

## An Automated Technique to Quality Control Radar Reflectivity Data

VALLIAPPA LAKSHMANAN

*Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma, and National Oceanic and Atmospheric Administration/National Severe Storms Laboratory, Norman, Oklahoma*

ANGELA FRITZ

*Valparaiso University, Valparaiso, Indiana*

TRAVIS SMITH

*Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma, and National Oceanic and Atmospheric Administration/National Severe Storms Laboratory, Norman, Oklahoma*

KURT HONDL

*National Oceanic and Atmospheric Administration/National Severe Storms Laboratory, Norman, Oklahoma*

GREGORY STUMPF

*Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma, and National Weather Service/Meteorological Development Laboratory, Silver Spring, Maryland*

(Manuscript received 10 October 2005, in final form 22 June 2006)

### ABSTRACT

Echoes in radar reflectivity data do not always correspond to precipitating particles. Echoes on radar may result from biological targets such as insects, birds, or wind-borne particles; from anomalous propagation or ground clutter; or from test and interference patterns that inadvertently seep into the final products. Although weather forecasters can usually identify and account for the presence of such contamination, automated weather-radar algorithms are drastically affected. Several horizontal and vertical features have been proposed to discriminate between precipitation echoes and echoes that do not correspond to precipitation. None of these features by themselves can discriminate between precipitating and nonprecipitating areas. In this paper, a neural network is used to combine the individual features, some of which have already been proposed in the literature and some of which are introduced in this paper, into a single discriminator that can distinguish between “good” and “bad” echoes (i.e., precipitation and nonprecipitation, respectively). The method of computing the horizontal features leads to statistical anomalies in their distributions near the edges of echoes. Also described is how to avoid presenting such range gates to the neural network. The gate-by-gate discrimination provided by the neural network is followed by more holistic postprocessing based on spatial contiguity constraints and object identification to yield quality-controlled radar reflectivity scans that have most of the bad echo removed while leaving most of the good echo untouched. A possible multisensor extension, utilizing satellite data and surface observations, to the radar-only technique is also demonstrated. It is demonstrated that the resulting technique is highly skilled and that its skill exceeds that of the currently operational algorithm.

### 1. Introduction

From the point of view of automated applications operating on weather data, echoes in radar reflectivity

may be contaminated by undesirable echoes. Some of these applications require that echoes in the radar reflectivity moment correspond, broadly, to precipitation, or “weather.” By removing ground-clutter contamination, rainfall obtained from the radar data of the National Weather Service (NWS) Weather Surveillance Radar-Doppler 1988 (WSR-88D) can be improved (Fulton et al. 1988; Krajewski and Vignal 2001). A large

---

*Corresponding author address:* V. Lakshmanan, NSSL/RRDD, 120 David L. Boren Blvd., Norman, OK 73071.  
E-mail: lakshman@ou.edu

number of false positives for the “mesocyclone detection algorithm” (Stumpf et al. 1998) are caused in regions of clear-air return (McGrath et al. 2002; Mazur et al. 2004). A hierarchical motion estimation technique segments and forecasts poorly in regions of ground clutter (Lakshmanan 2001). Hence, a completely automated algorithm that can remove regions of nonprecipitating echo, such as ground clutter, anomalous propagation, radar artifacts, and clear-air returns, from the radar reflectivity field would be very useful in improving the performance of other automated weather-radar algorithms.

Steiner and Smith (2002) describe the causes, effects, and characteristics of such contamination in weather-radar data. Several research groups have endeavored to determine individual features and combinations of features that can be used to remove range gates of radar reflectivity data that correspond to “bad echoes.” Local neighborhoods in the vicinity of every range gate in the three WSR-88D radar moments (reflectivity, velocity, and spectrum width) were examined by Kessinger et al. (2003) and used for automated removal of nonprecipitating echoes. They achieved success by examining some local statistical features (the mean, median, and standard deviation within a local neighborhood of each gate in the moment fields) and a few heuristic features. Steiner and Smith (2002) introduced the “spinChange” (hereinafter referred to as SPIN), which is the ratio of gate-to-gate differences in an  $11 \times 21$  local neighborhood that exceed a certain threshold (2 dBZ in practice) to the total number of such differences. Kessinger et al. (2003) introduced the “SIGN,” which is the average of the signs of the gate-to-gate difference field within the local neighborhood. Steiner and Smith (2002) used a decision tree to classify range gates into two categories (precipitation and nonprecipitation), and Kessinger et al. (2003) used a fuzzy-rule base using features that included the SPIN feature introduced by Steiner and Smith (2002). In addition to these features based on a single elevation scan, some vertical-profile features were also used—the maximum height of a 5-dBZ echo was used by Steiner and Smith (2002). Kessinger et al. (2003) discussed the use of vertical differences between the two lowest reflectivity scans. Zhang et al. (2004) used the features introduced in Steiner and Smith (2002) but calculated them with respect to physical height instead of radar tilts (elevation angles).

Neural networks (NNs) have been utilized in a variety of meteorological applications. For example, NNs have been used for prediction of rainfall amounts by Venkatesan et al. (1997) and for the diagnosis of tor-

nado probability from mesocyclones by Marzban and Stumpf (1996). In fact, Cornelius et al. (1995) attempted to solve the radar quality problem using NNs. However, the performance of the NN was no better than a fuzzy-logic classifier (C. Kessinger 2004, personal communication), and the NN attempt was dropped in favor of the much more transparent fuzzy-logic approach described in Kessinger et al. (2003).

Greco and Krajewski (2000) also used NNs to classify pixels in WSR-88D data as either anomalous propagation/ground clutter (AP/GC) or good echo. The technique of this paper goes beyond that work by following a formal mechanism for evaluating and selecting input features to the NN. We also describe how to account for statistical anomalies in the way the NN inputs are computed, especially at the edges of echoes. A method of selective emphasis is followed here to ensure good performance on significant echoes. Last, the technique described in this paper removes or retains entire echo regions, not just individual pixels.

A particular challenge in the quality control (QC) of radar reflectivity data is that errors in the QC process can be additive from the point of view of downstream applications. This effect is particularly noticeable in applications that accumulate radar reflectivity data in space or time. For example, an application that combines radar data from four radars to form its results will be affected by QC errors in any of the four radars. In a similar way, an application that accumulates data from one of the radars over time will be faced with increasing likelihood that the input data are incorrect as the time window of accumulation increases. Radar data mosaics and precipitation accumulation algorithms, therefore, have very stringent requirements on data quality. To put these requirements in perspective, consider using a hypothetical single-radar QC algorithm that is correct 99% of the time. For an application that combines data from 130 WSR-88Ds in the continental United States, the odds of there being bad data somewhere in the country at any time would be 73% ( $1 - 0.99^{130}$ ). A 3-h accumulation of precipitation based on data from a single radar collected every 5 min will be wrong 30% ( $1 - 0.99^{36}$ ) of the time.

Section 2 is organized as follows. Following a description of the machine learning technique used, the full set of candidate features is described in section 2a. The training process itself is described in section 2b. Based on the results of training, feature selection was implemented and the training was repeated on the selected set of features. The feature selection and the final list of inputs are described in section 2c. Preclassification methods were devised to ameliorate shortcomings in the way the local features were computed. The result of

the classifier is also subject to spatial postprocessing. Both the preprocessing and the postprocessing are described in section 2d. Section 3 describes the results achieved on an independent test set and discusses the implications. Section 4 summarizes the contents of the paper.

## 2. Method

We used a resilient back-propagation NN (RPROP) as described in Riedmiller and Braun (1993) with one hidden layer. Every input node was connected to every hidden node, and every hidden node was connected to the output node. In addition, there was a short-circuit connection from the input nodes directly to the output node to capture any linear relationships; that is, the network was “fully connected” and completely “feed forward.” Every hidden node had a “tanh” activation function, chosen because of its signed range. The output unit had a sigmoidal activation function,  $g(a) = (1 + e^{-a})^{-1}$ , so that the outputs of the networks could be interpreted as posterior probabilities (Bishop 1995). Each noninput node had, associated with it, a bias value that was also part of the training.

The error function that was minimized was a weighted sum of the cross entropy [which Bishop (1995) suggests is the best measure of error in binary classification problems] and the squared sum of all of the weights in the network:

$$E = E_e + \lambda \sum w_{ij}^2. \quad (1)$$

The first term is a variation of the cross-entropy error suggested by Bishop (1995) and is defined as

$$E_e = - \sum_{n=1}^N c^n [t^n \ln y^n + (1 - t^n) \ln(1 - y^n)], \quad (2)$$

where  $t^n$  is the target value of the  $n$ th set of input features, called a pattern. The target value for a pattern is 0 if the range gate has bad echoes and 1 if it has “good echoes.” Here,  $y^n$  is the actual output of the NN for that pattern input;  $N$  is the total number of patterns. The cost  $c^n$  captures the importance of that pattern (this process of selective emphasis is explained in more detail in section 2b). The second, squared-weights, term of Eq. (1) attempts to reduce the size of the weights and thus improves generalization (Krogh and Hertz 1992). The relative weight  $\lambda$  of the two measures is computed every 50 epochs within a Bayesian framework with the assumption that the weights and the errors have Gaussian distributions so that the ratio of their variances gives a measure of by how much to decay the weights (MacKay 1992; Bishop 1995). We started by weighing

the sum of weights 2 times as much as the cross-entropy term ( $\lambda = 2$ ), we updated  $\lambda$  based on the distribution of the weights and errors every 50 epochs (one epoch is a complete pass through the entire set of training patterns), and we stopped the learning process at 1000 epochs. We chose the final weights of the network from the epoch at which the *validation entropy error* was minimum, as will be discussed in section 2b.

Doppler velocity data can be range folded (aliased). For WSR-88D, at the lowest tilt, the velocity scan has a shorter range than does the reflectivity scan. We therefore divided the training range gates into two groups—one in which velocity data were available and another in which there was no Doppler velocity (or spectrum width) information. Thus, two separate NNs were trained. In real-time operation, either the with-velocity or the reflectivity-only network is invoked for each range gate depending on whether there are velocity data at that point.

At the end of training, the with-velocity network had 28 inputs, 4 hidden nodes, and 1 output and the reflectivity-only network had 22 inputs, 8 hidden nodes, and 1 output.

### a. Input features

We chose as potential inputs to the NN features culled from earlier quality-control attempts—the AP and GC algorithm of Steiner and Smith (2002), the Next Generation Weather Radar (NEXRAD) radar echo classification (REC) algorithm (Kessinger et al. 2003), and the inputs used in the quality-control preprocessing in NEXRAD precipitation products (Fulton et al. 1998). These included the data value, the mean, the median, and the variance computed in local neighborhoods of each of the three moments (reflectivity, velocity, spectrum width) at the lowest tilt of the radar. In addition, we took the same four values for the second lowest tilt of the radar. We also utilized some of the textural features that have been found to be useful in discriminating between precipitation and AP/GC. These were the SPIN (Steiner and Smith 2002), the gate-to-gate average square difference (Kessinger et al. 2003), and the SIGN (Kessinger et al. 2003). We included the vertical gradient (difference between the reflectivities at the two lowest scans; Fulton et al. 1998) as another input to the NN.

In order that it can operate in real time, the QC algorithm should not wait for the end of a volume scan before being able to produce QCed reflectivity scans. To operate tilt by tilt in real time, it should be able to compute a feature such as the vertical difference (which depends on the two lowest tilts) even before the second tilt arrives. Otherwise, downstream algorithms will be

faced with a 20–60-s delay. Therefore, all of the vertical features were computed in a “virtual volume” (Lynn and Lakshmanan 2002) sense; that is, the latest available elevation scan at any instant in time was used to fill in the volume. For example, the vertical gradient would be updated twice within a volume scan: once at the end of the first elevation scan when the higher elevation scan used was from the last volume scan and again when the second elevation scan from the volume scan arrived. As a result, in real time, the QC of the first elevation scan is performed using the vertical-difference feature computed from the staggered difference but the QC of all the higher tilts is performed using the aligned difference. Because the NN is trained on such staggered differences, it learns to account for them.

In addition to these discriminants described in the literature, we utilized the following others:

- 1) the local minimum of the variance, calculated by first computing the variance within a neighborhood of every gate and then computing the local minimum in neighborhood around each gate of the variance field (This quantity is computed for the lowest two reflectivity scans and the lowest velocity and spectrum width scans. For example, the variance of the reflectivity field could be large either because of high texture, such as in AP/GC, or because of edge effects. Using the minimum variance avoids large variance close to the edge of echoes.),
- 2) the mean and local variance computed in a  $5 \times 1$  neighborhood, that is, along a radial for the lowest reflectivity scan (many problems in radar data, especially hardware issues, may impact a single radial while leaving adjacent radials unaffected),
- 3) the maximum vertical reflectivity, over all of the elevations (like all vertical features, the vertical maximum is computed in a virtual volume sense),
- 4) the fraction of range gates in a  $5 \times 5$  local neighborhood that had values greater than 5 dBZ in the lowest reflectivity scan and in the vertical maximum field (this feature, referred to as “echo size” in this paper, is useful in detecting speckle),
- 5) a weighted average of the reflectivity (dBZ) values over all the elevations, where the weight of each data point is given by the height of that range gate above the radar (This takes into account the entire vertical profile instead of just the first two elevations. A large weighted average is associated with tall thunderstorms and provides a counterbalance to the large horizontal gradients associated with such thunderstorms. AP/GC, on the other hand, has small weighted averages but also has large horizontal gradients.),
- 6) echo-top height, defined as the maximum height of reflectivity above 0-, 5-, and 10-dBZ thresholds (convective storms tend to have large values of echo-top heights),
- 7) the physical height at which the maximum reflectivity echo occurs at this range and azimuth from the radar (higher echoes are typically more reliable),
- 8) the change in the maximum vertical reflectivity from this virtual volume to the last complete volume (good echoes rarely have large temporal gradients),
- 9) the outbound distance along a radial to a gate with zero velocity, that is, at a particular gate, the distance to a gate further away from the radar with zero velocity (A gate with zero velocity is likely to be AP/GC. Gates between the radar and the AP gate are likely to also be AP, making this input potentially useful.), and
- 10) the inbound distance along a radial to a gate with an echo-top over 3 km, that is, at a particular gate, the distance to a gate closer to the radar that has an echo top higher than 3 km (this input is potentially useful because gates that come after such a high echo gate are likely to be good).

All of the NN inputs were scaled such that each feature in the training data exhibited a zero mean and a unit variance when the mean and variance are computed across all patterns in the entire with-velocity training set. The mean and variance computed on the training set are used to scale the run-time values.

Histograms of a few selected features are shown in Fig. 1. Note that these features are not linear discriminants by any means—it is the combination of features that gives the NN its discriminating ability. The histograms in Figs. 1c and 1d illustrate the result of selective emphasis (explained in section 2b) that we adopt during the training so that larger reflectivities are not automatically accepted.

We used these features to train an NN and then removed the features one at a time. If the resulting NN had essentially the same performance after a feature was removed, that feature was permanently removed. The final list of features used is listed in section 2c.

#### b. Training process

With the NN architecture arbitrarily fixed as above, the training process consists of choosing 1) the number of hidden nodes, 2) the weights of each link between nodes, and 3) the set of input features to use. This section describes our approach to selecting the weights and number of hidden nodes of the NN. Section 2c

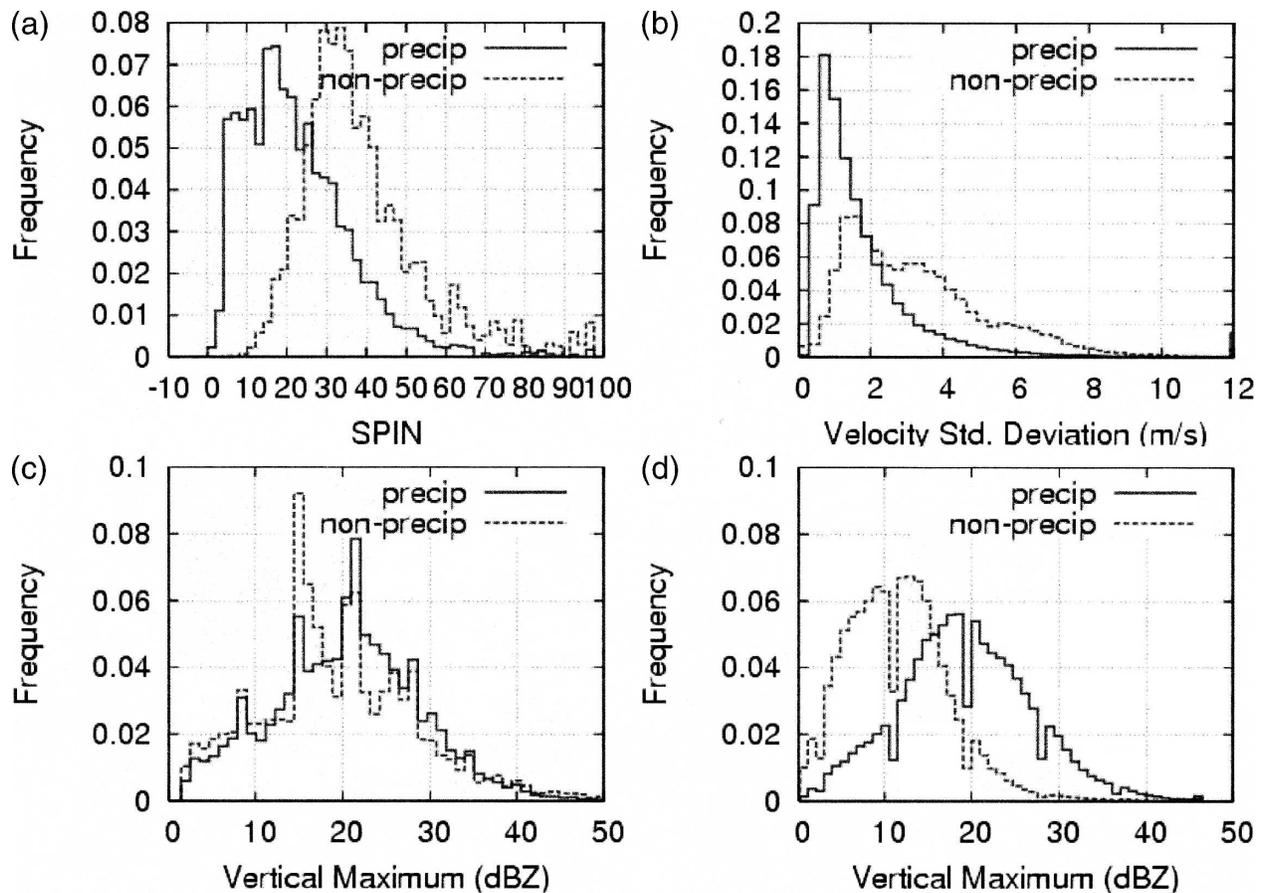


FIG. 1. Histograms of selected features over the entire training dataset: (a) SPIN, (b) standard deviation of velocity, and (c) vertical maximum. Note that, as a result of careful construction of the training set and selective emphasis (see section 2b), the histograms in (c) are nearly identical—this is not the a priori distribution of the two classes because AP is rare and clear-air return tends to be smaller reflectivity values. (d) The a priori distribution of the vertical maximum values in the training set before selective emphasis.

describes the method of choosing the input features. An NN is, essentially, an evidence-based framework. The weights were chosen to minimize the cross-entropy measure on a carefully selected set of radar volume scans. These volume scans covered a variety of phenomena that the resulting network was expected to discriminate with skill (see Table 1).

A human interpreter examined these volume scans and drew polygons using the “warning decision support system-integrated information” (WDSS-II) display (Hondl 2002) to select bad-echo regions (see Fig. 2). The expert determined this based on analyzing loops of reflectivity data, examining the velocity data, and considering outside factors such as the presence or absence of terrain and seasonal effects. For situations in which the expert was doubtful, the echoes were accepted as good echo so that the QC procedures would err on the side of caution. An automated procedure used these human-generated polygons to classify every range gate

of the training dataset into the two categories (precipitating and nonprecipitating).

The data we have are not representative of true a priori probabilities, because each of the scenarios is a rare event. Patterns are assigned different importance factors  $c^n$  [see Eq. (2)]. It is easy to see that if the cost factors  $c^n$  are positive integers, the cost factor can be moved out of the error equation by simply repeating the  $n$ th pattern  $c^n - 1$  times.

The importance factors are chosen based on the vertical maximum of the reflectivity (a process we call “selective emphasis”) to satisfy two goals: 1) to achieve an a priori probability, given any reflectivity value, of close to 0.5, so that the network is forced to rely on other factors and 2) to achieve better performance on strong echoes than on weak ones on the grounds that strong echoes are more important.

In the velocity network (a proxy for range gates close to the radar; see section 2), precipitating echoes are

TABLE 1. The volume scans used to train the WSR-88D QC NN described in this paper.

Radar	Location	Volume-scan start time	Reason chosen
KABR	Aberdeen, SD	2028:46 UTC 28 Jul 2005	Biological
KAMA	Amarillo, TX	2047:19 UTC 29 Apr 2004	Sun spikes, small cells
KBHX	Eureka, CA	1617:09 UTC 2 May 2003	Stratiform
KFDR	Fredrick, OK	1419:38 UTC 3 May 2005	Weak convection
KFWS	Fort Worth, TX	0358:51 UTC 19 Apr 1995	Biological, ground clutter
KFWS	Fort Worth, TX	1555:37 UTC 19 Apr 1995	Weak convection
KFWS	Fort Worth, TX	0459:04 UTC 20 Apr 1995	Insects, convection
KFWS	Fort Worth, TX	1855:57 UTC 20 Apr 1995	AP, squall line
KICT	Wichita, KS	0609:24 UTC 24 Apr 2003	Convection
KLIX	New Orleans, LA	1409:42 UTC 17 Aug 2005	Weak convection
KLSX	St. Louis, MO	0537:17 UTC 10 Nov 1998	Clutter near radar, convection
KMVX	Grand Forks, ND	1955:47 UTC 1 Jan 2005	Snow
KTLX	Oklahoma City, OK	1251:21 UTC 18 Oct 2002	Stratiform, sun spike
KTLX	Oklahoma City, OK	1613:43 UTC 19 Oct 2002	Convection, small cells
KTLX	Oklahoma City, OK	1829:11 UTC 29 Apr 2004	Sun spikes, biological
KTLX	Oklahoma City, OK	1805:40 UTC 5 May 2005	Test pattern
KUDX	Rapid City, ND	0229:57 UTC 15 Feb 2005	Snow
KYUX	Yuma, AZ	1511:46 UTC 1 Sep 2005	AP far away from radar

repeated  $d/20$  times while nonprecipitating echoes are repeated  $d/10$  times, where  $d$  is the reflectivity value. Thus, AP with large reflectivity values is emphasized, as are strong reflectivity cores. The emphasis is done in such a way that the resulting a priori probability in the

training set at any reflectivity value is close to 0.5, yielding a balanced distribution of patterns. In the reflectivity-only network, nonprecipitating echoes are repeated  $3d/5$  times. As can be seen from Eq. (2), the repeating of patterns has the same effect as imposing a cost factor

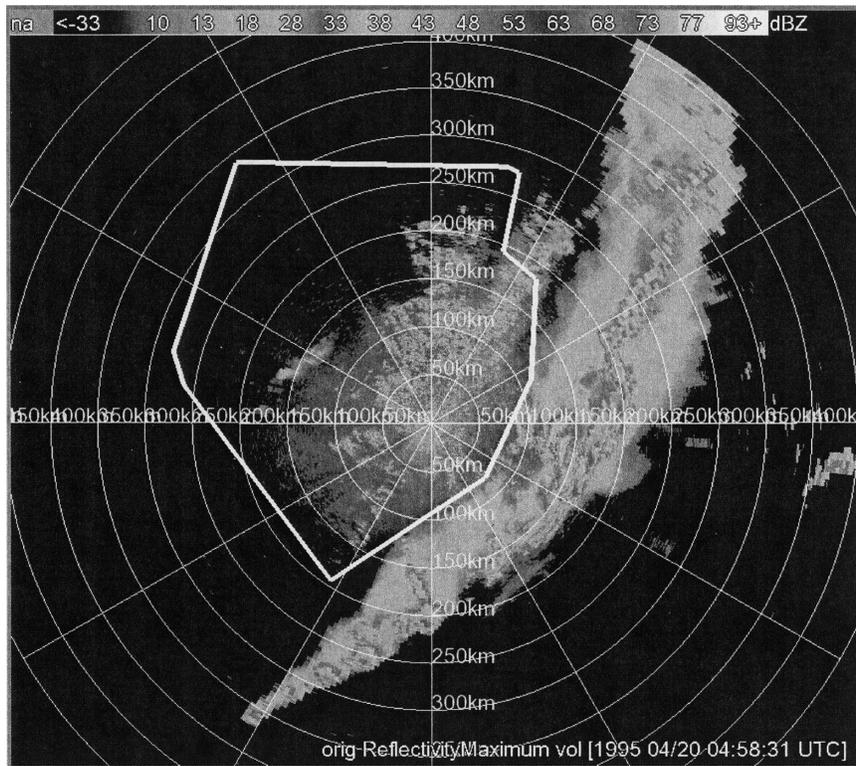


FIG. 2. An expert manually established the truth of the training and validation volume scans by drawing polygons around echoes that he believed did not correspond to precipitation.

TABLE 2. The volume scans used as a validation set to test the generalization of an NN.

Radar	Location	Volume-scan start time	Reason chosen
KAMA	Amarillo, TX	1053 UTC 24 May 1994	AP
KAMA	Amarillo, TX	2042:49 UTC 29 Apr 2004	Weak convection, interference
KAMA	Amarillo, TX	1408:11 UTC 29 Aug 2005	AP far away from the radar
KDDC	Dodge City, KS	0151:51 UTC 21 Apr 2005	Convection, interference
KDLH	Duluth, MN	1956:45 UTC 1 Jan 2005	Snow
KFSD	Sioux Falls, SD	2035:28 UTC 28 Jul 2005	Biological
KHDX	Holloman, NM	1029:36 UTC 28 May 2005	Interference
KINX	Tulsa, OK	1831:06 UTC 29 Apr 2004	Weak convection, sun strobes
KLNX	North Platte, NE	0253:23 UTC 15 Feb 2005	Snow
KSGF	Springfield, MO	1857:09 UTC 17 Aug 2005	Weak convection, biological
KTLX	Oklahoma City, OK	0806:51 UTC 11 Jun 2003	Weak convection
KTLX	Oklahoma City, OK	1124:43 UTC 11 Jun 2003	Sun strobe
KTLX	Oklahoma City, OK	0918:27 UTC 13 Jun 2003	Biological, convection

to each pattern. We are, in essence, assigning a higher cost to misclassifying high-dBZ range gates than to misclassifying low-dBZ range gates. The histograms in Figs. 1c and 1d show the effect of this selective emphasis—note that the histograms in Fig. 1c are shifted to the right (higher reflectivity values) and have approximately equal probabilities at all reflectivity values (balanced distribution).

Some range gates can be classified very easily because they are obvious. To aid processing efficiency, both in the training stage and in the running stage, we preclassify such range gates using criteria described in section 2d. These criteria also account for shortcomings in the way that the input features used for classification are computed. Such range gates are not presented to the NN in training, and range gates that match these criteria are preclassified the same way in run time as well. The preclassification process is described in more detail in section 2d.

The training process is an automated nonlinear optimization during which we need to find weights that yield a good discrimination function on the training data but are not so overfit to the training data that they do poorly on independent data.

A validation set can ensure a network's generalization, typically through the use of early stopping methods (Bishop 1995). In the NN literature, a validation set is also utilized to select the architecture of the NN (Masters 1993). We used a validation set that consisted of features derived from several volume scans that exhibited AP, convection, clear-air return, stratiform rain, and radar test patterns (see Table 2).

During each training run, the RPROP algorithm was utilized to optimize the weights on the training data cases. At each iteration of the optimization process, the performance of the weights on the validation set was checked. Even though the training entropy error may

continually decrease, the validation entropy error typically does not (see Fig. 3a). We select the weights corresponding to the stage at which the validation entropy error is minimum; thus the use of a quasi-independent validation set helps to prevent the weights being overfit to the training set. We trained each network with different numbers of hidden nodes and selected the number of nodes for which the validation entropy error is minimum, as shown in Fig. 3b. Thus, we used the validation set both to determine when to stop within a training run and to pick the final architecture of the NN. Other than to choose the number of hidden nodes, we did not consider any alternate network topologies (such as two layers of hidden nodes) because, in theory at least, a single hidden layer is enough to interpolate any continuous function to arbitrary accuracy (Bishop 1995).

Note that the cross-entropy error criterion for the with-velocity network is less than that of the reflectivity-only network. This is a measure of how useful the velocity information is to classification accuracy.

We used a testing set, independent of the training and validation sets, as described in section 3, and it is this independent set for which the results are reported.

### c. Feature selection

In the process of training the networks, each of the computed inputs was removed and the NN was reoptimized. The probability of detection (POD) of precipitating echoes and the false-alarm rates (FAR) for both networks (the with velocity and the reflectivity only; see section 2) were noted. If removing the input had no significant negative effect on the four statistics (POD and FAR on the two networks), the input was permanently removed.

For example, using this process, it was found that retaining just the mean and variance in the local neigh-

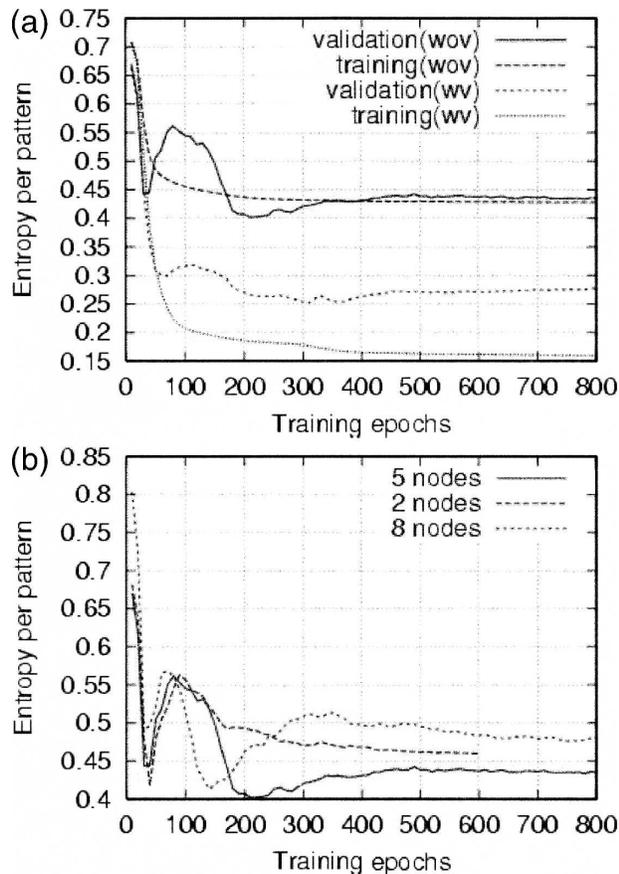


FIG. 3. Use of a validation set to decide when to stop the training and to decide on the number of hidden nodes. The y axis is  $E_c/N$  [see Eq. (2)]. (a) Training error vs validation error. Note that the training error continues to decrease but the validation error starts to increase after a while, showing that the training is becoming counterproductive; “wv” refers to the NN with velocity data, and “wov” refers to the NN without velocity data (see section 2 for the description of why there are two networks). (b) Validation error curves for different numbers of hidden nodes. The number of hidden nodes at which the validation error was minimized was chosen as the network architecture.

borhood was enough—use of the median did not improve the capability of the NN to learn the data, as measured by the cross entropy. We discovered that the use of some features hurt trainability, especially when the features in question were highly effective in a univariate sense. This was probably because the NN got caught in local minima. In such cases, we reformulated the feature to avoid this problem—the echo size, echo top, and vertical difference were defined so that their use does not lead to poor performance.

We discovered that computing the echo top with a threshold of either 0 or 10 dBZ, depending on whether the ground temperature at the radar site was below or above 10°C, respectively, greatly improved the perfor-

mance. We also found that computing the echo top only for elevations above the lowest elevation improves the reliability of this measure. Thus, the echo-top parameter referred to in this paper is the maximum physical height to the top of an echo when it reaches either 0 or 10 dBZ in elevation scans other than the lowest scan of elevation of reflectivity.

The vertical difference between the two lowest elevation scans leads to poorly performing NNs, probably because it is such a good univariate classifier. We defined the vertical difference as the difference between the lowest elevation scan and the first elevation scan that crosses 3 km at this range; thus at far ranges the vertical difference is always zero. This feature is used only close in to the radar where the difference is between the lowest scan and a scan at a reliable height.

The local neighborhood for all the statistics, including the SPIN, was chosen to be  $5 \times 5$ , after we experimented with  $3 \times 3$  and  $7 \times 7$  sizes. Choosing the neighborhood size for each statistic individually might yield better results, but the additional complexity may not be worth the reward. In the case of SPIN, for example, the  $11 \times 21$  neighborhood size used by Steiner and Smith (2002) yielded a slightly better performance, but the computation then takes nearly 8 times as long.

Following this process of pruning the input features, we were left with the following 28 input features at every gate—from the lowest velocity scan: 1) value, 2) local mean, 3) local variance, 4) difference, and 5) minimum variance; from the lowest spectrum width scan: 6) value; from the lowest reflectivity scan: 7) local mean, 8) local variance, 9) difference, 10) variance along radials, 11) difference along radials, 12) local minimum variance, 13) SPIN, and 14) inflections along radial (Kessinger et al. 2003); from the second-lowest reflectivity scan: 15) local mean, 16) local variance, 17) local difference, 18) variance along radials, 19) difference along radials, and 20) minimum variance; within a virtual volume scan: 21) vertical maximum, 22) weighted average, 23) vertical difference, 24) echo top, 25) height of maximum, 26) echo size in vertical composite, 27) inbound distance along radial to 3-km echo top, and 28) outbound distance along radial to gate with zero velocity. The “difference” statistic is simply the difference between the value at that gate and the mean within the neighborhood and is used to decorrelate the pixel value from the local mean.

The reflectivity-only NN, used where there are no velocity data or where the velocity data are range folded, had only 22 input features because the first six features (corresponding to velocity and spectrum width) are unavailable. If removing a reflectivity-based

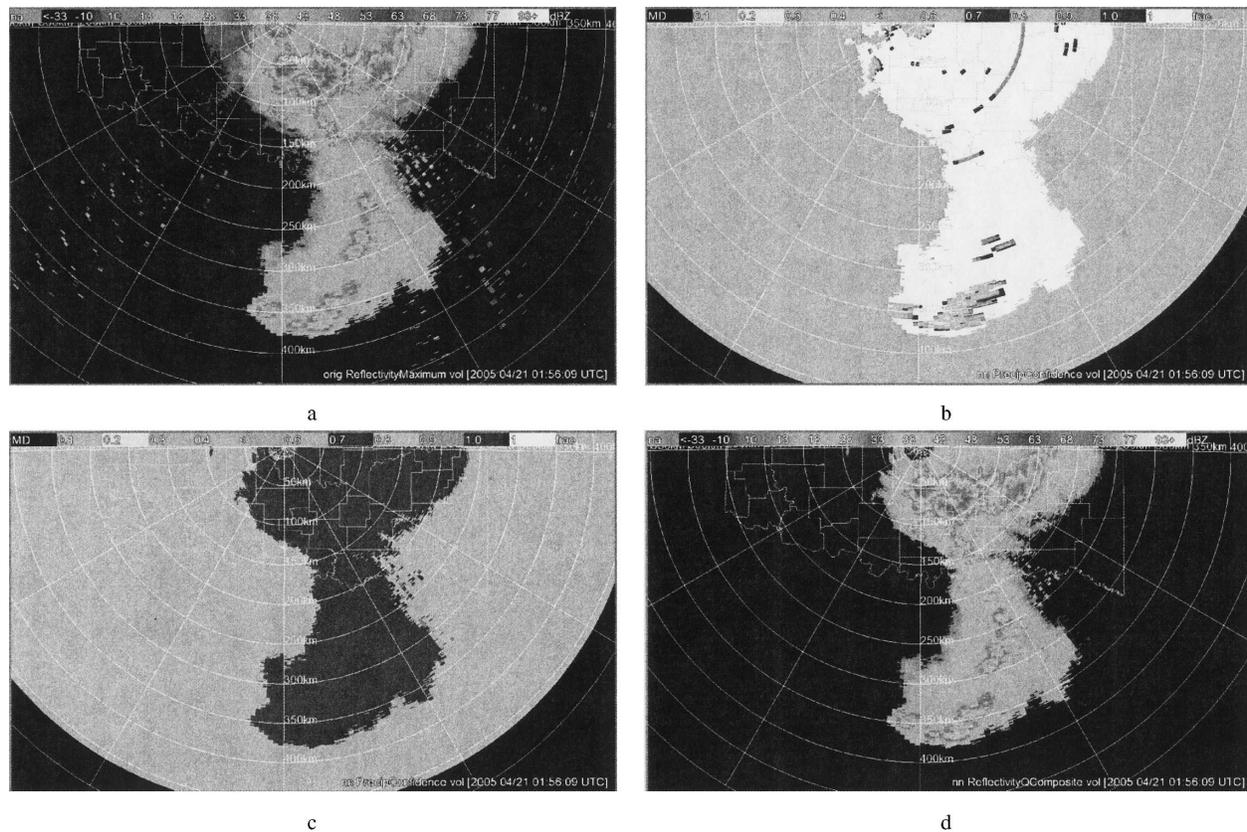


FIG. 4. The echo-size parameter, a measure of how much of a pixel's neighborhood is filled (see section 2a), is used during preprocessing to avoid presenting the NN with statistics computed in poorly filled areas. Postprocessing the NN confidence field based on spatial analysis reduces random errors. (a) Original image, (b) precipitation confidence before postprocessing, (c) precipitation confidence after postprocessing, and (d) final QCed product.

feature affected either the with-velocity network or the reflectivity-only network, that feature was retained.

#### d. Preprocessing and postprocessing

The NN simply provides a function that maps the input features to a number in the range [0, 1]. The quality of the output is only as good as the input features. In general, a more reliable input space yields a better-performing NN.

Many of our input features are computed in local neighborhoods around a particular gate. Such features exhibit unwanted characteristics near the edges of storms. Therefore, the echo-size parameter is used to preclassify such edge range gates. Only gates for which the echo size is greater than 0.9 are presented to the NN. Gates with echo sizes less than 0.5 are classified as nonprecipitation—this is akin to median filtering the input data. Gates with echo sizes between 0.5 and 0.9 simply pick up whatever the majority of their neighbors are classified as. The effect of this median filtering and assigning a “do not care” status is shown in Figs. 4a and 4b.

In radar reflectivity data, range gates are set to be invalid in two situations: 1) the range gate is not sensed by the radar at all, such as if the gate in question is out of the radar range at that elevation angle, and 2) the range gate in question was sensed, but the resulting echo's signal was not much stronger than the noise normally expected. To improve the robustness of the statistics being computed, we sought to distinguish between these two situations. We set all range gates in the reflectivity fields that could conceivably have had a radar return (those range gates with a height below 12 km) and that had a radar return below the noise threshold (and was therefore set to missing) to be zero. Thus, the only missing-data values correspond to atmospheric regions that are not sensed by the radar at all. At this stage, range gates that are expected to be beam blocked, applying the technique of O'Bannon (1997), are also set to have missing data. Values with 0 dBZ are used in computing local features, but missing data are not.

Features such as sun strobos are rare even within the context of an elevation scan and are difficult, therefore, for a gradient-descent function such as an NN to opti-

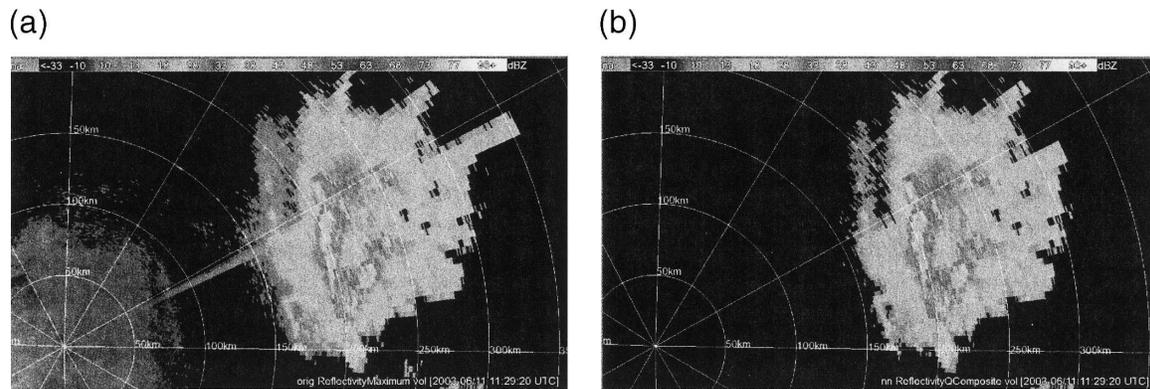


FIG. 5. A heuristic method is used to identify radials that are completely filled and monotonically increasing. Such radials are replaced by a linear interpolation of their (good) neighbors. (a) Original data, and (b) QCed data.

mize. A heuristic method consequently is employed to detect and remove such radials. We look for radials in which more than 90% of the gates are filled with reflectivity values greater than 0 dBZ and whose values are linearly increasing (with a correlation coefficient greater than 0.8). If less than four radials in the elevation scan are mostly filled, then those radials with a 90% fill and 0.8 correlation coefficient are removed from the data and are replaced by interpolating neighboring radials, as shown in Fig. 5. A future study could examine whether the use of sunrise and sunset times and the position of the sun relative to the radar position can be used to form an estimate of regions where sun strokes are likely.

The echo-top parameter is an extremely accurate univariate classifier once small echoes (echo size less than 0.5) and radial artifacts are removed from the presented input vectors. Thus, if a range gate has an echo top greater than 3 km, it is preclassified as being precipitation. As a result of preclassification, the range gates presented to the NN are the “hard” cases—those range gates with an echo top less than 3 km, with radials that look reasonable and are within an area of echo.

Although the NN computes the posterior probability that, given the input vector, the range gate corresponds to precipitating echoes, adjacent range gates are not truly independent. In fact, the training process, using polygons to delineate good and bad areas, underscores the fact that the QC needs to be performed on a spatial, rather than on a range gate, basis.

A simpler form of the texture segmentation introduced in Lakshmanan et al. (2003) is performed on the vertical maximum field. The simplification is that only a single scale of segmentation, rather than the multiscale segmentation described in that paper, is desired. Then, the mean of the NN output within each cluster is determined, and if the mean is below a certain threshold

(arbitrarily chosen to be 0.5) then the entire cluster is removed as not corresponding to precipitation.

Figure 4c demonstrates how postprocessing the resulting field has removed random<sup>1</sup> errors within the clusters to yield an almost completely correct classification.

### 3. Results and discussion

A diverse set of 16 volume scans (independent of those chosen for training and validation, shown in Table 3) were chosen, and bad echoes were marked on these volume scans by a human observer. An automated routine then used these polygons to create a “target” set of elevation scans. These are the products that an ideal QC algorithm would produce. The same volume scans were QCed using the NN introduced here and using the Build-7 version of the radar echo classifier implemented in the “open radar products generator” (Jain et al. 1998) of the NEXRAD system.

Two sets of comparisons were performed. From the cleaned-up elevation scans, “reflectivity composite” (vertical maximum) and “vertical integrated liquid” (VIL; Greene and Clark 1972) products were created. These products were then compared range gate by range gate with the same products created from the target elevation scans. The comparisons were formulated as follows: the probability of detection is perfect (1.0) if all precipitating echoes are retained in the final products. The false-alarm ratio is perfect (0.0) if none of the contaminants remain in the final products. The

<sup>1</sup> Statisticians distinguish between random errors and systemic (or systematic) errors. Systemic errors are errors resulting from an identifiable cause and can often be resolved. On the other hand, random errors are a result of the underlying probability distributions, such as the overlap between them, and cannot be avoided.

TABLE 3. The volume scans used to evaluate the WSR-88D QC neural technique described in this paper.

Radar	Location	Volume-scan start time	Reason chosen
KLBB	Lubbock, TX	0144 UTC 5 Oct 1995	AP
KTLX	Oklahoma City, OK	1416 UTC 16 Jun 1996	AP, rain
KUEX	Hastings, NE	0231 UTC 13 Jun 2002	Strong convection
KICT	Wichita, KS	2032 UTC 19 Apr 2003	Weak convection
KDVN	Davenport, IA	0436 UTC 1 May 2003	Weak convection
KAMA	Amarillo, TX	2150 UTC 3 May 2003	Biological
KTLX	Oklahoma City, OK	2231 UTC 30 Apr 2004	Strong convection
KFDR	Fredrick, OK	0259 UTC 16 Jul 2004	Bats
KINX	Tulsa, OK	1121 UTC 17 Aug 2004	Speckle
KCYS	Cheyenne, WY	0057 UTC 21 Sep 2004	Sun ray
KFFC	Atlanta, GA	1615 UTC 27 Sep 2004	Stratiform rain
KICT	Wichita, KS	0801 UTC 26 Oct 2004	Stratiform rain
KILX	Lincoln, IL	2356 UTC 26 Oct 2004	Stratiform rain
KHDX	Holloman, NE	1034 UTC 28 May 2005	Hardware fault
KDGX	Jackson, MS	2215 UTC 7 Jun 2005	Weak convection

critical success index (CSI) is a combination of these two measures (Donaldson et al. 1975) and is formulated similarly, that is, as a measure of the algorithm's skill in detecting and retaining precipitation echoes. The Heidke skill score (HSS) compares the performance of the algorithm with a theoretical algorithm that operates purely on chance (Heidke 1926). Because the HSS requires the number of nonprecipitating range gates that were correctly identified, this number was taken as the number of range gates from the radar that had an echo in the range  $(-\infty, 0)$  dBZ in both the original and the quality-controlled reflectivity composite fields (or  $VIL = 0$  in the case of the VIL fields).

Using the reflectivity composite as a verification mechanism ensures that we are assessing the algorithm across all reflectivity values, regardless of the height at which they occur. VIL, on the other hand, serves as a measure of the presence of significant echo. Because a good QC technique should not remove good data, using VIL as a verification measure is a way to assess the extent to which good data are retained.

The statistics were computed on the test dataset with one of the test cases removed at each time. This "leave one out" procedure, also termed jackknifing, can be used to estimate the standard error of a statistic (Efron and Tibshirani 1997). What is reported in Table 4 is the mean skill score achieved and the 95% confidence interval assuming a normal distribution of skill scores. A more complete statement of the results, including the individual cases, is available in Fritz et al. (2006).

#### a. Assessing performance of a QC algorithm

In this paper, the performance of the QC technique was assessed by a comparison with the skill of a human expert. Thus, the numbers in Table 4 reflect the per-

formance of the algorithm assuming that the human expert is perfect. To reduce expert bias, a different human expert than the one who provided the truth for the training and validation cases analyzed the test cases.

Other techniques of assessment have been followed in the literature. Kessinger et al. (2003) validated their technique by comparing the results of their QC technique running on polarimetric radar data with the results of a hydrometeor classification algorithm on the assumption that the hydrometeor classifier is perfect. This was reasonable because their study focused on AP/GC mitigation and polarimetric algorithms do very well in identifying AP/GC. Our goal in this research was to achieve a QC algorithm that worked on a variety of real-world observed effects, including hardware faults and interference patterns. It is unclear what a hydrometeor classification algorithm would provide in that case. Because there are few polarimetric radars in the United States at present, such an approach would also have limited the climatological diversity of test cases.

Another way that QC algorithms have been validated is by comparing the result of a precipitation estimate from the radar fields with observed rain rates (Robinson et al. 2001; Krajewski and Vignal 2001). One advantage of this approach is that it permits large volume studies—for example, Krajewski and Vignal (2001) used 10 000 volume scans, albeit all from the same radar and all in warm weather. The problem with such a verification method is that the skill of the QC technique has to be inferred by whether precipitation estimates on the QCed data are less biased (as compared with a rain gauge) than precipitation estimates carried out on the original data. Such an inference is problematic—reduction in bias can be achieved with unskilled operations, not just by removing true con-

TABLE 4. The QC NN described in this paper was compared with the REC, the operational algorithm implemented on the NEXRAD system. Reflectivity composite and polar VIL products were computed from the QCed reflectivity elevation scans and were compared range gate by range gate. The quoted skill scores are listed to two significant digits with a 95% confidence interval. The confidence interval was estimated by jackknifing the data cases.

Product	Data range	Measure	No QC	REC	QC NN
Composite	>0 dBZ	CSI	$0.61 \pm 0.06$	$0.59 \pm 0.057$	$0.86 \pm 0.011$
		FAR	$0.39 \pm 0.06$	$0.4 \pm 0.06$	$0.02 \pm 0.0072$
		POD	$1 \pm 0$	$0.96 \pm 0.0031$	$0.88 \pm 0.0088$
		HSS	$0.89 \pm 0.02$	$0.88 \pm 0.019$	$0.98 \pm 0.0016$
Composite	>10 dBZ	CSI	$0.68 \pm 0.071$	$0.66 \pm 0.069$	$0.96 \pm 0.0083$
		FAR	$0.32 \pm 0.071$	$0.32 \pm 0.073$	$0.02 \pm 0.007$
		POD	$1 \pm 0$	$0.94 \pm 0.0023$	$0.92 \pm 0.0039$
		HSS	$0.93 \pm 0.017$	$0.93 \pm 0.016$	$0.99 \pm 0.0011$
Composite	>30 dBZ	CSI	$0.92 \pm 0.02$	$0.84 \pm 0.014$	$1 \pm 0.000 72$
		FAR	$0.08 \pm 0.02$	$0.09 \pm 0.011$	$0 \pm 0.000 57$
		POD	$1 \pm 0$	$0.92 \pm 0.0065$	$1 \pm 0.000 29$
		HSS	$1 \pm 0.000 64$	$0.99 \pm 0.000 52$	$1 \pm 0$
Composite	>40 dBZ	CSI	$0.91 \pm 0.023$	$0.8 \pm 0.013$	$1 \pm 0.000 38$
		FAR	$0.09 \pm 0.023$	$0.1 \pm 0.0074$	$0 \pm 0.000 39$
		POD	$1 \pm 0$	$0.88 \pm 0.0088$	$1 \pm 0$
		HSS	$1 \pm 0.000 16$	$1 \pm 0.000 18$	$1 \pm 0$
VIL	>0 kg m <sup>-2</sup>	CSI	$0.53 \pm 0.16$	$0.48 \pm 0.13$	$1 \pm 0.0011$
		FAR	$0.47 \pm 0.16$	$0.49 \pm 0.15$	$0 \pm 0.000 53$
		POD	$1 \pm 0$	$0.9 \pm 0.0078$	$1 \pm 0.000 84$
		HSS	$0.97 \pm 0.0091$	$0.97 \pm 0.0085$	$1 \pm 0$
VIL	>25 kg m <sup>-2</sup>	CSI	$1 \pm 0.0022$	$0.65 \pm 0.033$	$0.99 \pm 0.0027$
		FAR	$0 \pm 0.0022$	$0.19 \pm 0.025$	$0 \pm 0.0022$
		POD	$1 \pm 0$	$0.76 \pm 0.026$	$1 \pm 0.000 75$
		HSS	$1 \pm 0$	$1 \pm 0$	$1 \pm 0$

tamination. This is because the mapping of reflectivity  $Z$  to rainfall  $R$  is fungible. Thus, if the chosen  $Z$ - $R$  relationship leads to overestimates of precipitation, a QC algorithm that removes good data along with AP/GC will lead to lower bias. In a similar way, if the chosen  $Z$ - $R$  relationship leads to underestimates of precipitation, a QC algorithm that does not remove all the contaminants will lead to a lower bias. If the  $Z$ - $R$  relationship is estimated directly from the data, the very process of fitting can lead to incorrect inferences because a contaminated field may end up being fit better than a true one.

Thus, assessing QC algorithms by comparing them with a human expert provides a comparison that is not subject to biases in other sensors or algorithms. The drawback is that automated large-scale studies of the kind performed by Krajewski and Vignal (2001) are not possible because of the reliance on a human expert to provide "truth." Because establishing the truth of data by hand takes inordinate amounts of time, the number of data cases for which truth is defined has to be small for such a direct evaluation method.

The size of our training, validation, and testing sets—about 12 volume scans each—may seem insignificantly small. When we started the study, we imagined that we would have to build an enormous database of artifacts

and good cases from radars all over the country over many years to build a robust algorithm. Instead, we found that an algorithm trained on only a few selected volume scans could do surprisingly well on data from anywhere in the United States. This result was because, while we needed to ensure, for example, that we had a stratiform rain case in our training, it did not matter from where in the country the case came. Thus, we took care to select the limited volume scans carefully, keeping in mind the underlying diversity of the data. We have verified that the algorithm does work equally well on radars across the country by monitoring a 1-km-resolution multiradar mosaic of QCed radar data (Lakshmanan et al. 2006) from 130+ radars across the United States. This multiradar mosaic, which we have been monitoring since April of 2005, was available on the Internet at the time of writing (<http://wdssii.nssl.noaa.gov/>). The experience gained in thus selecting a representative training set came in handy in when choosing a diverse testing set as well and explains the relatively small size of our testing set.

#### b. Comparison with the operational algorithm

The measures of skill on the reflectivity composite product serve as a proxy for visual quality, and the

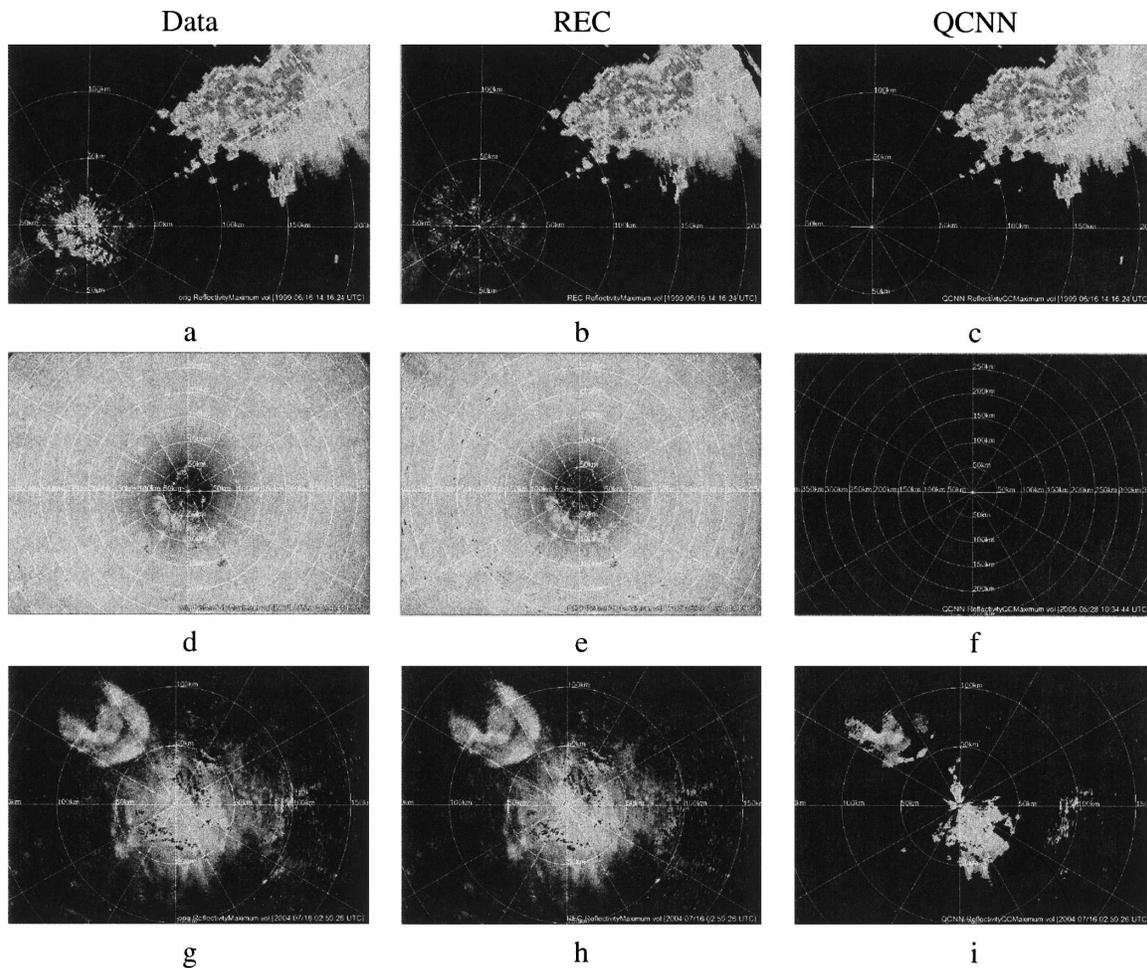


FIG. 6. The performance of the QC NN and the REC demonstrated on three of the independent test cases. (a) Reflectivity composite from the KTLX radar at 1416 UTC 16 Jun 1999 showing both ground clutter and good echoes. The range rings are 50 km apart. (b) Result of QC by the REC. (c) Result of QC by the NN. All of the AP has been removed and the good echo has not been touched. (d) Reflectivity composite from KHDX at 1034 UTC 28 May 2005 showing a hardware fault. (e) Result of QC by the REC. (f) Result of QC by the NN. All of the data from this problematic volume scan have been wiped out. (g) Reflectivity composite from KFDR at 0259 UTC 16 Jul 2004 showing biological contaminants. (h) Result of QC by the REC. (i) Result of QC by the NN—the algorithm performance is poor in this case.

measures of skill on the VIL product serve as a proxy for the effect on severe weather algorithms.

Range gates with VIL greater than  $25 \text{ kg m}^{-2}$  represent areas of significant convection. It is curious that when VIL is directly computed from the reflectivity data, without an intermediate QC step, the FAR is near zero for values greater than  $25 \text{ kg m}^{-2}$ . This is because the VIL product is based on a weighted integration in which echoes aloft receive large weights. The VIL is not affected so much by shallow, biological contamination or by AP/GC close to the ground. In other words, VIL values greater than  $25 \text{ kg m}^{-2}$  are obtained in regions for which there are high reflectivity values in several elevation scans—a condition not seen in bad-echo regions.

The effect of assigning higher costs [ $c_n$  in Eq. (2)] to higher reflectivity values during the training process is clear in the final results; the QC NN makes most of its mistakes at low values of reflectivity and is very good at reflectivity values above 30 dBZ.

As can be readily seen, both from Table 4 and from the examples in Fig. 6, the NN outperforms the fuzzy-logic automated technique of Kessinger et al. (2003), which is one of a number of algorithms that perform similarly (Robinson et al. 2001).

Note that the operational REC algorithm, like most earlier attempts at automated QC of radar reflectivity data, was designed only to remove AP and GC. Other forms of the REC (used to detect insects and precipitation) have been devised but have not been imple-

mented operationally (C. Kessinger 2006, personal communication). It is unclear how these individual algorithms would be combined to yield a single all-purpose QC algorithm. When QCed data are used as inputs to algorithms, such as precipitation estimation from radar, all forms of contamination need to be removed, not just AP. Hence, our goal was to test against all forms of contamination. The REC columns in Table 4 should not be used as an assessment of the skill of the algorithm itself (such an assessment would concentrate only on AP cases). Instead, the skill scores reported for REC should be used as an indication of the results that will result if the REC AP detection algorithm is used as the sole form of QC in a real-world situation in which all kinds of contaminants are present. In other words, the operational algorithm may or may not be a poor performer (our test set did not include enough AP cases to say one way or the other); it is just that an AP-only approach is insufficient as a preprocessor to weather-radar algorithms.

#### c. Effect of postprocessing

The precipitation confidence field that results from the NN is subject to random “point” errors. So, the field is postprocessed as described in section 2d and demonstrated in Fig. 4. Such spatial contiguity tests and object identification serve to remove random error, yielding QCed fields for which good data are not removed at random range gates. However, such spatial processing could also have side effects, especially where there are regions of bad data adjacent to a region of good echo. Because of their spatial contiguity, the two echo regions (good and bad) are combined into a single region and the output of the NN is averaged over all the range gates in that region. This procedure typically results in the smaller of the two regions being misclassified. We believe that such cases are relatively rare. One such case from 1993 (shown in Fig. 7) demonstrates a situation in which there is AP embedded inside a larger precipitation echo. The AP range gates are correctly identified by the NN, but they are misclassified at the spatial postprocessing stage. A multiscale vector segmentation, as in Lakshmanan et al. (2003), based on both the reflectivity field and the identified precipitation confidence field, could help to resolve such situations satisfactorily. We plan to implement such a segmentation algorithm.

Figure 7 also illustrates another shortcoming of the technique described in this paper. If there is AP at lower elevation angles and precipitation echoes at higher elevation angles over the same area, the desirable result would be to remove the AP at lower elevation scans but to retain the precipitation echoes aloft.

However, because the technique described in this paper uses all of the elevations within a virtual volume scan to classify range gates on a two-dimensional surface, the resulting classification is independent of the elevation angle. Thus, all of the range gates would either be removed (because of the AP below) or retained (because of the precipitation aloft).

#### d. Multisensor QC

The radar-only QC technique described above, and whose results are shown in Table 4 and Fig. 6, performs reasonably well and is better than earlier attempts described in the literature.

We have been monitoring a 1-km-resolution multi-radar mosaic of QCed radar data (Lakshmanan et al. 2006) from 130+ radars across the United States for nearly a year as of this writing. This continuous monitoring led to the identification of some systemic errors in the radar-only technique. Many of these errors occur at night in areas subject to “radar bloom” (biological contamination in warm weather, as shown in Fig. 6g) and when radars experience hardware faults or interference from neighboring radars that cause unanticipated patterns in the data. When considering data from just one radar at a time, the incidence of such errors may be acceptable. However, when merging data from multiple radars in real time from 130 WSR-88Ds, such as in Lakshmanan et al. (2006), the errors are additive—at any point in time, one of these systemic errors is 130 times as likely. Therefore, the quality of radar data feeding into multiradar products has to be much better.

One potential way to improve the quality of the reflectivity data is to use information from other sensors (Pamment and Conway 1998). In real time, we can compute the difference

$$T = T_{\text{surface}} - T_{\text{cloudtop}}, \quad (3)$$

where the cloud-top temperature  $T_{\text{cloudtop}}$  is obtained from the 11- $\mu\text{m}$  Geostationary Operational Environmental Satellite infrared channel and the surface temperature  $T_{\text{surface}}$  is obtained from an objective analysis of surface observations. In the ideal case, this difference should be near zero if the satellite is sensing the ground and should be greater than zero if the satellite is sensing clouds—in particular, high clouds. In practice, we use a threshold of 10°C. Unlike the visible channel, the IR channel is available throughout the day but has a poor temporal resolution and high latency (i.e., it arrives after significant delays) in real time. Therefore, the satellite data are advected using the clustering and Kalman filtering technique described in Lakshmanan et al.

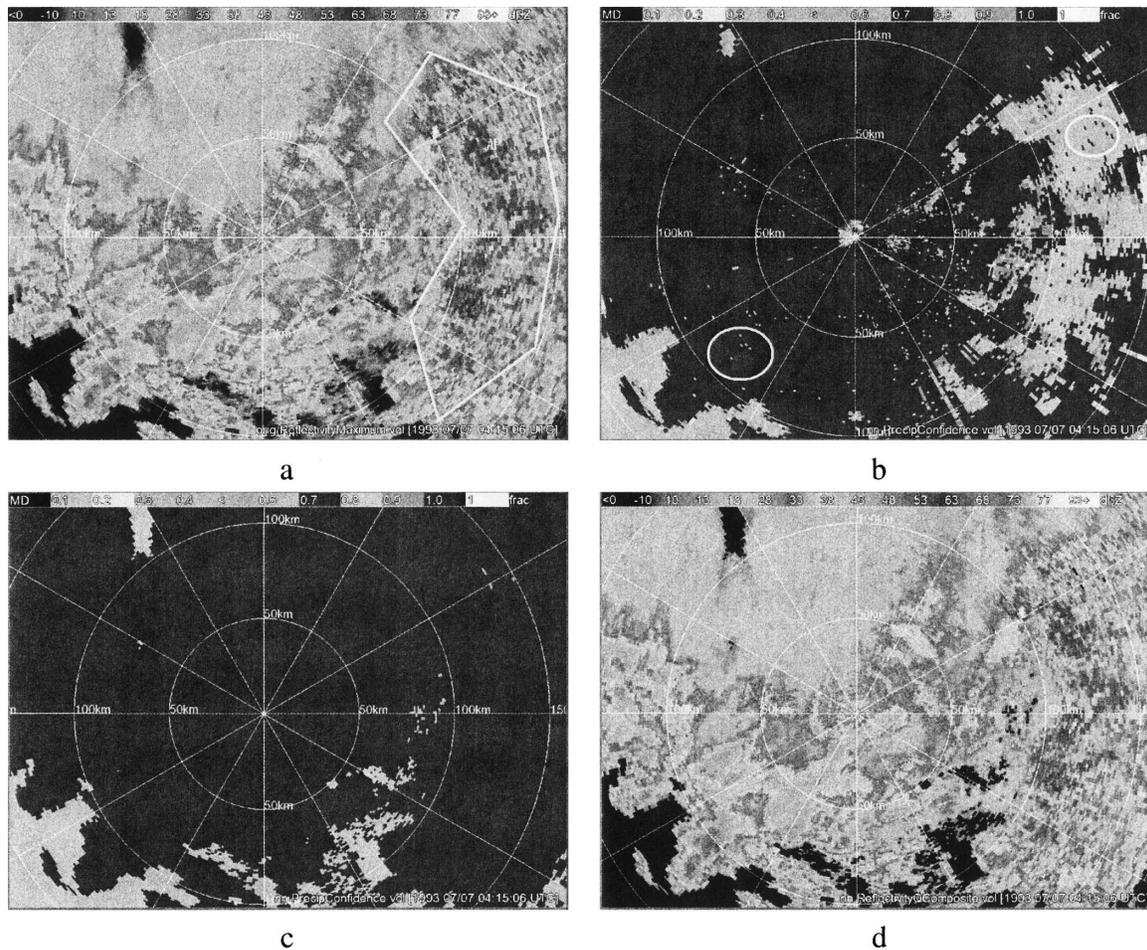


FIG. 7. An instance in which spatial postprocessing is detrimental. (a) A radar composite from KLSX at 0409 UTC 7 Jul 1993 showing anomalous propagation (marked by a polygon) embedded inside precipitation echoes. (b) The precipitation confidence field from the NN showing that the AP and precipitation areas have been correctly identified for the most part. The dark areas are good echo. Note, however, the random errors in classification in both the AP and precipitation areas (circled). (c) The result of spatial postprocessing; because the precipitation area is much larger than the AP area, the AP area is also identified as precipitation. Use of a vector segmentation algorithm may help to resolve situations like this one. (d) The resulting QCed composite field has misidentified the AP region as precipitation.

(2003) to match the current time before the difference is computed.

If there is any overlap between a region of echo sensed by radar and clouds sensed through the temperature difference, that entire region of echo is retained. If there is no overlap, the entire region is deleted, as shown in Fig. 8.

Although in this paper we have shown this multisensor QC operating on radar data combined from multiple radars, the multisensor QC technique can be applied to single radar data also (Lakshmanan and Valente 2004).

The trade-offs of using such a multisensor QC technique are not yet completely known. Although sporadic errors left by the radar-only technique are corrected by

the multisensor component, the multisensor QC technique of the form discussed leaves much to be desired. Satellite IR data are of poor resolution (approximately 4 km every 30 min); the surface temperature data we use are even coarser. Many shallow storms and precipitation areas, for example, lake-effect snow, are completely missed by the satellite data. The advection may be slightly wrong, leaving smaller cells unassociated with the clouds on satellite. Thus, although the multisensor aspect can provide an assurance that clear-air returns are removed, it can also remove many valid radar echoes. An examination of its performance over a climatologically diverse dataset is required.

An implementation of this QC technique for WSR-88D data was available (free for research and academic

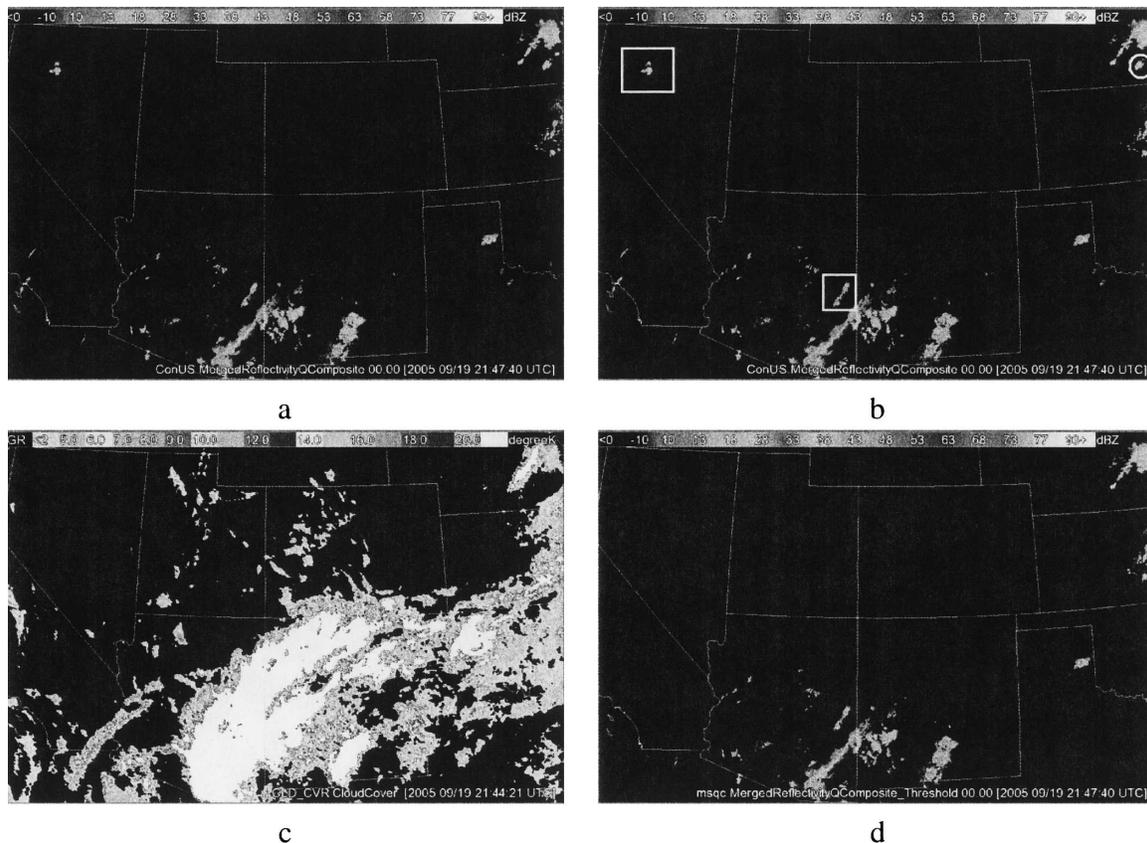


FIG. 8. (a) Part of a radar image formed by combining QCed radar data from all of the WSR-88D radars in the continental United States at approximately  $1 \text{ km} \times 1 \text{ km}$  every 2 min. The detail shows Nevada (top-left of image) to the Texas Panhandle (bottom-right of image). (b) Two good echoes are marked by rectangles (in Nevada and Arizona). One bad-echo region is marked by a circle (in Nebraska, top-right). (c) Cloud-cover product formed by taking a difference of cloud-top temperature detected by satellite and surface temperature from an objective analysis of surface observations. Notice that the echo in Arizona has cloud cover; the marked echoes in Nevada and Nebraska do not. (d) Applying the cloud-cover field helps to remove some bad echo in Nebraska while retaining most of the good echo. However, good echo in Nevada has been removed because it was not correlated with cold cloud tops in satellite IR data.

use) on the Internet at the time of writing (<http://www.wdssii.org/>).

#### e. Research directions

We have focused on developing a single algorithm that could be used to perform QC on WSR-88D data from anywhere in the continental United States and in all seasons. It is possible that a more targeted QC application (e.g., targeted to the warm season in Florida) might work better because it would limit the search space for the optimization procedure. On the other hand, there should be no reason to target a QC application by the type of bad echoes expected. There is little that the end user of two QC products—one for AP/GC and the other for radar test patterns—can do other than to use a rule of thumb to combine these into a single product. A formal optimization procedure

should, in theory, do better at arriving at such a combined product.

It was pointed out in this paper that when combining radar data from multiple radars the errors are additive. Therefore, the QC procedure, although adequate for single-radar products, needs to be further improved for multiradar products. The use of satellite and surface data in real time was explored, but some limitations were observed. A multisensor approach merits further study.

This research utilized only a selected set of training, validation, and independent test cases. In such a selection, the climatological probabilities of the various errors are lost. A climatological study of all the WSR-88D radars in the continental United States, with the actual probability of the various problems that a QC technique is expected to handle, may permit a better design of QC techniques. We believe that our technique of

selected emphasis to balance out distributions and to emphasize strong reflectivity echoes is a better approach, but this assertion is unproven.

#### 4. Summary

This paper has presented a formal mechanism for choosing and combining various measures, both proposed in the literature and introduced in this paper, to discriminate between precipitation and nonprecipitating echoes in radar reflectivity data. This classification is done on a region-by-region basis rather than range gate by range gate. It was demonstrated that the resulting radar-only quality-control mechanism outperforms the operational algorithm on a set of independent test cases. The approach of this paper would be just as applicable to data from a newer generation of radars or to data from radars in different areas of the world.

*Acknowledgments.* Funding for this research was provided under NOAA-OU Cooperative Agreement NA17RJ1227, FAA Phased Array Research MOU, and National Science Foundation Grants 9982299 and 0205628. Angela Fritz was supported by the NSF's Research Experiences For Undergraduates (REU) Program 0097651.

#### REFERENCES

- Bishop, C., 1995: *Neural Networks for Pattern Recognition*. Oxford University Press, 482 pp.
- Cornelius, R., R. Gagon, and F. Pratte, 1995: Optimization of WSR-88D clutter processing and AP clutter mitigation. Forecast Systems Laboratory Tech. Rep., 182 pp.
- Donaldson, R., R. Dyer, and M. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. Preprints, *Ninth Conf. on Severe Local Storms*, Norman, OK, Amer. Meteor. Soc., 321–326.
- Efron, B., and R. Tibshirani, 1997: Improvements on cross-validation: The 632+ bootstrap method. *J. Amer. Stat. Assoc.*, **92**, 548–560.
- Fritz, A., V. Lakshmanan, T. Smith, E. Forren, and B. Clarke, 2006: A validation of radar reflectivity quality control methods. Preprints, *22d Int. Conf. on IIPS for Meteorology, Oceanography and Hydrology*, Atlanta, GA, Amer. Meteor. Soc., CD-ROM, 9.10.
- Fulton, R., D. Breidenback, D. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377–395.
- Greco, M., and W. Krajewski, 2000: An efficient methodology for detection of anomalous propagation echoes in radar reflectivity data using neural networks. *J. Atmos. Oceanic Technol.*, **17**, 121–129.
- Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water—A new analysis tool. *Mon. Wea. Rev.*, **100**, 548–552.
- Heidke, P., 1926: Berechnung des erfolges und der gute der windstarkvorhersagen im sturmwarnungsdienst (Calculation of the success and the goodness of strong wind forecasts in the storm warning service). *Geogr. Ann.*, **8**, 301–349.
- Hondl, K., 2002: Current and planned activities for the warning decision support system-integrated information (WDSS-II). Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 146–148.
- Jain, M., and Coauthors, 1998: Software development of the NEXRAD open systems radar products generator (ORPG). Preprints, *14th Int. Conf. on IIPS for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., 563–566.
- Kessinger, C., S. Ellis, and J. Van Andel, 2003: The radar echo classifier: A fuzzy logic algorithm for the WSR-88D. Preprints, *Third Conf. on Artificial Applications to the Environmental Sciences*, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, P1.6.
- Krajewski, W., and B. Vignal, 2001: Evaluation of anomalous propagation echo detection in WSR-88D data: A large sample case study. *J. Atmos. Oceanic Technol.*, **18**, 807–814.
- Krogh, A., and J. Hertz, 1992: A simple weight decay can improve generalization. *Advances in Neural Information Processing Systems*, J. E. Moody, S. J. Hanson, and R. P. Lippman, Eds., Vol. 4, Morgan Kaufmann, 950–957.
- Lakshmanan, V., 2001: A hierarchical, multiscale texture segmentation algorithm for real-world scenes. Ph.D. thesis, University of Oklahoma, 115 pp.
- , and M. Valente, 2004: Quality control of radar reflectivity data using satellite data and surface observations. Preprints, *20th Int. Conf. on IIPS for Meteorology, Oceanography, and Hydrology*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, 12.2.
- , R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atmos. Res.*, **67**, 367–380.
- , T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time, three dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity, velocity and derived products. *Wea. Forecasting*, **21**, 802–823.
- Lynn, R., and V. Lakshmanan, 2002: Virtual radar volumes: Creation, algorithm access and visualization. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., CD-ROM, P5.3.
- MacKay, D. J. C., 1992: A practical Bayesian framework for backprop networks. *Advances in Neural Information Processing Systems*, J. E. Moody, S. J. Hanson, and R. P. Lippman, Eds., Vol. 4, Morgan Kaufmann, 839–846.
- Marzban, C., and G. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar-derived attributes. *J. Appl. Meteor.*, **35**, 617–626.
- Masters, T., 1993: *Practical Neural Network Recipes in C++*. Morgan Kaufmann, 493 pp.
- Mazur, R. J., V. Lakshmanan, and G. J. Stumpf, 2004: Quality control of radar data to improve mesocyclone detection. Preprints, *20th Int. Conf. on IIPS for Meteorology, Oceanography, and Hydrology*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, P1.2.
- McGrath, K., T. Jones, and J. Snow, 2002: Increasing the usefulness of a mesocyclone climatology. Preprints, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., CD-ROM, 5.4.
- O'Bannon, T., 1997: Using a terrain-based hybrid scan to improve WSR-88D precipitation estimates. Preprints, *28th Conf. on Radar Meteorology*, Austin, TX, Amer. Meteor. Soc., 506–507.
- Pamment, J., and B. Conway, 1998: Objective identification of echoes due to anomalous propagation in weather radar data. *J. Atmos. Oceanic Technol.*, **15**, 98–113.

- Riedmiller, M., and H. Braun, 1993: A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proc. IEEE Conf. on Neural Networks*, San Francisco, CA, IEEE, 586–591.
- Robinson, M., M. Steiner, D. Wolff, C. Kessinger, and R. Fulton, 2001: Radar data quality control: Evaluation of several algorithms based on accumulating rainfall statistics. Preprints, *30th Int. Conf. on Radar Meteorology*, Munich, Germany, Amer. Meteor. Soc., 274–276.
- Steiner, M., and J. Smith, 2002: Use of three-dimensional reflectivity structure for automated detection and removal of non-precipitating echoes in radar data. *J. Atmos. Oceanic Technol.*, **19**, 673–686.
- Stumpf, G., A. Witt, E. D. Mitchell, P. Spencer, J. Johnson, M. Eilts, K. Thomas, and D. Burgess, 1998: The National Severe Storms Laboratory mesocyclone detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 304–326.
- Venkatesan, C., S. Raskar, S. Tambe, B. Kulkarni, and R. Keshavamurthy, 1997: Prediction of all India summer monsoon rainfall using error-back-propagation neural networks. *Meteor. Atmos. Phys.*, **62**, 225–240.
- Zhang, J., S. Wang, and B. Clarke, 2004: WSR-88D reflectivity quality control using horizontal and vertical reflectivity structure. Preprints, *11th Conf. on Aviation, Range and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, P5.4.