Automated Retrieval of Cloud and Aerosol Properties from the
ARM Raman Lidar. Part II: Extinction

TYLER J. THORSEN
Department of Atmospheric Sciences, University of Washington, Seattle, Washington

QIANG FU
Department of Atmospheric Sciences, University of Washington, Seattle, Washington, and College of Atmospheric Sciences, Lanzhou University, Lanzhou, China

(Manuscript received 15 September 2014, in final form 12 August 2015)

ABSTRACT

A feature detection and extinction retrieval (FEX) algorithm for the Atmospheric Radiation Measurement Program’s (ARM) Raman lidar (RL) has been developed. Presented here is Part II of the FEX algorithm: the retrieval of cloud and aerosol extinction profiles. The directly retrieved extinction profiles using the Raman method are supplemented by other retrieval methods developed for elastic backscatter lidars. Portions of features where the extinction-to-backscatter ratios (i.e., lidar ratios) can be obtained are used to infer the lidar ratios for the regions where no such estimate can be made. When neither directly retrieved nor an inferred value can be determined, a climatological lidar ratio is used. This best-estimate approach results in the need to use climatological lidar ratios for less than about 5% of features, except for thin cirrus at the ARM tropical western Pacific Darwin site, where above 12 km, about 20% of clouds use a climatological lidar ratio. A classification of feature type is made, guided by the atmosphere’s thermodynamic state and the feature’s scattering properties: lidar ratio, backscatter, and depolarization. The contribution of multiple scattering is explicitly considered for each of the ARM RL channels. A comparison between aerosol optical depth from FEX and that from collocated sun photometers over multiple years at two ARM sites shows an agreement (in terms of bias error) of about −0.3% to −4.3% (relative to the sun photometer).

1. Introduction

An essential prerequisite for deriving atmospheric radiative heating profiles is an accurate knowledge of cloud and aerosol (particulate) extinction. From a remote sensing prospective, lidars have the capability to provide unparalleled range-resolved observations of particulate extinction. However, lidars fundamentally measure backscattered energy—not extinction—and for widely prevalent single-channel elastic backscatter lidars (e.g., Campbell et al. 2002; Winker et al. 2010), extinction must be obtained by assuming the ratio of particulate extinction to backscatter (i.e., the lidar ratio; Klett 1981; Fernald 1984). High-spectral-resolution lidars (HSRL; e.g., Hair et al. 2008; Grund and Eloranta 1991) and Raman lidars (RL; e.g., Goldsmith et al. 1998; Matthais et al. 2004) are more advanced lidars that can intrinsically separate signals from molecules and particles, allowing for directly measured particulate backscatter and extinction coefficients. These advanced lidars also provide direct retrievals of the lidar ratio—an indicator of the target’s microphysical properties (e.g., Burton et al. 2014) and a critical input for elastic backscatter lidar retrievals.

The work here centers around developing an automated extinction retrieval algorithm for the Atmospheric Radiation Measurement Program’s (ARM; Ackerman and Stokes 2003) Raman lidars (Goldsmith et al. 1998) that have operated at the ARM Southern Great Plains (SGP) site near Lamont, Oklahoma (36.61°N, 97.49°W); the Darwin, Australia, tropical western Pacific (TWP) site (12.43°S, 130.89°E); and as part of the third ARM Mobile Facility (AMF3) currently stationed in Oliktok Point (OLI), Alaska (70.50°N, 149.89°W). The TWP RL will soon be moved to the eastern North Atlantic (ENA).
site on Graciosa Island in the Azores (39.09°N, 28.03°W). The ARM RL and its automated algorithms were originally designed to measure/retrieve water vapor and aerosol properties (Turner et al. 2002). However, recent studies have shown the RL to be capable of producing high-quality observations of clouds as well (Wang and Sassen 2002; Dupont et al. 2011; Thorsen et al. 2013). Therefore, the main goal of this series of papers is to develop a new automated algorithm for feature detection and extinction retrieval (FEX), in order to help fully realize the potential of the ARM RL. The FEX algorithm objectively identifies features (i.e., clouds and aerosols) and retrieves their extinction and backscatter profiles over the full extent of the troposphere. Complete details of feature detection are given in Thorsen et al. (2015, hereafter Part I), while Part II presented here focuses on the retrieval of particulate extinction. The intent is to run FEX operationally within the ARM Data Management Facility (DMF), with the output being made available to the general user community through the ARM website (http://www.arm.gov/).

A description of the ARM RL system and a review of methods for inverting the elastic and Raman lidar equations are given in sections 2 and 3, respectively. The retrieval methodology is given in section 4, including a summary of Part I, since feature detection and retrieving extinction are intertwined (to fully comprehend the content presented in this paper, we recommend the reader review Part I in its entirety). Section 4 describes how extinction is retrieved by FEX both directly using the Raman method and by other methods developed for elastic backscatter lidars in order to obtain the best possible extinction estimate for all detected features. In support of accurate extinction retrieval, a classification of feature type is made and multiple-scattering effects are explicitly considered. In section 5 multiple years of data at both the SGP and TWP sites are analyzed. Presented are the frequency of feature types at these sites and the frequency with which different types of retrievals are performed. The need to correct for multiple-scattering effects is justified a posteriori by presenting the errors introduced by ignoring its effects. The retrieval of aerosol optical depth is validated against collocated sun photometers. Finally, a summary and conclusions are given in section 6.

2. The ARM Raman lidar

Table 1 lists the characteristics of the ARM RL and the detection channels used in this work. The basic design is given in Goldsmith et al. (1998), although the original system has since evolved through various upgrades and modifications (Ferrare et al. 2006; Newsom et al. 2010; Eichmann et al. 2011; Thorsen et al. 2013). The RL at the ARM SGP site has been in near-continuous operation since 1998. Additional ARM RLs were deployed at the Darwin TWP site in December 2010 and at the AMF3 OLI site in October 2014, both with nearly the same design as the SGP RL. While the ARM RL contains temperature (Newsom et al. 2013) and water vapor (Turner et al. 2002) channels, only the elastic and nitrogen channels are used for this work. Backscattered returns are collected using both a narrow field of view (FOV; referred to as the “high channels”) and a wide FOV (referred to as the “low channels”). Throughout this paper the prefix “high” is dropped when referring to the high channel, while the prefix “low” will be included when referring to low-channel signals.

### Table 1. Specifications of the ARM RL transmitter and receiver channels used in this study.

<table>
<thead>
<tr>
<th>Laser wavelength</th>
<th>355 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receiver FOV (narrow)</td>
<td>0.3 mrad</td>
</tr>
<tr>
<td>Receiver FOV (wide)</td>
<td>2 mrad</td>
</tr>
<tr>
<td>Data acquisition</td>
<td>Simultaneous</td>
</tr>
<tr>
<td>Pulse repetition frequency</td>
<td>30 Hz</td>
</tr>
<tr>
<td>Beam divergence</td>
<td>0.1 mrad</td>
</tr>
<tr>
<td>Telescope</td>
<td>61 cm</td>
</tr>
<tr>
<td>Pulse accumulation time</td>
<td>10 s</td>
</tr>
<tr>
<td>Range resolution</td>
<td>7.5 m</td>
</tr>
<tr>
<td>Channels (0.3-nm bandpass filters)</td>
<td></td>
</tr>
<tr>
<td>High elastic parallel (S_{\lambda_0}), copolarized, narrow FOV</td>
<td>355 nm</td>
</tr>
<tr>
<td>High elastic perpendicular (S_{\lambda_0}), cross polarized, narrow FOV</td>
<td>355 nm</td>
</tr>
<tr>
<td>High nitrogen (S_{\lambda_0}), narrow FOV</td>
<td>387 nm</td>
</tr>
<tr>
<td>Low elastic (S_{\lambda_0}^L), wide FOV</td>
<td>355 nm</td>
</tr>
<tr>
<td>Low nitrogen (S_{\lambda_0}^N), wide FOV</td>
<td>387 nm</td>
</tr>
</tbody>
</table>

3. Inversion of the elastic and Raman lidar equations

The zenith-pointing lidar equation for the background-subtracted signal received in the elastic channel (at the laser wavelength \(\lambda_0\)) can be written as (Measures 1984)

\[
S_{\lambda_0}(z) = N_{\lambda_0} \frac{S_{\lambda_0}}{z^2} D_{\lambda_0}(z) F_{\lambda_0}(z) \beta_{m\lambda_0}(z) + \beta_{p\lambda_0}(z) T_{m\lambda_0}(0, z) T_{p\lambda_0}(0, z),
\]

where \(N_{\lambda_0}\) is the number of photons received from height \(z\), \(D_{\lambda_0}\) is the system constant incorporating all the lidar parameters that describe the characteristics of the optics and detectors, \(S_{\lambda_0}\) is the number of transmitted
photons, and \( O(z) \) is the overlap function that describes the overlap between the laser beam and receiver FOV. Terms \( \beta_{m, \lambda_0}(z) \) and \( \beta_{p, \lambda_0}(z) \) are the molecular and particulate backscatter coefficients, respectively. Terms \( T_{m, \lambda_0}(0, z) \) and \( T_{p, \lambda_0}(0, z) \) are the transmission due to molecules and particulates,

\[
T_{m, \lambda_0}(0, z) = \exp \left[ - \int_0^z \alpha_{m, \lambda_0}(z') \, dz' \right] \tag{2}
\]

and

\[
T_{p, \lambda_0}(0, z) = \exp \left[ - \int_0^z \alpha_{p, \lambda_0}(z') \, dz' \right], \tag{3}
\]

where \( \alpha_{m, \lambda_0}(z) \) and \( \alpha_{p, \lambda_0}(z) \) are the molecular and particulate extinction coefficients, respectively. The quantity

\[
\beta_{p, \lambda_0}(z) = \frac{S_{p, \lambda_0}(z) \exp \left\{ -2 \int_{z_0}^z \left[ K_{p, \lambda_0}(z') + K_m \beta_{m, \lambda_0}(z') \right] \, dz' \right\}}{\beta_{p, \lambda_0}(z_0) + \beta_{m, \lambda_0}(z_0) - 2 \int_{z_0}^z K_{p, \lambda_0}(z') S_{p, \lambda_0}(z') \exp \left\{ -2 \int_{z_0}^{z'} \left[ K_p(z'') + K_m \beta_{m, \lambda_0}(z'') \right] \, dz'' \right\} \, dz'} \tag{5}
\]

where \( K_p(z) \) is the lidar ratio, that is, the ratio of particulate extinction to particulate backscatter

\[
K_{p, \lambda_0}(z) = \frac{\alpha_{p, \lambda_0}(z)}{\beta_{p, \lambda_0}(z)}. \tag{6}
\]

The \( K_m \) is the ratio of molecular extinction to its backscatter calculated following Bucholtz (1995), and \( z_0 \) is the height of the boundary condition. The single-scattering signal \( S_{p, \lambda_0}(z) \) is related to the measured signal by Eq. (4). Equation (5) illustrates the problem with all solutions to the elastic lidar equation: there are two unknowns, the particulate extinction and the backscatter coefficient, but only one equation.

This significant limitation of elastic backscatter lidars is alleviated by the use of either an HSRL or a Raman lidar: both measure a signal independent of the particulate backscatter. For the signal in the Raman nitrogen channel, which only contains Raman backscatter due to nitrogen molecules at the shifted wavelength \( \lambda_{N_2} \), the lidar equation can be written as

\[
S_{N_2}(z) = \frac{\sigma N_{N_2} \lambda_{N_2}^2}{2} O_{N_2}(z) F_{N_2}(z) \beta_{N_2, \lambda_{N_2}}(z) T_{m, \lambda_0}(0, z) \times T_{p, \lambda_0}(0, z) T_{m, \lambda_{N_2}}(0, z) T_{p, \lambda_{N_2}}(0, z). \tag{7}
\]

Compared to Eq. (1), extra transmission terms are needed to account for the different wavelength of the

\[
\beta_{p, \lambda_0}(z) = [\text{SR}_{E, N_2}(z) - 1] \beta_{m, \lambda_0}(z), \tag{10}
\]

where \( \text{SR}_{E, N_2}(z) \) is a scattering ratio (Part I); that is, the ratio of total (particulate and molecular) backscattering to just molecular backscattering at \( \lambda_0 \), derived from both the elastic and nitrogen channels,
Equation (11) is obtained by combining Eqs. (1) and (7) (Cooney et al. 1969; Melfi 1972; Part I) along with using Eqs. (4) and (9) to account for multiple scattering. The \( C_{E,N_2} \) is a calibration constant. Via Eqs. (8) and (10), a Raman lidar can directly retrieve profiles of the particulate extinction and backscatter—we refer to this as the Raman method. However, the application of the Raman method is limited by the relatively weak strength of Raman scattering relative to elastic scattering. Noise is particularly an issue for the extinction retrieval since in Eq. (8) the slope of the signal is needed, which greatly amplifies the noise.

Using only the elastic channel signal, a layer-average lidar ratio can be determined using the transmission-loss method (e.g., Fernald et al. 1972; Young 1995; Spinhirne et al. 1996; Comstock and Sassen 2001; Chen et al. 2002). By taking the ratio of the elastic signal [Eq. (1)] in clear sky [i.e., \( \beta_{p,\lambda_0}(z) = 0 \)] above a feature layer (at \( z = z_f \)) to that below it (at \( z = z_b \)), the particulate optical depth of the layer can be solved for

\[
\tau_f(z_b, z_f) = \frac{1}{2} \ln \left[ \frac{S_{\lambda_0}^b(z_f) z_b^2 O_{\lambda_0}(z_f) \beta_{m,\lambda_0}(z_f) T_{m,\lambda_0}^b(0, z_f)}{S_{\lambda_0}^b(z_b) z_b^2 O_{\lambda_0}(z_b) \beta_{m,\lambda_0}(z_b) T_{m,\lambda_0}^b(0, z_b)} \right] .
\] (12)

where the optical depth is the integral of the extinction coefficients:

\[
\tau_f(z_b, z_f) = \int_{z_b}^{z_f} \beta_{p,\lambda_0}(z') dz'.
\] (13)

The quantity \( \overline{\beta_{p,\lambda_0}} \) is the averaged value of the lidar ratio over the whole layer. By iterating, Eqs. (5) and (13) can be used to solve for a layer-averaged lidar ratio. In situations when the lidar ratio cannot be determined using Eq. (12) as a constraint, retrievals using only the elastic channel signal must assume a lidar ratio profile. In these situations, FEX attempts to infer the lidar ratios from other regions where a direct retrieval was made; otherwise, climatological values are used.

The retrieval of particulate extinction and backscatter in this work is performed for both FOVs of the ARM RL. At the expense of increased noise, using the low-channel signals has the benefit of achieving complete overlap sooner (i.e., at a lower height above the system), allowing for a more accurate retrieval in the near field than the high channels. For the ARM RL, the high channels achieve complete overlap by 5 km and the low channels by 800 m (Goldsmith et al. 1998). Quantities retrieved from low-channel signals are denoted by a superscript \( L \) (e.g., \( K^L_{p,\lambda_0}, \beta^L_{p,\lambda_0}, \alpha^L_{p,\lambda_0} \)), while those retrieved from high-channel signals have no superscript.

4. Retrieval algorithm

Part I describes in full detail the initial processing, determining calibration constants, and detecting features. In summary, photon-counting profiles from the MERGE product (Newsom et al. 2009) are accumulated to specified time and height bins—here 2 min and 30 m are used. Random signal uncertainty in each accumulated photon-counting profile is also calculated along with the signal-to-noise ratio (SNR). Molecular scattering terms (e.g., \( \beta_{m,\lambda_0} \) and \( \alpha_{m,\lambda_0} \)) are determined following Bucholtz (1995) using radiosonde observations. Climatological values of the aerosol Ångström exponent \( [\alpha(z)] \) are obtained from collocated Cimel sun photometer data (Holben et al. 1998). Overlap functions [e.g., \( O_{\lambda_0}(z) \)] are determined, along with all calibrations constants (e.g., \( C_{E,N_2} \)). Features are identified using range-dependent thresholds applied to multiple quantities. Of these quantities, the scattering ratio profile from the high elastic and nitrogen channels [Eq. (11)] and from the low elastic and nitrogen channels \( (SR_{E,N_2}^L) \) are needed along with the high-channel volume depolarization ratio \( (\delta) \).

The FEX extinction retrieval is summarized in Fig. 1 as a flowchart. The approach is to obtain an estimate of the particulate backscatter and lidar ratio profiles for all features using multiple retrieval methods and to combine these into a single best estimate. The extinction retrieval is iterative since both the feature classification and the multiple-scattering model require knowledge of the extinction profile itself. Iterations are also required since the methodology for feature detection (Part I) relies on extinction and vice versa.

a. Feature classification

Each pixel where a feature is detected by FEX is classified as either aerosol, rain (including virga), liquid cloud, ice cloud (including snow), or horizontally oriented ice (HOI) cloud. In addition to providing essential information about the target, this classification serves several practical purposes for the algorithm itself. Separating clouds from aerosols allows for the correct Ångström exponent profile to be used in Eqs. (8) and (11). For the purpose of particulate extinction retrieval,
this classification provides a guide for the appropriate lidar ratio to be inferred or assumed from a climatology when no direct retrieval can be made [section 4b(2)]. Furthermore, since the retrieval of lidar ratios using the Raman method relies on smoothing to overcome insufficient SNR, a feature classification allows different feature types to be smoothed separately. Finally, knowledge of the feature type is important for modeling multiple scattering [section 4b(3)] to determine both the appropriate particle size and backscatter phase function.

Although a more accurate determination of feature type could be made by incorporating other remote sensors, such as a cloud radar and/or a microwave radiometer (e.g., Shupe 2007), we have chosen here to design a scheme that relies only on temperature and humidity profiles from radiosondes in addition to the RL measurements. Because only lidar signals are used, mixed-phased pixels (i.e., those containing both ice and liquid) are not identified since liquid will dominate the total backscatter, and the received signal will be indistinguishable from that observed in a volume containing purely liquid. In lieu of more advance classification schemes, such as those based on neural networks (e.g., Bankert 1994; Miller and Emery 1997) or

FIG. 1. Flow diagram for the extinction retrieval in the FEX algorithm. Blue and red boxes denote the retrievals of particulate backscatter coefficient and lidar ratio, respectively. The best-estimate calculations occur in boxes with yellow backgrounds. Both the lidar ratio and particulate backscatter coefficient are retrieved at the laser wavelength $\lambda_0 = 355$ nm. The second subscripts indicate the channels used to derive each quantity with $E$ denoting the elastic channel, $N$ denoting the nitrogen channel, EN denoting that both the elastic and nitrogen channels are used, and BE denoting a best estimate derived from the combination of several quantities.
probabilistic methods (e.g., Baum et al. 1997; Hu et al. 2009; Liu et al. 2009), the classification here is based on a set of rules and empirical thresholds determined by both the expected scattering properties and established literature. To illustrate the choices for the empirical thresholds, the corresponding distributions of the retrieved feature properties themselves are given from the SGP site (Figs. 2 and 3). The distributions of these properties are similar at the TWP site (not shown). The same set of rules/thresholds is used for both the SGP and TWP sites.

FEX’s classification is guided by temperature, wet-bulb temperature, the depolarization ratio (δ), the best-estimate particulate backscatter (βpBE; see section 4b(1)), and directly retrieved lidar ratios obtain from either the Raman method Kp,EN or the transmission-loss method Kp,E [see section 4b(2)]. Temperature is taken from radiosonde profiles interpolated to the RL heights/times. The wet-bulb temperature is calculated using the interpolated radiosonde temperature and relative humidity profiles following Stull (2011). Both the particulate backscatter and lidar ratio are unavailable on the first iteration as no extinction retrieval has been performed yet (i.e., Fig. 1). Therefore, for the first iteration, the total attenuated backscatter is used in lieu of the particulate backscatter and no lidar ratio values are used. The lack of lidar ratio estimates hampers, but does not preclude a classification from being made as will become apparent below. Also note that both the particulate backscatter and lidar ratio used here are single-scattering values since the contribution of multiple scattering in each channel’s signal is accounted for [see section 4b(3)].

While it is possible to design a classification scheme using the particulate depolarization ratio, whose value

Fig. 2. Classification of aerosol and cloud by depolarization ratio (δ) and particulate backscatter (βp) thresholds for (a) wet-bulb temperature (Tw) above and (b) below 0°C. Modifications to the thresholds depending on the lidar ratio (Kp) are shown as dashed white lines. (c),(d) The histograms of in-feature depolarization ratios and best-estimate particulate backscatter in the corresponding temperature regimes from August 2008 through July 2013 at the SGP site. Frequency in (c) and (d) is normalized to the most frequency-occurring bin and presented on a logarithmic scale.
does not depend on the particle number concentration, we instead use the volume depolarization ratio. This is because the calculation of particulate depolarization ratio is numerically unstable (e.g., Cairo et al. 1999), especially for weakly scattering features, resulting in a significant fraction of unusable values when classification is done on a per-pixel basis.

Temperature is used to provide an absolute constraint on the possible cloud phase. Only liquid is likely to exist when the wet-bulb temperature is above 0°C, while only ice can exist at temperatures below the level of homogeneous freezing of -40°C. For the intermediate temperature range, either liquid or ice is allowed to exist and the details of separating the two are given in section 4a(3).

1) AEROSOL AND CLOUD

The classification starts by making an initial partitioning of clouds and aerosols using the particulate backscatter and depolarization ratio thresholds depicted in Fig. 2. The thresholds depicted using black lines in Fig. 2 are used regardless of the value of the lidar ratio, while thresholds given as white dashed lines are used when the lidar ratio becomes larger than either 40 or 60 sr. The reasoning for these modified thresholds for larger values of the lidar ratio will be discussed below.
When the wet-bulb temperature is above freezing, aerosol is separated from liquid cloud using a particulate backscatter threshold (Fig. 2a) since we expect the former to have smaller backscatter coefficients than the latter. This difference in backscatter can be seen in Fig. 2c, which gives the depolarization and backscatter histogram for all SGP features that occur when the wet-bulb temperature is above freezing.

For wet-bulb temperatures below freezing, cloud and aerosol are separated using both backscatter and depolarization thresholds as shown in Fig. 2b (black line). This allows for the possibility of cloud of either phase since ice, like aerosol, also has relatively small backscatter coefficients. However, ice is expected to have larger depolarization ratios than most (but not all) aerosol. This can be seen by comparing the two-dimensional (2D) histogram for features at these lower temperatures (Fig. 2d) to that for temperatures where mostly liquid cloud is expected (Fig. 2c). Regardless of the lidar ratio, features with $\delta$ less than 0.09 and relatively small backscatter coefficients are initially considered to be aerosol (Fig. 2b).

As shown by Omar et al. (2009), all types of aerosols have depolarization ratios less than this with the exception of dust aerosol. Pure dust can have depolarization ratios as large as 30% (Omar et al. 2009; Liu et al. 2011), which is comparable to that of ice clouds (e.g., Sakai et al. 2003). Because of this, the lidar ratio is used to help make the distinction between ice and dust. Modeling calculations and lidar observations have shown that typical lidar ratios of pure dust vary between 40 and 70 sr at visible and ultraviolet wavelengths (Sakai et al. 2002; Liu et al. 2002; Murayama et al. 2003; Anderson 2003; De Tomasi et al. 2003; Amiridis et al. 2005; Dubovik et al. 2006; Müller et al. 2007; Burton et al. 2012), larger than expected from most liquid or ice clouds [section 4b(2)]. Therefore, when the lidar ratio is greater than 40 sr, the maximum allowable aerosol $\delta$ is increased to 30% (dashed white in Fig. 2). An increase to 40% is made when the lidar ratio is greater than 60 sr since we expect almost no clouds to have lidar ratios larger than this [section 4b(2)].

In addition to an increase in the depolarization ratio threshold used to separate cloud and aerosol, the backscatter coefficient threshold is also increased for larger values of the lidar ratio. The increased backscatter threshold is used in both temperature regimes (Figs. 2a and 2b). Increasing the aerosol backscatter threshold allows for the identification of possible instances of optically thicker dust or smoke. Smoke typically has lidar ratios greater than $60$ sr (Müller et al. 2000; Peppler et al. 2000; Franke et al. 2001; Mattis 2003; Balis 2003; Burton et al. 2012).

The initial cloud and aerosol classification made using the thresholds in Fig. 2 can result in parts of clouds, typically the edges that have both small backscatter and depolarization values, to be falsely identified as aerosol. To remedy this, several steps are taken. First, height (time) layers are defined as consecutive range (time) bins containing a feature. If a pixel belongs to both height and time layers that are both mostly cloudy, then the pixel itself is also considered to be cloud. Second, aerosol pixels that are completely surrounded by cloudy pixels are changed to cloudy pixels. Finally, two-dimensional feature “objects” are determined: defined as regions of connected pixels in the current day being processed. Connectivity is defined using an eight-pixel neighborhood: that is, a pixel is considered connected to another pixel when any one of its eight neighbors—either the pixel in the next higher or lower height bin, the pixel in the preceding or following time bin, or the four pixels on the diagonals—contain a feature. If the majority of pixels in a feature object are identified as cloud, then all pixels in the object are changed to cloud. This object test is not applied to boundary layer aerosol, defined as 2D objects that contain at least one pixel below 400 m. Clouds are commonly found embedded in boundary layer aerosol, which would be identified as a single-feature object, and if enough pixels are cloudy, then the entire object could be erroneously changed to cloud.

2) RAIN

After classifying pixels as cloud or aerosol, the next step is to identify rain. The presence of rain is only allowed when the wet-bulb temperature is greater than 0°C. Therefore, rain must be distinguished from aerosol and liquid cloud. Rain can have a depolarization ratio and backscatter coefficient similar to either aerosol or liquid cloud; therefore, the lidar ratio is mainly used to identify its presence.

Figure 4 shows the theoretical lidar ratio for liquid spheres from Mie theory (Wiscombe 1980) for 355 nm and an index of refraction of $1.357 + i2.416 \times 10^{-9}$ (Segelstein 1981). The theoretical lidar ratios in Fig. 4 are given as a function of the median volume radius used in an assumed normalized gamma size distribution (e.g., Bringi and Chandrasekar 2001) for multiple values of the shape parameter $\mu$. The variation of the lidar ratio with $\mu$ and median radii in Fig. 4 embodies a wide range of observed distributions of liquid clouds and rain (Miles et al. 2000; Bringi et al. 2003). Also shown in Fig. 4 are the median lidar ratios retrieved from liquid clouds from the SGP and TWP RL [section 4b(2)]. These retrieved values of the lidar ratio agree well with the theoretical values, considering that the median droplet radius of most liquid clouds lies between 2 and 13 $\mu$m (Miles et al. 2000). When the median radius becomes larger than
about 400 $\mu$m, the lidar ratio begins to decrease with increasing droplet size. This decrease in the lidar ratio is used to identify the presence of rain in FEX: bins with $K'_{\chi,\text{EN}} < 12 \text{ sr}$ are initially considered to be rain. The median rain lidar ratio retrieved by FEX is similar at each site with both the SGP and TWP values being lower than predicted from Mie theory (Fig. 4). This is likely because the assumption of spherical drop shapes, and therefore Mie theory, is not valid for drops above a radius of 0.5 mm (Beard 1976). Less restrictive scattering calculations are needed to validate the retrieved values of the lidar ratio for rain, but such work is beyond the scope of this study.

After the initial identification of rain with $K'_{\chi,\text{EN}} < 12 \text{ sr}$, several requirements must be met to avoid falsely identifying aerosol as rain. Layers of rain—that is, consecutive range bins containing rain—must occur immediately below a cloud layer that fully attenuates the lidar signal, or the rain itself must completely attenuate the lidar signal (full attenuation is defined as when the high-elastichannel SNR is $<1$). Finally, any cloud layers immediately below rain layers are changed to rain, and if rain occurs in the lowest bin where a retrieval of the lidar ratio using the high-channel signals is made—that is, 1.5 km [see section 4b(2)]—then the remaining features at lower heights are assumed to also be rain.

### 3) Cloud Phase

Cloudy pixels are further classified into liquid, ice, or HOI using the particulate backscatter and depolarization ratio thresholds in Fig. 3. As mentioned previously, all clouds that occur above a wet-bulb temperature of $0^\circ\text{C}$ are considered to be liquid. When the wet-bulb temperature is below $0^\circ\text{C}$ and the temperature is above $-40^\circ\text{C}$, both liquid and ice can exist. In this temperature regime, depolarization is especially useful as only nonspherical particles like ice induce a depolarization (Sassen 1991). However, multiple scattering in liquid clouds also causes depolarization (Carswell and Pal 1980; Sassen 1991). The $\delta$ and $\beta_{p,\text{BE}}$ thresholds defined in Fig. 3a are based on an examination of the distribution of their values in pixels identified as cloud (Fig. 3c). Clouds form two distinct distributions with ice clouds associated with lower backscatter and higher depolarization and vice versa for liquid clouds. The black lines in Figs. 3a and 3c denote the thresholds defined to separate liquid from ice where the allowable depolarization for a liquid pixel increases with increasing backscatter to allow for multiple-scattering effects.

After applying the thresholds in Fig. 3a, the classification of ice/liquid cloud is finalized by applying a mode filter. For each pixel, the phase of cloudy pixels is replaced by the most common phase in a 210-m by 14-min window centered on the pixel. This filtering is done both to reduce noise in the classification and to correctly identify the edges of liquid clouds, which have small backscatter coefficients and are commonly initially misidentified as ice.

HOI particles are also identified by FEX. The beam angle for the TWP RL is about $4^\circ$–$5^\circ$ off zenith and the SGP RL is about $1^\circ$–$2^\circ$ off zenith (D. Turner 2013, personal communication). This near-zenith geometry makes it possible for a significant portion of the laser beam to be scattered perpendicular to the surface of HOI. Scattering from HOI particles has fundamentally different properties than scattering by randomly oriented ice (ROI); namely, for HOI, the polarization of the incident beam is preserved. Therefore, when the liquid–ice thresholds in Fig. 3a are applied, HOI will commonly be falsely identified as liquid.

In addition to small values of depolarization in HOI, the lidar ratio can be very small (e.g., Platt 1978). Therefore, $K'_{\chi,\text{EN}}$ is used to distinguish between HOI and liquid. As shown above, a lidar ratio threshold of 12 sr falls below all theoretical values that occur in liquid clouds; therefore, like with rain identification, pixels with $K'_{\chi,\text{EN}} < 12 \text{ sr}$, are classified as HOI. However, unlike rain, HOI only exists when the wet-bulb temperature is below $0^\circ\text{C}$ and $\beta_{p,\text{BE}}$ must be relatively large.
The HOI backscatter and depolarization requirements defined in Fig. 3a are similar to the limits found by Noel and Chepfer (2010). FEX’s HOI threshold of 12 sr corresponds to identifying volumes with ice crystals tilted to within 12° or less of the horizontal (Platt 1978), which should be sufficient for detecting HOI since previous lidar observation have shown that the majority of ice crystals wobble within 2.5° from the horizontal plane (Sassen 1991; Sassen and Comstock 2001). For temperatures below −40°C, where no liquid cloud exists, HOI is identified using only backscatter and depolarization thresholds (Fig. 3b).

4) EXAMPLE

Figure 5 gives the inputs into and results of FEX’s feature classification for 10 May 2011 at SGP. All feature categories identified by FEX are present in this example. Ice clouds are present above about 9 km. A layer of aerosol exists throughout the day. At both the beginning and end of the day, this aerosol layer is topped by mixed-phase cloud layer containing ice, liquid, and HOI. Several periods of rain also occur toward the end of this day. The stark contrast in the lidar ratio (Fig. 5c) leveraged by FEX to identify HOI and rain is clearly visible.

b. Extinction

An extinction retrieval with the highest possible accuracy is achieved by using the nitrogen channel signal to obtain the extinction and backscatter coefficients. However, this accuracy comes at the expense of less precision due to the larger amount of signal noise. On the other hand, retrievals using only the elastic channel give more precise values but with less accuracy because of the uncertainty in specifying the lidar ratio profile. Therefore, FEX seeks to create a best estimate using the methodology discussed in this section with the aim of creating a balance between precision and accuracy. At
the same time, FEX also provides values and their uncertainties from all retrieval methods so researchers can tailor a best estimate to their requirements, if need be. Specifically, the processing works toward retrieving the lidar ratio profile and the particulate backscatter coefficient as depicted in Fig. 1.

Both the lidar ratio and particulate backscatter coefficient are retrieved at \( \lambda_0 = 355 \text{ nm} \), though in Fig. 1 and in the following sections the subscript \( \lambda_0 \) is omitted from all symbols. Instead, the subscripts given indicate the channels used to derive each quantity with \( E \) denoting the elastic channel, \( N \) denoting the nitrogen channel, \( EN \) denoting that both the elastic and nitrogen channels are used, and \( BE \) denoting a best estimate.

1) PARTICULATE BACKSCATTER

The particulate backscatter coefficient can be directly obtained from the scattering ratio using both the elastic and nitrogen channels [Eq. (10)] for both the high (\( \beta_{p,BE} \)) and low (\( \beta_{p,EN}^L \)) channel signals. For the best estimate (\( \beta_{p,BE} \)), \( \beta_{p,EN}^L \) is used where the low-nitrogen-channel SNR is greater than 3; otherwise, \( \beta_{p,BE} \) is used. When the high-nitrogen-channel SNR is less than 3, the best estimate is taken from the Fernald method [Eq. (5); \( \beta_{p,BE} \)] using the lidar ratio best estimate from the methods discussed in the next section.

2) LIDAR RATIO

To obtain the extinction coefficient from the nitrogen channel signal [Eq. (8)] for a wide range of features requires some amount of smoothing/averaging. Instead of retrieving a smoothed extinction coefficient, FEX instead retrieves at efficient as depicted in Fig. 1.

\[ k(z', t', \delta z, \delta t) = \exp \left\{ -\frac{9}{2} \left[ \left( \frac{z'}{\delta z} \right)^2 + \left( \frac{t'}{\delta t} \right)^2 \right] \right\}, \] (15)

where \( z' \) and \( t' \) are

\[ z' = \left[ \frac{-(\delta z - 1)}{2} + n \mid n = 0, 1, \ldots, \delta z - 1 \right] \] (16)

and

\[ t' = \left[ \frac{-(\delta t - 1)}{2} + n \mid n = 0, 1, \ldots, \delta t - 1 \right], \] (17)

respectively. The kernel \( s \) is the same size as kernel \( k \) with \( s(z', t', \delta z, \delta t) = 1 \).

Equations (14)–(17) employ the knowledge of feature location by considering only pixels with the same feature present. In addition, potential large biases in the overlap function, which can be greatly exaggerated since it appears in the slope term in Eq. (8), are avoided by setting \( M(z, t) = 0 \), where \( z < z_0 \). For the high channel \( z_0 = 1.5 \text{ km} \) and for the low channel \( z_0 = 0.5 \text{ km} \). Above these heights the standard deviation of overlap functions derived by FEX are less than 10%. When the low-channel overlap function cannot be determined (see Part I), \( z_0 = 0.8 \) (i.e., the height above which the low channel achieves complete overlap). Using a Gaussian kernel over a moving average better preserves the spatial structure since pixels closer to the pixel being smoothed are likely to possess more similar properties. In lieu of more advanced nonlinear filters, using a linear convolution to apply the smoothing kernel allows for straightforward propagation of uncertainty. Smoothing is performed separately for each feature type to avoid mixing signals from dissimilar features.

Multiple profiles of smoothed high- and low-channel extinction and backscatter coefficients are calculated for the window sizes defined in Table 2. The ratio of these smoothed extinction and backscatter coefficients are
used to obtain six sets (one for each smoothing level) of smoothed lidar ratio profiles for each feature type. Standard uncertainty propagation (Bevington and Robinson 2002) is used to obtain the random error in these smoothed lidar ratio profiles.

The lidar ratio profiles obtained from the multiple levels of smoothing are combined into single-low-channel \( (K'_{p,EN}) \) and single-high-channel \( (K_{p,EN}) \) estimates. In each pixel, the lidar ratio obtained using the least amount of smoothing needed to achieve a random error less than 30% is used. These estimates of the lidar ratios from the high and low channels are then merged into a single estimate: \( K'_{p,EN} \). Pixels with values of the lidar ratio from both the high and low channels are merged 1) using \( K'_{p,EN} \) below 5 km when FEX cannot directly determine either the high elastic or nitrogen overlap function (Part I), 2) the channel with the least amount of smoothing is used, and 3) when the same amount of the smoothing is performed, the value with the smaller random error is used. Small gaps in \( K'_{p,EN} \) of less than 1 km and 20 min are filled using bilinear interpolation (error estimates are also interpolated). Like with smoothing, each feature type is processed separately when interpolating.

Layer-averaged values of the lidar ratio are determined using the transmission-loss method. The optical depth of layers consisting of a single-feature type and bounded by clear sky both above and below is determined using Eq. (12). To reduce the amount of random noise, Eq. (12) is evaluated by taking the median value of both the above-layer \( (z = z_1) \) and below-layer \( (z = z_s) \) terms over a 500-m portion of clear sky. The layer optical depth \( \tau_i \) and the directly retrieved lidar ratio profiles using the Raman method \( K'_{p,EN} \) are used as constraints on the Fernald solution to the elastic lidar equation [Eq. (5)]. When using the Fernald solution, the boundary condition at \( z = z_0 \) for the layer is set in the clear sky above (i.e., the far field), which ensures numerical stability (Klett 1981). For layers with a \( \tau_i \) estimate, an initial guess at the layer-averaged lidar ratio \( K_{p,E} \) is made and the particulate backscatter is solved for using Eq. (5). Iterations are performed by adjusting the layer-averaged lidar ratio \( K_{p,E} \) and resolving for the particulate backscatter until Eq. (13) is satisfied (i.e., \( K_{p,E} \) changes by less than 0.05 sr). When making adjustments to \( K_{p,E} \) during these iterations, bins with valid (i.e., less than 30% random error) directly retrieved \( K'_{p,EN} \) lidar ratios are not modified. For the lidar ratio best estimate, only \( K_{p,E} \) with relative errors less than 30% are used. Uncertainty in \( K_{p,E} \) is calculated by determining the standard deviation over the 500 m of the above-layer and below-layer terms and then performing standard error propagation using the median values.

| Table 3. Median and standard deviation of the directly retrieved lidar ratios for each feature type from December 2010 through December 2014 at the TWP site and from August 2008 through July 2013 at the SGP site. Values for HOI at TWP are not shown due to its small sample size. |

<table>
<thead>
<tr>
<th></th>
<th>Samples Median Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWP</td>
<td></td>
</tr>
<tr>
<td>Aerosol</td>
<td>18,968,239 40.26 20.56</td>
</tr>
<tr>
<td>Rain</td>
<td>2,568,387 5.50 4.32</td>
</tr>
<tr>
<td>Liquid</td>
<td>1,797,032 17.78 6.38</td>
</tr>
<tr>
<td>Ice</td>
<td>8,193,245 26.59 6.54</td>
</tr>
<tr>
<td>HOI</td>
<td>2,004,941 4.00 2.68</td>
</tr>
<tr>
<td>SGP</td>
<td></td>
</tr>
<tr>
<td>Aerosol</td>
<td>57,184,716 50.08 24.70</td>
</tr>
<tr>
<td>Rain</td>
<td>1,310,865 5.52 4.78</td>
</tr>
<tr>
<td>Liquid</td>
<td>2,109,185 18.88 8.66</td>
</tr>
<tr>
<td>Ice</td>
<td>14,999,902 22.03 7.44</td>
</tr>
</tbody>
</table>

The directly retrieved lidar ratios with less than 30% relative uncertainty—that is, the combined estimates of the lidar ratios from the Raman method \( K'_{p,EN} \) and the layer-averaged values \( K_{p,E} \) from the transmission-loss method—are used to infer the values for pixels where no estimate exists. Feature types are considered separately and four types of averages are determined, defined as follows:

1) layer averaged: using all directly retrieved lidar ratios in each vertical feature layer (defined as consecutive range bins containing a single-feature type);
2) object averaged: using all directly retrieved lidar ratios in each 2D feature object (objects are composed of a single-feature type connected in 2D; see section 4a for more details on how objects are defined);
3) profile averaged: using all directly retrieved lidar ratios in each profile;
4) daily averaged: using all directly retrieved lidar ratios in the same day that they are processed.

For all four types of averages, at least 25% of the pixels must have directly retrieved values. For example, for a profile-averaged lidar ratio to be used, at least 25% of all bins in the profile for a single-feature type must have a directly retrieved value. Inferred lidar ratios using these four averages are assigned with preference to using layer-averaged, object-averaged, profile-averaged, and finally daily-averaged values. When none of these averaged values can be used, a climatological value of the lidar ratio is used based on the feature classification as given in Table 3. Table 3 gives the median and standard deviation of all directly retrieved lidar ratios from August 2008 through July 2013 at SGP and December 2010 through December 2014 at TWP. The probability density functions (PDF) of these lidar ratios for these same periods are given in Fig. 6. The statistics in Table 3 and Fig. 6 represent some of the most comprehensive measurements of directly retrieved lidar ratios to date, particularly for clouds. Further analysis of these lidar ratios will be the subject of a future paper.
Figure 7 gives the lidar ratio retrieved by applying the Raman method to the high- and low-channel signals along with the resulting best-estimate lidar ratio on 25 December 2012 at the Darwin TWP site. Also given in Fig. 7, to help provide context, is the scattering ratio derived using only the elastic channel (Part I) and the classification of feature type. Throughout this day, a large amount of cirrus exists, along with boundary layer aerosol with a few liquid clouds embedded in it and a small amount of midlevel liquid cloud toward the end of the day. The lidar ratio of the aerosol layer in the best estimate (Fig. 7e) is obtained at various levels of smoothing using the Raman method applied to the high-channel signals above 1.5 km (shades of red in Fig. 7f). Below 1.5 km, the Raman method applied to the low-channel signals is used (shades of blue) with small gaps being filled in by interpolated values (purple). Below 500 m, where the Raman method is not attempted, a layer-averaged value is typically used (brown). In general, larger amounts of smoothing are required to obtain the lidar ratios for the cirrus layer. However, there exist portions of the cirrus layer at higher heights and those that are more tenuous for which the Raman method cannot be used to obtain a lidar ratio with a random error less than 30%. In some instances the transmission method can be used to fill in the remainder of the layer (green). For the rest of the cirrus layer, inferred values of the lidar ratio are used. In this example, no features require the use of a climatological lidar ratio.

3) MULTIPLE SCATTERING

The retrievals of the particular backscatter and lidar ratio discussed above require knowledge of the single-scattering signals, $S_{l0}^i(z)$, $S_{lN}^i(z)$, $S_{l0}^{ll}(z)$, and $S_{lN}^{ll}(z)$. However, only the total signal (i.e., single plus multiple scattering) is measured and therefore we introduce the multiple-scattering function $F(z)$ to account for multiple-scattering effects [e.g., Eqs. (4) and (9)]. Although it is traditional for multiple scattering to be accounted for by a factor that modifies the particulate extinction coefficient (e.g., Kunkel and Weinman 1976; Pal and Carswell 1976; Platt 1973; Young and Vaughan 2009), it is not necessary to do so. Conceptually, the increased signal strength as photons remain in the receiver FOV manifests itself as a combination of an apparent decrease in the extinction coefficient and an apparent increase in the backscatter coefficient. Furthermore, introducing a function that modifies the entire lidar equation, as is done in this work, is more numerically convenient. Multiple-scattering models require single-scattering values as inputs. Therefore, any explicit treatment of multiple scattering requires iterations. If multiple scattering is accounted for by modifying the particulate extinction coefficient, then the
Fig. 7. FEX’s retrievals on 25 Dec 2012 at the Darwin TWP site. (a) The scattering ratio derived using only the elastic channel; (b) the classification of features as either liquid cloud (red), rain (purple), ice (black), HOI (brown), or aerosol (gray); (c) the lidar ratio derived using the Raman method applied to the high-channel signals; (d) the lidar ratio derived using the Raman method applied to the low-channel signals; (e) the best-estimate lidar ratio; and (f) flag indicating the type of retrievals used in the best-estimate lidar ratio. Shades of red and blue in panel (f) denote the level of smoothing (Table 2) used for the Raman method for the high- and low-channel signals. Other colors in (f) denote those pixels where the lidar is determined by interpolation (purple), transmission-loss method (green), layer-averaged (brown), object-averaged (pink), profile-averaged (orange), daily-averaged (gray), and climatological values (black).
The H06 model, whose parameterizations of ice and used to set the proper near-backscatter phase function in well; therefore, the feature classification made by FEX is elastic channel contains backscatter from particulates as liquid cloud phase functions are given in Hogan (2008). The H06 model does not provide any parameterization function near 180° for the elastic and nitrogen channel signals as the phase Separate multiple-scattering functions are also needed functions are calculated for the high and low channels. Separate multiple-scattering functions are also needed for the elastic and nitrogen channel signals as the phase function near 180° differs in each channel. For purely molecular backscatter, as occurs in the nitrogen channel, the phase function is near isotropic around 180°. The elastic channel contains backscatter from particulates as well; therefore, the feature classification made by FEX is used to set the proper near-backscatter phase function in the H06 model, whose parameterizations of ice and liquid cloud phase functions are given in Hogan (2008). The H06 model does not provide any parameterization of the phase function for rain and aerosol; therefore, the near-backscatter phase function for these bins are treated as liquidlike and isotropic, respectively. The H06 model also requires the particle size, specifically the equivalent-area radius, to determine the distribution of the forward-scattered photons. The particle sizes used for each feature type and their respective references are given in Table 4. The ice particle size is parameterized as a function of extinction. Other feature types use a single size for all multiple-scattering calculations. Table 4 contains a combination of particle sizes expressed as effective median volume and mean radii, all of which are assumed to be approximately equal to the equivalent-area radius. Finally, the multiple-scattering model requires molecular backscatter and extinction coefficients, which are calculated from radiosonde observations, and the single-scattering particle-backscatter and extinction coefficients. The latter requirement makes iterations necessary. An initial guess at the multiple-scattering functions is made using the best-estimate particulate backscatter coefficients and lidar ratios derived using the total signals. On the next iteration, these functions are used to obtain the single-scattering signals needed for the extinction retrievals with iterations continuing until all the bins in all four multiple-scattering functions change by less than 0.1%. During these iterations, the ice particle sizes are updated with each new extinction best-estimate profile.

Potential biases introduced by the multiple-scattering functions in the best-estimate extinction were investigated. Two runs of FEX were performed with multiple years of TWP and SGP data using the high/low value or plus/minus one standard deviation of the particle size or fit coefficients as given in Table 4. A third run of FEX was also performed to test the sensitivity to the near-backscatter phase function by using a liquidlike phase function for aerosols and an isotropic one for all other features. For aerosols, it is found that the assumed particle size and near-backscatter phase function introduce a median uncertainty of less than ±0.04% in the best-estimate extinction. The median extinction uncertainty due to multiple-scattering assumptions in all other features is less than ±2%. The exception is for the near-backscatter phase function for rain: assuming that it is isotropic instead of liquidlike introduces a median uncertainty of about 5%. All these uncertainties due to the choice of particle size or near-backscatter phase function are at least an order of magnitude less than the typical multiple-scattering corrections themselves (section 5c).

The efficacy of the multiple-scattering functions was tested by comparing both the total (e.g., $S_{\text{tot}}$) and

<table>
<thead>
<tr>
<th>Feature type</th>
<th>Site</th>
<th>Size (µm)</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ice and H0I</td>
<td>TWP</td>
<td>43.29(±3.783)α^{0.2514(±0.0861)}</td>
<td>Power law derived from effective radius in situ data where α is extinction (km⁻¹)</td>
<td>Table 2 in Fu (1996)</td>
</tr>
<tr>
<td></td>
<td>SGP</td>
<td>6.54 (±1.91)</td>
<td>Effective radius from multiple retrieval algorithms at TWP and SGP</td>
<td>Fig. 4 in Zhao et al. (2012)</td>
</tr>
<tr>
<td>Liquid</td>
<td>TWP</td>
<td>5.70 (±1.11)</td>
<td>C-band polarimetric radar retrievals near the TWP site of median volume drop radius</td>
<td>Table 1 in Thurai et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>SGP</td>
<td>6.54 (±1.91)</td>
<td>W-band Doppler radar spectra retrievals at SGP of median volume drop radius</td>
<td>Fig. 6 in Giangrande et al. (2012)</td>
</tr>
<tr>
<td>Rain</td>
<td>TWP</td>
<td>670 (±135)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SGP</td>
<td>770 (±165)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerosol</td>
<td>TWP</td>
<td>2.19 [1.75 2.55]</td>
<td>Mean radius from AERONET global climatology</td>
<td>Table 2 in Omar et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>SGP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Particle radii used in the multiple-scattering model. Standard deviations are given in parentheses. For aerosols, the mean size is calculated from Omar et al. (2005) by taking the average of the coarse-/fine-mode radii weighted by the coarse-/fine-mode fraction. The bracketed values for aerosols give the minimum- and maximum-weighted average sizes in the aerosol categories from Omar et al. (2005).
single-scattering (e.g., $S^0_0$) low-channel signal to the high-channel signal. This comparison is given in Fig. 8 for in-cloud signals for multiple years of TWP data. Only signals above 5 km are used in this comparison to avoid introducing potential biases from the high-channel overlap functions, and the low-channel signals have been calibrated relative to the high-channel signals (Part I). Data are from December 2010 through December 2014 at the TWP site. Summary statistics are given in the bottom-right corner of each panel: the sample size ($N$), the slope of the linear regression line, the correlation coefficient ($r$), the root-mean-squared error [RMS, calculated as RMS difference/mean (high channel)], the bias error [calculated as mean(low channel − high channel)/mean(high channel)].

5. Results

All results presented in this section, unless otherwise specified, are from data averaged to 2 min and 30 m from August 2008 through July 2013 at the SGP site and December 2010 through December 2014 at the TWP site. During these periods the RL experienced an uptime of 97% and 52% at the SGP and TWP sites, respectively.

a. Vertical occurrence of features

Figure 9 gives the occurrence of liquid cloud, ice cloud (including both ROI and HOI), HOI, rain, and aerosols at both the TWP and SGP sites. Several expected differences exist between the two sites, such as a higher frequency of rain and ice cloud at the TWP site. The clouds at TWP show the characteristic tropical trimodal distribution (Johnson et al. 1999) with a lower-level peak in liquid cloud amount, a midlevel peak in both liquid and ice cloud near the melting level, and an upper-level peak in ice cloud below the tropopause. More aerosols exist below about 2 km at TWP; while more aerosols exist above 2 km at SGP. A nontrivial amount of HOI occurs at SGP with its frequency comparable to
About 90% of aerosols are able to use both directly retrieved lidar ratios and backscatter coefficients [with the exception of below 500 m for the lidar ratio; see section 4b(2)]. A directly retrieved particulate backscatter coefficient is used for about 80% of liquid clouds. For the liquid cloud lidar ratio, the solar background exerts some influence since about 80%–90% of lidar ratios are directly retrieved at night compared to 50% during the daytime, when there is more of a reliance on layer- and daily-averaged values.

With ice clouds, there is a clear influence of the solar background in both the particulate backscatter and lidar ratio processing choices. For both the particulate backscatter and lidar ratio, about 75% of nighttime ice clouds use a directly retrieved value below about 15 km. At SGP the amount is larger: about 85% of nighttime ice clouds use directly retrieved values, regardless of height. At TWP during the daytime, the majority of ice clouds above 5 km require the Fernald solution to obtain the particulate backscatter coefficient. Similar daytime frequency profiles exist at SGP, although the change to the majority of ice cloud using the Fernald solution occurs above 8 km instead. For the ice cloud lidar ratio retrieval during the day (Fig. 11c), the Raman method is used for the majority of profiles only at lower heights, and its use steadily decreases to zero at about 14 km. At these higher heights, there is a heavier reliance on transmission-loss estimates, and object and daily averages. The use of climatological values is needed for less than 5% of ice clouds below 12 km. However, above 12 km the presence

**Fig. 9.** The vertical occurrence of feature types at the (a) TWP site and (b) SGP site. The feature types given are liquid cloud (red), rain (purple), ice (black), HOI (brown), or aerosol (gray). The ice occurrence profile includes both ROI and HOI. The aerosol occurrence profiles have been scaled by a factor of 0.25.
of optically thin cirrus makes climatological lidar ratios necessary for about 20% of ice clouds. At SGP, where cirrus are typically optically thicker, there is no such increase in the amount of climatological lidar ratios with height during the day; about 5% of ice clouds need to use a climatological value regardless of height.

Figure 12 gives the amount of smoothing required to obtain lidar ratios from the Raman method applied to the high-channel signals at TWP. Here, day and night are not shown separately since requiring a fixed maximum uncertainty makes for a similar pattern of smoothing levels. For liquid cloud (Fig. 12b), the majority of the lidar ratios can be obtained with either no smoothing or at the second level of smoothing (0.3 km × 10 min). For aerosol and ice cloud, almost no lidar ratios can be obtained without some smoothing. Both aerosol and ice cloud smoothing amounts increase with height. The overall pattern of smoothing amounts for the low-channel Raman method or at SGP is similar to those presented in Fig. 12.

c. Multiple-scattering errors

An offline run of FEX with no multiple-scattering correction was performed to examine biases introduced by neglecting the effects of multiple scattering. We focus only on the effect of multiple-scattering on the best-estimate extinction produced by FEX, although errors due to ignoring multiple scattering vary with retrieval method and the quantity being analyzed. For example, for the particulate backscatter coefficient obtained from the scattering ratio using both the elastic and nitrogen channels, there is a partial cancellation of multiple-scattering effects (Wandinger 1998). But quantities that use only a single signal, such as the Fernald solution or the Raman method for extinction, will have larger contributions from multiple scattering. In addition,
quantities using low-channel signals will have larger multiple-scattering effects than quantities using high-channel signals. Here we focus only on the best-estimate extinction, which is a combination of multiple quantities from both FOVs (section 4b).

A box plot of the relative errors \([\text{extinction without a multiple-scattering correction} - \text{extinction with multiple-scattering correction}] / \text{extinction with multiple-scattering correction}\) is given in Fig. 13. Similar error distributions exist at TWP and SGP. Multiple-scattering effects for aerosols are small, with upper/lower quartiles of about \(\pm 2\%\). For hydrometeors, the failure to account for multiple scattering typically causes extinction to be biased low, although appropriate corrections vary in both range and sign. The large size of rain translates to the largest errors with median errors around \(-35\%\). Typical errors in liquid and ice cloud extinction range from about \(-30\%\) to near 0\%. At SGP, HOI has a wide range of errors with a median close to 0\%. The errors in extinction presented in Fig. 13 illustrate the importance of explicitly treating multiple scattering when using systems similar to the ARM RL for hydrometeor retrievals.

d. Sun photometer comparison

Validating FEX’s cloud extinction retrieval is difficult since no instrument at the ARM sites makes comparable measurements. However, for aerosol extinction, the ARM sites are equipped with Cimel sun photometers, which operate as part of the Aerosol Robotic Network (AERONET; Holben et al. 1998), allowing for the comparison of daytime aerosol optical depth as shown in Fig. 14. The coincident sun photometer optical depth is calculated at 355 nm by applying the 380–340-nm Ångström exponent to the 340-nm channel optical depth. Level 1.5 cloud-screened sun photometer data are used (Smirnov et al. 2000), and profiles where the RL-detected
clouds are excluded. The comparison in Fig. 14 depends on several portions of FEX, including the extinction retrieval itself, classifying the feature as aerosol, and detecting features since extinction retrievals are only performed where a feature is detected.

As shown in section 5b, directly retrieved aerosol extinction coefficients are not always possible. Therefore, to evaluate what is expected to be the most accurate profiles from FEX, Figs. 14a and 14c compares only the subset of FEX profiles where all aerosol bins above 1.5 km (i.e., the lowest height used in the high-channel Raman method lidar ratio retrieval) have directly retrieved extinction. In this comparison, bias errors are $-4.9\%$ and $-3.0\%$ (relative to the sun photometer) at TWP and SGP, respectively. Figures 14b and 14d show the same comparison, but they include all profiles regardless of the type of extinction retrieval performed. The agreement here is very similar to the directly retrieved extinction comparison, suggesting that the methodology used to build the best estimate does not introduce significant biases.

Figure 15 gives the same comparison as Figs. 14b and 14d but for RL data accumulated to 10-min time bins. In this comparison, improved agreement is found: bias errors are $-4.3\%$ and $-0.3\%$ at TWP and SGP, respectively. Compared to the 2-min data, the RL-FEX aerosol optical depth increases slightly when using 10-min bins; this is an indication that a small amount of aerosol goes undetected at 2 min. Further accumulation of signal beyond 10 min results in little change in the RL-FEX optical depth (not shown).

We consider the biases presented in Figs. 14 and 15 to be reasonably good considering that the sun photometer is measuring optical depth along a different path through the atmosphere and the lack of direct lidar ratio estimates from the RL at lower heights. Relative to the 14-channel NASA Ames Airborne Tracking Sunphotometer (AATS-14), Schmid et al. (2006) report AERONET biases of 6% and 7% at 340 and 380 nm, respectively. Comparisons of AATS-14 to

---

**Fig. 12.** The frequency of smoothing levels (Table 2) used in the lidar ratio derived from the Raman method using the high-channel signals at the TWP site. Increasingly darker shades of red represent larger amounts of smoothing. In each height bin, frequency is normalized by the total number of (a) aerosol, (b) liquid cloud, or (c) ice cloud.

---

**Fig. 13.** Box plot of the relative error in the particulate extinction retrievals caused by neglecting multiple scattering at the (left side) TWP and (right side) SGP sites. The colored boxes give the interquartile ranges with the thick black line denoting the median. The whiskers give the 5th and 95th percentiles. The error in HOI for TWP is not shown due its small sample size.
other instruments measuring optical depth by Schmid et al. (2006) find that a 15%–20% bias is typical among different measurements of visible optical depth, and that errors are likely larger in the ultraviolet. Turner et al. (2002), using the original RL extinction algorithm, performed a similar comparison at SGP, finding a bias of about 30%.

6. Summary and conclusions

Presented is an automated retrieval of extinction from the Atmospheric Radiation Measurement Program’s (ARM) Raman lidars (RL), which is Part II of the feature detection and extinction retrieval (FEX) algorithm. Part I focused on feature detection. The intent is to run FEX operationally within the ARM Data Management Facility (DMF) with the output being made available to the general user community via the ARM website (http://www.arm.gov/).

The objective of this work is to obtain the best estimate of particulate backscatter and lidar ratios for all detected features. Depending on the SNR, the particulate backscatter is directly determined from the scattering ratio derived from both the high/low elastic and high/low nitrogen channels, or from the Fernald solution to the elastic lidar equation using the best-estimate lidar ratio profile. The best-estimate lidar ratio profiles are directly retrieved using a combination of the elastic and nitrogen channel signals with adaptive amounts of smoothing applied or the layer-averaged lidar ratio using the transmission-loss method. The uncertainty is required to be less than 30% for both types of retrievals. When this is
not possible, directly retrieved lidar ratios are used to infer lidar ratios for the day being processed. When neither directly retrieved nor an inferred value can be determined, a climatological lidar ratio is used. Multiple years of processed data at both the SGP and TWP show that climatological values of the lidar ratio are only necessary for less than about 5% of features, except for optically thin cirrus at the TWP. There, above 12 km, around 20% of clouds are processed using a climatological lidar ratio.

The process of retrieving extinction in FEX is supported by a classification of feature types: aerosol, water cloud, ice cloud, rain, and HOI. The contribution of multiple scattering is explicitly considered in each range bin. We show that errors in extinction due to ignoring multiple-scattering effects are significant for hydrometeors. For clouds, most corrections fall in the range from 0% to 30%, while median errors in rain are about 35%. Errors in aerosol extinction due to multiple scattering are relatively small (±2%). The accuracy at both the TWP and SGP sites in aerosol optical depth is established through a comparison with collocated multiyear sun photometer observations.

The continuously operated, automated ARM RLs provide an enormous wealth of water vapor, temperature, aerosol, and cloud data that has been unmatched outside the ARM program until only recently (Reichardt et al. 2012). The work described in this series of papers has greatly improved the quality of aerosol and cloud data, particularly the latter, to help fully realize the exceptional abilities of this instrument. The comprehensive set of lidar ratios has the potential to improve the extinction retrievals from existing elastic lidar datasets (e.g., CALIPSO). The opportunity now exists to study the variability of these lidar ratios, as their representativeness of changes in microphysics is relatively unknown, particularly for ice clouds. This work also has the capability to serve as an automated retrieval framework for other Raman lidars or HSRLs, including potential future advanced spaceborne lidars.

Acknowledgments. The Raman lidar, radiosonde, and sun photometer datasets were obtained from the ARM data archive (www.archive.arm.gov). We thank R. Hogan for making his multiple-scattering model freely available. This work was greatly improved by the excellent reviews of M. Vaughan, J. Campbell, and one anonymous reviewer. This research was supported by the Office of Science (BER), U.S. Department of Energy under Grant DE-SC0010557. It was also supported in part by NSFC Grant 41430425.

REFERENCES


