Classifying Proximity Soundings with Self-Organizing Maps toward Improving Supercell and Tornado Forecasting

CHRISTOPHER J. NOWOTARSKI AND ANDERS A. JENSEN
Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 4 December 2012, in final form 27 February 2013)

ABSTRACT

The self-organizing map (SOM) statistical technique is applied to vertical profiles of thermodynamic and kinematic parameters from a Rapid Update Cycle-2 (RUC-2) proximity sounding dataset with the goal of better distinguishing and predicting supercell and tornadic environments. An SOM is a topologically ordered mapping of input data onto a two-dimensional array of nodes that can be used to classify large datasets into meaningful clusters. The relative ability of SOMs derived from each parameter to separate soundings in a way that is useful in discriminating between storm type, location, and time of year is discussed. Sensitivity to SOM configuration is also explored. Simple skill scores are computed for each SOM to evaluate the relative potential of each variable for future development as a method of probabilistic forecasting. It is found that variance in SOM nodes is reduced compared to the overall dataset, indicating that this is a viable classification method. SOMs of profiles of wind-derived variables are more effective in discriminating between storm type than thermodynamic variables. The SOM method also identifies meteorological, geographic, and temporal regimes within the dataset. In general, conditional probabilities of storm-type occurrence generated using SOMs have higher skill when wind-derived variables are considered and when forecasting nonsupercell events. Storm-relative wind variables tend to have better skill than ground-relative wind variables when forecasting non-supercells, whereas ground-relative variables become more important when forecasting tornadoes.

1. Introduction

A wide breadth of work examining supercell thunderstorms has established links between the characteristics of such storms and their environments. Climatological studies, often using proximity sounding data, have unearthed relationships between both kinematic and thermodynamic properties of a storm’s inflow environment and its potential for producing supercells and tornadoes (e.g., Fawbush and Miller 1952; Darkow 1969; Darkow and McCann 1977; Maddox 1976; Brooks et al. 1994; Rasmussen and Blanchard 1998; Thompson et al. 2003; Craven and Brooks 2004; Davies 2004). Two assumptions underlying the use of proximity sounding data are that each sounding is adequately representative of a storm’s inflow environment and that this environment is homogeneous. Spatial variations in the storm environment, such as outflow boundaries, drylines, fronts, and elevation changes have been shown to influence storm properties (e.g., Atkins et al. 1999; Rasmussen et al. 2000; Ziegler et al. 2010; Markowski and Dotzek 2011; Houston and Wilhelmson 2012); however, it is often difficult to predict their storm-relative location and assess their influence in advance. Furthermore, such features may only be of importance if the background environment, as diagnosed by proximity soundings, does not preclude the possibility of supercells or tornadoes. Proximity soundings obtained from radiosonde observations or from global and regional forecast models are readily available, and forecasters continue to rely on these data when issuing probability forecasts of severe thunderstorms and tornadoes (e.g., convective outlooks and thunderstorm/tornado watches).

The most common forecast parameters derived from proximity soundings generally use one or more of the available thermodynamic and wind shear variables to assess the potential of the storm environment for producing severe weather. A variety of useful parameters such as the significant tornado parameter (STP) and supercell composite parameter (SCP) have been developed and implemented operationally (Thompson...
et al. 2002). The utility of these composite parameters depends on the depth or location of the vertical layer that is considered and the phenomena being forecast. For instance, Craven and Brooks (2004) have shown that using the 0–1-km depth is effective in discriminating between nontornadic and tornadic environments, whereas Houston et al. (2008) found the bulk wind difference over the 0–5-km layer to be most effective for discriminating between nonsupercell and supercell environments. In an effort to ameliorate the “one size fits all” nature of selecting arbitrary levels to evaluate a vertical profile, Thompson et al. (2007) evaluated shear and helicity over the “effective inflow layer.”

Considering the ambiguous and somewhat arbitrary nature of condensing a vertical profile into a point value, important information about the shape of the vertical profile may be lost when relying on composite parameters. Composite profiles have been used in the past to examine this problem (Maddox 1976; Kerr and Darkow 1996; Markowski et al. 2003). However, this technique often relies on classification by storm type (nonsupercell, nontornadic supercell, tornadic supercell, etc.) or tornado intensity (Fujita scale) before averaging variables. Therefore, if multiple regimes exist for the same storm type, they may be “averaged out.” Jensen et al. (2012) applied the self-organizing map (SOM) statistical technique (Kohonen 1995) to ozonesonde profiles. SOMs use a learning algorithm to cluster profiles into a specified number of nodes, the topology of which is determined by the algorithm, not the user. When applied to vertical profiles of relevant supercell and tornado environmental variables, storm environments may be classified and analyzed considering both the magnitude and shape of the vertical profile of each variable, eliminating the need for preclassification by storm type. The usefulness of an SOM depends on the relative percentage of the storm types in each node. If enough nodes are used in the SOM, multiple regimes in the vertical profile of a sounding-derived variable for each storm type may become apparent.

In this study, we utilize SOMs to cluster two-dimensional profiles of different thermodynamic and kinematic variables applicable to supercell and tornado prediction to analyze 1185 severe storm environments derived from a Rapid Update Cycle-2 (RUC-2; Benjamin et al. 2004) proximity sounding dataset (Thompson et al. 2003, hereafter T03; Thompson et al. 2007, hereafter T07). By comparing the percentages of the resultant storm types [nonsupercell (NS), nontornadic (NT) supercell, weakly tornadic (WT) supercell, and significantly tornadic (ST) supercell] of the profiles that are classified into each SOM node with the percentage of each storm type in the entire dataset, we are able to assess the relative utility of each variable in discerning supercell and tornado environments. Furthermore, varying the number of nodes in each SOM and the depth of the layer used to create the SOM allows us to find the configuration that maximizes the ability of the SOM to discriminate between storm types. The SOMs are also capable of discriminating between multiple regimes of variables for each storm type. When compared with composite hodographs, common bulk parameters, date, time, and location for their member events, the SOM nodes contain information on the larger weather patterns that lead to superfics and tornadoes.

Section 2 describes the proximity sounding dataset, the variables derived from these soundings, interpolation methods, and the SOM analysis method. Results of the SOM technique and sensitivities of these results to the SOM configuration for different sounding-derived variables are presented in section 3. Conclusions and the potential for future application of the SOMs in severe weather prediction are discussed in section 4.

2. Data and methodology

a. Proximity sounding data

The self-organizing maps in this study were created using the RUC-2 proximity sounding dataset collected from April 1999 through June 2001 by T03, with additional data collected from January 2003 through March 2005 (T07). Overall, there are 1185 proximity soundings in the contiguous United States (Fig. 1) of which there are 250 from discrete NS environments, 101 from marginal (MR) supercell environments, 443 from NT supercell environments, 278 from WT [tornadoes rated as category 0 or 1 on the Fujita scale (F0 and F1)] supercell environments, and 113 from ST (F2 or stronger tornadoes) supercell environments.

The data were collected and quality controlled as in T07 with only surface-based, discrete, right-moving, cyclonic supercells considered. The soundings are derived from hourly RUC-2 analyses and either interpolated to the nearest surface observing site (1999–2001 data) or taken from a point forecast (2003–05 data), both within 30 min and 40 km upstream of the radar-identified storm. Discrete storms were identified using Weather Surveillance Radar-1988 Doppler (WSR-88D) reflectivity and velocity data. To be considered a supercell, a storm was required to display reflectivity characteristics (a hook echo or inflow notch) and must have displayed at least 0.002 s$^{-1}$ cyclonic azimuthal shear in the 1-km resolution velocity data for at least 30 min. In this study, storms that were marginally supercellular (having peak cyclonic shear less than 0.002 s$^{-1}$) were considered in the NS category for
the sake of clarity, and the results do not change significantly when they are considered as an independent category. When multiple supercells occurred within 185 km of each other over a 3-h period, only a sounding representative of the most intense supercell’s inflow was added to the database. As discussed in Houston et al. (2008), these soundings are more correctly interpreted as representative of events rather than individual storms. For more details on the methodology concerning the collection and filtering of the dataset and associated errors, the reader is referred to T03 and T07.

The distribution of each storm type in the entire dataset is shown in Fig. 2a, and the time of day and day of year of each sounding is shown in Fig. 2b. Though events from nearly all times of day and seasons are considered, the majority occur in the late afternoon/early evening hours and during the late spring and early summer months. As in T03 and T07, soundings with zero surface-based convective available potential energy (SBCAPE) are excluded in an effort to eliminate “elevated” supercells (Colman 1990) from the database. However, a fair number of soundings with large surface-based convective inhibition (SBCIN) still remain in the dataset, potentially allowing storms that are somewhat decoupled from the surface.

b. Variables considered

The sounding data were interpolated to 100-m increment height levels above ground level (AGL) from 0 to 16 km before 21 variables relevant to supercell and tornado formation were computed at each height. Thermodynamic variables include potential temperature \( \theta \), equivalent potential temperature \( \theta_e \), \( d\theta/dz \), \( d\theta_e/dz \), Brunt–Väisälä frequency \( N^2 \), and relative humidity.

Fifteen variables were also derived from the RUC-2 wind profiles and observed storm motions. These include ground-relative (GR) and storm-relative (SR) wind speed and direction \( \mathbf{v}_e = \mathbf{v} - \mathbf{c} \), where \( \mathbf{v} \) is the horizontal wind vector and \( \mathbf{c} \) is the observed storm motion vector, individual GR and SR wind vector components \( (u, v, u_e, v_e) \), total vertical wind shear \( dv/dz \), and vector component vertical wind shear \( (du/dz, dv/dz) \). Profiles of the streamwise vorticity \( \omega_s = -|\mathbf{v}_e| \phi/dz \) and crosswise vorticity \( \omega_c = |\mathbf{v}_e| \phi/dz \), where \( \phi \) is the orientation angle of the storm relative wind (see Markowski et al. 2003, their Fig. 1), as well as ground-relative helicity density \( \text{GRHD} = \mathbf{v} \cdot \omega \) and storm-relative helicity density \( \text{SRHD} = \mathbf{v}_e \cdot \omega \), are computed following the methodology of Markowski et al. (2003). These vertical profiles were used as the input data for individual SOMs of each variable.

c. SOM method

An SOM is a topologically ordered mapping of input data onto a two-dimensional array of nodes (Kohonen 1995, page 77). SOMs begin with a user-defined map size; the size of the map being the number of nodes to create (e.g., a \( 3 \times 3 \) SOM has 9 nodes). The nodes of an SOM are first initialized to be representative of the input data. For instance, if the input data are profiles of a variable \( (x_1, x_2, x_3, \ldots, x_n) \) at heights \( (h_1, h_2, h_3, \ldots, h_n) \), then each SOM node is set to some unique initial value \( (r_1, r_2, r_3, \ldots, r_n) \) at heights \( (h_1, h_2, h_3, \ldots, h_n) \). The initialized SOM nodes are also known as reference vectors and each reference vector proceeds to learn iteratively from the input data. In general, each vector of input data is chosen stochastically and compared to each node.
using Euclidean distance, where the Euclidean distance between an input data vector and a node reference vector is \( \sqrt{(x_1 - r_1)^2 + (x_2 - r_2)^2 + \cdots + (x_n - r_n)^2} \). The node that is closest (smallest Euclidean distance) to the input vector is known as the best-matching unit (BMU). The nodes then become more like the input data where the degree of learning depends on how far a node is from the current BMU, with closer nodes learning more from the data. By varying the degree to which each node learns from the input data depending on the distance from the BMU, the SOM becomes topologically organized with similar nodes being adjacent to each other. Figure 3 conceptually illustrates the SOM competitive learning process for one input data vector. To improve computational efficiency, multiple comparisons and adjustments with different input data are performed simultaneously. After the nodes have learned from each input vector, the process is reiterated until the nodes converge to a final map. For more information on the implementation of the SOM method, see Vesanto et al. (2000).

One benefit of SOMs is that we only specify a number of nodes and make no prior assumptions about what the final map will look like. The entire input dataset can be organized objectively by using SOMs because the SOM only knows values of the input data at different heights and nothing about storm type. We employ \( 3 \times 3 \) SOMs because we have four categories (NS, NT, WT, ST); therefore, 9 SOM nodes are enough to identify different regimes for each storm type without having so many nodes that data visualization becomes difficult. We also utilized \( 2 \times 2 \) and \( 4 \times 4 \) SOMs to test the sensitivity of the SOM to the number of nodes. We create an SOM for each different variable up to 1, 3, 6, and 16 km AGL in order to understand how variables will cluster differently considering only the boundary layer, within and above the boundary layer, up to midlevels, and for the entire profile depth for different storm types (sensitivity to the height considered and number of nodes is discussed in section 3c). While the SOM nodes (reference vectors) provide useful information, SOM nodes are not real data; they learn from real data. Therefore, the SOM nodes are not plotted. Each input data profile is clustered according to its BMU, and the mean and standard deviation for each of these clusters is plotted. In this way the actual data are visualized.

3. Results

a. Evaluating SOM performance

To be considered a useful analysis technique, the self-organizing map of any variable must function both as a classification scheme for the profiles and as a method for discriminating between (and therefore predicting) the resulting storm type.

1) VARIANCE REDUCTION

As a classification technique, the primary goal of the SOM is to explain the variance in the dataset by grouping similar profiles into clusters described by their BMU. Thus, in a successful SOM, the variance in the subset of profiles matching a particular node should be smaller than the variance in the entire sample. For each

---

1 The mean profile in each cluster is very similar to, though not exactly the same as, the SOM node.
cluster in the SOM, we compute the height average of the variance at each level over the depth of the profile considered in the SOM. We then compare the average variance of all the clusters in an SOM with the height-averaged variance of the entire sample. Table 1 displays the percent reduction in the average variance of selected SOMs relative to the overall dataset [variance is only listed for those SOMs that had the five highest percentage difference spreads for each type; see section 3a(2)]. SOMs, on average, reduce the variance by 54.9%. The SOM for over the lowest 1 km is best in reducing variance (92.8% reduction), while the SOM of GRHD over the lowest 6 km is the worst (only 11.7% reduction). In general, a larger portion of the variance in the dataset is reduced by applying the SOM method to the lowest 1 km of the profiles rather than the lowest 3 or 6 km. The best-performing SOMs, in terms of their ability to classify variable profiles, are those where the largest difference in the profiles is due to the mean magnitude of the variable over the depth the SOM considers (e.g., , wind speed) rather than vertical gradients or the shape of the profile (e.g., , shear variables). In some cases, such as potential temperature or wind speed, vertical gradients may be more important in determining storm type. Thus, the SOMs of most interest for the purpose of discriminating between storm types may not necessarily be those that result in the most variance reduction.

FIG. 3. Conceptual diagram illustrating the SOM competitive learning process. Two nodes, or reference vectors (black profiles), of an SOM with N nodes are shown (top) before and (from middle to bottom) after learning. For a given input data vector (dashed gray line), each node is adjusted toward the data vector (center). The node initially closest (Euclidean distance over the entire profile depth) to the data vector is the BMU. The BMU is adjusted more than other nodes because of its closer proximity to the input vector. This is done for each input data vector and the entire process is repeated many times. After learning, the data vector is clustered with its BMU.
2) PERCENTAGE DIFFERENCE

The SOM is only a useful forecasting tool if the resulting classification of data also distinguishes between environments for different storm types. For instance, if the SOM reduces variance considerably but the percent of profiles in each cluster resulting in a particular storm type is the same as in the whole dataset, then the SOM is useless in predicting storm type. To assess the skill of each cluster of an SOM in differentiating between environments favorable or unfavorable for a particular storm type, we compare the percentage of profiles in each cluster matching that storm type with the percentage in the entire sample. If the percentage difference (\( \frac{\text{percentage of storm type in cluster} - \text{percentage of storm type in dataset}}{\text{percentage of storm type in dataset}} \)) is especially positive (negative) in any cluster, then a profile the SOM places in that cluster is more (less) likely to produce the storm type in question. In other words, if a cluster has a large, positive (negative) percentage difference in the number of ST profiles, then that cluster is a good (poor) representation of ST environments.

Once the percentage difference is acquired for each cluster, we evaluate the predictive ability of the SOM for each storm type by calculating the “spread” (\( \text{maximum percentage difference minus minimum percentage difference} \)).

\[ \text{spread} = \text{max} - \text{min} \]

---

2 Only clusters containing at least 5% of the dataset (60 profiles) are considered because the percentage difference may be skewed for smaller clusters.
The clusters reveal patterns in the shape of the wind speed profiles (i.e., the location of speed shear). Such distinctions may be useful in delineating between environments of particular storm types. Figure 4 shows that profiles in nodes 2 and 3 have similar values of 0–6-km speed shear, but it is evident that node 3 has more shear than node 2 in the lowest 1 km. Consequently, node 3 has a higher percentage of ST than node 2, which is dominated by NT supercells (Fig. 5). This difference is also reflected in the composite hodographs of each cluster (Fig. 6), which show similar 0–6-km shear but more curvature (i.e., storm-relative helicity) in the lowest 1 km of node 3. The composite sounding for both nodes is associated with an SCP consistent with supercells (Fig. 6), yet node 3 (SCP = 7.0) has a higher value than node 2 (SCP = 2.9). STP in node 3 (STP = 1.6) is more than twice that of node 2 (STP = 0.6), consistent with the higher frequency of ST supercells in that cluster. Figures 7 and 8 show the location and time-date, respectively, of the profiles in each cluster. Though most storms in these nodes occur in the late afternoon and evening hours, those events matching node 3 appear to occur over a more narrow geographic area (confined to the eastern Great Plains, and the Mississippi and Ohio River valleys) than node 2. Furthermore, node 3 events are generally limited to late winter, spring, and fall months, whereas node 2 events also occur during the summer months.

Profiles matched to nodes 3 and 6 are the most likely to produce ST events (Figs. 4 and 5). However, despite being rarer, profiles in node 3 reflect environments that are more likely to produce ST supercells. This distinction is interesting, considering the identical STP values for both nodes (Fig. 6). The average SBCAPE for profiles in node 6 is nearly 500 J kg⁻¹ higher than in node 3, perhaps due to the fact that more events in this node occur later in the year (more events between May and September than node 3). Furthermore, the average

difference) for each storm type. SOMs with a large (small) spread for a given storm type are deemed to be good (bad) at distinguishing profiles associated with that specific storm type. If the average spread for all storm types is large (small), then the SOM is judged to be effective (ineffective) in distinguishing between all storm types.

Table 1 lists the variables with the largest spread of percentage difference overall and for individual storm types. It should be noted that good performance by an SOM does not necessarily indicate that variable is more important in determining storm characteristics and evolution, but rather that it is better suited for prediction via the SOM method. In general, the wind-derived variables are better at discriminating between storm-type than thermodynamic variables (based on average spread). Of all of the SOMs, SR and GR wind speeds at 0–6 km are most effective in distinguishing between storm types overall and have high spreads for each individual storm type. SOMs at 0–6 km are generally better at clustering by storm type than their 0–1- and 0–3-km counterparts, though 0–1- and 0–3-km SOMs tend to become more important in discriminating profiles associated with tornadic storms. With the exception of 0–1-km streamwise vorticity, which is ranked highest in distinguishing ST supercell environments, simple profiles of wind magnitude and component strength tend to be more effective than more complicated kinematic variables such as vorticity and helicity density. The vertical gradient of the potential temperature (related to static stability) is the most skillful of any of the thermodynamic variables and does particularly well in discriminating NT profiles. Particularly poor discriminators of storm type include \( \theta_e \), wind direction, relative humidity, and \( \omega_c \) (not shown).

b. SOM attributes

1) 0–6-KM GROUND-RELATIVE WIND SPEED SOM

By the aforementioned measures, the \( 3 \times 3 \) SOM of GR wind speed \( |v| \) over the lowest 6 km performs well in clustering profiles and discriminating between storm types. For the sake of brevity, we demonstrate the potential of only one SOM in classifying profiles of only one variable into clusters and describe the qualitative skill of these regimes in discriminating between storm types. However, similarly illuminating results can be drawn from many of the other SOMs we produced.

Figure 4 shows clusters of GR wind speed profiles matching each node of the SOM with their average and standard deviation. Due to the topology of this SOM, the nodes are organized such that profiles with the lowest near-ground speeds are in clusters on the left (nodes 1, 4, and 7), with generally increasing low-level wind speed toward the rightmost SOM nodes. Wind speed at 6 km increases in nodes from bottom to top. As a result, those profiles with the most 0–6-km wind shear are in the top right, while those with the least are in the bottom left. Figure 5 shows the distribution of storm types in each node (black bars) and the difference (gray bars) in the percentage of each type in the node from its percentage in the entire dataset (Fig. 2a). In general, nodes with less (more) wind speed shear in the bottom left (top right) are associated with a greater frequency of NS (ST) events. A transition of NS–NT–WT–ST occurs between these two extremes. Average SBCAPE and SBCIN do not appear to have a consistent trend related to wind speed or shear (Fig. 6). However, the nodes with higher overall wind speed (and shear) tend to have lower lifting condensation levels (LCLs), indicating a relationship between wind and LCL.
SBCIN of node 6 is higher than node 3, which is expected given the increased number of events that occur in the overnight and early morning hours (0300–1200 UTC) relative to node 3. The geographic distribution of events in node 6 also includes more events in the central and northern Great Plains, indicating the climatological shift in tornado frequency toward the northwest during the late spring and summer months (Brooks et al. 2003).

Hodographs from two profiles in the dataset corresponding to two prolific tornado outbreaks over the observed period have been plotted in Fig. 6. In node 3, the ST event RUC hodograph from 2000 UTC 4 May 2003 near Independence, Kansas, has been plotted. On this day alone (part of a larger outbreak sequence from 3 to 11 May 2003) there were 81 confirmed tornadoes (25 were significant, 4 were rated F4) in the central United States resulting in 38 fatalities, with many tornadoes in eastern Kansas and western Missouri (Schneider et al. 2004). As expected, this hodograph is similar to the composite hodograph for the cluster it matches (node 3). Though the hodograph is shifted slightly eastward (toward the more positive u direction) the magnitude of the wind speed over the lowest 6 km, which is the only factor taken into consideration in this SOM, is very similar. The overall character of the event is a good match with this cluster considering matches to this node are rare but the probability of ST events is much higher in this node than in the entire dataset.

The second RUC hodograph matches node 6 and was taken at 2300 UTC 3 May 1999 near Oklahoma City, Oklahoma (Fig. 6). This day was part of another historic tornado outbreak, and the sounding is representative of the environment of a long-track F5 tornado that passed through the Oklahoma City metropolitan area causing 36 fatalities (Brooks and Doswell 2002). In this case, the hodograph shape below 6 km (the only levels considered by this SOM) compares favorably with the composite hodograph for the cluster to which it is matched. Though within one standard deviation of the composite hodograph, the ground-relative wind speeds in this hodograph are stronger over the lowest 3 km than the composite. Thus, this particular hodograph still resulted
in a ST event despite the slightly lower ST probability in this node than node 3. It is interesting to note that in this case, though the wind speed is generally stronger than the average for node 6 over the depth considered by the SOM, the shape of the profile still caused this event to be matched to that node.

Another profile shape pattern evident in the SOM is a relative maximum in wind speed below 1 km in nodes 4, 5, and 8 (Fig. 4). The jet in node 4 is located at 500 m, with wind speed decreasing from 5.4 to 4.8 m s\(^{-1}\) at 1100 m (before increasing again farther aloft). However, the jet is more pronounced and slightly higher in node 5 (winds decreasing from a jet of 10.4 m s\(^{-1}\) at 800 m to 9.6 m s\(^{-1}\) at 1700 m). Though both clusters represent regimes that have similar geographic locations (Fig. 7) and times–dates (Fig. 8), it appears that the intensified low-level jet in node 5 is more likely to result in WT than in node 4, which has a higher probability of NS storms (Fig. 5). Node 8 also has a weak low-level jet at 800 m but the lack of deep-layer shear suggests storms matching this node are even more likely to be NS than in node 4.

2) 0–1-KM STREAMWISE VORTICITY SOM

The 0–1-km streamwise vorticity SOM has the highest spread in percentage difference for ST events. Figure 9 shows node 7 (panels a–c) and node 9 (panels d–f) from this SOM. Node 9 has positive \(v_s\) over the lowest 1 km, especially below 200 m (Fig. 9d). In this case, the SOM does a particularly good job of identifying a node with profiles that are likely to produce a tornado (77.8% of all profiles in this node are associated with a WT or ST supercell). This node also has the highest frequency of ST supercells of any node in any SOM we tested (36.1%). This is a percentage difference of 27% over the frequency in the entire dataset (Fig. 9e). Though matching a profile to this SOM node is rare (only 72 of 1185 profiles match this node), the distribution suggests if such a profile of \(v_s\) occurs with a discrete thunderstorm, it will have a comparatively high probability of resulting in

![Fig. 5. The number of each storm type matching each node (black bars) and the percentage difference of the number of each storm type matching each node relative to the total number of each storm type (gray bars). RPSS and BSS for each storm type are also listed for each node.](image-url)
a tornado. The composite hodograph, moderate average SBCAPE, low-average LCL, and composite STP of 1.3 for this cluster are all consistent with environmental parameters associated with ST events (Fig. 9f).

This SOM also reveals subtle differences between NT and ST supercell environments. Though smaller than in node 9, the composite STP for node 7 is still relatively large (0.9) and the hodograph exhibits strong curvature over the lowest 3 km (Fig. 9c). However, this cluster contains profiles with near-zero $\omega_s$ (Fig. 9a) and large $\omega_c$ (not shown) in the lowest 300 m, which is manifest in the lowest segment of the hodograph where the shear is roughly parallel to the low-level storm-relative winds. This results in a 33% higher probability of NT than node 9 and an ST supercell frequency only 2% higher than in the entire dataset (Fig. 9b). The classification in this SOM suggests that $\omega_s$ in the lowest few hundred meters may be critical to the development of significant tornadoes. Furthermore, such a distinction may not be fully recognized by a forecast parameter like STP.

3) 0–6-KM $d\theta/dz$ SOM

Although SOMs of wind-derived variables generally discriminate between storm types best, SOMs of thermodynamic variables may contain meaningful information. When only wind information is considered, some profiles that are limited by thermodynamic characteristics such as strong surface static stability may be wrongly assumed to have a conditionally high probability of producing a tornado. Figure 10 shows two nodes from the 0–6-km $d\theta/dz$ SOM. Based on only the composite hodographs of these clusters, one might assume that the profiles in node 9 (Fig. 10d), which have larger 0–1-km SR helicity than those in node 5 (Fig. 10b), are more likely to be tornadic.
However, $d\theta/dz$ in node 9 is more positive in the lowest 1 km than in node 5 (Figs. 10a,e). This indicates strong near-surface stability for the profiles in node 9, which is evident in their higher average SBCIN value (Figs. 10b,f). Consequently, despite a “better” hodograph, node 9 is less likely to produce tornadoes than node 5 (Figs. 10c,g) and is dominated by NT supercells. In contrast to node 5, many of the events in node 9 occur in the overnight hours (Figs. 10d,h). Though our methodology has eliminated storms with zero SBCAPE, those profiles with nonzero SBCAPE but strong SBCIN may still correspond to quasi-elevated storms. It is possible node 9 with its strong 0–1-km SRH, low-average LCL, and strong SBCIN is composed of many nocturnal low-level jet scenarios wherein strong surface stability inhibits tornadoes despite a favorable wind profile.

c. SOM sensitivity to number of nodes and height

1) SENSITIVITY TO THE NUMBER OF NODES

One valuable feature of self-organizing maps is that the number of nodes is user defined. Including more nodes typically allows for more subtle data features to be seen. However, using too many nodes makes data visualization difficult and leaves more nodes with too few profiles clustered to that node, making the statistics of those nodes unreliable. Reducing the number of nodes means that each node is a BMU for more profiles such that variations between the nodes result from more pronounced data features. Figure 11 shows the $2 \times 2$ SOM for 0–6-km GR wind speed, which displays similar topology to the same SOM with 9 nodes ($3 \times 3$; Fig. 4). The $2 \times 2$ SOM is broken up into one node that is mostly nonsupercell environments (node 3) and three that are predominately supercell environments (nodes 1, 2, and 4). Nodes to the right have higher 0–1-km wind shears, whereas nodes above have higher 0–6-km wind shears. Node 1 includes 418 profiles and roughly matches the distribution percentages of the entire dataset, which is expected because of the large number of profiles matching that node. This SOM also includes a node with an increased percentage of ST (node 4), meaning the data feature that results in ST supercells (ostensibly strong 0–1-km speed shear) is still prominent enough to be identified with only four nodes. Going from a $3 \times 3$ to a $2 \times 2$ SOM for 0–6-km GR wind speed leads to a decrease in the largest percentage of NS storms (Fig. 4, node 7; Fig. 11, node 3). Having more nodes allows for easier classification of NS-type profiles in this case due to a node with almost no wind shear (Fig. 4, node 7).
Generally, adding more nodes increases the percentage of each type of storm in the node most favorable for that storm type. However, when a $4 \times 4$ SOM (not shown) is used, some nodes are often redundant (i.e., they are so similar that it is not clear they represent different regimes) or the number of matching profiles is often below our threshold of statistical significance (5% of the total number of profiles).

2) **SENSITIVITY TO THE DEPTH OF THE PROFILES**

Self-organizing maps show the most prominent features of a dataset, which makes changing the SOM depth important. Because wind speed near the tropopause can be an order of magnitude higher than surface wind speed, we create SOMs over the lowest 1, 3, and 6 km to eliminate the influence of the jet stream. We also create a 16-km-deep SOM to see how it compares with SOMs over lower heights. Figure 12 shows a $2 \times 2$ SOM for 0–1-km GR wind speed. Increasing 1-km wind shear corresponds to an increase of tornadic potential; node 2 shows the highest percentage of both WT and ST supercells. Progressing the depth of the $2 \times 2$ GR wind SOM from 1 to 16 km displays a continual increase in the percentage of NS storms in the most favorable node for that storm type (Fig. 12, node 1; Fig. 13, node 3; Fig. 11, node 3; and Fig. 14, node 1). This percentage of NS storms increases overall from 56.3% to 82.4%, which means diagnosing NS storm environments becomes statistically easier using the entire 0–16-km profiles of GR wind speed. Roughly the opposite is true for ST supercells. While the highest percentage of ST supercells is node 4 in the 0–3-km GR wind SOM, increasing SOM height from here leads to a lower percentage of ST supercells in the ST supercell node. Figure 13, which displays the 0–16-km GR wind SOM, reveals one of the important features in the data is the wind maximum near 11 km. While the SOM still does a good job of classifying profiles over this depth, this SOM generally provides less useful information for distinguishing storm type than do other SOMs.
d. Potential forecasting skill

We stress the subsequent analysis is not intended as a verification of forecast performance, but rather uses forecast skill scores as a means of investigating the potential predictive ability of each SOM. This method is unrealistic in that it assumes a perfectly reliable forecast by training our SOM with the same data used in the verification; however, the value of this methodology is in assessing the relative skill of each SOM in an idealized forecasting scenario.

1) RANK PROBABILITY SKILL SCORES

To assess the predictive capabilities of each SOM, we assume a forecaster may use the percentages of storm type in an individual cluster as the basis for issuing probabilities of those occurring for a profile that matches that cluster (on the condition that a discrete storm forms). For example, if a given SOM cluster is composed of 100 profiles associated with 5 NS, 25 NT, 30 WT, and 40 ST storms, then we assume that for any profile matched to that SOM node the forecaster would issue 0.05 NS, 0.25 NT, 0.30 WT, and 0.40 ST probability forecasts. By in turn defining each profile as a forecast event with conditional forecast probabilities issued as such, we can calculate a rank probability score (RPS; Wilks 2006) for each profile. It should be noted that this method assumes an ordering of classes such that the range of environmental characteristics supportive of WT supercells falls between ST supercells and NT supercells, whereas NT environments fall between WT supercell and NS storm environments. We then compute an average RPS for each cluster as well as the entire SOM. Comparing the average RPS for each SOM with the average RPS using the probability density function (PDF) of the entire, unclassified dataset (Fig. 2a) as the forecast probability yields the ranked probability skill score (RPSS). An RPSS of 1 indicates perfect skill, an RPSS of 0 indicates no skill relative to the overall dataset, and a negative score indicates worse skill than the overall dataset.

The average RPSS for each node shows that every SOM we tested has better forecast skill than the PDF of the entire, unclassified dataset. The best-predicting SOMs are those for SRHD (RPSS = 0.29), jv (RPSS = 0.29), and sr (RPSS = 0.28) over the lowest 6 km of the profiles. The worst-predicting SOMs were those of ue over the lowest 3 km (RPSS = 0.02) and 1 km (RPSS = 0.02). In general, SOMs based on information derived from the wind profile outperform those that use thermodynamic variables. Furthermore, SOMs of SR wind variables have a tendency to outperform their GR wind counterparts. The best thermodynamic SOM is du/dz over the lowest 6 km (RPSS = 0.11). The mean RPSS of all thermodynamic SOMs is 0.07, while the mean RPSS for all wind-derived SOMs is 0.18. When all storm types are considered, SOMs over the lowest 6 km generally have slightly more skill (the mean RPSS of all 0–6 km
SOMs is 0.16) than over the lowest 3 km (mean RPSS = 0.15) or lowest 1 km (mean RPSS = 0.13).

Surprisingly, SOMs that consider vertical profiles of wind speed (both SR and GR) are more effective than those that only consider shear. However, profiles of wind speed with height contain information about both speed and vertical speed shear, whereas profiles of wind shear contain only information about shear. This suggests that, at least in the application of the SOM technique, the average wind speed over the depth or surface wind speed (in addition to shear) may be an important consideration in discriminating between storm types.

It should be noted that the foregoing analysis focuses on average skill over all nodes in an SOM. Some individual nodes often show poor skill, especially when the PDF of storm types in that cluster is similar to that of the dataset as a whole. The best RPSSs for individual nodes occur when the distribution of storm type in each node differs the most from the climatology of the entire dataset. The majority of underperforming nodes are found in thermodynamic SOMs or those that only consider wind direction with height. An example of RPSS for each node of an SOM is shown in Fig. 5. The RPSS for the SOMs with the best overall and individual storm type percentage difference spreads is also listed in Table 1.

2) BRIER SKILL SCORES

Whereas the RPSS is a useful tool for characterizing the overall capability of an SOM in providing probability guidance for all four storm types, the Brier skill score (BSS; Wilks 2006) is useful in assessing the predictive capacity of the SOM for individual storm types. Similar to the method for calculating RPSS, we assume that each profile represents an individual forecast and the forecast probability vector is generated from the PDF of the cluster in the SOM to which that profile is matched. However, the probability vector for the BSS is based on a dichotomous event (i.e., either the storm type in question occurs or it does not). Like RPSS, the reference score is generated using the PDF of the entire dataset.

In general, SOMs are less capable of discriminating supercell events (NT BSS = 0.05, WT BSS = 0.05, ST BSS = 0.06) than nonsupercell events (NS BSS = 0.23). The best predictors for NS storms are $|v|_{\text{SR}}$ over the lowest 6 km (BSS = 0.53), 3 km (BSS = 0.50), and 1 km (BSS = 0.50). Similarly, the best predictors for NT supercells are also SOMs of $|v|_{\text{SR}}$ over the lowest 6 km (BSS = 0.14), 3 km (BSS = 0.12), and 1 km (BSS = 0.12). SRHD over the lowest 6 km (BSS = 0.10) and 3 km (BSS = 0.09), $v$ over the lowest 3 km and 1 km, and $\theta_e$ over the lowest 6 km (BSS = 0.09) are the best SOMs for WT supercells. Interestingly, GR wind variables seem more useful in predicting ST supercells than their SR counterparts, with the best ST supercell SOMs being $|v|$ over the lowest 3 km (BSS = 0.14) and 6 km (BSS=0.14), as well as GRHD over the lowest 1 km (BSS = 0.12). The BSS for the SOMs with the best individual storm type percentage difference spreads is also listed in Table 1.

When BSS is computed for individual clusters of an SOM, BSS improves for each storm type in each node...
the more the frequency of that type differs from the frequency in the entire dataset. Some clusters yield a near-perfect or perfect (BSS = 1) score. This occurs when a cluster of profiles contains none that result in the specified storm type. When this occurs, both the probability forecast for that event and the observation are 0, resulting in a perfect score (see Fig. 5, node 7). Similarly, when the percentage of storm type in each cluster matches the frequency in the entire dataset, the BSS for that storm type in that node is 0 (see Fig. 5, node 2). In many SOMs, one cluster is dominated by NS storms, resulting in a high probability forecast of such events (see Fig. 5, node 7). Presumably, this is due to the more stark contrasts between nonsupercell environments and supercell environments than the distinction between tornadic and nontornadic supercell environments. This physical difference is reflected in the higher skill scores for NS events.

4. Summary and conclusions

Self-organizing maps were used to classify profiles of 21 relevant thermodynamic and wind-derived variables from RUC-2 proximity soundings. Once sorted, the percentages of profiles matching a given node that resulted in NS, NT, WT, and ST storm types were compared with the percentages of each storm type in the entire dataset. In this way, we were able to determine which SOMs were most effective in discriminating between storm types. We also examined the potential for practical use of the SOM technique by using the storm-type PDF of each node as a means of issuing a conditional probability forecast for profiles fitting that node.

The results indicate that the SOM method has potential uses as both a classification technique for research purposes and as a means of generating conditional probability forecast products. In general, SOMs of wind-derived variables were better at identifying nodes or regimes linked to each storm type, whereas SOMs of thermodynamic variables were less useful. In many cases, we found that simpler variables such as wind speed and shear magnitude were more useful than more complicated variables like helicity density or streamwise vorticity in discerning regimes corresponding to individual storm types. These results are consistent with those of previous studies [most recently Grams et al. (2012)]. Though their study included
nondiscrete forms of convection and did not consider the entire shape of variable profiles, they found that Heidke skill scores showed kinematic parameters to better discriminate between tornadic and nontornadic severe storms than thermodynamic parameters (with 500-hPa wind speed scoring best overall). Furthermore, we found there seems to be little advantage to using storm-relative quantities over ground-relative quantities to predict storm type. Though slightly less effective for all storm types, ground-relative winds outperformed storm-relative winds in identifying significant tornado environments. Similarly, Markowski et al. (2003) found differences between storm-type composite profiles of ground-relative wind speed to be larger than their storm-relative counterparts.

Overall, the best SOM in terms of storm-type discrimination was the $3 \times 3$ SOM for 0–6-km GR wind speed. This SOM and others illustrate the skill of this technique in sorting profiles into nodes corresponding to different storm types, while also identifying geographic and temporal regimes of potential interest to the research community. Specific regimes related to large-scale flow such as the presence of a nocturnal low-level jet or the positions of storms relative to an upper-air trough (evident in SOMs of $u$- and $v$-wind components) are also often represented in individual SOM nodes.

SOMs were sensitive to the number of nodes and the height over which profile data were considered. Whereas having fewer nodes results in easier data visualization and identification of the most prominent features of the dataset, increasing the number of nodes generally leads to better discrimination between storm types. However, when too many nodes are considered, the data become more difficult to visualize, there are often redundant nodes, and more nodes have a statistically insignificant number of matching profiles. We find that for classifying the four storm types in this data sample, a $3 \times 3$ SOM appears to be the most effective means of analyzing the profiles. SOMs were better able to discriminate between supercells and nonsupercells when a greater depth of the profile was considered, but discrimination between nontornadic and tornadic supercells improved when shallower depths were used.

We employed a simple approach in developing an objective measure of forecasting potential. The skill statistics presented in this study are not meant as an absolute measure of the forecasting skill, but they do measure the relative ability each SOM might have as a
source of guidance in issuing conditional probabilities for supercell and tornado forecasts. Traditional measures of forecasting skill suggest that SOMs improve forecasting skill when the frequency of storm type in a node is used as the basis for conditional probabilities rather than the frequency distribution of the overall dataset. Furthermore, 0–6-km SOMs tend to have better skill than shallower SOMs, and SOMs of wind-derived variables generally have higher skill than SOMs of thermodynamic variables. The best forecasting skill is in NS regimes, while supercell regimes result in worse forecasting skill. As suggested by Markowski et al. (2003), NT and WT environments are statistically similar such that differentiating between them may be difficult. ST supercells by their nature are rare, such that a high degree of false alarms can result in poor skill scores. Yet, the SOM method has been shown to be effective in identifying regimes in which tornadoes are much more likely than climatology would suggest.

The primary advantage of SOMs is their ability to recognize patterns in the shape of a profile rather than simply considering bulk quantities like 0–6-km wind shear or helicity. Thus, as shown with the 0–6-km GR wind speed SOM, slight differences in the location and intensity of features in a profile are recognized and often affect the resulting PDF of storm type. This is especially evident in the 0–1-km streamwise vorticity SOM, where the near-ground streamwise vorticity appears to be critical to the prediction of ST supercells. Such distinctions may be less perceptible using established parameters like SCP or STP. Other studies have shown that hodograph shape in the lowest few hundred meters may be important to tornadoes (e.g., Miller 2006; Esterheld and Giuliano 2008). Togstad et al. (2011) suggested that “subtle differences in hodograph structure and shape in the lowest 1 km of the atmosphere may be very important in discriminating between tornadic and nontornadic storms... Such hodograph details are difficult (if not impossible) to incorporate into a statistical approach...” The results presented herein support their assertion that low-level wind details are essential in discerning tornadic environments. We believe the SOM technique, with further refinement, could be a feasible means of incorporating these data into a statistical forecast.

Considerable work is required before the SOM technique will be useful to operational forecasters. The
authors plan to extend this method to profiles of multiple variables in one SOM such that hodographs (considering both $u$ and $v$ in the same SOM) or combined temperature–dewpoint profiles may be considered together. This study examines the SOM technique wherein four storm types including NS are considered, but we suspect that a study eliminating the NS type might lead to new conclusions as to which SOMs are most effective for issuing tornado probabilities conditioned on the development of a supercell rather than simply a discrete storm. In this study, SOMs are generated using each variable independently. However, as shown when comparing the SOM of 0–6-km $\frac{du}{dz}$ with the composite hodograph of each node, a high probability of a particular storm type associated with one variable might be counterbalanced by a low probability associated with a different variable. As such, we plan to develop a system that appropriately weights the probabilities derived from SOMs of each variable into a single conditional probability estimate. The goal we envision for this method is a forecast product similar to the current SPC product based on the work of Togstad et al. (2011). Such a system should input real-time model analysis or forecast profiles at each grid point into an SOM or suite of SOMs trained on a much larger (and more representative of climatology) dataset than the RUC proximity soundings used here. The SOMs might then issue a conditional probability at each grid point that could be objectively analyzed in order to provide guidance to severe weather forecasters.

Acknowledgments. We wish to thank Rich Thompson and Roger Edwards at the Storm Prediction Center for collecting and supplying this dataset. We are also grateful to Harold Brooks and Paul Markowski for their encouragement and helpful suggestions for improving this work. We also thank the three anonymous reviewers, George Young, Mike Coniglio, Jim Marquis, and Ryan Hastings, for their opinions and constructive comments. This study was funded by NSF Grant AGS-0644533.

REFERENCES


