Radial Velocity and Wind Measurement with NIMA–NWCA: Comparisons with Human Estimation and Aircraft Measurements

STEPHEN A. COHN

National Center for Atmospheric Research, Boulder, Colorado*

ROBERT K. GOODRICH

National Center for Atmospheric Research, and*
Department of Mathematics, University of Colorado, Boulder, Colorado

CORINNE S. MORSE, ELI KARPLUS, STEVEN W. MUELLER, LARRY B. CORNMAN, AND R. ANDREW WEEKLEY

National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 3 April 2000, in final form 22 August 2000)

ABSTRACT

The National Center for Atmospheric Research (NCAR) Improved Moments Algorithm (NIMA) calculates the first and second moments (radial velocity and spectral width) of wind-profiler Doppler spectra and provides an evaluation of confidence in these calculations. The first moments and their confidences are used by the NCAR Winds And Confidence Algorithm (NWCA), to estimate the horizontal wind. NIMA–NWCA has been used for several years in a real-time application for three wind profilers in Juneau, Alaska. This paper presents results of an effort to evaluate the first moments produced by NIMA and horizontal winds produced by NIMA–NWCA through comparison with estimates from ''human experts'' and also presents a comparison of NIMA–NWCA winds with in situ aircraft measurements. NIMA uses fuzzy logic to separate the atmospheric component of Doppler spectra from ground clutter and other sources of interference. The fuzzy logic rules are based on similar features humans consider when identifying atmospheric and contamination signals in Doppler spectra. Furthermore, NIMA attempts to mimic the human experts' assignment of confidence to the moments. A Human Moment Analysis (HMA) tool was developed to assist the human experts in quantifying moments. This tool is described and a methodology of tuning NIMA rules based on human truth specification is presented. NIMA performed well on a dataset specifically chosen to be difficult. The average absolute error between the HMA estimate and NIMA-derived radial wind estimate was slightly more than 0.3 m s^{-1} when data with low NIMA confidence were excluded, which is comparable to the Doppler spectrum resolution. The correlation between winds derived from NIMA–NWCA and from HMA first-moment estimates exceeded 0.96 when the data with low NWCA confidence were excluded. The correlation coefficient between NIMA winds and in situ measurements by aircraft was 0.93 when aircraft winds that were believed to be accurate were used.

1. Introduction

UHF boundary layer wind profilers were first designed for automated monitoring of winds in the Tropics, with a time resolution of 30–60 min (Ecklund et al. 1990; Carter et al. 1995). However, with their proliferation, these profilers are frequently being used in much more demanding applications. The National Center for Atmospheric Research (NCAR) Improved Moments Algorithm (NIMA) was developed to calculate the first and second moments (radial velocity and spectral width) of Doppler spectra from boundary layer wind profilers and, in conjunction with the NCAR Winds and Confidence Algorithm (NWCA), ultimately to provide accurate wind and turbulence measurements along with an associated confidence in these measurements. NIMA is described by Cornman et al. (1998), with details and recent changes described by Morse et al. (2001, manuscript submitted to *J. Atmos. Oceanic Technol.*, hereinafter MGC), and NWCA is described in Goodrich et al. (2001, manuscript submitted to *J. Atmos. Oceanic Technol.*, hereinafter GMCC). NIMA uses fuzzy logic

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Stephen A. Cohn, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: cohn@ucar.edu

and global image processing to separate atmospheric signals (clear air and precipitation) from noise, stationary and moving clutter targets, and radio frequency interference (RFI). Zimmerman (1996) and Yager et al. (1987) include an overview and history of the development of fuzzy logic. The development of NIMA was motivated by the requirements for accurate, high-time-resolution wind and turbulence measurements. Specifically, it was designed for operational use at airports in close proximity to complex terrain. Most existing profiler moment estimation algorithms could not meet these stringent requirements. NWCA was developed in conjunction with NIMA to take advantage of the moment confidence estimates in calculating horizontal winds from the moments and to provide a confidence estimate for the resulting winds.

The purpose of this work is to present the results of quantitative evaluation of the NIMA radial velocity (first moment of the Doppler spectrum) and horizontal wind estimates given by NWCA. This evaluation is made by comparing the human experts' estimates of the Doppler spectra first moments and winds to the NIMA–NWCA-derived values. Comparison is also made with a limited number of in situ winds measured with an aircraft.

NIMA also assigns a ''confidence'' or quality control value to each radial velocity (first moment), and NWCA assigns a confidence to each horizontal wind value. The confidence estimation algorithm also utilizes fuzzy logic. First moment confidences are based on characteristics of the chosen atmospheric signal and the presence and proximity of interfering signals. A first-moment estimate will have high confidence if it is far from RFI and ground clutter, if there is local continuity in time and space of the first-moment estimates, and if there is a high signal-to-noise ratio (SNR). High confidence also requires that the spectral noise exhibits appropriate statistics, that the estimate is not based on data interpolated or extrapolated from surrounding data, and that the several methods used to compute the first moment agree well (MGC). Wind confidences are based on testing the assumptions made in the wind model (GMCC). The confidence is not directly equivalent to an estimate of the standard deviation or standard error of the measurement, but it does serve a similar purpose, allowing for the evaluation of whether the measurement is appropriate for a specific application. For example, in an aviation warning system, wind shear calculated from high-confidence winds may be used to issue a wind shear alert, but an alert based on low-confidence winds would be suppressed. A brief description of the NIMA and NWCA algorithms is given in section 2. Section 3 describes the dataset used for the NIMA evaluation, and the results of this evaluation are presented in section 4.

2. NIMA–NWCA summary

NIMA uses fuzzy logic and global image processing to identify atmospheric and contamination signals in Doppler spectra. The latest implementation of NIMA, which has been used with three wind profilers [part of the Juneau, Alaska, field program from which our evaluation data set was drawn], is described in detail in MGC. Briefly, it identifies the atmospheric feature within a set of range-gated Doppler spectra and uses this feature as a basis for calculating spectral moments for each range. The algorithm uses clues similar to those used by humans to distinguish features within an image.

The current implementation of NIMA includes recognition of the clear-air or precipitation components of Doppler spectra, stationary or slowly moving ground clutter, and RFI. It uses clues that include time and range continuity and the amplitude, the shape, and so on, of spectral features. For example, RFI is likely to be present within a narrow Doppler velocity interval at all ranges; the ground clutter signal is likely to be centered near zero Doppler velocity, have a large magnitude, and have a very sharp curvature and gradient in the velocity direction and sometimes also in the range direction. A fuzzy logic algorithm assigns a weight to each characteristic, for example, curvature or gradient, and combines all the appropriate information to choose the most likely atmospheric feature (or clutter or RFI feature). The moments of the signal feature are computed, and moments from successive beams are combined by NWCA to measure the horizontal wind. To do this, the wind is assumed to be linear and stationary during the measurement period, which can be as short as 2 min. These assumptions are tested in NWCA and are used as part of the wind confidence estimation.

Similar fuzzy logic rules are used to assign a confidence to these moments. Moment confidence is based on a combination of characteristics of the feature and also on how close in Doppler velocity the feature is to clutter or RFI features. The horizontal wind confidence is based on the confidence of the moments and also on time continuity of the moments in a beam direction and the symmetry of moments from geometrically opposite beam directions. Details of the NWCA wind algorithm are described in GMCC.

NIMA is unique as an algorithm to process wind profiler Doppler spectra. The spectral processing component of the Profiler Online Program (POP; Carter et al. 1995) does not utilize time or range continuity and relies on the premise that the atmospheric signal is stronger than any interfering signal. May and Strauch (1989) discuss two methods to estimate moments from Doppler spectra. However, they are concerned with numerical methods to find the moment of a distribution and do not consider interfering signals. Weber et al. (1993) introduced an algorithm that utilizes continuity but apply it to derived moments or winds rather than spectra. Clothiaux et al. (1994) utilized range continuity and a neural network classifier in a feature-based approach to discriminate against ground clutter. None of this earlier work applies global image processing.

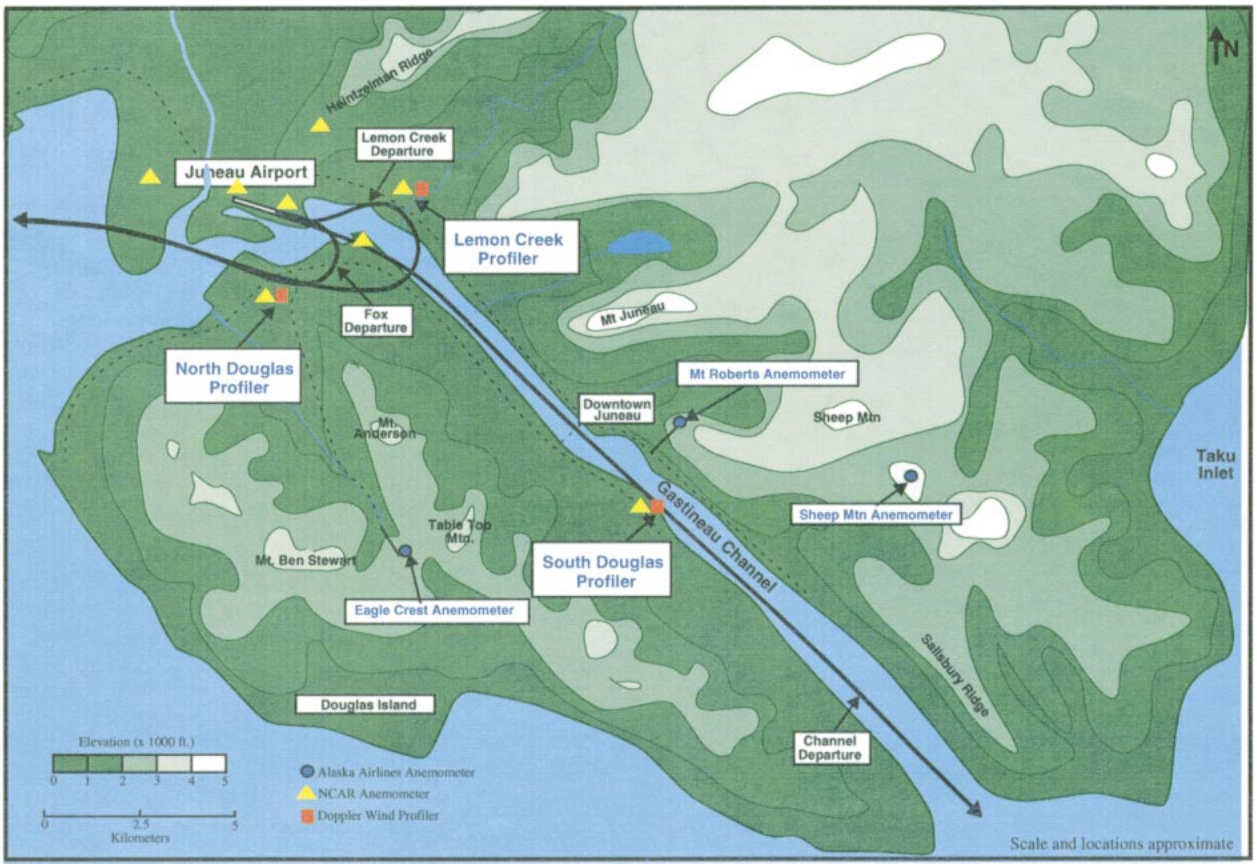


FIG. 1. Topographic map of the Juneau, AK, area. Wind profilers are located at Lemon Creek, South Douglas Island, and North Douglas Island. The three aircraft departure patterns are known as the Fox, Lemon Creek, and Channel departures.

3. Data

a. The Juneau field experiment

During the past three years, NIMA–NWCA has been used to process Doppler spectra and to estimate winds from three 915-MHz boundary layer wind profilers in Juneau in real time. These wind profilers are components of a turbulence and wind shear warning system being developed for the Juneau airport. Juneau is located in coastal southeast Alaska amid rugged peaks that rise about 1000 m from the Pacific Ocean. The locations of the three wind profilers, at North Douglas Island, Lemon Creek, and South Douglas Island, are shown in Fig. 1, along with an indication of the surrounding terrain. Other sensors in the vicinity of the airport include an array of anemometers, both near sea level and atop the nearby peaks. Typical aircraft departure tracks (known as Fox, Lemon Creek, and Channel departures) are also shown. Peaks approximately 1000 m high can be seen on either side of the Gastineau Channel (Fig. 2). In this photograph, oriented toward northwest, downtown Juneau is located just ahead of the nose boom of the research aircraft and the Juneau airport is just west of the far end of the channel. The proximity of this terrain to the airport and the turbulence and shear associated with flow over this terrain can create hazards during approaches and departures.

During February–April of 1998, an intensive data collection campaign occurred, which included a limited number of aircraft flights to measure wind and turbulence. Because these in situ verification data are available, much of the detailed analysis of NIMA–NWCA output is for this period. The wind profilers are operated with a repeating four-beam-direction sequence (30 s dwell time per beam), approximately north, west, south, and east. Doppler spectra are generated with approximately 0.3 m s^{-1} spectral resolution and are averaged using the Intermittent Clutter Removal Algorithm (ICRA; Merritt 1995). No vertically pointing beam is used, although vertical air motion is derived by combining data from opposing beam directions. This sequence was chosen because a rapidly updating series of horizontal wind is desired for the warning system. These profilers are operated with 60-m range resolution from a lowest gate of about 120 m and a maximum range of approximately 2.5 km.

b. Human-expert Doppler spectral moments

To test the performance of the NIMA first-moment algorithm, a large set of data was analyzed by humans

FIG. 2. Photograph oriented NW up the Gastineau Channel. Downtown Juneau is just past the nose boom of this research aircraft. The Juneau airport is west of the far end of the channel, and the peaks surrounding the channel are approximately 1 km high.

thoroughly familiar with wind profiler spectra. A ''Human Moment Analysis'' (HMA) tool was developed to assist in this process. The HMA tool displays Doppler power spectra (dB) and allows the human expert to specify lower and upper velocity bounds within which a clean atmospheric signal is present. Input is through a mouse and cross hair. Moments are then produced automatically by HMA after the selected interval is fitted with a least squares quadratic polynomial to the dB field. The first and second moments are estimated from the parameters of the quadratic fit. The fitting process assumes that the linear spectra will be approximately Gaussian in shape so that the log data will be quadratic. The human expert may also view a full height profile of spectra as well as a sequence of beam directions to take into account range and time continuity. The expert tries to place the velocity bounds so that ground clutter and other contamination are excluded from the fit. If the human expert is not satisfied with this calculated first moment, the expert can directly place a first moment anywhere in the spectrum.

An alternate analysis process was considered in which the human would specify the exact lower and upper bounds of the signal and the data between these bounds would be used to calculate moments. This process may be more accurate with clean spectra, but the fitting process has a significant advantage when the spectrum is partially overlapping clutter. In this situation, the fitting process essentially assumes the region of the spectrum obscured by clutter is symmetric with the observable region of the spectrum.

A typical screen of the HMA tool is shown in Fig. 3. As spectra on the right are analyzed, the first moment and spectral width are shown on the contour plot (on the left) using horizontal bars. The first moment is indicated near the center of the horizontal bar for a given range. The ends of each horizontal bar are placed one sigma unit on either side of the first moment. Here, sigma is the square root of the second central moment.

In addition to using the HMA tool to calculate spectral moments, the human expert provides a confidence index between 0 and 1 for the moments. This is a subjective estimate but is based on specific factors. The first-moment confidence is high for a clean spectrum, but it is lowered by factors such as the presence of multiple peaks between the cutoffs, nearby clutter, low SNR, or a poor quadratic fit.

The HMA algorithm of fitting a quadratic to the power spectrum is reasonable for clear air spectra, which are expected to be approximately Gaussian because of the spectrum of turbulent velocities and the wind profiler antenna pattern, but it is far from optimal for the case of rain spectra. In the case of rain or snow, the vertical motion of the hydrometeors must be accounted for to obtain reliable winds. The shape of rain power spectra is generally not Gaussian, but is determined by the raindrop size distribution that determines the drop fall speed relative to the air. Often these spectra will be skewed toward lower velocities because of a larger number of small raindrops. This possibility did not appear to change significantly the HMA first-moment estimate from what the human expert would have chosen if not

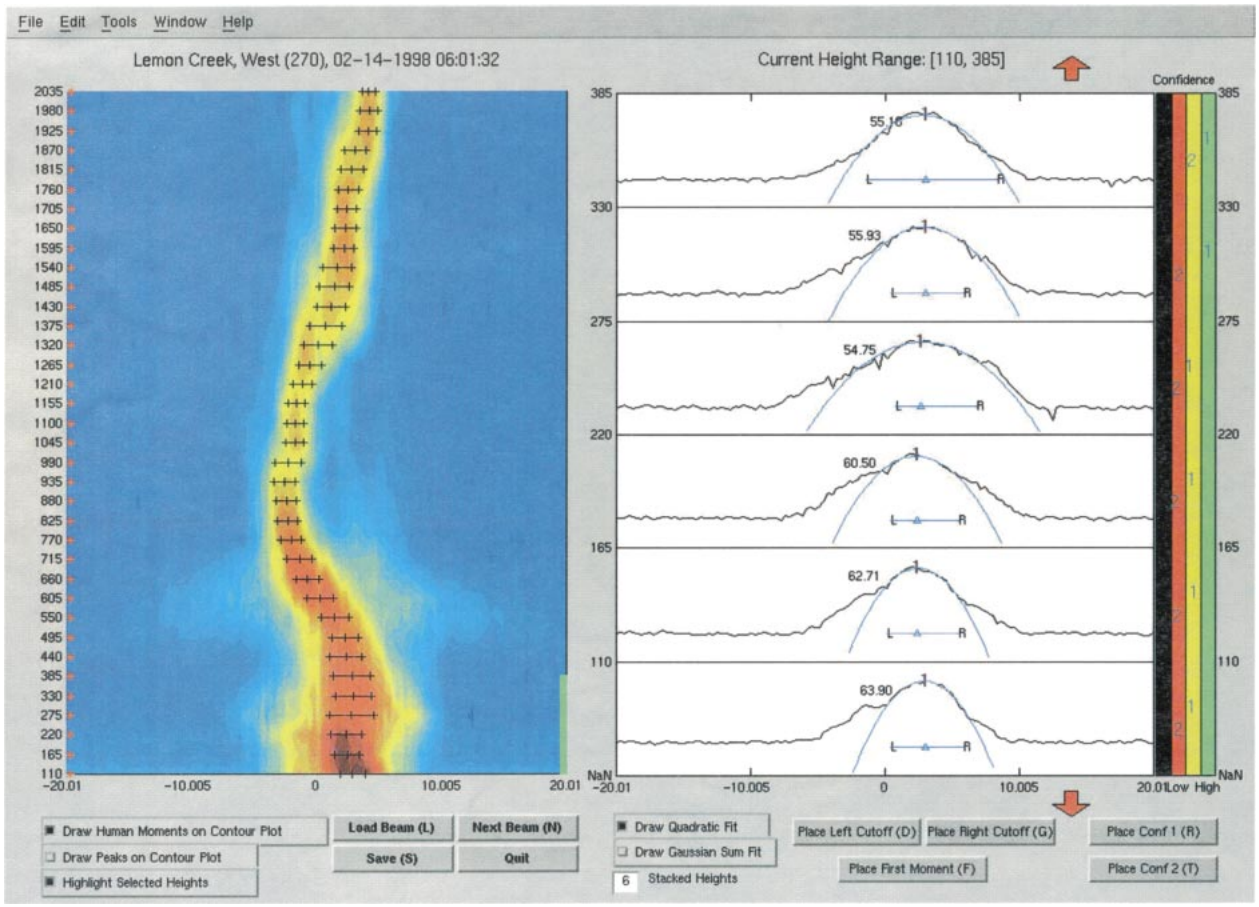


FIG. 3. A sample screen from the Human Moments Analysis tool. Human truth specifiers set a left and right signal cutoff for each spectrum in the right window, and HMA computes moments assuming a Gaussian-shaped signal. Moments are displayed for the entire profile on the left. Humans also provide first and second moment confidences using the scale at far right.

restricted by the fitting process. However, the fitting process almost always results in an underestimate of the second moment in rain. Because the assumption that the rain spectrum is Gaussian is a poor one, the human expert would typically lower the first- and second-moment confidence in rain.

Over 172 profiles (of 36 ranges) of spectra were selected for human truth specification from the Juneau dataset. These profiles were chosen to test NIMA under a variety of conditions, with the greatest number selected from times when the research aircraft was flying near one of the profilers. This choice allows an additional, independent verification of performance. The aircraft itself did not introduce much clutter in the Doppler spectra because ICRA (Merritt 1995) was used in forming averaged spectra. Profiles were also chosen at times when NIMA produced winds with large changes over short times. These large temporal discontinuities were generally found to occur when the atmosphere was dry and the SNR was low. They were chosen to test the NIMA first-moment confidence index and the NWCA wind confidence index in these conditions (i.e., to verify that NIMA–NWCA correctly labeled these data as having poor quality). Several cases of RFI were also selected to test the NIMA RFI algorithm. RFI is common in the Juneau data, because two of the profilers are in direct line of sight with each other. Cases of very strong winds [the Taku winds local to Juneau, e.g., Dierking (1998)] and of heavy rain were also included. Several cases were intentionally selected for which the NIMA winds were known to differ from other sensors (primarily anemometers on nearby mountain peaks). The verification dataset is not intended to be representative of typical conditions in Juneau, but rather consists mostly of difficult cases and cases for which aircraft data were available. Also, during the period when the aircraft was in Juneau, the wind profiler on the south side of Douglas Island was not producing reliable winds because of faulty switches that caused the beam pointing to be wrong. These winds were included as part of our study to test the NWCA wind confidence algorithm. There are a total of about 6300 spectra in the verification set. Profiles were analyzed with HMA in sets of four to give all beam directions (tilted to the north, east, south, and west), allowing for the comparison of an

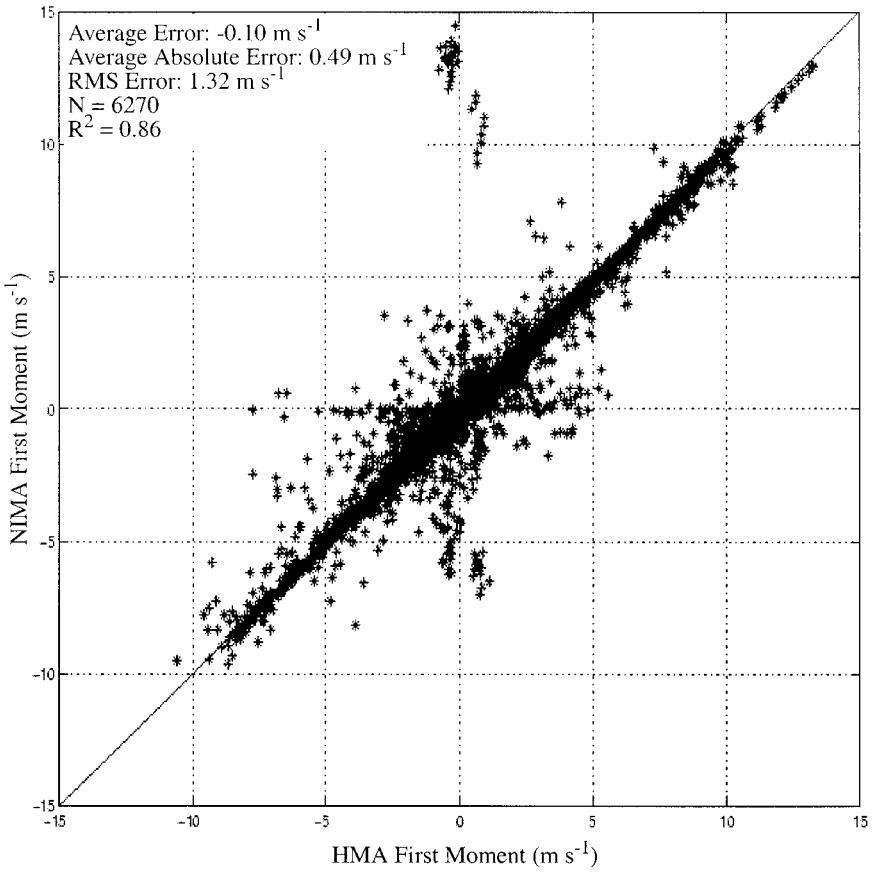


FIG. 4. Scatterplot comparing NIMA first moments with HMA first moments. The average absolute error is 0.49 m s^{-1}, and the rms error is 1.32 m s^{-1}. About 86% of the data are explained by a regression line with a slope of 0.95 (we would expect a slope of 1). The data used in these comparisons contain more cases of RFI and low SNR than is typical for Juneau.

HMA-generated wind with the NIMA–NWCA-produced winds.

c. Aircraft wind measurements

During the 1998 field project, a Citation II research aircraft operated by the University of North Dakota collected in situ wind and turbulence measurements in the vicinity of the Juneau airport. Research flights were typically on days with moderate or high winds and the potential for strong turbulence. The flight tracks were often close to the wind profiler sites, and nearly 32 h of wind measurements, with 1-s time resolution, were collected by the aircraft. The horizontal winds measured by this aircraft when it was near a wind profiler are a further check of NIMA–NWCA output winds. However, the aircraft wind cannot be considered to be precise ''truth'' for the profiler measurements because of known problems with aircraft measurements during turns, climbs, and descents, and the small aircraft sampling volume relative to the wind profilers.

4. NIMA Performance

a. First-moment comparison: NIMA and HMA

To evaluate the performance of the NIMA first-moment algorithm, the NIMA-produced first moments were compared with those produced by a human expert (using HMA). These comparisons are statistical. Scatterplots were generated, and average absolute errors (the average of the absolute value of the difference between the HMA and NIMA moments) and variances were computed.

A scatterplot comparing NIMA first-moment estimates on the vertical axis and HMA first moments on the horizontal axis is presented in Fig. 4. Notice the large cluster of points along the diagonal line, indicating that the HMA and NIMA first moments are generally in agreement. The clusters of points far from this diagonal are outlier cases that usually occur in situations of low SNR. There are also infrequent cases in which NIMA chooses a part of the Doppler spectra dominated by ground clutter or a point target, and occasionally where RFI was selected. Most of the outlier points in this figure occur when either the HMA first moment or

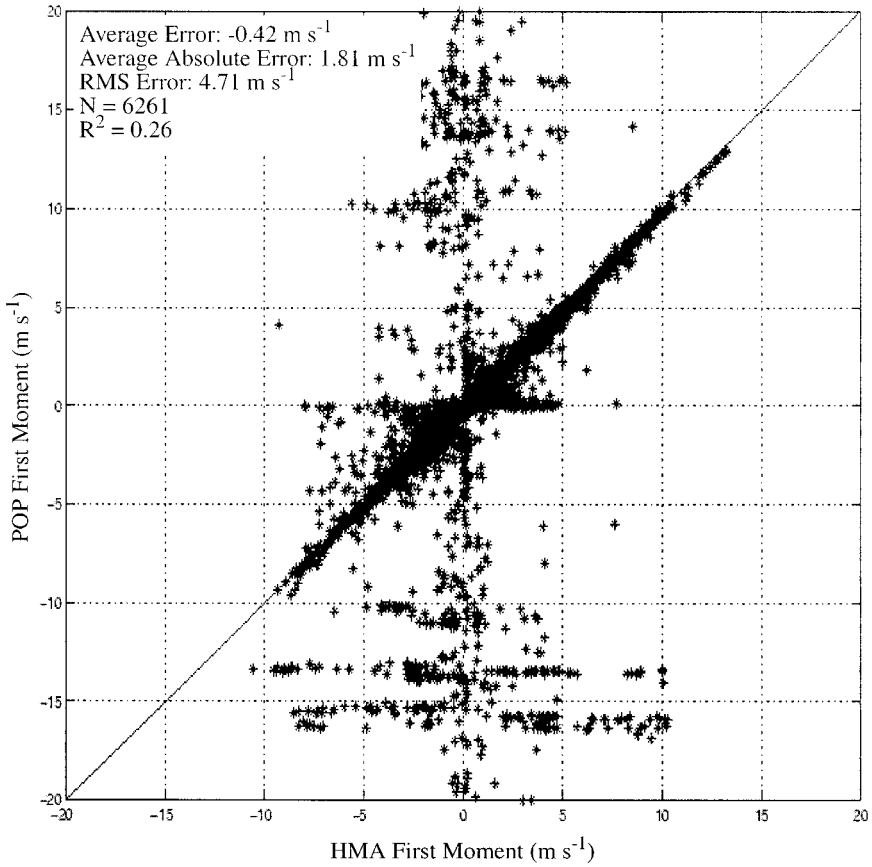


FIG. 5. This scatterplot compares POP first-moment estimates to HMA first moments. The large cluster of points along the diagonal line shows that there are many points at which the HMA and POP first moments are close to being equal. Points far from this line, at which POP and HMA first moments are different, usually occur in situations of low SNR or where POP chooses part of the Doppler spectra dominated by ground clutter, a point target, or RFI signals.

the NIMA first moment is close to zero. This result is likely to occur in low SNR when the human is at a loss to find the signal region (a common default is for the human to place the first moment near zero and give a low confidence to the result) or when NIMA chooses ground clutter that is close to zero and the human successfully avoids the clutter.

Note that Fig. 4 uses all the points in the HMA dataset, which includes 6270 human truth-specified first moments. The average absolute error in radial velocity of all these points is 0.49 m s^{-1}, and the root-mean-square error is 1.32 m s^{-1}. The amount of data explained by the regression line is 86% (a perfect result would be a value of 100%). The regression line has a slope of 0.95, close to the optimal value of 1.0.

As Fig. 4 also illustrates, the NIMA algorithm occasionally produces large first-moment outliers. However, as discussed below, these outliers are often recognized by NIMA as problematic and are given a low confidence. For comparison, the first moments computed by POP for the same dataset are shown in Fig. 5. The POP algorithm is a peak-finding algorithm that is used in most wind profilers for real-time measurements. There are many more cases in which the human first moment was placed near zero but POP chose values far from zero. Often POP selects RFI, ground clutter, or point targets. Note that wind-profiler processing typically includes a 30-min consensus of moments, and many of these outliers would be removed through the consensus process or could be removed using a minimum SNR threshold. However, in difficult conditions there can be sufficient outliers that a consensus is not reached and some acceptable data could be rejected. Under difficult conditions, NIMA moments still result in a temporally consistent set of winds in most cases.

b. First-moment confidence performance

The same data of Fig. 4 are shown color-coded by NIMA first-moment confidence in Fig. 6. Confidences of 1.0 represent near certainty, and those close to zero represent little certainty in the first moment. This color-coded figure shows that many of the large outliers have low NIMA first-moment confidence [for example, the

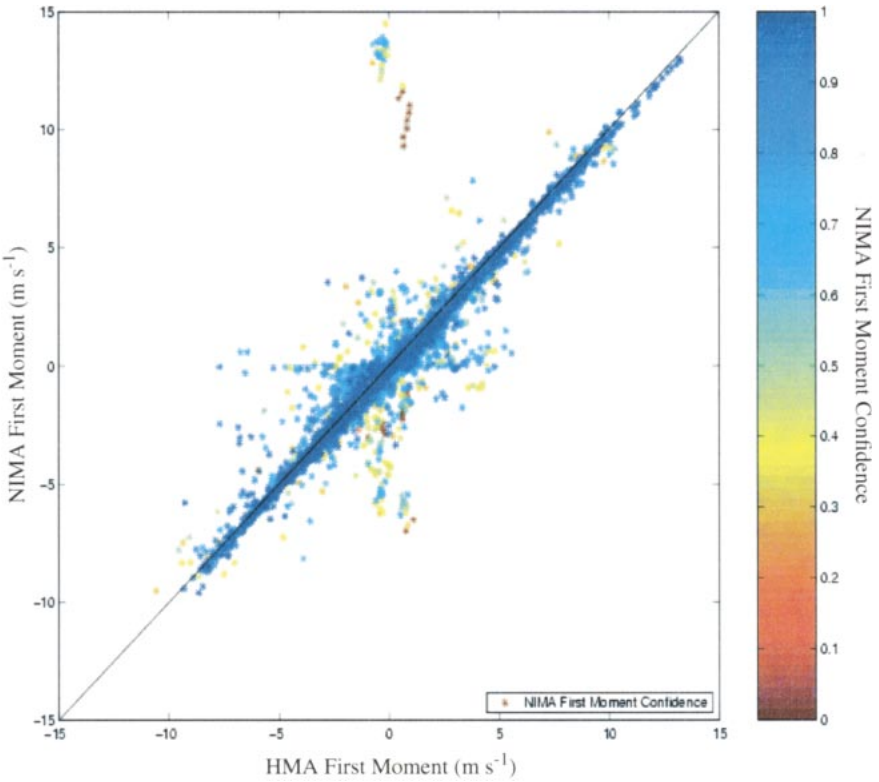


FIG. 6. The same data as in Fig. 4, but color coded by NIMA first-moment confidence. This coding shows that many outliers have low NIMA first-moment confidence [e.g., the points near (0, 10) in the scatterplot]. We can improve the overall accuracy of the result by excluding points with confidence below a chosen threshold of NIMA first-moment confidence.

cluster of points located near (0, 10) in the scatterplot]. This result suggests that if the NIMA first-moment data with confidence below a chosen threshold were excluded, the remaining data would have a higher average accuracy. The error (the average of the absolute value of the difference between the HMA and NIMA first moments) should decrease as the NIMA confidence threshold is increased. A further test of the NIMA confidence algorithm is to compare the data agreement using a higher NIMA confidence threshold with agreement using a higher human (HMA) confidence threshold.

In conducting this evaluation, it is necessary to represent the NIMA and HMA confidences on a common scale. The HMA confidences tended to be lower than those from the NIMA algorithm. Because of this fact, the moments are ranked by their HMA or NIMA confidence value, and fractions of the ranked data (rather than specific confidence thresholds) are considered when comparing the performance of the NIMA confidence to HMA confidence. The distribution of NIMA first-moment errors as a function of HMA confidence threshold is shown in Fig. 7. In the contour plot on the left, the left vertical axis shows the percent of data excluded based on low HMA first-moment confidence. Thus, at the level of 0.2 on the vertical axis, 20% of the data with the lowest HMA confidence have been excluded. A cross section at 0.2 would show the distribution of errors of all data in the top 80%, as ranked by HMA confidence. The right vertical axis gives the HMA confidence threshold corresponding to the fraction of points on the left axis. Thus, if we consider only data with an HMA confidence greater than 0.232, we have excluded 20% of the data and a contour at that height would be the distribution of errors (the signed differences of HMA first moments minus NIMA first moments). As we move up the vertical axis, more and more outliers are excluded. At 0.5 (HMA confidence of 0.411), a significant number of large outliers have been excluded. Above 0.7, only 30% of the data are left, and all errors are close to zero.

The right panels of this figure (Figs. 7b–d) show three ways to represent errors as a percent of data excluded. The top figure shows the average absolute error in radial velocity as a function of the fraction of data excluded. The blue curve in this graph shows an error of just over 0.3 m s^{-1}, with an excluded fraction of 0.2. This result means that if data with an HMA confidence above 0.232 (excluding the lowest 20% of the data ranked by HMA first-moment confidence) are included, the average of the absolute error for the remaining data is only slightly above 0.3 m s^{-1}, which is the velocity resolution of the Doppler spectra. The blue curve of the middle graph is

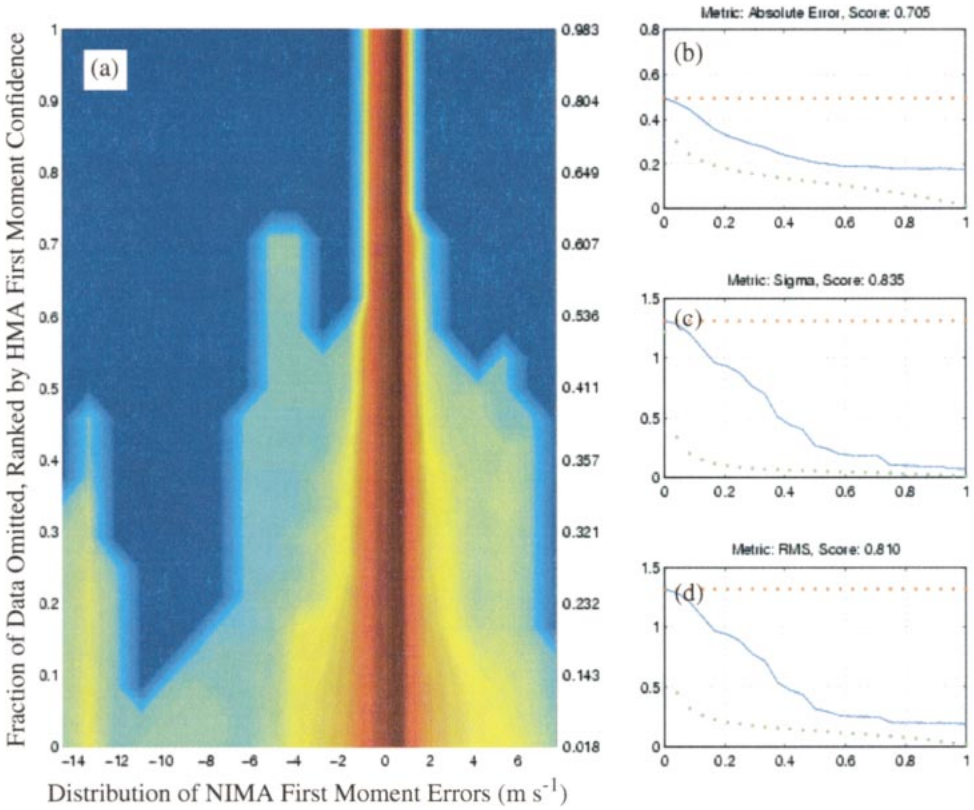


FIG. 7. (a) Contoured NIMA moment error distributions as a function of the fraction of data excluded when ranked by HMA confidence. The right scale indicates the HMA confidence threshold corresponding to the fraction excluded. (b) Average absolute error between NIMA and HMA moments as a function of the fraction of data excluded (blue). (c) Central second moment as a function of the fraction of data excluded. (d) Rms error as a function of the fraction of data excluded. See text for further explanation.

the second moment of the distribution as a function of the amount of data excluded, and the blue curve in the lowest graph is rms error as a function of the amount of data excluded. The second moment and rms error statistics are sensitive to outliers. Notice the steep drop in both statistics until 0.5 (when 50% of the data are excluded) as the large outliers on the left of the contour plot (with errors larger than -10 m s^{-1}) are removed from the data. These large outliers are the points in Fig. 6 around (0, 14). These points come from a dry case (with weak SNR) selected to test the NIMA confidence algorithms. Also, notice that both graphs are nearly level when more than 70% of the data are excluded, because the last of the outliers in the contour are removed. These graphs indicate that humans have skill in assigning confidence, given that the errors certainly decrease with increasing confidence. Some caution is needed here, because human verification is considered to be absolute truth. For example the outliers located near -14 m s^{-1} in the contour map are not excluded until nearly one-half the data have been rejected. This result means that the human expert had some confidence that these moments were correct.

A similar analysis as above, but now with NIMA first-moment confidence, is shown in Fig. 8. Notice that the contours indicate that the large outliers on the left (near -14 m s^{-1}) are eliminated by removing only 20% of the data as ranked by NIMA first-moment confidence. However, some outliers near -4 m s^{-1} persist at all confidence levels. From the blue curve in Fig. 8b, the error can be seen to be only slightly greater than 0.3 m s^{-1} after the lowest 20% of the data (as ranked by NIMA first-moment confidence) have been excluded. Notice that the lowest 20% of the data have a confidence threshold of 0.676 (the right vertical axis of the contour plot). This fact illustrates the necessity of studying the data as a function of the proportion of data excluded rather than directly from the confidence values. That is, the confidence scale is not necessarily linear. Note that the graphs for the variance and rms error do not drop off as rapidly beyond 0.2 as for the HMA first-moment confidence. This result suggests that there is room for improvement in the NIMA first-moment confidence algorithm.

A confidence index should have the property that errors (uncertainty) should decrease as confidence increases. The NIMA first-moment confidence index certainly has this property. If confidence were to make no dif-

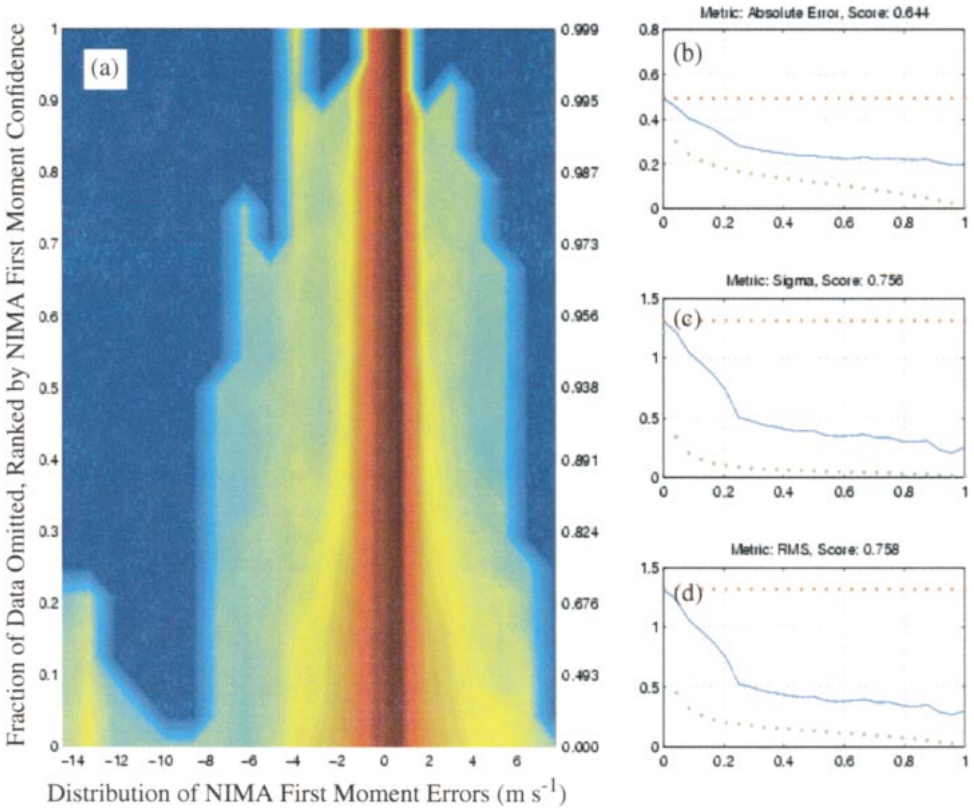


FIG. 8. Similar to Fig. 7, but excluding data ranked by low NIMA confidence.

ference in reducing error, then the blue curves would stay flat, as indicated by the red curves on the right-hand side of Figs. 7 and 8. Thus, if the blue curve remains close to the red curve, then increasing the confidence has no effect on performance and the confidence index has low skill. If the blue curve goes above the red curve, this means that, although higher confidence has been assigned, there is more error than in the original dataset. This outcome would be viewed as negative skill for such points. If confidence were perfectly correlated to error, an ordering of the data by error would be expected. This situation is indicated by the green curve in the right-hand graphs. Thus a blue curve that is close to the green curve indicates that the confidence algorithm has high skill. Of course, this assumes the human values are the truth.

A skill score based on these ideas has been developed and is indicated above each plot on the right side of Figs. 7 and 8. The score especially penalizes large errors at high confidence, but is related to the closeness of the blue and green curves. By this score, the human experts perform better than NIMA. This outcome is certainly expected, because, for this application, the fuzzy algorithms can only approach the skill of humans. The skill score can be used in evaluating the confidence maps in the NIMA fuzzy logic algorithm. As improvements are made in the confidence algorithms in NIMA, higher skill scores should result. One caveat is that, because the distribution of errors will be different for different datasets, the skill score can only be used to compare two confidence indices for the same set of moments and associated errors. This result is illustrated in Fig. 9, which shows an analysis of POP first-moment errors using SNR as a confidence index. This confidence index has a high skill score because the errors in first moments drop rapidly with increasing SNR. However, it is clear that the errors on the vertical axes of the graphs on the right of the figure are much larger than those in Fig. 8, so a more detailed comparison is needed. For example, removing 20% of the data in Fig. 8 by NIMA first-moment confidence results in an average absolute error of slightly more than 0.3 m s^{-1}. After removing 20% of the data by ordering with SNR, the remaining data for POP in Fig. 9 have an average absolute error of around 0.75 m s^{-1}. The differences are even more noticeable if rms and variance are compared, because POP has more outliers for this particular dataset.

c. *Wind magnitude and vector difference comparison*

Measurement of the horizontal wind vector is the objective of most wind-profiler measurements. To gauge the reliability of NWCA winds derived from NIMA first moments, a comparison with winds derived from the

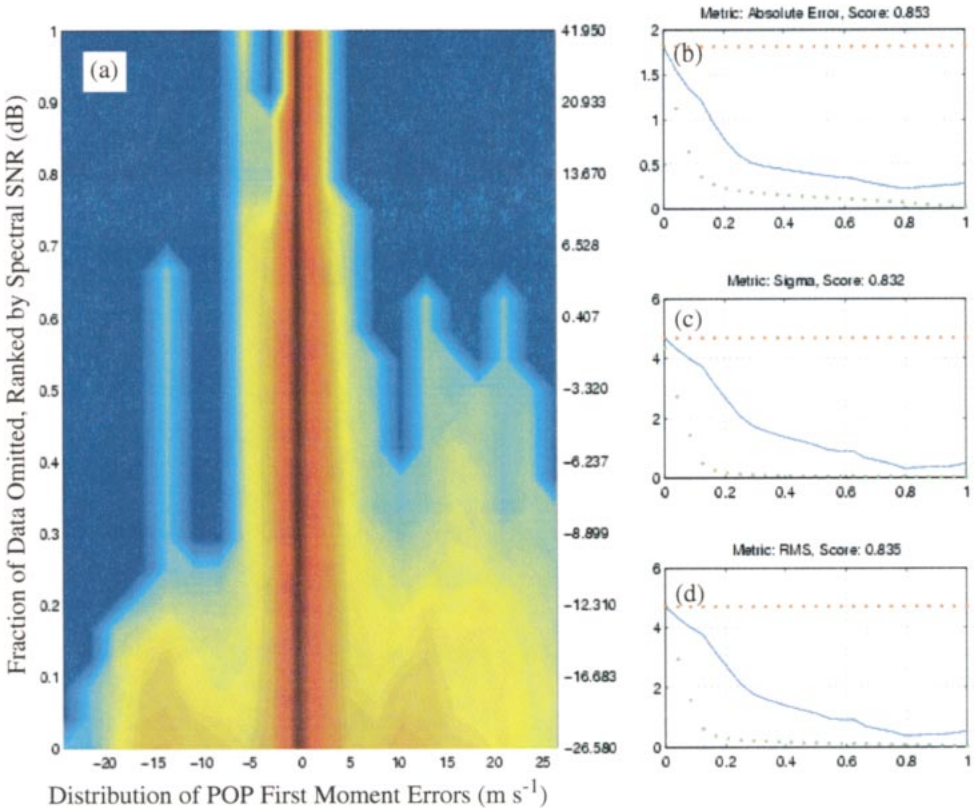


FIG. 9. Similar to Fig. 7, but contours of POP moment errors are shown as a function of fraction of data excluded by low SNR.

HMA data and, in section 4e, with winds measured by aircraft has been made. In computing winds, NWCA combines moments from a single set of the four beam directions and also provides a confidence based not only on the confidence in the moments but also continuity in time and range. To test the wind-confidence algorithm, data with low NWCA wind confidence are removed and the agreement of the remaining winds with those produced by HMA (also from a set of four beams) is evaluated. When low-confidence winds are removed, the remaining winds are reliable.

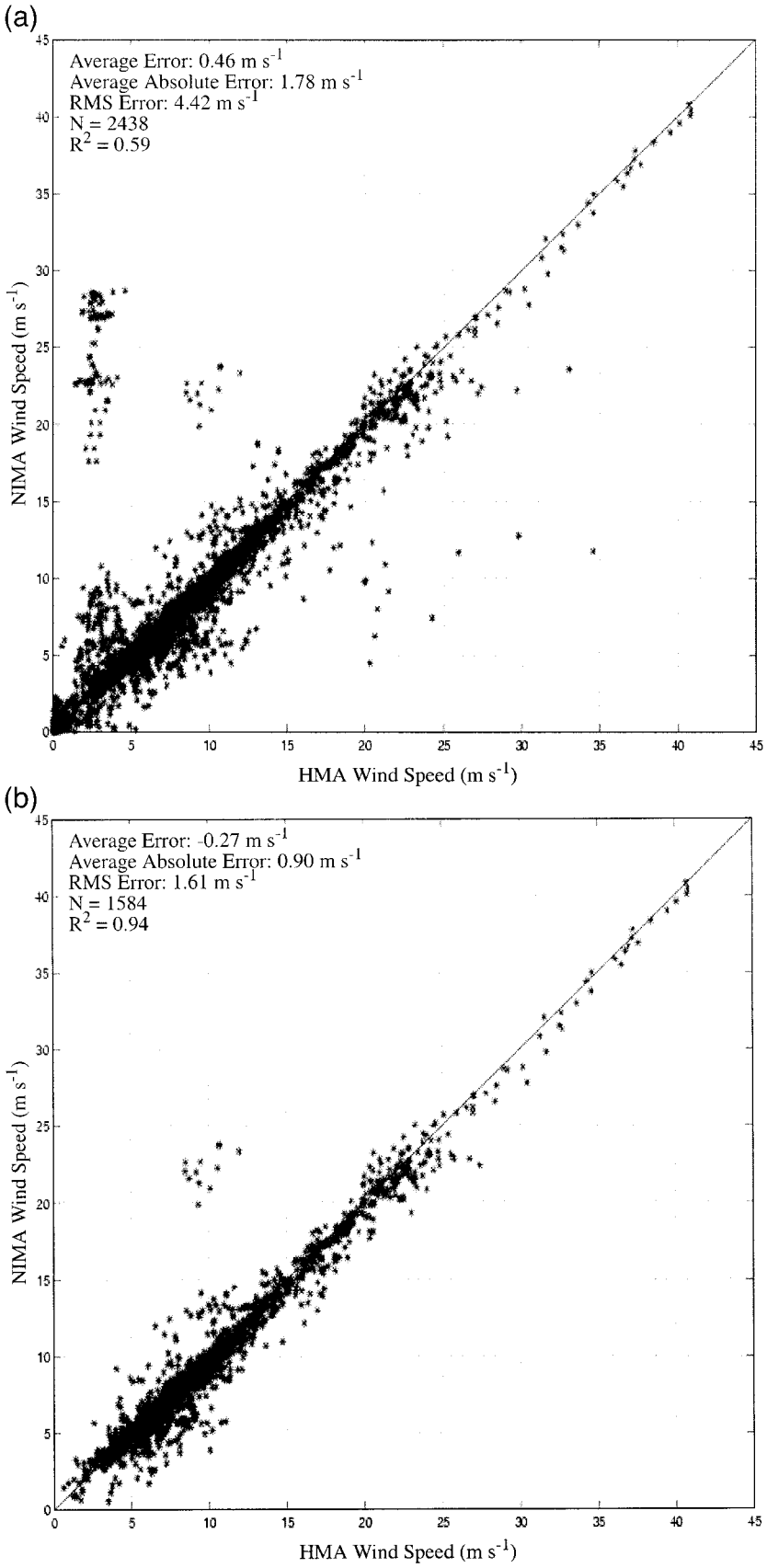
We compare NIMA–NWCA-derived wind speeds and HMA-generated wind speeds in Fig. 10a. Although the majority of wind speeds agree closely, the correlation is only 59% because of outliers. The large errors at the left of the scatterplot are low SNR cases for which NIMA selected a nonatmospheric signal. Many of the other points (35%) are from the malfunctioning South Douglas profiler, where the expected symmetry between opposing beams did not exist. This figure shows that it is necessary to remove poor-quality data to improve performance, because most applications would require higher correlation between NWCA winds and truth. The results of a linear regression between wind speed determined by HMA and by NWCA with the South Douglas data removed are seen in Fig. 10b. Notice the regression now explains 94% of the data. Most, but not all, of the outliers have been removed. The remaining outliers were low-SNR cases selected to test the NWCA wind-confidence algorithm.

d. Horizontal wind-confidence performance

The plot on the left of Fig. 11 represents the distribution of the magnitudes of the vector differences between winds estimated by HMA and NIMA–NWCA winds. The cross section of this plot at, for example, level 0.2 is the distribution of wind errors after 20% of the data as ranked by the NWCA wind confidence have been removed. The right axis shows that this corresponds to removing all data with an NWCA wind confidence below 0.497. The effect of removing a fraction

→

FIG. 10. Comparison of wind speed determined by the NIMA–NWCA algorithm and found from HMA moments. (a) All data are included; 59% of the data are explained by linear regression. (b) Data from the South Douglas wind profiler have been excluded because of intermittent mechanical switch failure; 94% of these data are explained by linear regression.



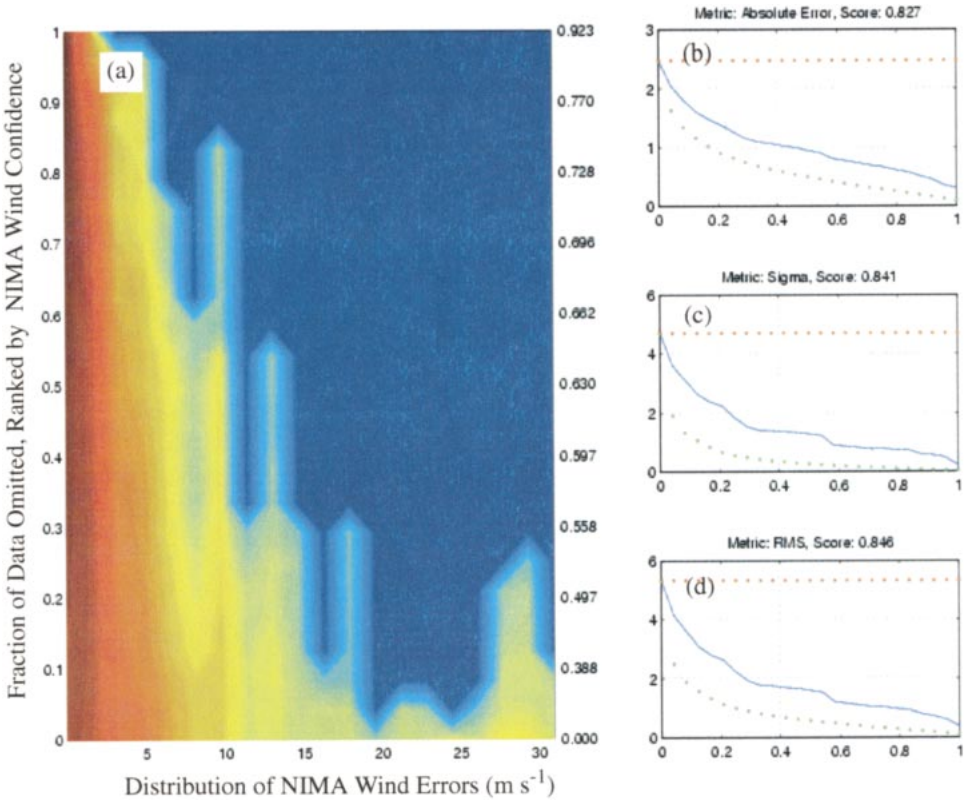


FIG. 11. (a) Contoured NIMA–NWCA wind error distributions as a function of the fraction of data excluded when ranked by low NWCA wind confidence. The right scale indicates the NWCA confidence threshold corresponding to the fraction excluded. (b) Average absolute error between NIMA–NWCA and HMA winds as a function of the fraction of data excluded. (c) Central second moment as a function of the fraction of data excluded. (d) Rms error as a function of the fraction of data excluded.

of the data when ranked by NWCA wind confidence is seen in several error metrics in Figs. 11b–d.

The data in Fig. 11 use all of the winds estimated by HMA, including the South Douglas profiler data for which there was a switching problem. Notice that, as the confidence index is increased, the number of large outliers is reduced. Also, the error metrics decrease as the confidence index increases. For example, when 40% of the data are removed, the average absolute error is reduced to approximately 1 m s^{-1}, the value of sigma reduces to about 1.4 m s^{-1}, and the root-mean-square error reduces to about 1.8 m s^{-1}. The bottom two curves on the right of the figure give an indication of the effect of outliers. The contour plot on the left shows that a large number of outliers have been removed when 40% of the data are removed (including most of the South Douglas data, which account for more than 30% of the total). These graphs show that NWCA wind confidence has skill in removing poor-quality data, and the remaining data have a significant reduction in error when compared with the HMA-produced winds. Although this dataset contains the South Douglas data when the winds were not reliable, the NWCA confidence index is able to remove these and other data that have large errors when compared with the HMA winds.

The NIMA–NWCA wind speed and HMA wind speed are compared in Fig. 12 after removing the 30% of data with the lowest NWCA wind confidence. This removal was done by applying a confidence threshold of 0.557. About 96% of the data are explained by linear regression, with an average absolute error of 0.74 m s^{-1}. Also, notice that the number of outliers has been reduced when compared with Fig. 10a. In fact, the results are slightly better than when all of the South Douglas data in the regression (Fig. 10b) were removed. This result shows that, even if a profiler has a switching problem and the winds are not reliable, the NWCA wind confidence algorithm is likely to detect the poor quality of the resulting winds.

The mean NWCA wind confidence for the truth-specified data set is 0.66, whereas a mean of 0.78 was found when a 9-month-long dataset was analyzed with NIMA–NWCA. This result suggests that the truth-specified dataset contains many difficult cases. Examining the distribution of confidences for the full 9 months, we find that applying an NWCA confidence threshold of 0.557

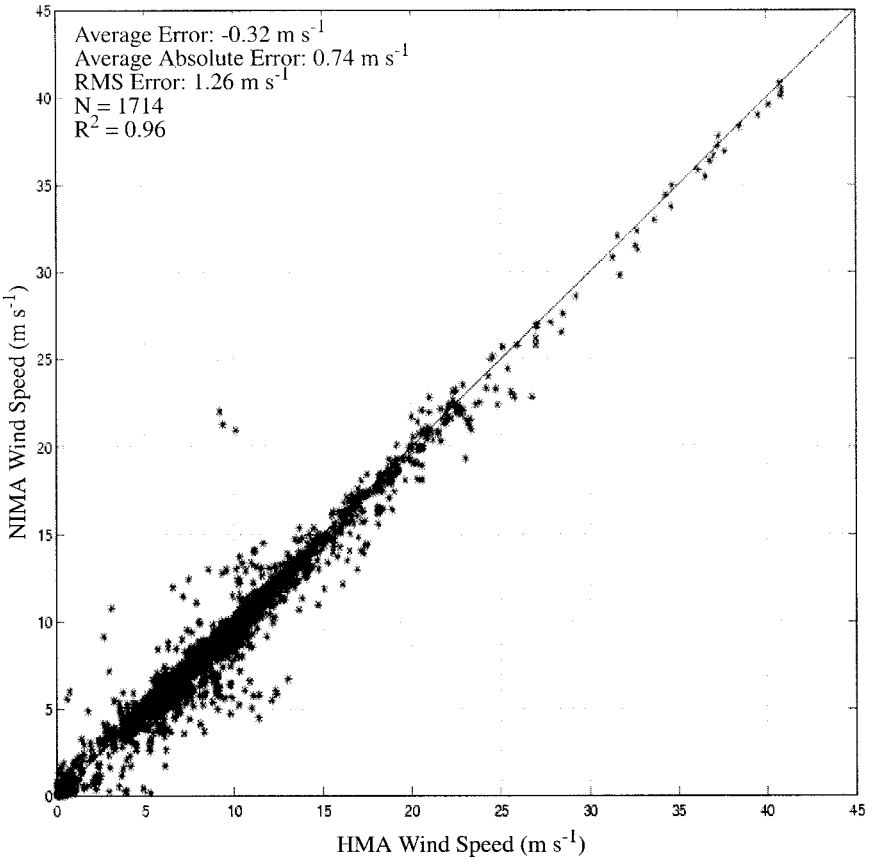


FIG. 12. Similar to Fig. 10a, but data with NWCA wind confidence less than 0.557 are excluded, removing approximately 30% of the data; 96% of data variability is accounted for by linear regression.

removes only 3% of the data. Because similar performance of confidence is expected for the truth-specified and 9-month datasets, it is expected that removing 3% of the 9-month dataset would result in an average absolute error of about 0.74 m s^{-1} in the wind speed and about 1.1 m s^{-1} in the vector wind.

e. Aircraft wind comparison

Comparison between the profilers and aircraft winds is useful because these measurements are completely independent, unlike the HMA winds, which are derived from the same profiler spectra as the NIMA–NWCA winds. However, winds measured by instruments at separated locations around Juneau are highly variable because of the complex terrain. This variation can be true of elevated winds as well because of waves and turbulence generated by flow over the mountains. A scatterplot of 124 horizontal wind measurements by the North Douglas wind profiler and by the aircraft at nearly the same place and time is shown in Fig. 13. There is no comparison with the Lemon Creek wind profiler, because the aircraft was always turning near this site, which affected the aircraft data quality, nor with the South Douglas wind profiler, because of the beam-steering problem at that site. The North Douglas comparison shows general agreement but with some large outliers and an overall variance of $R^2 = 0.64$. To create this comparison, aircraft data (taken at 1-s resolution) within 1 km horizontally and 30 m vertically of the center of the profiler pulse volume were averaged. The color-coded wind profiler wind confidence shows that most points have a high confidence, including the outliers. This comparison can not separate errors due to the profiler from those due to the aircraft. However, the worst outliers are from two days (9 March and 17 March 1998). The wind profiler spectral data have been examined closely, and the data were found to be generally clean, and so a calibration problem in the aircraft measurement is suspected. When data from these two days (14 circled points in Fig. 13) are removed, the remaining data have a variance of $R^2 = 0.86$.

5. Discussion and conclusions

An evaluation of the NIMA–NWCA algorithms to produce accurate first moments and rapid-update winds from boundary layer wind profiler spectral data has been

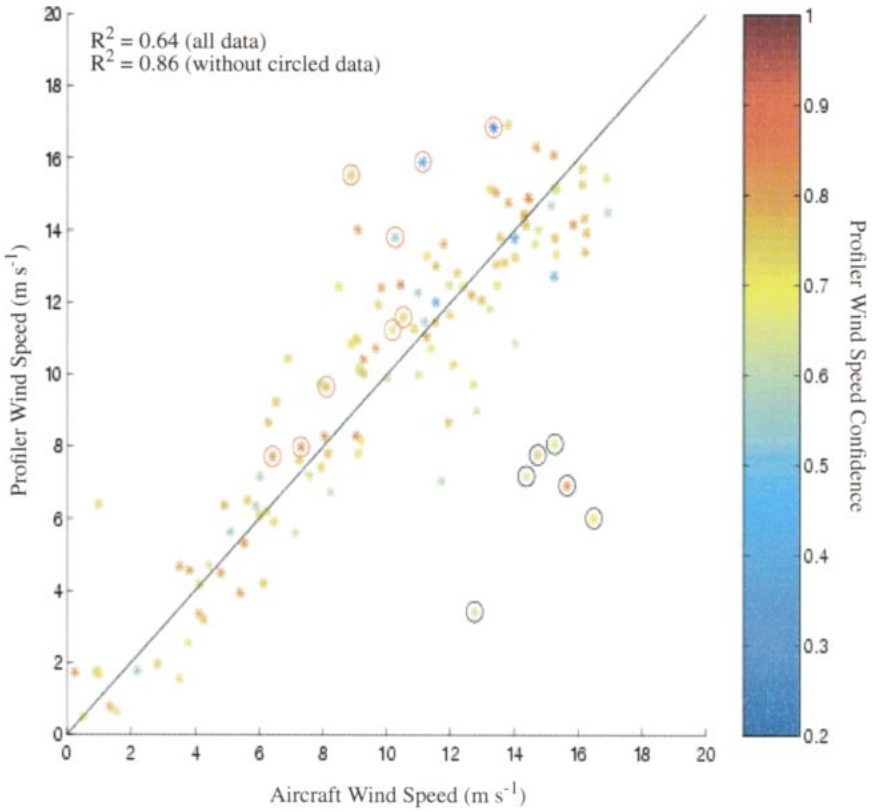


FIG. 13. Comparison of wind speed determined by the NIMA–NWCA algorithm and by an average of nearby aircraft measurements; 64% of the data are explained by linear regression. Circled points are from 9 Mar 1998 (black circles) and 17 Mar 1998 (red circles) and are suspect; 86% of the data are explained by linear regression if these points are removed.

presented. These algorithms have been designed to perform in real time within challenging environments and without human intervention. Although it has been demonstrated that the NIMA algorithm performs well, there are cases when the data quality is poor and the rapid-update winds are unreliable. To address this problem, a fuzzy logic wind-confidence algorithm has been developed and evaluated. One common criticism of rapid-update winds from a four- or five-beam wind profiler is the assumption of stationarity of the wind field during the beam cycle. The NWCA confidence algorithm partially addresses this concern by testing for this assumption and lowering confidence in the wind if it is not met. The NIMA–NWCA algorithm has been shown to have skill in recognizing poor data. For the dataset analyzed, we found that removing the 25% of data with the least confidence as determined with the NIMA algorithm reduces the RMS error (judged against a human-specified truth) from about 1.3 to 0.5 m s^{-1}. On a longer and less demanding dataset, we found that, if the lowest 3% of data (as defined by the NWCA confidence algorithm) in the Juneau environment are removed, the error in the wind will be on average about 0.74 m s^{-1}. We have shown that human confidence outperforms the NWCA wind-confidence algorithm. During this study, we further discovered that the NIMA first-moment confidence algorithm, which effects the NWCA wind confidence, could be improved to mimic more closely the human expert.

The reliability of winds and the confidence associated with NIMA–NWCA results make the Juneau wind profilers a more useful tool for critical applications such as an airport warning system. Although the NIMA–NWCA algorithm has been used primarily with data collected in Juneau with UHF boundary layer wind profilers, the same algorithm has been applied to other boundary layer profiler data. For these data, we found that the algorithm needed minor modifications to adapt to situations not seen in Juneau. We expect that, as more diverse data are analyzed, NIMA–NWCA will become more robust and more easily applied to Doppler spectra from any environment and wind profiler, including for example, the NOAA wind-profiler network. Further evaluation is needed to test NIMA performance on spectra from a vertical beam. Future development is expected to include application to temperature measurements using radioacoustic sounding system (e.g., Angevine et al. 1994) spectra and development of a precipitation recognition algorithm to improve performance in rain, snow, and when multiple peaks are present.

Acknowledgments. We thank all the participants in the 1998 field experiment for their effort, especially the staff of the NCAR Research Applications Program and of the University of North Dakota. We thank A. Rodi of the University of Wyoming for additional aircraft data processing. We also thank J. Stone for helping with the data processing. This research was in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

REFERENCES

Angevine, W. M., W. L. Ecklund, D. A. Carter, K. S. Gage, and K. P. Moran, 1994: Improved radio acoustic sounding techniques. *J. Atmos. Oceanic Technol.,* **11,** 42–49.

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.

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.

Cornman, L. B., R. K. Goodrich, C. S. Morse, and W. L. Ecklund, 1998: A fuzzy logic method for improved moment estimation from Doppler spectra. *J. Atmos. Oceanic Technol.,* **15,** 1287–1305.

Dierking, C. F., 1998: Effects of a mountain wave windstorm at the surface. *Wea. Forecasting,* **13,** 606–616.

Ecklund, W. L., D. A. Carter, B. B. Balsley, P. E. Courier, J. L. Green, B. L. Weber, and K. S. Gage, 1990: Field tests of a lower tropospheric wind profiler. *Radio Sci.,* **25,** 899–906.

May, P. T., and R. G. Strauch, 1989: Examination of wind profiler signal processing algorithms. *J. Atmos. Oceanic Technol.,* **6,** 731–735.

Merritt, D. A., 1995: A statistical averaging method for wind profiler Doppler spectra. *J. Atmos. Oceanic Technol.,* **12,** 985–995.

Weber, B. L., D. B. Wuertz, D. C. Welsh, and R. McPeek, 1993: Quality controls for profiler measurements of winds and RASS temperatures. *J. Atmos Oceanic Technol.,* **10,** 452–464.

Yager, R. R., S. Ovchinnikov, R. M. Tong, and H. T. Nguyen, 1987: *Fuzzy Sets and Applications, Selected papers by L. A. Zadeh.* John Wiley and Sons, 685 pp.

Zimmerman, H. J., 1996: *Fuzzy Set Theory and its Applications.* 3d ed. Kluwer Academic, 435 pp.