Which Polarimetric Variables Are Important for Weather/No-Weather Discrimination?

Valliappa Lakshmanan, Christopher Karstens, John Krause, Kim Elmore, and Alexander Ryzhkov

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and National Severe Storms Laboratory, Norman, Oklahoma

Samantha Berkseth

Valparaiso University, Valparaiso, Indiana

(Manuscript received 23 September 2013, in final form 20 January 2015)

ABSTRACT

Recently, a radar data quality control algorithm has been devised to discriminate between weather echoes and echoes due to nonmeteorological phenomena, such as bioscatter, instrument artifacts, and ground clutter (Lakshmanan et al.), using the values of polarimetric moments at and around a range gate. Because the algorithm was created by optimizing its weights over a large reference dataset, statistical methods can be employed to examine the importance of the different variables in the context of discriminating between weather and no-weather echoes. Among the variables studied for their impact on the ability to identify and censor nonmeteorological artifacts from weather radar data, the method of successive permutations ranks the variance of Zdr, the reflectivity structure of the virtual volume scan, and the range derivative of the differential phase on propagation [PhiDP (Kdp)] as the most important. The same statistical framework can be used to study the impact of calibration errors in variables such as Zdr. The effects of Zdr calibration errors were found to be negligible.

1. Introduction

Quality control algorithms to remove nonmeteorological echoes, such as ground clutter and bioscatter (e.g., Grecu and Krajewski 2000; Steiner and Smith 2002; Kessinger et al. 2003; Zhang et al. 2004; Lakshmanan et al. 2007, 2010; Gourley et al. 2007; Lakshmanan et al. 2012; Chandrasekar et al. 2013), are an essential preprocessing step for many automated weather applications that employ weather radar data. Recently, Lakshmanan et al. (2014) described a quality control (QC) technique that uses polarimetric radar variables (both at a range gate and at range gates surrounding it in three dimensions) to discriminate between weather and no-weather echoes. In this paper, we use the term *polarimetric radar variables* to refer to all variables available from a polarimetric radar, that is, to both "standard moments" (reflectivity, Doppler velocity, and spectrum width) and "polarimetric moments," such as Zdr and RhoHV. Because the QC algorithm of Lakshmanan et al. (2014)—that is, quality control neural network dual polarization (QCNNDP)—was developed by optimizing its weights of the discrimination scheme chosen by optimization over a large, diverse training dataset, it is possible to employ statistical methods to study the importance of different variables for the purpose of discriminating between meteorological and nonmeteorological echoes.

While this paper focuses on variable importance solely for discriminating between meteorological and nonmeteorological echoes, it is our hope that the methods that we employ will prove useful to determine empirically the relative importance of polarimetric radar variables for tasks such as hydrometeor classification (Park et al. 2009), where a citizen-science project is collecting large datasets of surface hydrometeor types (Elmore et al. 2014). Indeed, the methods of assessing variable importance that we employ in this paper can be applied to any task where predictor variables are combined in an optimizable manner to approximate a "truth" value.

The rest of this paper is organized as follows. A formal definition of variable importance and a background of the statistical terms and concepts are provided in section 1a.

Corresponding author address: Valliappa Lakshmanan, University of Oklahoma, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: lakshman@ou.edu

DOI: 10.1175/JTECH-D-13-00205.1

The QC algorithm that the statistical analysis methods rely on is described in section 1b. Various methods of assessing variable importance are described in section 2, starting with noncontext-specific methods in section 2a, through univariate methods based on probability distributions in section 2b and a rigorous but computationally infeasible method based on retraining the QC algorithm on subsets of the training data in section 2c, to methods based on permuting variable values within the training dataset in sections 2d and 2e. The variable importance framework is used to answer the topical question of the impact of Zdr calibration errors in section 3a.

a. Variable importance

A formal, mathematical definition of variable importance is provided by van der Laan (2006), where the importance of variable A used in predicting quantity Y is defined in terms of model P for predicting Y given a vector of variables W (of which A is one) as

$$\psi(P)(a) = E_{P*}\{E_P[(Y \mid A_a, W) - E_P(Y \mid A_{\varnothing}, W)]\} \tag{1}$$

where E_P is the predicted value of Y using the model P and $P*$ is a distribution of (W, Y) that is a known function of P. In other words, the importance of a variable is the difference in predictive skill between when the variable is retained in the prediction ($A = A_a$) and when the variable is removed ($A = A_{\varnothing}$). A simpler definition of the importance of variable A in the prediction of Y is provided by Breiman (2001) as the extent to which the prediction of Y would suffer (measured in terms of some standard skill score) if the value of A were to be improperly assigned. Variable importance has found extensive use in biostatistics (Strobl et al. 2007).

Although Breiman (2001) introduced the variable importance method using random forests as the prediction model and classification rate as the skill score, the techniques he pioneered are valid for any machine learning method where the model parameters are obtained by computing the best fit to a training dataset. In this paper, we use a neural network as the prediction model and the Heidke skill score (Heidke 1926) as the measure of skill in order to determine variable importance. We would caution that variable importance is measured on the training dataset to which the model was fit, because the definition of variable importance implicitly assumes the presence of both a prediction model P and the set of values on which to estimate it (the training data). In particular, variable importance should not be measured on an independent dataset. The robustness of the ranking of variables can also not be determined by jackknifing the training dataset and leaving individual cases out because that changes the training dataset.

Interested readers may wonder why simpler methods may not be used, especially because they may be familiar with these statistical techniques from the related problem of feature selection, the problem of choosing which variables to use as inputs to a prediction model. For completeness, and to reassure skeptical readers, we carry out three of these methods in this paper and show their results on polarimetric variables.

First, principal components analysis (PCA) is often carried out to determine the variables that capture the most variance on a dataset. But because the quantity being predicted, Y, is not taken into account by PCA, PCA cannot be used to determine which of the high-variance variables actually have an impact on the quantity being predicted. Thus, high-variance variables in a training dataset may not even be *relevant* to the problem (Hand et al. 2001).

If the prediction problem is one of classification, then histograms of a variable in the two classes are often examined to see if there is a clear separation between the two classes. If, for example, low values of A are associated with $Y = 0$ and high values of A with $Y = 1$, then A is a highly relevant variable. It is possible to measure the extent of this separation using statistical methods (more on this later), but it does not serve well as a measure of variable importance because the input variables are often closely related. A simple example might serve to illustrate this. Suppose that $Y = 1$ when $(A^2 - A) \geq 100$ and most of the time when $B \leq 0$. Let us assume that we wish to determine the importance of three variables—A, B, C—in predicting Y and further that C and A are highly correlated: $C = A^3 - 1$. The univariate skill of A, B, C (the skill when using A, B, and C individually, such as by examining the separation in their histograms or using just the one variable to determine Y) will all be high. However, in terms of importance, we should choose either A or C first and B next because A and C will tend to provide the same information. In the literature on variable importance, this is often referred to as "masking" as in "A and C mask each other." It should be noted that both A and C are relevant to the problem of predicting Y—it is just that once we have A, we no longer need C. The histogram-based or univariate-skill-based methods are better than plain PCA in that they yield measures of variable relevance, that is, a measure of whether that variable by itself is suitable to consider as an input but cannot be used to assess variable importance because these methods are susceptible to the masking problem.

A third statistical approach may be also familiar from the literature on feature selection. One way to determine if a feature is useful is to leave it out,[1] retrain the neural network, and measure its skill. The extent to which the skill of the neural network suffers is a measure of how important the variable is. This is truly a measure of variable importance as depicted by Eq. (1), but it does not scale because of the curse of dimensionality. Consider a problem where we wish to rank N variables. There are 2^N potential combinations of variables that we would have to assess. This is completely infeasible for realistic values of N.

Because of these reasons, the variable importance statistical framework introduced by Breiman (2001), formalized in van der Laan (2006), and improved by Radivojac et al. (2004) is what we will use in this paper.

b. The QC algorithm

QCNNDP (Lakshmanan et al. 2014) operates on six moments available from the WSR-88D polarimetric radar: reflectivity (Z), velocity (V), spectrum width (SPW), correlation coefficient (RhoHV), differential reflectivity (Zdr), and differential phase on propagation (PhiDP). Because PhiDP values within a radial depend on an arbitrary starting angle, the range derivative of PhiDP [termed Kdp as a shorthand, although no sophisticated method of imposing monotonicity, such as that of Wang and Chandrasekar (2009) and Giangrande et al. (2013), is carried out] was used instead of the PhiDP. All other moments are their raw values as reported by the radar (e.g., Doppler velocity values are not dealiased).

At each range gate, a set of features is collected. The first set of features corresponds to the values of the radar moments at the given range gate: the absolute of the Doppler velocity value (absvel), SPW, the reflectivity value (dBZ; refl),[2] Zdr, RhoHV, and the Kdp. The absolute value of Zdr (abszdr) and the gate-to-gate azimuthal shear computed from Doppler velocity (azshear) are also used as input features. In addition, a set of vertical features is computed from the virtual volume of reflectivity [i.e., the approach of Lynn and Lakshmanan (2002) is followed whereby, rather than wait for the highest tilt to construct a volume scan, a virtual volume of reflectivity is built that consists of the latest scans from the radar at each tilt of the volume coverage pattern being followed]. These are as follows:

(i) The maximum reflectivity value (dBZ) in the vertical column that includes the given range gate (dbzcomp)

(ii) The maximum height at which the Z value is greater than -14 dBZ (“height”) and is greater than 0 dBZ (htgood)

(iii) The Z value (dBZ) at 3 km in height from tilts greater than 1° (dbz3km)

(iv) The difference in Z value between the lowest tilt and the next higher tilt, whose elevation is more than 1° (delta)

The third set of features consists of the local variance of Z and Zdr (refvar and zdrvar) computed in a 5×5 neighborhood centered around the range gate. Finally, the result of a “simple” classifier (the number of criteria met where the criteria consist of thresholding RhoHV, |Zdr|, and reflectivity at 0.9, 2.3 dB, and 3 dBZ, respectively) is also computed and used as an input to a neural network.

The set of computed features at a range gate are presented to a neural network whose architecture is such that the probability that a range gate has meteorological echo is computed from the features x_i using

$$P_{\text{pixelwise}} = S\left(\sum_j \left\langle b_{h_j} + w_j \tanh\left\{b_0 + \sum_i [w_{ij} S(x_i + b_i)]\right\}\right\rangle\right), \tag{2}$$

where S is the sigmoid or logistic function $S(x) = 1/(1 + e^{-x})$ and tanh is the hyperbolic tangent function $\tanh(x) = (e^{2x} - 1)/(e^{2x} + 1)$. The weights (the w), biases (b and b_h), and the number of hidden nodes (the number of js) were obtained by numerical optimization on the training dataset shown in Table 1 [this table is larger than the one shown in Lakshmanan et al. (2014) because of incremental improvements in the QC algorithm] with the target classification (or “truth value”) for each range gate determined by a human expert on the basis of surface observations.

2. Measuring variable importance

Variable importance was assessed in several different ways: through using correlograms, creating subsets of variables, analyzing probability distributions, and conducting permutation tests. Because each of these methods makes some assumptions, it is useful to examine the results from all of them to gain a holistic view of variable importance. The methods below are arranged such that those discussed earlier in this section have shortcomings that are addressed by the methods addressed later.

[1] Note that this is different from jackknifing the training dataset by leaving individual cases out—here we are jackknifing the training dataset by leaving individual input variables out.

[2] All computations on the reflectivity field are carried out in reflectivity decibels.

TABLE 1. Data cases used for assessing variable importance.

Radar	Location	Date/time (UTC)	Description
KABR	Aberdeen, SD	11 Aug 2012/0800	Insects and light rain
KARX	La Crosse, WI	26 Jun 2012/1900	Insects and light rain
KARX	La Crosse, WI	26 Sep 2012/2000	Insects
KCLE	Cleveland, OH	24 Feb 2012/1300	Rain
KCLE	Cleveland, OH	24 Feb 2012/1800	Rain
KCLE	Cleveland, OH	24 Feb 2012/2300	Snow
KDLH	Duluth, MN	26 Sep 2012/1700	Insects
KEWX	Austin, TX	6 May 2012/0600	Instrument artifacts and storms
KEWX	Austin, TX	6 May 2012/0900	Instrument artifacts and storms
KEWX	Austin, TX	31 Mar 2013/2300	Storms and three-body hail spike
KFSD	Sioux Falls, SD	4 Aug 2012/0000	Line of storms
KFTG	Denver, CO	24 Feb 2013/2100	Snow
KGYX	Portland, ME	24 Feb 2013/1800	Snow
KHTX	Huntsville, AL	11 Aug 2012/0000	Storms
KJAX	Jacksonville, MS	23 Sep 2012/0800	Birds
KJAX	Jacksonville, MS	26 Sep 2012/1800	Ground clutter and storms
KJGX	Robins Air Force Base, GA	20 Sep 2012/0200	Birds
KLOT	Chicago, IL	4 Aug 2012/2000	Storms
KLSX	St. Louis, MO	26 Sep 2012/1900	Storms
KLSX	St. Louis, MO	24 Feb 2013/2000	Insects
KLTX	Wilmington, NC	17 Sep 2012/1600	Storms and sea clutter
KMHX	Morehead City, NC	19 Sep 2012/2200	Storms
KMRX	Knoxville, TN	24 Jun 2012/2300	Insects and storms
KNQA	Memphis, TN	21 Sep 2012/1000	Storms and AP
KOTX	Spokane, WA	26 Sep 2012/1600	Ground clutter
KPDT	Pendleton, OR	20 Jul 2012/1000	Rain
KRLX	Charleston, WV	1 Oct 2012/0800	Ground clutter and rain
KSGF	Springfield, MO	29 Feb 2012/0600	Line of storms
KTFX	Great Falls, ND	26 Sep 2012/1500	Insects and rain
KTLX	Oklahoma City, OK	19 May 2013/2200	Mix of weather and turbines
KTLX	Oklahoma City, OK	20 May 2013/2000	Line of storms
KTYX	Ft Drum, NY	24 Feb 2013/1900	Snow
KVNX	Vance Air Force Base, OK	20 May 2011/0800	Convection
KYUX	Yuma, AZ	23 Jul 2012/1500	Ground clutter
KYUX	Yuma, AZ	23 Jul 2012/1600	Insects and rain
KYUX	Yuma, AZ	26 Sep 2012/2100	Insects
KYUX	Yuma, AZ	29 Sep 2012/0900	Ground clutter

a. Correlogram

PCA was applied to the training dataset (leaving out the truth value) and the variables arranged in the order of the weight of that variable in the principal component. Figure 1 shows the variables in PCA order. The majority of the variance within the training data is explained by the local variance of reflectivity, by the azimuthal shear and by the local variance of Zdr and RhoHV—that is, by the "noisy" variables that are formed by differentiating smoother moment fields. These variables convey a lot of information [in the information-theoretic Shannon entropy (Shannon 1948) sense]. The PCA order is not very useful, however, to decide on variable importance because it is independent of whether echo is due to meteorological phenomena or not.

In Fig. 1 the section below the diagonal also depicts the Pearson correlation coefficient ρ of each pair of predictors using pie charts. A pie chart filled from 1200 to 1500 indicates a ρ of 0.25, whereas a pie chart filled from 12 to 2100 (clockwise) indicates a ρ of 0.75. On the other hand, a pie chart filled from 0900 to 1200 shows a ρ of -0.25. Because the sign of ρ does not matter much, one is looking primarily for circles that are almost fully filled. The results are mainly as expected. For example, Zdr and abszdr are correlated to about 0.5, which is as expected. The graph of this relationship is shown by a red line in Fig. 1 above the diagonal—this, too, is what one would expect from the relationship between a variable and its absolute value. The ellipse that forms the background of the red line shows the distribution of all the points—something that, admittedly, would make more sense with axes that show the actual values. Examining this diagram, the only high correlations (magnitude > 0.5) are between the following pairs of variables:

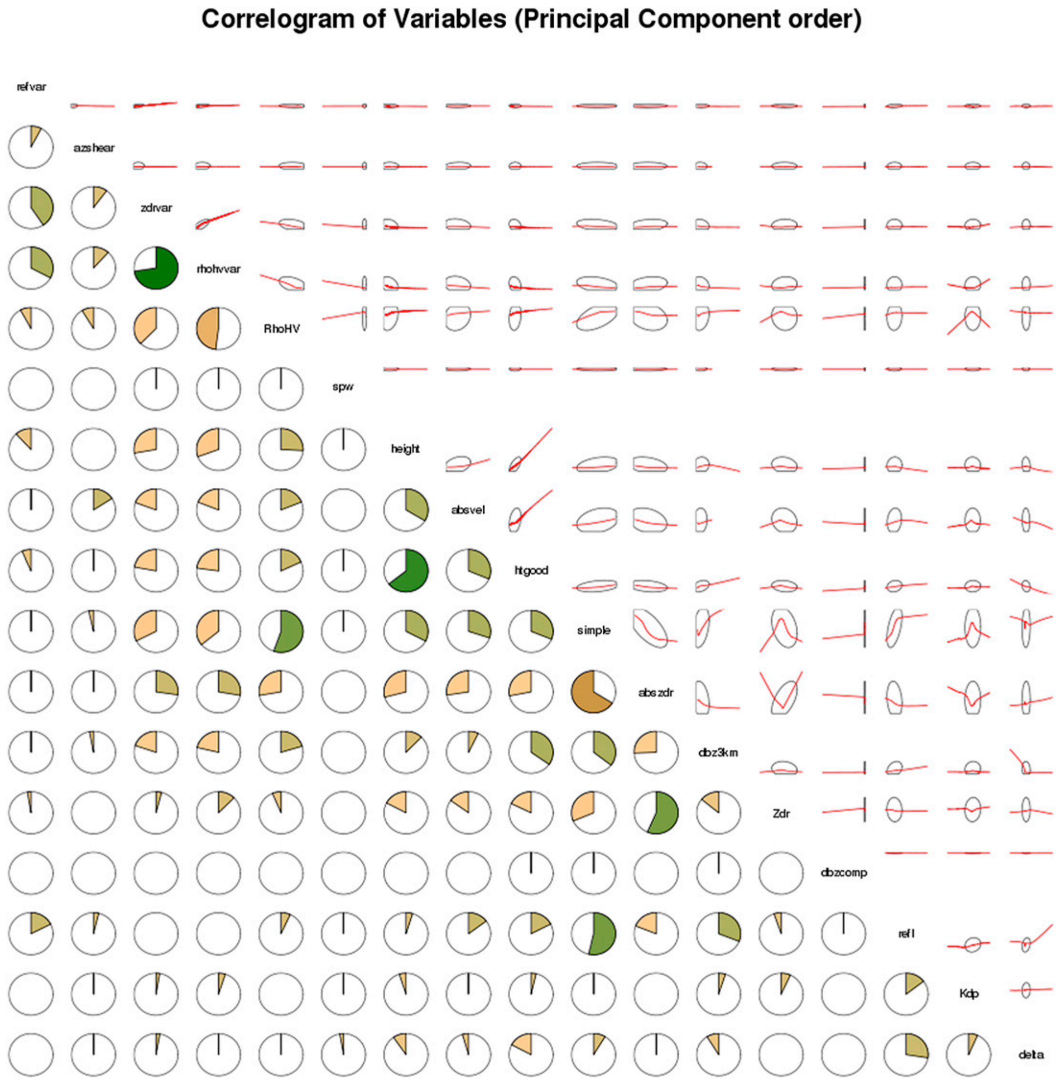

FIG. 1. Correlogram of the predictor variables showing the Pearson correlation coefficient (filled circles), and the relationship between variables (red lines) and their distribution (ellipses). The variables are (top to bottom) refvar, azshear, zdrvar, rhohvvar, RhoHV, spw, height, absvel, htgood, simple, abszdr, dbz3km, Zdr, dbzcomp, refl, Kdp, and delta.

(i) RhoHV and the simple classifier
(ii) abszdr and the simple classifier
(iii) reflectivity and the simple classifier
(iv) abszdr and Zdr
(v) zdrvar and rhohvvar
(vi) height of 0 (htgood) and of $-14\,\mathrm{dB}Z$ (height)

Because the simple classifier is defined as the number of criteria met where the first criterion is that RhoHV is greater than 0.9, the second is that |Zdr| is less than 2.3 dB, and the third is that the reflectivity is greater than $3\,\mathrm{dB}Z$, the high correlation between the simple classifier and these three variables is quite explicable. The presence of such a high correlation indicates that QCNNDP could omit the simple classifier from the set of predictands. The high correlation between the heights of 0 and of $-14\,\mathrm{dB}Z$ can be explained by observing that taller storms have high values of both these variables and that shallow systems have low values of both of them. The fact that the values are closely related seems to imply that any one of these variables might be enough. On the other hand, it might simply be the case that the high correlation exists only in regions far away from the radar where echo heights are poorly sampled. The high correlation between zdrvar and rhohvvar arises from their coincidence in situations of bioscatter (where both are high) and of precipitation (where both are low). Again, perhaps using just one of the variables carries enough information that the other is redundant.

Using PCA ranking to assess variable importance suffers from nonspecificity. It is possible to determine which variables contribute the most to the information content of the dataset but not to determine whether the information content of those variables is relevant to a particular problem. For example, knowing that zdrvar ranks highly in terms of PCA does not signify whether that variable contributes to the discrimination of weather echoes from nonweather ones [Hand et al. (2001) illustrate this problem on synthetic data]. After all, adding a column of white noise will result in a column that contributes heavily to the variance of a dataset while not contributing at all to anything meaningful. The correlation and relationship between variables that is shown in the correlogram of Fig. 1 is useful, however, to determine how unique a particular variable is.

b. Kullback–Leibler J measure

One way to assess variable relevance to a classification problem is to examine the class-conditional distributions $P(X = x \mid Y = 0)$, where X is the variable in question and Y is the labels (weather or no weather) associated with the pixel to which the X value corresponds. If $P(X = x \mid Y = 0)$ and $P(X = x \mid Y = 1)$ are extremely similar, then the variable does not possess any discrimination power *by itself*. Note that because we are not considering pairs of variables, looking at the class-conditional distributions of the variable does not provide information on whether a combination of variables possesses high discrimination skill.

To measure the divergence between two probability distributions, one can use the symmetric Kullback–Leibler J measure (Jeffreys 1946; Lin 1991) to obtain the importance of the ith variable, where

$$J_i = \sum_x [P(X_i = x \mid Y = 0) - P(X_i = x \mid Y = 1)] \log_2 \frac{P(X_i = x \mid Y = 0)}{P(X_i = x \mid Y = 1)}. \quad (3)$$

When the two probability distributions are identical over the full range of x, it is easy to see that $J_i = 0$, whereas if the distributions are fully divergent, then the measure tends toward infinity.

The probabilities were approximated by computing class-conditional histograms of the training dataset using 10 equally spaced bins between the minimum and maximum values in the ith column and smoothing the histogram by a Gaussian kernel density function with a bandwidth of two bins.

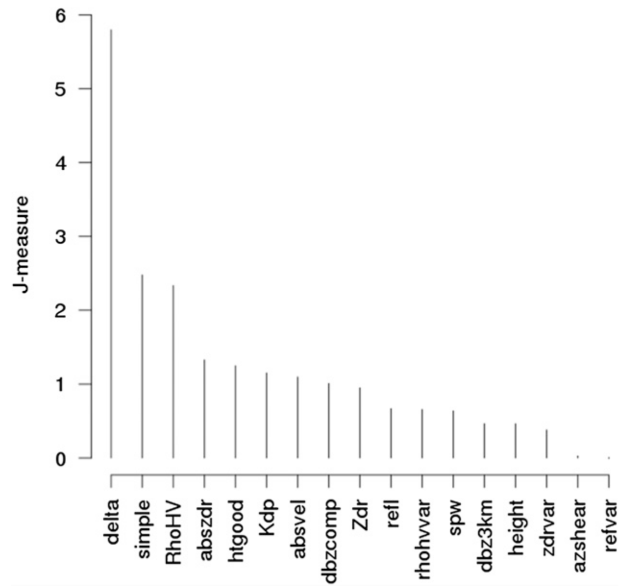

FIG. 2. Variable relevance assessed using the Kullback–Leibler J measure.

The variables ranked by the J measure are shown in Fig. 2. It is clear that variables such as the difference in reflectivity values between the two lowest tilts (delta, used in Grecu and Krajewski 2000) and the value of RhoHV (used in Gourley et al. 2007; Tang et al. 2014) are highly relevant to the problem of discriminating between weather and no-weather echoes. As will be seen later in the paper, though, these variables are not the most important ones for this discrimination.

To understand the difference between variable relevance and variable importance, it might help to note that because the J measure is a measure of how separate the probability distributions for meteorological and nonmeteorological echoes are, it is a measure of *univariate* skill. In Lakshmanan et al. (2014), univariate skill was measured by computing the Heidke skill score (HSS; Heidke 1926) at various thresholds and finding the threshold at which the HSS was maximum for each variable; based on this, the simple classifier that thresholded the data on reflectivity, Zdr, and RhoHV was devised. The J measure illustrates why this strategy worked: two of the three variables (RhoHV and Zdr) rank in the top half in terms of univariate discrimination skill, while the third variable (reflectivity) was used as a way to select candidate weather pixels. Indeed, the combination of the three thresholded variables (simple) ranks above any one of them.

However, while univariate skill is a measure of variable relevance, it does not translate directly into variable importance because the same information may be

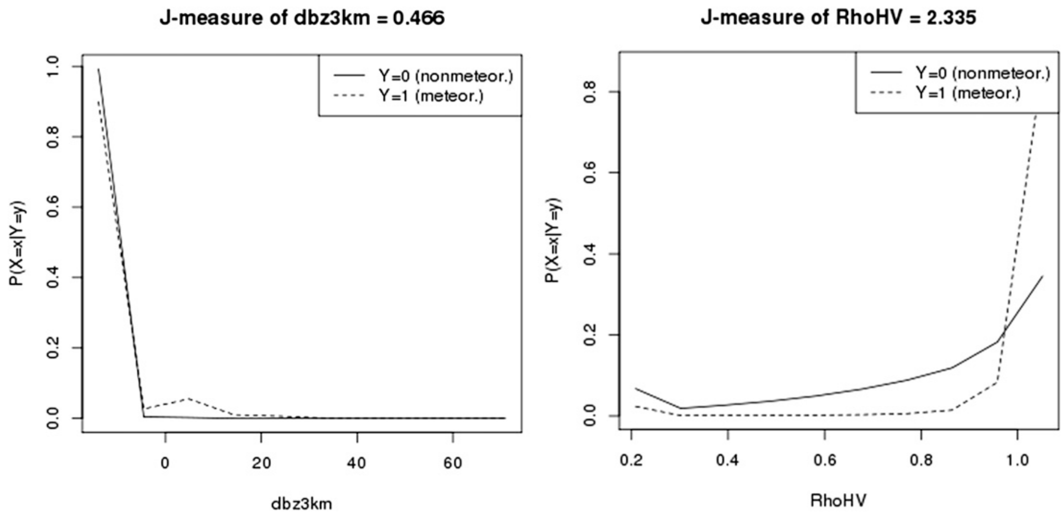

FIG. 3. The J measure is computed from the class-conditional distributions of a variable but does not take into account separability over specific ranges or the uniqueness of a variable's contribution. The class-conditional distributions of two variables with very different J measures are shown: (left) dbz3km and (right) RhoHV.

provided by multiple variables, thus making any single variable in the set unimportant. Another reason is that looking at variable relevance in terms of class-conditional distributions fails to take advantage of natural splits that occur in the data. Take, for example, the two variables shown in Fig. 3: while the distributions of the dbz3km variable are poorly separated over small values of the variable, the separability for midranges is higher and possibly provides unique skill (i.e., no other variable serves to separate those groups of pixels as well). The RhoHV distributions are better separated (with a higher J measure than dbz3km) at all ranges. However, the same separation possibly could be achieved by some other polarimetric variable. Thus, while the J measure can provide an indication of variable relevance, it should not be used by itself to gauge variable importance in a realistic discrimination algorithm.

c. Subsets

Given that the neural network is optimized from the training dataset, the importance of any particular variable can be gauged by removing that variable from the dataset, as suggested by Eq. (1), and retraining the neural network on the reduced dataset (which will be a subset of the original dataset). The drawback of this method is that it is extremely time consuming. The graph in Fig. 4, for example, took four full days of continuous computation on a powerful workstation. Readers will realize that the subsets shown in Fig. 4 are only those subsets with one of the variables removed, but to be complete, one would have to consider removing sets of two variables, sets of three variables, etc. With N variables, there are 2^N such subsets.[3] When 18 subsets take 4 days of continuous computation, a complete analysis of the method of subsets cannot be scalably carried out, since it would take 160 yr (although the time can be reduced by carrying out this embarrassingly parallel job on more computers).

Examining Fig. 4 points to another drawback of the method of subsets. The removal of variables such as dbz3km or htgood results in a significant reduction in the skill of the trained neural network but note that removal of the Zdr or the simple classifier results in an increase in the skill. This is counterintuitive, since the original neural network could have attained the skill of the new neural network by simply assigning the weight of the Zdr variable to zero. There are two possible explanations of how this situation could arise. One is that the original neural network may have been caught in a local minimum and that removing the variable in question enabled the gradient–descent optimization to find a better minimum. Another is that this is simply an artifact of the random nature of the neural network training and that if several more runs are carried out, then the error bars of the HSS will overlap. Naturally, with the expense of the method of subsets, carrying out hundreds of neural network training runs for each subset in order to determine the variability of the HSS that would be achieved on each

[3] This can be obtained by adding ${}^{N}C_1 + {}^{N}C_2 + \cdots + {}^{N}C_N$, but a simpler method is to observe that in each subset, each variable can either be included or not, so there are $2 \times 2 \times 2 \ldots$ or 2^N possible subsets.

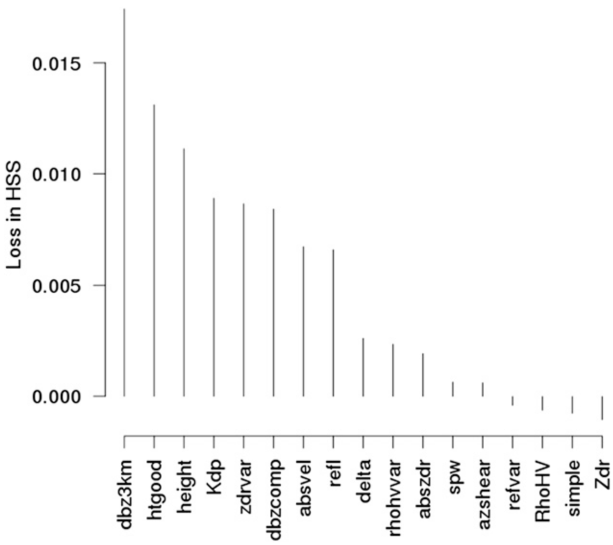

FIG. 4. Variable importance assessed by removing a variable and retraining the neural network: loss in HSS vs variables.

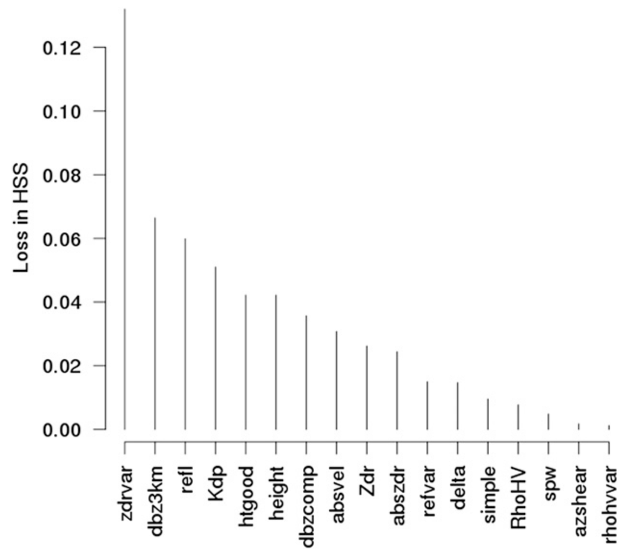

FIG. 5. Variable importance assessed by scoring the QC algorithm performance after permuting a variable: loss in HSS vs variables.

subset is quite infeasible. However, the fact that the simple classifier also showed up in the list of highly correlated pairs (see Fig. 1) indicates that this is one variable that could be safely removed.

Another interesting inference is available from Fig. 4. It may be noted that the absolute value of Zdr proves reasonably useful (in the middle of the range), while the usefulness of Zdr is quite low. This indicates that decorrelating the two variables may be quite useful. For example, rather than using Zdr and |Zdr|, it might be better to employ |Zdr| and sign(Zdr) as the two input variables.

At the other end of the scale, the most important variables all have to do with the virtual volume of reflectivity: dbz3km, the height of 0 dBZ (htgood), and the height of −14 dBZ turn out to be the three most important variables. Because the training dataset consists of range gates remaining after preclassifying echoes with dbz3km above 11 dBZ as being good (Lakshmanan et al. 2014) (i.e., the training dataset consists only of echoes for which dbz3km is less than or equal to 11 dBZ), the importance of dbz3km is being understated here. The most important polarimetric variables are Kdp and the variance of Zdr.

The fact that the zdrvar is more important than rhohvvar (with which it is highly correlated) indicates this is the variable to keep. Considering that calibration of Zdr is difficult to get right, and that there may be systematic issues with the calibration of Zdr (Ryzhkov et al. 2005), it is telling that the variance of Zdr (which is relatively immune to calibration issues) is more important than Zdr. However, there is enough concern about Zdr calibration that this issue will be addressed more thoroughly in section 3a.

It should also be noted that the magnitude of the loss in HSS is quite small. The HSS when using all the predictor variables is 0.75. Thus, a loss in HSS caused by the removal of the dbz3km variable is less than 2%. This indicates that the information provided by any individual variable is mostly redundant. No single variable is indispensable.

d. Permutation

A less computationally expensive way to rank variables by importance is to permute the values of a variable within the training dataset and then evaluate the skill of the original neural network on the permuted values (Radivojac et al. 2004). Essentially, to determine the importance of, say, Zdr, the Zdr values are randomized in a way that retains the original distribution of Zdr values and the QC algorithm carried out. The extent to which the QC algorithm is affected by the randomization of Zdr values is dependent on how important Zdr is.

The variables ranked by order of importance as assessed by the permutation test are shown in Fig. 5. Five of the top six places here go to the same set of variables as in the method of subsets, although their relative rankings are somewhat different. This is encouraging, as unlike PCA or the Kullback–Leibler J measure, the method of permutation provides an accurate measure of variable importance in the context of a specific application. Because there is no reoptimization–retraining in this method, the computational efficiency is driven by the speed at which one can permute variables within a training dataset and evaluate its skill—the graph in Fig. 5 was created in about an hour. Thus, the

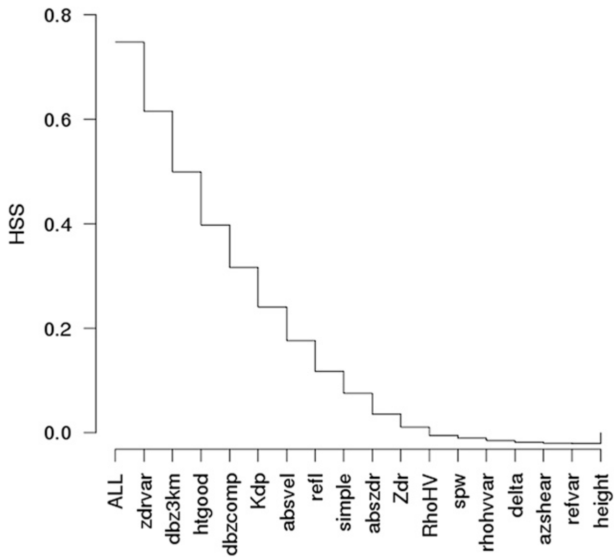

FIG. 6. Variables ranked by permuting subsets. The successive loss in HSS as each variable is removed is shown. Variables are removed in order of importance, so that earlier removals result in greater loss in skill.

permutation method is an accurate, computationally efficient way to study variable importance.

e. *Permutation rank*

Because the permutation method is computationally efficient, it becomes quite feasible to carry out a systematic process to rank variables by importance using this inductive procedure:

(i) The most important variable (rank of $k = 1$) is obtained using the permutation method described in the previous section.

(ii) Given the k most important variables, one can find the $(k + 1)$th most important variable by keeping the k variables permuted and finding the next most important variable by permuting each of the remaining variables one by one. The variable that results in the greatest loss in skill carries the $(k + 1)$th rank.

If one has N variables, then this process requires that $N(N + 1)/2$ permutations be carried out.

The most important variable was obtained by permuting all the columns one by one and finding the one that resulted in the greatest loss in skill (see Fig. 5). We obtained the second most important variable by keeping the zdrvar column permuted and permuting each of the other columns one by one in order to find the column that resulted in the greatest loss in skill. This is the second most important variable. From the two most important variables, we obtained the three most important variables, and so on. The result of this process is shown in Fig. 6, where the best variable found at each step of the inductive process is listed in the x axis and the HSS achieved after removing that variable is shown on the y axis. As each variable is removed, the HSS keeps falling.

Figure 6 can be interpreted as follows: if one were able to utilize only five variables in a polarimetric QC method, then the variables to choose should be the local variance of Zdr, a reflectivity at 3 km height, a height of reflectivity of at least 0 dBZ, the maximum reflectivity within the vertical column, and the range derivative of PhiDP (Kdp). Note that this is different from picking the five most important variables from Fig. 5 because that fails to account for whether a pair of highly ranked variables serves to classify (or misclassify) the same set of pixels. In other words, the ranking in Fig. 5 does not take into account intervariable correlations in the context of the weather/no-weather discrimination problem, whereas the ranking in Fig. 6 does. Figure 5 is just the first step of the process that was used to create Fig. 6.

It should be noted that there is a difference in methodology between Figs. 5 and 6 and that difference is reflected in the way the graphs are drawn. In Fig. 5, we studied the impact of permuting the variables one by one while keeping all the other variables at their original values. Hence, the y axis of that graph is the loss in HSS attributable to each variable in turn. Once it was decided that zdrvar was the most important variable, one can move on to creating Fig. 6. The zdrvar values were kept permuted. The rest of the variables were permuted one by one. It was then found that dbz3km was the next best variable (i.e., it resulted in the most loss in skill). At the next step, the two most important variables were kept permuted and the remaining $n - 2$ variables were permuted one by one to choose the third most important variable. Hence, Fig. 6 shows a stepwise reduction in HSS as the neural network sees fewer and fewer good variables.

The most important variables are the variance of Zdr, variables relating to the 3D reflectivity structure (reflectivity at 3 km above ground level, height of 0-dBZ echo, and maximum reflectivity within the 3D column), and the range derivative of PhiDP (Kdp). It is not surprising, therefore, that previous QC methods have used combinations of these variables to good result. Nonpolarimetric QC methods (Grecu and Krajewski 2000; Steiner and Smith 2002; Lakshmanan et al. 2007) have relied heavily on the structure of the 3D reflectivity field. In their fuzzy logic polarimetric radar QC method, Gourley et al. (2007) use the spatial variance of Zdr and PhiDP but no variables relating to the 3D reflectivity structure. In their physical rules-based method, Tang et al. (2014) employ RhoHV primarily (i.e., neither the variance of Zdr nor the range derivative of PhiDP), but their corrections involve the use of reflectivity at high altitudes to indicate whether low RhoHV values could be due to the presence of hail.

f. Limitations

One caveat that should be kept in mind is that the results of this paper, as with most studies that make conclusions on the basis of data, are limited by the scope and diversity of the dataset that was used. If there is a specific class of phenomena that is not represented in our dataset (e.g., derechos), then variables important in correctly classifying echoes for that phenomenon may be deemphasized in so far as that phenomena differ from others in its class. Because we have found that the QC of derechos requires no special handling and that the QC processes that work for other weather phenomena also work for derechos, this caveat does not apply to derechos. However, as we operate QCNNDP in real time, we discover instances where the QC algorithm does not work well and add new data cases to the training dataset. It is certainly possible that adding data cases of some specific meteorological or nonmeteorological phenomenon in this manner will alter the rankings. Similarly, choosing different subsets of the dataset would result in different rankings. For example, different variable rankings would result if the dataset were to be restricted to only winter weather or only to times of bird migration.

Also, the ranking of variables in this paper is limited to the list of variables that the current literature suggests as being important. For example, we have noticed that the noise correction applied to RhoHV can lead to RhoHV values greater than 1 and electronic interference echoes tend to have RhoHV greater than 1 due to overcorrection for noise (Ivic and Melnikov 2013). If, in addition to the radar moments currently being distributed, a gate-by-gate noise estimate were to also be available, then it might prove important to weather/no-weather discrimination. Also, the variable importance ranking in this paper reflects the variables as they are currently computed. Variable transformations, outlier detection, and better noise correction may improve the discrimination capability of a variable beyond the rank suggested in our study. Of course, as will be shown in section 3b, such transformations may also reduce the rank of a variable.

In other words, then, the process of assessing variable importance described in this paper should be carried out again if shortcomings of the QC algorithm are identified or new discriminant variables are proposed.

3. Extensions

We set out to study the importance of the different polarimetric variables in the context of the QC problem in order to provide guidance to other designers of QC methods, especially ones based on physical understanding, such as that of Tang et al. (2014), and fuzzy logic ones, such as that of Gourley et al. (2007). However, the variable importance framework can be employed to address questions such as the impact of Zdr calibration and to reduce the number of input variables in an automated learning method.

a. Impact of calibration problems

The variable importance framework described above may be used to explore the sensitivity of the QC algorithm to miscalibration of Zdr values. Recent research indicates that Zdr calibration errors on the WSR-88D network are normally distributed with nearly all the calibration errors between -1 and 1 dB, but severe hardware problems can cause calibration errors as low as -2 dB (R. Lee and J. Cunningham, NOAA's Radar Operations Center, 2013, personal communication). To study the impact of Zdr calibration errors, we added a bias within the range -2 to 2 dB to the Zdr values in the training set (thus affecting the values of Zdr, abszdr, and the simple classifier). Then, the originally trained neural network was carried out on the biased dataset and the Heidke skill score was calculated. The results are shown in Fig. 7.

As expected, as the miscalibration of Zdr increases, the resultant loss in skill becomes greater. It appears that a positive bias in Zdr value causes a slightly greater loss in skill than a negative bias of the same magnitude. However, in all cases, the loss in skill is quite small, less than 0.02 (the actual HSS is about 0.75). In other words, even a severe miscalibration of Zdr, such as a -2 dB calibration error, results in only a 2%–3% loss in skill. When the magnitude of the Zdr bias is below 1 dB (as is characteristic of the WSR-88D network), the loss in HSS is negligible. Discriminating between weather and no-weather echoes, therefore, is quite insensitive to Zdr biases.

We would like to caution that this result should not be overinterpreted. First, it is limited to the question of the impact of Zdr calibration errors on quality control applications—applications using Zdr values for other applications may be affected more. Second, this result does not mean that Zdr calibration errors can be ignored—it could also be that Zdr differences of less than 1 dB are discounted by the neural network precisely because of endemic Zdr calibration problems. Perhaps, if unbiased Zdr data had been available to the neural network training, a variable such as |Zdr| would have been emphasized more. In this respect, it is telling that the variance of Zdr, which is impervious to calibration errors, is the most important variable and that the range of Zdr

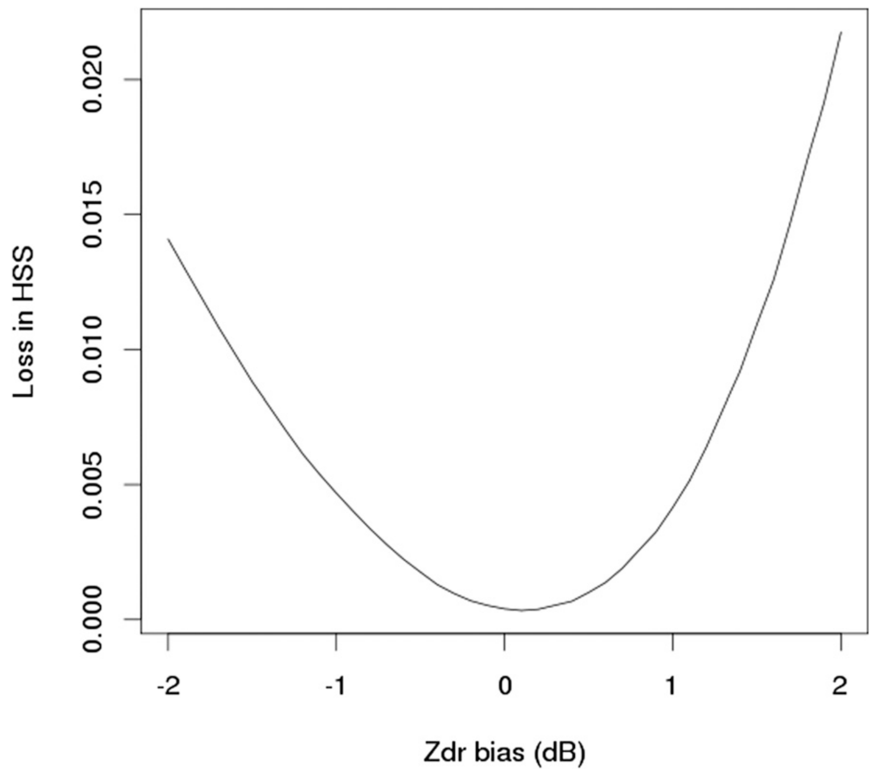

FIG. 7. Impact of Zdr miscalibration: loss in HSS vs Zdr bias.

biases that have almost no impact on QC algorithm performance is the same −1- to 1-dB range that actual WSR-88D calibration biases lie within.

b. *Improved computation of Kdp*

Considerable research efforts have been focused on improving the computation of specific differential phase or Kdp. In this paper, we defined Kdp as the range derivative of PhiDP and computed it as simply a gate-to-gate difference along the radial, making sure to skip over gates where the reflectivity was below the signal-to-noise threshold.

Theoretically, PhiDP ought to be monotonous and so Kdp should always be a positive value. In practice, the presence of strong horizontal or vertical gradients of precipitation within the beam sampled produce oscillations in PhiDP. This nonuniform beamfilling (NBF) becomes more common as the beam broadens with range and often produces negative values of Kdp (Ryzhkov 2007). Kdp can also be corrupted by PhiDP wrapping (crossing from 359° to 1°, a 2° change will result in a value of −358), whether the radar is operating in simultaneous transmitting mode or alternative transmitting mode, depolarization due to canted ice crystals (Ryzhkov and Zrnić 2006), and the transition from a regime of weather into nonweather or its opposite.

Would an improved Kdp computation improve the performance of the QC network and/or increase the relative importance of Kdp? The simple range-derivative Kdp computation was replaced with a method patterned after Ryzhkov et al. (2000) using a nine-gate average and a threshold of RhoHV ≥ 0.9. The Kdp obtained was the slope of a least squares linear fit to the PhiDP. This is similar to the computation used on the WSR-88D network for convective locations ($Z > 40$ dBZ).

The impact of improving the Kdp computation is shown in Fig. 8. The HSS, at 0.74, is approximately the same as the HSS achieved with a simple range derivative of PhiDP. Counterintuitively, a smoother Kdp calculation leads to the Kdp variable becoming much less important—the variable drops from being the fifth most important (see Fig. 6) to being one of the least important. It appears that, for the QC application at least, noise-tolerant ways of computing Kdp are actually counterproductive. Again, we would like to caution that our finding is limited to the QC problem and that other applications of polarimetric data may benefit by more noise-tolerant algorithms.

c. *Reduced feature set*

One potential use of ranked variable importance is in the selection of features for the neural network. A

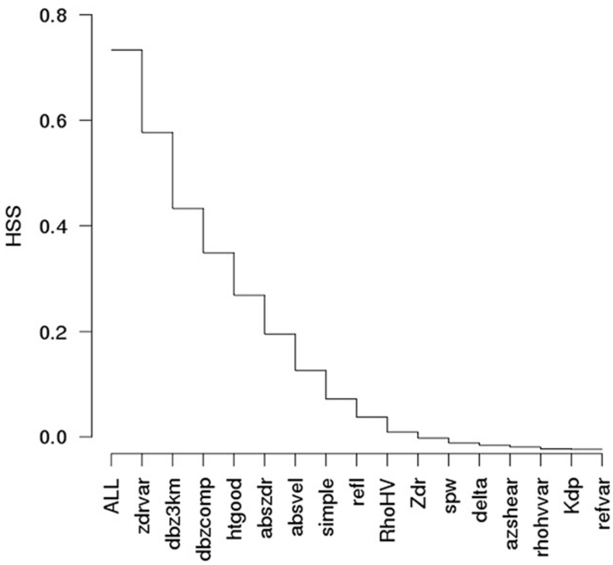

FIG. 8. Rankings based on a noise-tolerant method for computing Kdp: HSS vs variables. Rather surprisingly, a more noise-tolerant method for computing Kdp does not result in the Kdp variable becoming more useful.

neural network with a reduced number of input variables may perform better than one with the full set of inputs because of the curse of dimensionality, whereby a neural network's mapping of the input space becomes sparser as the number of input variables increases (Bishop 1995). Consequently, a reduced variable set was created with only high-importance inputs (we retained the variables that explained all but 0.1 of the HSS) and where only one variable of highly correlated pairs was retained.

The new set of input features at each range gate included 1) absvel ($\mathrm{m\,s^{-1}}$), 2) refl (dBZ), 3) Zdr (dB), 4) RhoHV, 5) Kdp, 6) dbzcomp, 7) dbz3km, 8) htgood, and 9) zdrvar in a 5×5 neighborhood around the range gate. A new neural network was trained with this reduced set of input variables. The importance of the various variables in this reduced input space is shown in Fig. 9c. Readers will note that the starting HSS of 0.72 is slightly less than the HSS of 0.75 achieved by the NN that operates on the full parameter set. Moreover, the lower HSS on the training dataset is not compensated by a more general mapping of the reduced parameter space: on the same case of independent cases used by QCNNDP (Lakshmanan et al. 2014), the HSS of the NN trained with a reduced set of variables is 0.76, whereas the HSS of the NN training using all the variables was 0.82 (see Figs. 9b and 9c). The resulting NN is about 10% more computationally efficient because variables that are not going to be used do not have to be computed. However, because the original implementation was also capable of operating in real time, there is little practical benefit to this increase in computational efficiency.

Of course, other reduced parameter sets are possible. In particular, one may omit fewer of the original variables, or decorrelate pairs of highly correlated variables (instead of dropping one variable from each pair).

d. *Variance of Zdr versus RhoHV*

A somewhat controversial result of this paper is that the variance of Zdr is a more important variable than RhoHV for the purpose of discriminating between weather and no-weather echoes (see Fig. 6). Three arguments are likely to be raised: 1) RhoHV and zdrvar are highly correlated, so how can one be better than the other? 2) Many successful QC algorithms have been devised based on RhoHV, so how can it be the case that RhoHV is not very important? 3) What explains the dramatic difference in ranking between these two essentially equivalent variables? We will address these three arguments in turn.

First, it may seem unlikely to many readers that zdrvar and RhoHV provide different information. After all, the spatial variance of radar variables is caused by both spatial nonuniformity of the fields and statistical fluctuations due to limited dwell times, and Melnikov and Zrnić (2004) shows that the variance of Zdr is fundamentally tied to RhoHV:

$$\sigma^2(\mathrm{Zdr}) = 21.34\frac{1-\rho_{\mathrm{hv}}^2}{M\sigma_{\mathrm{vn}}}, \tag{4}$$

where σ^2 is the variance, ρ_{hv} is RhoHV, M is the number of radar samples in a dwell time period, and σ_{vn} the normalized spectrum width given by $4WT/\lambda$, where W is the spectrum width, T is the pulse repetition period and λ is the radar wavelength. However, the equation given above is valid only in areas of high signal-to-noise (SNR) and the relationship holds only in homogeneous media. If the medium is not statistically homogeneous (e.g., in the presence of ground clutter), then Zdr can vary dramatically from gate to gate regardless of the magnitude of RhoHV. So, it is possible that zdrvar is statistically different from RhoHV, at least in such situations.

Second, that RhoHV is not an important discriminator seems counterintuitive because successful radar QC algorithms such as that of Tang et al. (2014) have been designed that rely solely on RhoHV. To understand this conundrum, readers need to distinguish between the rankings in Figs. 2 and 6. Readers should also note that, as a univariate linear discriminator, RhoHV is superior to the variance of Zdr (this is shown in Fig. 2). It is only with the use of multiple variables and a nonlinear discriminator that the variance of Zdr becomes more

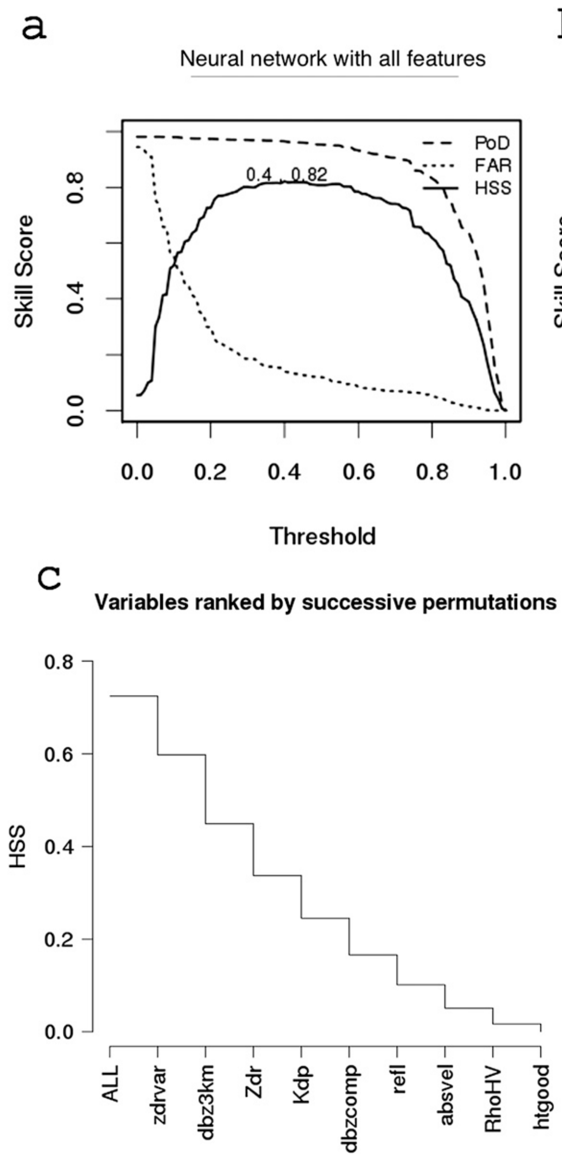

FIG. 9. Less important variables were pruned and a neural network was trained on a reduced feature set. The resulting neural network has a slightly worse performance, but it is more computationally efficient. (a) Performance of original QCNNDP, from Lakshmanan et al. (2014), evaluated on an independent dataset. (b) Performance of QCNNDP operating with a reduced number of inputs. (c) Relative importance of variables on the reduced-feature-set algorithm: HSS vs variables.

important (this is shown in Fig. 6). For example, echoes that are aloft and possess a low variance of Zdr are statistically likely to be weather. For such echoes, which have been correctly classified on the basis of dbz3km and zdrvar, the value of RhoHV is essentially not required. Thus, even though RhoHV is a highly relevant indicator by itself, it can be unimportant when considered as part of a nonlinear combination of discriminator variables.

Third, what explains the dramatic difference in ranking? In the case of closely related variables, a marginally more useful variable may completely mask the less useful one. Therefore, zdrvar does not have to be dramatically better than RhoHV, just slightly better. In general, weather echoes have RhoHV close to unity and a low zdrvar. However, human-made structures (such as towers) and mountains without vegetation can have RhoHV close to unity (indicating weather echoes) but the spatial variance of Zdr in such areas is high (correctly indicating nonweather echoes). Further, in areas of partial beam-filling and behind hail cores, RhoHV values can be very low (erroneously indicating nonweather echoes) but the variance of Zdr would be relatively low (correctly indicating weather echoes). Also, it should be realized that we are looking at a combination of discriminator variables that includes echo heights. The main core of the storm is easily classified by the volumetric part of the algorithm but the fringes of the storm, any outflow boundaries, and nearby insects are not. In these areas of weaker signal, RhoHV has a known tendency to be biased upward due to overcorrection for noise in the

WSR-88D (Ivic and Melnikov 2013). Until this problem is fixed, it can lead to false positive detections of weather data when RhoHV is used. On the other hand, a similar overcorrection of noise to Zdr leaves the variance of Zdr unaffected. Finally, the coding of Zdr and RhoHV values into 8-bit numbers on the WSR-88D works against RhoHV—the range of Zdr (-7 to $+7$ dB) allows for a very wide and resolvable range, whereas useful values of RhoHV get compressed into a much smaller range. Thus, even though the two variables are statistically equivalent in areas of high SNR, the resolution of Zdr makes it more useful as a discriminator. The abovementioned four, admittedly minor, factors lead to the variance of Zdr being slightly better than RhoHV. Once, the variance of Zdr is selected, though, the fact that RhoHV and zdrvar are mostly correlated works against RhoHV—there are few pixels that RhoHV can classify correctly that dbz3km, zdrvar, etc. have not already correctly classified.

4. Summary

Statistical methods can be employed to examine the correlation between the different variables, and principal components analysis (PCA) can help determine which variables contribute most to the information of a dataset. However, such methods do not take into account the context of a particular problem. In this case, we were interested in the variables most important to the problem of discriminating between weather and no-weather echoes. The Kullback–Leibler *J* measure can be used to determine the difference between class-conditional probabilities of each of the variables. Variables ranked by their *J* measure provide an indication of the univariate skill provided by that variable. However, the *J* measure does not take into account the practical consideration that real-world QC methods take advantage of multiple input variables, and that several variables (even if their values are uncorrelated) may tend to classify or misclassify the same set of pixels.

Therefore, the radar data quality control algorithm of Lakshmanan et al. (2014), which was created by optimizing its weights based on a large and diverse training dataset, was used to assess variable importance. Subset retraining methods, such as jackknifing and leave-one-out methods, are the most accurate means of assessing variable importance, but they prove computationally infeasible at assessing the importance of the different radar-observed variables for the purpose of discriminating between weather and no-weather data. However, the permutation approach of Radivojac et al. (2004) by which an already devised QC algorithm is carried out on permuted versions of the training dataset can be used to determine the most important variable in a computationally efficient way. The permutation approach by itself does not take into account redundancy between variables in the context of a specific problem. However, the permutation approach can be carried out in a systematic process in which the important variables found in earlier iterations are held constant and the next most important variable is determined. Using this permutation-rank approach, the most important variables for the purposes of discriminating between weather and no-weather echoes were determined. The ranking of the considered variables is shown in Fig. 6.

In brief, the most important variables are the variance of Zdr, variables relating to the 3D reflectivity structure (reflectivity at 3 km above ground level, height of 0-dB*Z* echo, and maximum reflectivity within the 3D column), and the range derivative of PhiDP (Kdp).

While not the main focus of this paper, it was realized that the variable importance framework introduced in this paper can be used to study the impact of Zdr calibration errors on the performance of the QC algorithm. The effect of Zdr calibration errors on weather/no-weather echo discrimination was found to be relatively negligible.

Acknowledgments. Funding for the authors was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-OU Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. Samantha Berkseth was supported by the National Science Foundation Research Experiences for Undergraduates (REU) Program AGS-1062932. We thank Robert Lee and Jeffrey Cunningham of NOAA's Radar Operations Center for the helpful discussions regarding Zdr biases on the WSR-88D network.

REFERENCES

Bishop, C. M., 1995: *Neural Networks for Pattern Recognition.* Oxford University Press, 482 pp.

Breiman, L., 2001: Random forests. *Mach. Learn.,* **45,** 5–32, doi:10.1023/A:1010933404324.

Chandrasekar, V., R. Kernen, S. Lin, and D. Moisseev, 2013: Recent advances in classification of observations from dual polarization weather radars. *Atmos. Res.,* **119,** 97–111, doi:10.1016/j.atmosres.2011.08.014.

Elmore, K., Z. Flamig, V. Lakshmanan, B. Kaney, V. Farmer, H. D. Reeves, and L. Rothfusz, 2014: mPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.,* **95,** 1335–1342, doi:10.1175/BAMS-D-13-00014.1.

Giangrande, S., R. McGraw, and L. Lei, 2013: An application of linear programming to polarimetric radar differential phase processing. *J. Atmos. Oceanic Technol.,* **30,** 1716–1729, doi:10.1175/JTECH-D-12-00147.1.

Gourley, J. J., P. Tabary, and J. P. du Chatelet, 2007: A fuzzy logic algorithm for the separation of precipitating from non-precipitating echoes using polarimetric radar observations.

J. Atmos. Oceanic Technol., **24,** 1439–1451, doi:10.1175/JTECH2035.1.

Grecu, 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, doi:10.1175/1520-0426(2000)017<0121:AEMFDO>2.0.CO;2.

Hand, D., H. Mannila, and P. Smyth, 2001: *Principles of Data Mining*. MIT Press, 546 pp.

Heidke, P., 1926: Berechnung des Erfolges und der gute der Windstarkvorhersagen im Sturmwarnungsdienst. *Geogr. Ann.,* **8,** 301–349.

Ivic, I. R., and V. Melnikov, 2013: Multi-lag hybrid correlation coefficient estimator. *36th Conf. on Radar Meteorology,* Breckenridge, CO, Amer. Meteor. Soc., 248. [Available online at https://ams.confex.com/ams/36Radar/webprogram/Paper228447.html.]

Jeffreys, H., 1946: An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London,* **186A,** 453–461, doi:10.1098/rspa.1946.0056.

Kessinger, C., S. Ellis, and J. Van Andel, 2003: The radar echo classifier: A fuzzy logic algorithm for the WSR-88D. *Third Conf. on Artificial Intelligence Applications to the Environmental Science,* Long Beach, CA, Amer. Meteor. Soc., P1.6. [Available online at https://ams.confex.com/ams/annual2003/techprogram/paper_54946.htm.]

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. J. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Climatol.,* **46,** 288–305, doi:10.1175/JAM2460.1.

——, J. Zhang, and K. Howard, 2010: A technique to censor biological echoes in radar reflectivity data. *J. Appl. Meteor. Climatol.,* **49,** 435–462, doi:10.1175/2009JAMC2255.1.

——, ——, K. Hondl, and C. Langston, 2012: A statistical approach to mitigating persistent clutter in radar reflectivity data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.,* **5,** 652–662, doi:10.1109/JSTARS.2011.2181828.

——, C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *J. Atmos. Oceanic Technol.,* **31,** 1234–1249, doi:10.1175/JTECH-D-13-00073.1.

Lin, J., 1991: Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory,* **37,** 145–151, doi:10.1109/18.61115.

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., P5.3. [Available online at https://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_47546.htm.]

Melnikov, V., and D. Zrnić, 2004: Simultaneous transmission mode of the polarimetric WSR-88D: Statistical biases and standard deviations of polarimetric variables. NOAA/National Severe Storms Laboratory Tech. Rep., 84 pp. [Available online at https://www.nssl.noaa.gov/publications/wsr88d_reports/SHV_statistics.pdf.]

Park, H., A. Ryzhkov, D. Zrnić, and K.-E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS. *Wea. Forecasting,* **24,** 730–748, doi:10.1175/2008WAF2222205.1.

Radivojac, P., Z. Obradovic, A. K. Dunker, and S. Vucetic, 2004: Feature selection filters based on the permutation test. *Machine Learning: ECML 2004; 15th European Conference on Machine Learning,* J.-F. Boulicaut et al., Eds., Lecture Notes in Computer Science, Vol. 3201, Springer, 334–346.

Ryzhkov, A. V., 2007: The impact of beam broadening on the quality of radar polarimetric data. *J. Atmos. Oceanic Technol.,* **24,** 729–744, doi:10.1175/JTECH2003.1.

——, and D. Zrnić, 2006: Depolarization in ice crystals and its effect on radar polarimetric measurements. *J. Atmos. Oceanic Technol.,* **23,** 1256–1267, doi:10.1175/JTECH2034.1.

——, ——, and R. Fulton, 2000: Areal rainfall estimates using differential phase. *J. Appl. Meteor.,* **39,** 263–268, doi:10.1175/1520-0450(2000)039<0263:AREUDP>2.0.CO;2.

——, S. E. Giangrande, V. M. Melnikov, and T. J. Schuur, 2005: Calibration issues of dual-polarization radar measurements. *J. Atmos. Oceanic Technol.,* **22,** 1138–1155, doi:10.1175/JTECH1772.1.

Shannon, C., 1948: A mathematical theory of communication. *Bell Syst. Tech. J.,* **27,** 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x.

Steiner, M., and J. Smith, 2002: Use of three-dimensional reflectivity structure for automated detection and removal of nonprecipitating echoes in radar data. *J. Atmos. Oceanic Technol.,* **19,** 673–686, doi:10.1175/1520-0426(2002)019<0673:UOTDRS>2.0.CO;2.

Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinf.,* **8,** 25, doi:10.1186/1471-2105-8-25.

Tang, L., J. Zhang, C. Langston, J. Krause, K. Howard, and V. Lakshmanan, 2014: A physically based precipitation–nonprecipitation radar echo classifier using polarimetric and environmental data in a real-time national system. *Wea. Forecasting,* **29,** 1106–1119, doi:10.1175/WAF-D-13-00072.1.

van der Laan, M. J., 2006: Statistical inference for variable importance. *Int. J. Biostat.,* **2,** doi:10.2202/1557-4679.1008.

Wang, Y., and V. Chandrasekar, 2009: Algorithm for estimation of the specific differential phase. *J. Atmos. Oceanic Technol.,* **26,** 2565–2578, doi:10.1175/2009JTECHA1358.1.

Zhang, J., S. Wang, and B. Clarke, 2004: WSR-88D reflectivity quality control using horizontal and vertical reflectivity structure. *11th Conf. on Aviation, Range, and Aerospace,* Hyannis, MA, Amer. Meteor. Soc., P5.4. [Available online at https://ams.confex.com/ams/11aram22sls/techprogram/paper_81858.htm.]