

A Fuzzy Logic Method for Improved Moment Estimation from Doppler Spectra

LARRY B. CORNMAN

National Center for Atmospheric Research, Boulder, Colorado

ROBERT K. GOODRICH

Department of Mathematics, University of Colorado, Boulder

CORINNE S. MORSE

National Center for Atmospheric Research, Boulder, Colorado

WARNER L. ECKLUND

Boulder, Colorado

(Manuscript received 4 June 1997, in final form 2 January 1998)

ABSTRACT

A new method for estimating moments from wind measurement devices that measure Doppler spectra as a function of range is presented. Quite often the spectra are contaminated by a wide variety of sources, including (but not limited to) birds, aircraft, velocity and range folding, radio frequency interference, and ground clutter. These contamination sources can vary in space, time, and even in their basic characteristics. Human experts analyzing Doppler spectra can often identify the desired atmospheric signal among the contamination. However, it is quite difficult to build automated algorithms that can approach the skill of the human expert. The method described here relies on mathematical analyses, fuzzy logic synthesis, and global image processing algorithms to mimic the human expert. Fuzzy logic is a very simple, robust, and efficient technique that is well suited to this type of feature extraction problem. These new moment estimation algorithms were originally designed for boundary layer wind profilers; however, they are quite general and have wide applicability to any device that measures Doppler spectra as a function of range (e.g., lidars, sodars, and weather radars).

1. Introduction

To fully exploit the utility of Doppler measurement devices for the real-time measurement of winds, wind shear, and turbulence, it is important that the data from these devices be quality controlled. These devices include lidars, sodars, weather radars, and wind profilers. Depending on the source of the contamination, quality control must be applied at the time series, spectral, moment, or wind estimation level. The contamination sources are quite varied, including but not limited to stationary and moving ground clutter, sea clutter, low signal-to-noise ratio, radio frequency interference (RFI), velocity and range folding, and aircraft and other moving point targets (e.g., birds). It is unlikely that a single method can solve all of the quality control problems that these contamination sources generate, hence a com-

bination of methods applied sequentially from the time series level to the wind estimation is required.

This paper describes a new algorithm [the National Center for Atmospheric Research Improved Moment Algorithm (NIMA)] that can often produce accurate moments even when there is a significant amount of spectral-level contamination. The techniques are quite general in that they only assume that the measurement device produces Doppler spectra as a function of range. However, a specific application to Doppler wind profilers is the main focus of this work. NIMA is a fully automated, real-time algorithm that attempts to mimic the feature detection processing of a human expert. One of the main features of NIMA is its fuzzy logic infrastructure. It is a highly parallel and adaptable methodology; modules can be altered, added, or removed with ease. The unique aspect of fuzzy logic techniques is the retention of maximal information through the composition step. That is, there is no arbitrary thresholding prior to the composition step and hence "subcritical" though still useful information can be used in the final decision process.

Corresponding author address: Larry Cornman, NCAR/RAP, PO Box 3000, Boulder, CO 80307-3000.
E-mail: cornman@ucar.edu

NIMA is based upon a combination of mathematical analysis, fuzzy logic techniques, and image processing methods. The general idea is to break down the Doppler velocity–range coordinate space into local elemental units wherein various mathematical quantities (e.g., gradient and curvature) are computed. These data are then synthesized using a fuzzy logic technique to generate a local, composite *membership* value, which indicates how similar the data in the given region are to the properties of the desired signal. Once these localized computations are completed, global image processing methods are then applied to extract the desired portion of the Doppler spectra: the atmospheric wind signature region. The Doppler moments are then calculated from the appropriately identified atmospheric signal.

In the past, wind profiler quality control efforts have been focused at the moment and wind estimation levels (Brewster 1989; Wilfong et al. 1993; Weber et al. 1993). Unfortunately, if the contamination has occurred before the moments or winds are estimated via the standard methods, the data can be severely corrupted and these quality control methods can only alleviate some of the problems. Recently, wind profiler quality control techniques using spectral-level and time series–level data have been developed (Clothiaux et al. 1994; Merritt 1995; Jordan et al. 1997). While these methods are successful in dealing with certain types of contamination, there is still a need for more systematic and broadly applicable techniques. It is hoped that NIMA is a constructive step in this direction.

2. Addressing the moment estimation problem

If Doppler wind profilers could provide rapid (1–3 min) estimates of winds, wind shear, and turbulence, then they would have a much greater applicability in automated, real-time environments. Satisfying these requirements requires that two problems be simultaneously addressed: improving the accuracy of the Doppler moments and increasing the update rate and accuracy of the wind and turbulence estimates.

The most commonly used moment-finding technique utilizes the peak of the spectra as the starting point for finding the atmospheric signal region (Strauch et al. 1984; Brewster 1989; Carter et al. 1995). However, if the peak is associated with noise or contamination, the resulting moments will be erroneous.

The atmospheric signal is clearly visible on the right-hand side of Fig. 1; however, the peak-finding algorithm has selected portions of the ground clutter signal (around zero velocity) and a point target (left-hand side at 700–900-m range). Furthermore, existing quality control techniques have generally been used in postprocessing applications. The fully automated techniques described here are capable of approaching the skill level of the human expert and thus are suitable for some real-time applications. The approach taken in this work relies on local mathematical analysis combined with fuzzy logic

and global image-processing techniques to simulate the methodology of the experts.

a. Fuzzy logic algorithms

Fuzzy logic has evolved into a very useful tool for solving complex, real-world problems. Fuzzy logic is well suited to applications in linear and nonlinear control systems, signal and image processing, and other data analysis problems (Klir and Folger 1988; Kosko 1992). The pioneering work in this area was done by L. Zadeh beginning in 1965 (Yager et al. 1987; Klir and Yuan 1996). The strength of fuzzy logic algorithms lies in their ability to systematically address the natural ambiguities in measurement data, classification, and pattern recognition. While fuzzy logic algorithms have been successfully applied in the engineering sciences, the use of these techniques in the atmospheric sciences has been fairly limited (Fujibe 1989; Delanoy and Troxel 1993; Albo 1994). The fuzzy logic techniques used in this work follow that of Delanoy and Troxel (1993) and Albo (1994). The reader should refer to those works for a more complete description of these techniques. Due to the inherent ambiguity in many aspects of atmospheric data measurement, analysis, detection, and forecasting algorithms, fuzzy logic has the potential to become a very useful tool in this field.

There are a variety of ways that fuzzy logic can be implemented, including fuzzy neural networks, fuzzy expert systems, and fuzzy inference systems. The current implementation is a variant of a fuzzy inference system and builds on two basic processes: *fuzzification* and *composition*. The first step, fuzzification, performs the conversion of measurement data into scaled, unitless numbers that indicate the correspondence or “membership level” of the data to the desired feature. The composition step combines the membership values from a number of different data types in a systematic fashion. The correspondence between a data value and the degree to which that data belongs to a certain class is quantified by the application of a prescribed functional relation or *membership function*.

b. Global processing

It is important to realize that the application of the membership functions and the composition step is done at the point level. A task still remains—joining the point data into global features. There are a number of image-processing techniques applicable to this type of feature extraction and characterization. A method that has been successfully applied to weather radar images is the “linear chain” algorithm (Rogers et al. 1991). In this method, point data are thresholded, combined, and then classified as distinct, spatially contiguous features. In this application, the “point data” are the total membership values.

Another image-processing technique is applied to

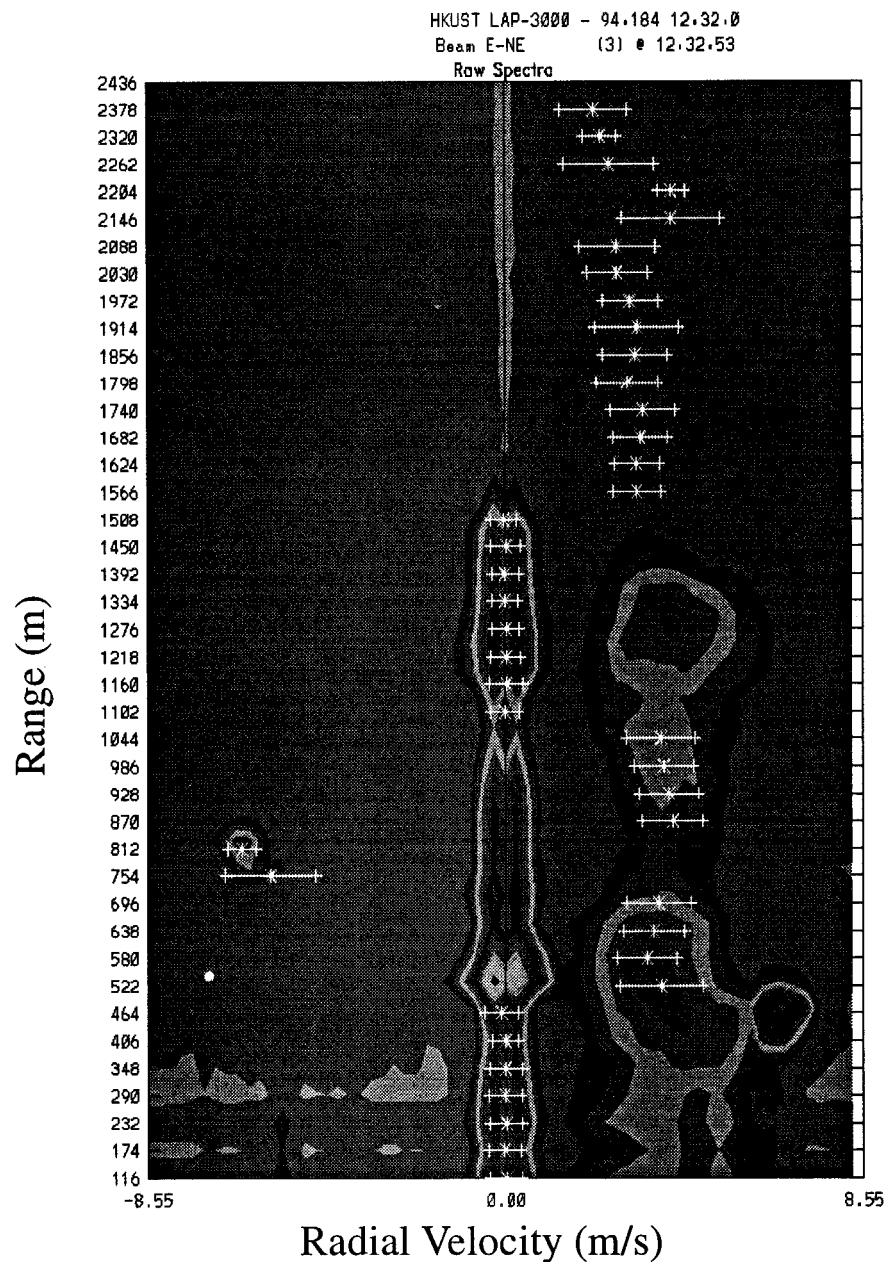


FIG. 1. Contour plot of Doppler spectra (dB) as a function of radial velocity and range. A clearly discernible ground clutter signal is centered at zero radial velocity, a point target (perhaps a bird) can be seen on the left-hand side at 700–900 m, and the atmospheric signal can be seen on the right-hand side of the figure. The stars are located where the peak-finding algorithm has computed the first moments and the plus marks are located at \pm the spectrum width (square root of the second moment).

the membership values before the feature-building step. This procedure, *density weighting*, is used to sharpen distinct regions by weighting each point by the (normalized) number of neighboring points whose total membership value is above a specified threshold. This technique does a good job of removing lower-

intensity points on the boundary of a feature since their initial values are weighted (multiplied) by numbers that are less than one. Mathematically, this can be codified for a general $n \times m$ rectangle with p points above the threshold. If the original value at the central point is A , then its density weighted value is $A(p/nm)$.

3. The NCAR improved moment algorithm (NIMA)

a. General process

The first step in analyzing the spectral data is to replace the one-dimensional range-gated spectra (the so-called stacked spectra) with a two-dimensional image. At each range, the Doppler spectra can be considered a two-dimensional curve, $g(x)$, where x is the radial velocity and $g(x)$ is the signal power at that velocity. The collection of these two-dimensional curves as a function of range can be thought of as describing a three-dimensional surface, $g(x, y)$. That is, $g(x, y)$ is the signal power at radial velocity x and range y . Figure 1 is an example of this surface (in a contour map presentation). As the coordinate axes do not have the same units, a coordinate scaling is applied: $\gamma_V \Delta V = \Delta x$; $\gamma_R \Delta R = \Delta y$, where γ_V and γ_R are the scale factors for Doppler velocity V and range R , respectively. The signal power is scaled in a variety of ways (e.g., linear, power law, or logarithmic scaling) in order to accentuate certain characteristics of the data. This can be quite useful when the data have a large dynamic range.

The two-dimensional image is segmented into a collection of overlapping subregions. For each of the subregions, the data are first filtered and then a number of local mathematical quantities (e.g., gradient and curvature) are computed. The computational method used in this stage of the processing is a two-dimensional least squares quadratic analysis. The derived mathematical quantities and other information from the data [e.g., signal power or signal-to-noise ratio (SNR)] are then combined into a single quantity at each given point using fuzzy logic techniques and then a threshold is applied. Finally, these local data values are synthesized into distinct, global features using image-processing methods. A large variety of features can be created: birds, aircraft, radio frequency interference, clutter, and, of course, atmospheric. The nonatmospheric features are masked out, leaving a set of candidate atmospheric features. Typically, there will be a set of candidate atmospheric features and hence further processing is required to select the correct one. A fuzzy algorithm aimed at identifying the atmospheric features is then applied, resulting in a "atmospheric feature score." The fields used in this step include range coverage and composite membership value of all points in a given feature. The highest scoring feature is assumed to be the atmospheric feature. It should be noted that there will be instances in which the atmospheric signal is so obscured by contaminants or low SNR, that even a human expert will be unable to identify the atmospheric signal. In this case, the algorithm will still designate the highest scoring feature as the atmospheric signal; however, an associated quality control metric will indicate the problem with the feature identification (see the discussion on "confidence indices" below).

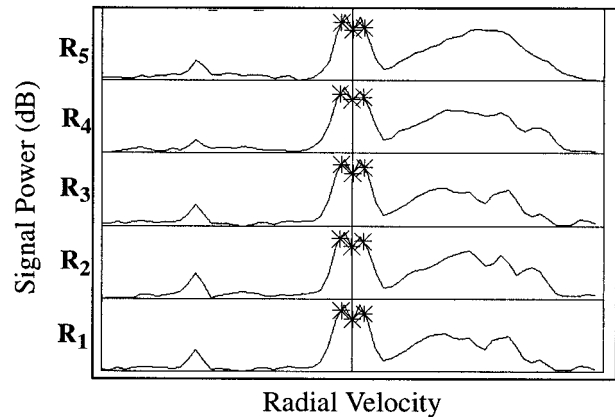


FIG. 2. Example of ground clutter signal (around zero) and atmospheric return (on right-hand side) as a function of radial velocity and range.

b. An example problem: Extracting the atmospheric signal in the presence of ground clutter

The algorithm concepts introduced above can be more easily understood in the following simplified example. Ground clutter can be a serious source of data contamination, appearing as a large-amplitude signal centered at zero velocity and extending for a number of ranges above the ground. Figure 2 illustrates (in stacked-spectra format) the first five ranges from Fig. 1 and shows the prototypical features of a ground clutter signal.

Mathematically, the ground clutter signal can be described as a set of points that 1) are close to symmetric around zero, 2) have large slopes along the radial velocity axis (positive on one side of zero and negative on the other), 3) have small slopes along the range axis, 4) have a large negative curvature along the velocity axis (in the proximity of zero radial velocity), and 5) have small curvatures along the range axis. The ground clutter feature can be visualized as a ridge line parallel to the range axis and centered at zero radial velocity. Unfortunately, all of the above-mentioned characteristics of the ground clutter signal are often quite similar to the atmospheric signal. They tend to differ only in magnitude: ground clutter signals have "large" slopes and curvatures, whereas atmospheric signals have "not-too-large" slopes and curvatures. Problems of this type, where the desired features are similar in structure to a contamination source and vary only in degree, are well suited to fuzzy logic techniques. Furthermore, on the edges of the clutter signal the slopes might be not-too-large, that is, identical to those associated with the atmospheric signal. At those locations, the point-by-point total membership values may not provide enough information to be able to differentiate the clutter from the atmospheric signal, and hence global feature extraction techniques are required.

In practice, a number of derived fields from a local quadratic surface fit and other characteristics of the data

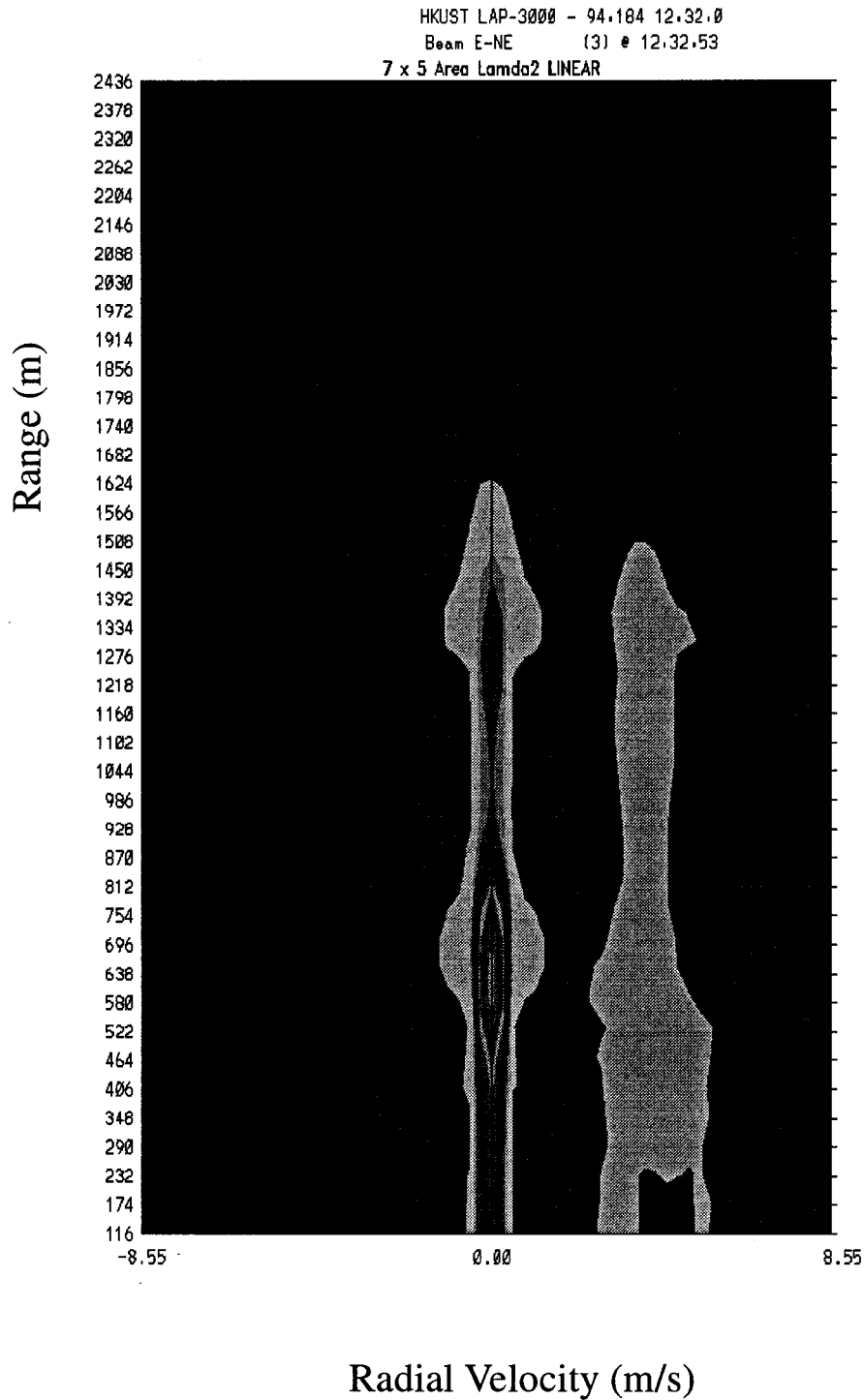


FIG. 3. Contour plot of the curvature field for the spectral data in Fig. 1.

are used. For illustration purposes, only a small subset of these fields are used in this example. Figure 3 illustrates a contour plot of the curvature field for the data in Fig. 1.

As expected, the ground clutter exhibits large cur-

vature values near zero radial velocity, which is quite distinct from the curvatures associated with the atmospheric signal. However, it is also apparent that there are regions in both features where the curvatures are similar. These data illustrate a strong ground clutter sig-

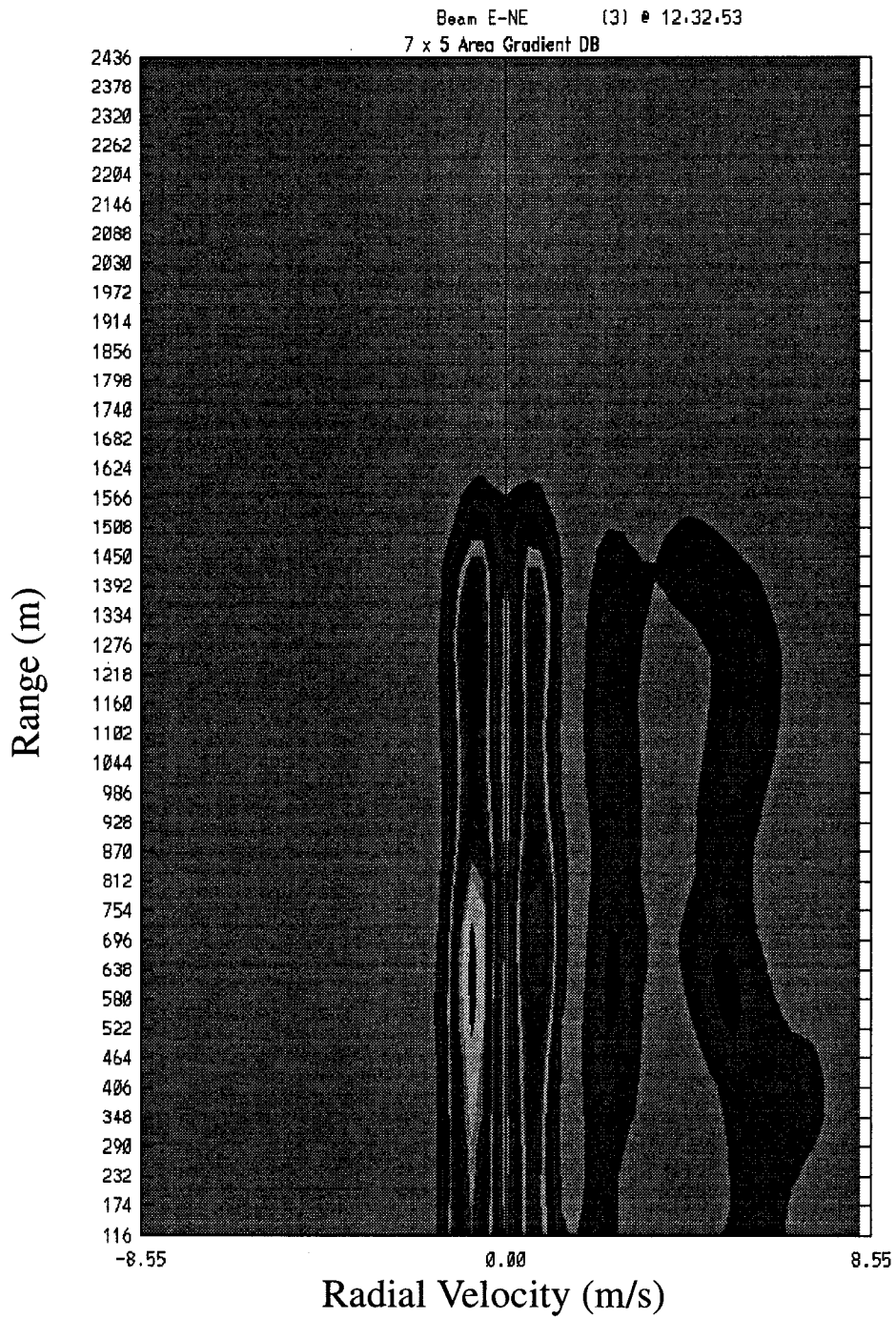


FIG. 4. Contour plot of the gradient field for the data shown in Fig. 1.

nal that is well separated from the atmospheric signal. Unfortunately, this is not always the case, and hence differentiating between these two types of signals can be problematic. Moreover, in the limit of overlaying signals, differentiation may be impossible. However, the resultant error in the calculation of the first moment will not be large in this case.

Figure 4 illustrates a contour plot of the magnitude

of the local gradient vector. Analogous to the curvature field, the ground clutter signal exhibits much larger gradient values than the atmospheric signal, although there are regions that have similar values.

The curvature and gradient fields for the atmospheric signal and ground clutter are complementary, that is, the regions where the curvatures are large are where the gradients are small, and vice versa (cf. Figs. 3 and 4).

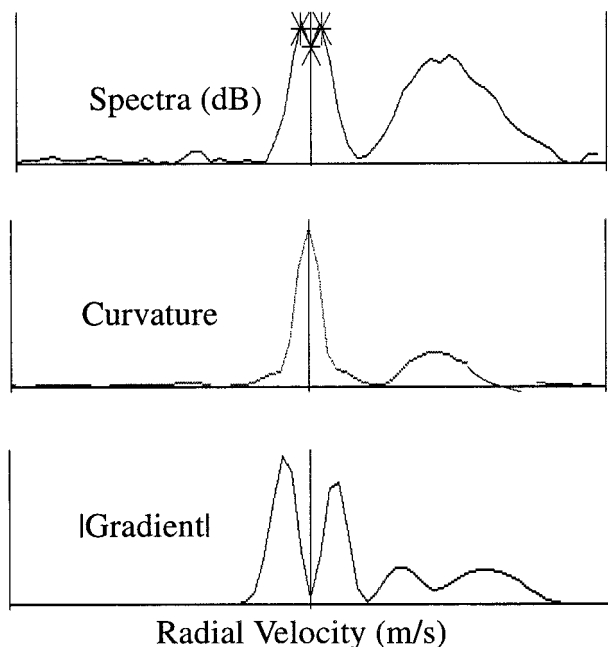


FIG. 5. Top to bottom: spectra, curvature, and gradient fields at 1334-m range for the example case.

This implies that neither field by itself is sufficient to identify the full extent of the ground clutter or atmospheric feature. This situation is clearly illustrated in Fig. 5. The top panel shows the raw spectra (in decibels) from the 1334-m range. The ground clutter signal (centered around zero velocity) and atmospheric signal (on the right-hand side) are easily identified. The middle panel shows the curvature field, and the bottom panel shows the gradient field. For the atmospheric feature and especially for the ground clutter feature, it can be seen how the two derived fields supplement each other and why both fields are required to fully characterize the features. The combination of these two disparate fields, the curvature and the gradient, is accomplished through the use of the fuzzy logic membership func-

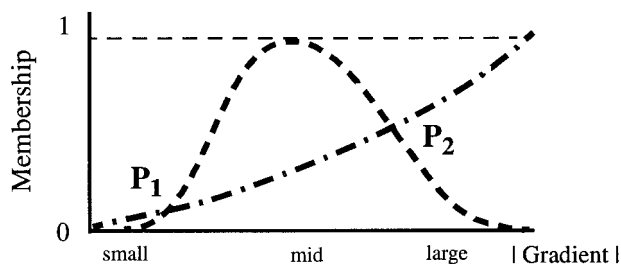


FIG. 6. Example of membership functions for the magnitude of the gradient vector of the spectral signal. The dashed curve is for the atmospheric signal identification and the dot-dash curve is for the ground clutter identification. Here, P_1 and P_2 indicate locations where the two membership functions intersect. The value of the gradient at P_2 is G_2 . As opposed to P_1 , the membership value associated with this gradient level is nontrivial.

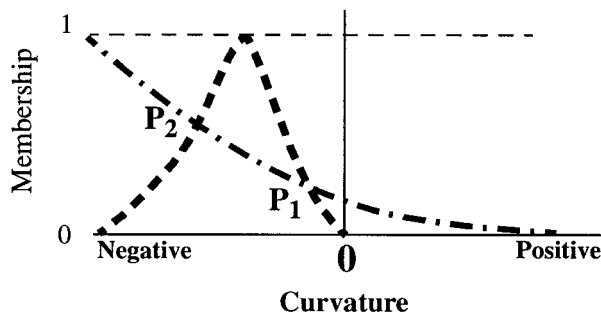


FIG. 7. Example of membership functions for the curvature of the spectral signal. The dotted curve is for the atmospheric signal identification and the dot-dash curve is for the ground clutter identification. Here, P_1 and P_2 indicate locations where the two membership functions intersect. The value of the gradient at P_2 is κ_2 . As opposed to P_1 , the membership value associated with this curvature level is nontrivial.

tions. That is, the two fields are both mapped into zero-to-one fields, which can then be arithmetically combined.

c. Membership functions

In this simplified example problem, the identification process is applied to both the atmospheric and ground clutter signals. Four membership functions are used for the ground clutter signal: gradient, curvature, distance from zero radial velocity, and symmetry around zero radial velocity. For the atmospheric signal, membership functions for the gradient and curvature are employed. In practice, there is a large number of membership functions that are employed. As described below, the membership functions for the atmospheric signal identification are quite different from those used for the ground clutter identification.

A large positive or negative slope parallel to the radial velocity axis is a strong indicator of ground clutter (or relatively stationary point targets). In contrast, moderate slopes are associated with the atmospheric signal. In the ground clutter identification process, points where the magnitude of the gradient vector is large are given large membership values. In the atmospheric feature identi-

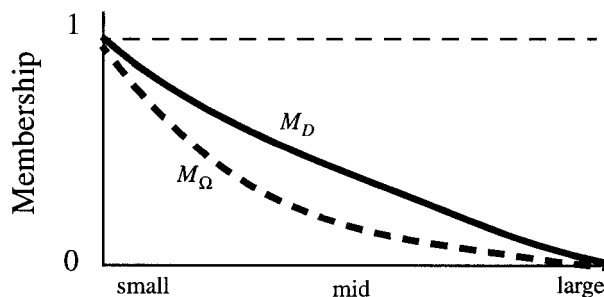


FIG. 8. Membership functions for the distance to zero radial velocity (solid line) and the symmetry around zero radial velocity (dashed line). Both fields are used for ground clutter identification.

fication process, points where the gradient is moderate are given large membership values. Slopes that are close to zero can be located near the peak of clutter or atmospheric signals or even in regions dominated by noise. (The slope calculations are from a local least squares analysis, so that noise regions can have small slopes when smoothed with the least squares fit.) These points might be thought of as “null” indicators and hence in both feature identification processes, they are given membership values close to zero. Based upon these facts, ground clutter and atmospheric membership functions for the magnitude of the gradient vector might appear as in Fig. 6.

Ground clutter (and point targets) will exhibit regions with large, negative curvature. In contrast, the atmospheric signal will be associated with moderate, negative curvatures. The membership functions corresponding to the curvature field are shown in Fig. 7.

With each of these fields (gradient and curvature), there are values where the membership functions have similar or even identical values.

There are two other membership functions that are

used for ground clutter identification in this example problem: the distance from zero radial velocity and the symmetry of the spectra around zero radial velocity. The first field, denoted D , is predicated on the assumption that ground clutter will be close to zero radial velocity. The second field, denoted Ω , is calculated using $\Omega(v) = |S(v) - S(-v)|$ and is based on the assumption that the ground clutter will be fairly symmetric around zero radial velocity. The membership functions for these two fields are illustrated in Fig. 8.

d. Combining membership values

At this stage, the individual membership values at a given point are combined into a total membership value (M_T) for that point. Typically, there will be a large number of membership fields to combine. However, in this simplified example, the total membership value for ground clutter identification, $M_T^C(x)$, at a given point, x , in the radial velocity–range coordinate space would be given by a linear combination of the four membership values:

$$M_T^C(x) = \frac{\alpha_G^C \times M_G^C(x) + \alpha_\kappa^C \times M_\kappa^C(x) + \alpha_D^C \times M_D^C(x) + \alpha_\Omega^C \times M_\Omega^C(x)}{\alpha_G^C + \alpha_\kappa^C + \alpha_D^C + \alpha_\Omega^C}, \quad (1)$$

where the subscripts G , κ , D , and Ω refer to gradient, curvature, distance, and difference, respectively. Similarly, the total membership value for the atmospheric signal identification, $M_T^A(x)$, is given by

$$M_T^A(x) = \frac{\alpha_G^A \times M_G^A(x) + \alpha_\kappa^A \times M_\kappa^A(x)}{\alpha_G^A + \alpha_\kappa^A}. \quad (2)$$

It is clear that using this synthesis methodology, it is relatively simple to add other membership fields. The ease of extensibility is one of the strengths in the current implementation. Other membership fields that might be added include different amplitude scalings of the data (e.g., linear, logarithmic, power law), spectral data from a given radial beam direction as a function of time, and spectral data from an opposing beam. The latter two fields can be implemented with a two-dimensional membership function, that is, reflecting the expected correlation (or noncorrelation) between phenomena. For example, it is reasonable to expect that ground clutter and atmospheric signals are fairly consistent from (temporal) sample to sample. Therefore, temporal consistency can be used to increase the likelihood that these two phenomena are correctly identified. The multiple-time field can also be used as a test of temporal stationarity in the wind field. In this application, a time average of atmospheric signal membership field is compared to the current field. Furthermore, the temporal analysis can

also be used to detect the onset of precipitation. The opposing-beam comparison can be implemented in an analogous fashion to test for spatial homogeneity in the wind field and to increase the confidence in the identification of atmospheric and ground clutter signals.

Figures 9 and 10 illustrate contoured total membership values for ground clutter and atmospheric signal, respectively. While the desired phenomena have been highlighted, they have not been unambiguously identified. On the right-hand side of Fig. 9, portions of the atmospheric signal that have nontrivial membership values can be seen. That is, some of the atmospheric signal looks similar to a weak ground clutter signal. If an a priori threshold was applied to this total membership field, depending on the threshold level, either portions of the ground clutter would be missed (high threshold) or portions of the atmospheric signal would be retained (low threshold). Furthermore, the atmospheric signal is typically more subtle than a strong ground clutter signal and, as can be seen from Fig. 10, it is more difficult to differentiate between the atmospheric signal and a variety of similar features. For example, a small region at the edge of the ground clutter signal between 1600 and 1700 m is indistinguishable, locally, from the atmospheric signal. To resolve this type of problem, it is required to move from local analysis to global analysis. That is, the point-by-point total membership values are collected into coherent, distinct features.

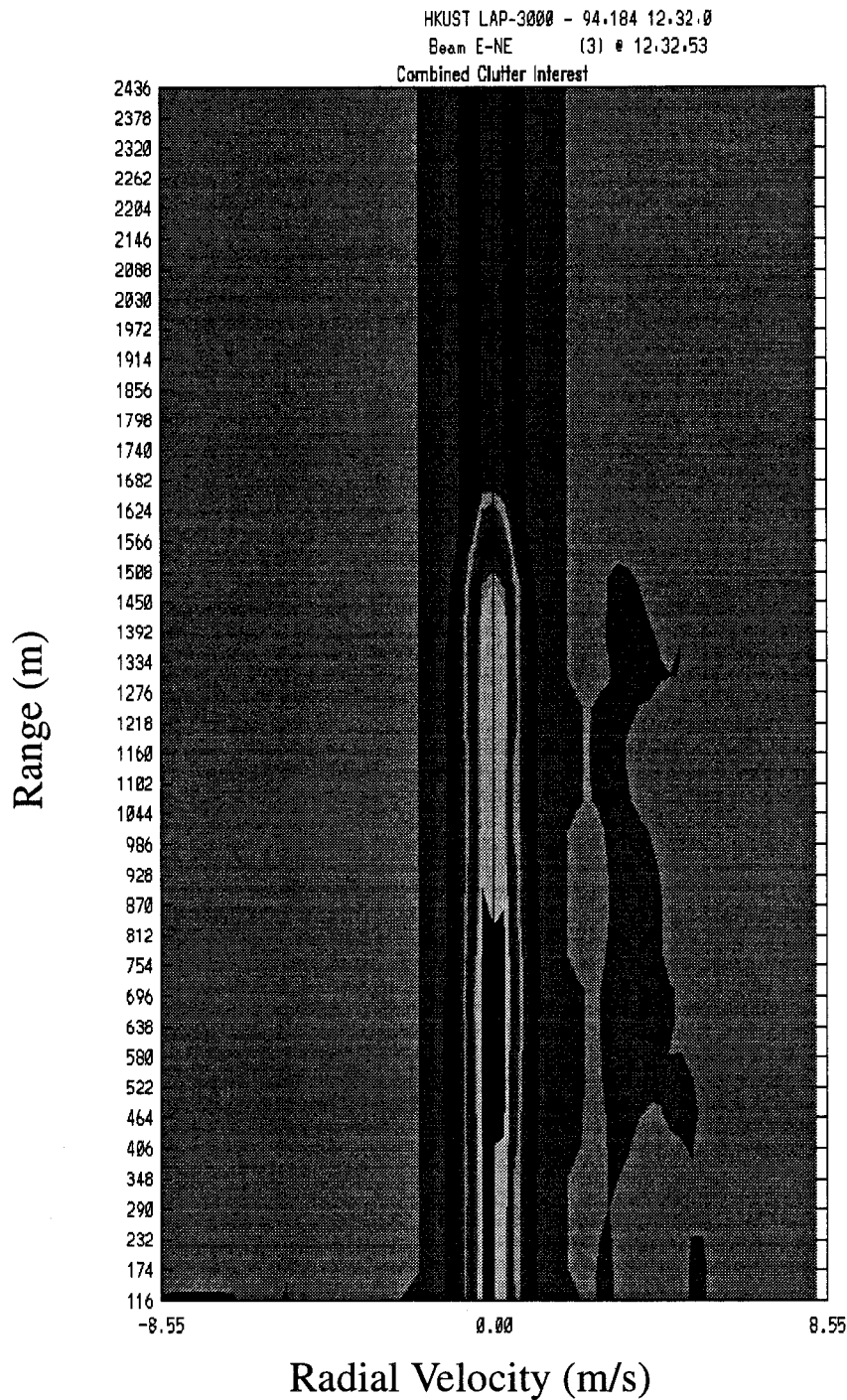


FIG. 9. Contour plot of the total membership values for ground clutter identification.

e. Feature identification

The first step in the global analysis is the application of density weighting. As discussed above, this procedure removes isolated, low-intensity (here, small total membership values) features while leaving coherent, high-intensity features intact. Figure 11 illustrates the result

after application of density weighting to the data in Fig. 10. Note that the density weighting produces the desired effect.

A comparison of Fig. 11, the density-weighted total membership field for the atmospheric signal, and the original data, Fig. 1, reveals that the ground clutter signal has been significantly attenuated. A peak-search al-

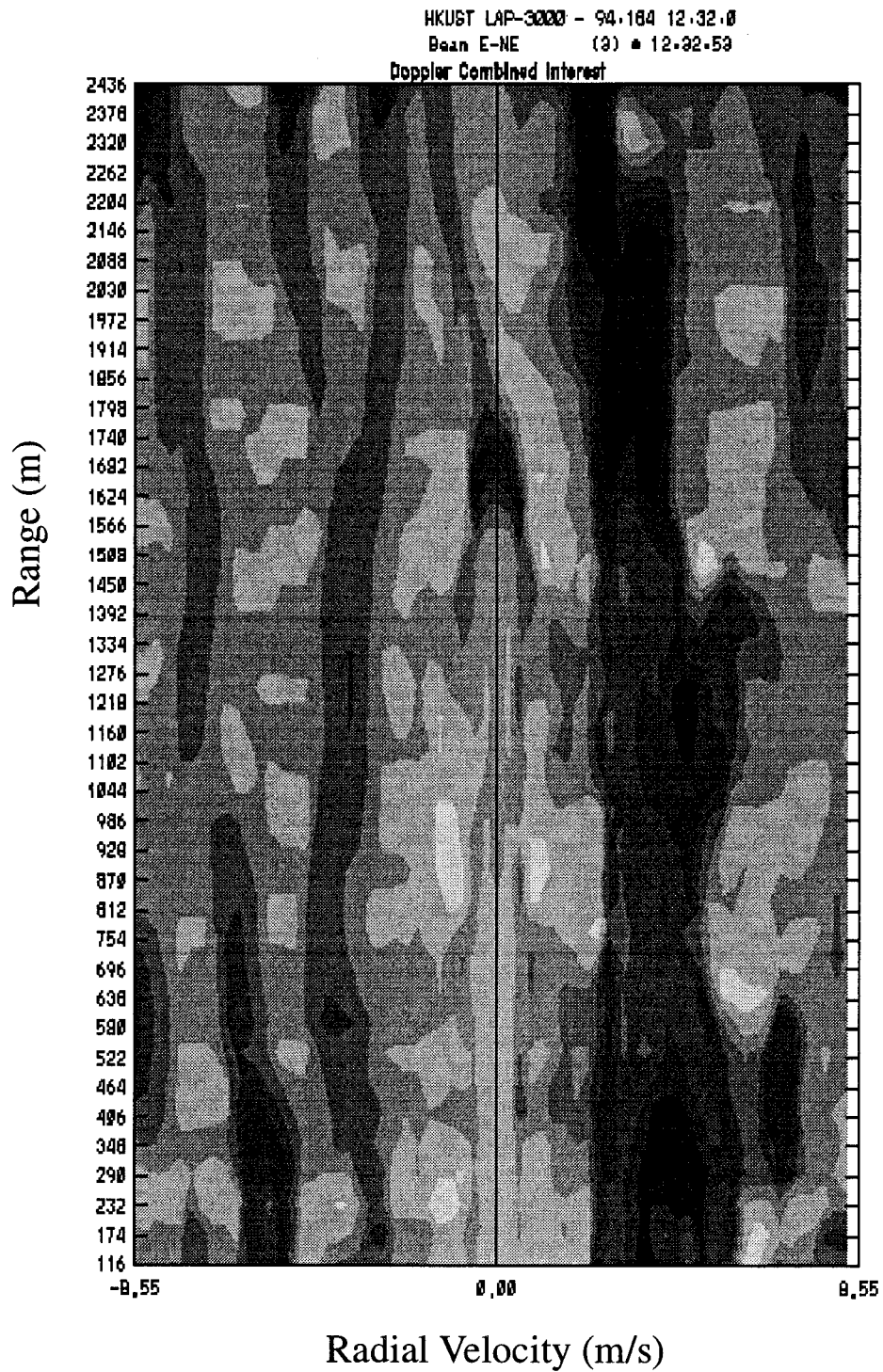


FIG. 10. Contour map of the total membership values for the atmospheric signal identification.

gorithm applied to the data in Fig. 11 would suffice in locating the correct region of the atmospheric signal. These peak locations could then be used with the original spectra in computing the moments. That is, these peaks in the total membership field could replace peaks

calculated directly from the spectra and then the moments would be calculated in the normal way (Carter et al. 1995) from the original spectra. In this example, the ground clutter and atmospheric signals are quite distinct in character and well separated in Doppler velocity

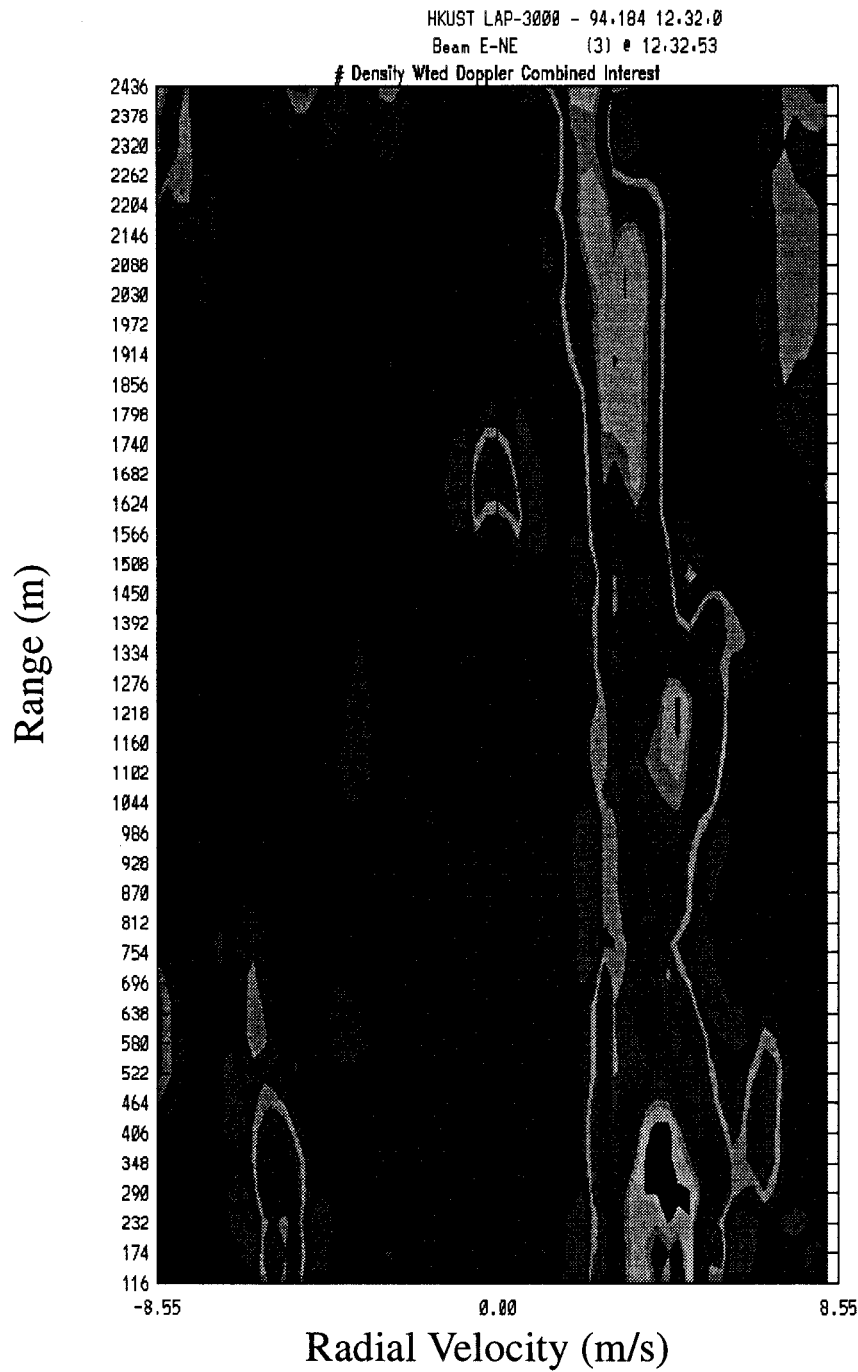


FIG. 11. Contour map of the total membership values for the atmospheric signal (cf. Fig. 10) after density weighting.

so that the separate step of identifying the ground clutter signal is not necessary. However, this is not always the case, and differentiating between the atmospheric signal and the contaminants can be important. For these more problematic cases, a peak-search algorithm might produce erroneous values and hence the necessity of the feature extraction step. Furthermore, a peak-search al-

gorithm may choose a peak that appears similar to an atmospheric signal (e.g., the feature in Fig. 11 along the zero-velocity axis at 1600–1700 m).

It is important to realize that the contours in Fig. 11 are merely graphical constructs, the information contained within the data is still at the point level, not at the feature level. The link between the point values and

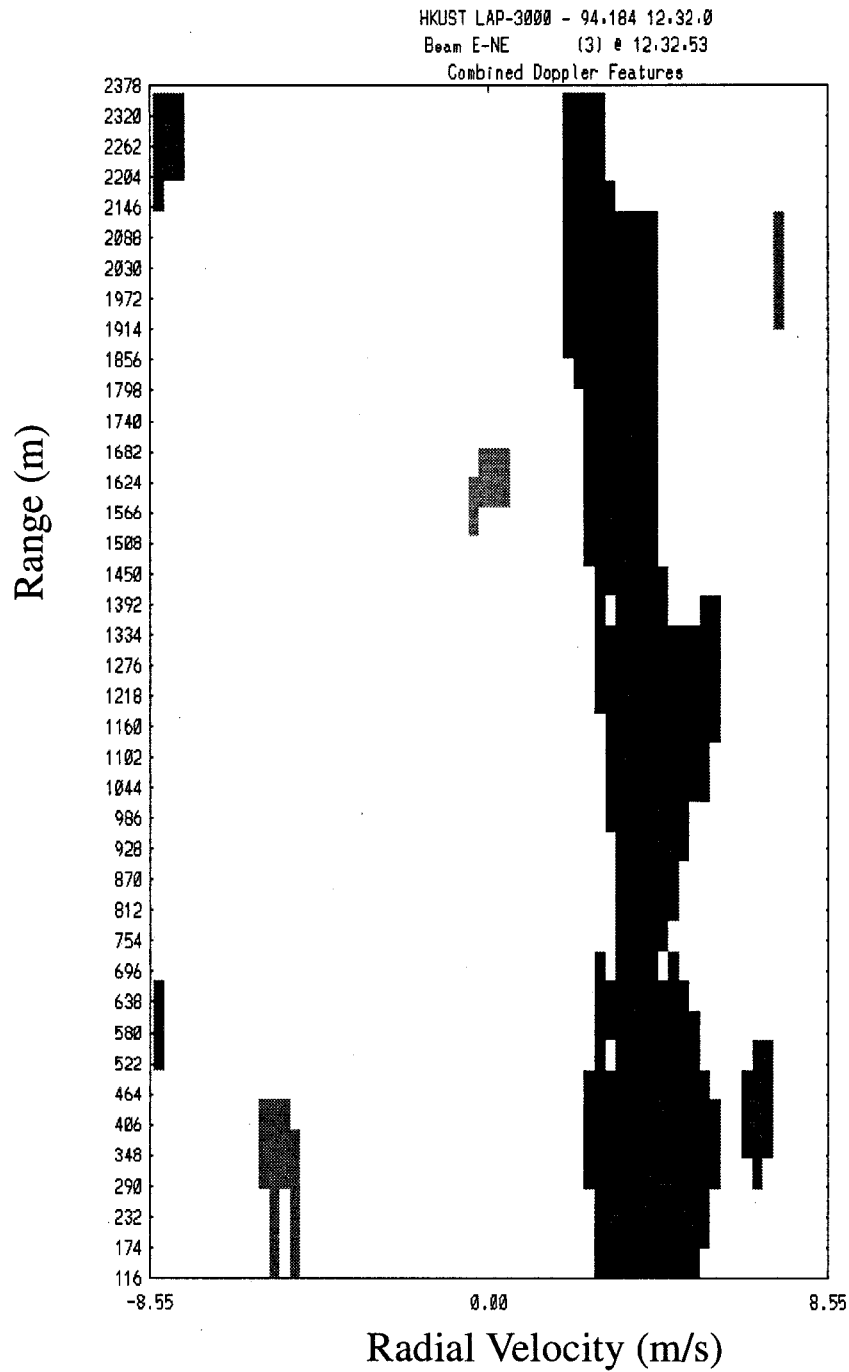


FIG. 12. Candidate features derived from the density-weighted, atmospheric signal, total membership field.

features is the linear-chain algorithm. Those points that have above-threshold total membership values are collected into distinct, coherent sets of data, so-called candidate features. It is important to note that until this point, strict thresholding of the data has been avoided, thus minimizing the loss of information. Figure 12 illustrates the results of the linear-chain algorithm applied to the data in Fig. 11.

Various rules are then applied to these candidate features to remove the “false” ones and retain the “true” one. For example, a reasonable a priori assumption might be that an atmospheric feature extends for a substantial number of ranges. A membership function that relates to the number of ranges that a particular feature occupies is then created. Assigning a “score” to each feature based on the sum of all the total membership

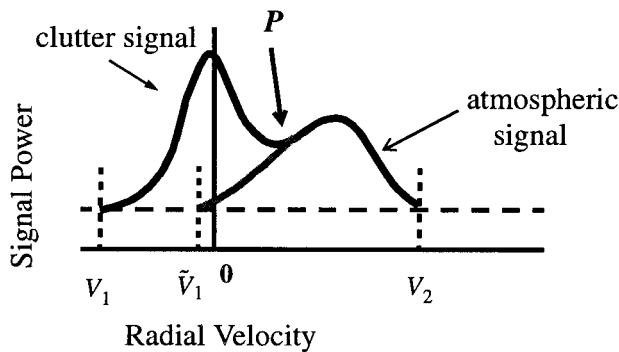


FIG. 13. Expanded region of overlapping clutter and atmospheric signal (dark curve), including extrapolation (light curve) of the atmospheric signal through the clutter. The noise level is indicated by the horizontal dashed lines.

values (from the atmospheric feature identification process) of its constituent points is another discrimination technique. These two simple rules would probably be sufficient to identify the correct feature in Fig. 11. More sophisticated rules may be required for complicated cases.

4. Moment calculations

Once the atmospheric feature has been identified, the moments are calculated in a similar fashion to existing methods. There are certain details of this procedure (related to the region over which the moments are computed) that are somewhat different than the standard method and deserve some discussion. The standard method determines the portion of the spectra over which the moments are computed by starting at the spectral peak and following the spectra down (on both sides of the peak) to the noise floor. Obviously, if the spectral peak is associated with a contamination source (e.g., ground clutter), this method can give erroneous moment values. The feature extraction techniques described above can usually distinguish between the desired atmospheric and contamination signals. This knowledge can then be applied to produce more accurate moments. It should be noted that for some profilers, a site-adaptable parameter (the range below which ground clutter is expected to be present) is used in an attempt to mitigate this problem. That is, below this a priori range value, a search for the spectral peak is made away from zero velocity. This method does often work; however, it fails if the ground clutter contamination extends above this range, if there is contamination from moving clutter sources (e.g., cars, birds, aircraft, etc.), or if the atmospheric signal is at zero velocity.

To elucidate the nonstandard techniques that are used here, consider the following example. Figure 13 illustrates what a typical spectra looks like with overlapping atmospheric and clutter signals (dark, thick lines).

The standard method would choose the peak of the clutter signal as the starting point and follow the signal

down to the noise level (horizontal dashed line). These intersection points then define the cut-off velocities (V_1 , V_2), and the moments would then be computed using the standard method (Brewster 1989). This approach will overestimate the zeroth and second moments and bias the first moment toward the clutter signal. Since the fuzzy logic feature detection method will differentiate between the clutter and atmospheric signals, this information can be exploited to avoid these errors.

The process adopted here occurs in two steps. Starting from the maximum of the atmospheric signal feature (as in the standard method), an attempt is made to follow the signal down to the noise level. As seen in Fig. 13, this would be successful on the right side of the signal, yielding the cut-off velocity V_2 . However, following the signal down on the left side, a ground clutter feature would be intersected before reaching the noise level. The feature detection method is not designed to find the precise extent of these features, hence a more accurate determination of the intersection point (P in Fig. 13) is required. This can be accomplished using similar mathematical and fuzzy logic techniques as used above, that is, the point P can be described as being along an “upside-down” ridge line (or valley) running roughly parallel to the range axis. Once this intersection point is found, an extrapolation of the atmospheric signal down to the noise level is performed (thin gray line extending from the left-hand side of the atmospheric signal). This results in the left-hand cut-off velocity \tilde{V}_1 , as seen in Fig. 13. In this manner, the errors that result from computing the moments over the merged signals is reduced considerably. This same methodology can be applied to overlapping (clear air) atmospheric and precipitation signals.

Once the first moments (radial velocities) are computed for each range, a final quality control procedure is performed: verifying the continuity of the moments as a function of range. While allowing for a natural range of variability over range, this step ensures that any outliers that passed through, or resulted from, the antecedent processing steps are removed. This continuity check utilizes a fuzzy logic approach, with membership fields generated by the chi-square value from local linear and quadratic least squares fits to the moments as a function of range. Another membership field is computed from the deviation of the given moment from a local median value. When an outlier is detected, an interpolated value is used as a replacement moment. If the outlier occurs near the edge (i.e., lowest or highest ranges) of the domain, an extrapolated value is used. A similar process is employed for the second moments.

Figure 14 illustrates the final result for this example case. It should be noted that the underlying spectral data has been processed through a two-dimensional median filter, hence the difference with the original data (cf. Fig. 1). This filtering step is performed to ensure accurate estimates from the least squares analysis. A comparison with Fig. 1 shows that the three problem areas

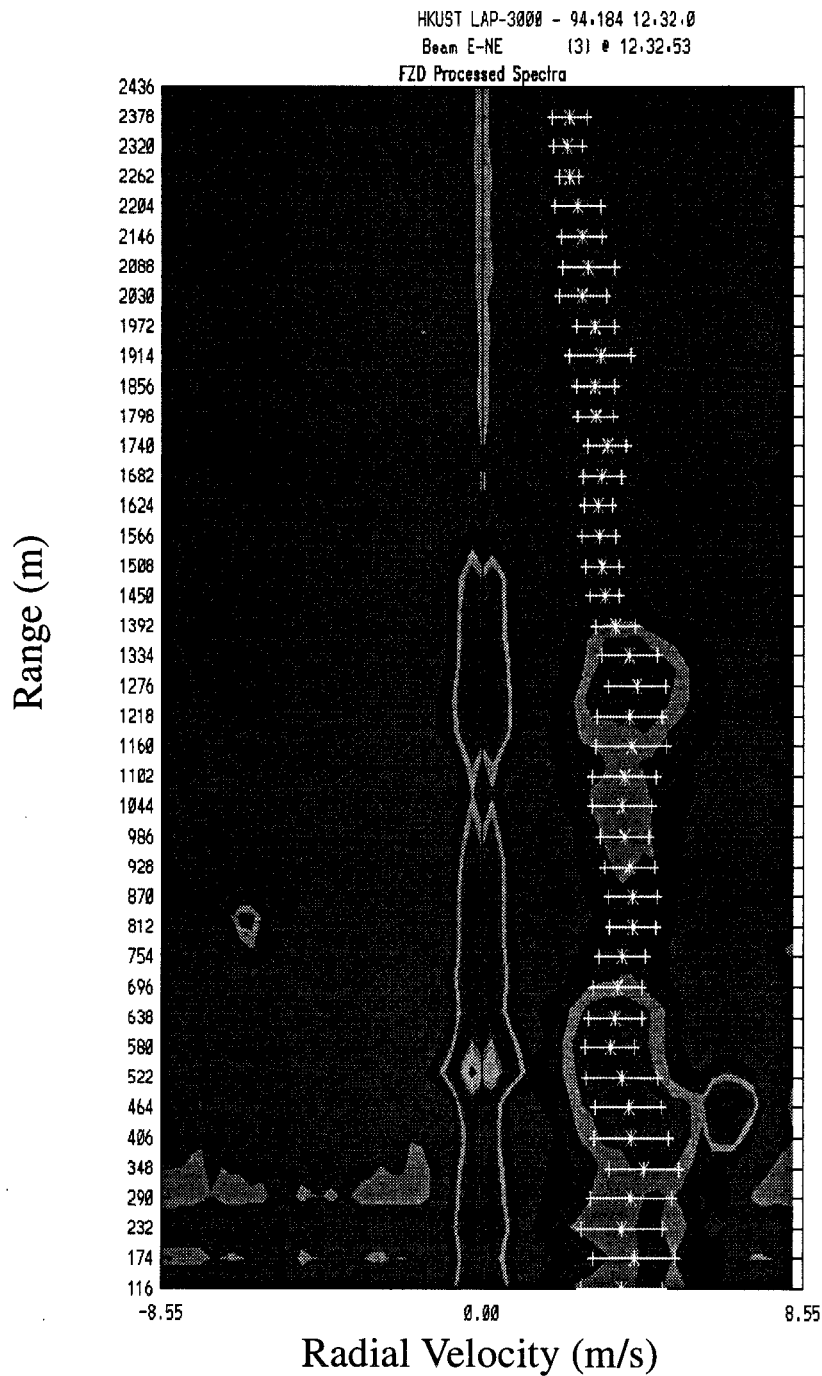


FIG. 14. Reprocessed Doppler moments using the NIMA algorithm.

in this example (ground clutter, point target at 750–850 m, and low SNR at the highest ranges) have been mitigated, and hence accurate moments can then be calculated.

5. Results

The algorithms described in this paper attempt to model the methodology of a human expert in estimating

moments from Doppler spectra. In this section, a quantitative comparison is performed between the first moments produced by human experts with those produced by NIMA. (To be precise, one of the human experts chose the portion of the spectra over which the moments were calculated, while the other expert estimated the moments themselves.) A similar comparison is made between the human expert-produced moments and those

produced by the Profiler On-line Program (POP) (Carter et al. 1995). POP is a peak-finding algorithm, which is one of the standard moment-generating algorithms used in profiler data processing. No independent source of “ground truth” is used herein, and as mentioned above, the goal of the NIMA is to approach the level of the human expert. Therefore, the moments from the human expert are considered to be truth, and a difference between the human expert moment and an algorithm moment will be called a difference or error. The data used in this analysis were specifically chosen because of the complicated spectra and high levels of contamination. In fact, there are several instances in which the Doppler spectra were so contaminated that even the human expert was uncertain of the atmospheric signal.

Two sets of data are used in the comparisons. The first dataset is from a 1299-MHz profiler located at Sha Lo Wan, China. The second is from a National Oceanic and Atmospheric Administration 449-MHz profiler located at Point Loma, California. The data from the Sha Lo Wan profiler were collected on 7–8 June and 3 July 1994. The June data were collected during the passage of Tropical Storm Russ. The chosen data represent a challenge to the algorithms because the data were contaminated by velocity folding, undetermined noise, large vertical shears, and ground clutter that often extended over a kilometer in range. From the Sha Lo Wan dataset, the human expert produced first moments for 45 beams, with 20 ranges per beam for a total of 900 points.

The data from the Point Loma profiler were collected on 27 September 1995. The human expert produced first moments for 44 beams (again 20 ranges per beam) for a total of 880 points. There was a persistent vertical shear associated with the marine boundary layer inversion that the NIMA algorithm initially identified as clutter. Once a simple parameter adjustment to handle this problem was made, the NIMA parameters were fixed and not modified during the analysis. It should also be noted that the algorithm parameters for the 449- and 1299-MHz devices were identical, that is, NIMA is insensitive to the specific type of profiler.

Figure 15 is a scatterplot generated from the combined dataset. The radial velocity estimates from the human experts are plotted relative to the horizontal axis and the POP radial velocity estimates are relative to the vertical axis. Note that there is a significant clustering along the $y = x$ line, but a large number of outliers (i.e., points far from the $y = x$ line) also exist. This implies that when the contamination level is small (i.e., the peak of the spectra is associated with the atmospheric signal), POP produced accurate moments. However, as the POP algorithm does not attempt to identify the source of the peak (i.e., contamination versus atmospheric), it will calculate erroneous moments when the peak is not associated with the atmospheric signal.

Large differences can be seen in Fig. 15 along the horizontal line $y = 0$. In these instances, POP has calculated the first moments from the spectral points near

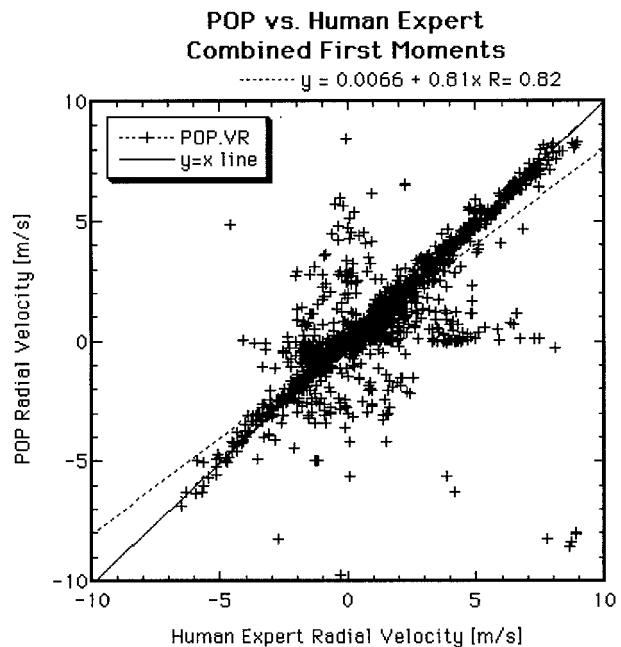


FIG. 15. Scatterplot of radial velocities (first moments) generated by the human experts (horizontal axis) and the POP algorithm (vertical axis). The solid line is the $y = x$ line, and the dotted line is the “best fit” line to the data.

zero velocity, typically associated with ground clutter, and the human expert has selected velocities away from zero. Other large differences were caused by velocity aliasing.

Figure 16 is a scatterplot of the radial velocities calculated by the human expert versus those estimated by NIMA. As with the scatterplot with the POP data, there is a significant clustering along the $y = x$ line; however, the noteworthy difference lies in the reduction of the number and severity of outliers. Table 1 summarizes a few simple statistical measures that compare the two algorithms. These statistics indicate that there is a stronger agreement between NIMA and the human expert than exists between POP and the human expert. The statistic that is most noticeable here is the large variance in the radial velocities between POP and the human expert. This enhanced variance is due to the large number of outliers generated by the POP algorithm.

In Fig. 17, vertical profiles of radial velocities produced by the human expert, POP, and NIMA are shown for the Sha Lo Wan data. The human expert did not produce a profile below 500 m in this case because of excessive ground clutter. These data are from the Sha Lo Wan profiler during the passage of Tropical Storm Russ, during which the horizontal wind speed (from aircraft and anemometer measurements) was between 25 and 35 m s^{-1} . The Nyquist velocity for the Sha Lo Wan data was 8.55 m s^{-1} . This is a relatively small Nyquist value, and hence during periods of strong wind velocity aliasing can be expected. Referring to Fig. 17,

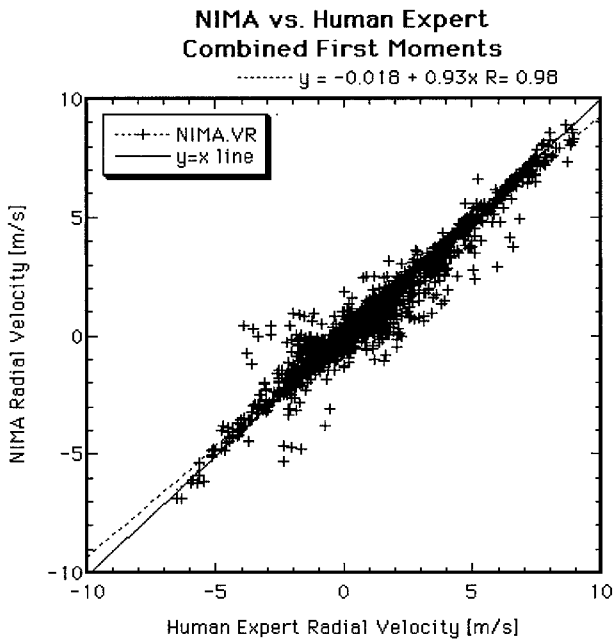


FIG. 16. Scatterplot of radial velocities (first moments) generated by the human experts (horizontal axis) and the NIMA algorithm (vertical axis). The solid line is the $y = x$ line, and the dotted line is the “best fit” line to the data.

note that the radial velocities from POP and NIMA correspond quite well with the values produced by the human expert at most heights. The few heights at which the POP algorithm is in significant disagreement with the human expert are associated with the velocity aliasing. POP does not have a dealiasing algorithm, whereas NIMA has a fuzzy-based dealiasing algorithm and thereby was able to produce accurate moments in this instance. It should be noted that out of the 1780 data points used in this analysis, only five values were contaminated by velocity aliasing. These points are visible in the lower-right-hand corner of Fig. 15.

Strong ground clutter and point targets are two other sources of contamination that result in differences between POP and the human expert. The data in Fig. 18 are a good illustration of these problems. Note that these data are the same as in the example problem presented above. POP has mistakenly selected the strong return from a point target (perhaps a bird) in calculating the radial velocities at 700–800 m. In the height range 100–500 and 1100–1500 m, POP has used the strong signal associated with ground clutter to calculate the first moments. NIMA was able to identify the ground clutter and point target as contaminants. Even though the atmospheric return was at a lower signal strength, the algorithm did estimate the proper moment. At all of the other heights, the radial velocities from the human expert, POP, and NIMA are in very good agreement.

Typically, when computing the wind vector, a large number of radial velocity samples are employed (e.g., as in the POP consensus algorithm), and it is assumed

TABLE 1. Statistical comparison of radial velocity estimation.

Statistic	POP expert	NIMA expert
Correlation coefficient	0.82	0.98
Mean difference	0.2 m s ⁻¹	0.1 m s ⁻¹
Mean absolute difference	0.6 m s ⁻¹	0.3 m s ⁻¹
Variance	2.2 m ² s ⁻²	0.3 m ² s ⁻²

that most random outliers will be thrown out and not affect the wind estimate. Under many conditions these criteria are satisfied and accurate wind estimates are generated. However, if the goal is to produce short time-averaged winds or turbulence estimates (e.g., using second moments), even a few outliers can severely contaminate the results. That is, if some of the outliers are not identified as such, or there are too many of them to compute a stable estimate, the final result will be erroneous. Unfortunately, certain sources of contamination can be quite stable as a function of time (e.g., ground clutter), and, as can be inferred from Figs. 17 and 18, these outliers can even cause a long time-average wind estimation algorithm to fail.

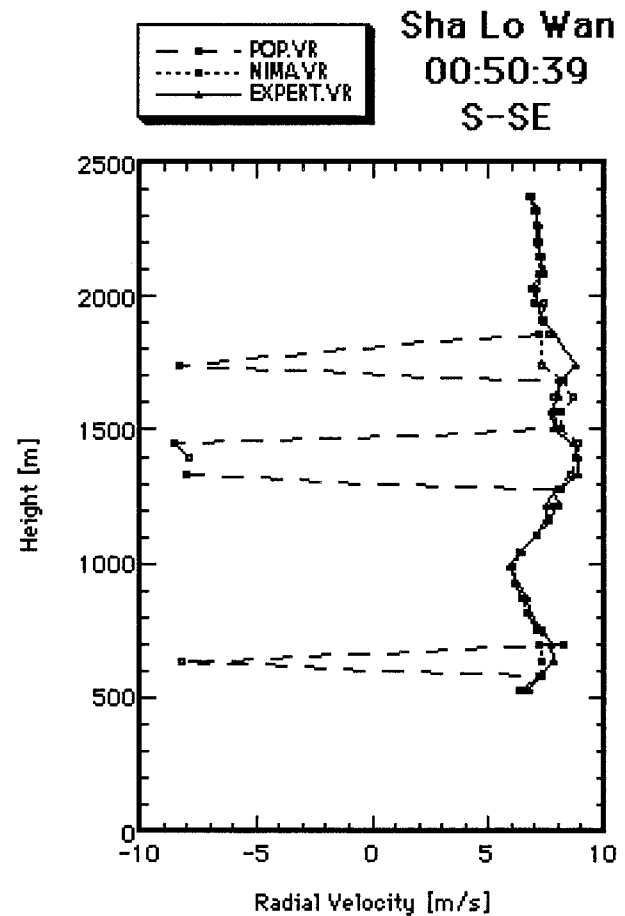


FIG. 17. Vertical profile of radial velocities from the human expert (solid line), POP (dashed line), and NIMA (dotted line). The large POP outliers are due to velocity aliasing.

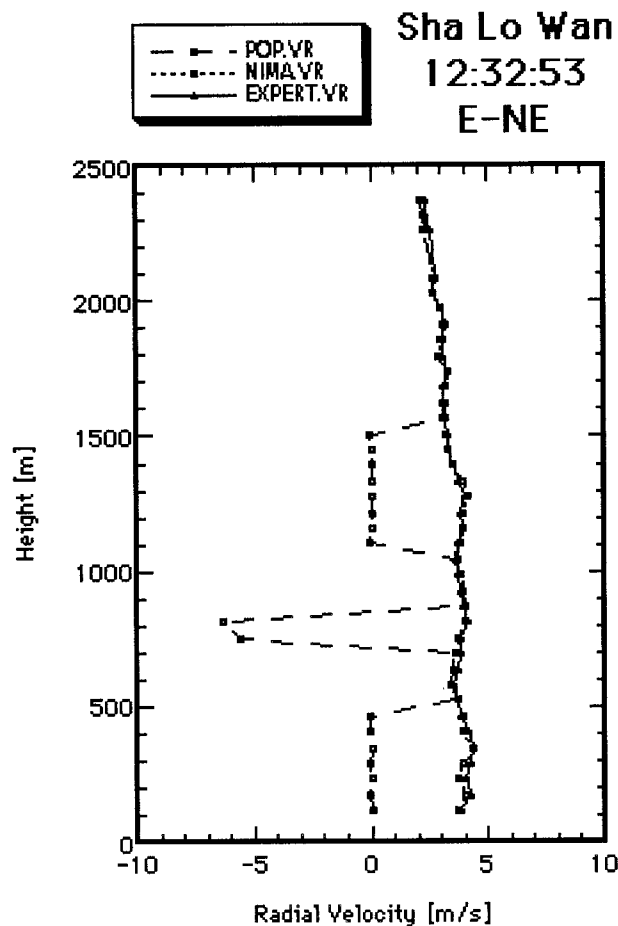


FIG. 18. Vertical profile of radial velocities from the human expert (solid line), POP (dashed line), and NIMA (dotted line). The large POP outliers are due to ground clutter and point targets at 700–800 m.

While the NIMA produces accurate moments in many cases, there are still deficiencies with the existing version. The largest differences in NIMA versus the human expert were associated with severe ground clutter at the first few range gates and significant vertical shear in the radial velocities. The clutter problem is demonstrated by the few outliers near the horizontal line $y = 0$ in Fig. 16. Nevertheless, the number and severity of these outliers is significantly less than with POP (cf. Fig. 15). The problem that NIMA had with some instances of strong vertical shear in the radial velocity is more subtle and produced the largest error in any of the data (4.2 m s^{-1}). This specific case, illustrated in Fig. 19, was from the Sha Lo Wan profiler; however, there were similar cases from the Point Loma data. Note that the large shear, as seen in this vertical beam, is centered around 500 m. POP and the human expert are in reasonable agreement, whereas NIMA apparently did not detect the shear. On the other hand, the shear at 700–900 m was correctly identified.

During an in-depth analysis of this case, it was discovered that the large shear at 500 m did not occur in

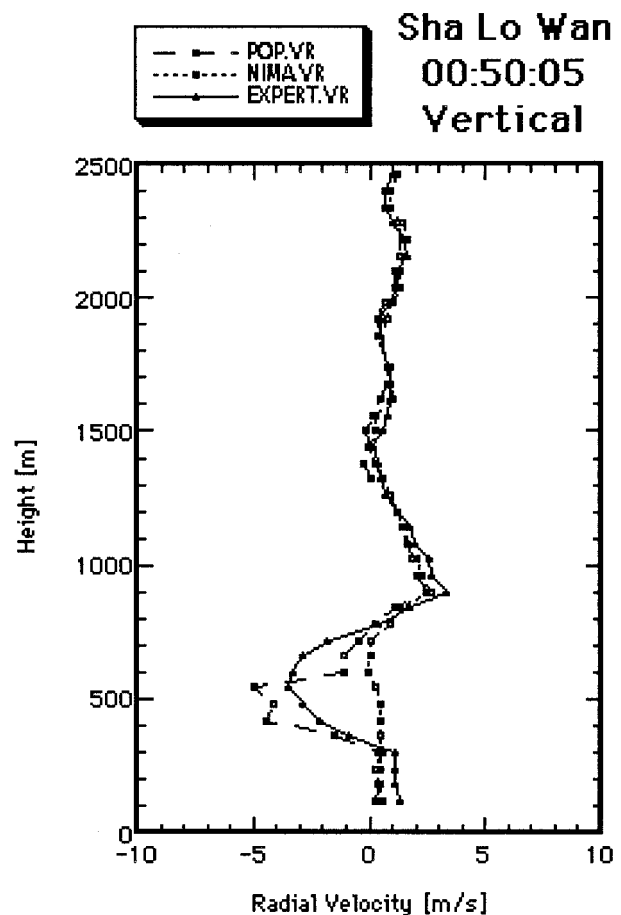


FIG. 19. Vertical profile of radial velocities from the human expert (solid line), POP (dashed line), and NIMA (dotted line). The large NIMA outliers are due to a large vertical shear in the radial velocity.

the previous vertical beam, nor in the following vertical beam. Given the strong winds flowing over the adjacent terrain, it is possible that this feature was associated with a mountain rotor. The feature detection routine produced a number of candidate features, and it was only in the final selection step that the shear feature was ruled out. That is, the algorithm did detect the shear; however, as NIMA gives weight to spatial and temporal continuity, among other qualities, in the final analysis the wrong feature was selected. Of course, the NIMA parameters associated with temporal and spatial continuity and feature selection could be modified to pick the correct feature in this case. However, this would most likely result in a large increase in false detections in other circumstances. This is a very difficult problem. If it is assumed that the atmosphere is reasonably consistent in terms of spatial and temporal continuity, then spurious contaminants are easily dismissed. However, transient phenomena, such as the vertical shear in this case, would be missed. On the other hand, if these criteria are relaxed, intermittent signals (which in some circumstances may be true atmospheric phenomena) will cause prob-

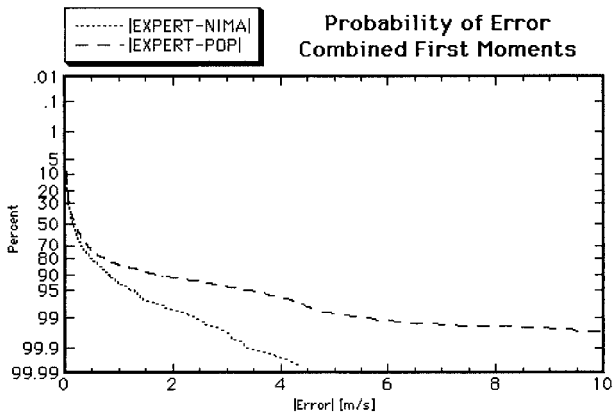


FIG. 20. Probability of absolute error greater than the specified value, calculated from the difference in first moments between the human expert and POP (dashed line) and the human expert and NIMA (dotted line).

lems. Without access to other information (e.g., from another sensor), this type of problem may be unsolvable. Nevertheless, there is typically enough information to determine if the algorithm is sure of its answer. For example, in this case the existence of poor data quality and multiple candidate features are indicators of a problematic situation. This information could then be incorporated into a quality control metric indicating the existence of questionable data.

Figure 20 is an illustrative comparison between POP and NIMA. This figure shows the probability of an absolute difference (or error) between NIMA and the human expert as well as those probabilities for the difference between POP and the human expert. These probabilities clearly distinguish the performance of the two algorithms in calculating first moments. For example, the probability that POP generates a first moment with an error greater than 2 m s^{-1} is 10%, whereas for NIMA this probability is 2%. On the other hand, the probabilities for an error greater than 4 m s^{-1} are 4% and 0.05%, respectively. This represents a factor of 2.5 decrease for POP, whereas for NIMA these numbers depict a factor of 40 decrease. Furthermore, it is clear from the figure that over the whole range of absolute errors, the rate of decrease for POP is significantly less than for NIMA. That is, the probability that a given radial velocity generated by POP is an outlier is far greater than for NIMA.

The results presented above dealt solely with the first moments generated by POP and NIMA. Second moments are calculated using a sum of squared velocities (weighted by the signal strength) between the cutoff velocities, hence errors in the cutoff velocities can result in significant errors in the estimation of second moments. An incorrect choice of cutoff velocities can occur if the wrong region of the spectra over which the second moment is calculated is selected, as in Fig. 13 where the atmospheric signal is overlapping a contamination

signal. POP selects the cutoff velocities as the points where the (assumed) atmospheric signal intersects the noise level. On the other hand, NIMA analyzes the signal to determine whether the cutoff velocities occur before the signal reaches the noise floor. This often results in more accurate cutoffs and hence more accurate second moment estimates. As second moments are useful in calculating turbulence from Doppler spectra, this step is quite important to ensure accurate turbulence estimates.

6. Conclusions

The utility of wind, wind shear, and turbulence estimates from Doppler moments is highly correlated to the quality of the moments. In this paper, a new spectral-level quality control algorithm, NIMA, has been described. This algorithm is based on a combination of mathematical analysis, fuzzy logic synthesis, and image-processing techniques. NIMA is a fully automated, real-time algorithm that attempts to mimic the feature detection processing of a human expert. One of the main features of NIMA lies in the fuzzy logic infrastructure. It is a highly parallel and adaptable algorithm; modules can be modified, added, or removed with ease.

It has been shown that even for highly contaminated wind profiler spectra, NIMA can produce first moments that compare well with the moments calculated by human experts. The main improvement between NIMA and one of the standard wind profiler moment estimation algorithms, POP, lies in the reduction of the number and severity of outliers. This is especially important for real-time applications, where short-time averages are desired. There are still shortcomings with NIMA, and a few areas targeted for improvement have been described. Clearly, additional verification is warranted, especially with independent data sources (e.g., lidar and aircraft data).

The most fundamental and important modification to NIMA will be the addition of quality control metrics or confidence indices. After the first and second moments are estimated, they are used in applications such as wind, wind shear, and turbulence estimation. Obviously, the accuracy of these derived products will be directly proportional to the accuracy of the moments. Although NIMA will reduce the number and magnitude of the moment estimation errors, there will be circumstances in which inaccuracies remain. This will be especially problematic when short time averages are desired. To mitigate the effect of questionable moment estimates on the derived products, some form of quality control is required. This is facilitated by assigning a confidence value for each moment.

When processing averaged Doppler spectra from wind profilers, there is enough time between beam acquisitions to apply a large number of quality control algorithms. With a scanning weather radar or unaveraged spectra from a profiler, fewer and more efficient

techniques may be required for near-real-time processing. However, as the speed of computation increases, such algorithms could be feasible and would certainly have a large impact on reducing the errors in moment estimation.

Acknowledgments. This research and development work is sponsored by the Royal Observatory Hong Kong and fulfills a part of Agreement ROC IV 7/92 VI. Per subsection 12.3 of the agreement, findings may only be distributed for scientific, academic, and research purposes. Any other use must be authorized by the authors and the Royal Observatory Hong Kong. This research is also sponsored by the National Science Foundation through an Interagency Agreement in response to requirements and funding by the Federal Aviation Administration's Aviation Weather Development Program. The views expressed are those of the authors and do not necessarily represent the official policy or position of the U.S. government.

The authors would like to thank Dr. Earl Gossard of NOAA/ETL for assuming the role of "human expert" for the Point Loma profiler data and Eli Karplus for assistance in preparing the figures. Finally, the authors would like to acknowledge Dr. Peter Neilley and Dr. Alexander Praskovsky of NCAR for their valuable comments in the preparation of this paper.

REFERENCES

- Albo, D., 1994: Microburst detection using fuzzy logic. National Center for Atmospheric Research, Boulder, CO, 49 pp. [Available from Research Application Program, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.]
- Brewster, K. A., 1989: Profiler training manual #2: Quality control of wind profiler data. National Weather Service, 39 pp.
- 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.
- Clothiaux, E. E., R. S. Penc, D. W. Thomson, T. P. Ackerman, and S. R. Williams, 1994: A first-guess feature-based algorithm for estimating wind speed in clear-air Doppler radar spectra. *J. Atmos. Oceanic Technol.*, **11**, 888–908.
- Delanoy, R. L., and S. W. Troxel, 1993: Machine intelligence gust front detection. *Lincoln Laboratory J.*, **6**, 187–211.
- Fujibe, F., 1989: Short-term prediction of rainfall patterns in central Honshu, Japan—Classification with the fuzzy c-means method. *J. Meteor. Soc. Japan*, **67**, 967–983.
- Jordan, J. R., R. J. Latatis, and D. A. Carter, 1997: Removing ground and intermittent clutter contamination from wind profiler signals using wavelet transforms. *J. Atmos. Oceanic Technol.*, **14**, 1280–1297.
- Klir, G. J., and T. A. Folger, 1988: *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, 355 pp.
- , and B. Yuan, Eds., 1996: *Fuzzy Sets, Fuzzy Logic and Fuzzy Systems. Selected Papers by Lofti A. Zadeh*. World Scientific Publishing Company, 826 pp.
- Kosko, B., 1992: *Neural Networks and Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, 449 pp.
- Merritt, D. A., 1995: A statistical averaging method for wind profiler Doppler spectra. *J. Atmos. Oceanic Technol.*, **12**, 985–995.
- Rogers, D., F. W. Wilson Jr., and R. K. Goodrich, 1991: A feature identification algorithm. Preprints, *25th Int. Conf. on Radar Meteorology*, Paris, France, Amer. Meteor. Soc., 115–118.
- Strauch, R. G., D. A. Merritt, K. P. Morgan, K. B. Earnshaw, and D. van de Kamp, 1984: The Colorado Wind-Profiling Network. *J. Atmos. Oceanic Technol.*, **1**, 37–49.
- Weber, B. L., D. B. Wuertz, D. C. Welsh, and R. McPeck, 1993: Quality controls for profiler measurements of winds and RASS temperatures. *J. Atmos. Oceanic Technol.*, **10**, 452–464.
- Wilfong, T. L., S. A. Smith, and R. L. Creasy, 1993: High temporal resolution velocity estimates from a wind profiler. *J. Spacecr. Rockets*, **30**, 348–354.
- Yager, R. R., S. Ovchinnikov, R. Tong, and H. Nguyen, Eds., 1987: *Fuzzy Sets and Applications. Selected Papers by L. A. Zadeh*. John Wiley, 684 pp.