An Evaluation of Snowband Predictability in the High-Resolution Rapid Refresh

JACOB T. RADFORD AND GARY M. LACKMANN

North Carolina State University, Raleigh, North Carolina

MARTIN A. BAXTER

Central Michigan University, Mt. Pleasant, Michigan

(Manuscript received 30 April 2019, in final form 29 July 2019)

ABSTRACT

Narrow regions of intense, banded snowfall present hazardous travel conditions due to rapid onset, high precipitation rates, and lowered visibility. Despite their importance, there are few verification studies of snowbands in operational forecast models. The objective of this study is to evaluate the ability of the High-Resolution Rapid Refresh (HRRR) model to predict snowbands in the United States east of the Rocky Mountains. An automated band-detection algorithm was applied to a 3-yr period of simulated and observed radar reflectivity to compare snowband climatologies. This algorithm uses the distributions of reflectivities in contiguous precipitation regions to determine a band intensity threshold. The predictability of snowbands on a case-by-case basis was also evaluated using an object-oriented approach. The distribution of HRRR forecast banding resembles that of the observations, but with a significant positive frequency bias. This may partially be due to underrepresentation of observed bands in our verification dataset due to limited radar coverage in portions of the central United States. On a case-by-case basis, traditional skill metrics indicate limited predictability, but allowing for small timing discrepancies dramatically improves scores. Object-oriented verification yields mixed results, with 30% of forecasts receiving a score indicative of a well-predicted event. However, 69% of cases have at least one forecast lead demonstrating skill, suggesting the HRRR is successful in depicting environments conducive to band formation. These results suggest adopting a probabilistic, ensemble approach, and indicate that the deterministic HRRR is best suited for the identification of regions of elevated snowband risk and not precise timing or location information.

1. Introduction

Narrow regions of intense, banded snowfall present hazardous travel conditions due to rapid onset, high precipitation rates, and lowered visibility. Though to the authors' knowledge there has been no quantification of the societal or economic impacts associated specifically with snowbands, annual economic costs of heavy snow events are estimated to be in the billions (Adams et al. 2004). Through direct and indirect means, winter precipitation in the United States accounted for approximately 900 fatalities per year between 1975 and 2011 (Black and Mote 2015). Furthermore, mesoscale precipitation bands account for ~22% of total annual precipitation occurrences (Fairman et al. 2016) and were found to occur in 85% of cold-season heavy precipitation events in the Northeast (Novak et al. 2004, hereafter N04) and 63% of central U.S. heavy snow events (Baxter and Schumacher 2017, hereafter BS17). Snowbands are on the order of hundreds of kilometers in length and tens of kilometers in width, leading to extreme gradients in impacts over a short distance along the minor band axis. The limited extent of snowbands means even the mere occurrence of bands is difficult to predict, not to mention the timing, location, and intensity (e.g., Nicosia and Grumm 1999; Evans and Jurewicz 2009; Novak and Colle 2012).

Single bands are most commonly the product of an ageostrophic, thermally direct, transverse frontal circulation initiated by frontogenesis, which occurs when the axis of dilatation is within 45° of the orientation of isotherms (Petterssen 1956; Sawyer 1956; Nicosia and Grumm 1999). This thermally direct frontal circulation acts to offset frontogenetical increases in the magnitude of the horizontal temperature gradient, thereby maintaining thermal wind balance. Assuming saturated

Corresponding author: Jacob T. Radford, jtradfor@ncsu.edu

DOI: 10.1175/WAF-D-19-0089.1

conditions, the rising branch of the circulation may release conditional instability (CI) and produce heavy precipitation north of the surface cyclone (Schultz and Schumacher 1999). Novak et al. (2009) concluded that equivalent potential vorticity (EPV) is reduced in the cyclone warm sector due to differential horizontal temperature advection. Vertical motion is stronger and smaller in scale when collocated with low EPV in the warm sector (Emanuel 1985).

N04 and BS17 have developed robust snowband climatologies. N04 developed a band climatology utilizing a quasi-objective classification system for northeast U.S. cold-season precipitation events. This system separates precipitation features into classes of nonbanded, single-band, multiband, narrow cold frontal, transitory, or undefined based on band intensity, length, width, and duration criteria using base reflectivity. N04 found that "single" bands, defined as "a linear reflectivity feature 20–100 km in width and greater than 250 km in length. . . [with] a minimum intensity of 30 dBZ along a majority of its length for at least 2 h," were the most common, with 55% of cases falling into this category. They also developed surface-cyclone-relative band composites that were used to conclude that 81% of bands occurred to the northwest of the corresponding cyclone center, which were predominantly located immediately off the U.S. East Coast. Mirroring N04, BS17 conducted a climatology for the central United States, separating cases into banded (66) and nonbanded (38). These events were identified manually using a similar single band definition to that of N04, and used to create cyclone-relative band composites. Contrasting N04, BS17 found that most central U.S. bands were located to the northeast of the surface cyclone and exhibited a wider variety of orientation angles owing to more diverse 500-hPa flow regimes. While the banding ingredients of moisture, lift, and instability were found to be consistent, the locations of these ingredients relative to the surface low varied between northwest- and northeast-quadrant bands.

Few studies have assessed the predictability of snowbands due to the recent advent of models with adequate grid spacing to explicitly depict banding. Instead, forecasters have relied on forecasts of environmental conditions commonly associated with bands to determine at-risk locations (Novak et al. 2006; Evans and Jurewicz 2009). Evans and Jurewicz (2009) concluded that the presence of model-predicted banding ingredients of frontogenetical forcing, moist symmetric instability (MSI), and saturation were found in both major and minor snowfall events, but the "magnitude, depth, and persistence" of these ingredients correlate to total snowfall. However, this correlation decreases significantly with lead times greater than 12 h. Evans and Jurewicz (2009) propose a forecast strategy of using a high-resolution deterministic model to identify "snowfall potential," then applying an ensemble to provide location and timing probabilities.

Novak et al. (2008) simulated a northeastern U.S. banding case and found that while a 4-km MM5 model configuration could represent realistic band development, the intensity was significantly underestimated and the band axis was displaced by approximately 50 km relative to observations. Novak and Colle (2012) assessed snowband predictability using a 16-member 12-km ensemble system. This study was largely proof-of-concept, verifying the performance of the multimodel ensemble in just three banded cases. While noting the limitations of the relatively coarse grid spacing and the need for a larger sample size, the authors conclude that predictability varies significantly from case to case and is very sensitive to initial conditions. Furthermore, it was determined that the multimodel ensemble may be more useful in deciding *if* a band will occur rather than when and where a band will occur.

Advances in computational efficiency have allowed for the development of operational models with significantly finer grid spacings, such as NCEP's High-Resolution Rapid Refresh (HRRR), developed by the Earth System Research Laboratory (ESRL; Smith et al. 2008; Benjamin et al. 2016). The implementation of operational convection-permitting models such as the HRRR may aid forecasters in assessing snowband predictability and improving "single" snowband forecasts (Evans and Jurewicz 2009; Novak et al. 2008; Novak and Colle 2012). It is recognized that models such as the HRRR have been widely used by operational forecasters for several years, but, to our knowledge, there has not been a systematic verification study for numerical forecasts of banded snowfall. First, we present an objective band-detection algorithm, then compare HRRR-forecast and observational snowband climatologies east of the Rockies, followed by verification based upon matching of forecast and observed band objects. This paper is divided into four additional sections: data and methods, snowband climatologies, snowband verification, and conclusions and future work.

2. Data and methods

Observed snowbands were identified from NEXRAD base reflectivity mosaic (N0Q) products archived by the Iowa State Environmental Mesonet (IEM). Mosaics were collected at hourly intervals for the 2015/16, 2016/17, and 2017/18 winter seasons, defined as 1 November through 31 March. These three seasons were chosen due to the limited availability of archived HRRR data prior

TABLE 1. Mean observed and forecast snowband characteristics for the winter seasons of 2015/16, 2016/17, and 2017/18.

	Observed	3-h lead	6-h lead	9-h lead	12-h lead	15-h lead
No.	329	392	408	462	482	445
Length (km)	373 ± 128	376 ± 142	384 ± 131	389 ± 136	387 ± 148	392 ± 146
Width (km)	80 ± 35	86 ± 41	86 ± 36	89 ± 38	85 ± 42	87 ± 43
Axis ratio	5.2 ± 2.1	4.8 ± 1.8	4.9 ± 1.8	4.9 ± 2.0	5.1 ± 2.1	5.2 ± 2.2
Area (km^2)	11 200 ± 9200	15 200 ± 14 000	15 300 ± 11 700	15 900 ± 12 700	15 400 ± 14 300	16 300 ± 14 900
Solidity	0.48 ± 0.10	0.58 ± 0.13	0.59 ± 0.14	0.59 ± 0.14	0.59 ± 0.14	0.59 ± 0.14
Centroid temperature (°C)	−5.5 ± 4.6	−4.8 ± 4.0	−5.0 ± 4.1	−5.5 ± 4.2	−5.3 ± 4.6	−5.1 ± 4.4
Intensity (dBZ)	24.3 ± 3.8	20.4 ± 4.9	18.9 ± 5.0	18.4 ± 5.2	18.1 ± 5.4	18.4 ± 5.3

to this time (Blaylock et al. 2017). The NEXRAD mosaics have a grid spacing of 0.5 km and an intensity interval of 0.5 dBZ.

Forecast snowbands were identified from the operational version of the HRRR produced at the National Centers for Environmental Prediction (NCEP) for the same period. Specifically, precipitation features were identified in the 1000 m above ground level (AGL) reflectivity field. The HRRR has a grid spacing of 3 km, is initialized hourly, and provides hourly forecasts through 15 h. The HRRR was upgraded during the study period and forecasts were extended to 18 h, but hours 16–18 are not evaluated. The upgrade also included changes to the physical parameterizations, which could potentially result in variability in forecast performance before and after the implementation. Storage and processing constraints limited archival to every third initialization. Areas of poor radar coverage and all locations west of 104°W were masked out of the forecast imagery to limit the impact of precipitation differences in areas of inadequate NEXRAD coverage. This was accomplished by drawing a 230 km radius (maximum NEXRAD range) out from the location of each NEXRAD station location and then masking any precipitation falling outside of these radii in the HRRR field. This mask reduces the possibility of erroneously identifying HRRR band frequency biases in regions of poor NEXRAD coverage.

The N0Q images were referenced against Real-Time Mesoscale Analysis (RTMA) 2-m temperature products to estimate precipitation type. Simple masks were created based on a temperature threshold of 0°C, with any precipitation occurring within the cold region assumed to be snow. We recognize that this assumption is not always valid, as precipitation with subfreezing 2-m temperatures may be in the form of freezing rain or ice pellets, and it is possible for precipitation type to be in the form of snow with temperatures slightly above freezing. Nevertheless, this strategy serves as an adequate first guess given the available observations. To maintain consistency with our observational treatment, forecast precipitation types were likewise estimated from the HRRR forecast 2-m temperature field. The RTMA and HRRR 2-m temperatures are comparable, with an average observed snowband centroid 2-m temperature of −5.5°C and an average forecasted snowband centroid temperature between −4.8° and −5.5°C, depending on forecast lead (Table 1). Previous regional climatological studies (e.g., BS17; Ganetis 2017) have used Automated Surface Observing Station (ASOS) and/or Cooperative Summary of the Day (COOP) data to narrow down precipitation events to those with snow as the predominant precipitation type. This can be an effective solution when the domain is limited in scope and precipitation types are assumed to be relatively homogeneous, but this method is difficult to apply to our broader domain encompassing multiple synoptic systems and precipitation types.

Our snowband size criteria matches that used by BS17, which required a 250-km length minimum and an aspect ratio exceeding three. The 250-km length requirement limits detected bands to the larger, "single" band category that is the focus of our predictability study, while the minimum aspect ratio of 3:1 better addresses the interpretation of a band as a precipitation feature that is significantly longer than it is wide. In addition, because our automated procedure uses bounding rectangles that tend to overestimate a region's width, a 3:1 aspect ratio gives more leeway to capture larger band features that are more likely to exhibit some degree of curvature. That said, bands exhibiting significant curvature are likely not captured by our definition, so bands in our dataset are quasi-linear. The prevalence of curved banding is unclear and the identification of such bands is a potential avenue for future work.

In previous works, the intensity criterion has generally been more varied than the size criterion. BS17, for example, applied a 25-dBZ minimum, reduced from N04's 30-dBZ minimum. Ganetis et al. (2018) implemented an adaptive threshold, where the minimum reflectivity is calculated as the upper sextile of all reflectivity values throughout the domain. This definition was developed in light of the consideration that humans do not identify bands according to a set

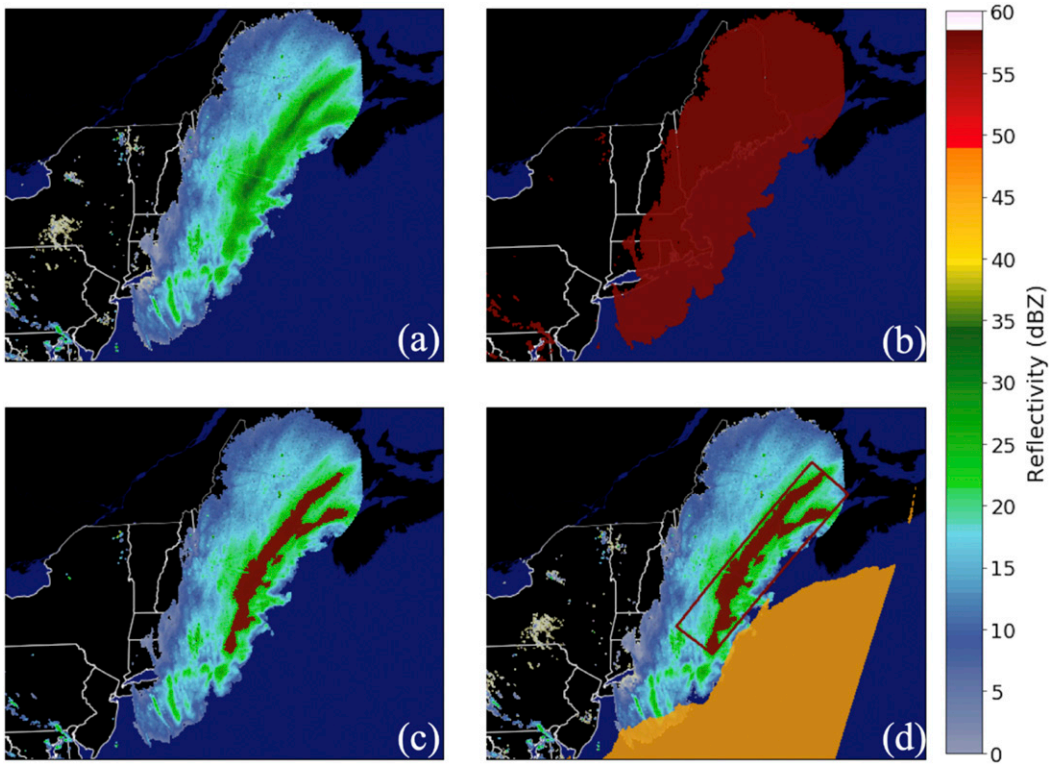

FIG. 1. A step-by-step visual representation of the band identification process. Beginning with (a) unaltered base reflectivity at 2300 UTC 9 Feb 2017, we threshold at 0 dBZ to identify (b) contiguous precipitation objects. Next, we threshold precipitation objects at 1.25 standard deviations above the object's mean to identify (c) embedded intense precipitation objects. Finally, (d) we apply a bounding rectangle to compare the region's length and width to our band definition and evaluate the fraction of the object's area falling below freezing (above freezing area is masked in orange).

reflectivity threshold or radar-display color value, but instead recognize bands as areas where reflectivity is locally maximized.

Our thresholding technique is most like Ganetis et al.'s (2018) adaptive threshold. We've applied a nested thresholding procedure, in which we first identify contiguous precipitation regions and then identify the embedded enhanced precipitation associated with banding. To identify contiguous precipitation features, the entire reflectivity field is thresholded at 0 dBZ, producing several distinct precipitation objects. A reflectivity value above each object's mean must be chosen in order to delineate a band; we selected a value of 1.25 standard deviations above each object's mean reflectivity based on favorable comparisons to manually identified bands and to tests using other standard deviation threshold values. Multiple thresholds were tested on several banded and nonbanded case studies across the domain, and the 1.25 standard deviation threshold was found to be an effective discriminant for what we subjectively identified as bands. Although we are confident that this method is an improvement on rigid reflectivity thresholds, it does not address the fact that the threshold choice is largely subjective.

Following digitization and bilinear spatial interpolation to the HRRR grid, N0Q images are smoothed using a 5×5 Gaussian filter, which helps to bridge small reflectivity discontinuities that may act to break up contiguous precipitation features. Band detection appeared to be relatively insensitive to filter width, but future detection work could benefit from a quantitative analysis of band frequency by filter size. From the original reflectivity image (Fig. 1a), precipitation objects are identified using a 0-dBZ threshold (Fig. 1b). Each object is compared against the 2-m temperature field to determine which grid cells will be classified as snow. Only cells with 2-m temperatures below 0°C are included in the calculation of a mean reflectivity. The reflectivity value 1.25 standard deviations above the mean reflectivity of each precipitation object is used as a threshold to reveal embedded precipitation features with a substantially higher reflectivity than their immediate surroundings (Fig. 1c). Once these precipitation features are identified, a rotated bounding rectangle is

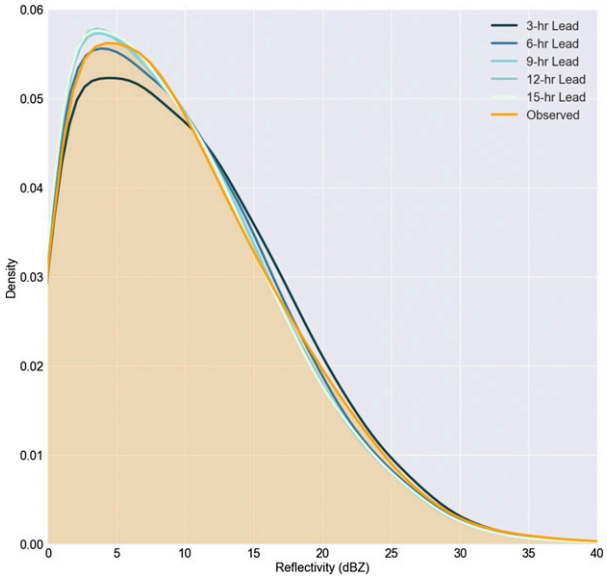

FIG. 2. Observed base and HRRR-simulated 1000-m reflectivity distributions for snow for the 2015/16, 2016/17, and 2017/18 winter seasons. Observed and HRRR forecast reflectivities at five different leads are shaded according to the legend.

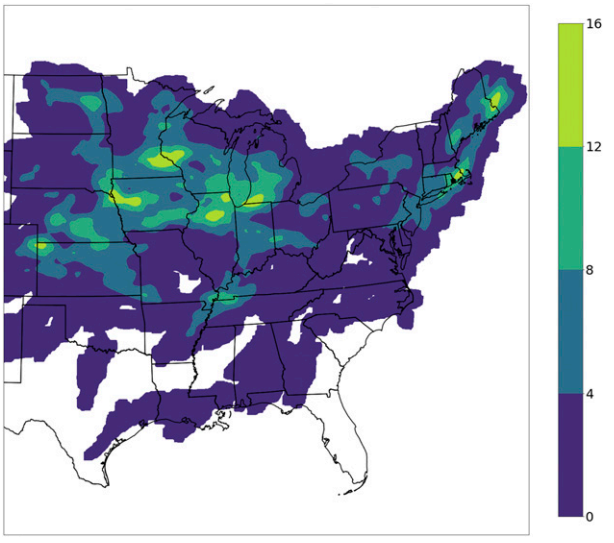

FIG. 3. The total number of hours banded snow was experienced for locations east of 104°W over the 2015/16, 2016/17, and 2017/18 winter seasons.

applied to each, allowing for measurements of attributes such as length and width. If over 50% of the precipitation object is below freezing and its bounding rectangle is 250 km in length with an aspect ratio of at least 3:1, it is classified as a snowband (Fig. 1d).

A major advantage of this thresholding technique is that it minimizes the influence of differences between observed base reflectivity and 1000-m simulated reflectivity. These two fields should not be directly compared without first acknowledging a key limitation: 1000-m reflectivity measures reflectivity at a constant height, while base reflectivity measures reflectivity at a constant elevation angle. Thus, equivalent values in the 1000-m and base fields should not necessarily be interpreted as perfect forecast accuracy. However, if we assume the distributions of the two fields to be similar, threshold values based on the mean and standard distribution should be comparable. The distributions of snow reflectivities above 0 dBZ are shown in Fig. 2. The distributions of the two fields are similar, with mean reflectivities of 10.1 dBZ and between 9.7 and 10.4 dBZ, depending on forecast lead time, for base and 1000-m forecast reflectivity, respectively.

3. Observed and forecast snowband climatologies

A total of 1003 snowband hours were identified from the observed base reflectivity east of the Rockies over the 2015/16, 2016/17, and 2017/18 winter seasons, using radar images at the top of every hour. When radar images only occurring at the top of every third hour are examined (to match HRRR initializations; e.g., 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC), 329 band hours were observed. Between 392 (3-h lead) and 482 (12-h lead) images with bands were identified from the HRRR-forecast 1000-m reflectivity, varying by lead time (Table 1). These values are 19% and 47% overestimates of observed band occurrence, respectively. Thus, the HRRR overestimates band frequency substantially, but the bias is reduced with decreasing forecast lead time. This frequency overestimate is not necessarily attributable to HRRR error, as overshooting in regions of sparse radar coverage may contribute, as will be investigated later in this section.

The smoothed distribution of band hours exhibits two local maxima (Fig. 3). One occurs over the coastal northeastern United States in an elongated region from New Jersey through eastern Maine. The second occurs in the central United States, extending from eastern Nebraska through Wisconsin and into Michigan. Both regions exhibit several locations receiving over 8 h of banding over the three seasons. These central and Northeast maxima largely corroborate the regional banding distributions presented in BS17 and N04. The only notable discrepancy between our distribution and these previous climatologies was a slight shift of the greatest density of Northeast snowbands from a region centered near Long Island, New York (shown in N04's Fig. 3), to eastern Massachusetts.

In a broad sense, the HRRR forecasts adequately capture the observed band distribution, particularly in the central United States (Fig. 4). A local maximum in

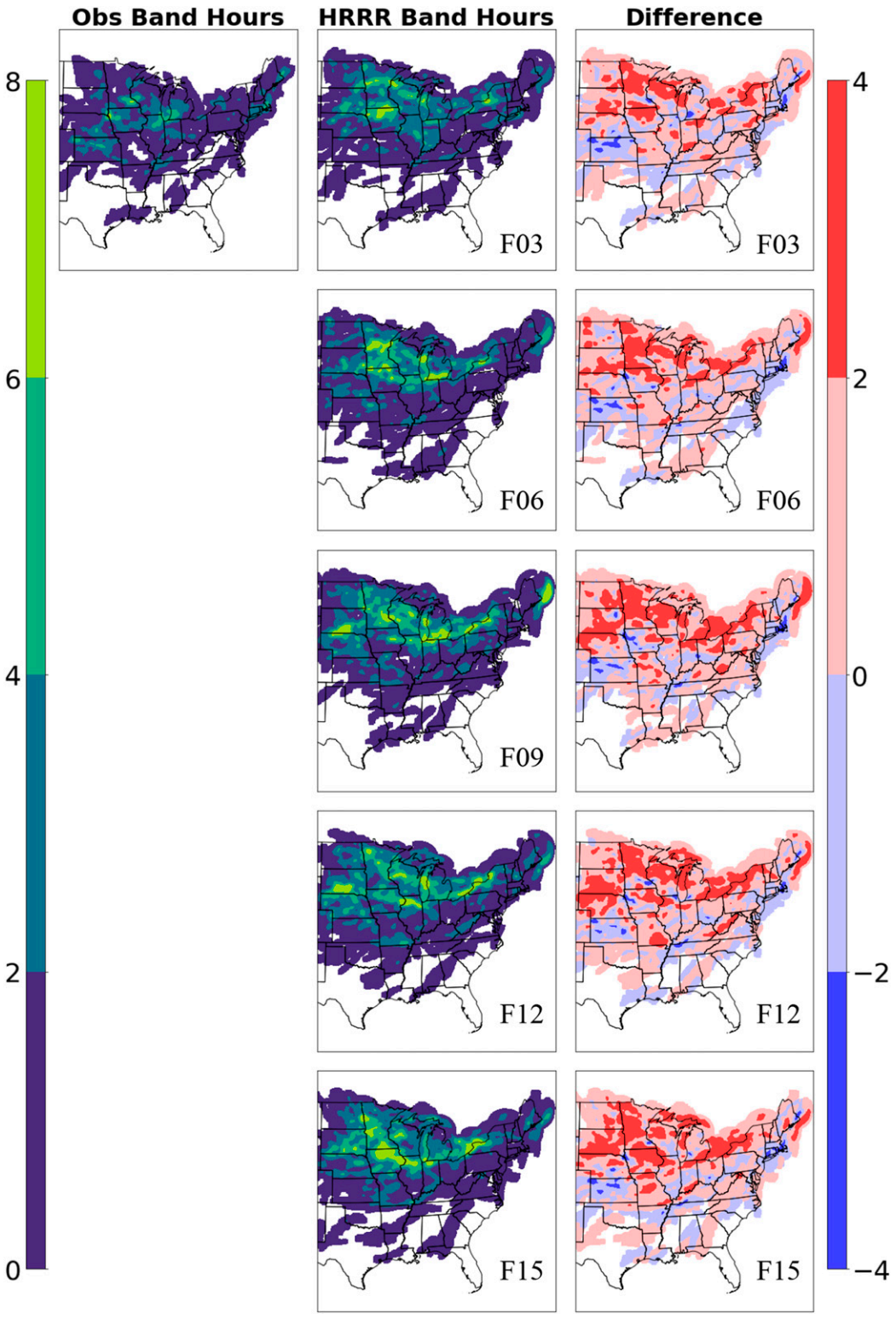

FIG. 4. (left) Observed distribution of banding, (center) forecast distributions of banding by lead, and (right) differences between forecast and observed banding frequencies. The units are hours of banding for the 2015/16, 2016/17, and 2017/18 winter seasons.

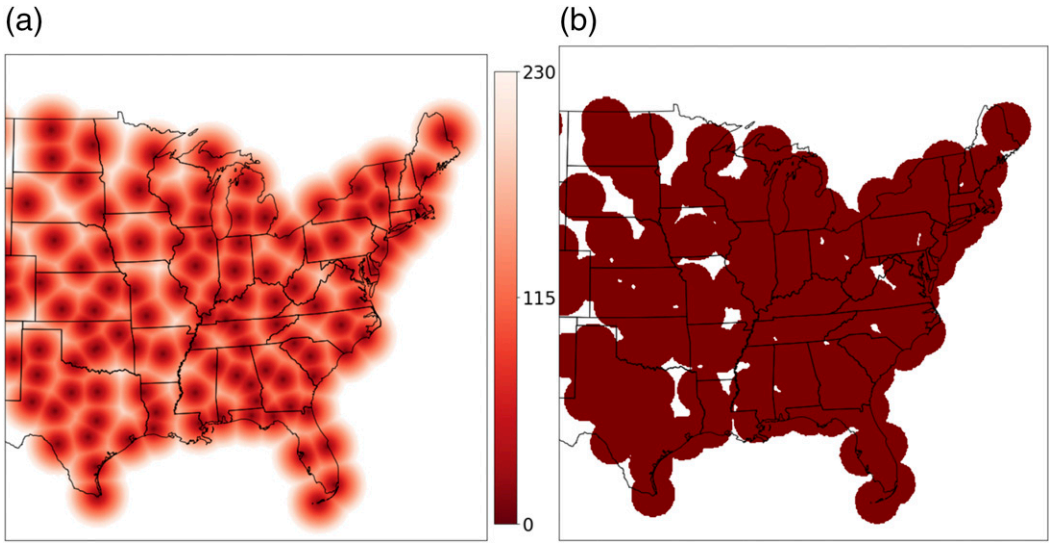

FIG. 5. (a) Distance (km) of each grid point from the nearest NEXRAD location and (b) a mask of locations with a beam height of less than 3 km.

the central United States similar to that in the observations appears in the HRRR forecast climatology at all lead times, with significant banding occurring between Nebraska and Michigan. The HRRR central U.S. maximum is oriented more zonally compared to the southwest to northeast orientation in the observed distribution. The observed New England coastal maximum is significantly underestimated by the HRRR. However, all lead times show a maximum in banding offshore of Maine, forming a couplet of underforecast banding along the coast and overforecast banding offshore. This could indicate that the HRRR is representing the same maximum, but has a northeastward location bias. Another reasonable interpretation would be that the observed coastal banding maximum extends offshore and coincides with the HRRR maximum, but the observational radar network has limited offshore range.

The forecast band distributions show widespread positive biases in band frequency, especially in the northern portion of the analysis domain (Table 1, Fig. 4). There are two regions in which the HRRR consistently overestimates relative band frequency. First, the HRRR exhibits a positive frequency bias in northern Wisconsin, Minnesota, and the Dakotas; however, these areas feature gaps in observational radar coverage. Figure 5a maps distance from the nearest radar site, while Fig. 5b demonstrates locations where radar beam height is less than 3 km AGL in altitude. Thus, the areas of most egregious contrast between model and observations may partially be a consequence of insufficient radar coverage rather than a model bias. Second is the HRRR's overestimate of banding along the Great Lakes coastlines, perhaps associated with lake-effect snowbands. The observed climatology exhibits a dearth of bands in lake-effect-prone areas compared to the HRRR climatologies. Our algorithm is not tailored for detection of lake-effect bands, as they are generally shorter than single bands and are not embedded within a larger stratiform precipitation region. However, larger lake-effect bands may still be identified. The forecast and observation lake-effect discrepancy could perhaps be explained by observational radars overshooting the tops of these shallow events, leading to band breakups and fewer observed features meeting the size minimum in the observed radar fields.

Band properties such as length, width, area, and intensity were calculated for detected bands in both forecast and observation. Observed band lengths ranged from the implemented minimum of 250 km to a maximum of 1033 km, with a mean length of 373 km. Observed band widths ranged from a minimum of 16 km to a maximum of 246 km, with a mean width of 80 km. The bands at the upper end of these ranges would more readily fall under the "meso-α" scale (Orlanski 1975; Houze et al. 1976; N04), but are included in this analysis as the distinction between the two is indefinite. Furthermore, the formation processes of these larger bands do not appear to be altogether different from more traditional, smaller single bands. For example, the largest band, which was 1033 km long and 246 km wide (keeping in mind the bounding rectangle width overestimation), occurred on 4 January 2018 to the northwest of an explosively intensifying offshore cyclone, following the formation of an occluded front (Fig. 6).

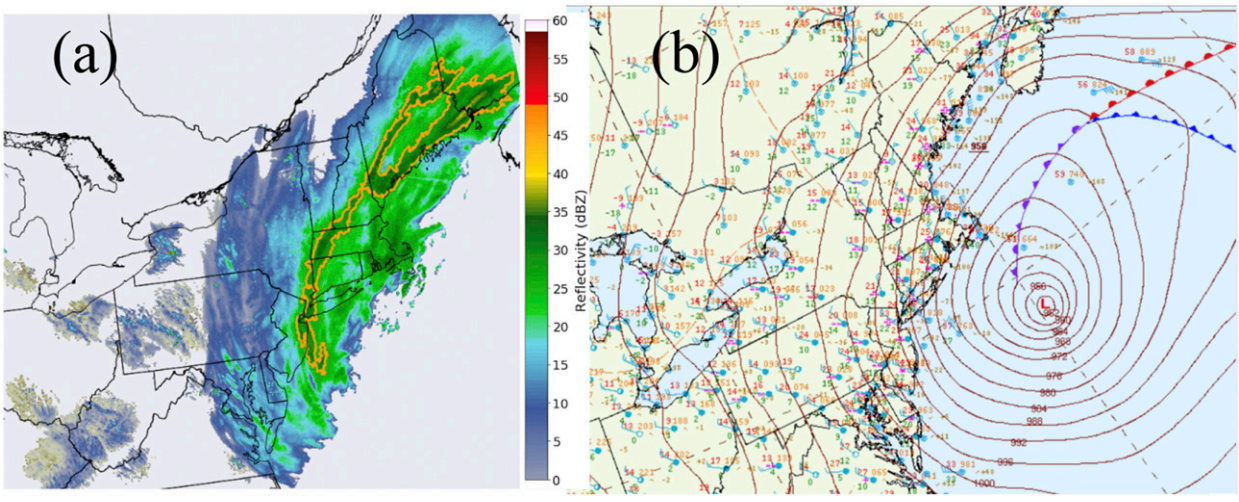

FIG. 6. (a) Observed reflectivity at 1800 UTC 4 Jan 2018 demonstrating a long snowband along the Northeast U.S. coastline. The identified snowband (contoured in orange) occurs on the northwest side of an occluded front associated with an intense offshore low, as seen in (b) the surface analysis, produced by the Weather Prediction Center (WPC; https://www.wpc.ncep.noaa.gov/html/sfc-zoom.php?h=18&y=2018&m=01&d=04).

Warm air wrapping around the cyclone brings lower static stability and elevated frontogenesis, leading to intense precipitation (N04; Schultz and Vaughan 2011).

Model forecast bands were slightly longer and wider than observed, with mean lengths ranging from an average of 376 km at the 3-h lead to 392 km at the 15-h lead and mean widths ranging from 86 km at the 3- and 6-h leads to 89 km at the 9-h lead (Fig. 7). Forecast band lengths ranged from the lower bound of 250 to 1408 km, while forecast band widths ranged from a minimum of 15 km to a maximum of 376 km. Band length and width distributions at each lead time and in the observations are shown in Figs. 7a and 7b, respectively. Aspect ratios averaged 5.2 for observed bands and between 4.8 at the 3-h lead and 5.2 at the 15-h lead for HRRR-forecast bands, indicating slightly more oblong objects in the observations.

There was a positive bias in forecast band area compared to the observations (Fig. 7c). While the mean observed area was 11 200 km^2, forecast band areas were between 15 200 and 16 300 km^2. Median forecast band areas ranged from 11 000 to 11 800 km^2, still a large positive bias when compared to the observed median of 8200 km^2. Percentage errors in the forecast mean band areas of 35%–45% far exceed the expected error from compounding length and width differences. The mean observed band solidity, or contour area divided by convex hull[1] area, was 0.48, compared with average solidities in the HRRR of between 0.58 and 0.59. In other words, observed snowbands were more complex, with more gaps and variation in contour coverage, while forecast objects were generally smooth. Thus, it appears that the HRRR demonstrates a substantial positive area bias, perhaps exacerbated by higher native resolution in the observational field compared to the HRRR grid spacing.

BS17 calculated a mean snowband length of 428 km, similar to our value of 373 km, and a mean width of 45 km, much narrower than our mean width of 80 km. Market and Becker's (2009) finding of a 61 km mean width falls roughly halfway in-between. However, BS17's narrower width is likely a result of manual versus automated band detection methods; while the human eye can effectively estimate the minor axis of complex features, bounding rectangles tend to overestimate feature width, especially in the event of band curvature, as shown in Fig. 8. On the other hand, manual methods must make a judgment call on where to measure the width of a precipitation feature. Our finding of a mean aspect ratio of 5.2 is in line with Fairman et al. (2016) and Market and Becker (2009), which found aspect ratios of 4.9 (mean) and 5.0 (median; mean not reported), respectively. BS17 found a much higher mean aspect ratio of 10.8, consistent with their narrower average bandwidth.

We recognize that there are additional complications in comparing a diagnostic field such as model-simulated reflectivity to observations. In particular, differences in microphysical parameterizations and their use in the calculation of radar reflectivity (e.g., Koch et al. 2005) can have a large influence on simulated reflectivities. Stratman et al. (2009) found that simulated reflectivity

[1] The convex hull of a contour is the smallest convex set containing all contour points.

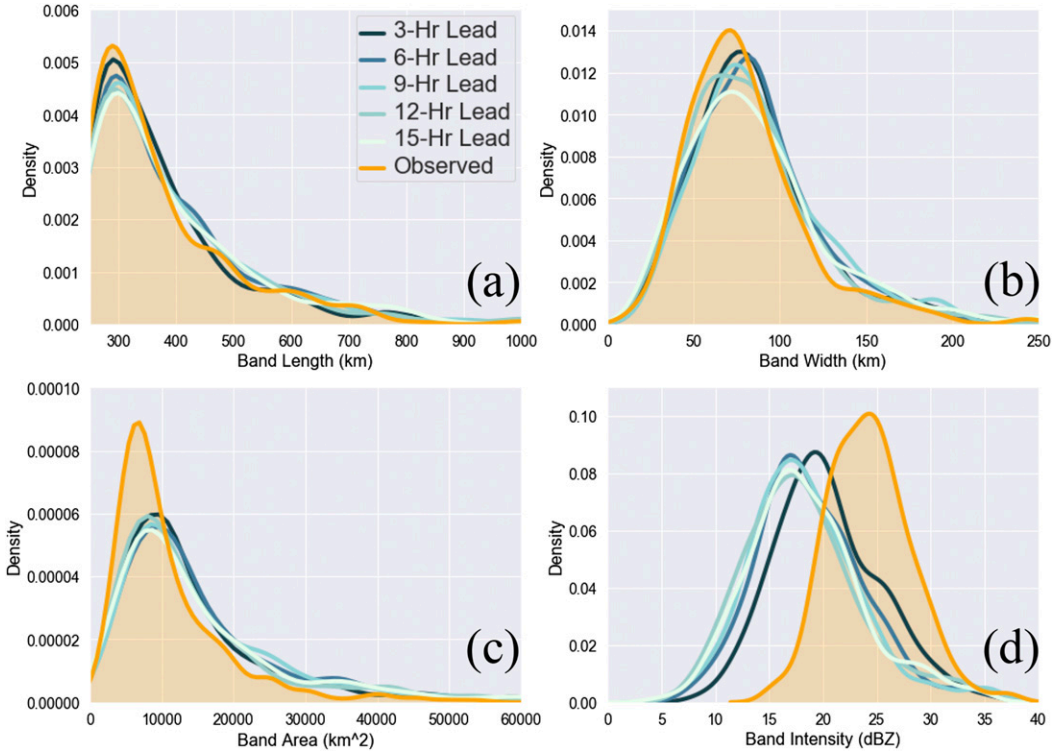

FIG. 7. Distributions of (a) length, (b) width, (c) area, and (d) mean intensity (computed as the mean reflectivity of all grid points composing the band). Observed and forecasted distributions are shaded according to the legend.

as calculated by the HRRR's Thompson microphysics scheme had a notable low bias at a 40-dBZ threshold when compared to WSR-88D radar data. While acknowledging these challenges, we compared mean intensities (defined as the mean reflectivity of all grid points composing the band) of forecasted and observed bands as a first attempt at evaluating forecasted band intensities. Forecast bands have mean intensities between 18.1 and 20.4 dBZ, well below the mean observed intensity of 24.3 dBZ. The full distribution of band reflectivities (Fig. 7d) is shifted toward lower values in the HRRR, seemingly implying that the HRRR is substantially underestimating banded precipitation intensity. The full distributions of below freezing forecast and observed reflectivities were similar in Fig. 2, suggesting that reflectivity differences are not solely due to differences between base and 1000-m reflectivities. A future research task could be to compare forecasted- and observed-band snow water equivalent to see if the intensity discrepancy is due to the assumptions used in simulated reflectivity calculations, or if it represents a negative intensity bias in the HRRR.

We next tested for regional variation in band properties. Bands were broken up into those occurring in the Northeast region (latitudes north of 36.5°N and longitudes east of 85°W), and those occurring outside this region, which mostly comprises central U.S. bands. This resulted in 281 Northeast bands and 722 "other" bands. Size differences were minor, but, as one might expect given the Northeast's nearby source of moisture, lower static stability, and greater ascent, the largest difference was found for average band intensity. While the mean intensity of the "other" band category was 24.0 dBZ, the mean intensity of the Northeast category

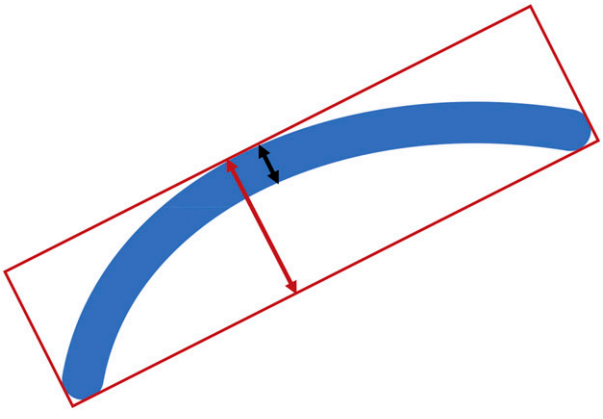

FIG. 8. Estimates of bandwidth using the human eye (black) and automated bounding rectangles (red). Bounding rectangles can significantly overestimate the width of objects with curvature.

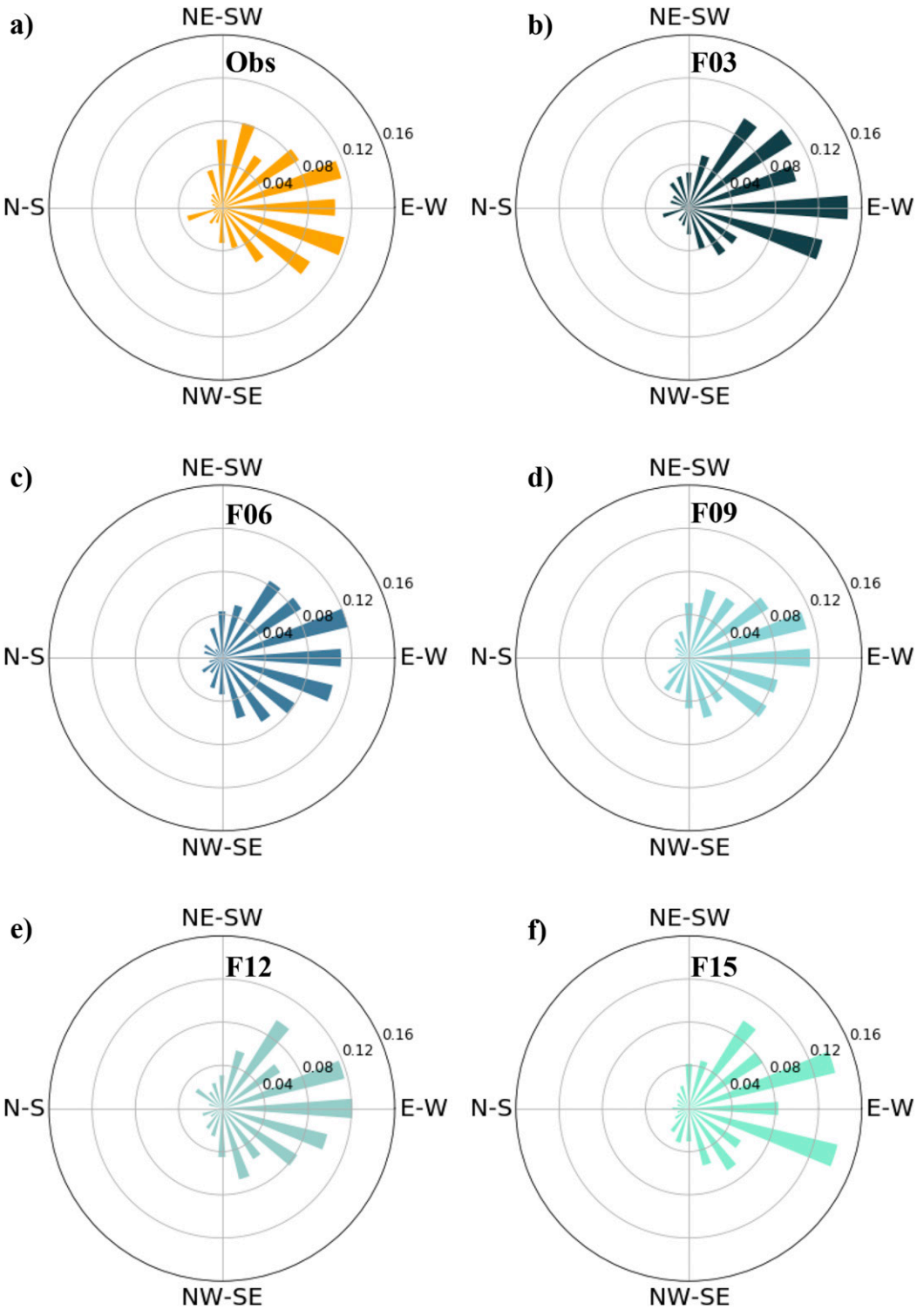

FIG. 9. Polar histograms of (a) observed and (b)–(f) HRRR-simulated band orientation angles.

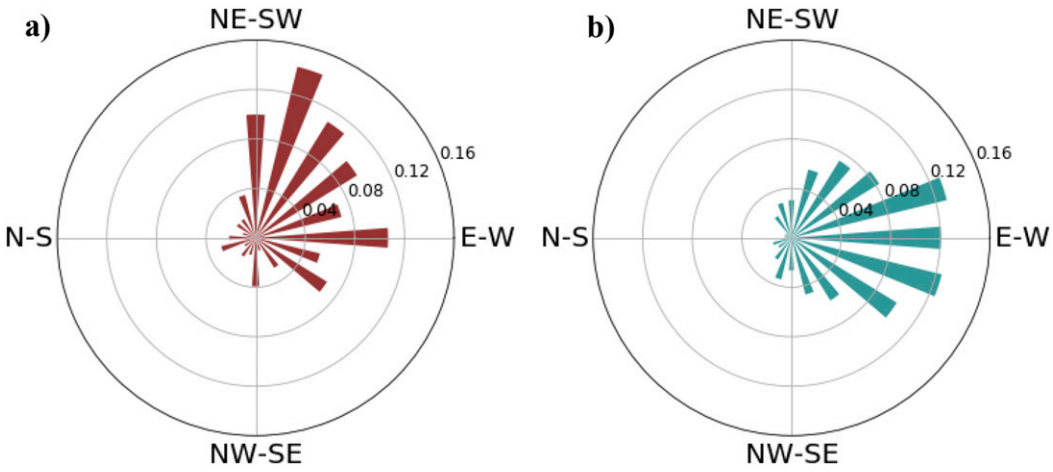

FIG. 10. Polar histograms of band orientation angles in (a) the Northeast region and (b) everywhere else for the 2015/16, 2016/17, and 2017/18 winter seasons.

was nearly 2 dBZ higher at 25.9 dBZ, traditionally a statistically significant difference ($t = 7.26$, $p < 0.0001$).

Given the consistent tracks of low pressure systems through the central United States and Northeast and the propensity for bands to form in regions of strong frontogenesis, it is reasonable to surmise that a pattern in band orientation would emerge. A polar histogram of observed band orientation angles normalized by total number of bands is shown in Fig. 9a. Most bands have a nearly straight east–west orientation with only minor tilts in the north–south direction. The most prominent orientation is tilted slightly from northwest to southeast. However, there is a significant portion of observed bands exhibiting a southwest to northeast orientation angle. The same orientation angle plots are shown for all five forecast lead times in Figs. 9b–f. The HRRR reproduces observed band orientation angles skillfully, with most forecast bands also exhibiting a nearly east–west orientation at all five lead times. Splitting these up by region, our results match BS17 and N04, with the Northeast category demonstrating southwest–northeast bands (Fig. 10a) and the ''other'' category demonstrating bands distributed about an east–west axis. BS17 posited that central U.S. storms moving southeast along the southern edge of the Alberta storm track exhibit the greatest confluence and midlevel frontogenesis downstream of the 700-hPa trough in a northwest–southeast orientation. This dilutes the otherwise dominant northeast–southwest oriented bands in southwesterly flow (BS17; Thomas and Martin 2007).

4. Snowband verification

While comparing HRRR-forecast and observed climatologies provides useful predictability information in a general spatiotemporal sense, it is more important that HRRR forecasts are accurate on a case-by-case basis. Do the individual bands predicted by the HRRR correspond to observed bands? The simplest approach to answering this question is to apply dichotomous contingency tables based only upon the presence of bands in order to calculate traditional verification metrics, such as probability of detection, false alarm ratio, and critical success index.

In our climatological comparisons, there were no spatial displacement restrictions for matching forecast and observed band objects. This means that bands could be matched from different locations in the domain, so long as they were detected at the same time. This raises concerns that results are influenced by false matches. To address these concerns, median matched band centroid displacements were calculated. The median displacement ranged from 164 km at the 3-h lead to 258 km at the 12-h lead. Furthermore, these are simple mass-weighted centroid displacements that are influenced by object shape and orientation. Thus, relatively small differences in spatial characteristics can lead to larger displacements. Object-oriented verification, to be presented shortly, further addresses spatial characteristics of matched objects, including object displacement.

According to probabilities of detection (POD), false alarm ratios (FAR), and critical success indices (CSI), the results using this dichotomous verification method depict limited band predictability. The HRRR's best POD performance is ~43% at the 12-h lead time. The worst POD performance is 37% at the 6-h lead time. On an encouraging note, 73% of observed bands were forecast in at least one of the HRRR forecast lead times. The FAR ranged from ~66% at the 3-h lead to 71.5% at the 9-h lead. It is important to note that the dramatic

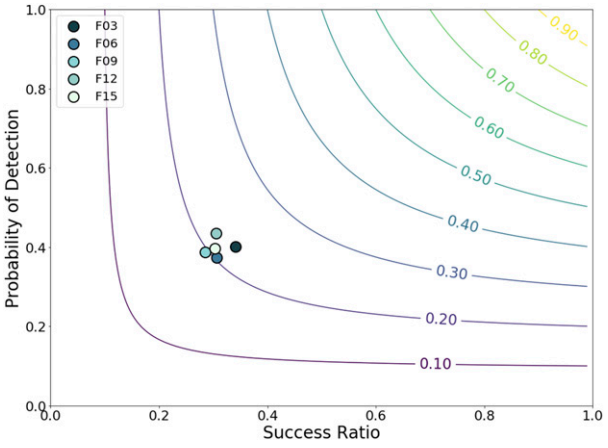

FIG. 11. Roebber's (2009) performance diagram for the 2015/16, 2016/17, and 2017/18 winter seasons at five different forecast leads. Success ratio (1 − FAR) is on the *x* axis, POD is on the *y* axis, and contours indicate CSI. Different forecast leads are shaded according to the legend.

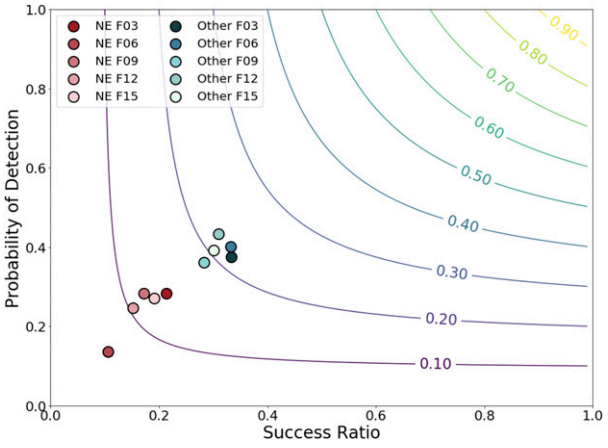

FIG. 12. As in Fig. 11, but breaking the set into two regions. Northeast and "other" category scores at five different forecast leads are shaded according to the legend.

overforecast in band frequency may at least in part be due to observational radar network coverage (Fig. 5). Probability of detection, success ratio (SR; the complement of FAR), and critical success index can all be visualized together using Roebber's (2009) performance diagram (Fig. 11). Points falling farther toward the upper-right corner of the diagram indicate better model performance. CSI scores were tightly packed, ranging from 0.20 at the 3-h lead to 0.22 at the 9-h lead. In other words, forecast bands rarely corresponded to observed bands and there is little variation with lead time.

Grouping events into hits and misses also allows us to evaluate the differing properties of observed bands that were correctly forecasted versus those that were not. The average hit had lengths varying by lead time of between 387 and 402 km, widths between 83 and 86 km, and areas of between 12 000 and 13 100 km^2. These values are larger than for missed bands, which had average lengths between 356 and 366 km, average widths between 76 and 78 km, and average areas between 9900 and 10 600 km^2. A traditional statistical significance threshold is only reached for length and width at the 3-h lead (length: $t = 3.0471$ $p < 0.005$; width: $t = 2.3971$ $p < 0.05$), though this does not necessarily mean there is no difference in reported size differences at other lead times, which exhibited greater variability (Amrhein et al. 2019). Differences in band intensities were generally small, but consistent across leads, with misses between 0.3 and 1.2 dB*Z* more intense than hits. This only reaches traditional statistical significance at the 12-h lead ($t = 2.7285$ $p < 0.01$), again with the caveat that subscribing to statistical significance thresholds is advised against by the statistics community (Amrhein et al. 2019; Hurlbert et al. 2019; Wasserstein and Lazar 2016).

Are bands more predictable in the Northeast or central United States? To address this question, we subsetted bands into 81 Northeast bands and 224 other bands based on a cutoff of 36.5°N and 85°W. The forecast skill for Northeast bands is lower than that of the "other" bands, with the Northeast band CSIs between 0.06 and 0.14, with the "other" band category featuring CSIs of between 0.19 and 0.22 (Fig. 12). The Northeast region's sample size is smaller than that of the "other" set, and thus a degree of increased variability is expected, but the causes of lower band predictability in the Northeast are worthy of additional study. Some potential hypotheses to investigate could be differing magnitudes of moist, midlevel frontogenesis and reduced EPV in the Northeast and central U.S. regions, the influence of terrain, and the HRRR's representations of Northeast coastal fronts and upstream convection. The latter is supported by Novak et al.'s (2009, 2010) findings of diabatic PV anomalies east of the primary NE bands, likely associated with convection driven by the warm Gulf Stream. It is well established that convection plays a factor in limiting predictability (Zhang et al. 2003), particularly for downstream precipitation (Mahoney and Lackmann 2007), so it is reasonable extrapolate these results to snowband predictability.

The dichotomous contingency table is rigid in its definition of what constitutes a hit. For example, a band may be forecast just an hour before or after a band was observed, but in this situation the model is penalized with both a false alarm and a miss since the forecast and observed band do not occur at precisely the same time. To account for this, we apply fuzzy verification

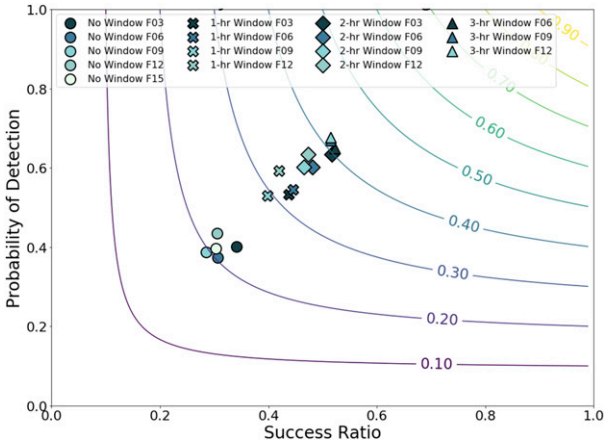

FIG. 13. As in Fig. 11, but applying fuzzy time windows. No, 1-, 2-, and 3-h time windows are symbolized according to the legend.

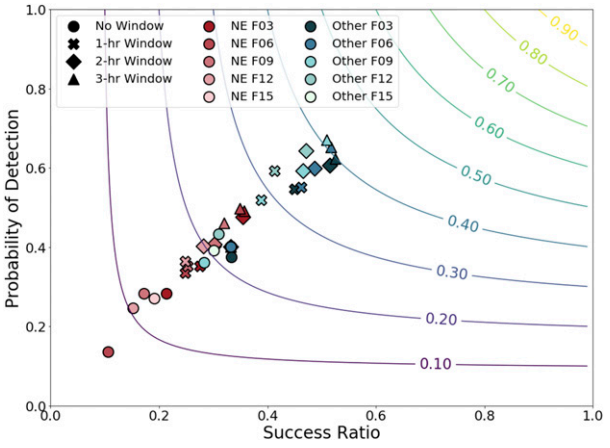

FIG. 14. As in Fig. 11, but broken down into regions with fuzzy time windows applied. No, 1-, 2-, and 3-h time windows are shown with different symbols for each lead time. The Northeast and ''other'' categories at different lead times are symbolized and color coded according to the legend.

(Ebert 2008), for which model skill is rewarded in these ''near-miss'' cases. Fuzzy verification was applied in the time domain. This entails creating small time windows, and, if a band is detected at any time in this window, a positive identification is recorded. Different window lengths may be chosen depending on the desired degree of leniency afforded to the model. In our case, 1-, 2-, and 3-h windows have been applied. Since these windows are symmetric, the maximum time difference between a forecast and observed band would be 2, 4, and 6 h, respectively.

With fuzzy time verification, POD increased from an across-lead mean of 39.9%–55.1% with a 1-h window, 61.8% with a 2-h window, and 66.6% with a 3-h window. Similarly, the FAR decreased from the initial 69.2%–57.5% with a 1-h window, 51.6% with a 2-h window, and 48.3% with a 3-h window. Consequently, CSI increased from 0.21 with no window, to 0.32 with a 1-h window, 0.37 with a 2-h window, and 0.41 with a 3-h window (Fig. 13). Our climatology combined with the significant increase in performance with the application of fuzzy time verification suggests that the HRRR may often capture the environmental conditions associated with heightened band risk and thus produce bands in temporal proximity to observed bands, but should not be expected to predict the precise timing of bands.

We again split bands into ''Northeast'' and ''other'' to determine if the difference in regional predictability persisted when fuzzy time verification was employed (Fig. 14). The lower predictability of Northeast bands extended to the 1-, 2-, and 3-h time windows. The Northeast region experienced very similar performance improvement to the ''other'' category, with CSI rising approximately 0.15 between the no-window scenario and the 3-h window. This supports the idea that the errors in forecast band timing for the Northeast are no smaller or larger than in other regions.

Fuzzy verification does not provide information on whether matched bands exhibit similar properties. The Developmental Testbed Center's (DTC) Method for Object-Based Diagnostic Evaluation (MODE) tool lets researchers match precipitation features based on the similarity between forecast and observed object properties (Davis et al. 2009). The choice of these parameters and their relative importance is somewhat arbitrary, but should be tailored to the specific phenomenon in question. The parameters we deemed most important to band forecast skill were centroid distance, area ratio, aspect ratio difference, and orientation angle difference, which we weighted equally. A tunable interest function was developed to produce an overall interest score (a dimensionless measure of forecast and observed object similarities) based on the four parameters:[2]

$$I_{\text{of}} = C_D \frac{150.0 - D_{\text{of}}}{600.0} + \frac{2.0 - \text{AS}_{\text{of}}}{8.0} + C_O \frac{90° - O_{\text{of}}}{360°} + \frac{1.0 - A_{\text{of}}}{4.0}, \quad (1)$$

where C_D = centroid distance confidence; D_{of} = centroid distance; AS_{of} = aspect ratio difference; C_O = orientation angle difference confidence; O_{of} = orientation angle difference; A_{of} = area ratio difference.

[2] We refer the reader to the Model Evaluation Tools user documentation (Gotway et al. 2018) for more details on how MODE calculates interest scores.

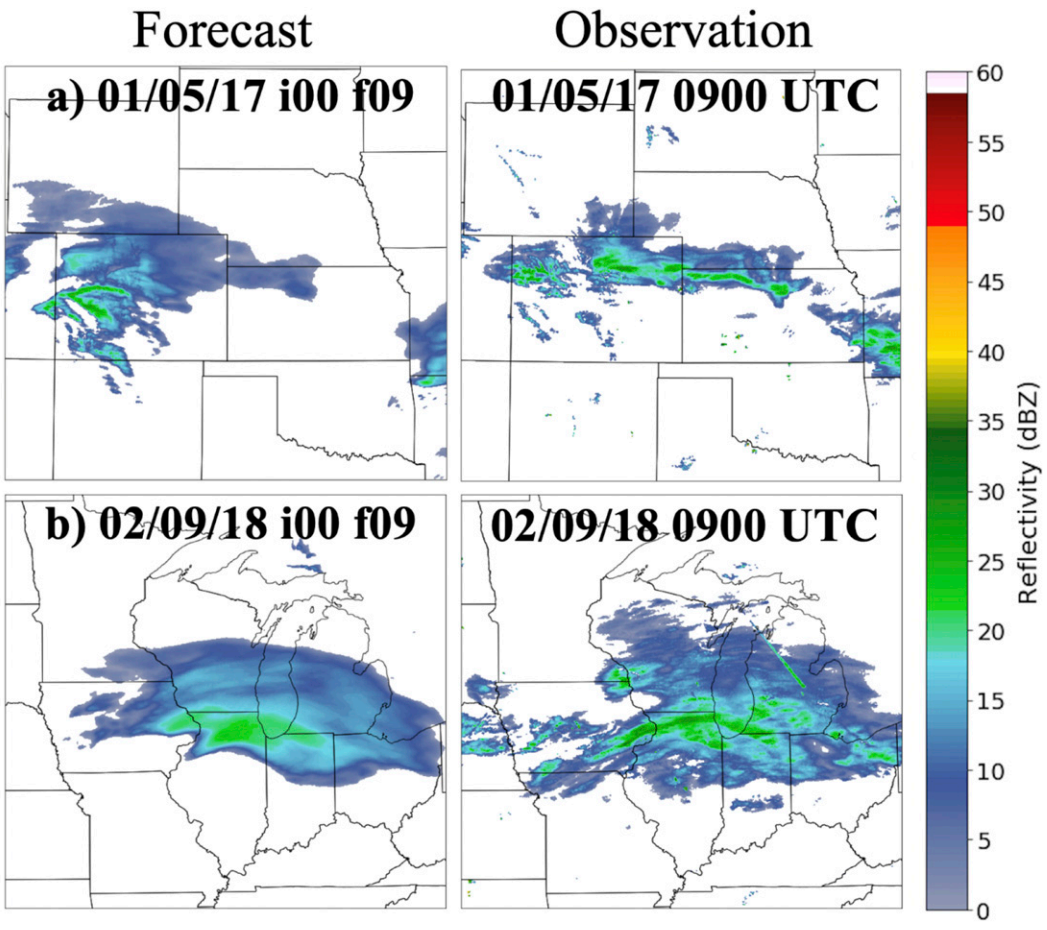

FIG. 15. (a) A comparatively poorly forecasted band case (0900 UTC 5 Jan 2017; interest = 0.0) and a comparatively well-forecasted band case (0900 UTC 9 Feb 2018; interest = 0.914) according to MODE interest scores. Both cases are shown at a 9-h forecast lead.

This interest function was designed specifically so that an interest score of 0.70 or greater correlates to a reasonably well-forecast banding event, similar to methods employed by Davis et al. (2009). Some examples of well- and comparatively poorly forecasted events are shown in Fig. 15. If errors were distributed equally between the four interest parameters, the 0.70 cutoff would equate to a centroid distance of 45 km, an area ratio of 0.70, an aspect ratio difference of 0.60, and an orientation angle difference of 27°. The mean interest for all 357 cases across all lead times was 0.54, while the median interest was 0.66. The distribution (Fig. 16) is highly skewed toward lower values due to the prevalence of cases in which one or more of the leads forecast no intense precipitation objects, leading to interests of 0.0. Breaking down the distribution by lead (Fig. 17), there is no trend with decreasing lead times, consistent with what was found in the traditional and fuzzy verification sections.

Using the 0.70 "well-forecast" threshold, the mean interest across leads indicate that 107 out of the 357 cases (30%) were well forecast. Out of the 357 cases, 247 (69%) had at least one forecast lead time with an interest greater than 0.70. This is similar to what was found in the traditional verification, in which 66% of cases were forecast in at least one forecast lead. Though less than ideal, this could help to signal forecasters that conditions are favorable for band development. Since there is no particular lead time that exhibits improved band forecast performance, it would be difficult for forecasters to determine which forecast lead will be correct; however, use of probabilistic visualizations such as a time-lagged ensemble could indicate areas where there is enhanced risk of banding. Examples of how these visualizations could assist in band forecast formation, timing, and duration are demonstrated in Figs. 3 and 4 of Novak and Colle (2012).

Well- and poorly forecasted events were separately composited through an affine coordinate translation to match centroids (Fig. 18). Figures 18a and 18b show band frequencies, with well-forecasted events demonstrating

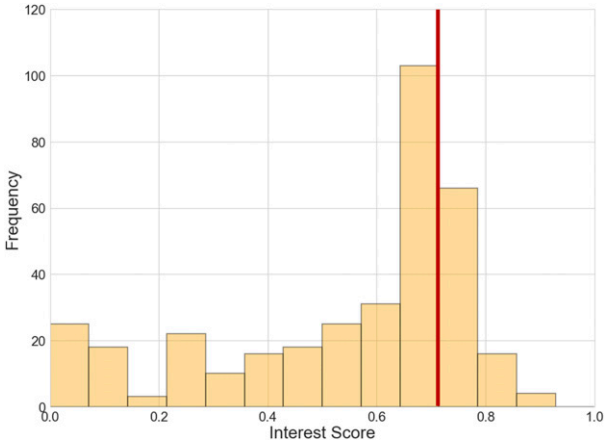

FIG. 16. The distribution of (across-lead) mean MODE interests between forecast precipitation objects and observed bands using equal weightings of interest parameters. The threshold for a well-forecast case is indicated by the bold red line.

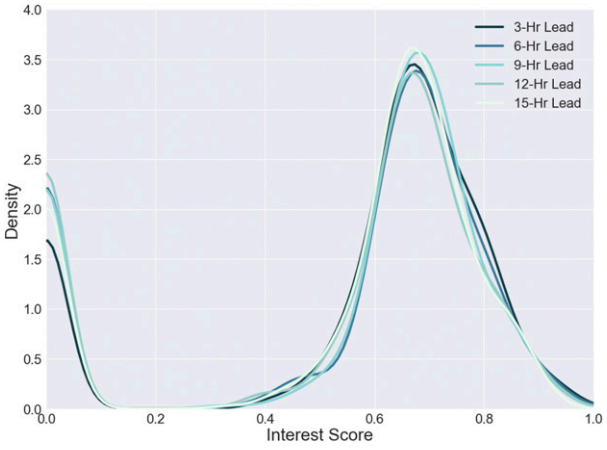

FIG. 17. Distributions of mean MODE interests at five different forecast leads (varying shades of blue).

a slightly more NW–SE orientation when compared to the west–east orientation of the poorly forecasted events. Figures 18c and 18d show the mean intensity of well- and poorly forecasted bands relative to the band centroid. There is no strong pattern to discern, but well-forecasted events are slightly less intense than poorly forecasted counterparts. Gathering information on the synoptic environments of these band cases will be of fundamental importance to further predictability analysis.

Finally, we investigate regional predictability based on MODE interest. Cases were split into 107 well-forecast cases and 250 poorly forecast cases based upon the 0.70 interest threshold. These well- and poorly forecast cases were then divided into ''Northeast'' and ''other'' subcategories. Unlike the results of traditional and fuzzy verification, there is no regional predictability dependence, with nearly identical proportions of well-forecast cases in the Northeast and elsewhere (27% vs 30%, respectively). The mean interest score in the Northeast was 0.54 compared to a mean of 0.55 elsewhere. Only cases with an observed snowband were evaluated here, thus the lack of regional interest variance seems to indicate that the poor Northeast predictability is associated with false alarms rather than missed forecasts.

5. Conclusions and future work

We have developed a robust, objective snowband definition that was then implemented into an automated detection algorithm to detect snowbands within reflectivity fields. This algorithm was first applied to observed base reflectivity and HRRR-forecast 1000-m reflectivity and supplemented with RTMA 2-m temperature fields to build region-independent forecast and observed snowband climatologies for the 2015/16, 2016/17, and 2017/18 winter seasons. Forecast and observed bands were then compared on a case-by-case basis to assess the predictability of bands by the HRRR. This was done using dichotomous contingency tables, fuzzy verification to give the HRRR a greater margin of error in time, and object-oriented verification using MODE to match precipitation features based on intrinsic properties of the objects such as location, orientation, area, and aspect ratio.

A total of 329 bands were detected in the observations, compared with between 392 and 482 in the model, constituting a large, positive bias. However, the largest overestimates in band frequency occur in regions of relatively sparse radar coverage. Reduced observational radar coverage and radar overshooting of shallow systems could preclude the detection of observed bands and lead to an overestimate by the HRRR. Band areas were forecast to be between 35% and 45% greater than observed, likely a combination of slightly longer and wider bands, along with higher average object solidities in the forecasts. Average band intensities were about 6 dBZ higher for observed bands. More investigation, such as comparisons between forecast and observed snow-water equivalent, will be necessary to clarify whether this is a true intensity bias, a consequence of assumptions used to calculate simulated reflectivity, or a difference between 1000-m and base reflectivities.

Considering results across all three verification methods, we reach the conclusion that the HRRR generally does not forecast snowband properties and timing with great accuracy. For every band that is forecast at the same time a band is observed (hits), there are nearly two bands forecast that were not observed (false alarms). However, allowing for small timing discrepancies of up to

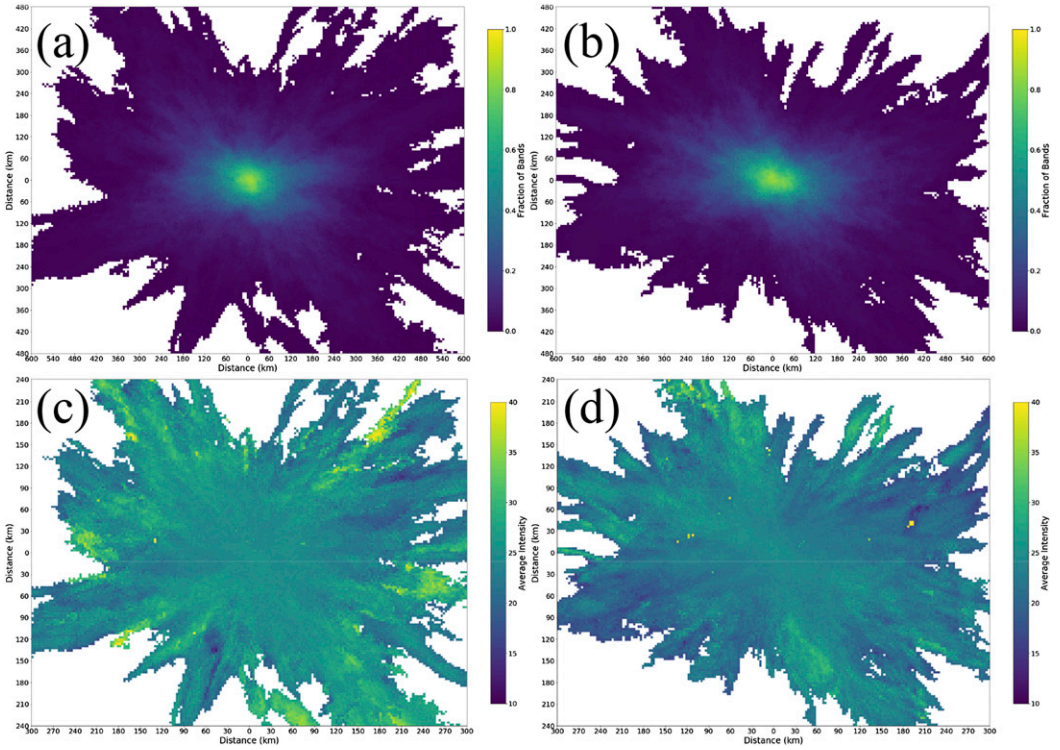

FIG. 18. Band-centroid relative composites demonstrating (a) poorly forecasted and (b) well-forecasted band frequency, and (c) poorly forecasted and (d) well-forecasted average band intensities. A MODE interest threshold of 0.70 was used as the cut-off for well- vs poorly forecasted band events.

3 h using a fuzzy verification approach significantly improved POD, FAR, and CSI. Object-oriented verification produced mixed results for HRRR performance. Only 30% of observed banding cases received an interest exceeding 0.70, a value corresponding to a reasonably well-forecast event. Though the fraction of well-forecast events across all lead times is low, 69% of cases received an interest of 0.70 for at least one lead time. There was no improvement with shorter forecast leads, a counterintuitive conclusion. There is perhaps some trade-off between model spinup impacting shorter leads and greater uncertainty impacting longer leads. Ideally, decreasing lead times would show greater consistency in band forecasts, but the presence of matching precipitation objects in at least one forecast lead could be beneficial to forecasters trying to assess if there is potential for a banding event. For example, the presence of a band at one lead time could be visualized on a time-lagged ensemble heat map, indicating a small probability of band formation. The HRRR may be useful to forecasters in determining when environmental conditions are favorable for banding, if not precisely when a band will develop, a conclusion also reached by Evans and Jurewicz (2009) and Novak and Colle (2012) for models employing convective parameterizations.

Band lengths, widths, areas, and solidities are relatively consistent between the Northeast and elsewhere, but bands in the Northeast are more intense, perhaps associated with the greater moisture supply, lower static stability, and stronger ascent near snowbands in the Northeast. While band orientations in the Northeast are primarily oriented from southwest to northeast, other bands are overwhelmingly oriented more zonally, results consistent with results of N04 and BS17. Northeast bands are less predictable than their central U.S. counterparts. CSIs for Northeast bands ranged from about 0.10 with no timing leniency to 0.25 with a 3-h window, while CSIs for other bands ranged from about 0.20 with no timing lenience to 0.40 with a 3-h window. Object-oriented verification, which only analyzed cases in which a band was observed (and therefore precludes false alarms) demonstrated nearly identical percentages of well-forecast cases between the two regions.

Armed with objective documentation of the difficulties in numerical prediction of banded snow events, additional research is needed to determine the factors that limit snowband predictability. More investigation into the synoptic environments associated with banding and topographical influences would be useful in determining why a large percentage of bands go undetected

even with a high false-alarm rate. A rich avenue for future work is to use the well- and poorly forecasted band dataset to develop more robust synoptic composites, empirical orthogonal functions, or self-organizing maps to identify robust signals that may distinguish well- and poorly forecasted banding events.

High-resolution NWP models such as the HRRR are just one tool in a forecaster's toolbox for forecasting snowbands. The HRRR may provide some value in identifying snowfall cases that have potential for band development, but it is inadequate for prediction of the precise band timing and location. HRRR guidance for high-impact, heavy snow events should be employed with these limitations in mind. Similar to other snowband predictability studies (e.g., Novak et al. 2008; Evans and Jurewicz 2009; Novak and Colle 2012), we recommend using the HRRR to identify regions of elevated risk for snowbands, rather than taking the timing, placement, and intensity at face value. Instead, application of high-resolution ensembles like the High-Resolution Ensemble Forecast (HREF) or the HRRR ensemble (HRRRE) could provide invaluable probabilistic information on the most likely location and time of band development, rather than simply signal the presence of environmental band ingredients.

Acknowledgments. This research was supported by NOAA Grant NA16NWS4680003, awarded to North Carolina State University. Additional support for this project was provided by the Developmental Testbed Center (DTC). The DTC Visitor Program is funded by the National Oceanic and Atmospheric Administration, the National Center for Atmospheric Research and the National Science Foundation. In particular, Tara Jensen, Jamie Wolff, John Gotway, and Randy Bullock provided advice on MODE usage. We thank Iowa State University for their base reflectivity mosaics and NCEP, Brian Blaylock, and the University of Utah for making HRRR data available. Finally, collaborators Jim Nelson, Sara Ganetis, Mike Erickson, Phil Schumacher, Mike Evans, and Jonathan Blaes contributed valuable advice throughout the research process. We thank Dr. Trevor Alcott and two anonymous reviewers for helpful comments and suggestions on our manuscript.

REFERENCES

Adams, R. M., L. L. Houston, and R. F. Weiher, 2004: The value of snow and snow information services. NOAA National Operational Hydrological Remote Sensing Center Rep., 49 pp.

Amrhein, V., S. Greenland, and B. McShane, 2019: Scientists rise up against statistical significance. *Nature*, **567**, 305–307, https://doi.org/10.1038/d41586-019-00857-9.

Baxter, M., and P. Schumacher, 2017: Distribution of single-banded snowfall in central U.S. cyclones. *Wea. Forecasting*, **32**, 533–554, https://doi.org/10.1175/WAF-D-16-0154.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Black, A. W., and T. L. Mote, 2015: Characteristics of winter-precipitation-related transportation fatalities in the United States. *Wea. Climate Soc.*, **7**, 133–145, https://doi.org/10.1175/WCAS-D-14-00011.1.

Blaylock, B., J. Horel, and S. Liston, 2017: Cloud archiving and data mining of High-Resolution Rapid Refresh forecast model output. *Comput. Geosci.*, **109**, 43–50, https://doi.org/10.1016/j.cageo.2017.08.005.

Davis, C. A., B. G. Brown, R. Bullock, and J. Halley-Gotway, 2009: The Method for Object-Based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Wea. Forecasting*, **24**, 1252–1267, https://doi.org/10.1175/2009WAF2222241.1.

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, https://doi.org/10.1002/met.25.

Emanuel, K. A., 1985: Frontal circulations in the presence of small moist symmetric stability. *J. Atmos. Sci.*, **42**, 1062–1071, https://doi.org/10.1175/1520-0469(1985)042<1062:FCITPO>2.0.CO;2.

Evans, M., and M. L. Jurewicz, 2009: Correlations between analyses and forecasts of banded heavy snow ingredients and observed snowfall. *Wea. Forecasting*, **24**, 337–350, https://doi.org/10.1175/2008WAF2007105.1.

Fairman, J. G., Jr., D. M. Schultz, D. J. Kirshbaum, S. L. Gray, and A. I. Barrett, 2016: Climatology of banded precipitation over the contiguous United States. *Mon. Wea. Rev.*, **144**, 4553–4568, https://doi.org/10.1175/MWR-D-16-0015.1.

Ganetis, S. A., 2017: The role of thermodynamic processes in the evolution of single and multibanding within winter storms. Ph.D. dissertation, Stony Brook University, 260 pp.

——, B. A. Colle, S. E. Yuter, and N. P. Hoban, 2018: Environmental conditions associated with observed snowband structures within northeast U.S. winter storms. *Mon. Wea. Rev.*, **146**, 3675–3690, https://doi.org/10.1175/MWR-D-18-0054.1.

Gotway, J. H., K. Newman, T. Jensen, B. Brown, R. Bullock, and T. Fowler, 2018: Model evaluation tools version 8.0 (METv8.0) user's guide. Developmental Testbed Center, 430 pp.

Houze, R. A., P. V. Hobbs, K. R. Biswas, and W. M. Davis, 1976: Mesoscale rainbands in extratropical cyclones. *Mon. Wea. Rev.*, **104**, 868–878, https://doi.org/10.1175/1520-0493(1976)104<0868:MRIEC>2.0.CO;2.

Hurlbert, S. H., R. A. Levine, and J. Utts, 2019: Coup de grâce for a tough old bull: "Statistically significant" expires. *Amer. Stat.*, **73**, 352–357, https://doi.org/10.1080/00031305.2018.1543616.

Koch, S. E., B. Ferrier, M. T. Stoelinga, E. Szoke, S. J. Weiss, and J. S. Kain, 2005: The use of simulated radar reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts. *11th Conf. on Mesoscale Processes*, Albuquerque, NM, Amer. Meteor. Soc., J4J.7, https://ams.confex.com/ams/32Rad11Meso/techprogram/paper_97032.htm.

Mahoney, K. M., and G. M. Lackmann, 2007: The effect of upstream convection on downstream precipitation. *Wea. Forecasting*, **22**, 255–277, https://doi.org/10.1175/WAF986.1.

Market, P. S., and A. E. Becker, 2009: A study of lightning flashes attending periods of banded snowfall. *Geophys. Res. Lett.*, **36**, L01809, https://doi.org/10.1029/2008GL036317.

Nicosia, D. J., and R. H. Grumm, 1999: Mesoscale band formation in three major Northeastern U.S. snowstorms. *Wea. Forecasting*, **14**, 346–368, https://doi.org/10.1175/1520-0434(1999)014<0346:MBFITM>2.0.CO;2.

Novak, D. R., and B. A. Colle, 2012: Diagnosing snowband predictability using a multimodel ensemble system. *Wea. Forecasting*, **27**, 565–585, https://doi.org/10.1175/WAF-D-11-00047.1.

——, L. F. Bosart, D. Keyser, and J. S. Waldstreicher, 2004: An observational study of cold season–banded precipitation in Northeast U.S. cyclones. *Wea. Forecasting*, **19**, 993–1010, https://doi.org/10.1175/815.1.

——, J. S. Waldstreicher, D. Keyser, and L. F. Bosart, 2006: A forecast strategy for anticipating cold season mesoscale band formation within eastern U.S. cyclones. *Wea. Forecasting*, **21**, 3–23, https://doi.org/10.1175/WAF907.1.

——, B. A. Colle, and S. E. Yuter, 2008: High-resolution observations and model simulations of the life cycle of an intense mesoscale snowband over the northeastern United States. *Mon. Wea. Rev.*, **136**, 1433–1456, https://doi.org/10.1175/2007MWR2233.1.

——, ——, and R. McTaggart-Cowan, 2009: The role of moist processes in the formation and evolution of mesoscale snowbands within the comma head of northeast U.S. cyclones. *Mon. Wea. Rev.*, **137**, 2662–2686, https://doi.org/10.1175/2009MWR2874.1.

——, ——, and A. R. Aiyyer, 2010: Evolution of mesoscale precipitation band environments within the comma head of Northeast U.S. cyclones. *Mon. Wea. Rev.*, **138**, 2354–2374, https://doi.org/10.1175/2010MWR3219.1.

Orlanski, I., 1975: A rational subdivision of scales for atmospheric processes. *Bull. Amer. Meteor. Soc.*, **56**, 527–530, https://doi.org/10.1175/1520-0477-56.5.527.

Petterssen, S., 1956: *Weather Analysis and Forecasting*. Vol. 1, McGraw-Hill, 428 pp.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Sawyer, J. S., 1956: The vertical circulation at meteorological fronts and its relation to frontogenesis. *Proc. Roy. Soc. London*, **234A**, 346–362, https://doi.org/10.1098/RSPA.1956.0039.

Schultz, D. M., and P. N. Schumacher, 1999: The use and misuse of conditional symmetric instability. *Mon. Wea. Rev.*, **127**, 2709–2732, https://doi.org/10.1175/1520-0493(1999)127<2709:TUAMOC>2.0.CO;2.

——, and G. Vaughan, 2011: Occluded fronts and the occlusion process: A fresh look at conventional wisdom. *Bull. Amer. Meteor. Soc.*, **92**, 443–466, https://doi.org/10.1175/2010BAMS3057.1.

Smith, T. L., S. G. Benjamin, J. M. Brown, S. Weygandt, T. Smirnova, and B. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh (HRRR) model. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1, https://ams.confex.com/ams/24SLS/techprogram/paper_142055.htm.

Stratman, D. R., and Coauthors, 2009: An examination of simulated reflectivity forecasts from a 10-member storm scale ensemble prediction system. *25th Conf. on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., P1.15, https://ams.confex.com/ams/89annual/techprogram/paper_150656.htm.

Thomas, B. C., and J. E. Martin, 2007: A synoptic climatology and composite analysis of the Alberta Clipper. *Wea. Forecasting*, **22**, 315–333, https://doi.org/10.1175/WAF982.1.

Wasserstein, R., and N. Lazar, 2016: The ASA statement on statistical significance and *p*-values: Context, process, and purpose. *Amer. Stat.*, **70**, 129–133, https://doi.org/10.1080/00031305.2016.1154108.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185, https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2.