolute backscatter intensity is color-coded, with the weakest scattering shown in blue and the strongest scattering shown in white.
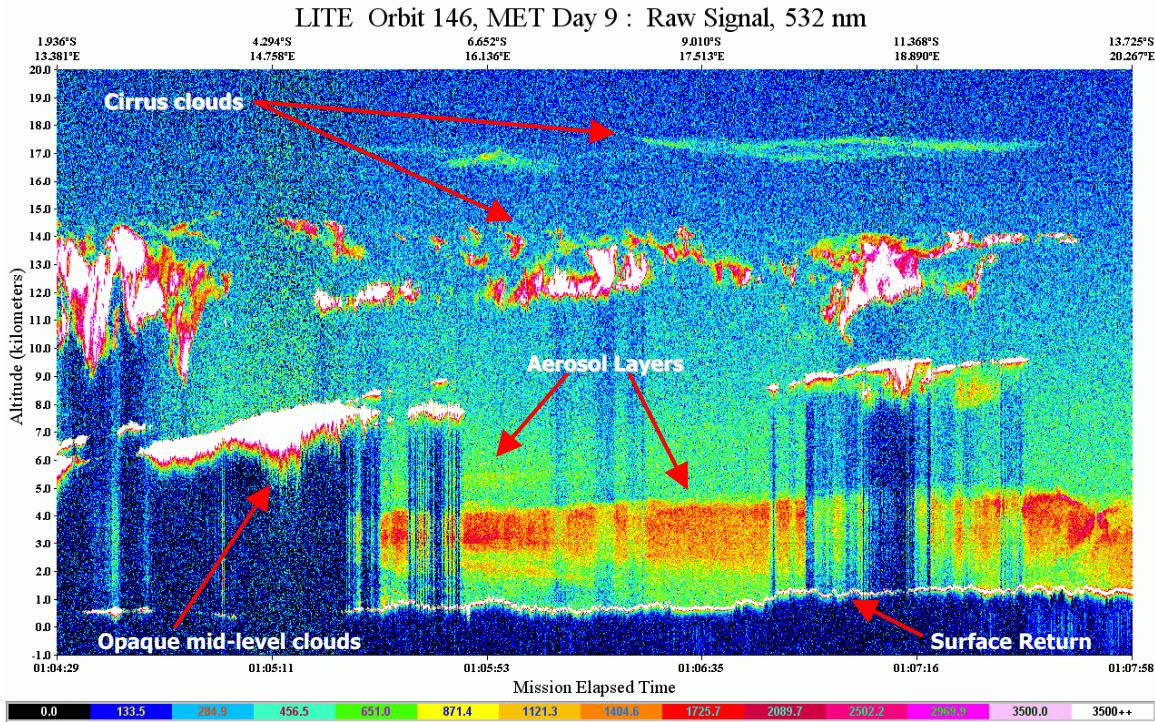
Figure 1.2: Space-based Lidar Scene from over South Central Africa

1.2. Goals

Even the most cursory examination of Figure 1.2 shows that clouds and aerosol layers can occur in a variety of characteristic spatial scales and display an enormous range of backscatter intensities. For example, the subvisible cirrus layer at 17 km is at times almost indistinguishable from the ambient molecular atmosphere (e.g. at MET 01:05:35). In contrast, the scattering intensity of mid-level clouds (e.g., at 7 km and MET 01:05:13) is frequently several orders of magnitude stronger than the molecular signal. Spatially, the fragmented cirrus layer at 14 km contains isolated patches spanning no more than a few hundreds of meters vertically and horizontally. The aerosol layer between the surface and ~4 km extends unbroken for hundreds of kilometers horizontally. An effective feature finding algorithm must be able to correctly locate the boundaries of features across this entire range of spatial scales and backscatter intensities.

Also worth noting in Figure 1.2 are the "shadows" that occur beneath optically thick features. These shadows occur in regions where the laser's energy has been substantially attenuated as the beam traverses a feature. As is evident at MET 01:06:15 and again at MET 01:07:00, as these shadows become more prominent, even the most robust features below become increasingly difficult to detect. In extreme cases – e.g., at MET 01:05:11 in Figure 1.2 – the beam is totally attenuated, so that even the surface of the Earth is no longer visible in the backscatter data. A truly first rate feature finder should be able to detect atmospheric features within the shadows beneath those features that are not completely opaque.

1.3. Background

Many of the feature finding techniques described in this document were first developed strictly for cloud detection as part of the FIRE and ECLIPS projects (Winker and Vaughan, 1994).

Following the flight of LITE in 1994, these algorithms were further developed and enhanced to reliably detect both cloud and aerosol layers in multi-wavelength data from space-based lidar (Platt et al., 1999). The feature finding algorithms developed for CALIPSO continue to build on this heritage. Built into the CALIPSO feature finder is a rudimentary understanding of the physics of light propagation through turbid media. The search routines employ an adaptive threshold technique that is combined with an iterated, multi-scale averaging scheme to locate feature boundaries across the full range of backscatter intensities.

Over the years, different researchers have proposed a number of techniques for locating feature boundaries within lidar data (Platt et al., 1994). The majority of the algorithms used within the lidar community fall into one of two categories: threshold methods and slope methods (a.k.a. zero crossing methods). Slope methods (e.g., Pal et al., 1992) make use of the range-dependent structure of the raw backscatter signal, P(r), from an up-looking lidar. For the up-looking viewing geometry the backscatter signal from a clear atmosphere decreases monotonically with range, so that the slope of the clear-air signal with respect to range, $dP(r)/dr$, is always negative. The presence of clouds and/or strong aerosols in a profile is characterized by an abrupt increase in signal level, such that at the near-range boundary of a feature (i.e., the base for an up-looking system) the sign of $dP(r)/dr$ changes from negative to positive. The slope method identifies the base of a layer using this "zero-crossing" behavior of the first derivative of the raw signal. An example of the slope method applied to up-looking lidar data is shown graphically in Figure 1.3.

In the basic slope method the location of the far-range boundary (i.e., feature top) is less well defined. For opaque features, an apparent top altitude, r_{top}, is determined to be that point where the raw signal drops below the signal level measured just beneath feature base. For transmissive features, the apparent top altitude is located at that point where $r_{top}^2 \cdot P(r_{top}) \leq r_{base}^2 \cdot P(r_{base})$. In general this approach results in a biased estimate of feature top: the apparent feature top is typically lower than the actual cloud top (Wang and Sassen, 2001). To remedy this deficiency, Wang and Sassen have proposed a more sophisticated, iterative approach to feature top determination. This approach includes comparisons to the expected "clear-sky" slope for transmissive features and to an empirically determined minimum reliable signal for opaque features.
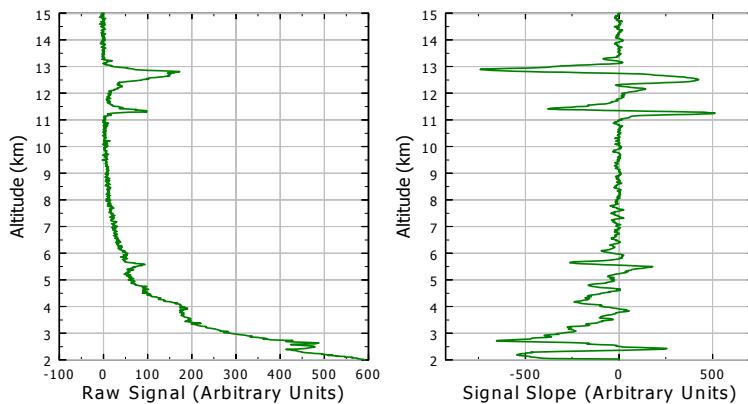

Figure 1.3: Slope method calculations for data acquired 4 September 2004 using the LaRC 8" up-looking cloud lidar.

Slope methods are especially well suited for use with up-looking, ground-based lidar systems. These instruments generally have a relatively high signal-to-noise ratio (SNR), which serves to minimize the number of false positives reported. Likewise, multi-layer profiles present no particular problem for a well-implemented slope algorithm. However, because the measured data must be differentiated numerically, the most effective implementations of the slope method are those for which the lidar generates high SNR data at a high vertical resolution. This high spatial resolution is necessary to permit well-localized estimates of the slope of the signal.

One practical advantage of slope methods is that they can be applied directly to uncalibrated raw data; no additional measurements, models, or assumptions are required. Threshold methods on the other hand require a second piece of information. As the name implies, threshold methods work by examining the signal at each range bin to see if it exceeds some (computed or specified) expectation of the maximum signal value that could be measured in "clear air". Data points are examined sequentially, and feature boundaries are determined by locating those regions for which all profile data exceed a threshold value for a distance greater than some predetermined minimum. Over the years several variants of the threshold technique have been proposed. Some early threshold methods (e.g., Melfi et al., 1985) applied an arbitrary, fixed threshold value to the range-corrected raw backscatter data. Winker and Vaughan (1994) examined the attenuated scattering ratios (see Section 3.2.1) using a range-dependent array of threshold values derived from either rawinsonde soundings or a molecular density model. Clothiaux et al. (1998) construct and continually update an archetypal clear-sky model to use in the threshold scheme they apply to micropulse lidar data.

In general, slope methods are not easily adapted for use with space-based lidars. This is due in part to the assumption of a specific profile geometry. For space-based lidars, $dP(r)/dr$ is no longer a decreasing function of range. For systems using longer wavelengths (e.g., 532 nm and greater) this is of little consequence, as $dP(r)/dr$ can be replaced with $dP(z)/dz$ which in clear air is strictly less than zero for either viewing geometry.[2] However, to make use of the zero-crossing behavior the profile of derivatives would have to be scanned beginning at the far end of the profile; that is, in the sub-surface region of a space-based profile. Beginning at R=R$_{far}$ means that any scan will be required to detect feature base first. Thus adopting this strategy would appear to introduce an additional potential for error, as for down-looking systems bases are more difficult to detect than tops precisely because the signal has been attenuated in transit through the feature.

More relevant for CALIOP, signal quality concerns further argue against the use of a slope method. While high SNR is required for the slope method to function properly, CALIPSO's orbital velocity of ~7 km/sec combined with the need to retrieve feature boundaries at high vertical and horizontal resolutions allows only minimal averaging. Figure 1.4 shows the results obtained from a typical set of slope calculations done using a LITE profile obtained by averaging over 7.4 km horizontally (10 laser pulses; see Figure 1.4a). Despite the fact that LITE data has a substantially higher SNR than CALIOP, it is still quite obvious (Figure 1.4c) that the level of

---

[2] Note though that for the shorter wavelengths – e.g., the 355 nm channel used in LITE – P(r) is no longer monotonic and therefore neither $dP(r)/dr$ nor $dP(z)/dz$ is a uniformly decreasing function, a condition which violates the fundamental assumption underlying the slope method.

random noise in the signal renders the slope information ambiguous at best. In the clear air regions there is no evidence of the uniformly negative values on which the slope method depends; instead, the slope oscillates wildly about zero.
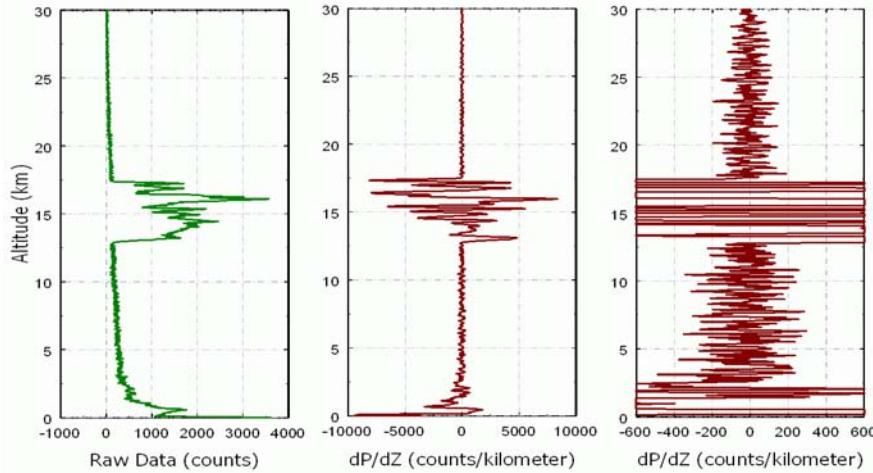
Figure 1.4: Slope calculations (dP/dz) computed over 0.25-km (17 data points) using LITE data: (a) raw data from LITE orbit 14; (b) slope with respect to altitude of the data shown in (a); and (c) a close-up view of the "clear air" slopes shown in (b).

1.4.   Scene Processes vs. Profile Processes

The vast majority of published feature finding strategies – including slope methods, threshold methods, and even new approaches adopting wavelet techniques (Davis et al., 2000) – are what might be called *profile processes*. In each of these schemes the actual feature-finding techniques are a form of one-dimensional line search applied to individual lidar profiles. Similarly, processes that can take advantage of the contextual information provided by a continuous time series of profile measurements are called *scene processes*. In this sense, the standard example of a scene process would be the edge detection techniques used in image processing, which are generally implemented as two-dimensional gradient search routines (e.g., Gonzalez & Wintz, 1977). While a number of sophisticated edge detection algorithms are available, with respect to the spatial analysis of lidar data it is critical to understand that not all edges and/or gradients are of equal interest. This situation is illustrated by the results shown in Figure 1.5. The upper portion of Figure 1.5 shows an extensive dust layer measured during LITE orbit 83. The lower portion shows the corresponding feature "boundaries" determined by the application of edge detection techniques derived from the work of Suk and Hong (1984). As seen most clearly in the central portion of the figure, edges are identified not only at the boundaries between features and clear air, but also within the interior regions of layers, wherever the signal exhibits an unusually steep gradient. This sort of behavior is part of the reason that studies comparing the two approaches (e.g., Galbraith, 1996) generally conclude that image processing techniques should be rejected in favor of established threshold methods. Therefore, in order to take full advantage of the contextual information offered by horizontal proximity, the CALIPSO feature-finder employs an iterated, multi-scale profile process in the analysis of each scene.
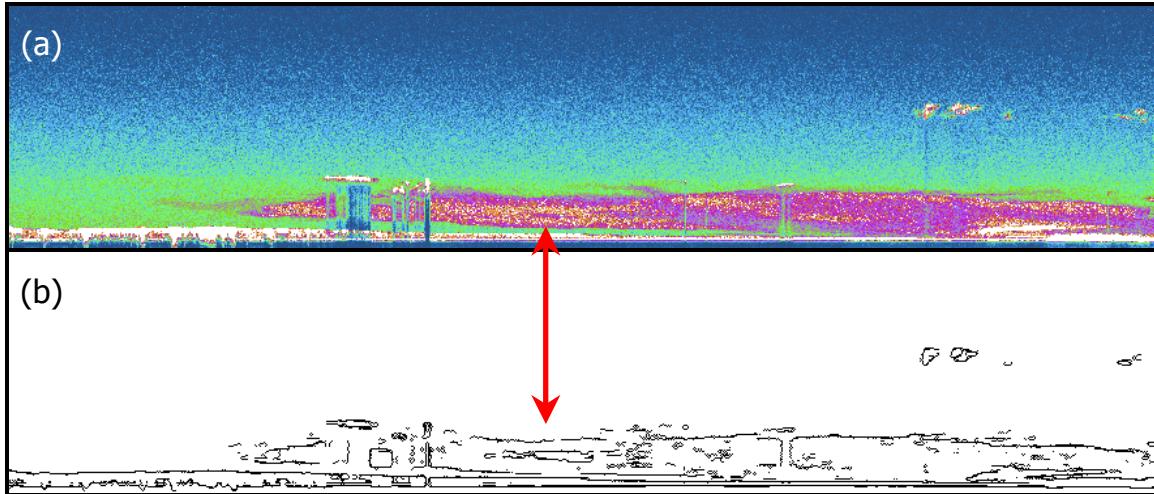
Figure 1.5: (a) 532 nm attenuated backscatter coefficients measured during LITE orbit 83; the purple and yellow colors show an extensive layer of Saharan dust; (b) results of the Suk and Hong edge detection algorithm when applied to the data shown in (a). Implementation of the edge detection algorithm and the associated images are both courtesy of Steven P. Palm, SSAI and NASA GSFC.

2. Detection Theory

By considering the cloud detection problem to be one of detecting a target against a noisy background (where the noise may be background noise, detector noise, or the statistical noise of the lidar signal) we can estimate the threshold of detectability for a given lidar instrument using standard target detection theory (Kingston, 1979). For aerosols, and especially for weak aerosol layers, the return signal must be heavily averaged before detection is possible. So in the case of greatest interest – determining the ultimate detection limits – we are in a regime where Gaussian statistics are a good approximation. For the 1064 nm channel, using an avalanche photodiode (APD) detector, the detector noise current is large enough that Gaussian statistics are always a good approximation. Strong cloud returns, which are much greater than the background return, can be identified with minimal or even no averaging. The photomultiplier (PMT) signal statistics in this case are not Poisson but Neyman type-A (Teich, 1981), because of the multi-stage multiplication process. However, in nearly all cases Gaussian statistics are a reasonable approximation (Liu, 2002). Both Poisson and Neyman-A distributions are skewed toward large values and Gaussian statistics provide a conservative estimate of detectability. Therefore, in the following we assume detector noise is Gaussian and that signal levels are high enough that both signal and background noise are also Gaussian.

The primary sources of noise of concern are detector dark current and amplifier noise, statistical fluctuation of the sunlit background, and the statistical fluctuation of the lidar return signal itself. On moonless nights, the noise in receiver channels using PMTs is essentially that due to the statistical fluctuations in the optical signal itself. During daytime, the noise in the PMT channels is determined by the background illumination. The noise in the 1064 nm channel, which uses an APD, is dominated by dark current at night. During daytime, both dark current and the solar background can be significant. Standard statistical theory has been used to derive criteria for differentiating the enhanced signal from aerosol or cloud features from random noise spikes.

This is then used to estimate the minimum detectable feature backscatter coefficients at 532 nm and 1064 nm as a function of the number of profiles averaged.

We define the standard deviation of the return signal from a molecular atmosphere, σ_m, as

$$\sigma_m = N_f \sqrt{\sigma_{sm}^2 + \sigma_b^2 + \sigma_n^2} \; . \tag{2.1}$$

where σ_{sm}, σ_b, and σ_n are the standard deviations of the statistical fluctuations of the lidar return signal, the solar background illumination, and the dark current, respectively. N_f is the detector noise factor. For the return from a feature we have

$$\sigma_c = N_f \sqrt{R \sigma_{sm}^2 + \sigma_b^2 + \sigma_n^2} \; . \tag{2.2}$$

where R is the scattering ratio within the feature. The probability distributions for the signal from within the feature, s_c (Rayleigh plus cloud), and the molecular signal, s_m (Rayleigh only), are shown in Figure 2.1. The molecular signal, s_m, can be written as the number of photoelectrons detected in a range bin, n_m. The magnitude of the RMS noise, σ_{sm}, is the square root of n_m. Similarly, the other noise terms, σ_b and σ_n, can be expressed in terms of the number of detected photoelectrons.

In general, a threshold signal, s_t, is defined such that if the signal at some range bin exceeds the threshold we identify that event as a feature. Then the probability of detection, P_d, is

$$P_d = 1 - \frac{1}{\sqrt{2\pi \sigma_c^2}} \int_{s_t - s_c}^{\infty} e^{-s^2/2\sigma_c^2} \, ds \tag{2.3}$$

and the false alarm probability, P_{fa}, (the probability the event is only a noise spike) is

$$P_{fa} = \frac{1}{\sqrt{2\pi \sigma_m^2}} \int_{s_t - s_m}^{\infty} e^{-s^2/2\sigma_m^2} \, ds \; . \tag{2.4}$$

These relations are shown graphically in Figure 2.1. The shaded areas to the right of s_t in Figure 2.1 represent P_d and P_{fa}. By increasing s_t, the false alarm rate is reduced while also decreasing the detection probability. The integrals in equation (1) and (2) can be easily evaluated using the complementary error function, erfc(x)

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} \, dt \; , \tag{2.5}$$

which allows (1) and (2) to be rewritten as (respectively)

$$P_d = 1 - \frac{1}{2} \text{erfc}\left(\frac{s_c - s_t}{\sqrt{2}\,\sigma_c} \right) \tag{2.6}$$

and

$$P_{fa} = \frac{1}{2} \text{erfc}\left(\frac{s_t - s_m}{\sqrt{2}\,\sigma_m} \right) . \tag{2.7}$$

From this, we see that it is convenient to parameterize P_d and P_{fa} in terms of, respectively,

$$x_d = (s_c - s_t) / \sigma_c \qquad (2.8)$$

and

$$x_{fa} = (s_t - s_m) / \sigma_m . \qquad (2.9)$$

As an example, $x_d = x_{fa} = 1.28$ gives a probability of detection of 90% and a false alarm probability of 10%.
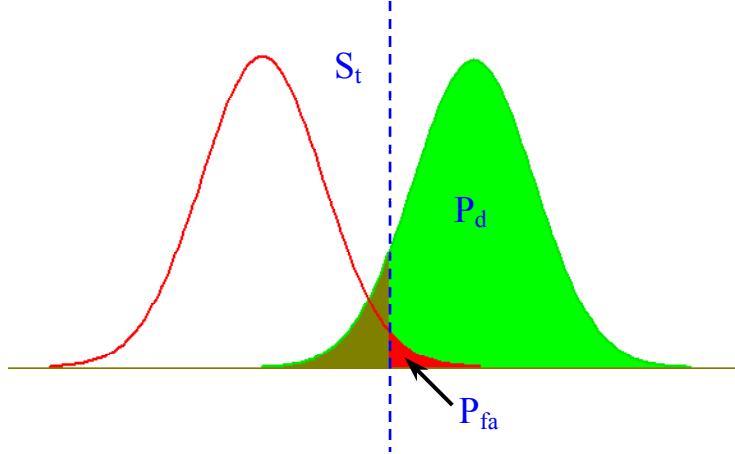
Figure 2.1: Detection probability diagram

2.1.   Target Discrimination

Using the definitions for x_{fa} and x_d given above, we can write

$$s_t = s_c - x_d \, \sigma_c = x_{fa} \, \sigma_m + s_m . \qquad (2.10)$$

Writing $s_c = R \, s_m$ and expanding σ_c we can write:

$$R \, s_m - x_d \, N_f \sqrt{R \sigma_{sm}^{\ 2} + \sigma_n^{\ 2} + \sigma_n^{\ 2}} - x_{fa} \, \sigma_m - s_m = 0 . \qquad (2.11)$$

Solving this equation for R gives the minimum scattering ratio which is detectable for the given noise levels. The solution to the above equation is:

$$R_{min} = 1 + \frac{x_d^2 N_f^2}{2 s_m} + \frac{x_{fa} \sigma_m}{s_m} + \sqrt{\left(1 + \frac{x_d^2 N_f^2}{2 s_m} + \frac{x_{fa} \sigma_m}{s_m}\right)^2 + \left(\frac{x_{fa} \sigma_m}{s_m} + 1\right)^2 + \left(\frac{x_d N_f}{s_m}\right)^2 \left(\sigma_b^2 + \sigma_n^2\right)} \qquad (2.12)$$

where the criteria for false alarms and probability of detection are specified by x_{fa} and x_d.

The minimum detectable particulate backscatter coefficient, β_{min}, is then given by

$$\beta_{min} = \beta_m \left(R_{min} - 1\right) . \qquad (2.13)$$

The SNR required to detect a feature is not a constant but varies with the scattering strength of the feature and with the detection criteria specified by x_d and x_{fa}. As the feature scattering increases, the contrast with respect to the noisy background increases and the feature can be detected with lower SNR. Figure 2.2 shows how the required SNR varies with backscatter

strength for features located at two altitudes. The SNR required for detection decreases with altitude because the molecular component of the return signal, which acts as a noisy background, decreases.
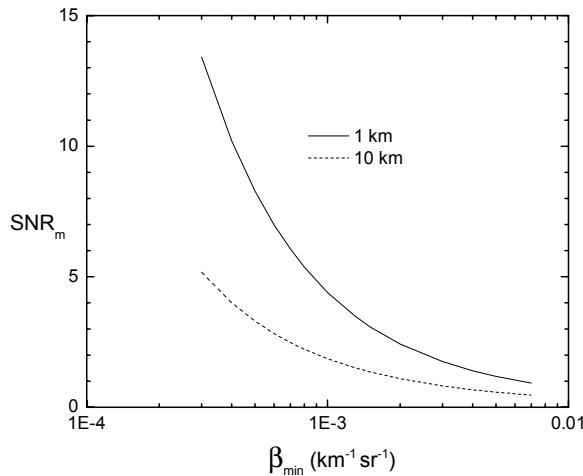
Figure 2.2: SNR required to detect layer with 532 nm backscatter coefficient of β_{532} (km^{-1} sr^{-1}) at two altitudes, 1 km and 10 km, parameterized in terms of the SNR in clear air outside the layer, SNR$_m$. This example for night conditions with 90% detection probability and 10% false alarm rate ($x_d = x_{fa} = 1.28$).

In the limiting case where background light, detector noise, and dark current are all negligible, R_{min} can be written in terms of the SNR in the clear air outside the feature, SNR$_m$. In this case, equation (2.12) becomes:

$$R_{min} = \left(\frac{x_d N_f}{2\,SNR_m} + \sqrt{1 + \left(\frac{x_{fa} N_f}{SNR_m} \right) + \left(\frac{x_d N_f}{2\,SNR_m} \right)^2} \right)^2 \qquad (2.14)$$

which applies to the PMT channels at night.

Detection sensitivity for the CALIOP 532 nm channel computed using equation (2.12) is shown in Figure 2.3 for day and night as a function of horizontal averaging. These predictions are for a detection probability of 90%, a false alarm rate of 10%, and vertical averaging to a 60 m resolution. If additional averaging is performed in the vertical, requirements for horizontal averaging are reduced. A column albedo of 5% is assumed. Horizontal red bars indicate typical scattering coefficients for different cloud types and for aerosols, based on observations by LITE. The two ends of the bar for aerosol represent the 5th and 95th percentiles of the aerosol distribution derived from nighttime LITE observations averaged to 7 km horizontal resolution.
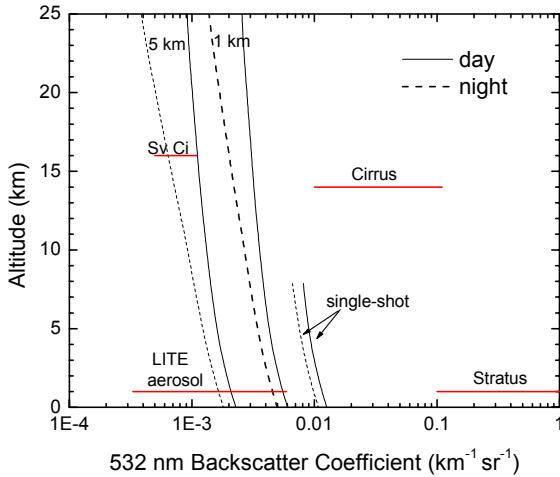
Figure 2.3: Detection sensitivity of CALIOP 532 nm channel on single shot profiles and profiles horizontally averaged to 1 km and 5 km

Minimum detectable backscatter coefficients decrease with altitude because the magnitude of the molecular component of the backscatter signal, which represents a background noise, decreases. Virtually all boundary layer clouds will be detectable at the full resolution of the downlinked data (30 m x 333 m), which allows precise cloud-clearing of boundary layer aerosol data. Only unusually high aerosol concentrations will be detectable at the highest resolution however. Figure 2.3 also indicates that most cirrus can be detected, day or night, at the full resolution of the downlinked data (60 meters by 1 km above 8 km altitude). Even some subvisible tropopause cirrus is detectable at 5-km horizontal resolution, but further averaging is required for good detection of aerosol.

Figure 2.4 shows the increase in sensitivity gained by averaging to 20 km or 80 km resolution. At 80-km resolution, CALIPSO achieves roughly the same sensitivity as LITE at a 7 km resolution.
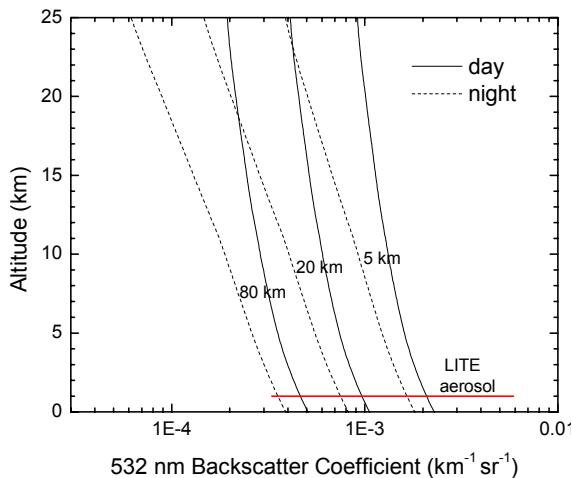
Figure 2.4: Detection sensitivity of CALIOP 532 nm channel on profiles horizontally averaged to 20 km and 80 km

Figure 2.5 compares the sensitivity of the 532 nm and 1064 nm channels. The APD detector used in the 1064 nm channel has much higher dark noise than the PMT detectors used in the 532 nm channels. The sensitivity of the 1064 nm channel is limited in most situations by the detector dark current, so the sensitivity shows much less variation between day and night and with altitude than the 532 nm channel. The APD detector has much higher quantum efficiency than the PMT detectors however, so the 1064 nm channel has better sensitivity than the 532 nm channel in the boundary layer, where signals are stronger. The 532 nm channel is much more sensitive than the 1064 nm channel in the upper troposphere and stratosphere, especially at night.
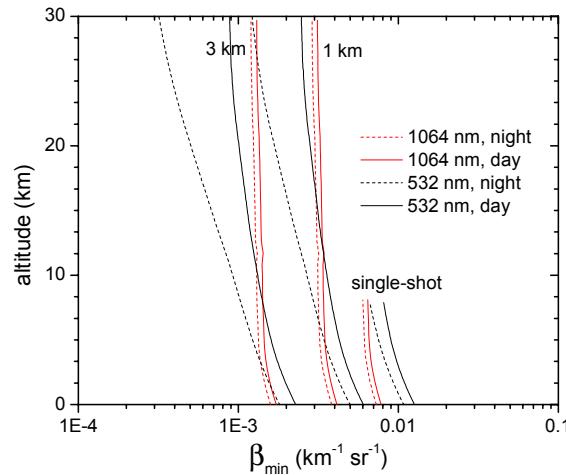
Figure 2.5: Comparison of detection sensitivity of 532 nm and 1064 nm channels

2.2. Enhancements to simple threshold detection

A practical detection algorithm requires a number of enhancements to simple threshold detection. Some of the more important ones are briefly outlined here, and discussed in more detail later.

(a) In practice, the finder algorithm must have a very low false-alarm rate to be useful. If the discrimination threshold is increased to make the false alarm rate very small, the probability of missed detection becomes unreasonably large. The true statistics of the background are quasi-Poisson with a long tail that exacerbates the false alarm problem. Even if the statistics of the background (solar plus noise) are strictly Gaussian, special measures would still have to be taken to reduce the false alarm rate. The vertical resolution of the lidar, 30 m, is less than the thickness of most layers found in the atmosphere – typically 200 m or greater for thin cirrus and at least several hundred meters for aerosols. Noise is uncorrelated from bin to bin, whereas a feature will fill at least a few bins and produce a return signal that exhibits a greater degree of correlation than noise. Therefore, requiring that a layer have a two or more successive samples which exceed the threshold provides significant discrimination against noise spikes.

As an example, consider a case where the signal threshold is set equal to the mean Rayleigh signal, $s_t = s_m$, which gives $P_{fa} = 0.5$. Because we assume the noise fluctuations are uncorrelated between samples, the probability of N successive samples exceeding the threshold is P_{fa}^{N}, as shown in Table 2.1.

Table 2.1: Detection probability as a function of averaging interval

N	$(P_{fa})^N$	$N\Delta z$ (meters)
1	0.50	30
2	0.25	60
4	0.0625	120
6	0.0156	180

The probability of N successive samples all exceeding s_t decreases rapidly as N increases, so that requiring features to have a minimum number of successive samples above threshold greatly reduces the false alarm rate.

(b) Strongly scattering cloud layers – such as altocumulus – which are thinner than this minimum thickness may be encountered. Because of this, a second discrimination threshold is defined for layers thinner than the 'minimum' layer thickness. This second threshold is set large enough to rule out false alarms from background noise.

(c) Because the signal noise varies with the background illumination, the threshold must be recomputed on each profile.

(d) Even in clear air the computed threshold must be altitude dependent. It must also be adjusted for attenuation below layers with significant attenuation.

3. Physical Models I: the CALIPSO Profile Scanner

This section will describe the general architecture of the CALIPSO profile-scanning algorithm, and will include flow charts depicting all major components. All necessary assumptions will be identified, and an overview of the required inputs and outputs will be provided. Detailed descriptions of inputs and outputs are contained in PC-SCI-202 Part 5 (Lidar Level 2 Data Products Appendix).

Figure 3.1: Simple Profile Scanning Flowchart

At the most abstract level (see Figure 3.1), the profile-scanning algorithm must accomplish three basic tasks: (1) retrieve the necessary input data; (2) search the profile(s) to find all features within that data; and (3) export the search results to the overall data processing system. Section 3.1 below will list all of the requirements for the "Retrieve Data" task. The mechanics of finding features within lidar profiles will be described in Section 3.2. Section 3.4 will describe the various output data products that will be compiled and made available both to end-users and to other algorithms.

3.1. Required Input Data

The profile scanner is applied to a range-resolved array of *attenuated scattering ratios*. For each profile examined these data are calculated as part of the algorithm's internal data storage. Forming the attenuated scattering ratios requires a profile of attenuated total backscatter

coefficients (from the Level 1 processing) and access to the same meteorological profile data used during the Level 1 processing (see PC-SCI-202, Part 5: Ancillary Data Products). All meteorological parameters are derived from model data obtained from NASA's Global Modeling and Assimilation Office (GMAO). The raw data from the GMAO is geolocated and regridded as required by the CALIPSO meteorological data server. Often referred to as the CALIPSO "Met Manager", this server functions as the sole source for all CALIOP-compatible meteorological profiles used anywhere within the CALIOP Level 1 and Level 2 data analyses.

A complete list of all input data required by the profile scanner is given below:

- an altitude array (z)

- profiles of attenuated backscatter coefficients for the 532 nm parallel $\left(\beta'_{532,\parallel}(z)\right)$ and perpendicular $\left(\beta'_{532,\perp}(z)\right)$ channels and the 1064 nm channel $\left(\beta'_{1064}(z)\right)$

- profiles of "clear air" volume backscatter coefficients $\left(\beta_{\text{air},\lambda}(z)\right)$, volume extinction coefficients $\left(\sigma_{\text{air},\lambda}(z)\right)$, and two-way transmittances $\left(T^2_{\text{air},\lambda}(z)\right)$ at each wavelength (from the met manager; note that in this context "clear air" includes contributions from molecules and ozone only)

- knowledge of the surface altitude (from a digital surface elevation map) and footprint position information (latitude & longitude)

- data acquisition time and a knowledge of the on-board lighting conditions (i.e., day or night) to help select the appropriate search constants

- knowledge of the background signal levels and the RMS variation of the baseline signals (computed on-board the satellite, and retrieved from the Payload Health & Status Data (PHSD))

- base altitude of the 532-nm calibration region (typically 30 km)

- an approximating function, $S^*(T)$, that will return the estimated effective cloud extinction-to-backscatter ratio, $S_c(T)$, for temperature T, where $T = T(z)$

- temperature profiles, $T(z)$ (note that if $S^*(T)$ is constant, the temperature profile requirement would vanish)

- a set of user-specified coefficients, described below, that controls the behavior of the scanning algorithm.

3.2. Finding Features in Lidar Profiles

The CALIPSO profile scanner is a threshold algorithm, similar to that used to detect clouds in radar data (Uttal et al., 1993) and ground-based lidar data (Winker & Vaughan, 1994). In concept, the threshold method is quite simple. A threshold level is established (either defined arbitrarily or chosen according to some heuristic), and the profile data is scanned beginning in some "known to be clear" section of the return. Data points are examined sequentially, and feature boundaries are located by determining those regions where the profile data exceeds the

threshold value for all points within some predetermined altitude range (i.e., some minimum feature thickness).

By default, the CALIPSO profile scanner is applied to 532 nm attenuated scattering ratios. While there are a number of different representations of the lidar backscatter data to which we could apply the profile scanning technique, the 532 nm attenuated scattering ratios offer certain structural advantages that allow us to implement a robust search routine.

3.2.1. Retrieving an Attenuated Scattering Ratio Profile

We construct the attenuated scattering ratios by dividing the measured total attenuated backscatter coefficients at 532 nm by a model of the attenuated backscatter coefficients that would be expected in perfectly "clear air". To compute the 532 nm total attenuated backscatter coefficients we simply add the contributions from the parallel and perpendicular components, as in equation (3.1).

$$\beta'_{Total}(z) = \beta'_{\parallel}(z) + \beta'_{\perp}(z) = \left(\beta_p(z) + \beta_m(z)\right) \cdot T_p^2(z) \cdot T_m^2(z) \cdot T_{O_3}^2(z). \qquad (3.1)$$

The subscripts in equation (3.1) represent scattering and/or attenuation due to particulates (p), molecules (m), and ozone (O_3).

Our clear air model is obtained from the CALIPSO Met Manager data, as follows:

$$\beta'_{air}(z) = \beta_m(z) \cdot T_m^2(z) \cdot T_{O_3}^2(z). \qquad (3.2)$$

All constituents of $\beta'_{air}(z)$ are derived from molecular and ozone number density profiles obtained from the Met Manager's meteorological data. (See the Met Manager ATBD for details.)

Dividing equation (3.1) by equation (3.2) gives us our expression for the attenuated scattering ratios.

$$R'(z) = \frac{\beta'_{Total}(z)}{\beta'_{air}(z)} = \left(1 + \frac{\beta_p(z)}{\beta_m(z)}\right) \cdot T_p^2(z). \qquad (3.3)$$
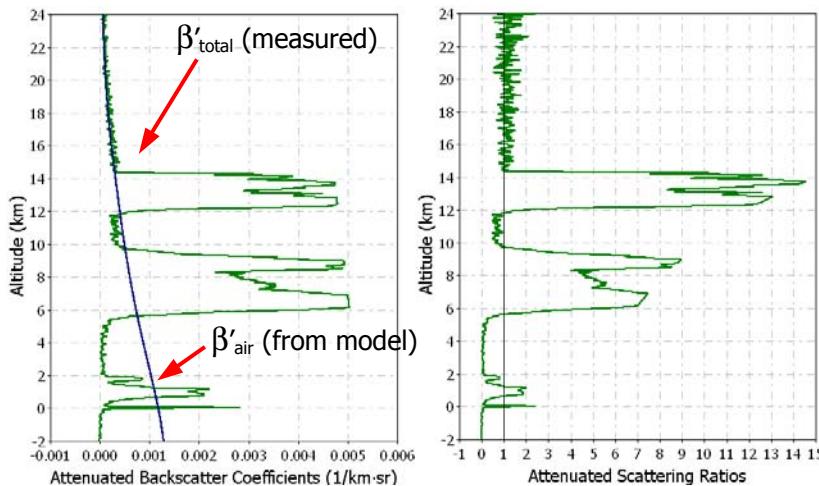

Figure 3.2: LITE Attenuated Backscatter Coefficients and Attenuated Scattering Ratios

The transformation from attenuated backscatter coefficients to attenuated scattering ratios is illustrated in Figure 3.2. We can see from equation (3.3) that in "clear air" – i.e., those regions where $\beta_p(z) = 0$ – the attenuated scattering ratio is equal to the two-way transmittance due solely to particulates $\left(T_p^2(z)\right)$. Consequently, if we assume noise-free data then within extended regions of clear air the slope of the attenuated scattering ratio with respect to altitude is zero. Despite the noise inherent in all real-world signals, this behavior is clearly evident in Figure 3.2 between 2 and 4 km, again between 10 and 12 km, and finally everywhere above 14.5 km. In contrast, the attenuated backscatter coefficients exhibit no such characteristics: signal levels change within each clear air region, and their rate of change (i.e., their slope) is different in each of the three signal regions specified above.

A profile of attenuated scattering ratios can thus be seen as a series of "features and flats". This sort of profile geometry is much more tractable than what is found in the attenuated backscatter coefficients, and is therefore much more amenable to the sort of fully automated search routines that can be implemented in a computer.

3.2.2. Threshold Array Basics

Due to the nature of the noise imposed on the lidar backscatter signal, a single attenuated scattering ratio threshold is generally inadequate for dealing with an extended altitude range. The CALIPSO threshold array is therefore a function of altitude, with lower threshold values in regions of (relatively) high clear air SNR, and higher threshold values in low clear air SNR (e.g., high altitude) regions.

Ideally the threshold array would be specified in terms of the expected variation of the "clear air" backscatter signal as a function of altitude. For a well-behaved signal we can partition the total contributions to this variation into two distinct categories. The first category consists of those noise sources whose contributions remain constant with respect to range from the lidar. Included in this segment of the noise budget are such things as detector dark current and the noise due to solar background contamination. In the second category we find those sources for which the contributions are range dependent. For the purposes of constructing a search threshold, the range-dependent component is assumed to be restricted to the Poisson- distributed photon arrival rates associated with backscatter of laser light from a molecular atmosphere having a range dependent number density. The following procedure assesses the noise contributions from both categories and combines them to compute a range-dependent *array of backscatter coefficient threshold values* to be used in scanning a lidar profile. For application in the profile scanner the backscatter coefficient threshold array is divided point-by-point by the clear air model obtained from the Met Manager to create an *attenuated scattering ratio threshold array*.

Procedure 1: Computing Threshold Values

A. The range-independent contributions can be approximated very accurately if we can determine the RMS deviation of the total signal in some altitude region where the expected backscatter contribution approaches zero. For the CALIPSO lidar this measurement is made on-board the satellite by computing the standard deviation of the signal in the upper baseline region (~70 to 80 km) for every laser pulse. This *measured backscatter variation* (MBV) is telemetered back to Earth at full resolution in the Payload Health and Status Data (PHSD).

If for some reason an on-board measurement of MBV is not available, a coarser, albeit still quite useable, approximation is obtained by computing the standard deviation of the measured attenuated backscatter coefficients over the highest CALIPSO altitude region (30.1 to 40.0 km).

Note that in either case the MBV is a standard deviation, with units of attenuated backscatter. Note too that, due to the nature of Poisson-distributed noise, simply adding a constant offset (i.e., MBV) to the clear air backscatter model will **not** produce an effective threshold array. For Poisson-distributed signal, as the signal level rises, so too does the magnitude of the corresponding noise envelope. This effect is demonstrated in Figure 3.3, which shows a "clear air" attenuated backscatter profile acquired during the LITE mission (in green) together with a clear air model (in blue) and a threshold array (in red) built using Th(z) = $\beta_{air}(z)$ + 3·MBV.[3] While the threshold array rises substantially above the noise envelope of the signal in the upper altitudes, below ~12 km it falls below the mean noise excursion, and substantially below the peak noise excursions.
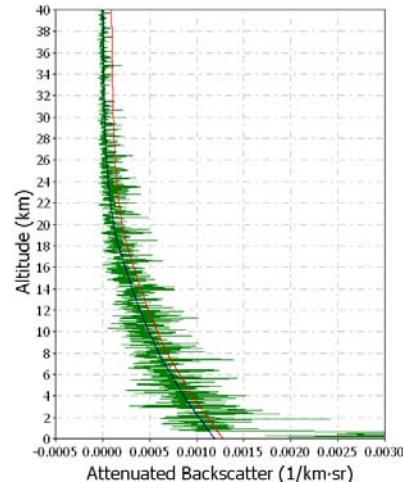
Figure 3.3: LITE Data with Threshold = $\beta_{air}(z)$ + 3·MBV

B. To compute the range-dependent component of the threshold array we start with the assumption that the molecular backscatter signal is Poisson distributed. Recall that for Poisson statistics, the signal-to-noise ratio (SNR) varies as the square root of the signal. Therefore we can compute the profile SNR relative to the SNR at the maximum sample altitude as follows

$$SNR_{relative} = \frac{\sqrt{\beta'_{air}(z)}}{\sqrt{\beta'_{air}(z_{max})}} . \tag{3.4}$$

Note that $SNR_{relative}$ is a decreasing function of altitude, and has a minimum value of 1.00 at the highest sample altitude (z_{max}).

---

[3] In the examples shown in this document, MBV is computed as the standard deviation of the attenuated backscatter coefficients between 30 and 40 km.

Once we have the relative SNR we can compute something that might be called the "relative standard deviation" or the *relative backscatter variation* (RBV). To do this, recall that

$$SNR(z) = \frac{mean(z)}{StDev(z)} \text{ so that } StDev(z) = \frac{mean(z)}{SNR(z)}.$$

Similarly

$$StDev_{relative}(z) = \frac{mean(z)}{SNR_{relative}(z)} \tag{3.5}$$

so that

$$RBV(z) = \frac{\beta'_{air}(z)}{SNR_{relative}(z)}. \tag{3.6}$$

Substituting (3.4) into (3.6) and simplifying yields

$$RBV(z) = \sqrt{\beta'_{air}(z) \cdot \beta'_{air}(z_{max})}. \tag{3.7}$$

RBV also has units of attenuated backscatter ($km^{-1} \cdot sr^{-1}$). As was the case with the MBV parameter, RBV alone is **not** sufficient to produce an effective threshold array. Figure 3.4 shows the same attenuated backscatter profile presented in Figure 3.3, only now the threshold array (in red) is computed using Th (z) = β_{air}(z) + 3·RBV(z). The behavior of the threshold in this case is essentially the opposite of that in Figure 3.3: the threshold is sufficiently offset from the clear air model in the lower altitudes, but falls unacceptably low as the altitude gets higher and higher.
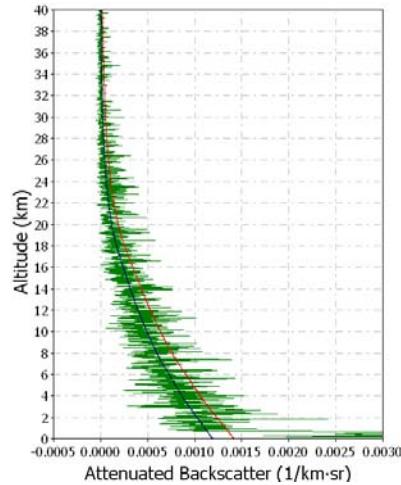
Figure 3.4: LITE Data with Threshold = β_{air}(z) + 3·RBV(z)

C. To combine contributions from MBV and RBV, select constants T_0 and T_1 and compute a threshold array as follows

$$\beta'_{Threshold}(z) = \beta'_{air}(z) + T_0 \cdot MBV + T_1 \cdot RBV(z). \tag{3.8}$$

T_0 and T_1 are programmable parameters fed to the algorithm via external configuration files. The initial values used in the CALIPSO mission will be selected based on retrievals using simulated data.

Figure 3.5 shows the results calculated using equation (3.8) with T_0=1.5 and T_1=2.5. The height of the threshold line is now much more uniform with respect to the extent of the noise excursions found at all altitudes throughout the measured data.
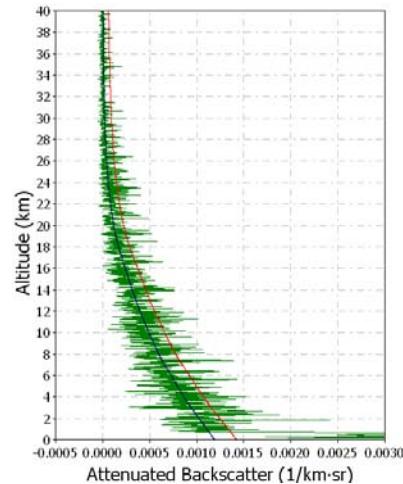
Figure 3.5: LITE Data with Threshold = $\beta_{air}(z) + 1.5 \cdot MBV + 2.5 \cdot RBV(z)$

The relative dominance of MBV or RBV within a given threshold array will depend on the lighting conditions. At night the background signal is very low, and so the width of the noise envelope about the mean attenuated backscatter signal will increase noticeably with range. Therefore, as we move lower in altitude the nighttime threshold value at any given altitude is increasingly influenced by the relative backscatter variation. During the day however, the noise envelope stays relatively constant, because the contributions from solar background light (i.e., MBV) overwhelm those inherent in the backscatter signal.

The adaptive, auto-scaling nature of the threshold scheme is demonstrated in Figure 3.6 using three profiles acquired during LITE. From left to right, the profiles shown are (a) a nighttime measurement of a thick aerosol layer, (b) a daytime measurement of "clear air", and (c) a daytime measurement of a bright stratus cloud. For all three cases, the threshold values (which are shown in red) were built using T_0=1.5 and T_1=2.5. Note that relative to the noise excursions all three thresholds fall in approximately the same place – that is, the noise peaks rise above the thresholds (as do the features in A and C), but the bulk of the noise and all of the clear air signal is confined below. Note too that the noise envelopes vary by almost two orders of magnitude from the nighttime measurement to the measurement of bright clouds in the daytime.
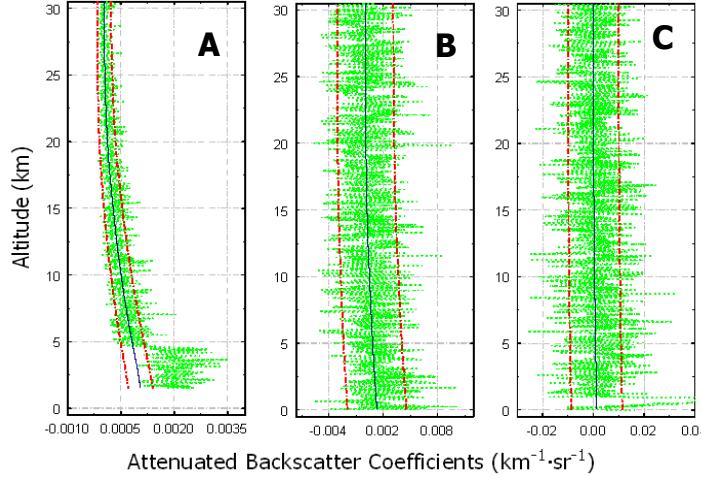
Figure 3.6: Thresholding technique for different lighting conditions; measured attenuated back-scatter coefficients are shown in green, the model attenuated backscatter coefficients are shown in blue, and the threshold lines are shown in red.

3.2.3.   Threshold Array Adaptations for CALIPSO

Thus far the threshold examples shown have used LITE data. Applying the threshold scheme to CALIPSO data requires some additional considerations. Unlike LITE, the CALIPSO data have been averaged both horizontally and vertically on-board the satellite, and as a consequence the threshold values given in equation (3.8) cannot be applied directly to an entire high-resolution profile. Corrections are required for regions having different averaging schemes. These corrections are given below in Table 3.1 for MBV computed using the signal in the region between 30-km and 40-km (6th column) and for MBV computed using the single-shot on-board measurements made in the baseline region (7th column).

Table 3.1: Full Resolution Threshold Corrections for Each Averaging Region

Altitude Region		Horizontal Resolution (km)	Vertical Resolution (meters)	30 m Points Per Resolution Element	Threshold Correction (30-40 km)	Threshold Correction (baseline)
Base	Top					
-2.0	-0.5	1/3	300	10	$\sqrt{15}$	$4\cdot\sqrt{5}$
-0.5	8.2	1/3	30	1	$5\cdot\sqrt{6}$	$\sqrt{2}$
8.2	20.2	1	60	6	5	$2\cdot\sqrt{3}$
20.2	30.1	5/3	180	30	$\sqrt{5}$	$4\cdot\sqrt{15}$
30.1	40.0	5	300	150	1	20

The necessity for these corrections and the effect of applying them can clearly be seen in Figure 3.7. Note in particular how the noise envelope of the full resolution backscatter signal expands abruptly at the 20-km on-board averaging boundary. Note too the corresponding step function increase in the threshold array. Similar threshold discontinuities appear at the 8-km and (less noticeably) 30-km on-board averaging boundaries.
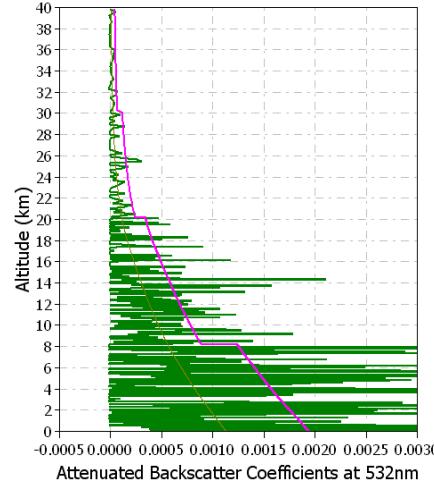
Figure 3.7: Simulated full resolution CALIPSO profile showing a measurement of clear air

3.2.4. Attenuated Scattering Ratio Threshold Arrays

For application in the profile scanner the backscatter coefficient threshold arrays are converted to attenuated scattering ratio threshold arrays. The threshold equation now becomes

$$R'_{Threshold}(z) = 1 + \frac{T_0 \cdot MBV(z) + T_1 \cdot RBV(z)}{\beta'_{air}(z)} .$$
(3.9)

Due to the corrections required to account for the varying on-board averaging resolutions, MBV is now expressed as a function of altitude.

A side-by-side comparison of the two threshold techniques is shown in Figure 3.8. Note that in both cases the threshold tracks the outline of the signal's noise envelope.
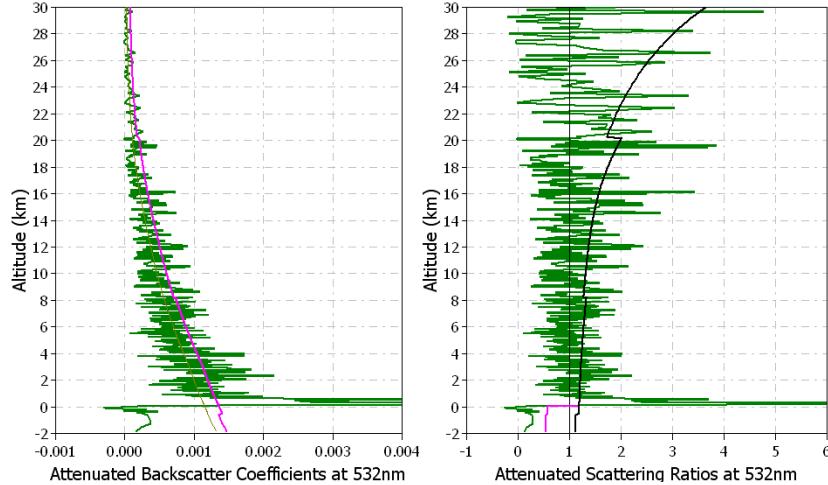
Figure 3.8: Threshold Arrays for Attenuated Backscatter Coefficients and Attenuated Scattering Ratios

3.2.5.   Selecting the Profile Scanning Search Constants

A family of dynamically selected coefficients governs the progress of the profile-scanning engine. These coefficients cannot be considered as constants, because (a) their value may vary from profile to profile, and (b) their value may change even within the scan of a single profile. Taken together, these coefficients form a set of programmable parameters that are supplied to the algorithm via external configuration files. As with the values for T_0 and T_1, the initial values for the profile scanner search parameters used in the CALIPSO mission will be selected based on retrievals using simulated data.

3.2.5.1.   Feature Thickness, Spike Thickness, and Spike Threshold Factor

The search for features in lidar backscatter data is complicated by the fact that Nature creates clouds and aerosol layers in an infinite array of sizes and shapes. Figure 3.9 shows a profile from LITE Orbit 35 that simultaneously depicts two extreme cases. The very thin "spike" just above 9 km is an strongly scattering alto-cumulus cloud. The broad, flat feature extending from 5.6 km down to the surface is a dust layer originating in the Western Sahara desert. We can parameterize the shape of features using an aspect ratio statistic:

$$\text{Aspect Ratio} = \frac{\text{Peak Attenuated Backscatter Coefficient}}{\text{Feature Geometric Thickness}}. \tag{3.10}$$

The aspect ratio of the dust layer in Figure 3.9 is ~0.0005. The aspect ratio of the alto-cumulus cloud is somewhere in excess of 0.032. (Because the LITE high-gain data saturates in strong features, we cannot measure the full magnitude of the backscatter peak.) A robust feature-finder must be able to reliably detect the boundaries of features at both extremes of the aspect ratio scale. To accomplish this we employ a trio of coefficients known as the *minimum feature thickness*, the *minimum spike thickness*, and the *spike threshold factor*.
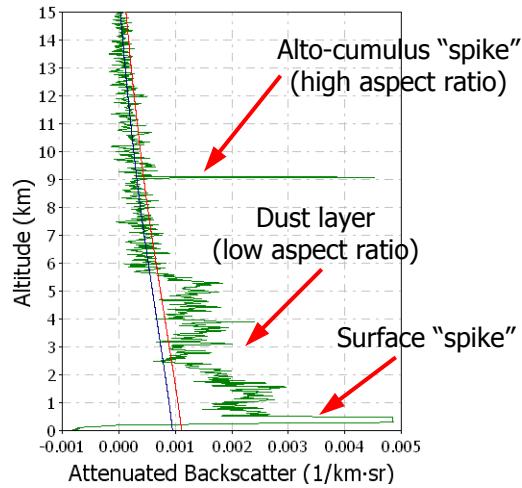
Figure 3.9: The Extremes of Feature Aspect Ratio

The minimum feature thickness is a measure of vertical distance, having units of length (kilo-meters). It is used in concert with the threshold array to identify most ordinary features. This is perhaps the most straightforward application of the general threshold technique. To qualify as a feature, the scattering ratios in a candidate region must exceed the threshold value for all data points within a vertical extent greater than or equal to the specified minimum feature thickness.

For the LITE data, a constant value of 150 to 200 meters was generally effective in correctly locating small features while simultaneously rejecting extended noise excursions. However, (a) feature finding for LITE was restricted to the troposphere, (b) the LITE data has a uniform vertical resolution of 15 meters, and (c) the LITE SNR was substantially higher than the SNR we can expect from CALIPSO. Because the CALIPSO data comes to us "pre-averaged", the CALIPSO profile scanner specifies a minimum feature thickness for each of the five on-board averaging regions. To select the optimum values we must take into account both the expected geometric depths of features within each region and the amount of vertical averaging done prior to the search.

Note that the most effective minimum feature thickness can be quite different for different threshold values. If the minimum feature thickness is made larger, the magnitude of the threshold constants T_0 and T_1 can frequently be reduced. Likewise, if the minimum feature thickness is made smaller, larger values of T_0 and T_1 are usually needed to filter out spurious features.

The high aspect ratio features referred to as "spikes" represent the one general class of features that are not detected using the minimum feature thickness parameter. As shown in Figure 3.9, a spike is a thin, dense cloud, one for which the measured vertical extent is less than the minimum cloud thickness, but which also contains at least one very large attenuated scattering ratio. To identify these features requires two additional parameters: a minimum spike thickness and a spike threshold factor. Like the minimum feature thickness, the minimum spike thickness is a vertical distance with units of length, such that $0 < \Delta Z_{spike} < \Delta Z_{feature}$. The spike threshold factor is a unitless floating-point number, always greater than one, that amplifies the threshold array within the boundaries of the prospective spike. For a region to qualify as a feature by virtue of being a spike,

- the attenuated scattering ratios within that region must all rise above the local threshold level for a vertical distance equal to the minimum spike thickness; AND

- at least one of the attenuated scattering ratios therein must also exceed the product of the threshold value and the spike threshold factor. Useful values for the spike threshold factor are generally in the region of 8 ± 2.

The spike identification process is illustrated in Figure 3.10 using the data shown earlier in Figure 3.9. The standard feature finder threshold is shown in bright red; the spike threshold, which is the product of the spike threshold factor and the standard threshold, is shown in dark red. In this example the minimum feature thickness is set to 180 meters, the minimum spike thickness is set to 90 meters, and the spike threshold factor is set to 5. The 9-km signal excursion does not have a vertical extent of 180 meters or more, and hence would initially be rejected by the scanner algorithm. However, it does exceed the standard threshold for the requisite minimum of 90 meters and at least one of the attenuated scattering ratios within that range exceeds the spike threshold, and thus it is correctly identified as a feature.
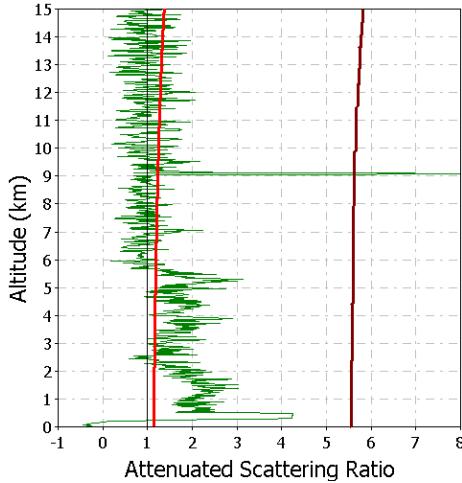
Figure 3.10: Detection of "spikes" in the attenuated scattering ratio regime; attenuated scattering ratios are shown in green, the standard feature finder threshold is shown in bright red, and the spike threshold is shown in dark red.

3.2.5.2.   Minimum Clear Air Distance

To establish a final feature base altitude, the profile scanner requires the specification of a minimum clear air distance. This parameter specifies the minimum "feature-free" vertical distance that must exist beneath each potential feature before a fixed base altitude can be assigned. In theory feature base is determined at that point where (a) the mean attenuated scattering ratio is less than or equal to the mean attenuated scattering ratio above the feature, and (b) the slope of the attenuated scattering ratios is zero. The measurements of mean and slope are both made over the specified minimum clear air distance. Practical methods for determining feature base altitude are discussed in detail beginning in section 3.2.7.

3.2.5.3.   Closing Gaps Between Features

When analyzing inhomogeneous features with highly variable backscatter coefficients (e.g., cirrus clouds), a "plain vanilla" threshold method will frequently find numerous separate, narrow, closely spaced peaks in a region where the human eye (guided by intuition and a global view of the data) discerns only a single cloud layer. Therefore, the ability to "close gaps" between similar features has been incorporated into the CALIPSO feature finder; that is, two features that are considered to be too close together are merged, so that the resulting single feature has the base altitude of the lower feature, and the top altitude of the upper feature.

At a minimum, the algorithm for closing gaps requires specification of some maximum distance over which gaps are to be closed. Note too that to have any effect, this distance must exceed the minimum clear air distance. More sophisticated gap closing algorithms also employ a feature similarity score. This value is computed from the spatial and optical properties of the two features under consideration, and is used to prevent the merging of dissimilar features (e.g., a

cirrus cloud and an elevated aerosol layer).[4] At present, the CALIPSO gap-closing algorithm does not incorporate similarity scores, and instead requires only the specification of a maximum gap-closing distance. Should the distance between the base altitude of the upper feature and the top altitude of the lower feature be less than the specified gap-closing distance, the two features are merge, and thereafter treated as a single feature.

3.2.5.4.   Search Start and Stop Altitudes

By default the CALIPSO profile scanner searches an altitude range from a maximum of 30-km[5] down to a minimum of −1.5 km. However, the search start and stop altitudes are programmable parameters. As will become clear later, a profile scan can be initiated over any subsection of an attenuated scattering ratio profile.

3.2.6.   Searching for Feature Top

Once we have specified a search start altitude, obtained a profile of attenuated scattering ratios, and computed an array of threshold values (see Sections 3.2.2 through 3.2.4), the first task of the profile scanner is to search for the boundary between "clear air" and the top of the next feature (if any). The search is conducted by scanning down through the profile, beginning at some altitude $Z = Z_{max}$ and ending at $Z = Z_{min}$. A feature top is located at the first point such that either

   (a) the attenuated scattering ratio exceeds the threshold value for the number of consecutive points required to span the minimum feature distance, or

   (b) the attenuated scattering ratio exceeds the threshold value for the number of consecutive points required to span the minimum spike distance AND at for at least one of those points the attenuated scattering ratio exceeds the threshold value by a multiplicative factor greater than the spike threshold factor.

By definition, the attenuated scattering ratio at the highest cloud top will be greater than 1.00.

The flowchart in Figure 3.11 gives a pictorial representation of the steps necessary to locate the initial feature boundary.

---

[4] An appropriate similarity score might be the inverse of the Euclidean distance between features in some (TBD) feature vector subspace, so that similar features (i.e., those that are "nearby" in the selected subspace) would have a high similarity score. This is a topic for future research.

[5] In practice the profile scan is initiated **no higher** than the base of the 532 parallel channel calibration region.
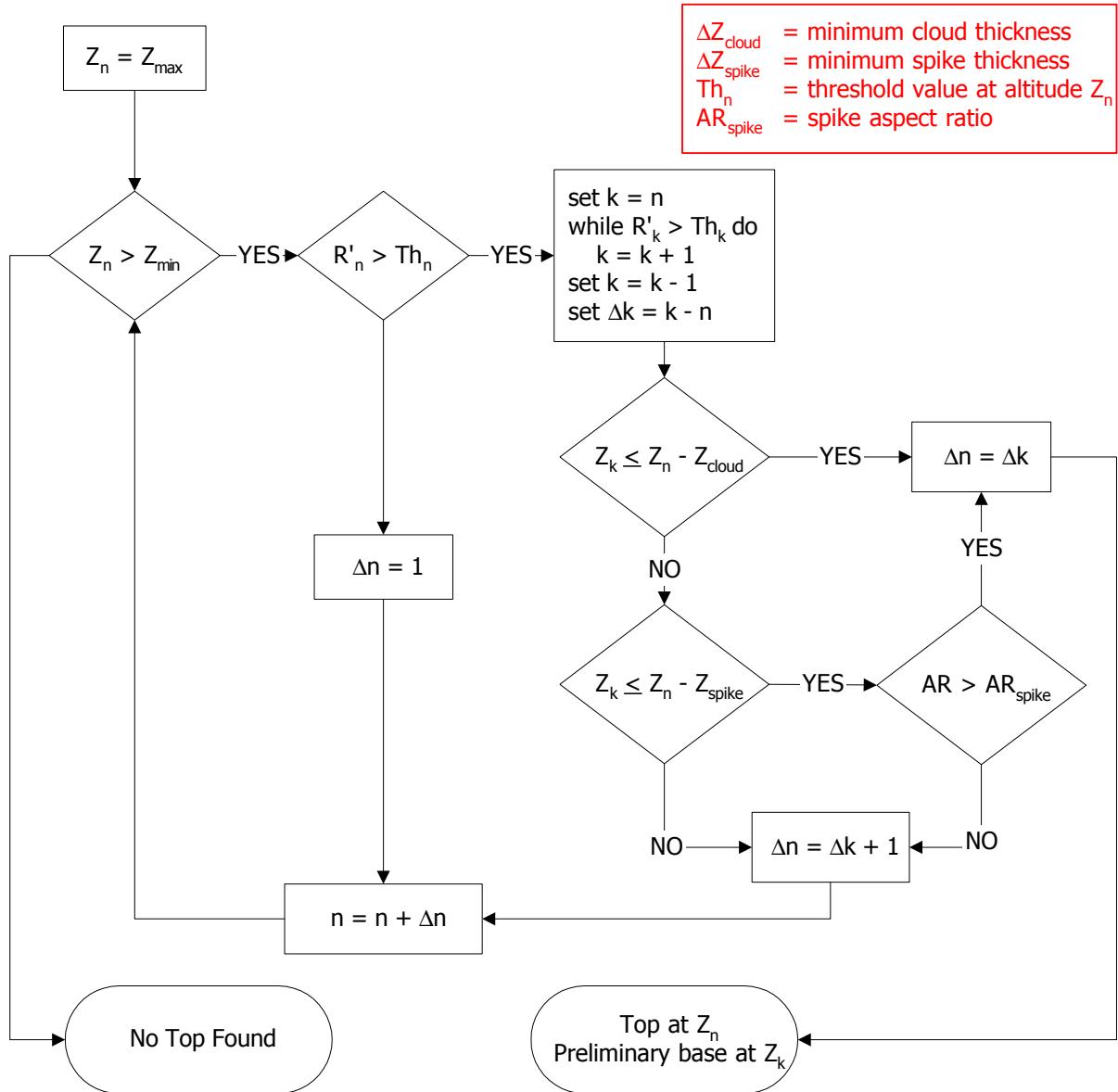
Figure 3.11: Locating Feature Top[6]

3.2.7. Estimating Feature Base

The initial estimate of cloud base occurs at that point where the value of the attenuated scattering ratio first drops below the threshold. However, because the signal is always contaminated with some amount of random noise, this first point may not correspond to the true base. The measured value at any range bin can be either higher (positive noise) or lower (negative noise) than the expected (i.e., "true") value. Negative noise excursions can result in in-cloud data values that fail to exceed the threshold value, thereby leading to a premature determination of

---

[6] In this example, the altitude array is structured so that the maximum altitude is at Z[0], and the minimum altitude is at Z[N]

base altitude. To combat this sort of error, we use a "probabilistic base locator". This algorithm chooses the first point below the threshold as an initial estimate of cloud base. It then "looks ahead" by examining an altitude range extending below the current estimate of cloud base for a depth equal to the minimum clear air distance (see section 3.2.5.2). The estimate of cloud base altitude is continually advanced (i.e., moved lower in altitude), so long as some pre-selected percentage of the points within the examination range remains above the threshold value. A semi-final estimate of cloud base is made at that altitude where the number of points above the threshold is less than the required percentage. In practice, effective values of this "look-ahead percentage" range between 50% and 70%.

3.2.8. Refining Estimates of Base & Top

A second and more complicated factor in base altitude determination is attenuation of the signal by the cloud. This is perhaps best explained by example. Suppose we have established a constant threshold value of 1.5 – that is, contiguous attenuated scattering ratios of 1.5 or greater are indicative of a cloud. Suppose now that we are asked to evaluate a data point within a cirrus cloud having a true scattering ratio, $1 + \beta_c / \beta_m$, of 2.5. If the optical depth of the cloud to this point is 0.5, then the attenuated scattering ratio will be $R' = 2.5 \cdot \exp(-2 \cdot \tau) = 0.92$. Clearly this fails our threshold test (i.e., 0.92 < 1.50). But, just as clearly, if we could compensate properly for the attenuation incurred within the cloud (or if this same ratio were encountered at the top of the cloud), this point would pass the threshold test.

To overcome the effects of in-cloud attenuation, we once again employ a "look ahead" strategy. Near the base of a cloud, the magnitude of the attenuated scattering ratios should fall off for some distance; that is, the slope of the attenuated scattering ratios with respect to range should be negative. At cloud base the magnitude should stabilize to some constant value and the slope should go to zero. We can refine our estimate of cloud base by examining the slope in the look-ahead region. As long as the slope remains negative, we revise our estimate of the base height to a new, lower altitude, and then repeat the process. When the slope goes to zero, we have located cloud base. This situation is illustrated by the attenuated scattering ratio profile shown in Figure 3.12.
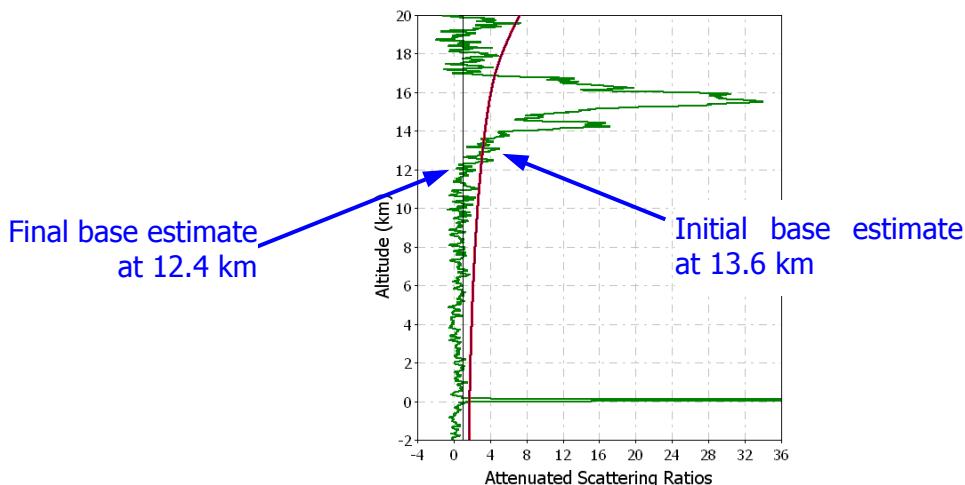
Figure 3.12: LITE Data from Orbit 24 (18.00° N, 107.86° E)

The complications addressed in Sections 3.2.7 and 3.2.8 operate entirely independently of one another, and so any "base locator" algorithm must be prepared to deal with both simultaneously.

3.2.9.   Updating the Threshold Array

In constructing our initial threshold array we account for signal attenuation due to molecular scattering and ozone absorption. However, below each feature this initial threshold array must be rescaled to account for any additional attenuation due to the propagation through the feature. Failure to do so will result in the profile scanner being unable to identify weaker but still prominent features that lie beneath moderately to highly attenuating features. The graphic on the left in Figure 3.13 clearly shows the unacceptable consequences that can arise when the threshold array is not properly rescaled. Note that the aerosol layer between 1.0 km and the surface falls entirely beneath the original threshold. The right-hand panel in Figure 3.13 demonstrates the results obtained when the initial threshold is properly rescaled to account for the attenuation of the overlying cirrus layer.
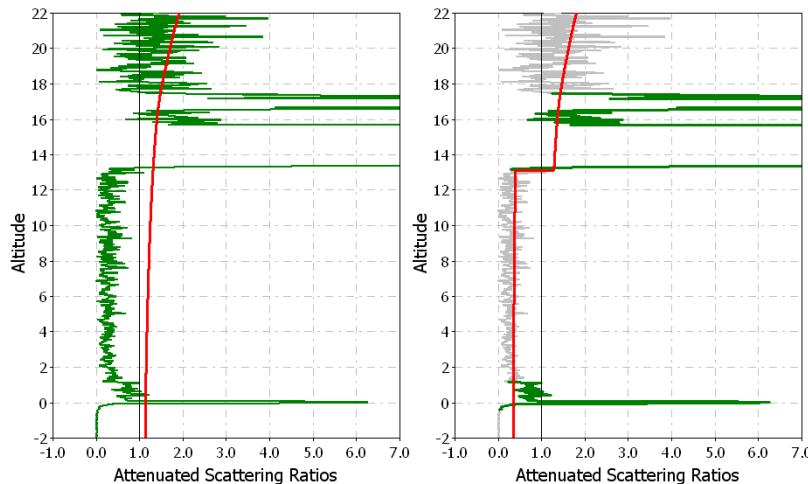
Figure 3.13: Threshold Update

The initial threshold array is computed assuming a particulate transmittance of $T_p^2 = 1$ at the base of the calibration region.[7] However, this assumption is rendered invalid immediately upon identifying the first feature, as T_p^2 will now be something less than one. To repair our threshold array for use on subsequent, lower altitude portions of the data, we need an estimate of the feature two-way transmittance. In general this estimate is easily computed directly from the attenuated scattering ratio data. Referring to equation (3.3) it is clear that the mean attenuated scattering ratio in the "clear air" region beneath a feature provides a direct measurement of the feature two-way transmittance. Because we have used a "look ahead" region to refine our estimate of feature base, we have also established a known-to-be-clear region over which we can compute the average ratio. As a consequence, once the profile scanner has made a final determination of feature base the next step is to estimate the feature two-way transmittance. We

---

[7] Properly speaking, any signal attenuation between the lidar and the calibration altitude is absorbed into the calibration constant, so that our 30-km calibration constant is actually $C_{30km} = C_{lidar} \cdot T_p^2 (30\ km)$

do this by computing the average value, $\overline{R}'$, of the attenuated scattering ratios over the minimum clear air region (see section 0) beneath our new feature. In this region we assume that $\overline{R}' \approx T^2_{feature}$. If $\overline{R}' \geq 1$ or $\overline{R}' \leq 0$, no change is made to the profile scanner's threshold array. However, for $0 < \overline{R}' < 1.0$ the *threshold array* is multiplied by $\overline{R}'$ from feature base downwards to the end of the array. The profile scanning then resumes using this modified threshold array.

Due to the nature of the random noise in our signal and to occasional errant behavior by the detectors (e.g., overshoot), this last step can cause some unpleasant problems. In particular, if our estimate of T^2_p is too low the modified threshold values will likewise be too low, which can subsequently lead to the false detection of a large number of spurious features. To guard against the possibility of this sort of unwanted behavior we use the feature integrated attenuated backscatter to develop a simple consistency check.

3.2.9.1. Feature Integrated Attenuated Backscatter

The feature integrated attenuated backscatter is defined by

$$\gamma'_{feature} = \int_{top}^{base} \beta_p(r) \cdot T^2_p(r) \, dr . \qquad (3.11)$$

As we will see presently, $\gamma'_{feature}$ is an extraordinarily useful quantity that has application in several different aspects of feature finding. Unfortunately, we cannot retrieve accurate values of γ' until after we've completed the extinction retrieval. However, we can make reasonably good *estimates* of the feature integrated attenuated backscatter using the attenuated scattering ratios and the *clear air trapezoid* technique.

We can make an initial approximation of $\gamma'_{feature}$ by first correcting the total attenuated backscatter coefficients for the molecular and ozone transmittances, then integrating between top and base, as in equation (3.12).

$$g = \int_{top}^{base} \frac{\beta'(r)}{T^2_m(r) \cdot T^2_{O3}(r)} \, dr = \int_{top}^{base} \beta_p(r) \cdot T^2_p(r) \, dr + \int_{top}^{base} \beta_m(r) \cdot T^2_p(r) \, dr . \qquad (3.12)$$

Alternately, the same expression can be derived using the product of the attenuated scattering ratios and the molecular backscatter coefficients, as in equation (3.13).

$$g = \int_{top}^{base} R'(r) \cdot \beta_m(r) \, dr = \int_{top}^{base} \beta_p(r) \cdot T^2_p(r) \, dr + \int_{top}^{base} \beta_m(r) \cdot T^2_p(r) \, dr . \qquad (3.13)$$

In either case, the estimate is too large due to the contributions from the feature-attenuated molecular backscatter term. To reduce the error in the estimate, we can approximate the integral on the far right side using the appropriate clear air trapezoid area. The general concept of the clear air trapezoid correction is illustrated in Figure 3.14. This figure shows a multi-featured profile of LITE attenuated scattering ratios acquired during orbit 129. The measured data is shown in green within the features and in pale gray elsewhere. The clear air trapezoids are shown as blue crosshatched regions.
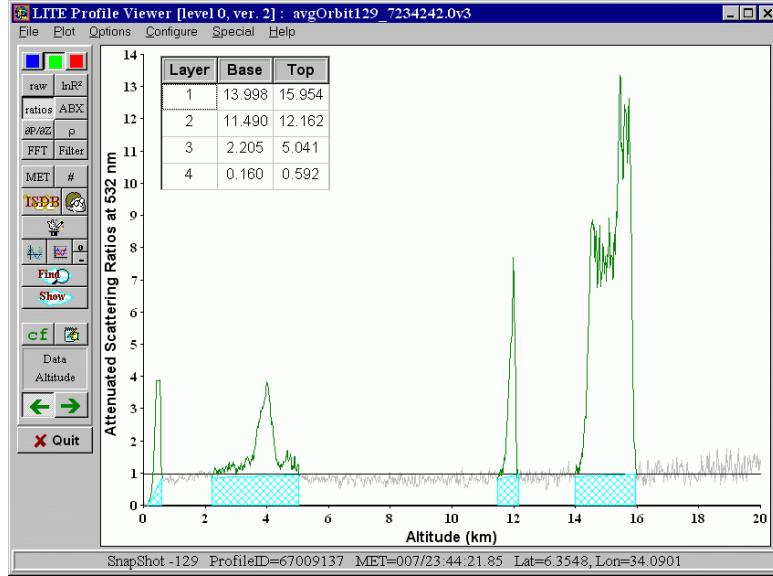
Figure 3.14: Clear Air Trapezoids

For each feature we can compute g as indicated in either equation (3.12) or equation (3.13). This calculation will return the total area under the signal curve between the specified boundaries (i.e., feature base and top). To estimate the fraction of the total area contributed by molecular scattering we construct a trapezoid representing our best guess at the area of "non-feature" contributions to the signal. The base of each trapezoid extends along the $R'(z) = 0$ line. The legs of the trapezoid are located at the feature base and top altitudes and extend up perpendicular from the zero line to the signal variable values at the base and top altitudes (e.g., $\mathcal{R}_{base}$ and $\mathcal{R}_{top}$, as in equation (3.14)).

Within the profile scanner, an operational estimate of $\gamma'_{feature}$ is constructed as follows:

$$\mathcal{R}_k = \beta_m(z_k) \cdot R'(z_k)$$

$$g = \frac{1}{2} \cdot \sum_{k=top+1}^{base} (z_{k-1} - z_k) \cdot (\mathcal{R}_{k-1} + \mathcal{R}_k) \qquad .$$

$$\gamma'_{feature} = g - \left(\frac{1}{2} \cdot (z_{top} - z_{base}) \cdot (\mathcal{R}_{top} + \mathcal{R}_{base}) \right)$$

(3.14)

3.2.9.2.   A Constrained Threshold Update Scheme

The equation relating $\gamma'_{feature}$ to feature optical depth (τ) and lidar ratio (S, a.k.a. extinction-to-backscatter ratio) is given by

$$\gamma'_{feature} = \left(\frac{1}{2 \cdot S} \right) \cdot (1 - \exp(-2 \cdot \tau)) = \left(\frac{1}{2 \cdot S} \right) \cdot (1 - T^2).$$

(3.15)

Development of (3.15) is given in two seminal papers dating from the early 1970s (Fernald et al., 1972 and Platt, 1973). During a profile scan, we can use equation (3.14) to compute $\gamma'_{feature}$ as

soon as we have located feature base. Using a *reasonable choice for S* we estimate a lower bound for $\langle R' \rangle$ using a rearranged version of equation (3.15):

$$T^2 \approx 1 - 2 \cdot \gamma'_{feature} \cdot S_{reasonable} . \qquad (3.16)$$

Note that generating a reasonable lower bound for $\langle R' \rangle$ requires that we be able to select an appropriate upper bound for $S_{reasonable}$.

When we begin a profile scan we assume that the two-way attenuation of the signal due to extinction by particulates is 1.00 ($\tau_p = 0 \Rightarrow T_p^2 = 1.00$). Within the profile scanning algorithm we maintain a current estimate of the particulate transmittance, $\tilde{T}_p^2$, that is updated after exiting each cloud layer. For a complete profile scan, the update scheme is as follows:

Procedure 2: Threshold Update

A.   Initialize $\tilde{T}_p^2$ to 1.00.

Repeat the following steps until the entire profile has been scanned:

B.   Locate feature base and top altitudes and compute $\gamma'_{feature}$

C.   Compute $\langle R' \rangle$ by averaging the attenuated scattering ratios between Z_0=base and Z_1=base+MinimumClearAirDistance.

D.   If $\langle R' \rangle$ is greater than $\tilde{T}_p^2$ or less than zero leave $\tilde{T}_p^2$ unchanged. No threshold update is required.

E.   If $\langle R' \rangle$ is greater than zero and less than $\tilde{T}_p^2$, select $S_{reasonable}$ and compute $\tilde{R}$ using equation (3.16); i.e., $\tilde{R} = 1 - 2 \cdot \gamma'_{feature} \cdot S_{reasonable}$

F.   Set $\tilde{T}_p^2$ to the LARGER of $\langle R' \rangle$ and $\tilde{R}$ and multiply the remainder of the threshold array by the revised value of $\tilde{T}_p^2$. The remainder of the threshold array is that part from Z_0=CloudBase downward to the end of the search region.

The strategy outlined above takes a conservative approach to feature finding. By restricting the maximum attenuation ascribed to any feature, we avoid the identification errors associated with an unreasonably low threshold array. However, weak features lying beneath very strong features with unusually large lidar ratios may be missed because the threshold is not allowed to drop sufficiently low. Strategies for retrieving these weaker features are considered in detail later in section 3.4.

3.2.9.3.   Strategies for Selecting $S_{reasonable}$

The range of naturally occurring lidar ratios within clouds is relatively small. Recent research reports values of $S_c^{-1} \approx 0.041 \pm 0.026$ for cirrus (i.e., $S_c \approx 24$; see Sassen & Comstock, 2001 and/or Eloranta, Kuehn, & Holz, 2000), and $S_c \approx 18$ for water clouds (Pinnick et al., 1983). For either ice clouds or water clouds, an appropriate upper bound for $S_{reasonable}$ lies in the range between 30 sr and 50 sr. The range of measured lidar ratios for aerosols is much wider (from

about 10 to 100; see Anderson et. al. 2000), and the mean $S_{aerosol}$ can be much larger ($\approx 66 \pm 16$; see Table 4.10 in Masonis, 2001). Clearly the upper bound for aerosol lidar ratio should be higher than that for clouds – say perhaps 80 sr. Furthermore, because multiple scattering cannot be neglected in the analysis of space-based lidar data, $S_{reasonable}$ must be considered as an effective lidar ratio; that is $S^* = \bar{\eta} S$, where $\bar{\eta}$ is the layer averaged multiple scattering factor such that $0 < \bar{\eta} \leq 1$.[8] The actual values of $S_{reasonable}$ are therefore somewhat lower than the corresponding cloud or aerosol lidar ratio limits.

Since the threshold constraint is applied to prevent overestimates of feature attenuation, one possible approach is to select $S_{reasonable}$ based on a coarse estimate of feature type. For example, suppose $\gamma'_{feature} > 0.0015$ and $\chi'_{feature} < 0.6$, indicative of a robust layer of small particles. We could therefore assume that all such layers are aerosols with $\bar{\eta} \approx 0.85$, and then set $S_{reasonable} = 70$ sr. In all other cases, we would assume the feature is a cloud $(\bar{\eta} \approx 0.75)$, and set $S_{reasonable}$ to 30 sr. (Note that if a feature is an aerosol of weak to moderate concentration, the actual attenuation will likely be small, and so should not be greatly different than the constrained attenuation we estimate using $S_{reasonable} = 30$.)

More sophisticated approaches are also possible. Several on-going investigations are examining the relationship between lidar ratio and the mid-cloud temperature. The current best estimate of cirrus cloud lidar ratios as a function of temperature, T, is given by

$$S_{cirrus}(T) = -1.259\,T - 6.698 \text{ for } -70° < T < -20°. \tag{3.17}$$

where T is given in degrees centigrade, and includes sign (Platt et al., 2002). However, employing this additional information in the selection of $S_{reasonable}$ requires not only a successful discrimination between clouds and aerosols, but also the ability to accurately assess cloud phase. These tasks are well beyond the scope of the feature-finding algorithm, and so until such time as a feature classification module can be incorporated, $S_{reasonable}$ is implemented as a single user-defined constant obtained from a configuration file at runtime.

3.2.9.4. Threshold Updates and the Minimum Clear Air Distance

Proper revision of the threshold level is probably the most challenging problem involved in adapting the threshold method for use with attenuated scattering ratios. Care must be taken to avoid computing the new value in regions where the signal is contaminated with aerosol, or where the profile is distorted due to (for example) an inadequate detector response time. In this respect, statistical concerns alone suggest that it is always better to use more data points rather than less. However, because we have a fixed altitude resolution, using additional data points can mean extending the minimum clear air distance to a perhaps unacceptable length.

3.2.10. Profile Scanner Flow Chart

A flowchart of the profile scanning engine discussed in this section is given below in Figure 3.15.

---

[8] A complete analysis of multiple scattering effects can be found in the CALIPSO Extinction ATBD (PC-SCI-203 Part 5).
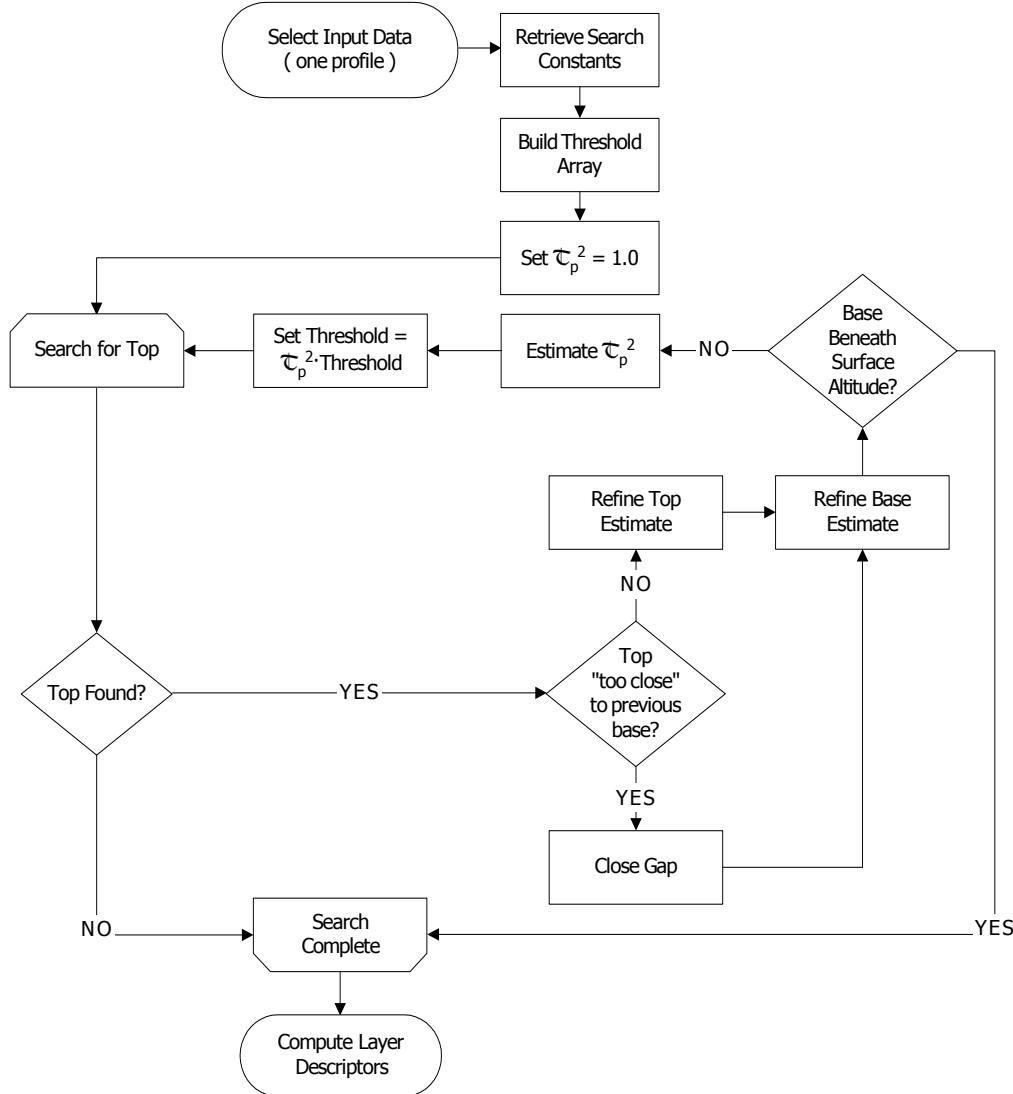
Figure 3.15: Profile Scanning Flowchart

Additional detail will be added to this chart in later sections of this document.

3.3. Assumptions

The feature-finder relies on the structure of the attenuated scattering ratio profile to extract base and top altitudes of cloud and aerosol layers. We assume that background and baseline subtractions have been done correctly, and that the density model we use is a suitable representation of the true molecular density in the measurement area, and that the system constant estimate is acceptably close to the true value.

Our feature-finder algorithms are built on an expectation of how the Earth and its atmosphere should appear to our lidar. That expectation takes the form of "clear air" scattering models built from meteorological data obtained from the GMAO via the CALIPSO Met Manager. Similarly, the height of the surface above mean sea level is assumed to be a known input obtained from a

high quality digital elevation map (DEM) such as the GTOPO30. Where these model inputs are in error our retrieval schemes can be expected to yield less than perfect results.

3.4. Internal and External Outputs

The primary output of the profile scanner is a set of *layer descriptors*. Layer descriptors describe what might be called the bulk properties of a feature; that is, they provide information about the geometric and optical characteristics ascribed to an entire feature. For example, perhaps the most fundamental of all spatial layer descriptors are the feature top and base altitudes. Likewise, optical depth and integrated attenuated backscatter are important optical layer descriptors. We record a comprehensive set of layer descriptors for each feature identified during a profile scan. In addition to the integrated properties traditionally used in lidar data analysis, we also compute a standard suite of descriptive statistics using several different realizations of the profile data within a feature. These statistics include the minimum and maximum values within the feature, the sample mean value and standard deviation computed between feature top and feature base, and, to further characterize the shape of the profile data, the centroid and a dimensionless "*skewness*" parameter (Brandt, 1999).

The following layer descriptors are reported for each feature:

- base and top altitudes

- integrated attenuated backscatter (γ') at 532 nm and 1064 nm (for the 532 channel we report the integrated *total* attenuated backscatter coefficient, as described in equation (3.14))

- attenuated backscatter statistics (i.e., min, max, mean, standard deviation, centroid, and skewness of the attenuated total backscatter coefficients at each wavelength)

- layer-integrated 532 nm volume depolarization ratio (δ_{layer})

- 532 nm volume depolarization statistics

- layer-integrated attenuated total color ratio (χ'_{layer}) for each feature

- attenuated total color ratio statistics

- for all transmissive layers (in general, all layers above the bottom one), an estimate of the two-way transmittance at 532 nm and its associated uncertainty

- feature aspect ratio at 532 nm and 1064 nm

- temperatures at the base, top, and mid-feature altitudes

Explicit formulas for calculating the layer descriptors listed above are given in section 6. Note too that the specification given here is not a complete list of all layer descriptors provided in the Lidar Level 2 data product. Additional layer descriptors are added by the scene classification algorithms (e.g., cloud-aerosol classification) and following the extinction retrievals (e.g., 1064 nm optical depth).

Each complete pass of the profile scanner produces an array of layer descriptors – i.e., one set for each feature found. In addition to this array of layer descriptors, the profile scanner outputs include all of the information necessary to locate the profile in time and space. Profile search parameters are included as well.

The following information is recorded for <u>each profile</u>:

- latitude and longitude (start and stop values)

- time (start and stop times)

- solar zenith and azimuth angles

- effective surface height

- threshold values and other "constants" used in the search routine (among other things, these parameters will provide information on the minimum scattering intensity that could be detected in any given region)

- number of features found

- an array of layer descriptors (one set for each located feature within the profile)

4. Physical Models II: Averaging to Retrieve Tenuous Features

The peak backscatter intensities of the features measured by space-borne lidar range over several orders of magnitude. Strongly scattering features such as stratus and fair weather cumulus are easily detected using a single laser pulse. For more tenuous features – e.g., thin cirrus clouds – the average of several laser pulses may be required to obtain the signal-to-noise ratio necessary to differentiate feature boundaries from the ambient scattering environment. The unambiguous detection of the very weakest features – faint aerosol layers and subvisible cirrus – may require averaging over a substantial number of pulses. To identify all of the features within a given scene at the maximum possible spatial resolution we employ a *SELECTIVE ITERATED BOUNDARY LOCATION* (SIBYL) scheme. The SIBYL algorithm makes multiple passes through a specified scene, constructing profiles of attenuated scattering ratios at a series of increasingly coarse spatial resolutions. Immediately after construction, each profile is scanned for the presence of clouds, aerosol layers, and/or surface returns using an enhanced version of the profile scanner described above. Using an *intensity clearing* technique, the backscatter data from those regions identified as containing a feature are removed from subsequent processing. As a consequence, features found at high spatial resolutions (i.e., with less averaging) will **not** be included in the profiles of attenuated scattering ratios scanned at coarser resolutions (i.e., more averaging).

The essential step required for SIBYL's intensity clearing scheme to be effective is the application of an attenuation correction immediately following the removal of features. For all transmissive features, the intensity clearing operation estimates the feature two-way transmittance and uses this value to renormalize the partial profile of attenuated scattering ratios that lies beneath the feature. This attenuation correction is fundamentally different from the one employed by the profile scanner. In the profile scanner the attenuation correction (i.e., the feature two-way transmittance) is measured over a very limited distance and applied to the threshold array. SIBYL's attenuation correction is determined only after the profile scanner has completed its work. As a consequence, SIBYL can optimize the estimation process and generate an attenuation estimate using (potentially) a much larger range of "clear air" data. This improved estimate is then applied directly to the backscatter data that will be used in the next iteration of the profile scanner. This renormalization step occurs for all but the lowest feature in

the region being analyzed: the lowest feature is by definition opaque.[9]   For opaque features occurring within the surface-attached aerosol layer (SAL; Wiegner, 2002), SIBYL also includes a cloud-clearing loop to separate aerosols from clouds and surface signals at the highest possible spatial resolution (see section 4.4).
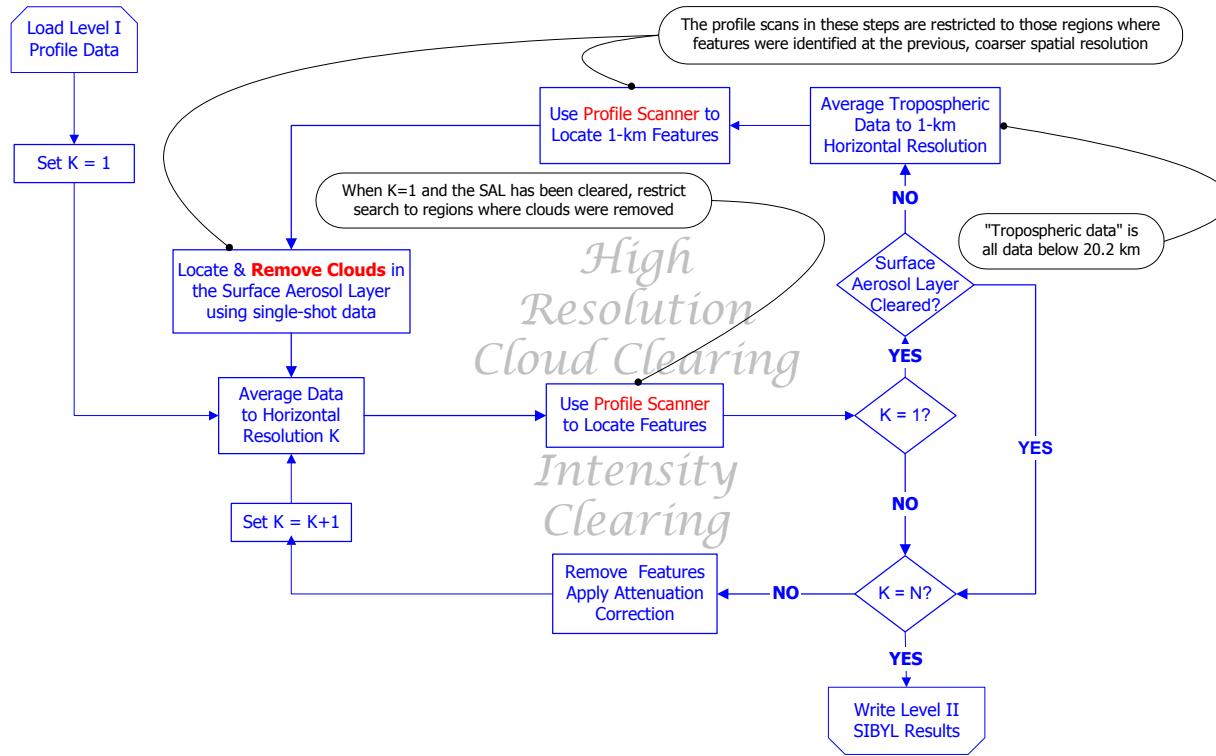
Figure 4.1: SIBYL flowchart; three horizontal resolutions are used (i.e., N=3), with an initial resolution (K=1) of 5-km, and subsequent resolutions of 20-km and 80-km.

Figure 4.1 shows the general flow of the SIBYL algorithm for an arbitrary number (N) of averaging levels. Note that the cloud-clearing loop executes once and once only, while the intensity-clearing loop occurs N-1 times. SIBYL currently uses three horizontal averaging intervals: a fundamental averaging distance of 5-km, an intermediate distance of 20-km, and a maximum horizontal averaging distance of 80-km. The 5-km interval was chosen to match the on-board data averaging scheme – the largest horizontal averaging distance in the data down-linked from the satellite is 5-km (see PC-SCI-201). Successive increments were selected based on the expected improvement in signal-to-noise ratio. Increasing our horizontal averaging distance by a factor of four each time should yield (on average) a factor of two increase in the profile SNR.

---

[9] Recall that a feature is ANY extended signal excursion rising significantly above the expected molecular backscatter level. By this definition, the surface return is clearly a feature…and a very opaque feature too!

4.1. Required Input Data

In one sense SIBYL is simply an averaging engine whose sole purpose is to build horizontally averaged lidar profiles to feed into the profile scanner. Consequently, SIBYL's input requirements are a superset of the profile scanner input requirements enumerated in section 3.1. The difference is that SIBYL requires the raw materials – e.g., profiles of attenuated backscatter coefficients for all three measurement channels – necessary to compile a complete set of profile scanner input parameters for any averaging distance up to and including the maximum horizontal averaging distance. SIBYL therefore ingests blocks of data that span an 80-km horizontal distance. SIBYL also requires access to the full range of data products delivered by the CALIPSO Met Manager.

4.2. Intensity Clearing

The intensity clearing technique is used to locate feature boundaries over a wide range of feature backscatter intensities. The initial step in the intensity clearing process is to identify strong features and then remove them from each profile. Before removal, the two-way transmittance of the feature is estimated. Immediately following removal, all data beneath the feature is corrected for the attenuation ascribed to the feature. Ideally, what remains following this procedure is a profile that represents the backscatter that we *would have measured had the feature not been present*. Adjacent intensity-cleared profiles can then be further averaged to increase the contrast between weaker features and the ambient molecular background. By applying this intensity clearing technique the data generated in subsequent averages is weighted equally with respect to feature attenuation throughout the entire profile altitude range.

The rationale for intensity clearing, its intermediate and final effects, and the procedures required to accomplish it are easier to explain and understand by using numerous pictures to augment a very few words. The graphs and tables that follow were all generated from the first 20-km horizontal segment of the simulated CALIPSO data pictured in Figure 4.2. In this simple scene we have a cirrus cloud overlying a highly absorbing aerosol layer. Both layers consist of a uniform distribution (with respect to altitude) of backscatter coefficients and a range-invariant lidar ratio. A summary of the "truth" parameters for the simulated features is shown in Table 4.1.
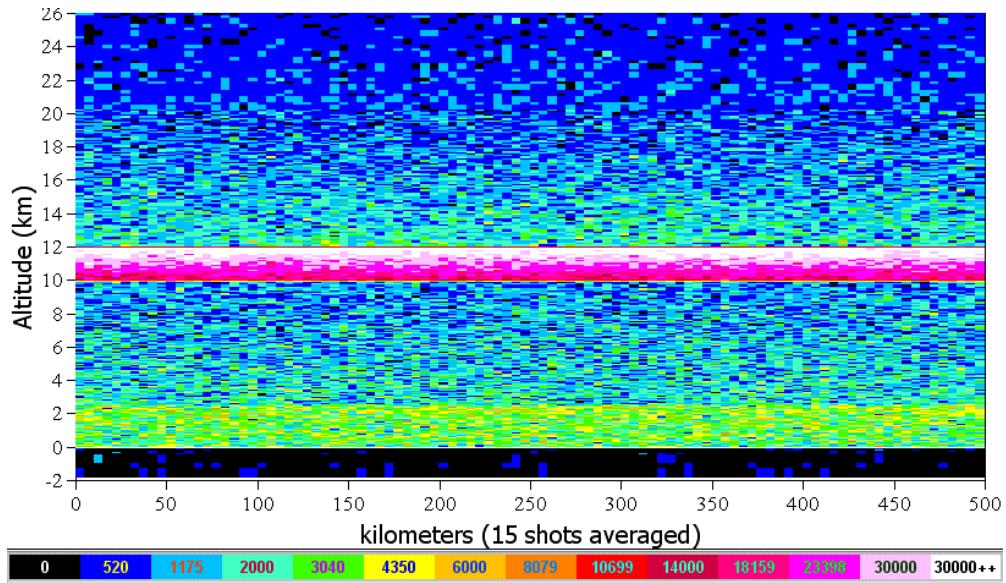
Figure 4.2: A Simple CALIPSO Simulation (raw backscatter signal)

Table 4.1: Simulation Truth Parameters

Parameter	Aerosol	Cloud
Base	0.0 km	10.0 km
Top	2.5 km	12.0 km
532 nm optical depth	0.20	0.50
532 nm backscatter coefficient	0.0013 km^{-1} sr^{-1}	0.008 km^{-1} sr^{-1}
532 nm lidar ratio	60.9 sr	25.0 sr

Figure 4.3 shows the first four attenuated scattering ratio profiles that will be used in the initial profile scans made by SIBYL. Each profile represents a 5-km horizontal average from the simulated data shown in Figure 4.2. Note that in profile view the cloud remains quite prominent, while the aerosol layer is essentially invisible. A close look at the bottom of each profile plot will also reveal the surface spike. Sensing the surface spike operationally confirms that the cloud layer is transmissive, and indicates to the profile scanner that after finding the feature boundaries it should also make an estimate of the two-way transmittance.
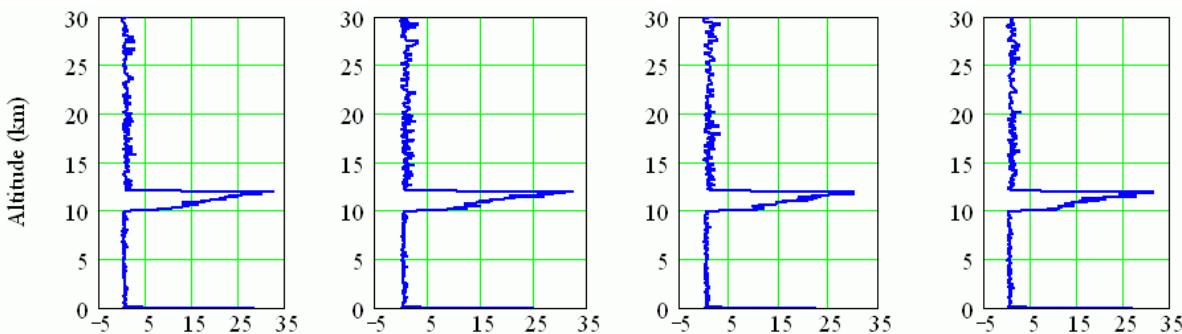
Figure 4.3: 5-km Attenuated Scattering Ratio Profiles

For the data shown in Figure 4.3 our SIBYL prototype returns the following cloud boundaries and optical depth estimates:

Table 4.2: Feature Finder Preliminary Results

Profile Number	Base	Top	Optical Depth
1 (left most)	9.915 km	12.015 km	0.487
2 (center left)	9.915 km	12.075 km	0.404
3 (center right)	9.915 km	12.015 km	0.516
4 (right most)	9.915 km	12.015 km	0.506

The next step of the process is to "remove" all features found as a result of the initial scan of the 5-km averaged data. Conceptually, the removal step amounts to replacing each feature with the appropriate values for "clear air". For the attenuated scattering ratios within the uppermost feature, this procedure is equivalent to replacing all values within the feature with a value of 1.00. (In the context of attenuated scattering ratios, 1.00 represents clear air.) A pictorial representation of the procedure is shown in Figure 4.4. Note in particular that the cloud peaks prominent in Figure 4.3 have now vanished, and have been replaced everywhere by "clear air" (i.e., $R'(z)=1.00$).
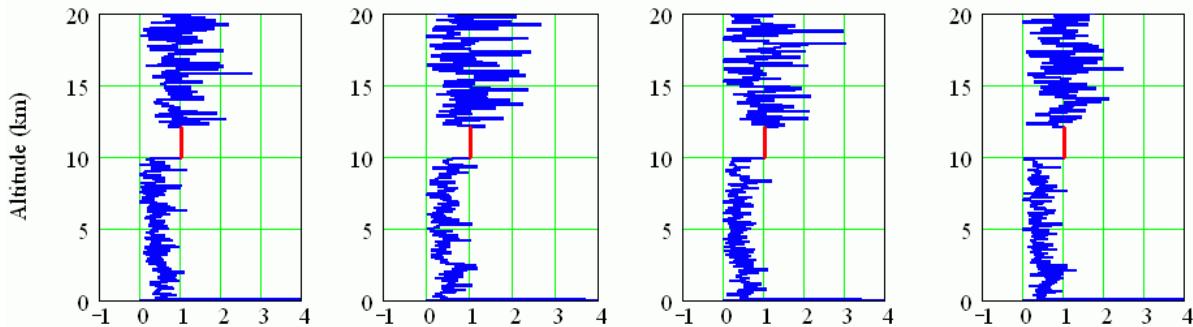
Figure 4.4: Feature removal at a 5-km horizontal resolution, with no attenuation correction applied

At this point we have accomplished the first part of the intensity-clearing goal: we have identified and removed all strong features in each profile. We now need to address part two, and apply an attenuation correction to the data beneath each removed feature so that our modified returns will appear as if the removed feature(s) had never been present. The effects of feature attenuation can be seen clearly in Figure 4.4 by looking at the step function structure of the attenuated scattering ratios between 5 and 15 km. Above feature base (10 km), the mean attenuated scattering ratio is approximately 1.00, indicative of clear air. However, below the base the mean attenuated scattering ratio is approximately 0.368 ($\exp(-2 \cdot \tau) = 0.368$ for $\tau = 0.50$). This discontinuity is more clearly evident in Figure 4.5, which shows the average of the four profiles displayed in Figure 4.4.
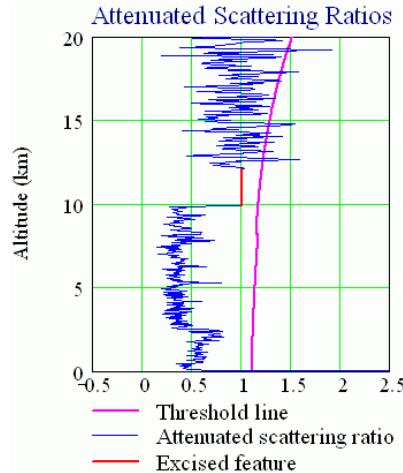
Figure 4.5: 20 km average, no attenuation correction applied

Figure 4.5 also shows the profile scanner threshold that would be computed for this particular profile using the technique described in section 3.2.4. Looking at the threshold applied to the uncorrected data, it is obvious that (a) the contrast between the aerosol layer and the ambient molecular background is such that the feature should be found, and (b) the attenuated scattering ratios within the aerosol layer all lie far below our current search threshold.

Fortunately, applying the necessary attenuation correction is fairly straightforward in concept (though in practice it can on occasion prove to be quite tricky to execute). We simply measure the mean attenuated scattering ratio in the clear air region beneath the feature, and use this result to "renormalize" all ratios beneath the feature.

The prototype version of the feature finder computes mean attenuated scattering ratios beneath each of the four 5-km profiles to be 0.378, 0.446, 0.356, and 0.353 respectively. After dividing each profile through by the appropriate correction factor, the profiles shown in Figure 4.4 look like this:
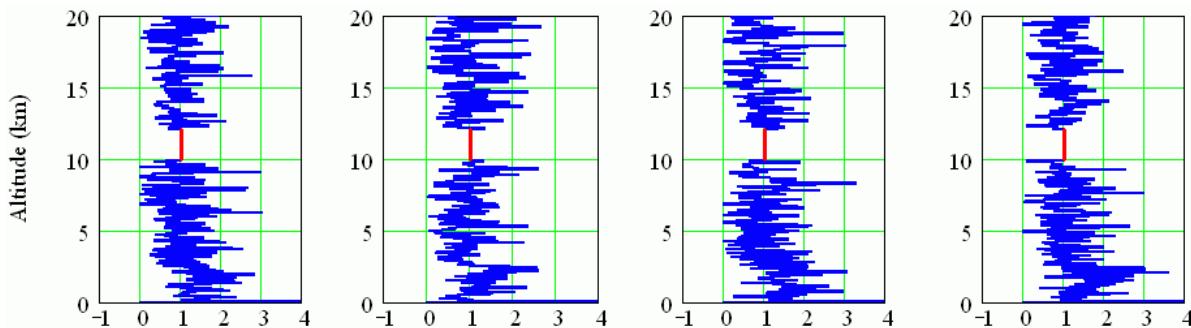
Figure 4.6: Feature Removal with Attenuation Correction

Now when we build the 20-km average we get the desired feature finder result, as shown in Figure 4.7. By including the attenuation correction as an integral component the intensity clearing applied to the 5-km data we are able to locate the underlying aerosol layer with excellent accuracy in the subsequent 20-km average.
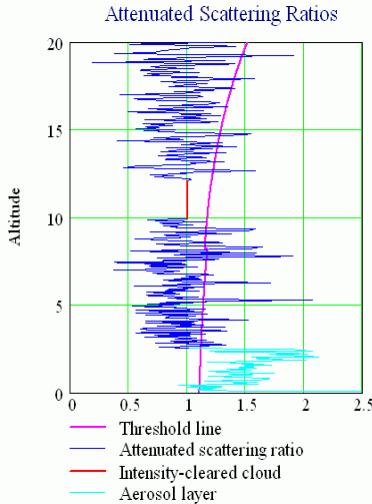
Figure 4.7: 20 km Average with Attenuation Correction Applied

SIBYL can be thought of as a *nested multi-grid feature finder*. The adaptive profile scanning engine presented in section 3 is applied repeatedly to profiles formed by averaging and smoothing over increasingly coarse spatial grids. Consequently, the process described above for moving from 5-km averages to 20-km averages is repeated at the next averaging level using the identical procedure. Four intensity-cleared 20-km profiles are averaged to form a single 80-km profile, and this more highly averaged, lower resolution profile is scanned for the presence of features by the same scanning engine used for all previous feature finding. The only deviations to the general procedure occur at the first and finest averaging resolution and at the last and coarsest resolution. The intensity clearing that follows each profile scan is not applied to the single profile produced at the coarsest spatial resolution, as it is not necessary for any subsequent processing. For the first average, an additional cloud-clearing loop is applied. This loop is described in detail in section 4.4.

4.3. Selection Strategies for "Clear Air" Regions

Effectively rescaling the profiles from which features have been removed requires an accurate estimate of the two-way transmittance for each feature. An initial estimate of each feature's two-way transmittance is made during the profile scanning process, immediately after feature base has been definitively located (see section 3.2.9). This first estimate is used to adjust the threshold level for the remainder of the profile scan. By necessity however (because we do not know in advance where the next feature is), this estimate is made over a limited altitude range in the region just beneath feature base. Therefore, when the scan of the entire profile is complete and the locations of all features detectable at that averaging resolution are known, we then employ a second scanning process to refine our two-way transmittance estimates by locating the most likely region of clear air between features. We illustrate this procedure with the following example.

Figure 4.8 shows the feature-finder results obtained from a scan of a 5-km averaged profile selected from the simulated CALIPSO data presented earlier in Figure 4.2. Because the look-ahead distance for this scan was set at 0.5-km, the initial estimate of feature two-way transmittance is computed in the region between 9.5 and 10.0 km; i.e., immediately beneath the base

of the cirrus layer. Clearly a more robust estimate can be made by computing the mean attenuated scattering ratio over a larger vertical distance. Nevertheless, it is not prudent to use the entire range between the base of one feature and the top of the next lower feature, as weaker layers that are not identified in the earlier, higher resolution scans can cause a significant underestimate of the two-way transmittance. To guard against this sort of error, we define a 'clear air analysis depth' whose magnitude is a linear function of the altitude gap between a given pair of successive features. We then treat this clear air analysis depth as a sliding window, moving it one range bin at a time through the gap between features, and computing a 'clear air score' for each interval. The final estimate for feature two-way transmittance is computed over the interval associated with the minimum clear air score.
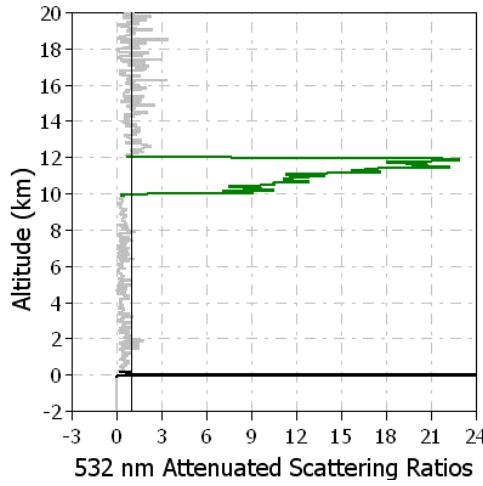
Figure 4.8: Simulated attenuated scattering ratios showing the detection of a cirrus cloud between 10-km and 12-km, and a surface return at 0-km.

To calculate the clear air score, we recognize that, in the absence of noise, the slope of the attenuated scattering ratios (dR'/dZ) in a purely molecular atmosphere should always be zero. In such case, the mean attenuated scattering ratio $(\langle R' \rangle)$ will equal the two-way transmittance of the overlying feature, and will always be less than or equal to one. Therefore, for each interval we compute both dR'/dZ and $\langle R' \rangle$. For those intervals where $0 < \langle R' \rangle \leq 1$, the value of the clear air score is identical to dR'/dZ. Otherwise, if $\langle R' \rangle \leq 0$ or $\langle R' \rangle > 1$, the clear air score is marked as being invalid. Calculating the 'clear air analysis depth', D_{clear}, requires a trio of user-supplied configuration constants such that

- D_{max} is the maximum allowed value of D_{clear};

- gap_{min} is the minimum permissible gap length (i.e., $gap_{min} \geq$ look-ahead distance);

- gap_{max} is the maximum gap width that to be considered.

D_{clear} is then selected according to

$$D_{clear} = \begin{cases} D_0 \text{ for } gap < gap_{min} \\ D_0 + \left(\dfrac{D_{max} - D_0}{gap_{max} - D_0} \right) (gap - D_0) . \\ D_{max} \text{ for } gap > gap_{max} \end{cases} \qquad (4.1)$$

where D_0 represents the look-ahead distance.

4.4. Boundary Layer Cloud-Clearing

As described above, the initial profile scan is performed on profiles averaged over 5 km along-track. However, clouds with horizontal extents less than 5-km are often found in the boundary layer, and may be located anywhere within the surface-attached aerosol layer. These clouds are usually strongly scattering and thus can be detected in single-shot profiles. To locate clouds within the SAL, a special feature detection loop is implemented in SIBYL. This loop examines 5-km averaged profiles and attempts to identify "hot spots" embedded within a newly located, as yet unclassified feature. These hot spots are defined as regions where the backscatter is so intense that we are confident the feature is either a cloud or the surface, based solely on the strength of the signal. By using the clearing technique described below we are able to separate cloud from aerosol at the highest lidar spatial resolution. This separation of clouds and aerosols is necessary to prevent cloud contamination of aerosol extinction retrievals and also aids the scene classification algorithm.

Intensity clearing is applied to features that are assumed to be homogeneous; that is, the transmissive features that we identify and remove in the intensity clearing process are assumed to be either pure cloud or pure aerosol. The application for this assumption is necessitated by the attenuation correction that is required in order to properly compensate for the energy dissipated through those features that are removed. Making reliable estimates of this attenuation involves a certain amount of horizontal averaging, and within the constraints imposed by a fully automated data processing scheme, SIBYL averages as little as possible for each transmissive feature located. Consequently, there is little to be gained by attempting to classify transmissive features at horizontal spatial resolutions finer than 5-km. Even in perfectly clear, totally dark conditions, CALIPSO cannot make accurate backscatter measurements of the Rayleigh atmosphere at single shot or even single kilometer resolution. (For example, see Figure 3.7.) It follows therefore that we cannot obtain a useable two-way transmittance measurement of an attenuated signal on a similar scale. This is a critical point, because the retrieval of extinction and backscatter coefficients from within transmissive features is constrained by the two-way transmittance measurement.

Note however that no attenuation correction is required, or even possible, for opaque features. Because the subsequent extinction retrieval is not constrained by an attenuation measurement, we can therefore take extra liberties in our horizontal averaging of the data. This works to our advantage in the boundary layer, where the features we detect are frequently heterogeneous and, because the aerosol layer is generally attached to the surface, almost always opaque. Figure 4.9 demonstrates a compelling example of the turbulent mix of feature types found in the boundary layer. The red and orange colors below 2.5 km indicate dense aerosol concentrations in the SAL. The semi-continuous thick white line at 0 km is the surface of the Atlantic Ocean south of Brazil. The elevated, intermittent, white "hot spots" embedded in the upper portion of the aerosol layer

are cumulus clouds. Clouds of this type can exist on very small spatial scales. To ensure the homogeneity of features identified as boundary layer aerosols, SIBYL includes a process dedicated specifically to identifying and ultimately removing clouds located within all boundary layer features.
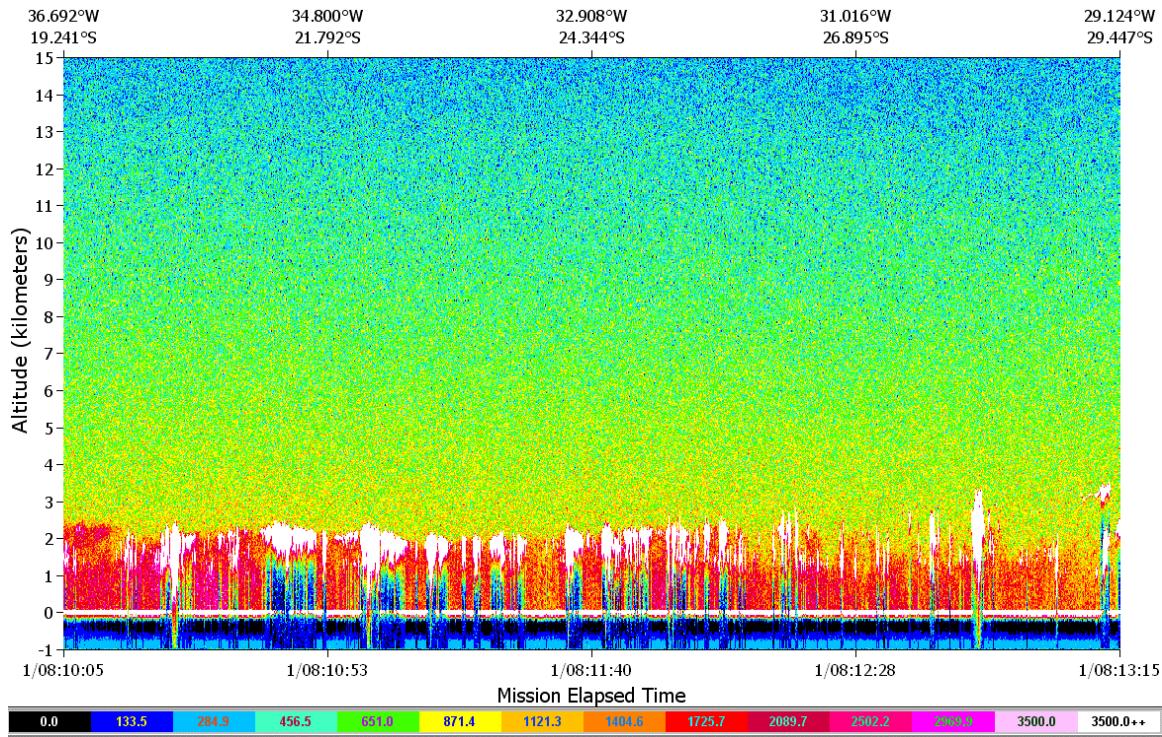
Figure 4.9: Clouds Embedded in Aerosol from LITE Orbit 22

4.4.1. Definition of Terms for Boundary Layer Features

Boundary layer features are (a) connected (i.e., vertically adjacent) to the surface, (b) thick (or at least substantially thicker than a pure surface return), and (c) opaque (because the surface is not transparent).

Homogeneous features are those which are either (a) all cloud (hot spot), or (b) all "not cloud". Recall that at this point no cloud vs. aerosol determination has been made for weak features: all we know is the location of those features that meet our defined criteria for hot spots (which are identified as either clouds or surface returns). For the purposes of computing an extinction profile, the final feature boundaries of a homogeneous feature are identical with the feature boundaries located by the 5-km profile scan. These are the boundaries that will be recorded in the low-resolution scene classification data product.

4.4.2. Rationale

The high-resolution cloud-clearing process (i.e., the upper loop in Figure 4.1) is a method for identifying the locations of any exceptionally strong backscatter within a feature. Based solely on their very high backscatter intensity, these regions are defined a priori as either clouds or surface returns. The motivating assumption underlying this operation is that these hot spots (i.e., clouds) will cover only some fraction of any 5-km sample region. Therefore, once the hot spots

are identified and removed, further averaging can be done in order to isolate and retrieve the scattering characteristics of the surrounding media (presumably aerosol).

Figure 4.10 presents an expanded view of a portion of the data shown in Figure 4.9. In this example, the aerosol layer is ubiquitous, while clouds occur in only 30% of the profiles. Unambiguous determination of the aerosol spatial boundaries requires that the aerosol-only portions of the return be separated from the cloud regions. This fact is clearly demonstrated in the second, seventh, eleventh, and twelfth 5-km segments, where the maximum cloud top protrudes above the top of the aerosol layer.
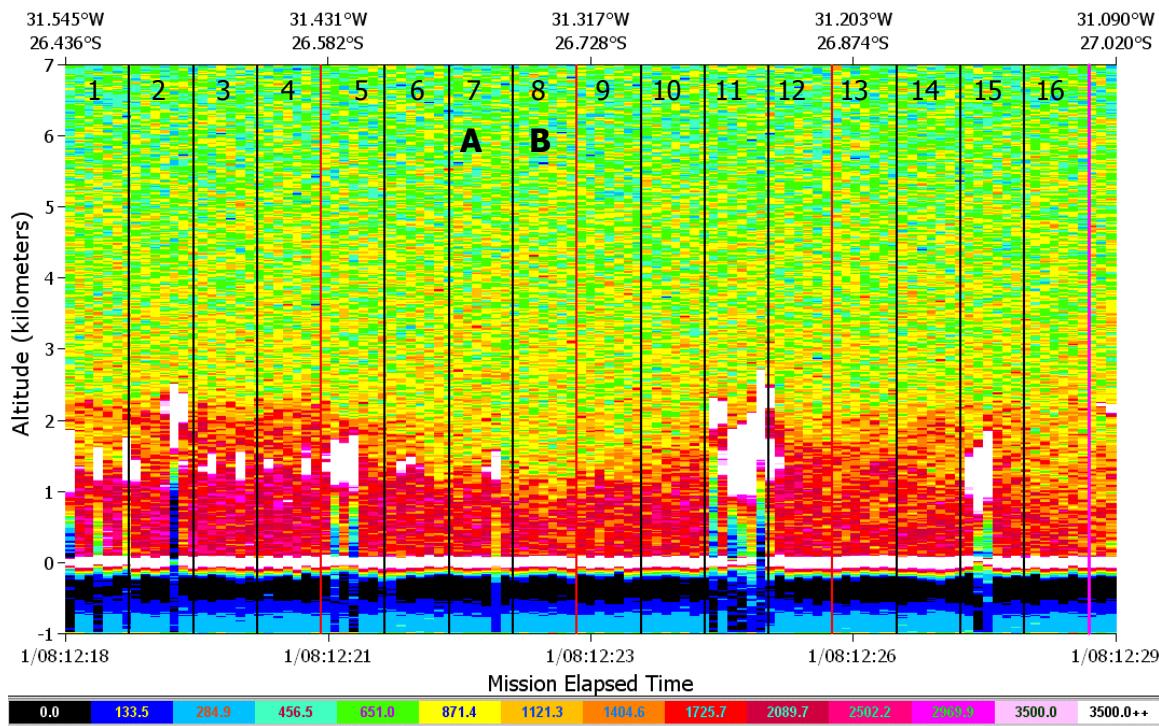
Figure 4.10: Spatial Scales of Cumulus Clouds – each vertical line represents a 5-km horizontal boundary; the red lines indicate 20-km boundaries, and the fuchsia line indicates the end of a full 80-km segment such as would be processed in a complete cycle of SIBYL.

4.4.3.   Procedure

The boundary layer features that are found at SIBYL's fundamental averaging resolution of 5-km may be a heterogeneous mix of clouds, aerosol layers, clear air, and surface returns. To separate the individual components – i.e., the contributions from each laser pulse – we construct a modified threshold array and re-examine those regions identified as containing features using backscatter data at the highest available horizontal resolution.

The differences between contiguous 5-km averages are illustrated in Figure 4.11. These two line plots represent averaged profiles taken from the segments labeled A and B in Figure 4.10. Profile B shows a homogeneous aerosol layer extending from approximately 1.0 km down to a very strong surface return centered at 0.0 km. This same aerosol layer is also present in profile A. However, due to the very brief appearance of a bit of fair weather cumulus in profile A, the top of the composite feature in profile A is approximately 0.5 km higher than the top of the

feature in profile B. Note that in both instances the feature comfortably exceeds the rather conservative threshold line established by the profile-scanning algorithm.
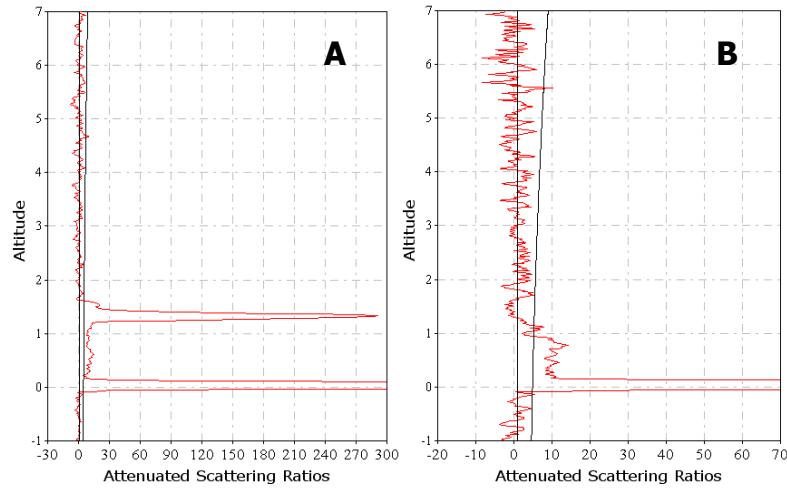
Figure 4.11: LITE 1064 nm attenuated scattering ratios for two consecutive 5-km segments; segment A shows fair weather cumulus embedded in aerosol, while segment B contains a single cloud-free aerosol layer

To separate regions of cloud and aerosol in profile A we now examine the full-resolution profiles from the region of data used to make the current 5-km average. A modified threshold array is constructed for each full-resolution profile, and a revised threshold algorithm is applied *only* to those altitude ranges in which a feature was located by the 5-km profile scan. All data points within the 5-km feature that exceed the modified threshold value are identified as "hot spots".

One prime target for hot spot removal is fair weather cumulus embedded in boundary layer aerosol. Because the expected particle size is significantly different between these two feature types, we can utilize the two-wavelength capability of our lidar to help differentiate one from the other. Instead of rescanning the 532 nm attenuated scattering ratios, the cloud clearing procedure is applied instead to the 1064 nm attenuated backscatter coefficients. Because the 1064 nm channel is much less sensitive to the scattering from small aerosol particles, the contrast between clouds and aerosols is greater at this wavelength. Furthermore, for high intensity features in the boundary layer the 1064 channel has superior SNR relative to the 532 channel.[10] Finally (and perhaps paradoxically), high intensity signals measured in the 1064 channel receive an additional SNR boost from the on-board averaging scheme. (Recall that in the lower troposphere the 1064 channel has been subjected to additional on-board averaging, so that its vertical resolution is 60 meters vs. 30 meters at 532 nm.)

Because molecular backscatter is essentially undetectable in single shot 1064-nm profiles, the threshold array used in this context is a simplified version of the attenuated backscatter coefficient threshold first presented in equation (3.8). For each full-resolution profile we construct a modified threshold array by replacing the weighted range-dependent molecular

---

[10] For low intensity features dark current dominates the 1064 signal. For high intensity signals the high quantum efficiency of the 1064 detector dominates.

model term $\left(T_1 \cdot RBV\left(z\right)\right)$ with a weighted constant term $\left(T_2 \cdot \beta_{MaxAerosol}\right)$.[11] The constant in this new term represents the largest backscatter coefficient we would expect to measure in an aerosol layer. The actual value(s) used for $\beta_{MaxAerosol}$ will be determined empirically from analyses of the available aerosol backscatter data (see Section 4.4.4).

$$PBL_{Threshold}\left(z\right) = T^2_{above} \cdot \left(\beta'_{1064,air}\left(z\right) + T_0 \cdot MBV_{1064}\left(z\right) + T_2 \cdot \beta_{MaxAerosol}\right). \tag{4.2}$$

The complete expression for the revised threshold is given above in equation (4.2). The T^2_{above} term represents the attenuation of the signal within the boundary layer due to any overlying features (e.g., cirrus layers), and is required to properly scale the resulting threshold array. For the purposes of cloud clearing, we can assume that the feature attenuation down to the top of the boundary layer is independent of wavelength. The attenuation due to large particles (clouds) will be essentially the same for both wavelengths. The attenuation from small lofted particles (aerosols) will in general be larger at 532 nm, but in this particular application an overestimation of the total column attenuation leads only to a slightly more conservative estimate of hot spot occurrence. Therefore, an acceptable attenuation correction can be obtained from the mean value of the 532 nm attenuated scattering ratios in the clear region immediately above the boundary layer feature.

$\beta_{MaxAerosol}$ and T_2 are both user-specified coefficients read into the SIBYL analysis from external configuration files. Note too that the value of $\beta_{MaxAerosol}$ can be expected to change as a function of season and/or geophysical location.

A succinct description of the cloud-clearing algorithm is provided by the flowchart shown in Figure 4.14. The details of the procedure are illustrated by Figure 4.12, which shows attenuated backscatter profiles for each of the seven LITE laser pulses used to construct Profile A in Figure 4.11. Upon decomposition, the cloud that is so prominent in Profile A is seen to be present only in the fourth and fifth high-resolution profiles. The black line shown in each plot represents a modified threshold array constructed according to equation (4.2). Hot spots are those data points that rise above the threshold lines. In this example, the surface spike is identified as a hot spot in all seven profiles, as is the cumulus cloud in profiles four and five.

---

[11] Like T_0 and T_1, T_2 is a strictly positive weighting constant.
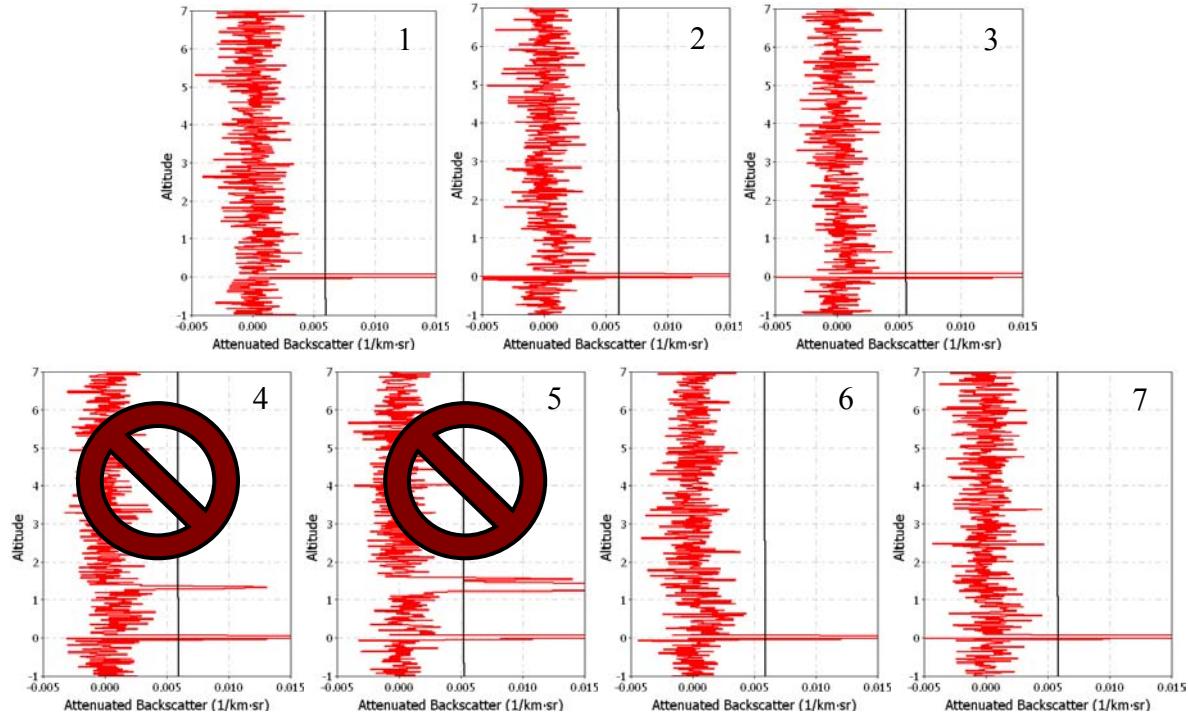
Figure 4.12: LITE High-resolution 1064 nm Attenuated Backscatter Coefficients

For subsequent processing within SIBYL, the region from the top of the uppermost hot spot down to the end of the profile is marked as 'cloud-contaminated' data and removed from consideration in subsequent processing. This step represents the actual cloud clearing.

After all high-resolution profiles have been scanned for hot spots, the data that remains within the original 5-km segment is reaveraged (i.e., all data marked as 'missing' is excluded). The profile scanner is now reapplied to the cloud-cleared segment of the profile to determine whether weaker features can be detected at the 5-km resolution using the reaveraged 532 nm attenuated scattering ratios. Because only the boundary layer features are subject to the cloud-clearing process the profile scanner can reuse the threshold array constructed for the scan of the original 5-km average. This secondary scan is depicted below in Figure 4.13.
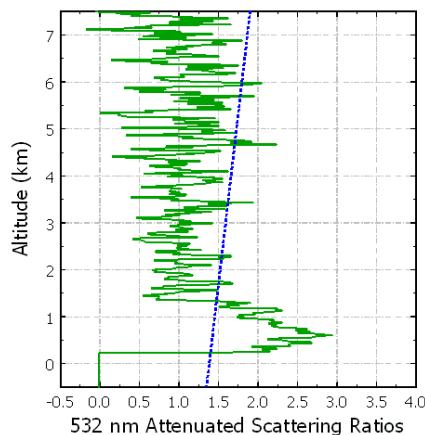
Figure 4.13: Cloud-Cleared and Reaveraged 532 nm Attenuated Scattering Ratios

At this point, those regions identified as hot spots have been *irrevocably removed* from the remainder of the feature finding process. (This data is, of course, tagged and stored, and will resurface later as necessary in other parts of the analysis process.) If, as in the example shown here, the aerosol layer is sufficiently robust to be identified by the secondary 5-km scan it is treated exactly as any other feature found at 5-km.
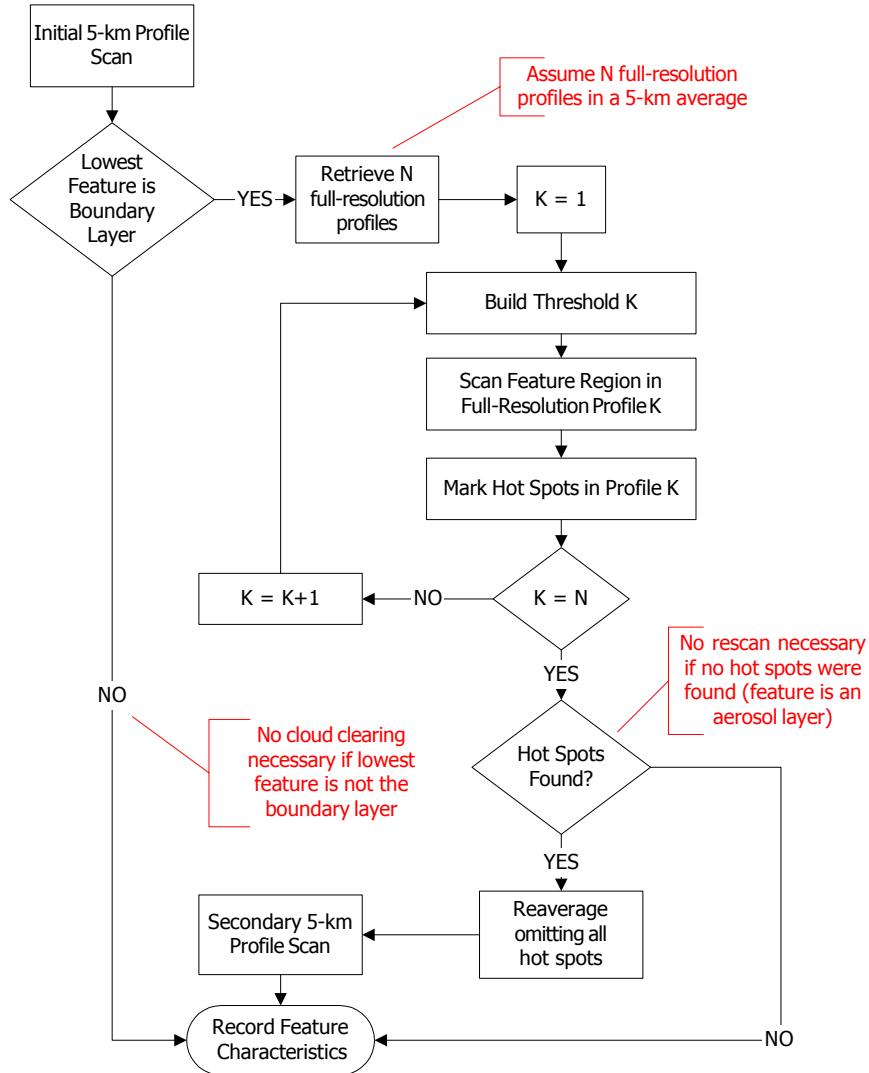
Figure 4.14: Boundary Layer Cloud-Clearing Flowchart

4.4.4.   Obtaining a Best Estimate for $\beta_{MaxAerosol}$

Prior to the start of the CALIPSO mission, $\beta_{MaxAerosol}$ was selected based on the distribution of aerosol backscatter coefficients derived from LITE data during studies conducted as part of the Global Aerosol Climatology Project (GACP; Mishchenko et al., 2002), shown in Figure 4.15. After launch, updates to this value will be made as required throughout the mission. These updates will be derived from aerosol distributions obtained from analysis of the CALIOP lidar data.
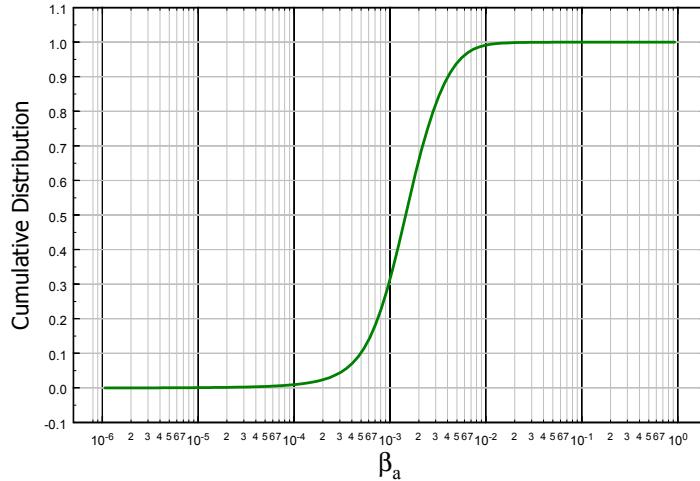
Figure 4.15: Distribution of aerosol backscatter coefficients retrieved from LITE data using S*=25

4.5.   False Positive Interdiction Using an Integrated Backscatter Threshold

False positives are noise excursions that pass all of the feature acceptance tests levied by the profile scanner.   The misidentification of false positives as features can arise from several different sources. As an example, the threshold damping technique describe in section 3.2.9 is aimed at eliminating one specific type of false positive that can occur beneath features associated with unusually low values of $\langle R' \rangle$.   However, additional strategies are also employed to further reduce the number of these spurious features that otherwise will contaminate the data set. Once again, the feature integrated attenuated backscatter plays a significant role in this activity. Because SIBYL makes multiple passes through the data, the algorithm can afford to "miss" faint features in the higher-resolution scans. Assuming these weaker features possess a certain degree of horizontal persistence, additional averaging will improve the contrast between the feature and the surrounding clear air. Thus weak features which are barely detectable at one resolution can be unambiguously identified at the next averaging step.

As we move through SIBYL's sequence of spatial resolutions, from 5-km though 20-km and ultimately to 80-km, we are focusing our detection efforts at each stage on progressively weaker and more tenuous features. The best example of this occurs in the first stage, which includes both the 5-km profile scan and the single profile hot-spot identification. Our goal here is to identify the boundaries of *strong features*. In fact, given our strategy for implementing the extinction retrieval (see PC-SCI-202 Part 4), we specifically DO NOT want to identify weak features at this highest of spatial resolutions. Extinction profiles within features are retrieved at the same horizontal resolution at which they are detected by SIBYL. If insufficient averaging is done for weak features, the SNR within attenuated backscatter profile will be unacceptably low, leading to an extinction retrieval of poor quality.   Therefore, to enable SIBYL to deliver only those features with adequate SNR to the extinction algorithms we impose a lower bound on the scattering intensity of the features that can be reported back from any profile scan. This lower bound is implemented by enforcing a minimum required value for the feature integrated attenuated backscatter. Feature candidates that fail to meet this predetermined γ' threshold value will be rejected. By adopting this strategy we simultaneously introduce yet another technique for

eliminating false positives. Given sufficient horizontal persistence, genuine features rejected at a higher resolution will be retrieved on a subsequent profile scan at a lower resolution, when additional averaging their contrast has improved and the γ' restrictions are less stringent.

SIBYL's initial profile scan provides an excellent example of how this process works. In the troposphere, the weakest clouds we will measure are those that are categorized as *subvisible*, *threshold visible*, or more generically as *thin cirrus*. Such clouds cirrus are common near the tropical tropopause (Beyerle et al., 1998). The threshold level at which a cloud becomes subvisible occurs at an optical depth of approximately 0.03. (Uthe & Russell, 1976; Sassen et al., 1989) Since cirrus cloud lidar ratios are generally in the range of $S_c \approx 25$ (e.g., Sassen & Comstock, 2001), we can use equation (3.15) to derive a table of integrated attenuated backscatter values typical of subvisible cirrus (see Table 4.3). Using this table, we can select a threshold value that will prevent subvisible cirrus from being detected at resolutions of less than 20 km, which will (a) improve the SNR of the profiles that will eventually be passed on the extinction retrieval and (b) reduce the frequency of false positives.

Table 4.3: Integrated Attenuated Backscatter for Subvisible Cirrus

Lidar Ratio	Optical Depth		
	0.020	0.025	0.030
20	9.80×10^{-4}	1.22×10^{-3}	1.46×10^{-3}
25	7.84×10^{-4}	9.75×10^{-4}	1.16×10^{-3}
30	6.54×10^{-4}	8.13×10^{-4}	9.71×10^{-4}

Tests with simulated CALIPSO data derived from LITE nighttime observations indicate that using a threshold value of $\gamma' = 0.0015$ at the 5-km averaging resolution effectively eliminates the large majority of false positives. Particularly during daytime, γ' values such as those shown in Table 4.3 are highly unlikely to be associated with genuine features unless substantial horizontal averaging has been performed.

The false positive rejection strategies described in this section have been integrated into the profile scanner used within the SIBYL algorithm. A flowchart diagramming the modified scanning routine is shown in Figure 4.16.
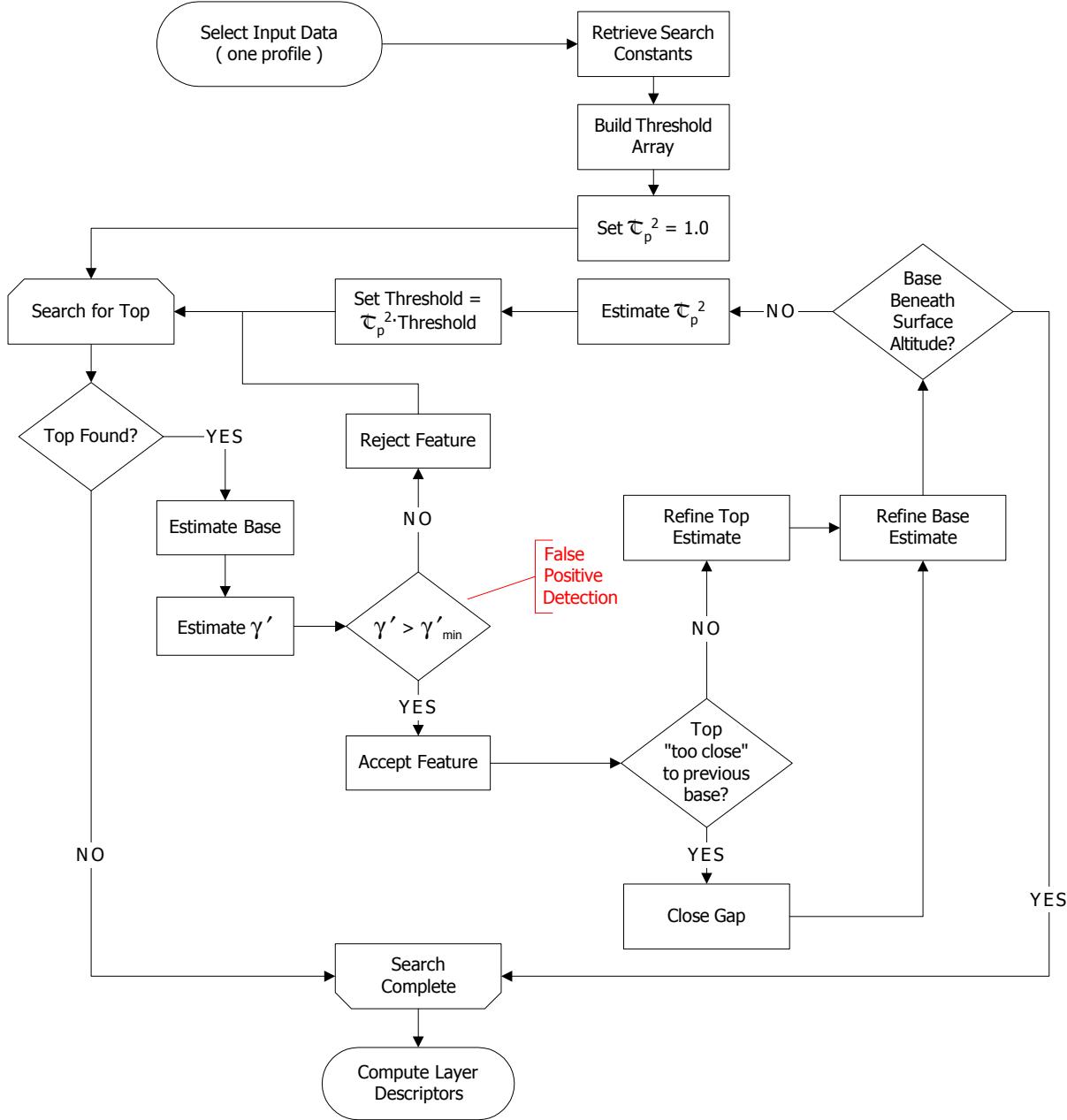
Figure 4.16: Profile Scanning Flowchart Customized for use by SIBYL

4.6. SIBYL Procedure and Detailed Flowchart

In principle, SIBYL can have an arbitrarily large number of averaging levels, as shown in the high-level flowchart presented in Figure 4.1. In practice however, we have limited the number of averaging levels to three: 5-km, 20-km, and 80-km. In this section we outline in detail the specific steps that are taken in a single cycle by the SIBYL procedure. These steps are also described pictorially in Figure 4.17.

Procedure 3: SIBYL: A Nested Multi-Grid Feature Finder

A. Select N consecutive *profile packages*. Each profile package consists of all full resolution profile data spanning a horizontal distance equal to the fundamental (i.e., the first and smallest) averaging distance. For CALIPSO the fundamental averaging distance is 5 kilometers. The number N is determined by dividing the maximum averaging distance, which for CALIPSO is 80 kilometers, by the fundamental averaging distance. For CALIPSO, N=16.

B. For n = 1…N do

B.1. Average over profile package n (horizontal averaging only) to produce a single profile having a uniform 5-km horizontal resolution

B.2. Use the profile scanner to locate features within the nth 5-km averaged profile (section 3)

B.3. If the profile scan locates any features, generate a set of five 1-km horizontal averages for all tropospheric data (20.2-km and lower) within the initial 5-km horizontal segment; search only the altitude regions identified as features in the 5-km scan to identify those features visible at 1-km resolution; calculate layer descriptors for each 1-km feature found

B.4. If the profile scan locates a boundary layer feature…

- Retrieve the constituent parts (i.e., the full resolution profiles) that were used to make the nth profile

- Using the full resolution data, search the altitude region identified as a boundary layer feature to identify hot spots (section 4.3)

- Calculate layer descriptors for each hot spot (section 6); mark hot spots as "missing data"

- Re-average profile package n, omitting **all hot spots** from the calculation (section 4.3)

- Re-apply the profile scanner to the reformed n^{th} profile (section 3)

B.5. Calculate layer descriptors for each feature found in the nth profile (section 6)

B.6. Remove all features found in the nth profile, applying the appropriate attenuation correction beneath each (section 4.1)

C. At the conclusion of Step B SIBYL will have generated N=16 intensity-cleared profiles. Each profile represents an average over the designated fundamental averaging distance (5-km). These N profiles are now partitioned into K groups such that the horizontal distance spanned by each group is equal to CALIPSO's intermediate averaging distance of 20 km. (K=4)

D. for k = 1…K do

D.1. Average all 5-km profiles in the kth group to produce a single profile having a horizontal resolution of 20-km

D.2. Use the profile scanner to locate features within the kth averaged profile (section 3)

D.3. Calculate layer descriptors for each feature found in the kth profile (section 6)

D.4. Remove all features found in the kth profile, applying the appropriate attenuation correction beneath each (section 4.1)

E. At the conclusion of Step D SIBYL will have generated another K=4 intensity-cleared profiles. Each of these new profiles represents an average over the designated intermediate averaging distance (20-km). All K of these profiles are now averaged a single profile having a horizontal resolution of 80-km.

F. Use the profile scanner to locate features within this single 80-km averaged profile (section 3)

G. Calculate layer descriptors for each feature found in the 80-km profile (section 6)

Retrieve 16 Level II Frames
1 Frame = 5 km = 15 laser shots

n = 1

Reaverage cloud-cleared profiles to 5-km

Cloud-clear boundary layer features by searching single shot data used to build profile n

Full resolution layer descriptors

Average horizontally over frame n to produce a single profile

Apply Profile Scanning Engine to find layers in profile n (5-km average)

Average frame data to 1-km; apply Profile Scanning Engine in 5-km feature regions

1-km layer descriptors

n = n+1

Features Found?

No

Yes

Cloud cleared?

No

Intensity Clearing

Yes

No

n = 16?

Compute layer descriptors; mark layer locations as "missing data" in frame n

5-km layer descriptors

Yes

k = 1

Build profile j by averaging intensity cleared data over 4 frames (4k-3 ... 4k)

Apply Profile Scanning Engine to find layers in profile k (20 km average)

Intensity Clearing

k = k+1

Compute layer descriptors; mark layer locations as "missing data" in profile k

20-km layer descriptors

No

k = 4?

Yes

Average 4 intensity-cleared 20-km profiles to produce a single 80 km profile

Apply Profile Scanning Engine to find layers in 16-frame (80 km) average

Compute layer descriptors

Done

80-km layer descriptors

Figure 4.17: SIBYL Internals

4.7.   Internal and External Outputs

For every instantiation of the profile scanner, SIBYL retrieves a comprehensive record of the search results returned by the scanning engine. This record includes information locating the profile in time and space and describing (via layer descriptors) all of the features that were found (see section 3.4 for details). After processing each block of input data, SIBYL generates a set of profile scanner results obtained from one 80-km averaged profile, four 20-km averaged profiles, sixteen 5-km averaged profiles, eighty 1-km averaged profiles, and 240 single-shot profiles. Over the course of each individual orbit, a sequential time history of these scanner results is stored in a number of HDF (hierarchical data format) files.[12] These are intermediate files which are used to produce the vertical feature mask and other archived data products described in the CALIPSO Data Products Catalog (PC-SCI-503).

5. Uncertainty Analyses

Our ability to accurately locate cloud and aerosol boundaries in backscatter lidar data is affected by several types of errors: timing and ranging errors, errors resulting from noise in the lidar signal, errors due to inaccuracies in the ancillary data, and errors due post-acquisition data processing. Timing errors are inherently very small, and ranging errors are expected to be less than the maximum range resolution of 30 m. This section analyzes the dominant uncertainties in the feature-finding process, which primarily arise from noise in the backscatter signal, and imperfect correction of the attenuation from overlying features.

5.1.   Minimum detectable backscatter coefficient

The amount of signal averaging required before retrievals are performed depends on the scattering strength of the atmospheric targets and the lidar system performance, as well as characteristics of the retrieval algorithm.   Predictions of detection sensitivity based on the theory presented in Section 2 are presented below. These results were derived for a detection probability of 90%, with a threshold chosen such that the false alarm probability is 10%. While a 10% false alarm rate is unacceptably high, various aspects of the SIBYL algorithm which are not modeled by the theory (e.g., the minimum feature thickness and the false positive rejection scheme) will significantly reduce the false alarm rate.

Predicted minimum detectable target strengths, in terms of 532 nm scattering ratios and 532 nm backscatter coefficients, are shown in Table 5.1 for all horizontal averages processed by the SIBYL algorithm. Table 5.1(a) lists the predicted sensitivities at altitude of 1 km for scans done at the full vertical resolution of the downlinked data (30 meters). Table 5.1(b) provides the same information for data averaged vertically to a resolution of 60 meters.

---

[12] More information about HDF can be found at the HDF web site maintained by the National Center for Supercomputing Applications; see http://hdf.ncsa.uiuc.edu/

Table 5.1(a) : Minimum detectable backscatter coefficient (km^{-1} sr^{-1}) at 532 nm, 1 km altitude, and full vertical resolution (30 meters), for both day and night

Laser Pulses Averaged	Horizontal Resolution (km)	Day		Night	
		R_{min}	β_{min}	R_{min}	β_{min}
1	0.333	14.22	1.91×10^{-2}	12.56	1.67×10^{-2}
3	1	7.06	8.75×10^{-3}	6.10	7.37×10^{-3}
15	5	3.17	3.14×10^{-3}	2.75	2.53×10^{-3}
60	20	1.98	1.41×10^{-3}	1.77	1.11×10^{-3}
240	80	1.46	6.68×10^{-4}	1.36	5.14×10^{-4}

Table 5.1(b) : as in Table 5.1(a), but for data averaged to a 60 meter vertical resolution.

Laser Pulses Averaged	Horizontal Resolution (km)	Day		Night	
		R_{min}	β_{min}	R_{min}	β_{min}
1	0.333	9.02	1.16×10^{-2}	7.84	9.89×10^{-3}
3	1	4.83	5.54×10^{-3}	4.16	4.57×10^{-3}
15	5	2.45	2.09×10^{-3}	2.15	1.66×10^{-3}
60	20	1.67	9.68×10^{-4}	1.52	7.50×10^{-4}
240	80	1.32	4.65×10^{-4}	1.25	3.56×10^{-4}

As can be seen from the tables, virtually all clouds in the boundary layer will be detectable at single-shot resolution. However, only unusually high aerosol concentrations will be detectable at this resolution. Horizontal averages of many kilometers will be necessary to detect aerosols. Additional averaging in the vertical can be performed to reduce the requirements for horizontal averaging.

Table 5.2 shows the predicted minimum detectable target strengths, in terms of scattering ratio at 532 nm and in terms of backscatter coefficients, at an altitude of 10 km for a vertical resolution of 60 meters for several horizontal averages. Single shot estimates are not shown, as only averaged profiles are downlinked above 8.2 km. Minimum backscatter coefficients are smaller than in Table 5.1, not only due to the on-board averaging that has occurred, but also because the Rayleigh noise at 10 km altitude is less than at 1 km altitude. Table 5.2 indicates that most cirrus can be detected, day or night, at the full resolution of the downlinked data (60 meters by 1 km). Even thin subvisible tropopause cirrus should be detectable from 30-shot averages. This analysis will be continued to help determine the horizontal and vertical averaging appropriate for computing the final cloud output data products.

Table 5.2 : Minimum detectable backscatter coefficient (km^{-1} sr^{-1}) at 532 nm and 10 km altitude, day and night

Laser Pulses Averaged	Horizontal Resolution (km)	Day		Night	
		R$_{min}$	β$_{min}$	R$_{min}$	β$_{min}$
3	1.00	7.66	3.58×10^{-3}	6.01	2.69×10^{-3}
15	5.00	3.46	1.32×10^{-3}	2.72	9.24×10^{-4}
60	20.00	2.13	6.04×10^{-4}	1.75	4.06×10^{-4}
240	80.00	1.54	2.88×10^{-4}	1.35	1.89×10^{-4}

5.2. Detection efficiency using simulated data

To verify that the performance of the CALIOP profile scanner meets or exceeds the theoretically predicted levels, an extensive simulation study was conducted. For this study we designed a prototype scene, with layer heights chosen to test detection efficiencies both in the PBL region (base = 1-km, top = 3-km) and in the mid-troposphere (base = 9-km, top = 11-km). Throughout this scene, feature backscatter coefficients were held constant vertically within each full resolution profile, but were varied horizontally in a step-wise fashion over a range spanning two orders of magnitude (0.0002 km^{-1} sr^{-1} to 0.05 km^{-1} sr^{-1}). A listing of the spatial and optical parameters used in building the prototype scene is provided in Table 5.3. A single realization of this scene is pictured in Figure 5.1(a).

Table 5.3 : Simulation parameters for the prototype scene shown in Figure 5.1

Segment	Horizontal Location	Vertical Location	Backscatter Coefficient	Optical Depth
1	0-km – 80-km	1-km – 3-km	2.0×10^{-4} km^{-1} sr^{-1}	0.008
2	80-km – 160-km	1-km – 3-km	5.0×10^{-4} km^{-1} sr^{-1}	0.02
3	160-km – 240-km	1-km – 3-km	1.0×10^{-3} km^{-1} sr^{-1}	0.04
4	240-km – 320-km	1-km – 3-km	2.0×10^{-3} km^{-1} sr^{-1}	0.08
5	320-km – 400-km	1-km – 3-km	5.0×10^{-3} km^{-1} sr^{-1}	0.2
6	400-km – 480-km	1-km – 3-km	1.0×10^{-2} km^{-1} sr^{-1}	0.4
7	480-km – 560-km	1-km – 3-km	2.0×10^{-2} km^{-1} sr^{-1}	0.8
8	560-km – 640-km	1-km – 3-km	5.0×10^{-2} km^{-1} sr^{-1}	2
9	640-km – 720-km	9-km – 11-km	2.0×10^{-4} km^{-1} sr^{-1}	0.008
10	720-km – 800-km	9-km – 11-km	5.0×10^{-4} km^{-1} sr^{-1}	0.02
11	800-km – 880-km	9-km – 11-km	1.0×10^{-3} km^{-1} sr^{-1}	0.04
12	880-km – 960-km	9-km – 11-km	2.0×10^{-3} km^{-1} sr^{-1}	0.08
13	960-km – 1040-km	9-km – 11-km	5.0×10^{-3} km^{-1} sr^{-1}	0.2
14	1040-km – 1120-km	9-km – 11-km	1.0×10^{-2} km^{-1} sr^{-1}	0.4
15	1120-km – 1200-km	9-km – 11-km	2.0×10^{-2} km^{-1} sr^{-1}	0.8
16	1200-km – 1280-km	9-km – 11-km	5.0×10^{-2} km^{-1} sr^{-1}	2

To quantify the sensitivity of the profile scanner and the effects of random noise on our detection frequency, we generated, analyzed, and compiled statistics for 100 realizations of the prototype scene. The SIBYL profile scanner was used to determine feature boundaries within each of these 100 instances at five different horizontal averaging intervals: full resolution, 1-km, 5-km, 20-km and 80-km. All profiles were scanned at the full vertical resolution of the downlinked data; no additional vertical averaging or smoothing was performed for any of the tests. The nested multi-grid search capabilities provided by the SIBYL algorithm (see section 4) were not used for these trials. However, the false positive rejection scheme described in section 4.5 was employed, and, as will be explained presently, had a noticeable effect on some of the results.
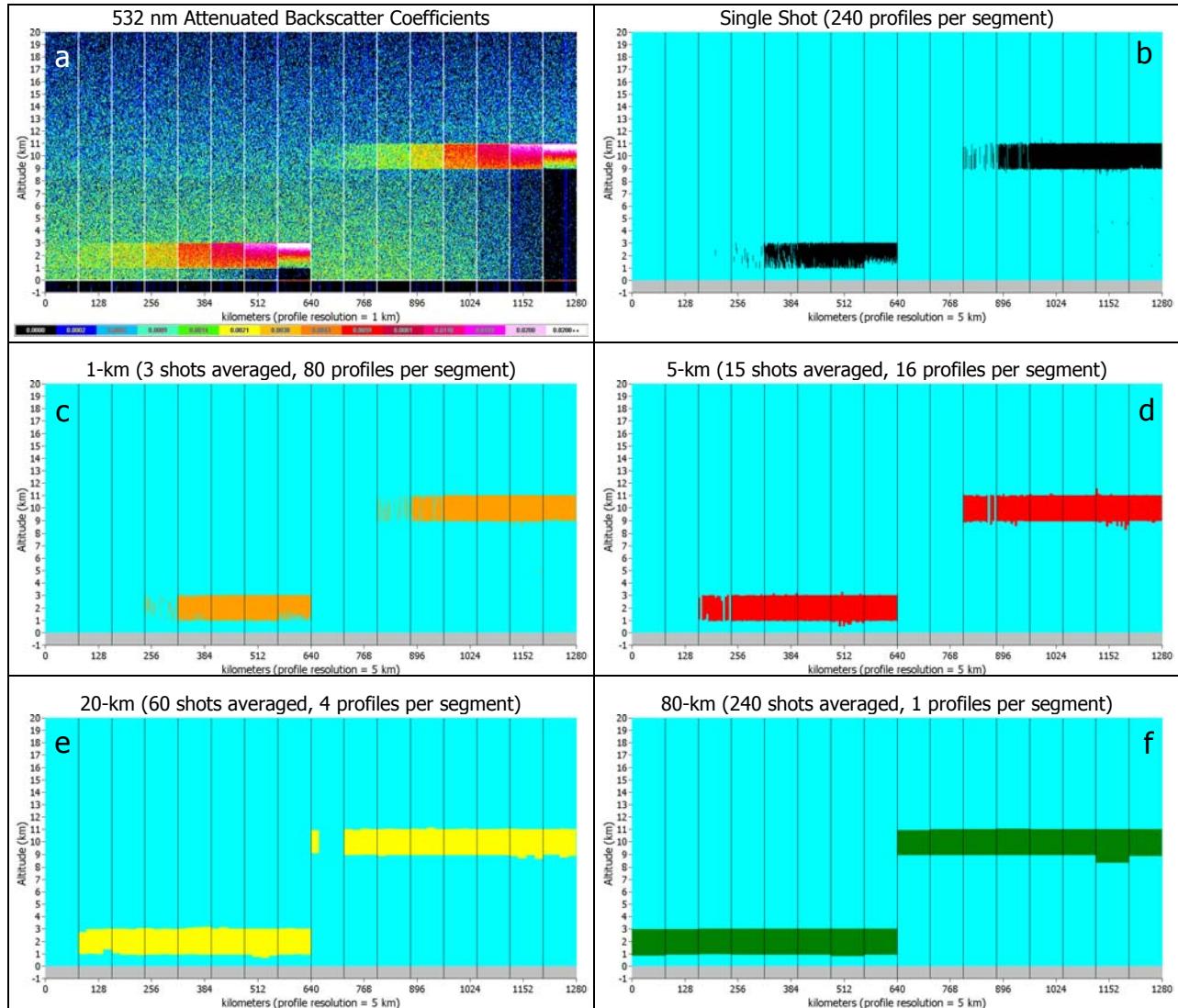
Figure 5.1 : Examples of (a) simulated nighttime measurements of attenuated backscatter coefficients at 532 nm; (b) results obtained searching full resolution profiles; (c) results obtained searching profiles averaged to 1-km horizontally; (d) results obtained searching profiles averaged to 5-km horizontally; (e) results obtained searching profiles averaged to 20-km horizontally; (f) results obtained searching profiles averaged to 80-km horizontally.

Typical results obtained in the analysis of a single nighttime realization are shown in Figure 5.1. Panels (b) through (f) show, respectively, the locations of features found at the full resolution of the downlinked data, and at horizontal averaging intervals of 1-km, 5-km, 20-km, and 80-km. Due to the configuration of the on-board averaging scheme, only the data below 8.2-km is downlinked at the full 1/3-km horizontal resolution of the lidar. Between 8.2-km and 20.2-km the data is averaged on-board the satellite to a nominal horizontal resolution of 1-km. Therefore, in panel (b) the features in segments 9 through 16 are retrieved using a 1-km horizontal averaging interval.

As shown in Table 5.1, the minimum detectable aerosol backscatter coefficient for boundary layer features at the 1/3-km averaging interval is 1.67×10^{-2} km^{-1} sr^{-1}. Thus in Figure 5.1(b), which shows the detection results for the search at 1/3-km, we would expect to see a 90% or better success rate in segments 7 and 8, where the particulate backscatter coefficients are 2.0×10^{-2} km^{-1} sr^{-1} and 5.0×10^{-2} km^{-1} sr^{-1}, respectively. Similarly, for the 1-km search shown in Figure 5.1(c), where the minimum detectable backscatter coefficient is 7.37×10^{-3} km^{-1} sr^{-1}, we would expect to see a 90% or better success rate in segments 6 through 8. We should also see this same 90% or better success rate in segments 5 through 8 for the 5-km search (see Figure 5.1(d)). For the 20-km horizontal averages (Figure 5.1(e)), a 90% or better success rate should occur in segments 4 through 8. At this resolution, the success rate should also approach 90% in segment 3, for which the particulate backscatter coefficient is 1.0×10^{-3} km^{-1} sr^{-1} and the minimum detectable backscatter coefficient (at the 90% level) is 1.11×10^{-3} km^{-1} sr^{-1}. Likewise, at the 80-km averaging interval (Figure 5.1(f)), the success rate should be close to 90% in segment 2, and should equal or exceed 90% in segments 3 through 8. In the example shown here, the theoretical expectations are met by the practical implementation in every instance.

Composite statistics computed over the full set of 100 realizations for the aerosol segments (i.e., 1–8) of the prototype scene are provided in Figure 5.2 (see the 'Detection Frequency' numbers in each panel). These results confirm the findings illustrated by Figure 5.1 and described in the preceding paragraph. In fact, the SIBYL profile scanner performs considerably better than the theoretical predictions for all horizontal averages. For example, at the 80-km averaging interval, the SIBYL algorithm achieves a 99% success rate when measuring an aerosol layer having a backscatter coefficient of 2.0×10^{-4} km^{-1} sr^{-1}. Theory predicts a lower success rate (90%) for a higher backscatter coefficient (5.14×10^{-4} km^{-1} sr^{-1}), and thus in this instance the sensitivity of the actual profile scanner exceeds the theoretical expectation by a factor of 2.5. This marked difference between theory and practice can be attributed to the fact the SIBYL can employ much lower threshold values than are specified in the theoretical studies. The false positives that would otherwise occur are eliminated by the enforcement of SIBYL's minimum feature thickness requirement; i.e., some number of consecutive data points must exceed the threshold before a feature is positively identified.

More insight into the performance levels that will be achieved by SIBYL can be obtained by comparing the test results to the predicted success rates for the 60-meter vertical averages given in Table 5.1. With one illuminating exception, the feature finder results achieved when scanning at a 30-meter vertical resolution – i.e., with no vertical averaging – exceed those predicted for the averaged profiles. The exception occurs in segment 3 during the 5-km search. While the layer backscatter coefficient in this region is just slightly below the 90% expected detection threshold (1.0×10^{-3} km^{-1} sr^{-1} vs. 1.101×10^{-3} km^{-1} sr^{-1}), the actual rate of detection is substantially lower (42%). This apparent sub-optimal performance is due to the "protective actions" enforced by

SIBYL's false positive rejection scheme. For series of tests reported here, the false positive threshold for the 5-km search was set at $\gamma'_{threshold}$ = 0.0015 sr^{-1}. The integrated attenuated backscatter for the feature in segment 3 is 0.002 sr^{-1}. Realizing that noise in the signal can prevent the profile scanner from identifying the full extent of a feature, and that, again due to the influence of noise, the integrated attenuated backscatter measurements will actually be randomly distributed about a mean of 0.002 sr^{-1}, a detection frequency of 42% appears to be quite reasonable given the relatively high value (with respect to the target) of the rejection threshold. When the test was repeated with the false positive rejection threshold halved, to a value of 0.0075, the 5-km detection frequency in segment 3 rose to 88%.

In the discussions above, a successful trial is said to occur whenever any fraction of a feature is correctly detected. However, because features have vertical extent, simple hit rate calculations are not sufficient for characterizing the efficacy of the detection scheme. A second critical issue must also be addressed: on those occasions when we detect a feature, we need to ascertain the vertical extent of the feature detected. The bar graphs shown in Figure 5.2 provide this information. For averaging intervals of 5-km and greater, if the detection frequency exceeds 90% the profile scanner will correctly identify the entire extent of the layer for all but the very weakest of features (see segment 1 for an averaging interval of 80-km). Paradoxically, when using averaging intervals of 1/3-km or 1-km, the layer fraction detected actually decreases for the most intense features. This situation, illustrated by the diagram in Figure 5.3, is caused by strong attenuation within the feature. When only a minimum of horizontal averaging is done, the SNR at the base of these features can become quite low, and hence the base estimation and refinement strategies described in sections 3.2.7 and 3.2.8 can terminate prematurely.

Table 5.4 lists detection frequencies and the mean feature size for the detected features for all segments and all averaging intervals. Examining the results shows that, for the lower backscatter coefficients, features at the higher altitude are detected more frequently than those at the lower altitude (e.g., compare the 5-km results in segment 3 to those in segment 11). This disparity in detection efficiency is due to the increased contrast with respect to the molecular atmosphere that a fixed backscatter coefficient will exhibit with increasing altitude. A backscatter coefficient of 2.0 x 10^{-3} km^{-1} sr^{-1} will generate an attenuated scattering ratio of ~2.8 at an altitude of 3-km, whereas at 11-km the attenuated scattering ratio produced by same backscatter coefficient will rise ~5.7.
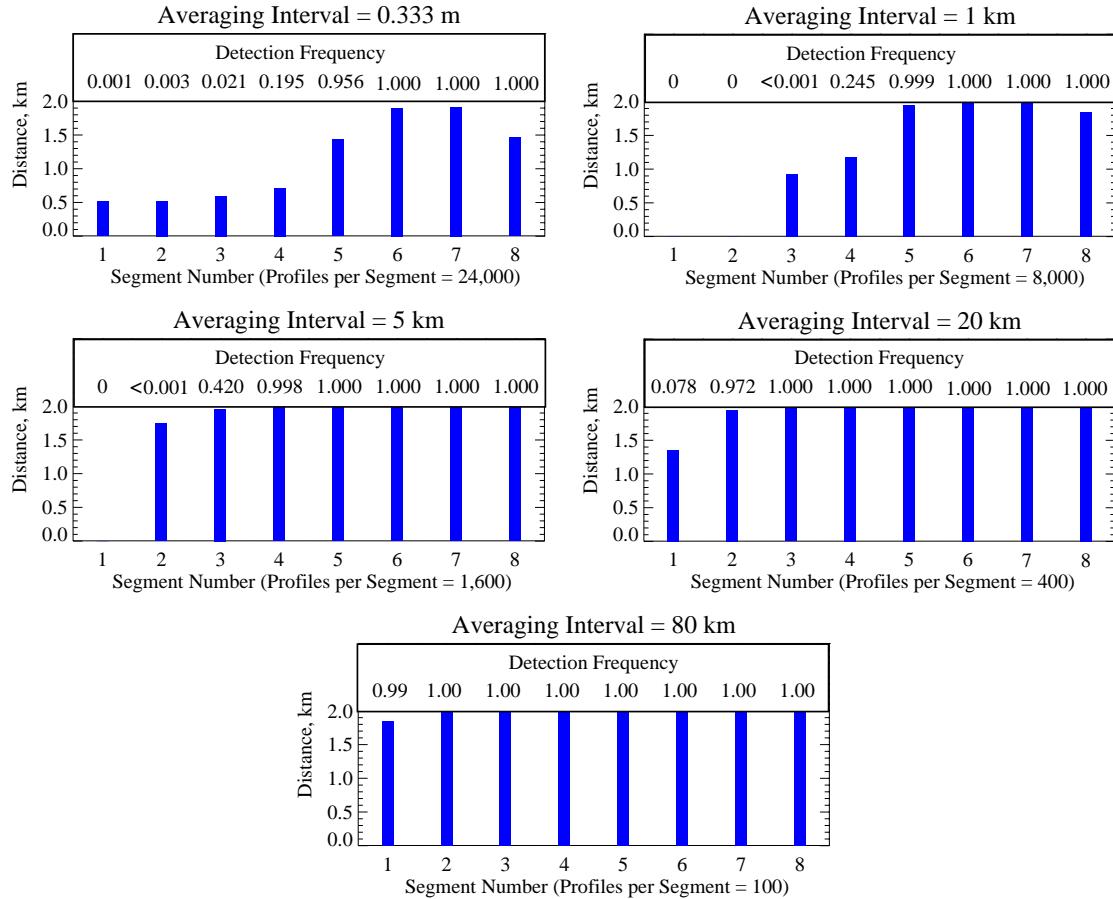
Figure 5.2 : Aerosol layer detection for the minimum detectable backscatter simulations; test segment numbers are listed along the bottom of each panel; the bar graphs plot the mean feature thickness for all features detected within each segment; detection frequency within each segment is listed above each bar.
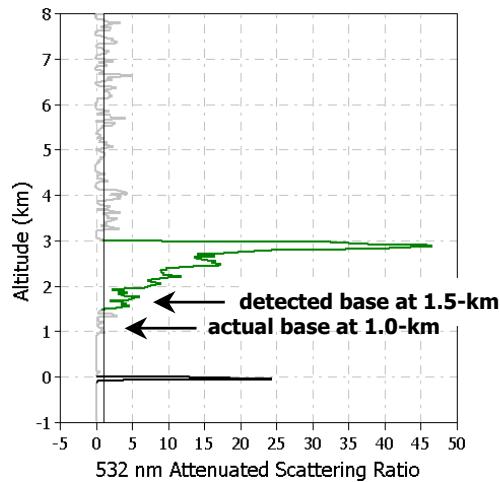
Figure 5.3 : Premature termination of base altitude search due to low SNR cause by strong attenuation within a feature is illustrated using a 1-km horizontal average extracted from the prototype scene shown in Figure 5.1(a).

Table 5.4: Detection results for all segments. The detection frequency number describes how often any fraction of a feature was detected within the target area (target areas for each segment are listed in Table 5.3); the mean thickness value describes the average vertical extent of all features detected within a segment.

| Segment	Detection Frequency / Mean Thickness				
	Single Shot	1-km	5-km	20-km	80-km
1	0.001 / 0.514	0.000 / 0.000	0.000 / 0.000	0.078 / 1.355	0.990 / 1.854
2	0.003 / 0.523	0.000 / 0.000	0.001 / 1.740	0.973 / 1.950	1.000 / 2.032
3	0.021 / 0.593	0.004 / 0.930	0.420 / 1.959	1.000 / 2.030	1.000 / 2.037
4	0.195 / 0.718	0.245 / 1.178	0.998 / 2.024	1.000 / 2.038	1.000 / 2.040
5	0.956 / 1.433	0.999 / 1.950	1.000 / 2.039	1.000 / 2.040	1.000 / 2.040
6	1.000 / 1.903	1.000 / 2.023	1.000 / 2.040	1.000 / 2.040	1.000 / 2.040
7	1.000 / 1.914	1.000 / 2.022	1.000 / 2.040	1.000 / 2.040	1.000 / 2.040
8	1.000 / 1.466	1.000 / 1.845	1.000 / 2.038	1.000 / 2.038	1.000 / 2.040
9	NA	0.000 / 0.000	0.000 / 0.000	0.330 / 1.943	1.000 / 2.081
10	0.003 / 1.622	0.003 / 1.622	0.010 / 2.081	1.000 / 2.100	1.000 / 2.119
11	0.223 / 1.711	0.223 / 1.711	0.844 / 2.087	1.000 / 2.116	1.000 / 2.116
12	0.948 / 1.984	0.948 / 1.984	1.000 / 2.109	1.000 / 2.121	1.000 / 2.120
13	1.000 / 2.104	1.000 / 2.104	1.000 / 2.114	1.000 / 2.119	1.000 / 2.118
14	1.000 / 2.113	1.000 / 2.113	1.000 / 2.114	1.000 / 2.117	1.000 / 2.119
15	1.000 / 2.113	1.000 / 2.113	1.000 / 2.115	1.000 / 2.122	1.000 / 2.125
16	1.000 / 2.091	1.000 / 2.091	1.000 / 2.114	1.000 / 2.120	1.000 / 2.119

5.3. Stability of the SIBYL Algorithm

The tests reported in section 0 verify the performance of the SIBYL profile scanning algorithm with respect to theoretically predicted minimum detectable backscatter coefficients. To assess the stability and end-to-end performance of the complete SIBYL algorithm, a second, more comprehensive set of tests was conducted. For this additional test sequence we used simulations derived from multi-layer LITE measurements of aerosols and broken cirrus clouds acquired over western Kazakhstan during orbit 79. The LITE source data is shown in panel (a) of Figure 5.4. The spatial and optical properties of the features in this scene were extracted and used as input to the CALIPSO lidar simulation software (Powell et al., 2002). Panel (b) of Figure 5.4 shows the "truth template" derived from the data in panel (a), with the green regions indicating the presence of features (either cirrus cloud or aerosol). Within this truth template, the aerosol layer extends uniformly from the surface (0-km) to 2.5-km, with an optical depth of 0.09 and a lidar ratio of 40 sr at 532 nm. The cirrus clouds are spatially intermittent, with a mean lidar ratio of approximately 20 sr, and an optical depth that varies from a low of ~0.01 (subvisible) to a maximum of 0.67.

The stability test was constructed in a manner similar to the earlier detection efficiency test. 100 realizations of the test scene were generated for both nighttime and daytime lighting conditions. Examples scenes can be seen in Figure 5.4(c) and Figure 5.4(d). Each scene was then analyzed using the SIBYL algorithm, configured as described in Table 5.5. By comparing the detection results to the truth template, three metrics were tabulated for each trial: the success rate, the false positive rate (which is equal to the total area of all false positives divided by the total "clear air"

area), and the missed feature or false negative rate (equal to the total area of all missed features divided by the total true feature area).

Table 5.5: Configuration of the SIBYL algorithm for the feature finder stability test; the numbers reported here are the default values for the production version of SIBYL.

| Parameter	Night Value	Day Value
Search Start Altitude	30.0 km	30.0 km
Search Stop Altitude	-1.5 km	-1.5 km
MBV	1.50	1.75
RBV	1.50	1.50
Spike Threshold Factor	10.0	50.0
$S_{reasonable}$	40.00	30.00
False Positive Threshold (γ)	0.00150	0.00150
Minimum Feature Thickness		
Lower stratosphere (30.1 km to 20.2 km)	540	540
Upper troposphere (20.2 km to 8.3 km)	240	240
Lower troposphere (8.3 km to –0.5 km)	180	180
Spike Thickness		
Lower stratosphere (30.1 km to 20.2 km)	360	360
Upper troposphere (20.2 km to 8.3 km)	120	120
Lower troposphere (8.3 km to –0.5 km)	90	90
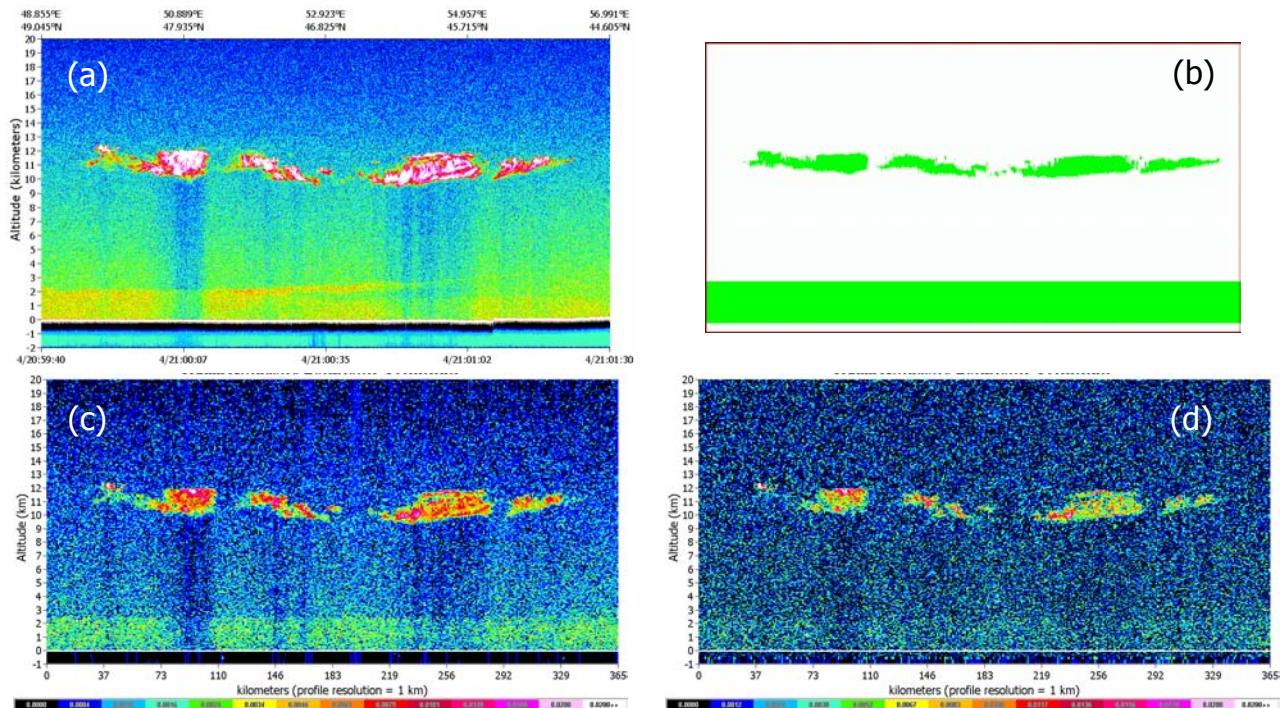
Figure 5.4 : Nighttime (panel (c)) and daytime (panel (d)) CALIOP simulations derived from the LITE measurements shown in panel (a). The green regions in panel (b) specify the locations of all cloud and aerosol layers within the test scene.

Figure 5.5 shows typical feature finder results for an arbitrarily selected nighttime (panel (a)) and daytime (panel (b)) instantiations of the test data. In these plots, the green color represents those areas where SIBYL correctly identified a feature. The white portions of the plots also indicate a successful outcome, as they represent areas of clear air, where SIBYL rightly did not identify a feature. False positives are shown in black. In the scene selected for this test, the intermittent nature of the clouds, combined with SIBYL's variable averaging scheme, leads to many more false positives being found in and around the cirrus deck as compared to within the homogeneous aerosol layer. Those areas where SIBYL failed to detect a feature are shown in red. While the aerosol detection rates are quite good during the nighttime example, due to the greatly increased levels of background noise, they deteriorate noticeably in the daytime example. For the nighttime scene, the false negative rate is 6.8% and the false positive rate is 0.9%. The corresponding numbers for the daytime scene are 47.7% and 0.8%, respectively.
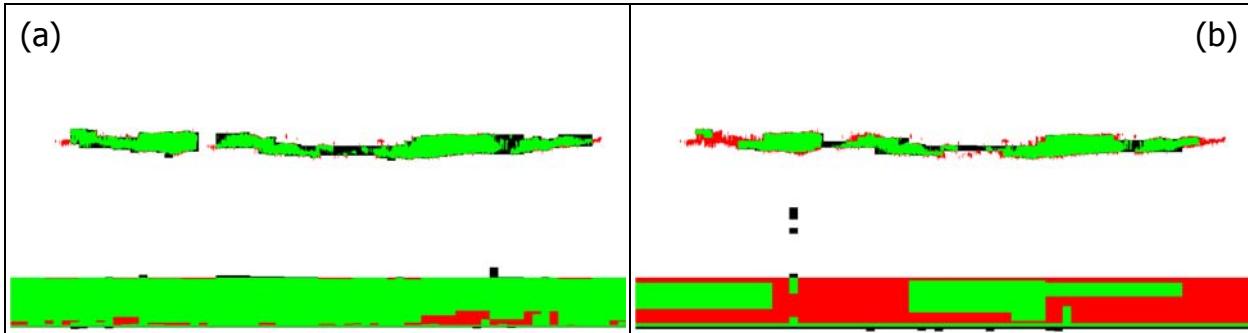
Figure 5.5: Representative feature finder results obtained using SIBYL for nighttime (panel (a)) and daytime (panel (b)) instantiations of the test scene. The green regions of the plots indicate areas where SIBYL successfully identified a feature; the black regions represent false positives; and the red portions indicate those areas where the feature detection algorithm failed.
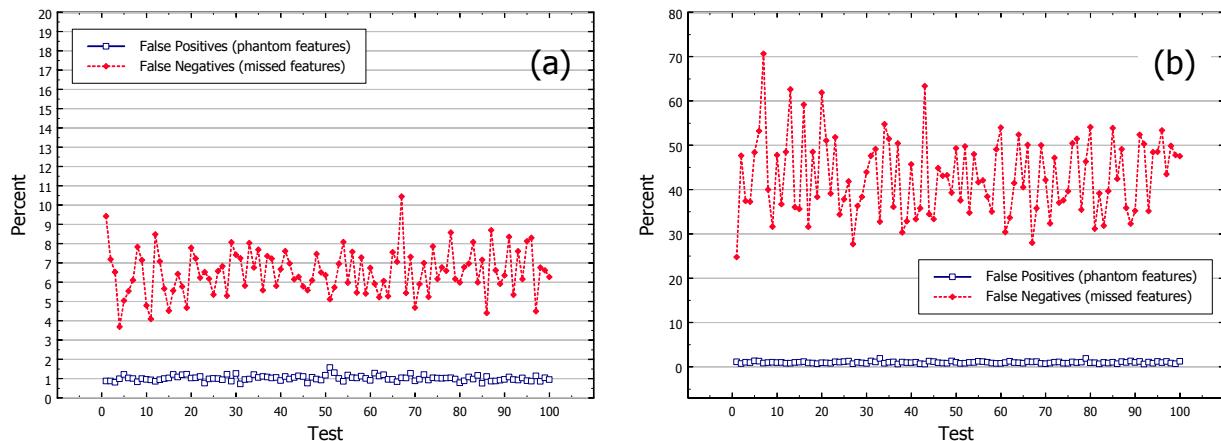
Figure 5.6: Composite results from the SIBYL stability test for all nighttime (panel (a)) and daytime (panel (b)) scenes. The false positive rate and the false negative (missed features) rate are seen to remain relatively steady throughout the entire sequence.

The results for all 100 scenes are shown in Figure 5.6. Panel (a) shows the nighttime results for both false positives (in blue) and false negatives (in red). The false negative rate computed over all nighttime scenes is 6.53% ± 1.40%. This translates into a 93% feature detection rate in a scene where the features are dominated by a relatively weak aerosol layer. (The aerosol back-

scatter coefficient is 5.0 x 10^{-4} km^{-1} sr^{-1}.) The mean nighttime false positive rate is 1.02% ± 0.14%. Taken together, these numbers show that SIBYL's performance during nighttime operations is accurate, stable, and reliable.

Panel (b) of Figure 5.6 shows the daytime results. As expected, the false negative rate for daytime detections is substantially higher than at night. However, the scene-to-scene variability, derived by dividing the standard deviation of the false negative rate by its mean, is almost identical for both the daytime and nighttime cases: 0.206 during the day versus 0.214 during the night. The false positive rate during daytime is 1.01% ± 0.22%; i.e., essentially identical to the nighttime value. Therefore we can conclude that while the accuracy of the detection scheme during daytime operations is degraded with respect to nighttime performance (due to poorer SNR caused by high levels of solar background radiation during daylight), the stability and reliability of the algorithm remain constant irrespective of lighting conditions.

5.4. Performance Examples

In this section we present one final example demonstrating of the performance of SIBYL. The simulated CALIOP data shown in Figure 5.7 was derived from measurements acquired during LITE orbits 54 and 129. This scene was deliberately designed as a stress test for the feature finder scheme. Feature placement (e.g., highly attenuating cirrus overlying optically thin aerosols), composition (e.g., fair weather cumulus embedded in aerosol), and horizontal scales were chosen so as to exercise all aspects of the SIBYL algorithm. The upper panel shows a time history of the data plotted at a 1-km horizontal resolution and full resolution vertically. The middle panel shows the feature boundaries reported by SIBYL. The lower panel provides the locations for missed features all and false positives. Overall, we judge SIBYL's performance on this scene to be excellent. The false negative rate is 12.2%, and the false positive rate is 1.9%. The false negatives occur predominately in regions where overlying features have significantly attenuated the backscatter signal; for example, see the region between 0-km and 200-km horizontally and 0-km 6-km vertically, where the combination of highly attenuating cirrus overlying a deep but weakly scattering aerosol layer results in a relatively large "missed features" area. The majority of the false positives are seen to be associated with faint and/or broken features whose boundaries intrude upon, but only partially fill, any given horizontal averaging region. Examples of this behavior occur in the very faint cirrus at ~12-km vertically and ~300-km horizontally, and in the broken cirrus clouds that occur at ~9-km vertically and 520-km horizontally.
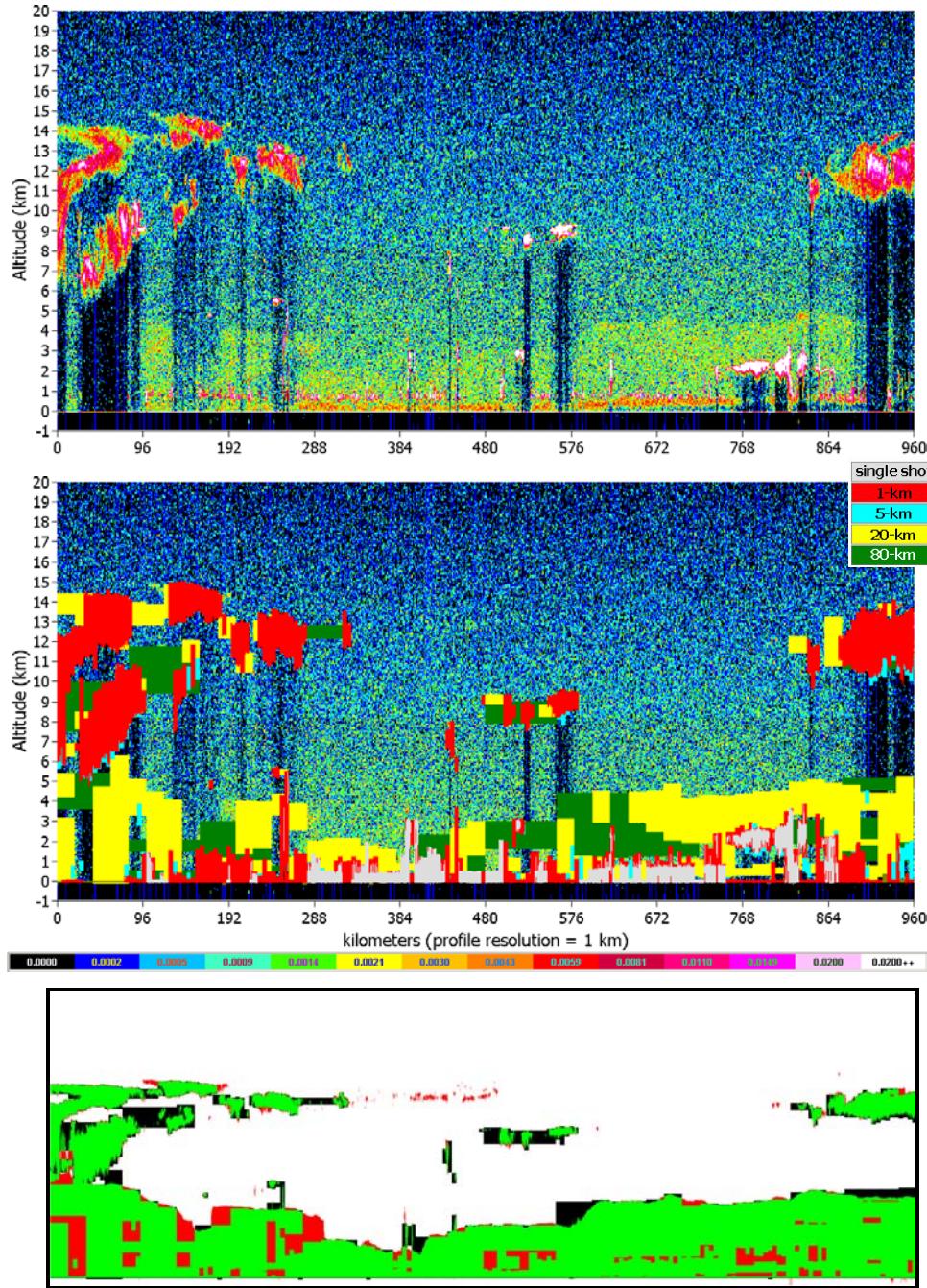
Figure 5.7: Images of simulated CALIOP data before (upper panel) and after (middle panel) being analyzed by SIBYL. The data in this scene was synthesized from retrievals performed on LITE measurements acquired during orbits 54 (clouds) and 129 (aerosols). In the middle panel, features detected at single shot resolution are rendered in pale gray; features detected at a 1-km horizontal resolution are rendered in red; 5-km features are shown in cyan; 20-km features are shown in yellow; and features detected at 80-km are shown in dark green. The lower panel presents an analysis of the feature finder results. The green regions indicate successful detection, black regions represent false positives, and red regions show missed features.

6. Computing Layer Descriptors

In this section we provide concise formulas for all of the layer descriptors that are computed by the feature finder algorithms and reported in the archived data products. Where appropriate, more detailed development is given elsewhere in this document. Note too that several of the "layer descriptors" are not scalar quantities, but instead contain an array of descriptive statistics calculated from the range-resolved backscatter data. We have defined a standard suite of statistics that consists of the minimum and maximum values within the feature, the sample mean value and standard deviation computed between feature top and feature base, and a dimensionless "skewness" parameter (Brandt, 1999). The general expressions used for computing estimates of the mean, standard deviation, centroid, and skewness are given below.

$$\text{mean: } \bar{x} = \frac{1}{N-1} \cdot \sum_{n=1}^{N} x_n \tag{6.1}$$

$$\text{standard deviation: } s = \sqrt{\frac{1}{N-1} \cdot \sum_{n=1}^{N} (x_n - \bar{x})^2} \tag{6.2}$$

$$\text{centroid: } C = \left(\frac{\sum_{n=1}^{N} x_n \cdot Z_n}{\sum_{n=1}^{N} x_n} \right) \tag{6.3}$$

$$m_3 = \frac{1}{N-1} \cdot \sum_{n=1}^{N} (x_n - \bar{x})^3$$

$$\text{skewness} = \frac{m_3}{s^3} \tag{6.4}$$

The summations indicated in equations (6.1) through (6.4) are taken over all data points in between and including feature top and feature base.

6.1. Layer base and top heights

Layer base and top heights are the most basic of all layer descriptors. Base and top heights are determined SIBYL's modification of the profile scanner (see Figure 4.16).

6.2. Renormalization region

The renormalization region is specified by the base and top altitudes of the region of "clear air" beneath a transmissive feature. For all transmissive layers SIBYL uses a search routine to determine a renormalization region. The two-way transmittance of the layer is computed in this region.

6.3. Horizontal averaging resolution

The amount of horizontal averaging that was required to detect the base and top altitudes of a given layer.

6.4. Attenuated backscatter statistics at 532 nm and 1064 nm

We characterize the feature backscatter intensity with our suite of descriptive statistics (equations (6.1) through (6.4)) from the *corrected total attenuated backscatter coefficients* at both 532 nm and 1064 nm. As shown in equation (6.5), the correction applied compensates for attenuation of the signal due to molecules and ozone. The magnitude of these effects at each range bin is computed from meteorological models.

$$
\mathsf{B}_{\lambda,k} = \frac{\beta'_\lambda(z_k)}{T^2_{m,\lambda}(z_k) \cdot T^2_{O_3,\lambda}(z_k)}
\tag{6.5}
$$

6.5. Integrated attenuated backscatter at 532 nm and 1064 nm

The calculations necessary to derive layer integrated attenuated backscatter from the 532 nm attenuated scattering ratios are covered in detail in Section 3.2.9.1. The procedure outlined there is equally applicable to the 1064 nm data. Alternately, the following calculations can be used to derive layer integrated attenuated backscatter from the (total) attenuated backscatter coefficients.

$$
g_\lambda = \frac{1}{2} \cdot \sum_{k=top+1}^{base} (z_{k-1} - z_k) \cdot \left(\mathsf{B}_{\lambda,k-1} + \mathsf{B}_{\lambda,k}\right)
$$

$$
\gamma'_{\lambda,feature} = g_\lambda - \left(\frac{1}{2} \cdot (z_{top} - z_{base}) \cdot \left(\mathsf{B}_{\lambda,top} + \mathsf{B}_{\lambda,base}\right)\right)
\tag{6.6}
$$

The uncertainty in $\gamma'_{\lambda,feature}$ is derived using standard statistical techniques (e.g., as in Bevington and Robinson (1992)), and is most readily developed using the alternate formulation given in equation (6.6):

$$
\Delta g_\lambda = \sqrt{\left(\frac{z_{top} - z_{base}}{2}\right)^2 \left(\Delta\mathsf{B}_{\lambda,top}^2 + \Delta\mathsf{B}_{\lambda,base}^2\right) + \frac{1}{4} \sum_{k=top+1}^{base} (z_{k-1} - z_k)^2 \left(\Delta\mathsf{B}_{\lambda,k-1}^2 + \Delta\mathsf{B}_{\lambda,k}^2\right)}.
\tag{6.7}
$$

Here $\Delta\mathsf{B}_{\lambda,k}$ represents the uncertainty in the B_λ value measured in the kth range bin, where

$$
\left(\frac{\Delta\mathsf{B}_{\lambda,k}}{\mathsf{B}_{\lambda,k}}\right)^2 = \left(\frac{\Delta\beta'_\lambda(z_k)}{\beta'_\lambda(z_k)}\right)^2 + \left(\frac{\Delta T^2_{m,\lambda}(z_k)}{T^2_{m,\lambda}(z_k)}\right)^2 + \left(\frac{\Delta T^2_{O_3,\lambda}(z_k)}{T^2_{O_3,\lambda}(z_k)}\right)^2.
\tag{6.8}
$$

6.6. Total (volume) depolarization statistics

Extinction coefficients and the associated two-way transmittances are not polarization-sensitive. Thus to obtain the total or *volume depolarization ratio profile* within a feature we simply divide the perpendicular channel attenuated backscatter coefficients by the corresponding parallel channel attenuated backscatter coefficients:

$$
\delta_v(z) = \frac{\beta'_{532,\perp}(z)}{\beta'_{532,\parallel}(z)}.
\tag{6.9}
$$

Having computed the volume depolarization ratio profile, we can then compute the suite of descriptive statistics described above.

6.7. Layer-integrated volume depolarization

Because the depolarization ratio is the quotient of two noisy measurements, the computed quantity has a greater uncertainty than does either of the constituent parts. As a consequence, a quantity of perhaps greater utility in the scene classification algorithms is the *feature integrated depolarization ratio*, defined as follows:

$$\delta_{\text{layer}} = \left(\left. \sum_{k=top}^{base} \beta'_{532,\perp}(z_k) \middle/ \sum_{k=top}^{base} \beta'_{532,\|}(z_k) \right. \right). \tag{6.10}$$

The relative error in δ_{layer} is derived using standard error propagation formulas (Bevington and Robinson, 1992):

$$\frac{\Delta\delta_{\text{layer}}}{\delta_{\text{layer}}} = \sqrt{\frac{\sum_{k=top}^{base} \Delta\beta'_{532,\perp}(z_k)^2}{\left(\sum_{k=top}^{base} \beta'_{532,\perp}(z_k)\right)^2} + \frac{\sum_{k=top}^{base} \Delta\beta'_{532,\|}(z_k)^2}{\left(\sum_{k=top}^{base} \beta'_{532,\|}(z_k)\right)^2}}. \tag{6.11}$$

6.8. Attenuated total color ratio statistics

To obtain the *attenuated total color ratio profile* within a feature we divide the 1064 nm corrected attenuated backscatter coefficients by the corresponding 532 nm corrected attenuated (total) backscatter coefficients:

$$\chi'(z) = \frac{B_{1064}(z)}{B_{532}(z)}. \tag{6.12}$$

Our standard suite of descriptive color ratio statistics is then computed using profiles of $\chi'(z)$.

6.9. Integrated attenuated total color ratio

The *feature integrated attenuated total color ratio* is analogous to the feature integrated depolarization ratio, and is computed as follows:

$$\chi'_{feature} = \left(\left. \sum_{k=top}^{base} B_{1064,k} \middle/ \sum_{k=top}^{base} B_{532,k} \right. \right). \tag{6.13}$$

Continuing the analogy to δ_{layer}, the relative error in $\chi'_{feature}$ is also derived using the same standard error propagation formulas:

$$\frac{\Delta\chi'_{feature}}{\chi'_{feature}} = \sqrt{\frac{\sum\limits_{k=top}^{base} B_{1064,k}^2}{\left(\sum\limits_{k=top}^{base} B_{1064,k}\right)^2} + \frac{\sum\limits_{k=top}^{base} B_{532,k}^2}{\left(\sum\limits_{k=top}^{base} B_{532,k}\right)^2}} \ . \tag{6.14}$$

6.10. Layer Two-way Transmittance and Error Estimate at 532 nm

The two-way transmittance of transmissive layers is computed by calculating the mean attenuated scattering ratio over the renormalization region.

$$\mathsf{T}_{layer}^2 = \left\langle R' \right\rangle_{\text{renormalization region}} = \left(\frac{1}{base-top}\right) \cdot \sum_{k=top}^{base} R'(z_k) \tag{6.15}$$

The error estimate is simply the standard deviation of the attenuated scattering ratio computed over the same region.

7. Practical Considerations

7.1. Application

Application of the SIBYL algorithm requires a contiguous segment of valid Level 1 attenuated backscatter profiles, together with the corresponding ancillary data, extending over a horizontal distance equal to the largest averaging interval specified in the main loop (i.e., with reference to Figure 4.1, the Nth interval). For the CALIPSO implementation of SIBYL, this interval is 80-km. 80-km segments containing invalid profile data or invalid ancillary information (e.g., baseline RMS values) will not be analyzed during the automated production processing.

7.2. Trouble Spots

SIBYL's spatial analysis is expected to be somewhat problematical in the following regions and/or seasons:

- polar troposphere – a review of the Geoscience Laser Altimeter System (GLAS; Spinhirne, 2002) data shows that this region frequently exhibits an extremely complex distribution of features; at present we have very little experience analyzing down-looking, multi-wavelength lidar measurements obtained in the polar troposphere; the closest available proxy is the LITE data from above 50º N and below 50º S

- polar stratosphere – given the current maximum horizontal averaging distance of 80-km, PSC identification will be confined to the more robust clouds; boundary location for faint PSCs will require additional off-line processing

- midlatitude & equatorial stratosphere – these regions will be relatively tractable if conditions remain quiescent; however, a major volcanic eruption could cause difficulties

7.3.   Identifying Surface Returns

Ultimately, the task of classifying atmospheric features as either clouds or aerosols is allocated to the scene classification algorithms. However, separating surface returns from atmospheric features is handled by the feature-finder algorithms. This section describes the algorithm used to do so.

Practical implementations of algorithms for identifying surface returns must be able to accommodate several situations. As illustrated in Figure 7.1, the surface spike in a profile can be obscured entirely by overlying layers, as in panel B. Much of the time, as in panels A and C, the surface spike has a magnitude much greater than the adjacent atmospheric return, allowing reliable detection. However, as illustrated in panel C, application of the standard threshold technique alone is frequently inadequate: both clouds and aerosols can be located immediately above or directly in contact with the Earth's surface. In these cases, surface detection requires an examination of the (potentially composite) as yet unclassified feature to determine if the whole or any part of it resembles a surface spike.
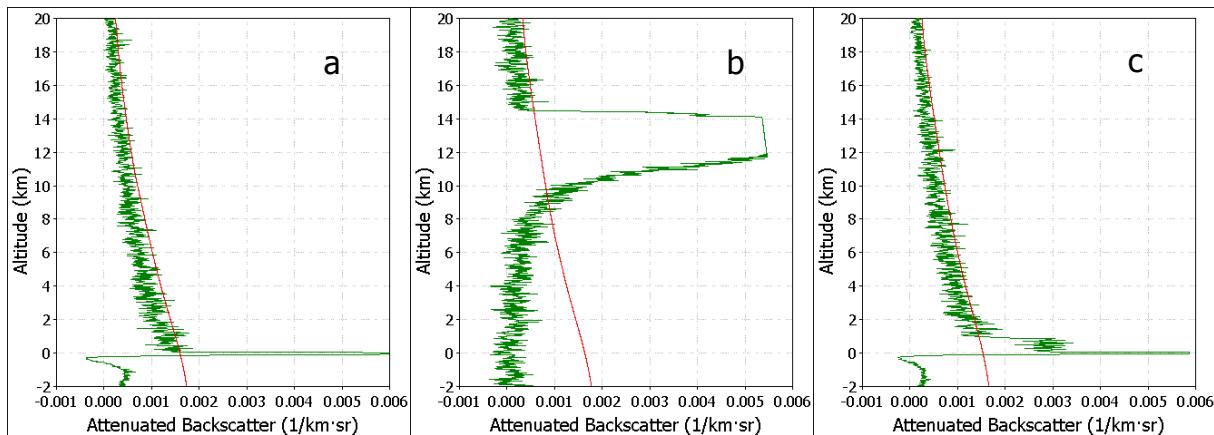

Figure 7.1: The three classes of surface returns: (a) a "pure" surface return, (b) no surface return detected, and (c) surface return in contact with an atmospheric feature. All profiles shown were acquired during LITE orbit 148, between 45° N and 35° N. The smooth red line in each image represents the initial profile scanner threshold that would be used in the search routine.

Identification of surface returns is accomplished using a high-resolution digital elevation map (DEM) and a model of the measured impulse response function of the lidar receiver. Because the intensity of any surface spike can be diminished due to attenuation by overlying features, the absolute magnitude of the surface signal is of relatively little importance. Instead we rely more on signal shape. Surface returns are defined as a significant and easily noticeable peak in the backscatter return. To be considered a valid surface return, this spike must occur in an altitude regime consistent with values obtained from the digital elevation map, and should have both a width characteristic of a surface return (as defined by the receiver's instrument response function) and a magnitude in excess of all other the signals in the selected neighborhood. The width of the search neighborhood is a location-dependent function of the uncertainties in satellite position and attitude (which introduce uncertainties in the laser footprint location), and the relative error associated with the DEM.

The flowchart shown in Figure 7.2 presents full details of the surface recognition algorithm. A sequence of tests is applied to lowest feature within a profile (i.e., F[min]) to determine which of the three categories of surface returns it best represents: surface return not found (e.g., Figure 7.1b), a "pure" surface return (e.g., Figure 7.1a), or a surface return in contact with a separate atmospheric feature (e.g., Figure 7.1c). If the lowest feature is classified as a composite – that is, as having atmospheric and surface components, as in category 3 – then the two components are separated and reported individually. As is standard with the CALIPSO algorithm architecture, the behavior of the surface detection algorithm is controlled by a set of programmable parameters that are supplied to the algorithm at runtime via external configuration files.
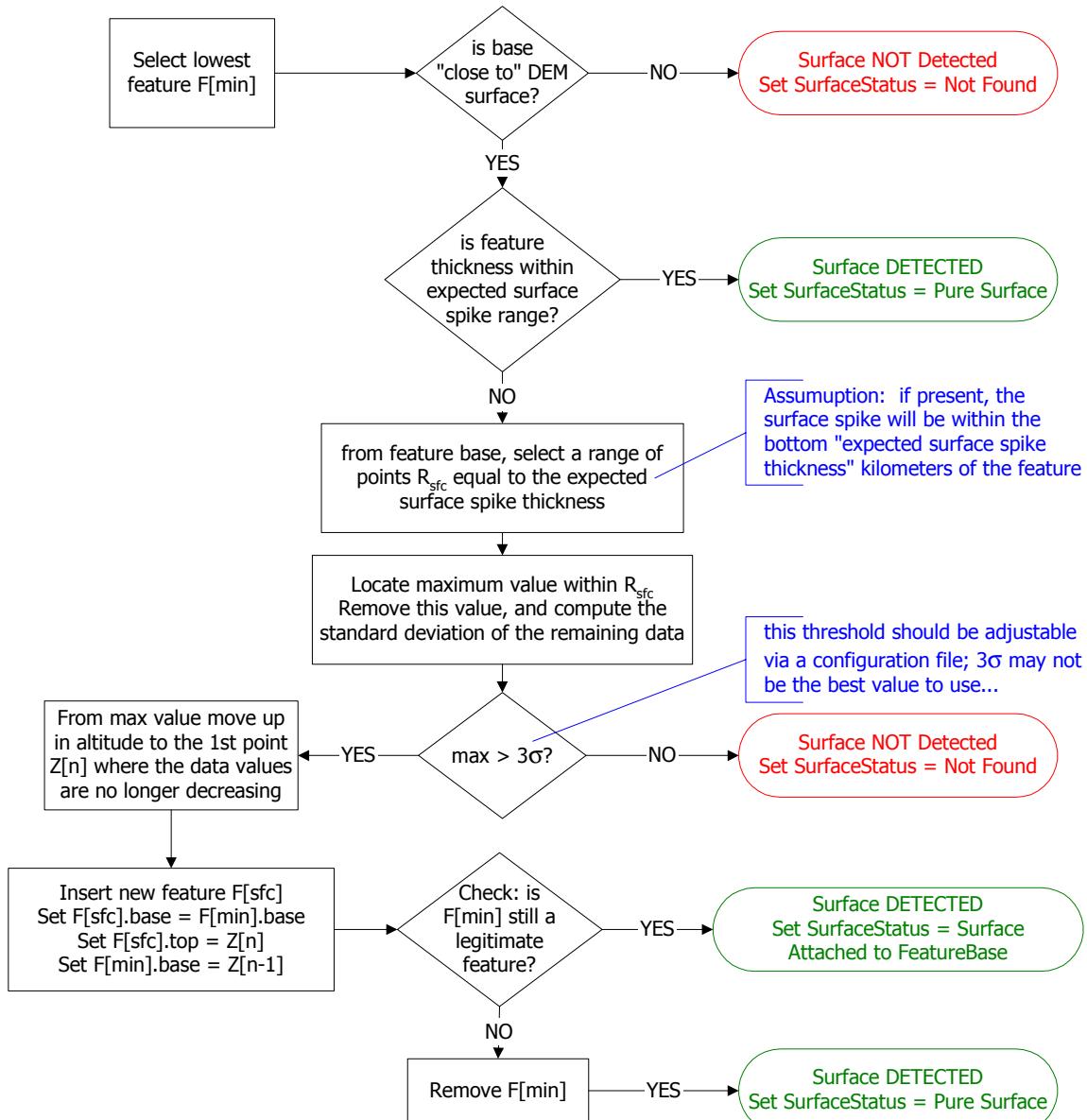

Figure 7.2: Surface spike identification procedure

7.4. Assessing Feature Opacity

Because the lidar signal can be completely attenuated by optically thick clouds, it is necessary to evaluate each base height to determine whether it indicates the true base or merely the point at which the lidar signal becomes completely attenuated. However, within the context of automated detection of features in backscatter lidar returns, the assessment of a cloud (or an aerosol layer) as being "opaque" is not necessarily straightforward. An object which appears opaque at one spatial resolution may be reclassified as transmissive upon further averaging. Opaque clouds filling only part of an averaging interval represent another complication.

Consider a sequence of unclassified features detected within a single lidar profile (e.g., as in Figure 1.1, Figure 3.2, Figure 3.13, and/or Figure 7.1). Clearly the lowest of the features within each profile are opaque, as no backscattered signal can be measured beneath them. Now consider Figure 7.3, which shows a profile acquired during LITE orbit 54. When analyzing this profile with the LITE profile scanner, two features are detected: an upper layer between 11.86-km and 13.26-km, and a lower layer between 7.48-km and 9.51-km. As no features are detected below 7.48-km, the lower layer appears to be opaque. However, in the context of the nested multi-grid averaging scheme used by SIBYL, it is possible that the surface (which, according to the DEM used by LITE, is at 0.3-km) could be detected during a subsequent search at 5-km, 20-km, or 80-km. That this is a realistic possibility – and, for this particular measurement, quite likely – can be seen by examining the upper panel of Figure 7.5, which presents a color modulated time history of the orbit 54 data from which Figure 7.3 was extracted. Features that initially appear opaque when profiles are searched at maximum spatial resolution can later be reclassified as transmissive upon subsequent analysis of the underlying data.
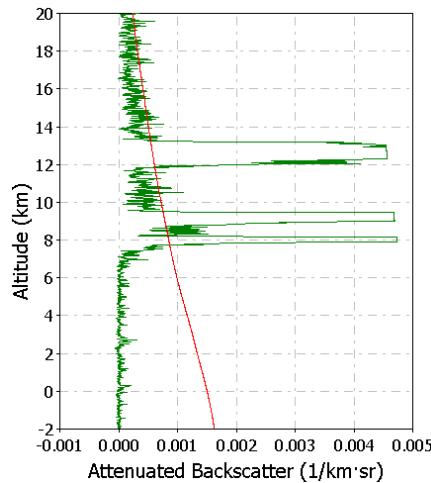
Figure 7.3: LITE profile from orbit 54 at 7.349° S, 52.846° W

To resolve the spatial resolution conundrum, the CALIPSO opacity flag is evaluated only for those features detected at horizontal averaging resolutions of 5-km or greater. For any 80-km segment of data, SIBYL will produce a uniform collection of profile descriptors: one set at 80-km, 4 sets at 20-km, and 16 sets at 5-km.[13] Data from the layer descriptors and profile

---

[13] We note that layer descriptors are also generated at 1-km and 1/3-km resolutions. However, these data are not germane to the current discussion, as the feature search for those resolutions is restricted to regions where features were previously identified at the 5-km resolution.

descriptors is combined to produce a time history of the "maximum penetration depth" (MPD). The MPD for any 5-km segment describes the lowest feature boundary detected within that segment during any of SIBYL's scans through the region; that is, from the fundamental averaging distance of 5-km out to the maximum averaging distance of 80-km. The definition of opacity is thus straightforward for features found at 5-km: if the MPD for that 5-km region is lower than the feature in question then that feature is transparent. At 20-km and 80-km, features are defined as transparent if 50% or more of the MPD values within the region are less than the feature base height (i.e., a lower feature was detected at least half of the time). The flowchart presented in Figure 7.4 gives additional details about the mechanics of the CALIPSO opacity algorithm. Figure 7.5 presents a conceptual diagram showing the application of the opacity algorithm to a segment of LITE data acquired during the nighttime portion of orbit 54.

Get Layer Descriptor Set → Select a complete set of SIBYL layer descriptors. For the currently defined averaging levels, a complete set includes all single shot, 1-km, 5-km, 20-km, and 80-km layer descriptors associated with a single 80-km horizontally averaged segment of data

Construct Maximum Penetration Profile → A "maximum penetration profile" is simply an array that describes the height of the lowest feature found at SIBYL's fundamental resolution (i.e., 5-km). For example, the max penetration profile (MPP) for the 5-20-80 averaging scheme will have 16 elements. The procedure for building this array is as follows:

[1]   fill all elements in the MPP with the base and top altitudes of the lowest feature found in the 80-km scan
[2]   selecting sequential sets of 4 MPP elements at a time, if the TOP of the lowest feature in the corresponding 20-km averaging lies beneath the current MPP top within the set, replace all set elements with the 20-km feature data
[3]   repeat step [2] working with individual MPP elements and the 5-km feature data

Assign 5-km Layer Opacities → For each 5-km set of layer descriptors, all layers found above the corresponding MPP element are transmissive.

Assign 20-km Layer Opacities

Assign 80-km Layer Opacities → Features averaged over more than SIBYL's fundamental averaging distance are defined as transmissive if 50% or more of the corresponding MPP elements (i.e., those that span the same horizontal distance) lie beneath the feature's base.

Figure 7.4: Layer opacity procedure
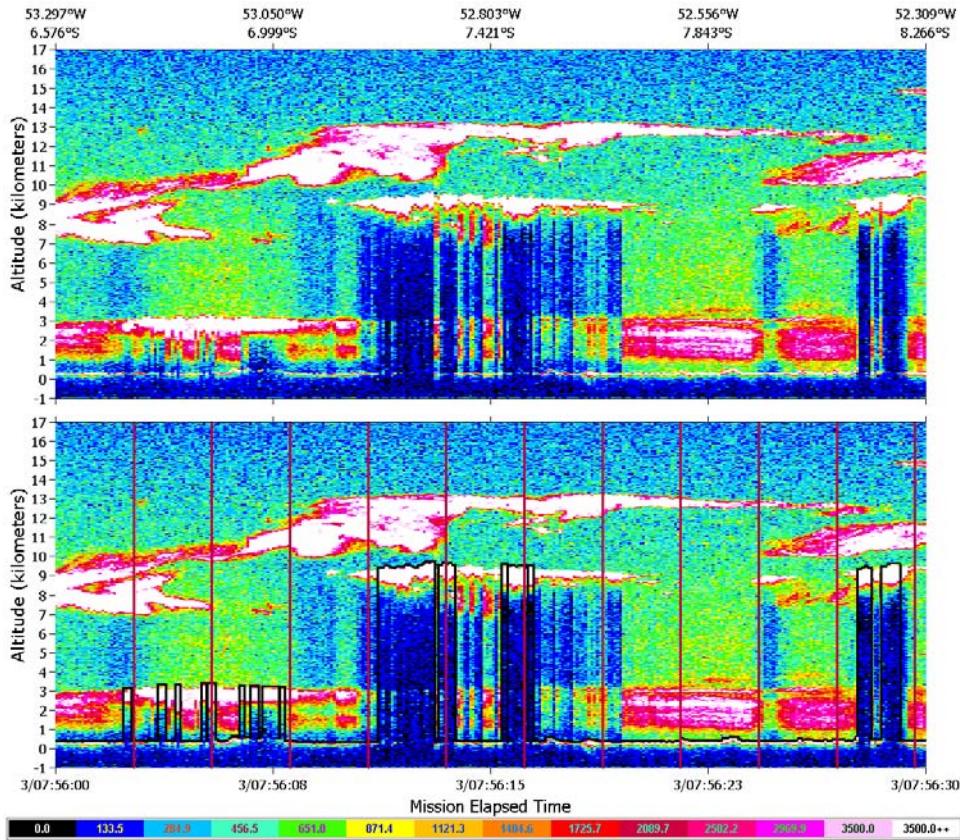
Figure 7.5: Conceptual diagram of the CALIPSO opacity algorithm applied to LITE data acquired during orbit 54. The upper panel shows the scene prior to application of the algorithm. In the lower panel, the black line demarcates the top of the "opaque layer". The vertical lines in the lower panel are drawn at 20-km intervals.

7.5. Proposal for Validation (preliminary suggestions)

Initial activities validation activities for the feature location algorithms will include:

- underflights of the orbit track using both up-looking and down-looking lidars

- statistical comparison of cloud heights with those obtained from ground stations (the CALIPSO data must be co-located and co-temporal to within 10-km)

- comparison of feature horizontal extent with satellite images (e.g., MODIS)

Additional detail can be found in the CALIPSO Science Validation Plan (PC-SCI-501).

7.6. Quality Control and Diagnostics

The general approach to quality control and diagnostic issues adopted within SIBYL is to examine the received backscatter signals and the subsequent feature-finder outputs to see whether they behave in a manner that is consistent with what one would expect from valid measurements. Any of the following conditions can indicate the possibility of an error:

- No features of any kind are found within a complete profile. At minimum, either a surface return or a single atmospheric feature should be detected for every laser pulse.

- The integrated attenuated backscatter summed over all features within a column/profile lies above the maximum possible value. In opaque layers, γ' asymptotes to a value of $1/2\bar{\eta}S$, which typically is less than 0.5. (S and η vary, depending on the characteristics of the layer and the lidar sensing geometry.) Values greater than the maximum asymptotic value indicate an error condition. The estimate of the maximum value will be revised after launch based on on-orbit data.

- Within any 80-km data segment passed to SIBYL, the mean molecular signal between 30-km and the top of the first feature must have (a) a mean attenuated scattering ratio close to 1 and (b) a positive correlation coefficient with respect to the molecular attenuated backscatter model. (Due to the high background levels present during daytime, this test might only be useful for nighttime data.)

- Within all features having a mean attenuated scattering ratio or 5 or greater, the 532 and 1064 signals should have a positive correlation coefficient.

The sections below discuss several techniques for implementing these quality checks.

7.6.1. Checking Feature-Finder Inputs

To retrieve accurate estimates of feature boundaries, the profile scanning algorithm depends on the fidelity of the meteorological data and on the accuracy of the calibration constants retrieved in the Level 1 analyses. The 532 nm calibration procedure guarantees that the mean attenuated scattering ratio computed over the calibration region (nominally 30-km to 34-km) will be equal to 1.0. The tests performed in this section will assess the linearity of the calibration with respect to altitude by analyzing clear air regions beneath the calibration regions. Two altitude regions will be tested, as follows:

(a) For all 80-km segments of nighttime data for which the highest feature top is less than 18-km, compute

  - the mean and standard deviation of the 532 nm attenuated scattering ratios; and

  - the correlation coefficient for the measured attenuated backscatter at 532 nm with respect to the 532 nm clear air model delivered by the CALIPSO Met Manager.

All quantities should be calculated between 18-km and 22-km vertically, and at a horizontal resolution of 80-km.

(b) For all 80-km segments of nighttime data with a maximum feature top less than 6-km, compute the statistics specified above (mean and standard deviation of $R'_{532}(z)$, and the correlation coefficient with respect to the "clear air" model) between 6-km and 9-km vertically at a horizontal resolution of 80-km for both the 532 nm data and the 1064 nm data.

Totals for all values calculated above are accumulated for each nighttime granule. These numbers are then used to generate the following quality assurance plots:

- histograms showing the distribution of mean attenuated scattering ratios computed in each altitude regime (i.e., 6-km to 9-km and 18-km to 22-km); and

- time series plots of mean attenuated scattering ratios and correlation coefficients for each altitude regime; error bars derived from the corresponding standard deviation calculation should be included for the mean attenuated scattering ratio plots.

Tables of summary statistics should also be produced for each granule. These tables should include the median, mean, and standard deviation of each set of histogram data, and the slope (with respect to time) of the mean attenuated scattering ratios and correlation coefficients. Assuming the SIBYL algorithm is performing effectively, the median values of the R'_{532} histogram should both be close to 1.0. Significant variation from 1.0 can mean either that the molecular and/or ozone number density profiles are in error, or that the feature finding algorithm is failing to identify high-altitude features.

7.6.2. Checking Feature-Finder Performance

To monitor the performance of SIBYL, we rely on statistical assessments of the features found within each orbit. For each 5-km column, we

(a) determine the number of cloud layers, the number of aerosol layers, and the total number of atmospheric features (i.e., the number of clouds plus the number of aerosols); and

(b) compute the sum of the feature geometric thicknesses found within the column for all features and again separately for both clouds and aerosols.

Table 7.1: Number of cloud layers per 7.4-km column derived from LITE measurements

Layers	Number	Percent
0	22972	18.47
1	58528	47.05
2	28851	23.19
3	10288	8.27
4	2855	2.30
5	699	0.56
6	147	0.12
7	38	0.03
8	8	0.01
9	1	0.00

The accumulated totals over any given time period should remain relatively stable and should closely resemble real-world distributions acquired by other instruments (e.g., LITE and GLAS). Using statistics derived from LITE measurements we can develop an initial expectation regarding the distributions of the number of features found within each profile (see Table 7.1) and the total geometric thickness of those features (see Figure 7.6). Similar statistics can also be derived from the GLAS data.
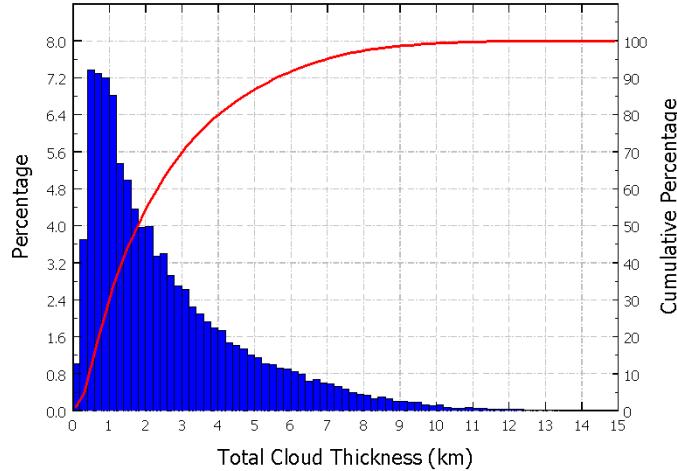
Figure 7.6: Total cloud geometric thickness per 7.4-km column derived from LITE measurements

7.7. Browse Products

The Level 1 lidar browse products consist of color-modulated time histories of attenuated backscatter coefficients for the three measurement channels: 532 nm parallel, 532 nm perpendicular, and 1064 nm total. For Level 2, these browse products are augmented by images of the vertical feature mask (VFM) data product, generated on time scales identical to those used in the Level 1 images. The VFM, which combines the information derived by SIBYL and the scene classification algorithms, contains one 16-bit integer for each lidar altitude resolution element in the downlinked (i.e., Level 0) data stream. Each of these integers is interpreted as bit-mapped set of scene classification flags. VFM images thus provide comprehensive pictures of the SIBYL results and the subsequent feature classifications at the highest possible spatial resolution. The descriptive information contained within these scene classification flags consists of Boolean-type classifications (e.g., a yes-no answer to a specific query) and simple categorizations (e.g., cloud, aerosol, "clear air") and/or assessments (e.g., poor, fair, good, excellent). Images representative of both the Level 1 browse products and the Level 2 VFM products are displayed in Figure 7.7. The VFM image shows the boundaries for all atmospheric features, color-coded according to feature type. Regions that do not contain features are also classified as being clear air, surface returns, subsurface data, or totally attenuated data.
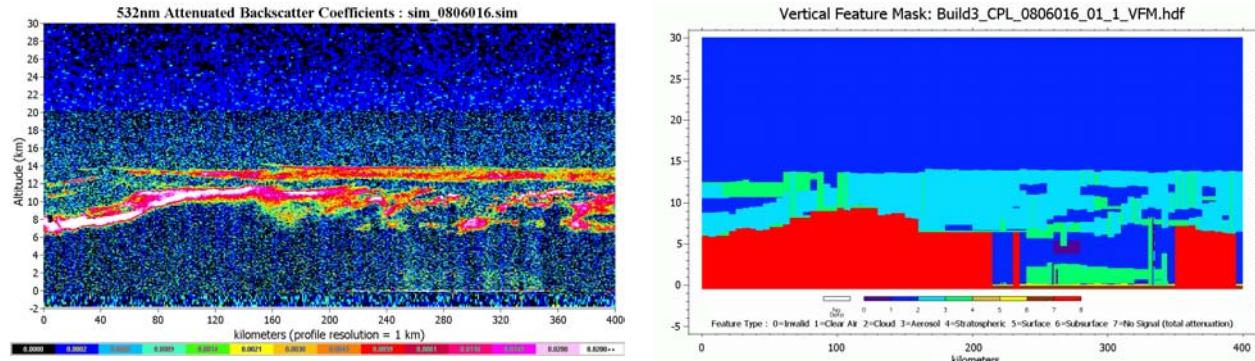

Figure 7.7: (a) Level 1 browse images; (b) Level 2 vertical feature mask image showing feature boundaries and feature type

Additional information describing the content and format of the VFM can be found in Part 6 of the CALIOP ATBD (see PC-SCI-202, Part 5).

---

8. References

Anderson, T. L, Masonis, S. J., Covert, D. S., Charlson, R. J., and Rood, M. J., 2000: "In situ measurement of the aerosol extinction-to-backscatter ratio at a polluted continental site", *J Geophys Res*, **105**, pp. 26907-26915.

Bevington, P. R and D. K. Robinson, 1992: <u>Data Reduction and Error Analysis for the Physical Sciences</u>, McGraw-Hill, 328 pp.

Beyerle, G., M. R. Gross, D. A. Haner, N. T. Kjome, I. S. McDermid, T. J. McGee, J. M. Rosen, H.-J. Schäfer, and O. Schrems, 2001: "A Lidar and Backscatter Sonde Measurement Campaign at Table Mountain during February–March 1997: Observations of Cirrus Clouds", *J Atmos Sci*, **58**, pp. 1275-1287.

Brandt, S., 1999: <u>Data Analysis</u>, Springer-Verlag, 652 pp.

Clothiaux, E. E., G. G. Mace, T. P. Ackerman, T. J. Kane, J. D. Spinhirne, and V. S. Scott, 1998: "An automated algorithm for detection of hydrometer returns in micropulse lidar data", *J Atmos Ocean Tech*, **15**, pp 1035-1042.

Davis, K. J., Gamage, N., Hagelberg, C.R., Kiemle, C., Lenschow, D. H., Sullivan, P. P., 2000: "An Objective Method for Deriving Atmospheric Structure from Airborne Lidar Observations", *J Atmos Ocean Tech*, **17**, pp. 1455-146.

Eloranta, E. W., R. E. Kuehn, and R. E. Holz, 2001: "Measurements of Backscatter Phase Function and Depolarization in Cirrus Clouds made with the University of Wisconsin High Spectral Resolution Lidar", in *Advances in Laser Remote Sensing: Selected Papers presented at the 20th International Laser Radar Conference (ILRC)*, A. Dabas, C. Loth, and J. Pelon (editors), ISBN 2-3702-0798-8, pp. 255-257.

Fernald, F. G., B. J. Herman, and J. A. Reagan, 1972: "Determination of Aerosol Height Distributions by Lidar" , *J Appl Meteorol*, **11**, pp. 482-489.

Galbraith, A., 1996: <u>Cirrus Cloud Detection with a Micro Pulse Lidar</u>, Masters Thesis, Department of Electrical and Computer Engineering, University of Arizona, 75 pp.

Gonzalez, R. C. and P. Wintz, 1977: <u>Digital Image Processing</u>, Addison-Wesley, 377 pp.

Kingston, R. H., 1978: <u>Detection of Optical and Infrared Radiation</u>, Springer-Verlag, 140 pp.

Liu, Z. and N. Sugimoto, 2002: "Simulation study for cloud detection with space lidar using analog detection photomultiplier tubes", *Appl Optics*, **41**, pp 1750-1759.

Masonis, S. J., 2001: <u>An empirical study of the lidar ratio and its variability, with implications for determining climate forcing by satellite-borne lidar</u>, Ph.D. dissertation, University of Washington, 257 pp. [Available from Bell and Howell Information and Learning, 300 North Zeeb Road., P.O. Box 1346, Ann Arbor, MI 48106-1346]

Melfi, S. H., J. D. Spinhirne, S-H. Chou, and S. P. Palm, 1985: "Lidar Observations of Vertically Organized Convection in the Planetary Boundary Layer over the Ocean", *J Clim Appl Meteorol*, **24**, pp. 806-821.

Mishchenko, M., J. Penner, and D. Anderson, 2002: "Global Aerosol Climatology Project", *J Atmos Sci*, **59**, pp. 249-249.

Pal, S. R., W. Steinbrecht, and A. I Carswell, 1992: "Automated method for lidar determination of cloud base height and vertical extent", *Appl Optics*, **31**, pp. 1488-1494.

Pinnick, R. G., S. G. Jennings, P. Chylek, C. Ham, and W. T. Grandy, Jr., 1983: "Backscatter and Extinction in Water Clouds", *J Geophys Res*, **88**, pp. 6787-6796.

Platt, C. M. R., 1973: "Lidar and Radiometric Observations of Cirrus Clouds", *J Atmos Sci*, **30**, pp. 1191-1204.

Platt, C. M. R., S. A. Young, A. I. Carswell, S. R. Pal, M. P. McCormick, D. Winker, M. D. Guasta, L. Stefanutti, W. L. Eberhard, R. M. Hardesty, P. H. Flamant, R. Valentin, B. Forgan, G. G. Gimmestad, H. Jager, S. S. Khmelevtsov, I. Kolev, B. Kaprielov, D.-R. Lu, K. Sassen, V. S. Shamanaev, O. Uchino, Y. Mizuno, U. Wandinger, C. Weitkamp, A. Ansmann, and C. Wooldridge, 1994: "The Experimental Cloud Lidar Pilot Study (ECLIPS) for cloud-radiation research", *B Am Meteorol Soc*, **75**, pp. 1634-1654.

Platt C. M. R., D. M. Winker, M. A. Vaughan, and S. D. Miller, 1999: "Backscatter-to-extinction ratios in the top layer of tropical mesoscale convective systems and in isolated cirrus from LITE observations", *J Appl Meteorol*, **38**, pp. 1330-1345.

Platt, C. M. R., S. A. Young, R. T. Austin, G. R. Patterson, D. L. Mitchell, and S. D. Miller, 2002: "LIRAD Observations of Tropical Cirrus Clouds in MCTEX. Part I: Optical Properties and Detection of Small Particles in Cold Cirrus", *J Atmos Sci*, **59**, pp. 3145-3162.

Powell, K. A., W. H. Hunt, and D. M. Winker, 2002: "Simulations of CALIPSO Lidar Data", in *Lidar Remote Sensing in Atmospheric and Earth Sciences*, L. R. Bissonnette, G. Roy and G. Vallée (editors), Defence R&D Canada - Valcartier, Québec, Canada, pp. 781-784.

Sassen, K., M. K. Griffin, and G. C. Dodd, 1989: "Optical scattering and microphysical properties of subvisual cirrus clouds and climatic implications", *J Appl Meteorol*, **28**, pp. 91-98.

Sassen, K. and J. Comstock, 2001: "A Midlatitude Cirrus Cloud Climatology from the Facility for Atmospheric Remote Sensing. Part III: Radiative Properties", *J Atmos Sci*, **58**, pp. 2113-2127.

Spinhirne, J. D., 2002: "Atmospheric Measurements by the 2002 Geoscience Laser Altimeter System Mission", in *Lidar Remote Sensing in Atmospheric and Earth Sciences*, L. R. Bissonnette, G. Roy and G. Vallée (editors), Defence R&D Canada - Valcartier, Québec, Canada, pp. 727-730.

Suk, M. and S. Hong, 1984: "An Edge Extraction Technique for Noisy Images", *Comput Vision Graph*, **25**, pp 24-45.

Teich, M. C., 1981: "Role of the doubly stochastic Neyman type-A and Thomas counting distributions in photon detection", *Appl Optics*, **20**, pp 2457-2467.

Wang, Z. and K. Sassen, 2001: "Cloud Type and Macrophysical Property Retrieval Using Multiple Remote Sensors", *J Appl Meteorol*, **40**, pp. 1665-1683.

Wiegner, M., W. Kumpf, V. Freudenthaler, I. Stachlewska, B. Heese, 2002: "Surface Aerosol Layer: Annual Cycle and Parameterization", in *Lidar Remote Sensing in Atmospheric and Earth Sciences*, L. R. Bissonnette, G. Roy and G. Vallée (editors), Defence R&D Canada - Valcartier, Québec, Canada, pp. 321-324.

Winker, D. and M. Vaughan, 1994: "Vertical distribution of clouds over Hampton, Virginia observed by lidar under the ECLIPS and FIRE ETO programs", *Atmos Res*, **34**, pp. 17-133.

Winker, D. M., R. Couch, and M. P. McCormick, 1996: "An overview of LITE: NASA's Lidar In-space Technology Experiment", *Proc IEEE*, **84**, pp. 164-180.

Uthe, E. E., and P. R. Russell, 1977: "Lidar observations of tropical high-altitude cirrus clouds", International Symposium on Radiation in the Atmosphere, Garmisch-Partenkirchen, W. Germany, Aug. 19-28, 1976; Proceedings, Princeton, Science Press, pp. 242-244.

Uttal, T., L. I. Church, B. E. Martner, and J. S. Gibson, 1993: CLDSTATS: A Cloud Boundary Detection Algorithm for Vertically Pointing Radar Data, NOAA Technical Memorandum ERL WPL-233.