Cluster Analysis Techniques to Separate Air Motion and Hydrometeors in Vertical Incident Profiler Observations

CHRISTOPHER R. WILLIAMS, WARNER L. ECKLUND, AND PAUL E. JOHNSTON

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

KENNETH S. GAGE

National Oceanic and Atmospheric Administration Aeronomy Laboratory, Boulder, Colorado

(Manuscript received 7 June 1999, in final form 12 October 1999)

ABSTRACT

Profilers operating in the UHF range are sensitive to both Bragg scattering from radio refractive index structure and to Rayleigh scattering from small point targets. Identification of the scattering process is critical for proper interpretation of these observations, especially the data collected from the vertical incident beam. This study evaluates the performance of Doppler velocity thresholds as a means to separate air motions from hydrometeor motions in vertical incident profiler observations. This evaluation consists of three different steps. First, using two collocated profilers operating at different frequencies, the observations are unambiguously identified as Bragg or Rayleigh scattering processes. Second, the observations are separated into either air or hydrometeor motion using only the data from one profiler. The third step quantitatively evaluates the performance of the single profiler separation techniques by counting the number of correct classifications and adjusting the count by the number of incorrect classifications.

Constant Doppler velocity threshold methods are acceptable methods to separate air motions from hydrometeor motions only after the *correct* threshold is determined. This study presents a cluster analysis method that robustly and objectively separates air from hydrometeor motions. The introduced cluster analysis produces two thresholds. The first threshold is a Doppler velocity threshold that is a function of reflectivity. The second threshold is the maximum reflectivity in which the Doppler velocity threshold divides the observations into two statistical distributions using the Kolmogorov–Smirnov statistical test. The cluster analysis method quantitatively performs better than constant Doppler velocity threshold methods, and is a repeatable, self-adapting, statistically based procedure.

1. Introduction

Profilers operating in the ultra-high frequency (UHF) range are sensitive to the backscattered power from structures at half the operating wavelength due to gradients of the radio refractive index in the atmosphere as well as to small point targets in the profiler pulse volume. This sensitivity enables the atmospheric refractivity turbulence (Hill 1978) as well as the vertical structure of precipitating cloud systems (Gage et al. 1994) to be studied using the same profiling Doppler radar. If the different scattering processes are not identified and separated, the analyzes of either the turbulence or precipitation vertical structure will be cross-contaminated.

The sensitivity of UHF profilers to atmospheric refractivity turbulence allow them to observe the ''clear-air'' motions overhead. The refractivity turbulence results from inhomogenities in humidity and temperature within the radar pulse volume (e.g., Gage 1990). Bragg scattering processes are detected by the profiler when energy is backscattered from refractive turbulence irregularities that have spatial structures equal to half the profiler operating wavelength. The backscattered energy from small distributed point targets (such as raindrops) in the profiler pulse volume results from Rayleigh scattering processes. These distributed targets may be hydrometeors (raindrops, snow flakes, ice particles), biological flyers (birds or insects), or airborne particles. The profiler's sensitivity to both Bragg and Rayleigh scattering processes permits the profilers to be used for estimating the vertical air motion and the vertical structure of precipitating cloud systems, as long as the two scattering processes can be identified and separated.

One method to identify Bragg and Rayleigh scattering processes is to sample the same portion of the atmosphere with two profilers operating at different frequencies. The difference in reflectivity between the two

Corresponding author address: Christopher Williams, CIRES/NOAA Aeronomy Laboratory, Mail Stop R/E/AL3, 325 Broadway, Boulder, CO 80303.
E-mail: chris@al.noaa.gov

profilers is zero for Rayleigh scattering observations and is a constant determined by the ratio of operating wavelengths for Bragg scattering observations (Rogers and Brown 1997). Absolute calibration of both profilers is not necessary to identify the scattering processes. The work of Gage et al. (1999) demonstrated that two collocated profilers operating at 915 and 2835 MHz could be used to unambiguously identify the Bragg and Rayleigh scattering processes over the Tiwi Islands north of Darwin, Australia.

Although it is not possible to unambiguously separate the scattering processes using a single frequency profiler, several analysis techniques have been proposed to separate the vertical air and hydrometeor velocities based on the characteristics of the observations. Ralph (1995) performed a thorough analysis to determine the vertical Doppler fall speed of rain and snow or ice when the returned power from the hydrometeors was equal to that from moderate and strong turbulence. This analysis provided the background necessary to use the vertical Doppler velocity as a threshold to separate air and hydrometeor motions. This work also provided constant Doppler velocity thresholds that separated hydrometeor and air motion observations for moderate and strong turbulent air motion conditions. These constant thresholds do not account for local atmospheric conditions and precipitation drop size distributions that may be significantly different than the midlatitude values used to develop the theory. In order to account for local atmospheric conditions, Ralph et al. (1996) introduced the dual-optimization method to determine the best constant Doppler velocity threshold that separates the air and hydrometeor motions. This method, applied at each altitude, examines Doppler velocity distributions during a finite time interval. This repeatable procedure uses the statistical properties of the hydrometeor and air motion distributions to determine the best Doppler velocity threshold to divide the two distributions. Steiner and Richner (1994) used a combination of reflectivity, Doppler velocity, and spectral width to develop a median parameter to separate air and hydrometeor motions. Estimating the scattering process benefits from the addition of the different spectral moments, even though the combination of observations from multiple altitudes decreases the performance of the separation method.

This paper uses the two frequency information from two collocated profilers to unambiguously separate the Bragg or Rayleigh scattering observations. Using only the information from a single frequency profiler, the constant and dual-optimization Doppler velocity thresholds that separate air and hydrometeor motions are evaluated. This paper also introduces an impartial method to separate the air and hydrometeor motions using cluster analysis techniques. This new method uses both the reflectivity and Doppler velocity observations to produce a velocity threshold that is a function of reflectivity. The cluster analysis also yields a reflectivity threshold above which two statistically separate distributions do not exist, and therefore, only hydrometeor motions are present. All three separation methods are evaluated using an objective goodness score that counts the number of correct separations adjusted for the number of incorrect separations.

The observations presented in this paper reflect 54 days from Manus Island, Papua New Guinea. Section 2 briefly describes these observations. Using the two-frequency technique described in Gage et al. (1999), the work in section 3 unambiguously identifies these observations as relation to either Bragg or Rayleigh scattering. The new cluster analysis method to separate air and hydrometeor motions using the Doppler velocity and reflectivity is presented in section 4. The independent goodness score used to evaluate the different techniques is discussed in section 5. The constant Doppler velocity threshold and dual-optimization methods are objectively evaluated in section 6. Section 7 compares previous separation methods and the cluster analysis method. Concluding remarks are made in section 8.

2. Manus Island datasets

A UHF profiler operating at 915 MHz, in support of the Tropical Ocean Global Atmosphere (TOGA) program, was installed on Manus Island, Papua New Guinea, in March 1992. This profiler has operated nearly continuously, providing horizontal winds, vertical motions, and RASS (Radio Acoustic Sounding System) estimated virtual temperature. The horizontal winds and virtual temperature observations during the TOGA Coupled Ocean–Atmosphere Response Experiment (COARE) have been quality controlled, and the archived datasets are available to the research community (Parsons et al. 1994; Riddle et al. 1996; Hartten 1998).

In February 1996, an S-band profiler operating at 2835 MHz was installed next to the 915-MHz profiler on Manus Island. For approximately two months, both profilers operated nearly continuously. The transmitted peak power of the S-band profiler (350 W) was less than the 915-MHz transmitted peak power of 500 W, so that both profilers had approximately the same sensitivity to Bragg scatter. The S-band profiler was much more sensitive to the hydrometeors than the 915-MHz profiler because of the wavelength dependence of the Rayleigh scattering. Table 1 summarizes system characteristics.

The UHF profiler sampled the atmosphere with two different vertical resolutions, either a 100-m pulse length resolution from 0.2 to 6 km or a 500-m pulse length resolution from 0.4 to 17 km. The UHF profiler determined the horizontal wind by using the Doppler beam swinging (DBS) technique to sample the atmosphere in two off-zenith beam directions and in the vertical beam direction. Combining three consecutive 30-s beam averages produced one complete triad of observations every 90 s. The DBS technique was immediately repeated for the other pulse length resolution. This six-beam swinging sequence was repeated for 25 min of every half hour to determine

TABLE 1. Typical parameters used at Manus Island for 915- and 2835-MHz profilers.

Parameter	S-band profiler	UHF profiler
Frequency	2835 MHz	915 MHz
Wavelength	10.6 cm	32.8 cm
Peak power	320 W	500 W
Antenna	3-m shrouded dish	2-m phased array
Beamwidth	3.2°	9°
Height resolution	100 and 500 m	100 and 500 m
Maximum height sampled	18.9 km for 500-m mode	17.3 km for 500-m mode
Maximum radial velocity	± 11 m s^{-1}	±11 m s^{-1}
Spectral points	128	128
Dwell time	~30 s	~30 s
Recording	Full spectra	Full spectra

the half hour horizontal winds as described by Carter et al. (1995) and Hartten (1998). During the other 5 min of the half hour, the profiler operated in the RASS mode, measuring the Doppler shift of the transmitted acoustic energy to determine the virtual temperature profile (May et al. 1990). During the 25 min the profiler was in the DBS mode, one vertical beam profile was obtained every 90 s and alternated between the 100- and 500-m pulse length resolutions.

The vertically pointing S-band profiler operated with the same vertical resolution as the UHF profiler to ensure similar pulse volumes. The dwell time of each S-band profile was about 30 s, and the sequence was nine 500-m pulse length resolution profiles followed by one 100-m pulse length resolution profile. Emphasis on the 500-m resolution allowed for focus on the weak precipitating cloud systems that advect over the profiler (Ecklund et al. 1999). These weak precipitating cloud systems are observed above the melting level and do not contribute to the surface precipitation, yet they do affect the surface radiation. The profilers recorded 128-point Doppler spectra at each range gate. While techniques at the NOAA (National Oceanic and Atmospheric Administration) Aeronomy Laboratory are being developed to routinely separate and identify the Bragg and Rayleigh scattering components resolved in the Doppler velocity spectra, this study uses only the moments determined from the dominant spectral peak. Also, this study analyzes only the 500-m pulse length profiles.

In order to compare the observations made by the different profilers, only the profiles that were within ±60 s of each other were deemed simultaneous profiles. During the 54-day period from 19 February to 23 April 1996, more than 15 000 simultaneous vertical profiles were recorded. Their reflectivity distributions, as a function of altitude, are shown in Fig. 1. The top two panels on the left (Figs. 1a,c) show the number of observations at each altitude, and the top two color panels (Figs. 1b,d) show the two-dimensional frequency distributions (2DFDs) of reflectivity (expressed in units of the water-equivalent reflectivity factor designated dBZ_e) from the 915- and 2835-MHz profilers, respectively. The warmest color in each panel represents the pixel with the maximum number of occurrences. The colors become cooler on a logarithmic scale to show more detail than linear color scales provide. The 2DFD allows comparison of both the reflectivity and the altitude dimensions (ordinate and abscissa) in contrast to Contoured Frequency by Altitude Diagrams (CFADS), which, as normalized by altitude, prohibit comparative distributions along the altitude dimension (Yuter and Houze 1995).

Figure 1f shows the 2DFD of the reflectivity difference of each simultaneous observation, and Fig. 1e shows the number of observations as a function of altitude. Two different distributions, resulting from the Bragg and Rayleigh scattering observations, can be identified in this panel. As described in the general overview of Rogers and Brown (1997), radars operating at different wavelengths have different sensitivities to Bragg and Rayleigh scattering. If the dominant scattering process is from small point targets—for example, raindrops or insects—then the scattering process would be Rayleigh scattering. The difference in reflectivity for Rayleigh scattering is zero. If the dominant scattering process is from the radar refractive index caused by humidity gradients and turbulence, then the scattering process would be Bragg scattering. The ratio of reflectivity factor Z for two profilers observing the same volume of Bragg scattering is

$$\frac{Z_{\mathrm{UHF}}}{Z_{\mathrm{Sband}}} = \left|\frac{\lambda_{\mathrm{UHF}}}{\lambda_{\mathrm{Sband}}}\right|^{11/3} = \left|\frac{33.0}{10.6}\right|^{11/3} = 64.3, \qquad (1)$$

which is the same as the difference in the reflectivity factor in log space (UHF–S band) of 18 dB [see Rogers and Brown (1997) and Gage et al. (1999) for more details]. In Fig. 1f, the observed reflectivity difference is near 18 dBZ and increases with altitude. This increase apparently results from the relative decrease in S-band reflectivity with altitude, consistent with the decreased inertial subrange of turbulence.

Two collocated profilers observing the same volume can unambiguously identify Rayleigh and Bragg scattering processes, but a single profiler can estimate only the scattering process. To avoid confusion in terminology, a single profiler estimate of the scattering process is termed either an *air* motion or a *hydrometeor* motion observation. If the single profiler estimate agrees with

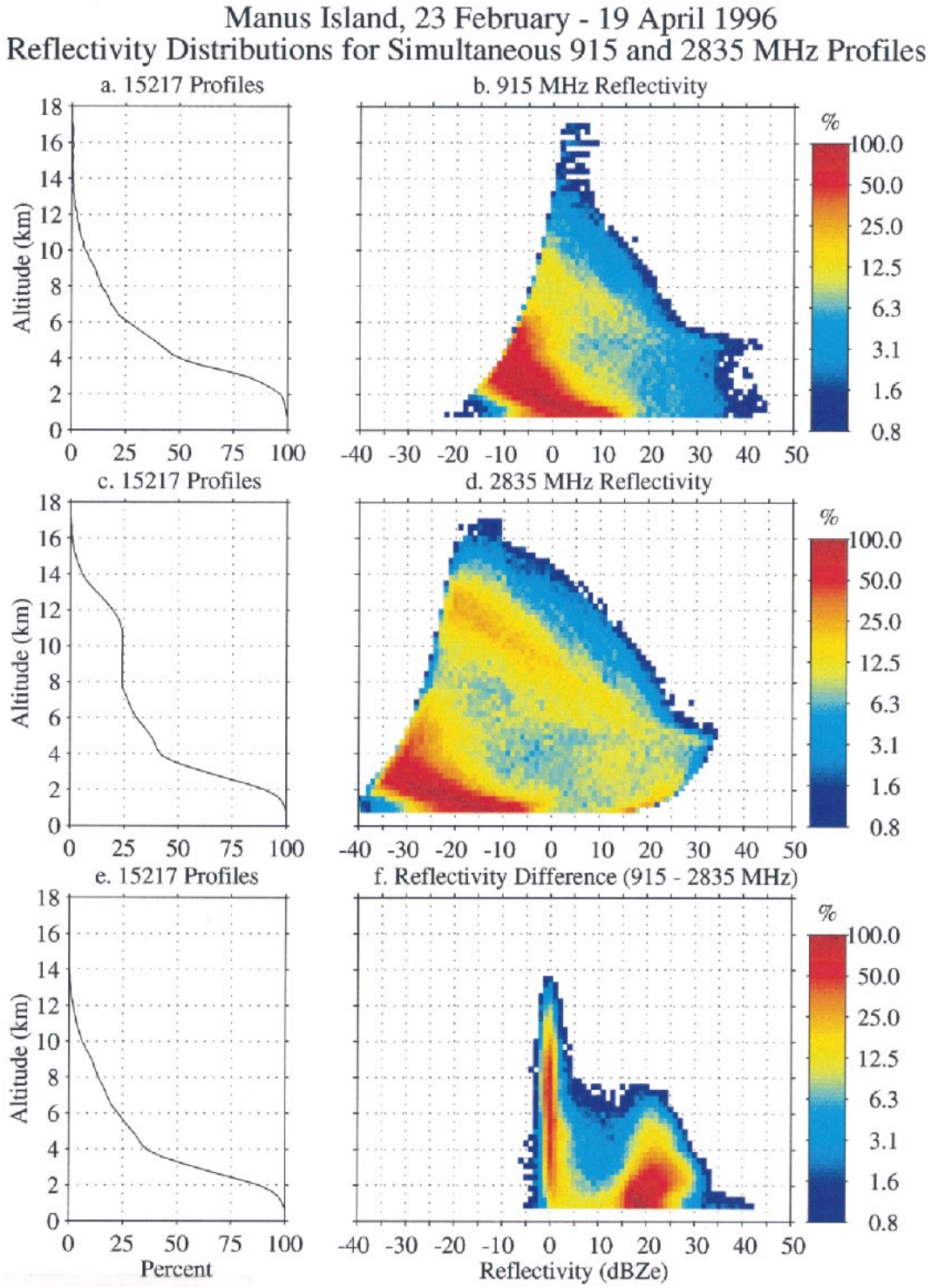

FIG. 1. Frequency distribution of simultaneous 915- and 2835-MHz observations. Occurrence at each altitude expressed as a percent of 15 217 simultaneous observations: (a) 915-MHz profiler, (c) 2835-MHz profiler, and (e) reflectivity difference (915 − 2835 MHz). Two-dimensional reflectivity distribution normalized by pixel with maximum occurrences (b) 915-MHz profiler, (d) 2835-MHz profiler, and (f) reflectivity difference (915 − 2835 MHz). The color scaling is logarithmic, and only pixels greater than 0.8% are shown.

the true scattering process, then there is a one-to-one correspondence with the terminologies.

3. Unambiguous Bragg and Rayleigh scattering identification

The first step in this study is to identify each observation as resulting from either a Bragg or Rayleigh scattering process. Since the two profilers have different sensitivities to Bragg and Rayleigh scattering, all three spectral moments from both profilers using the standard spectral moment processing routines are evaluated to ensure that the simultaneous observations result from the same scattering process. Only the observations with a difference in mean Doppler velocity less than 1 m s^{-1} and a difference in spectral width less than 0.5 m s^{-1} are retained and divided into Bragg and Rayleigh scattering groups based on their difference in reflectivity using the following criteria:

$$\text{Bragg scattering} \qquad Z_{32.8\text{ cm}} - Z_{10.6\text{ cm}} > 10 \text{ dB}.$$

$$\text{Rayleigh scattering} \qquad |Z_{32.8\text{ cm}} - Z_{10.6\text{ cm}}| \leq 2.5 \text{ dB}.$$

A binary value of ''Bragg'' or ''Rayleigh'' is assigned to each observation and is considered to represent the true scattering process of each observation. The balance of this manuscript deals only with the moments from the UHF profiler, because of the inherent redundancy of the S-band observations after the division into Bragg and Rayleigh scattering observations.

The vertical distributions of the Bragg and Rayleigh scattering observations are presented in Fig. 2. The percent occurrence of each scattering process is shown in Fig. 2a as a function of altitude. In this dataset, the Rayleigh scattering observations have the maximum number of 1716 observations at an altitude of 6 km. This is approximately 1 km above the 0°C isotherm. Note the decrease in percent occurrence with decreasing altitude, with the reduction of 25% at the altitude of 2 km, indicating the occurrence of fewer rain observations at 2 km than ice observations at 6 km. The maximum number of Bragg scattering observations occurs at 1.5 km, with 10 527 observations. Note that the Bragg percent occurrence drops to approximately 10% at 6 km, which corresponds to approximately the same number of Rayleigh scattering observations at this altitude.

The vertical distribution of Bragg scattering observations made by the UHF profiler is shown as a 2DFD in Fig. 2b in the units of refractivity turbulence structure constant C_n^2, which scales from the reflectivity factor given by Gossard (1990):

$$C_n^2 = 7.49 \times 10^{-16}\, \lambda^{-11/3} Z_{32.8\text{ cm}}. \qquad (2)$$

Near 1 km, C_n^2 has a mean value of just greater than 10^{-13}, which corresponds to a moderate refractivity turbulence index. The logarithmic color scale shows a uniform distribution with about a factor of 6 reduction in number count at 10^{-14} and 10^{-12} magnitudes. The term C_n^2 decreases with altitude approximately one order of magnitude per 1.5 km. Although the relative difference in reflectivity between two collocated profilers is used to separate the observations into Bragg and Rayleigh scattering, Fig. 2 relies on an absolute calibration estimated to be accurate to 1–2 dBZ_e.

The vertical distribution of Rayleigh scattering reflectivity shown in Fig. 2c as a 2DFD presents the wide range of reflectivities represented by precipitation processes. The occurrence of reflectivities from 5 to 35 dBZ_e below 5 km results from the predominant stratiform rain occurring at Manus Island. A few convective rain events with reflectivities greater than 40 dBZ_e do occur at Manus Island, though they are less frequent than stratiform rain events. Note that the radar bright band associated with the 0°C isotherm is identifiable near 5 km.

The 2DFDs shown in Figs. 1 and 2 indicate the vertical structure of the different scattering processes only in terms of the reflectivity. At each altitude, the Bragg and Rayleigh scattering processes have different three-dimensional distributions defined by their moments of reflectivity, mean Doppler velocity, and spectral width. Figure 3 shows the one-dimensional (1D) and two-dimensional (2D) representations of the Bragg and Rayleigh observations at the altitude of 2 km. The probability distribution functions (PDFs) show the 1D distribution of the moments and the scatterplots show the different views of the 2D distributions. Note that the spectral width PDF is repeated in Figs. 3a and 3g for symmetry and completeness. Note also that the Bragg and Rayleigh PDFs overlap for both the reflectivity and spectral width distributions, implying ambiguity in the separation of the Bragg and Rayleigh scattering observations using only one of these variables. In contrast, the Bragg and Rayleigh PDFs for the Doppler velocity distributions are more distinct and do not have as much overlap. Clearly, the 2D distribution of moments shown in the scatterplots, as well as the implied 3D distributions, define different clusters that represent the different scattering processes that are not identified as easily in the 1D probability distribution functions.

Consider the two clusters in the Doppler velocity versus reflectivity scatterplot (shown in Fig. 3d). One cluster has a positive correlation between reflectivity and downward Doppler velocity. This cluster is associated with hydrometeor observations. The correlation relates to larger drops having larger fall speeds and larger reflectivities (Ulbrich 1992). The other cluster, with near-zero mean Doppler velocity, is associated with air motions. The velocity variance of the air motion cluster is often correlated with reflectivity. At this altitude of 2.0 km, increased reflectivity and velocity variance may be associated with boundary-layer turbulence and overturning near cloud boundaries. If the separation between air and hydrometeor motions is not made correctly, a biased mean vertical air velocity will be obtained (Angevine 1997).

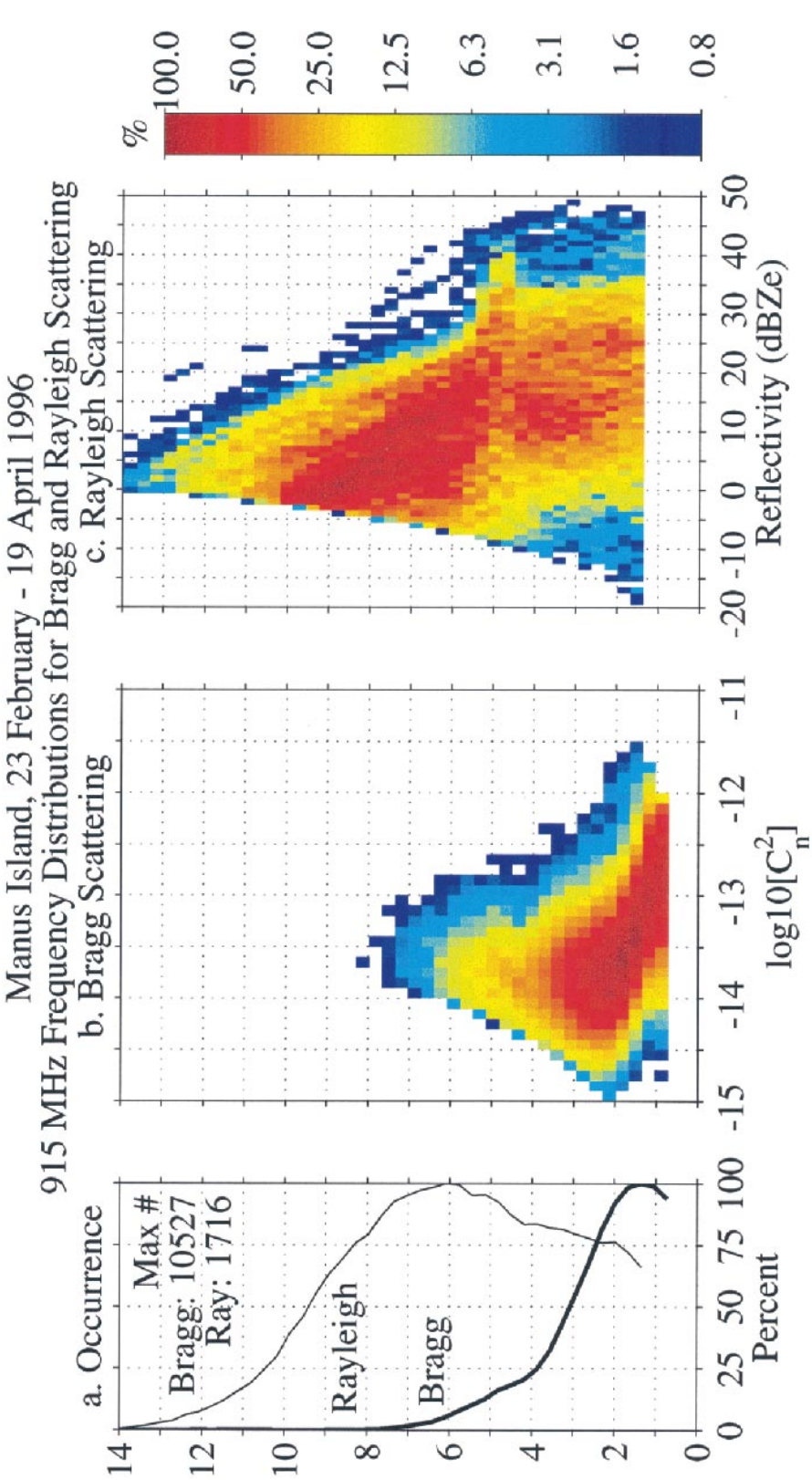

FIG. 2. Frequency distribution of 915-MHz Bragg and Rayleigh scattering observations. (a) Occurrence at each altitude expressed as a percent of a maximum of 10 527 Bragg observations and 1716 Rayleigh observations. Two-dimensional distribution of (b) Bragg scattering observations and (c) Rayleigh scattering observations. The color scaling is logarithmic, and only pixels greater than 0.8% are shown.

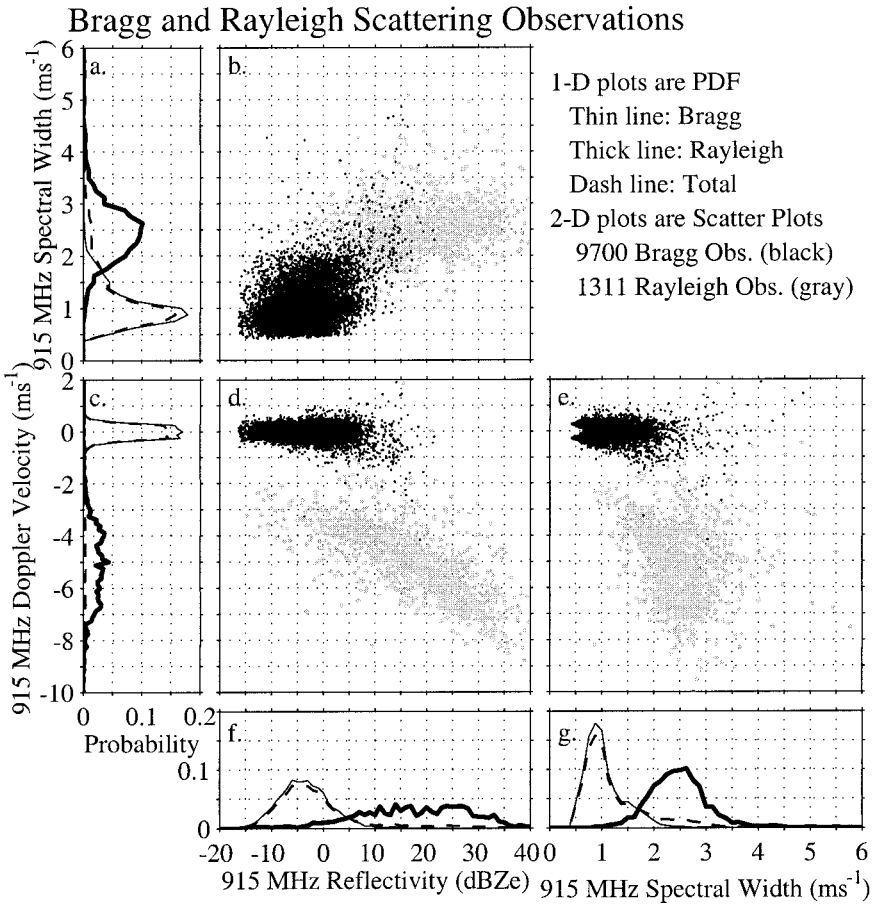

FIG. 3. Distributions of 915-MHz reflectivity, Doppler velocity, and spectral width for Bragg and Rayleigh scattering observations at 2.0 km. One-dimensional distributions normalized to unit area PDFs for Bragg (thin line), Rayleigh (thick line), and total distribution (dashed line) for [(a) and (g)] spectral width, (c) Doppler velocity, and (f) reflectivity. Scatterplot of Bragg (black dots) and Rayleigh (gray dots) in two dimensions for (b) spectral width vs reflectivity, (d) Doppler velocity vs reflectivity, and (e) Doppler velocity vs spectral width.

4. Cluster analysis using Doppler velocity and reflectivity space

As described in Ralph et al. (1996), the correlation between Doppler velocity and reflectivity of hydrometeors can be used to visually separate precipitation motions from air motions on a plot of the Doppler velocity versus reflectivity space (Fig. 3d). The difficult task is to determine a robust method to automatically perform the separation that our eyes perform with ease. The method outlined below follows target identification cluster analysis techniques.

In Fig. 3d, there are two clusters formed by the Bragg and Rayleigh scattering observations. The Bragg scattering cluster has a slight wedge shape, demonstrating velocity variance with increasing reflectivity. This reflectivity dependence prevents a constant velocity threshold to be drawn that cleanly separates the two clusters. An improved velocity threshold increases in magnitude (more negative) with increasing reflectivity. Also, above some large valued reflectivity threshold, no Bragg scattering observations exist. The cluster analysis presented herein determines the reflectivity-dependent velocity threshold and the reflectivity threshold using an objective, repeatable procedure. Neither training data nor subjective interpretations are necessary to implement this cluster analysis method.

The cluster analysis method consists of two different steps. The first step determines the velocity threshold separating the observations of similar reflectivities into two distributions. The second step is a statistical test to determine if the resulting two distributions come from one similar or two different populations. The first step determines the velocity threshold as a function of reflectivity, and the second determines the reflectivity threshold.

a. Velocity threshold determination (K-mean cluster analysis)

The velocity threshold is determined for small intervals of reflectivity. The interval of ± 5 dBZ_e is used in

this study. For each small reflectivity interval, there are typically two distributions in the Doppler velocity dimension. These two distributions correspond to the hydrometeor and turbulent air motions with similar reflectivity magnitudes. After determining the Doppler velocity threshold that separates these two distributions, the reflectivity interval is shifted to a new range, and a new velocity threshold is computed for this new reflectivity interval. Repeatedly shifting the reflectivity interval until the whole range of reflectivities are covered results in a Doppler velocity threshold as a function of reflectivity.

The velocity threshold for each reflectivity interval is determined using a constrained K-mean cluster analysis with $K = 2$. Or, in other words, two clusters of observations are determined by the mean value of the observations assigned to each cluster. We expect (and search for) only two clusters that have different mean Doppler velocities. We also expect the air motions to have a near-zero mean Doppler velocity, and the hydrometeor observations to cluster with a mean downward Doppler velocity. To ensure that the cluster analysis converges with the expected air motion cluster, the air motion cluster is forced always to have a mean Doppler velocity of zero. As an iterative process, the cluster analysis consists of five steps. These five steps are followed after each observation is initially assigned to either the air or hydrometeor cluster.

1) Determine the mean velocity of the air and hydrometeor motion clusters using all of the members assigned to each cluster.
2) Set the mean velocity of the air motion cluster to zero.
3) Determine the velocity difference between each observation and the two mean cluster velocities.
4) Reassign each observation to either the hydrometeor or the air motion cluster based on the minimum velocity difference between each observation to the two cluster mean velocities.
5) Repeat steps 1–4 until there is no change in the cluster membership.

After the cluster analysis has reached its final membership assignment, the Doppler velocity that is equidistant to both cluster mean values is assigned as the Doppler velocity threshold for this reflectivity interval. Since the air motion cluster mean value is constrained to be zero, the Doppler velocity threshold is equal to half the hydrometeor cluster Doppler mean velocity.

The Doppler velocity thresholds (solid lines) determined from the UHF observations at 2.0 and 4.8 km are presented in Fig. 4, along with the scatterplot of the Rayleigh and Bragg scattering observations. It must be emphasized that the Rayleigh and Bragg information is not used in determining the Doppler velocity threshold or the reflectivity threshold (dashed line, discussed in the next section) but is used *only* for the symbols in the scatterplot.

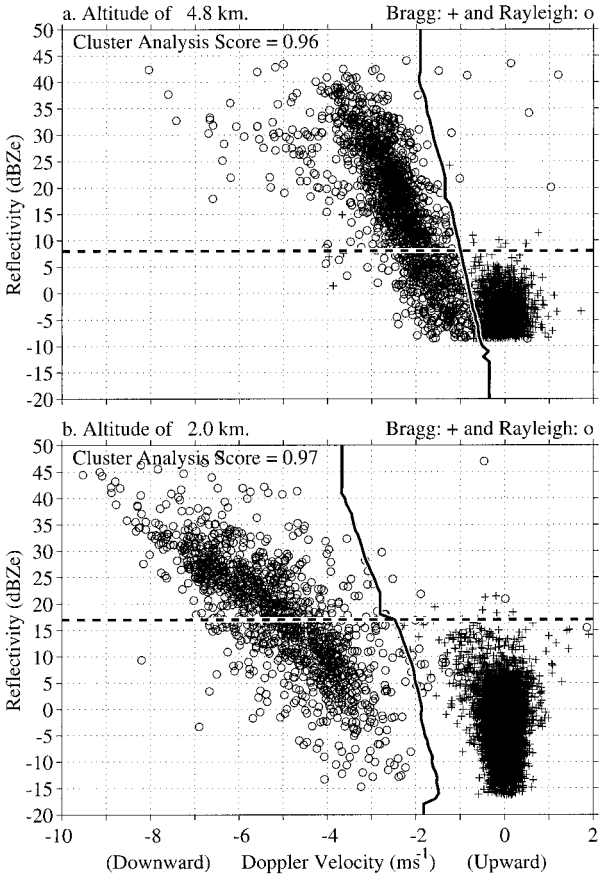

FIG. 4. Cluster analysis Doppler velocity threshold (solid line) as a function of reflectivity, and the cluster analysis reflectivity threshold (dashed line) plotted on the reflectivity vs Doppler velocity scatterplot, with the Bragg (plus) and Rayleigh (circle) observations at (a) 4.8 and (b) 2.0 km. The Bragg and Rayleigh scattering information is not used in determining the Doppler velocity or the reflectivity thresholds; it is used only for the symbols.

b. Reflectivity threshold (Kolmogorov–Smirnov statistical test)

The K-mean cluster analysis will always produce a Doppler velocity threshold that separates the observations into two clusters. This velocity threshold does not, however, imply that the observations are divided into two statistically different distributions. The K-mean cluster analysis is a robust estimator (there will be a threshold determined for every set of observations). The resulting clusters, however, need to be evaluated to determine if the two clusters represent two different distributions or are members of one common distribution. The Kolmogorov–Smirnov statistical test determines if the hydrometeor cluster probability distribution function is significantly different than the original, total PDF. The null hypothesis assumes that the hydrometeor cluster PDF equals the original PDF. If the null hypothesis is rejected at the 0.01 significance level, then the hydrometeor cluster distribution is deemed different than the original distribution. If the null hypothesis is not re-

jected at the 0.01 significance level, then the threshold Doppler velocity does not necessarily represent the division between two separate distributions. The reflectivity threshold shown in the scatterplot of Fig. 4 (dashed line) represents the maximum reflectivity interval in which there were two significant clusters.

In summary, the Doppler velocity threshold is a function of reflectivity and separates the observations into two clusters based on the observed Doppler velocity. The reflectivity threshold represents the maximum reflectivity in which the cluster analysis produces two statistically separate clusters.

5. Goodness score as a quantitative measure

The goodness score quantifies the agreement between the true scattering process, either Rayleigh or Bragg, and the single profiler estimated scattering process, either hydrometeor or air motions. The goodness score counts the number of correct single profiler estimates adjusted for the number of incorrect estimates.

The goodness score is constructed from three variables: two variables that represent the true scattering process, either Rayleigh (r) or Bragg (b), and one variable that represents the estimated scattering process, either hydrometeor (h) or air (a). Qualitatively, either estimate of the scattering process can be used in the goodness score. This study uses the hydrometeor (h) estimate. Using conditional probabilities, the goodness score is the number of Rayleigh observations identified as hydrometeor observations $N(h|r)$, adjusted by the number of Bragg observations identified as hydrometeor observations $N(h|b)$. This difference is normalized by the total number of Rayleigh observations $N(r)$ to ensure that the goodness score is less than or equal to unity. The goodness score is expressed as

$$G = \frac{N(h|r)}{N(r)} - \frac{N(h|b)}{N(r)}, \tag{3}$$

where the first term represents the favorable outcomes, and the second term represents the nonfavorable outcomes and appears as a penalty. The best score has a value of unity.

6. Objective evaluation of the constant Doppler velocity threshold separation method

This section evaluates objectively the performance of the constant Doppler velocity threshold methods by determining the goodness score for each altitude of the observations.

a. Constant velocity thresholds

The use of constant velocity thresholds to separate precipitation and air motion observations can be justified using physical reasoning (Ralph 1995). The best constant Doppler velocity thresholds will be altitude dependent. These thresholds will account for the decreases in atmospheric density with altitude and the changes in terminal fall speed of hydrometeors between their solid and liquid phases.

To evaluate the sensitivity of constant Doppler velocity thresholds to separate air and hydrometeor motions, several different thresholds were selected and applied to every altitude of the Manus Island dataset. The constant downward Doppler velocity thresholds of -0.5, -0.75, -1.5, and -3.0 m s^{-1} were selected. For each altitude and each threshold, every observation was assigned either to be a hydrometeor (h) or air motion (a) observation based on its Doppler velocity. The two terms of the goodness score G were determined and are plotted in Fig. 5. The fraction of Rayleigh scattering observations identified as hydrometeors $N(h|r)/N(r)$ is presented in Fig. 5a for each constant velocity threshold. Note the similarity between the fractions for each velocity threshold at low altitudes. For altitudes above the melting level (near 4.5 km), the performance varies between the different velocity thresholds. The large magnitude velocity thresholds (-1.5 and -3.0 m s^{-1}) are not capturing the ice and snow particles identified by the smaller magnitude velocity thresholds (-0.75 and -0.5 m s^{-1}). Figure 5b shows the fraction of Bragg scattering observations identified as hydrometeors $N(h|b)/N(r)$ or the penalty term of G. The smaller magnitude thresholds identify more Bragg scattering observations as hydrometeors at the lower heights than the larger magnitude thresholds. Figure 5c shows the total goodness score G, which is the difference of Figs. 5a and 5b. Note that the larger magnitude thresholds have better goodness scores at low altitudes. Conversely, the smaller magnitude velocity thresholds perform better at the higher altitudes. The large fall velocities of rain at the lower altitudes and the smaller fall velocities of snow and ice above the 0°C isotherm near 4.5 km explain this phenomenon.

One conclusion from this sensitivity test is that the constant Doppler velocity threshold is a good method to separate air motion from hydrometeor observations when the correct constant velocity threshold is selected. This sensitivity test does not, however, address how to determine the best constant velocity threshold at each altitude. The dual-optimization method, presented in the next section, offers a method for making that choice.

b. Dual-optimization method to determine Doppler velocity threshold

In order to achieve the best goodness score at each altitude, the Doppler velocity threshold must change with altitude to coincide with the physics of the observations. Ralph et al. (1996) proposed the dual-optimization method as a means of determining the best velocity threshold at each altitude. Assuming that the air motion and hydrometeor observations contribute to two different yet overlapping distributions in the Doppler

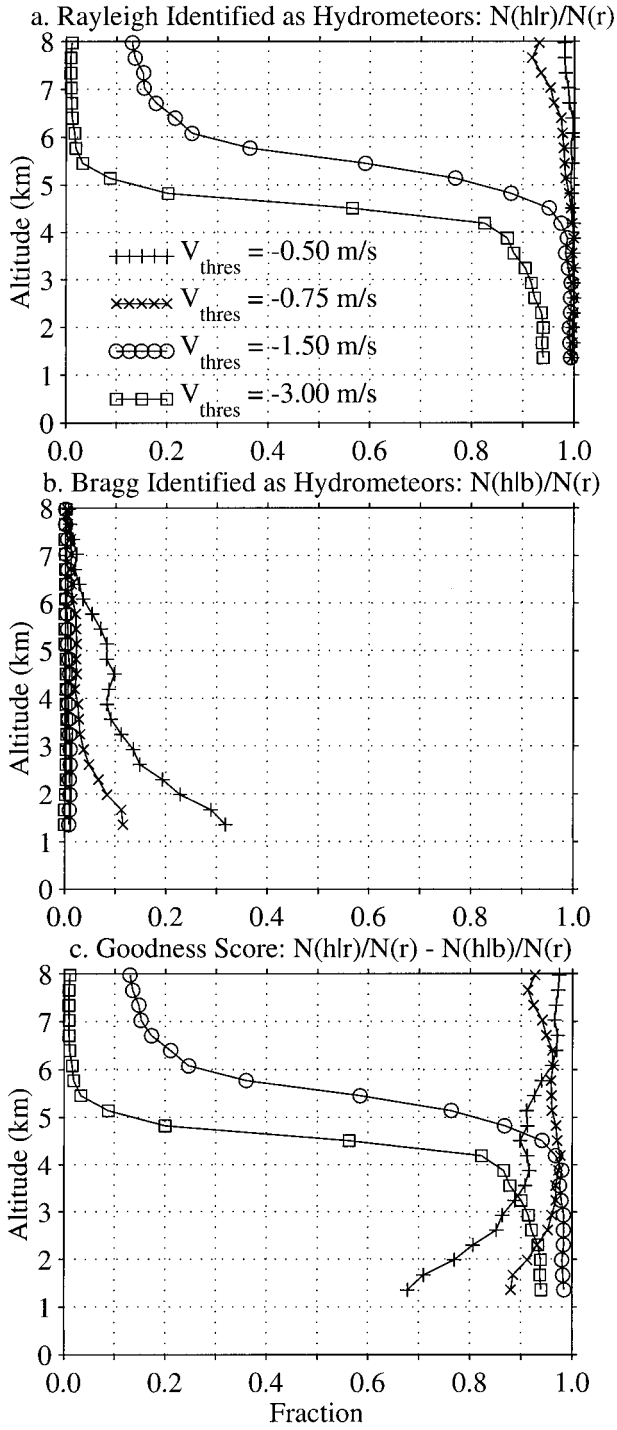

FIG. 5. Goodness score and its fractions for constant Doppler velocity thresholds. Velocity thresholds of −0.5 (plus), −0.75 (cross), −1.5 (circle), and −3.0 (square) m s^{-1}. (a) Fraction of Rayleigh scattering observations identified as hydrometeors, $N(h|r)/N(r)$. (b) Fraction of Bragg scattering observations identified as hydrometeors, $N(h|b)/N(r)$. (c) The goodness score: $G = N(h|r)/N(r) - N(h|b)/N(r)$.

velocity domain, the dual-optimization method estimates these two distributions from the observed Doppler velocity distribution. By assuming that the air motion Doppler velocity distribution is symmetric about its peak value, the portion of the air motion distribution overlapping with the hydrometeor distribution is estimated to be equal to its symmetric value. Thus, the estimated air motion Doppler velocity distribution is forced to be symmetric about the peak value. The estimated hydrometeor distribution is the difference of the total distribution and the estimated air motion distribution. Once the air motion and hydrometeor Doppler velocity distributions are estimated, then the fraction of observations misclassified as air or hydrometeor motions can be determined for each Doppler velocity threshold. The dual-optimization Doppler velocity threshold minimizes both the air and hydrometeor misclassified fractions. (See Ralph et al. (1996) for more details.)

Figure 6 shows the goodness score as well as the correct and penalty terms as a function of altitude for the dual-optimization Doppler velocity threshold (circles). The dual-optimization method does a good job identifying the Rayleigh scattering observations as hydrometeor observations $N(h|r)/N(r)$ except above 7 km. Here the method begins to fail because of the lack of Bragg scattering observations to correctly identify the air motion portion of the Doppler velocity distribution (which does not exist). The reduced performance above 7 km is not a limitation in separating the air and hydrometeor motions, because Bragg scattering is not expected to be detected at these altitudes with this profiler. The fraction of Bragg scattering observations identified as precipitation, shown in Fig. 6b, indicates that the dual-optimization method (circles) identifies a large fraction of Bragg observations as hydrometeors below 3 km. This is due to the low-magnitude Doppler velocity threshold determined by the dual-optimization misclassified fractions. This error is discussed in detail in section 7 when compared with the goodness scores determined from the cluster analysis. The goodness score, presented in Fig. 6c, indicates that the dual-optimization method (circles) estimates acceptable constant Doppler velocity thresholds above approximately 3 km. But below this altitude, the dual-optimization Doppler velocity threshold identifies too many Bragg scattering observations as hydrometeors to be a reliable method with this dataset.

7. Discussion

The results of the cluster analysis developed in section 4 are evaluated using the objective goodness score G. Figure 6 shows the goodness score as well as the correct and penalty terms as a function of altitude for the cluster analysis method (asterisks). The cluster analysis goodness score is nearly constant with altitude, indicating that the cluster analysis method adapts to the changes

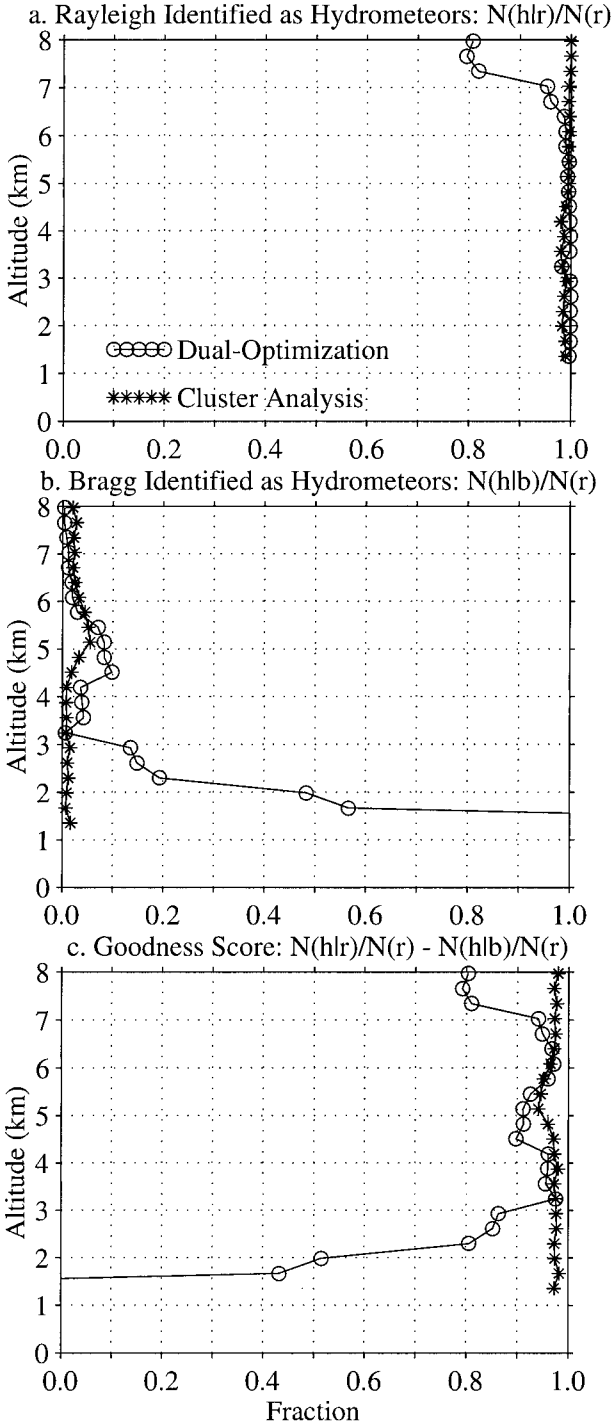

FIG. 6. The same as Fig. 4 except for the dual-optimization method (circle) and the cluster analysis method (asterisk).

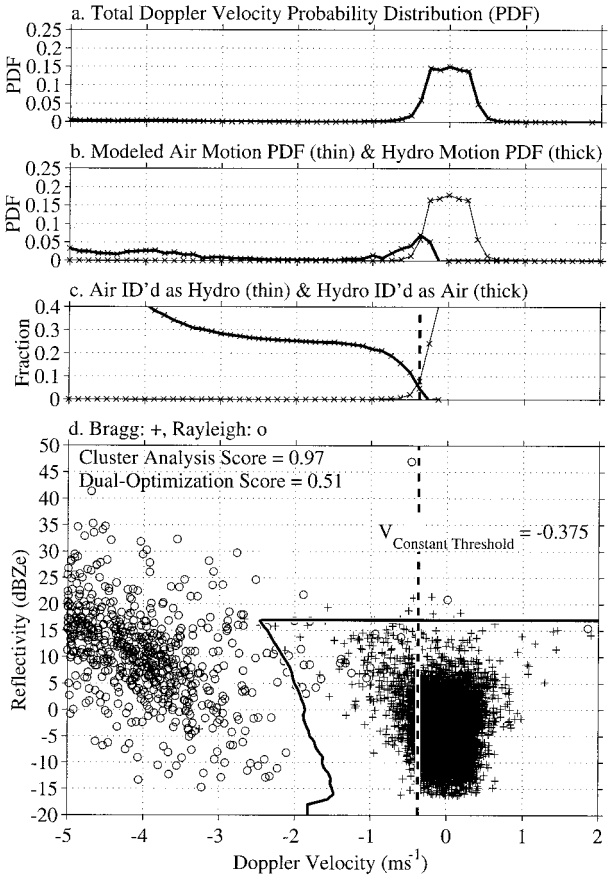

FIG. 7. Construction of dual-optimization constant Doppler velocity threshold for the observations at 2.0 km. (a) Total Doppler velocity distribution normalized to unit area (PDF). (b) Modeled air motion distribution (thin line) and constructed hydrometeor distribution (thick line) both normalized to unit area (PDF). (c) Fraction of modeled air motion observations identified as hydrometeors (thin line) and fraction of modeled hydrometeor observations identified as air motion (thick line), as a function of Doppler velocity threshold. (d) Cluster analysis thresholds (solid) and dual-optimization constant Doppler velocity threshold (dash) on the Bragg (plus) and Rayleigh (circle) scatterplot.

in the altitude-dependent Doppler velocity–reflectivity relationships. These altitude-dependent changes result from changes in atmospheric density and different Doppler terminal fall speeds associated with liquid and frozen water particles.

As demonstrated in Fig. 6b, the cluster analysis (asterisks) and dual-optimization (circles) methods produce different fractions of Bragg scattering observations identified as hydrometeors below approximately 3 km. This difference must be understood in order to evaluate the performance of both methods. The dual-optimization method is presented step-by-step in Fig. 7 using the observations from the altitude of 2.0 km. The total Doppler velocity distribution is presented as a PDF for the limited velocity range of -5 to $+1$ m s^{-1}. Though the resolution of 0.125 m s^{-1} is evident in Fig. 7a, the results are not sensitive to the level of quantization. The next step in the dual-optimization method estimates the air motion distribution by assuming that the distribution is symmetric about the peak value in the distribution, which is at zero Doppler velocity for this example. Figure 7b shows the modeled symmetric air motion dis-

tribution normalized to unit area. The hydrometeor distribution is constructed by taking the difference of the observed distribution and the modeled air motion distribution, and is also shown in Fig. 7b normalized to unit area. Using the normalized air and hydrometeor distributions, the fraction of air (hydrometeor) observations misclassified as hydrometeor (air) observations can be calculated for every possible Doppler velocity threshold. The misclassified fractions for this example are shown in Fig. 7c. The dual-optimization Doppler velocity threshold is defined where these two fractions are equal, which is at -0.375 m s^{-1}, for this example.

The dual-optimization constant Doppler velocity threshold of -0.375 m s^{-1} is plotted on the scatterplot of Fig. 7d, along with the cluster analysis Doppler velocity and reflectivity thresholds. The Rayleigh and Bragg information is used only to determine the symbols plotted in Fig. 7d and not to estimate the thresholds. Note that the constant dual-optimization Doppler velocity threshold intersects the cluster of Bragg scattering observations. This intersection results in the increase in the fraction of Bragg scattering observations identified as hydrometeors presented in Fig. 6b (circles). Examination of the dual-optimization Doppler velocity thresholds at all altitudes below 3 km indicates that the magnitude of the thresholds are too low to adequately separate the Bragg and Rayleigh scattering observations. This result may be related to the sensitivity of the 915-MHz profilers to refractive gradient turbulence compared with the 404-MHz profilers, from which data were used to develop the dual-optimization Doppler velocity threshold in Ralph et al. (1996). Since the true scattering process is known to be Bragg scattering, the increase in velocity variance with reflectivity is not due to biological targets (e.g., insects) but to Bragg scattering processes.

8. Concluding remarks

This paper focuses on evaluating and applying the techniques developed by Ralph (1995) and Ralph et al. (1996) using the high-power 404-MHz profilers operating in the central United States to observations made by low-power 915-MHz profilers operating in the Tropics. The 404-MHz profilers can observe Bragg scattering routinely up to 10 km above ground level, while the 915-MHz profilers routinely observe Bragg scattering up to 5 km (Carter et al. 1995). The difference in profiler sensitivity to Bragg scattering is a factor in the evaluations presented in this work, and it reflects the difficulties of applying techniques developed on one platform to another platform. The analysis presented here extends the work of Ralph et al. (1996). It is a necessary prerequisite to the application of the Ralph et al. (1996) techniques to tropical observations. Likewise, the techniques presented herein must be objectively evaluated using 404-MHz observations before general application to those data.

As is described in Gage et al. (1999), using collocated profilers operating at different frequencies allows the unambiguous identification of Bragg and Rayleigh scattering processes. The separation of the scattering processes is critical for proper interpretation of the observations. Having two profilers operating side by side is not, however, the normal mode of deploying profilers. So, methods that identify the scattering processes using data from only one profiler must be developed and independently evaluated.

One objective of this study was to evaluate the performance of constant Doppler velocity thresholds to separate air motions from hydrometeor motions in vertical incident profiler observations. This evaluation consisted of three different steps. First, the observations were unambiguously identified as resulting from Bragg or Rayleigh scattering by using two collocated profilers operating at different frequencies. The second step consisted of separating the observations into either air motions or hydrometeor motions using only the data from one profiler. The last step quantitatively evaluated the performance of the separation techniques by counting the number of correct classifications and adjusting this count by the number of incorrect classifications.

It was shown that the constant Doppler velocity threshold is an acceptable method to separate air from hydrometeor motions, if the correct threshold is selected. Choosing the correct Doppler velocity threshold is the dilemma at hand. The dual-optimization method introduced by Ralph et al. (1996) was a good method to determine the constant Doppler velocity thresholds between the altitudes above 3 km in this dataset. The dual-optimization method accounted for the transition region from liquid to frozen hydrometeors near the melting level. Below 3 km, the dual-optimization method determined velocity thresholds that were too small in magnitude, causing more Bragg scattering observations to be identified incorrectly as hydrometeor observations. This error was due to the assumed symmetrical shape of the air motion Doppler velocity distribution. This assumption may have been true for the midlatitude 404-MHz observations used in developing the dual-optimization method but was not true with the 54 days of tropical 915-MHz observations at Manus Island. The 915-MHz air motion observations had a large velocity variance and a skew to more downward motions at larger reflectivities. Since the two-frequency information was used to determine that these observations resulted from Bragg scattering processes, this skew in Doppler velocity relates to turbulent processes and not to backscattered energy from Rayleigh scatterers. Thus, Rayleigh scatterers—for example, insects—are not possible explanations for the asymmetric distribution of the Bragg scattering observations. The asymmetric frequency of occurrence of the larger reflectivity Bragg scattering observations, more downward than upward, is consistent with the larger refractive index induced by the downward phase of vertically propagating gravity

waves (Nastrom and VanZandt 1994). Nastrom and VanZandt's theory states that, on a statistical basis, the downward-moving air phase of vertically propagating gravity waves has larger reflectivities then the upward phase.

Another objective of this study was to present a cluster analysis method that separates robustly and objectively air from hydrometeor motions. The cluster analysis identified the clusters, or regions, of air and hydrometeor motions in the two-dimensional reflectivity–Doppler velocity domain. The cluster analysis impartiality adapts to the distributions embedded in the dataset. The introduced cluster analysis performs better than the accepted dual-optimization method at low altitudes, and has similar results at higher altitudes. The cluster analysis is able to distinguish between the air and hydrometeor motions even with the increased Doppler velocity variance, with increasing reflectivity observed in the Bragg scattering distributions. The cluster analysis uses the Kolmogorov–Smirnov statistical test to ensure that the determined Doppler velocity threshold separates the observations into two statistically different distributions. If the two distributions are not statistically different, then a reflectivity threshold is defined, indicating that only one distribution is present at all reflectivities greater than this threshold. The cluster analysis is a repeatable and a statistically sound procedure.

The cluster analysis method divides the observations into either hydrometeor or air motion observations by examining the distributions in the reflectivity–Doppler velocity domain. Observations with reflectivities greater than the reflectivity threshold are identified as hydrometeor observations, independent of the Doppler velocity and independent of the ambient air motion (e.g., updrafts or downdrafts). The observations with reflectivities less than the reflectivity threshold are divided into air or hydrometeor motion observations based on the measured Doppler velocity. This division is determined using small reflectivity intervals to account for the observed increased Doppler velocity variance with reflectivity in the air motion observations. Ambient air motions will affect the success of separating equal reflectivity air and hydrometeor motion observations. The ambient air motion contamination will cause the (hypothetical) quiet air clusters to have larger Doppler velocity variance, leading to the merging of the two clusters. The cluster analysis would still produce velocity thresholds that separate the observations into two clusters. The well-defined and separated Bragg and Rayleigh scattering clusters, which account for updrafts and downdrafts, represented in the reflectivity–Doppler velocity domain shown in Figs. 3, 4, and 7, suggest that the ambient air motion contamination is not a frequent phenomenon in this dataset.

The cluster analysis presented in this study, as well as the dual-optimization method, does not include the time–height continuity of the profiler observations. The observations at each height are analyzed in a statistical manner separately. Both analysis methods are statistical in nature and should be viewed as inputs to a larger, more detailed model that identifies and separates the observations into air and hydrometeor observations. These larger models could be neural networks or fuzzy logic models that incorporate many tests to reach a final separation conclusion. Advanced models should also incorporate the complete Doppler velocity spectra to identify the Bragg and Rayleigh scattering components that are resolved in the spectra.

Acknowledgments. We thank the Weather Service of Papua New Guinea for assistance with maintaining the day-to-day operations of the profiler site at Manus Island. The research reported herein was supported in part by the National Science Foundation under Grant ATM-9214800, and in part by the NOAA Office of Global Programs and DOE ARM.

REFERENCES

Angevine, W. M., 1997: Errors in mean vertical velocities measured by boundary layer wind profilers. *J. Atmos. Oceanic Technol.,* **14,** 565–569.

Carter, D. A., K. S. Gage, W. L. Ecklund, W. M. Angevine, P. E. Johnston, A. C. Riddle, J. Wilson, and C. R. Williams, 1995: Developments in UHF lower tropospheric wind profiling at NOAA's Aeronomy Laboratory. *Radio Sci.,* **30,** 977–1001.

Ecklund, W. L., C. R. Williams, P. E. Johnston, and K. S. Gage, 1999: A 3-GHz profiler for precipitating cloud studies. *J. Atmos. Oceanic Technol.,* **16,** 309–322.

Gage, K. S., 1990: Radar observations of the free atmosphere: Structure and dynamics. *Radar in Meteorology,* D. Atlas, Ed., American Meteorological Society, 534–565.

——, C. R. Williams, and W. L. Ecklund, 1994: A new tool for diagnosing tropical convective cloud systems. *Bull. Amer. Meteor. Soc.,* **75,** 2289–2294.

——, ——, ——, and P. E. Johnston, 1999: Use of two profilers during MCTEX for unambiguous identification of Bragg scattering and Rayleigh scattering. *J. Atmos. Sci.,* **56,** 3679–3691.

Gossard, E. E., 1990: Radar research on the atmospheric boundary layer. *Radar in Meteorology,* D. Atlas, Ed., American Meteorological Society, 477–527.

Hartten, L. M., 1998: Reconciliation of surface and profiler winds at ISS sites. *J. Atmos. Oceanic Technol.,* **15,** 826–834.

Hill, R. J., 1978: Spectra of fluctuations in refractivity, temperature, humidity, and the temperature–humidity cospectrum in the inertial and dissipation ranges. *Radio Sci.,* **13,** 953–961.

May, P. T., R. G. Strauch, and W. L. Ecklund, 1990: Temperature sounding by RASS with wind profiler radars: A preliminary study. *IEEE Trans. Geosci. Remote Sens.,* **28,** 19–28.

Nastrom, G. D., and T. E. VanZandt, 1994: Mean vertical motions seen by radar wind profilers. *J. Appl. Meteor.,* **33,** 984–995.

Parsons, D., and Coauthors, 1994: The Integrated Sounding System: Description and preliminary observations from TOGA COARE. *Bull. Amer. Meteor. Soc.,* **75,** 553–567.

Ralph, F. M., 1995: Using radar-measured radial vertical velocities to distinguish precipitation scattering from clear-air scattering. *J. Atmos. Oceanic Technol.,* **12,** 257–267.

——, P. J. Neiman, and D. Ruffieux, 1996: Precipitation identification from radar wind profiler spectral moment data: Vertical velocity histograms, velocity variance, and signal power—Vertical velocity correlations. *J. Atmos. Oceanic Technol.,* **13,** 545–559.

Rogers, R. R., and W. O. J. Brown, 1997: Radar observations of a major industrial fire. *Bull. Amer. Meteor. Soc.,* **78,** 803–814.

Riddle, A. C., W. M. Angevine, W. L. Ecklund, E. R. Miller, D. B.

Parsons, D. A. Carter, and K. S. Gage, 1996: In situ and remotely sensed horizontal winds and temperature intercomparisons obtained using Integrated Sounding Systems during TOGA COARE. *Contrib. Atmos. Phys.,* **69,** 49–61.

Steiner, A., and H. Richner, 1994: Separation of clear-air echoes from precipitation echoes in UHF wind profiler measurements. *Ann. Geophys.,* **12,** 497–505.

Ulbrich, C. W., 1992: Algorithms for determination of rainfall integral parameters using reflectivity factor and mean Doppler fall speed at vertical incidence. *J. Atmos. Oceanic Technol.,* **9,** 120–128.

Yuter, S. E., and R. A. Houze Jr., 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part II: Frequency distributions of vertical velocity, reflectivity, and differential reflectivity. *Mon. Wea. Rev.,* **123,** 1941–1963.