

Validation of Snow Multibands in the Comma Head of an Extratropical Cyclone Using a 40-Member Ensemble

RYAN CONNELLY AND BRIAN A. COLLE

School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York

(Manuscript received 31 October 2018, in final form 1 July 2019)

ABSTRACT

This paper investigates the ability of the Weather Research and Forecasting (WRF) Model in simulating multiple small-scale precipitation bands (multibands) within the extratropical cyclone comma head using four winter storm cases from 2014 to 2017. Using the model output, some physical processes are explored to investigate band prediction. A 40-member WRF ensemble was constructed down to 2-km grid spacing over the Northeast United States using different physics, stochastic physics perturbations, different initial/boundary conditions from the first five perturbed members of the Global Forecast System (GFS) Ensemble Reforecast (GEFSR), and a stochastic kinetic energy backscatter scheme (SKEBS). It was found that 2-km grid spacing is adequate to resolve most snowbands. A feature-based verification is applied to hourly WRF reflectivity fields from each ensemble member and the WSR-88D radar reflectivity at 2-km height above sea level. The Method for Object-Based Diagnostic Evaluation (MODE) tool is used for identifying multibands, which are defined as two or more bands that are 5–20 km in width and that also exhibit a >2:1 aspect ratio. The WRF underpredicts the number of multibands and has a slight eastward position bias. There is no significant difference in frontogenetical forcing, vertical stability, moisture, and vertical shear between the banded versus nonbanded members. Underpredicted band members tend to have slightly stronger frontogenesis than observed, which may be consolidating the bands, but overall there is no clear linkage in ambient condition errors and band errors, thus leaving the source for the band underprediction motivation for future work.

1. Introduction

Precipitation bands often occur within the comma head of extratropical cyclones (e.g., Novak et al. 2009; 2010), either as a large primary band and/or numerous finer-scale multibands. About 85% of cyclones in the Northeast contained some sort of banded structures (Novak et al. 2004), and 81% of the bands were located in the northwest quadrant of the cyclone. Meanwhile, continental cyclones have been shown to feature more bands in their northeast quadrant than in the northwest (Baxter and Schumacher 2017). Primary bands, those that generally exhibit lengths on the order of hundreds of kilometers and widths on the order of tens of kilometers, have been documented and studied extensively (Novak et al. 2004, 2006, 2008, 2009, 2010); however, much less focus has been given to multibands, which often have a width of 5–10 km (Novak et al. 2004; Ganetis et al. 2018), yet still can produce intense snowfall rates.

a. Multibanded environments and prediction

Multibanded precipitation presents a forecasting challenge, since relatively few studies have examined the environments favoring these finer-scale bands. The development of a primary band typically includes mid-level frontogenesis in a region of weak stability or instability (Novak et al. 2008). Earlier studies emphasized conditional symmetric instability (CSI) for snowbands (Shields et al. 1991; Martin 1998; Nicosia and Grumm 1999), which is favored in environments with strong baroclinicity and vertical wind shear. However, CSI is not a necessary condition for the formation of snowbands (Shields et al. 1991; Novak et al. 2004; Norris et al. 2014). Analysis of a few case studies showed that CSI is the dominant instability responsible for multiple bands (Shields et al. 1991; Xu 1992; Nicosia and Grumm 1999). Using a much larger climatology Ganetis et al. (2018) found that CSI was evident in the average vertical profile for cases with either multibands or both a primary band and multibands.

Theory and idealized simulations have also suggested that increased CSI can promote more narrow convective

Corresponding author. Dr. Brian A. Colle, brian.colle@stonybrook.edu

structures, thus favoring multibands. Xu (1992) used a viscous semigeostrophic model, in which moist geostrophic potential vorticity (MGPV) is allowed to be negative, and showed that multiple finescale rainbands develop on the warm side of a frontal boundary in the presence of negative MGPV. Other studies found that these bands can coexist with a shallow layer of conditional instability (Shields et al. 1991; Norris et al. 2014). Idealized numerical simulations of a cyclone show the importance of latent heat release and surface friction in allowing the warm frontal precipitation to develop into multibands (Norris et al. 2014). Shields et al. (1991) observed multiple finescale snowbands in the warm advection region of a cyclone well north of the surface warm front, in a region of little low-level frontogenesis or even slight frontolysis. They tied the formation and persistence of these multibands to an axis of boundary layer confluence, and noted that they behaved similarly to narrow cold-frontal rainbands (NCFRs). Ganetis et al. (2018) also showed that there is often very little frontogenesis and instability with these bands, so the exact genesis mechanisms for these bands are unclear.

Hobbs and Locatelli (1978) showed using radar observations that lifting in a region of potential instability produced cloud-top generating cells and streamers that merge into warm-frontal bands. More recent observations have included turbulence-induced waves, both in the boundary layer and at cloud top that cause mesoscale precipitation structures in winter storms, in which multibands could be considered part (Raubert et al. 2017). This “seeder-feeder” process (Bergeron 1950) may influence band development (Stark et al. 2013; Rosenow et al. 2014), which exists at about 7–9 km AGL and on a horizontal scale of 0.5–1.5 km (Rosenow et al. 2014).

Previous research concentrated on the ability of mesoscale models to predict primary bands through simulating a few case studies and an ensemble (Novak and Colle 2012). Multibands may be more difficult to predict since many of these bands are smaller in scale. Novak and Colle (2012) varied initial conditions (ICs) and lateral boundary conditions (BCs) from a few deterministic models, and as well as cumulus convection and microphysical parameterizations to understand the primary band predictions. Greybush et al. (2017) enhanced an ensemble to predict snowbands over the Northeast United States by using a stochastic kinetic energy backscatter (SKEBS) (e.g., Berner et al. 2011) and stochastic perturbations to physical tendencies (SPPT) (e.g., Palmer et al. 2009). In that study it was found that the use of SKEBS and SPPT produced ensemble spread that was comparable to that of a multiphysics ensemble and IC/BC ensemble.

b. Motivation

High-resolution mesoscale models can realistically predict primary bands within the comma head of East Coast winter storms if they accurately resolve the lift, moisture, and stability (Novak and Colle 2012). In contrast, there has not been a study investigating the capability of high-resolution models to predict multibands in the cyclone comma head for several cases. It is hypothesized that mesoscale models require at least 1–2-km grid spacing to simulate these bands, and that these bands may be intrinsically more difficult to simulate, since the forcing (lift) for these bands is not as clear as primary bands (Ganetis et al. 2018).

Also, earlier studies of validating snowbands within models took a subjective (e.g., human eye) approach to comparing the observed and simulated bands. In this paper some of the latest feature-tracking approaches will be applied in order to objectively quantify the difference between observed and simulated band numbers and structures. This will provide the first overall comprehensive validation of multibands for several U.S. East Coast winter storms using an ensemble approach varying both ICs and physics. Some of the potential ingredients for band formation are explored (moisture, frontogenesis, and stability) to potentially explain band differences between the model and observed. Overall, this study attempts to answer the following two general questions:

- 1) How well can a mesoscale ensemble simulate multibands, and are there any biases with these bands (e.g., number, width, length, and location)?
- 2) What physical processes may be deficient in the model that may be leading to problems with the multiband predictions?

Section 2 summarizes the data and methods used in this study, specifically, the sources for observed data, the objective tool used for multiband identification, the ensemble design, methods of verifying the ensemble, and some grid spacing and parameterization scheme sensitivities applied to the ensemble. Section 3 highlights the multiband statistics for the ensemble, while section 4 explores some of the physical processes when grouped by banded and nonbanded ensemble members. Finally, a summary of this study and conclusions are presented in section 5.

2. Data and methods

a. Observations and case selection

Radar data from the National Weather Service (NWS) Weather Surveillance Radar-1988 Doppler (WSR-88D) network was used at selected sites in the northeastern

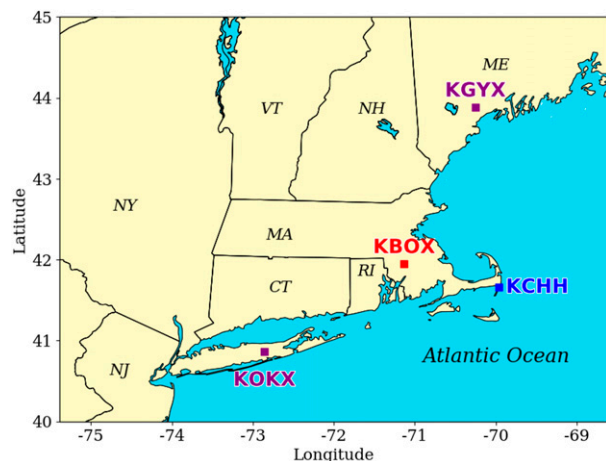


FIG. 1. Map of the four radar sites used for case selection, with upper air rawinsonde observation (raob) sites shown in blue, WSR-88D sites in red, and sites that contain both at the same location in purple.

United States (Fig. 1). Analyses from the NCEP Rapid Refresh model (RAP; Benjamin et al. 2016) at 13-km grid spacing and 25-hPa vertical intervals were used for three-dimensional comparisons with the Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model data.

Four cases of multibands were examined over the Northeast United States, which were selected based on lists of recent storms provided by the National Weather Service offices at Upton, New York; Taunton, Massachusetts; and Gray, Maine. Each case featured the development and persistence of multiband snow during 2014–17 for which we had RAP data, and the bands must occur sometime between 18 and 30 h after the 0000 UTC initialization, although bands between 12 and 36 h were included in the analysis. For the initial screening the cases must exhibit at least 2.54 cm of snow for the area of interest around the radar, and have at least 6 consecutive hours with 2 or more observed multibands that match the Ganetis et al. (2018) criteria, and at least 10 observed multibands in total during the period of banding. This resulted in 7 cases, but we favored the more active events, with 4 cases exhibiting at least 12.7 cm of snow, while another had less precipitation but at least 30 multibands.

The cases studied, the WRF initialization time used for each case, and the corresponding radar sites are summarized in Table 1.

b. Objective band identification

WSR-88D radar data is mapped from radial coordinates onto a Cartesian grid of 2-km spacing in the x and y direction using University Corporation for Atmospheric Research's (UCAR's) radx software (https://ral.ucar.edu/projects/titan/docs/radial_formats/radx.html). The radar domain for each case is a box with sides of 216 km, which is the largest square area that could be completely inscribed for the lowest elevation angle to reach 2 km above sea level. The height of 2 km was chosen to roughly maximize the radial range of data that could be plotted while not overshooting potentially important low-level features. The reflectivity data at both 2 km above sea level and a horizontal resolution of 2 km was objectively analyzed using NCAR's Method for Object-Based Diagnostic Evaluation (MODE; Davis et al. 2009) tool. The tool was used for each hour to identify band objects by applying a smoothing convolution with an input radius of grid points to both the observed or simulated radar reflectivity, and then filtering out values that fall below a given intensity threshold. The tool searches neighboring grid points and checks if they meet the given reflectivity threshold in order to combine those grid points into one discrete object. The convolution radius controls how many grid points away from the given grid point the tool should look. Thus, a smaller convolution radius results in a greater number of smaller objects being identified, while a larger radius results in fewer larger objects being identified that will also tend to be more smooth in appearance. The radius of the convolving disc is a user-defined parameter, so a few values were utilized for the model and observed (2, 4, and 8 km).

Choosing an adequate reflectivity intensity threshold proved difficult, since one value does not apply to all cases and at all times. Therefore, as in Ganetis et al. (2018), the threshold is set as a function of the reflectivity field at each time step. Multiple sets of MODE output were used for the upper quartile, upper sextile, upper octile, and upper decile of the reflectivity in the inner most nested WRF domain for each member (or observed) and each

TABLE 1. Table of cases, WRF initialization times, WSR-88D sites, and forecast hour window used for verification for each case.

Case	WRF initialization time	WSR-88D site	Forecast hour window
15–16 Feb 2014	0000 UTC 15 Feb 2014	KBOX	14–30
26–27 Nov 2014	0000 UTC 26 Nov 2014	KGYX	18–36
7–8 Jan 2017	0000 UTC 7 Jan 2017	KOKX	13–30
9–10 Dec 2017	0000 UTC 9 Dec 2017	KOKX	13–30

TABLE 2. Fraction of time that each of the MODE permutations are chosen as the “best,” expressed as a percentage of all WRF member hours, for each case and for all four cases combined. The UpQ, UpS, UpO, and UpD refer to the upper quartile, sextile, octile, and decile, respectively. The convolution disk radius used by MODE is listed as 2, 4, and 8 km.

	UpQ 2 km	UpQ 4 km	UpQ 8 km	UpS 2 km	UpS 4 km	UpS 8 km	UpO 2 km	UpO 4 km	UpO 8 km	UpD 2 km	UpD 4 km	UpD 8 km
15–16 Feb 2014	8%	10%	14%	8%	5%	10%	4%	9%	11%	5%	9%	8%
26–27 Nov 2014	7%	9%	20%	3%	10%	13%	4%	7%	11%	2%	8%	7%
7–8 Jan 2017	8%	5%	9%	9%	6%	15%	3%	5%	14%	8%	5%	14%
9–10 Dec 2017	7%	8%	13%	7%	5%	13%	6%	7%	9%	6%	8%	11%
All four cases combined	7%	8%	14%	7%	6%	13%	4%	7%	11%	5%	8%	9%

time. In total 12 permutations of MODE were run on all forecast hours for each WRF member as well as on the observed radar reflectivity field. These permutations were comprised of these three convolving disc radii and four statistical intensity thresholds as described above. A single forecast hour for a single member of the WRF ensemble will be referred to as a “member hour” in this paper. For each member hour, all 12 of the above MODE permutations were run and used for most of the verification below. The same 12 MODE permutations were also used for observed radar analysis for consistency during the validation.

A “best” member was also used for some validation comparisons below, so an automated procedure was defined to identify the best using the 12 MODE permutations. If a given permutation contained at least two objects that satisfied the criteria for a multiband, the sum of the area that those multibands covered was noted. If that sum was greater than the previously greatest value, this permutation overrode the previous one as the new best permutation. After all 12 had been checked, the best permutation is chosen systematically as the one whose identified multibands covered the most area. The best permutation is redefined for each forecast hour of each WRF member, as well as for each hour of the observed radar reflectivity. The areal coverage is used rather than a simple count of multibands because it was found that in some instances, especially at the lowest convolving disc radius and highest statistical intensity threshold, MODE would break apart features that a human would subjectively classify as one banded object into two or more multibands that did not fully connect. Table 2 shows the percentage each of the 12 different MODE permutations was considered the “best.”

Definitions were then applied to the MODE-identified objects in order to identify multibands. The definition was based on the length, width, and area of each object. The criteria for this algorithm were initially drawn from Novak et al. (2004). Multibands were at first defined as objects with a width of at least 5 km, but no more than 20 km, and whose aspect ratio of the long to short axis

exceeds 2:1. However, it was found that given the noisiness of radar reflectivity, very small objects were being classified as multibands that did not last at least 20–30 min when viewing the radar data. Therefore, it was decided after numerous tests that the best representation of multibands occurs when the width of MODE-identified objects is greater than 8 km, but less than 30 km. This is similar to the 10–50-km range used in Ganetis et al. (2018), but it was reduced to allow for some more bands to validate. An 8-km minimum width also allows the model to marginally resolve these features when run at 2-km grid spacing, but not setting the minimum width too large to eliminate many of the very narrow bands in the 2-km radar analyses. Since multibands are highly transient and difficult to track, adding a temporal longevity rule to the band identification would have added another degree of uncertainty to the analysis (same band or new band?), so we compared the bandlike features at each time.

Figure 2 shows an example of all 12 MODE analyses for a single forecast hour of a single WRF member as well as the simulated reflectivity (Fig. 2a) for the best permutation (Fig. 2f) at 2 km MSL at 2200 UTC 7 January 2017. The multibands are identified in each of the 12 MODE permutations but that the number of objects identified as multibands varies by permutation. The range of reflectivity thresholds at this time and for this WRF member ranges from 18.6 dBZ for the upper quartile to 20.2 dBZ for the upper decile. The number and shape of objects is relatively consistent for the three convolution radii for a given statistical reflectivity threshold (e.g., upper quartile), except that the bands appear “fatter” at the highest (8-km) convolution radius. However, the number of bands changes more significantly when moving from one statistical reflectivity threshold to the next, with the middle two (upper sextile and upper octile) generally resolving more bands than the lowest and highest thresholds (upper quartile and upper decile). This was generally found for other times and cases as well (not shown).

One limitation to this methodology is that radar reflectivity objects may intersect the edge of the 216 km ×

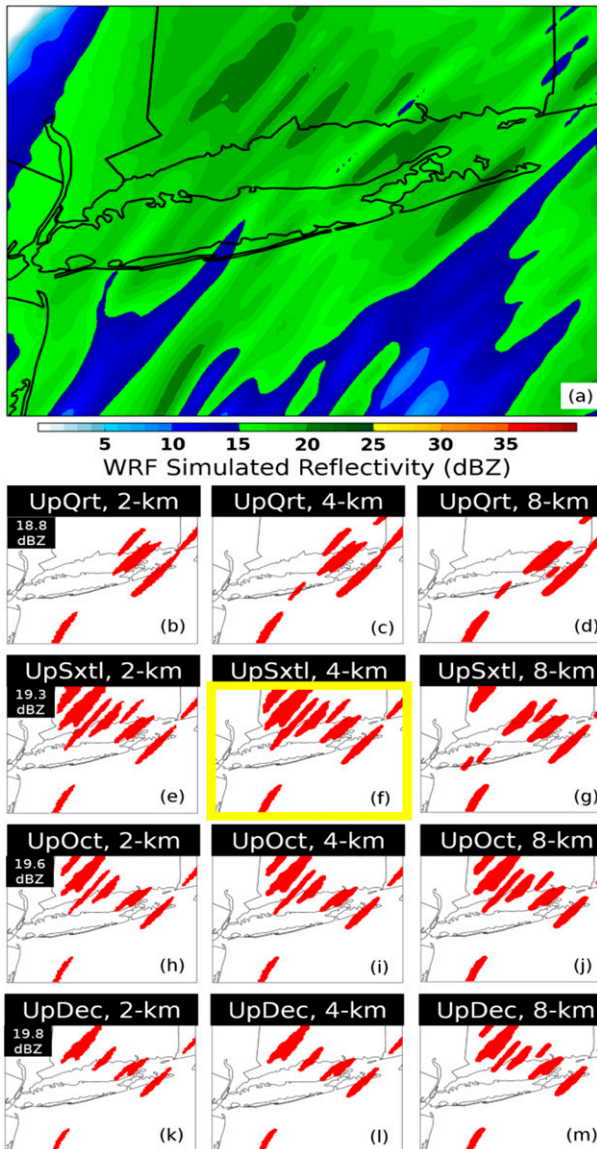


FIG. 2. (a) WRF-simulated reflectivity at 2 km MSL for a member at 2200 UTC 7 Jan 2017 (hour 22). (b)–(m) Band objects for the 12 different MODE permutations for the same WRF member and same forecast hour, showing the difference in number of bands identified. The “best” permutation is outlined in bold yellow in (f).

216 km verifying domain. For example, a large, nearly circular stratiform region that is only partially in the domain may result in MODE erroneously identifying a fine sliver of the precipitation region that intersects the domain as a precipitation band. To mitigate this issue, a further criterion is applied: an object is only allowed to be identified as a multiband if the distance from its centroid (an attribute written out by MODE) to the domain boundary is greater than half of the object length.

Counts and areal coverage of multibands are then recorded in both modeled and observed fields so that

WRF members can be sectioned into hit, miss, false alarm, and correct null groups for each forecast hour and for thresholds of 2, 3, and 4 multibands using the same custom algorithm operating on the observed reflectivity field from the time that is closest to the forecast hour in question. WRF’s reliability in simulating multibands will be further assessed by examining attributes of the multibands that WRF simulated. These attributes are calculated directly by MODE and viewable in the output text file for each member hour. The attributes chosen are 1) length, 2) width, 3) centroid latitude, 4) centroid longitude, and 5) areal coverage. Latitude and longitude are analyzed as the mean centroid position of all multibands in the domain for a given member hour or radar hour.

c. WRF ensemble setup

An ensemble was run for each of the four cases using WRF as summarized in Table 1. The 40-member ensemble uses the first five members of the GEFS reforecast at 1° grid spacing (GEFSR; Hamill et al. 2013) as in Greybush et al. (2017), which constructed a WRF ensemble down to 3-km grid spacing to simulate snowbands within two U.S. East Coast storms. These GEFSR members were used as initial conditions at 0000 UTC for each case (since GEFS reforecast is only run every 0000 UTC), and the GEFS forecast every 3 h were used for boundary conditions, as well as for initializing snow cover, soil temperature, and sea surface temperature (SST). For each GEFSR IC, four runs varied microphysics and planetary boundary layer parameterization schemes as other studies (Gallus and Bresch 2006; Jankov et al. 2007; Jones et al. 2007; Novak and Colle 2012; Schumacher and Clark 2014). The microphysical schemes used are Thompson (Thompson et al. 2008) and Morrison (Morrison et al. 2009). These are chosen because Thompson is a single-moment scheme, while Morrison is a double moment. In addition, the Morrison scheme assumes spherical ice and constant-density snow, while the Thompson scheme prescribes snow density as being inversely proportional to its diameter.

For the planetary boundary layer, the YSU (Hong et al. 2006), and MYNN2 (Nakanishi and Niino 2006) schemes are used. These two schemes also differ significantly, in that the YSU scheme is a “nonlocal” mixing scheme, while the MYNN2 scheme is a “local” scheme (Coniglio et al. 2013). These schemes were used form a matrix (Fig. 3), with four unique pairings of different physics to run with the five different ICs. For all runs, cumulus convection was parameterized using the Grell–Freitas scheme (Grell and Freitas 2014) on the outermost 18- and 6-km domains. This scheme is chosen to be in accordance with what is used on the outermost domain of the NCEP High Resolution Rapid Refresh

	GEFS 1	GEFS 2	GEFS 3	GEFS 4	GEFS 5
SKEBS+SPPT 1					
SKEBS+SPPT 2					
SKEBS+SPPT 3		SKEBS+SPPT Ensemble			
SKEBS+SPPT 4					
MYNN+THOM					
MYNN+MORR					
YSU+THOM		Classical Ensemble			
YSU+MORR					

FIG. 3. Diagram illustrating the matrix of WRF runs used in this study, which included 20 members of an ensemble using different GEFS forecast members (“classic” ensemble) with different physics, and a 20-member ensemble using both the GEFS members and the SKEBS+SPPT perturbations.

model (HRRR), which is a WRF-ARW run nested within the NCEP RAP model (Benjamin et al. 2016).

Stochastic perturbations have been shown to broaden ensemble dispersion, while also reducing model error that results from the failure to adequately represent subgrid-scale processes (Berner et al. 2011). Therefore, in addition to the runs described above, four runs for each of the GEFSR IC/BCs used the stochastic ensemble kinetic energy backscatter scheme and stochastic perturbation to physical tendencies. SKEBS is a technique that backscatters subgrid-scale energy onto the u - and v -component winds and temperature by randomly perturbing the temperature and streamfunction (Berner et al. 2011), while SPPT perturbs parameters within the physics schemes (Palmer et al. 2009). These members are referred to as the SKEBS+SPPT ensemble. For SKEBS+SPPT, the Thompson microphysical scheme and MYNN2 planetary boundary layer and surface layer scheme will be used in accordance with the parameterization schemes currently used by the operational HRRR. Thus, an additional 20 members with stochastic perturbations was applied using the same physics schemes. The SKEBS and SPPT setup was the same as described as the two WRF ensemble case studies of snowbands in Greybush et al. (2017), in which more details about this approach and parameters used can be found. They found the best results when IC/BCs were perturbed with the stochastic perturbations.

Each WRF member used the following domain setup and run times. The outermost domain is at 18-km grid spacing (Fig. 4). There are inner 6- and 2-km nests centered on the region of snow banding using one-way nesting. Numerous tests of sensitivity to grid spacing were completed prior to deciding on this particular WRF setup for the 26–27 November 2014 case as described in section 2e.

d. Synoptic setup

This section provides some background on the 26–27 November event, and then the other cases validated in

this study are briefly summarized. The frontogenesis in this paper uses the 2D Petterssen (1936) form [as in Novak et al. (2004), their Eq. (1)]. At the time of peak precipitation banding at 0500 UTC 27 November, a 998-hPa surface cyclone was located just east of the Northeast U.S. coast on the east side of this jet (Fig. 5a), with the cyclone located on the east side of this jet. The jet and cyclone was associated with an upper-level trough over the eastern United States, with a 700-hPa low center over southwestern Maine (Fig. 5b). Observed multibands had developed during the last few hours across southern Maine (Fig. 6a), and the multiband objects identified in MODE for the best member are in Fig. 6b. Figure 5c shows cross section A–A' through these bands at 0500 UTC using the RAP analysis, which illustrates weak frontogenesis along a well-defined frontal zone sloping from the surface up to 600 hPa to the northwest, while there is a layer of weak stability or conditional instability [negative saturation moist potential vorticity (MPV*)] between 600 and 500 hPa.



FIG. 4. WRF domain setup, with outermost domain at 18-km grid spacing, nest at 6 km, and innermost nest at 2 km. The 2-km nest is used for all verification unless otherwise specified.

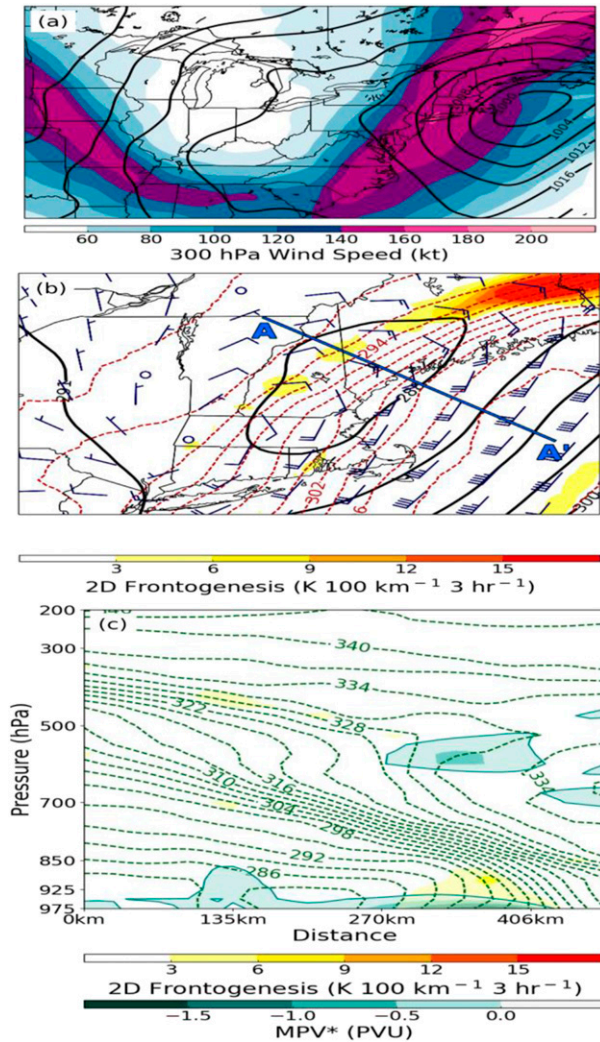


FIG. 5. (a) The 300-hPa wind speed (kt; 1 kt $\approx 0.51 \text{ m s}^{-1}$) and mean sea level pressure (hPa) at 0500 UTC 27 Nov 2014. (b) The 700-hPa geopotential height (dam), potential temperature (K), 2D frontogenesis [$K (100 \text{ km})^{-1} (3 \text{ h})^{-1}$], and wind vectors (kt) at same time. (c) Cross section across A–A' indicated in (b) showing 2D frontogenesis [yellow to orange shaded; $K (100 \text{ km})^{-1} (3 \text{ h})^{-1}$], saturated equivalent potential temperature (green dashed; K), and saturation potential vorticity (MPV*; PVU; 1 PVU = $10^{-6} \text{ K kg}^{-1} \text{ m}^2 \text{ s}^{-1}$, with negative in blue shades).

The 15–16 February 2014 banded event across eastern Massachusetts featured the deepest cyclone of the four events (minimum sea level pressure of $\sim 982 \text{ hPa}$; not shown). It rapidly deepened over the Gulf Stream and was located $\sim 300 \text{ km}$ southeast of the banding region in the left-exit region of the upper-level jet. There was a sloping region of frontogenesis [$>15 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$] above the banded region with conditional instability or CSI between 800 and 600 hPa as indicated by a region of $\text{MPV}^* < 0$. In contrast, the surface cyclones were weaker for the other three events (~ 1000 , ~ 1002 , and $\sim 1000 \text{ hPa}$

for the 26–27 November 2014, 7–8 January 2017, and 9–10 December 2017 events, respectively). The 26–27 November 2014 event initially featured the strongest frontogenesis [$>15 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$ over a much larger region] but this frontogenesis weakened considerably by the time multibands became the dominant precipitation mode. The surface cyclone for this event was not located in an upper-jet entrance or an exit region. For the 7–8 January 2017 event, the surface cyclone was located over coastal North Carolina, about 500 km southeast of the region of banded snowfall and in the right entrance region of the upper jet. There was weak, patchy CSI in approximately the 550–450-hPa layer above a region of frontogenesis [$\sim 6\text{--}9 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$] sloping upward to the west from about 750 to 600 hPa. As for the 26–27 November 2014 event, the surface cyclone was not located in either the jet entrance or exit region.

e. Sensitivity to grid spacing

For one of the cases (26–27 November 2014) a series of WRF sensitivity tests were completed to determine what grid spacing to use for the ensemble. For these set of sensitivity WRF runs the outer two domains were placed the same as in Fig. 4, but the inner most nest was centered over Maine using the same size as in Fig. 4. The three domain setups have grid spacings of (12, 4, and 1.33 km), (9, 3, and 1 km), and (18, 6, and 2 km). An additional 0.67-km domain was nested within the 2-km domain (boundaries about 50-km on from the sides of the 2-km domain). A convective scheme (CP) was run on the 18-, 12-, 9-, and 6-km domains. Although 6 km is not ideal for a CP scheme, we found that $<10\%$ of the precipitation in the cyclone comma head is produced by the domains using a CP scheme. Thus, the structural differences and number of bands within the comma head are mainly the result of model resolution, not partitioning between CP and explicit precipitation.

Figure 7 shows results for forecast hour 23 (0500 UTC 27 November). The 12-, 9-, and 6-km domains (Figs. 7a,d,g, respectively) are inadequate to explicitly resolve multibands. However, an increase in reflectivity intensity is observed at 9- and 6-km grid spacing (Figs. 6d,g) as compared to 12-km. The 4-km grid (Fig. 7b) lacks definition in the bands seen in the 3- and 2-km domains. (Fig. 7e,h). The innermost domain shows even more improvement in resolving the observed banded structures in Figs. 7c, 7f, and 7i. The finer 2-km grid spacing (Fig. 7h) shows more improvement in resolving multibands than the 3-km output (Fig. 7e), with at least three long, discrete bands. The largest improvement occurs with the 12-km run on the outer domain and 1.33-km inner nest (Fig. 7c), with several

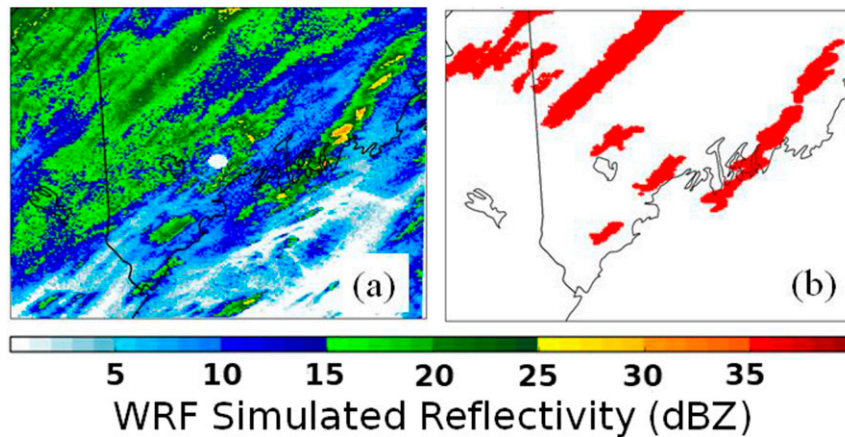


FIG. 6. (a) Observed reflectivity for 0459 UTC 27 Nov 2014. (b) MODE objects from (a) with the multibands colored red.

multibands resolved at this time. However, changes in how the model resolves multibands at 1 km (Fig. 7f) versus at 3 km (Fig. 7e) are limited as well as 2 km (Fig. 6h) versus at 0.67 km (Fig. 7i). Therefore, WRF run at 2-km grid spacing presents a reasonable grid spacing for resolving multibands without incurring unnecessary computational cost.

This experiment also illustrates how differences in the larger-scale (parent) domain can potentially impact the higher-resolution band predictions. The 12-, 9-, 6-km outer domains have different precipitation distributions, with the 12 km having less precipitation over the western part of the plotted domain while the 6 km has more precipitation. Presumably, this is the result of changes in the regional-scale forcing and moisture as a result of the different parent domain resolutions. As a result, there are more robust bands at 2-km grid spacing with the 18- and 6-km outer domains than the 1.33-km domain nested within the 12- and 4-km domains. This illustrates that the ambient conditions modified by the outer domain may be just as important as decreasing the grid spacing in changing the band predictions. A more detailed understanding of these domain differences on snowband development requires additional case simulations and analysis, which is beyond the scope of this paper.

3. WRF ensemble band verification

a. Band number verification

Figure 8 presents the band count by hour for each of the four cases. For the 2-km WRF it is constructed by taking the mean number of bands for each of the 40 members and the 12 MODE permutations for each member, while the observed is the mean number of its 12 MODE permutations. The observed mean frequency

is also shown as reference. For the 26–27 November 2014 and 7–8 January 2017 cases (Figs. 8b,d), there are distinct 5–6-h windows of peak multibanded number, while for the other two cases the temporal distribution is more flat (Figs. 8a,c). The observed peak is well defined also for the 26–27 November case and about the same time as the WRF, and in general for all cases and most times there are more multibands than what is simulated by WRF.

To quantify some of the accuracy in the 2-km WRF, hits, misses, false alarms, and correct nulls are defined here in accordance with a standard 2×2 verification contingency table (Barnes et al. 2009), such that a probability of detection (POD) and false alarm rate can be defined. Since timing of mesoscale features is challenging, the verification required the observed and simulated bands to occur from one hour before to one hour after. All 12 MODE permutations were used in the analysis. The POD for all four cases combined ranges from 0.32 to 0.69 for four- and two-band thresholds, respectively (not shown), and it reaches as high as 0.92 for at least two bands in the 26–27 November case. The FAR ranges over 0.12, 0.03, and 0 for the four-, three-, and two-band thresholds (not shown), but there are variations between the cases. The 15 February 2015 case has a false alarm rate of 0.60 and 0.13 for the four- and three-band thresholds, respectively.

Multiband count error was defined as the number of multibands in one WRF member at one forecast hour, minus the number of observed multibands at the corresponding verification time averaged for all 12 MODE permutations. This difference was then integrated across the forecast hour window used for each case and across all WRF ensemble members. Thus, every instance of a count error corresponds to a unique WRF member hour. For all four cases combined, the distribution of multiband count error exhibits a negative bias (Fig. 9a). The median count bias is -1 , the 10th percentile of the

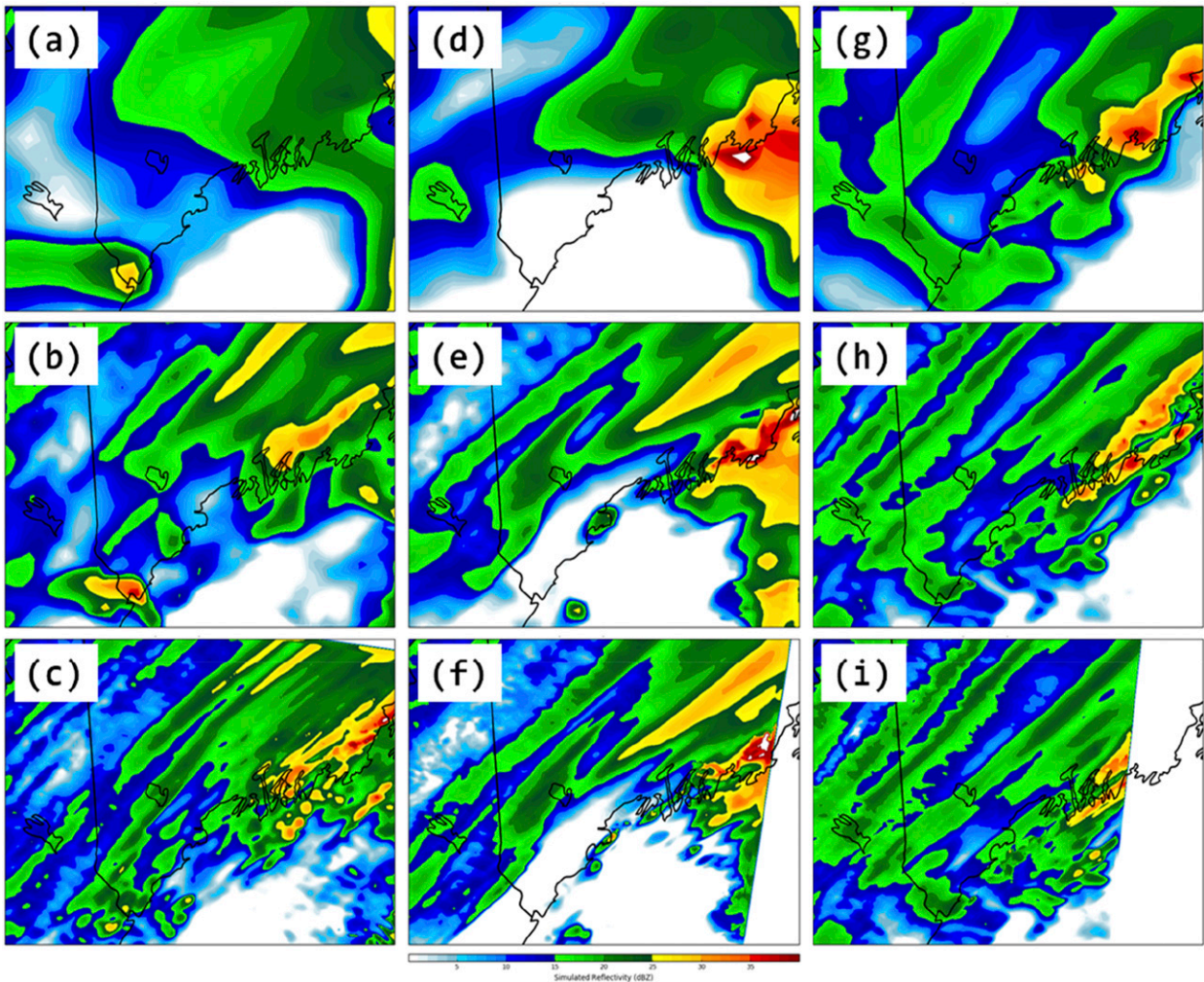


FIG. 7. WRF-simulated reflectivity at 2 km AGL at forecast hour 23 (0500 UTC 27 Nov 2014) for the (a) 12-km domain, (b) 4-km nest within (a), and (c) at 1.33-km nest within (b). (d)–(f) As in (a)–(c), but for the 9-, 3-, and 1-km domains. (g)–(i) As in (a)–(c), but for the 6-, 2-, and 0.67-km domains.

distribution is -4 , and the 90th percentile was $+1$. There is a tail toward a negative bias all four cases as well, with the mean bias between -0.44 and -1.38 . Figure 10 illustrates that all four cases when plotted separately also have a negative count bias, with much of the negative count errors ranging from one to three bands. Although the analysis in this study combines results for all 12 MODE permutations, it is worth noting that this bias toward too few bands is consistent regardless of MODE permutation (Table 3).

b. Bandwidth and length

Biases in length and width were calculated for each case in the same way as the multiband count error, except that the mean of a given attribute (e.g., length or width) was computed for a given WRF member hour. For example, if a WRF member at hour 24 features

four multibands, the lengths of those four were averaged together. That one number was then used to compare with the average of all of the radar observed band objects at that same time. Times when either the observations or the WRF ensemble member did not feature any multibands were dropped from these computations. For all four cases combined using all 12 MODE permutations, WRF exhibits a slight positive length bias (Fig. 9b), with a mean and median error of 5.31 and 2.54 km, respectively. The observed median length for these bands is 37.5 km, with a 90th percentile of 78.7 km and 10th percentile of 21.5 km. However, there is relatively large variability in errors between cases, with two of the four cases having median error near zero (Fig. 11). The same is true for the width errors (Figs. 9c), which have little overall bias and the cases range from slightly positive (2–4 km) to slightly

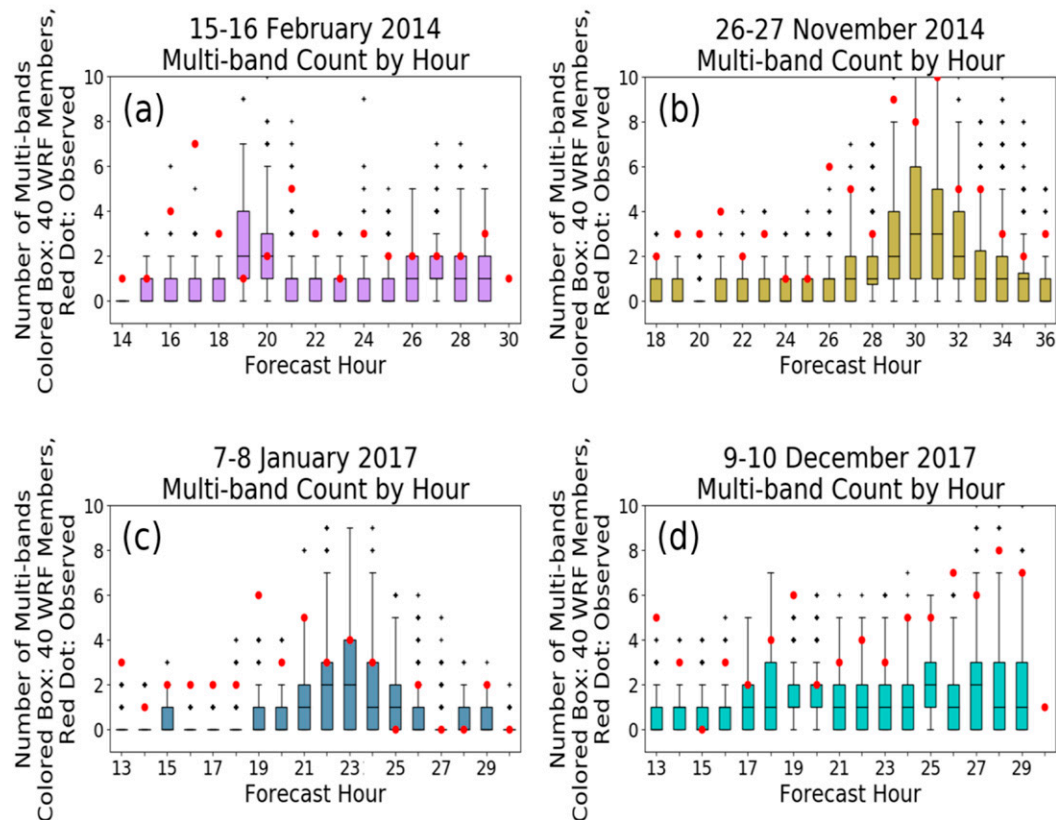


FIG. 8. Multiband count at each forecast hour for the WRF ensemble (shaded box-and-whisker plots using all 40 members and 12 MODE permutations) and observed multi-band count for the mean of the 12 MODE permutations (red dot) for (a) 15–16 Feb 2014, (b) 26–27 Nov 2014, (c) 7–8 Jan 2017, and (d) 9–10 Dec 2017. The box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data. Outlier points are those past the end of the whiskers.

(from -3 to 0 km) negative (Fig. 12). The observed median width is 13.2 km. The 90th percentile is 24.9 km and the 10th percentile is 8.7 km. The more positive width and length errors tend to be with those WRF permutations with the larger convolution radius, and many of these tend to be the best members ($\sim 14\%$ best for the 8 -km radius versus 3% – 5% for the 2 -km radius). Unlike for count error, multiband length and width errors are not consistent across all 12 MODE permutations. WRF would exhibit a positive length or width bias with some MODE permutations and a negative one with others (Table 3).

c. Band placement

Multiband placement within the verifying domain was also examined. The centroid latitude and longitude for each identified object was given in the MODE output; it was averaged together for each WRF member hour for those objects meeting the criteria defining multibands. The WRF ensemble performs fairly well in locating the multibands in the north–south direction (not shown). The distribution

of multiband mean centroid error for all four cases combined is nearly normal about 0 with a median bias of just $+0.03^\circ$ latitude. This near-zero bias may result from averaging of signs across the four cases. Three of the cases actually exhibit a negative centroid latitude bias, while the 15–16 February 2014 case exhibits a median centroid bias of $+0.20^\circ$ latitude (~ 22.2 km). The WRF ensemble performs worst for the 9–10 December 2017 case, with a centroid latitude error of -0.29° latitude (~ 32.2 km).

The mean multiband centroid longitude exhibits a positive (eastward) bias for all four cases combined with a wide distribution (Fig. 9d). The median error is $+0.22^\circ$ longitude to the east (~ 32 km, assuming a latitude of 40°N). The 10th percentile is -0.85° longitude, while the 90th percentile is $+1.17^\circ$. The same three individual cases that exhibit a negative (southward) latitude bias also have a positive (eastward) longitude bias (Fig. 13), while 15–16 February 2014 exhibits a negative (westward) longitude bias. The WRF ensemble performs worst for 9–10 December 2017, with a median

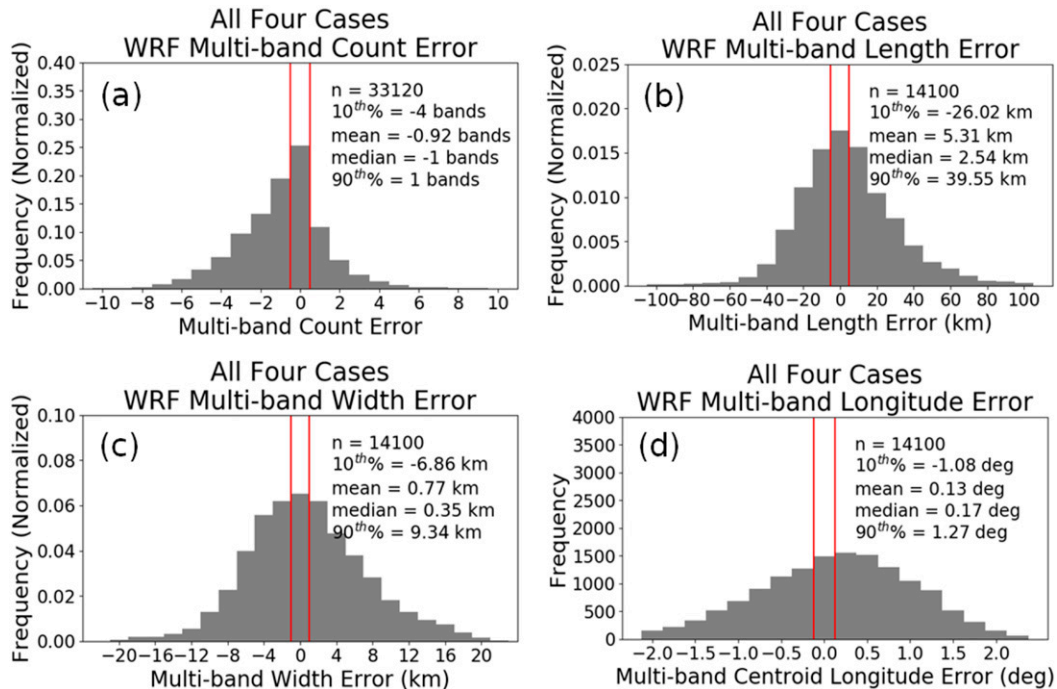


FIG. 9. Histogram showing errors (model minus observed) for band (a) count binned every 1, (b) length (km) binned every 10 km centered at 0 (e.g., from -5 to 5 km), (c) width (km) binned every 2 km centered at 0 (e.g., from -1 to 1 km), and (d) longitude ($^{\circ}$) binned every 0.25° centered at 0 (e.g., from -0.125° to 0.125°) for all four cases combined and all 12 MODE permutations. The term n is the product of WRF members \times forecast hours \times 12 MODE permutations \times 4 cases.

centroid bias of -0.53° longitude (~ 77 km at $\sim 40^{\circ}$ N). It performs best for 26–27 November 2014, with a median centroid bias of $+0.07^{\circ}$ longitude (~ 11 km at 43° N).

d. Comparison of multiband count error in WRF subensembles

The results above represent the full 40-member WRF ensemble. The 20-member SKEBS+SPPT ensemble was compared with the 20-member IC+physics ensemble. Individual microphysics and PBL schemes within the IC+physics ensemble were also compared with each other. For each ensemble group there is no statistically significant difference in multiband count error. The mean multiband count error using or not using the SKEBS is -1.88 and -1.91 , respectively. The p value of the two-sample t test is 0.80, indicating only a 20% confidence that WRF run with either SKEBS on or SKEBS off more adequately produces multibands when compared to the other. Similarly, the mean multiband count errors for the MYNN2 and YSU PBL scheme are -1.92 and -1.97 , respectively. The p value of the two-sample t test is 0.72. The mean multiband count errors for the Thompson and Morrison MP scheme are -1.94 and -1.84 , respectively, and the p value of the two-sample t test is 0.47. While the two microphysical schemes show

the greatest difference, that difference is still not statistically significant. Finally, the same test was completed for each subensemble of WRF with the initial and lateral boundary conditions varied among the five GEFS Reforecast perturbations used in the ensemble. There is no statistically significant difference found in multiband count error for any combination of two of the IC/BC members (e.g., GEF5R1 and GEF5R5). Overall, this suggests that the bias in band count underprediction highlighted above is not dependent on the ICs and physics used in this study.

4. Physical processes

This section explores some of the physical processes for the banded versus nonbanded WRF members. We compared the basic ingredients for precipitation (lift, moisture, and stability), with low- to midlevel frontogenesis as the potential lifting source. Banded times are those member hours that have at least two multibands as well as the two hour period before the initial formation for that same member hour. These extra times are included because it is believed that the environment immediately before multiband genesis is just as important as at the time of genesis. Likewise, nonbanded member hours are those with fewer than two multibands at that hour and in

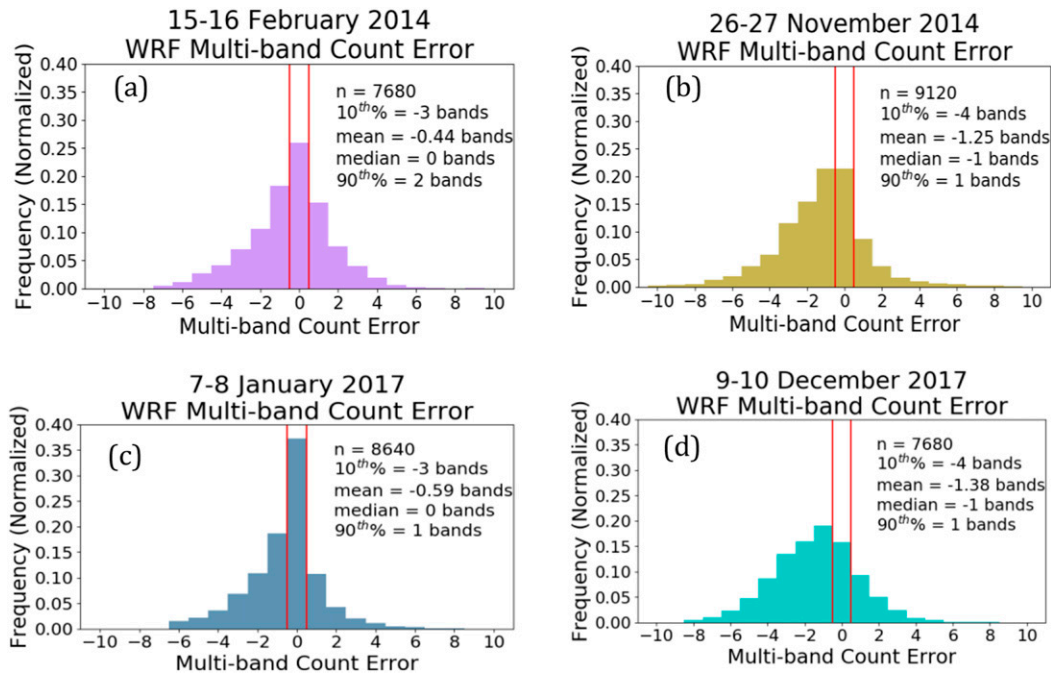


FIG. 10. As in Fig. 9a, but band count errors are separated for the individual cases: (a) 15–16 Feb 2014, (b) 26–27 Nov 2014, (c) 7–8 Jan 2017, and (d) 9–10 Dec 2017.

the hour before and hour after for a given member. This is done so as not to double count pregenesis member hours as nonbanded times and thus smooth out the difference signal.

Not all forecast hours are used. Instead, only member hours considered to be within the window of peak banding for each case are used. This filtering was done so that the beginnings and ends of each event would not bias the nonbanded statistics away from a faithful comparison to the banded time statistics. Multiband counts by hour (Fig. 7) were used for selecting the window of peak banding. The windows are forecast hours 18–21, 28–33, 21–26, and 17–28 for 15–16 February 2014, 26–27 November 2014, 7–8 January 2017, and 9–10 December 2017, respectively. The following results pertain only to member hours within each case's peak banded window. For sections 4a and 4b, some of the processes are plotted relative to the band in the center of the plot, while section 4c shows vertical

profiles of some band ingredients for a small region around the band.

a. Frontogenesis

Frontogenesis at 700 hPa was chosen because it is generally the level of maximum forcing across the cases (Novak et al. 2010). To obtain a smoother frontogenesis field not contaminated by perturbations around the band the outer 18-km WRF domain was used around each centroid. An exceedance probability is calculated for the banded versus the nonbanded members. For all four cases except 26–27 November 2014, a threshold of $5.0 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$ is chosen because it is the threshold that illustrates a useful amount of spatial variability in the probabilities. For 26–27 November 2014, the threshold is increased to $10.0 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$.

For all four cases the overall pattern is characterized by a band of maximum exceedance probability values oriented southwest to northeast (Fig. 14),

TABLE 3. Range in mean values from all 12 MODE permutations for the indicated multiband attribute for each case. Bold font indicates ranges that span 0.

	Multiband count error	Multiband length error (km)	Multiband width error (km)
15–16 Feb 2014	[−0.85, −0.09]	[−0.23, 14.45]	[−0.10, 3.59]
26–27 Nov 2014	[−1.80, −0.55]	[−8.46, 7.28]	[−1.21, 2.47]
7–8 Jan 2017	[−0.81, −0.30]	[6.63, 16.39]	[−2.73, 1.62]
9–10 Dec 2017	[−1.75, −0.93]	[−11.10, 15.42]	[−2.03, 1.54]

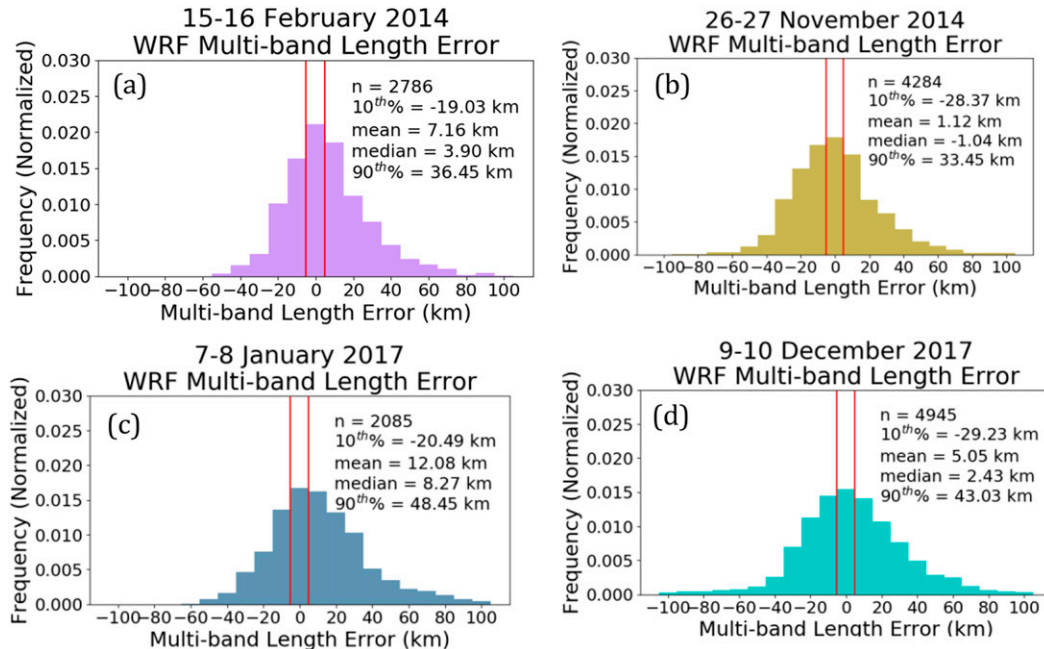


FIG. 11. As in Figs. 9b and 10, but band length errors (km) for the four individual cases.

centered on or near the mean object centroid. For 15–16 February 2014 (Figs. 14a,b), exceedance probabilities are ~10%–20% greater for nonbanded times than for banded times. The difference between the banded and nonbanded member-hour subsets is less clear for 26–27 November 2014 (Figs. 14c,d). For 7–8 January 2017 (Figs. 11e,f) and 9–10 December 2017 (Figs. 14g,h), the exceedance probabilities for the banded times are slightly greater than for nonbanded times. Probability values are much lower for the January and December 2017 cases, and are applied to a lower threshold of frontogenesis as discussed previously, indicating that the overall magnitude of frontogenesis is less for both of these cases than for 26–27 November 2014. Overall, although there is some suggestion that banded times have more frontogenesis, it is not true for all events and not statistically significant. This is consistent with Ganetis et al. (2018), which showed that frontogenesis may not be the dominant forcing for these multibands.

b. Moisture, stability, and vertical shear

The 2-km nest was used to obtain the moisture, stability, and vertical shear environment that produced the band. Differences in the spatial patterns of relative humidity at banded and nonbanded times are small (not shown), though there is some case-to-case variability. The largest difference between banded and nonbanded times occurs with 15–16 February 2014 (not shown), in which the probability of relative humidity exceeding 90% at 850 hPa is ~70%–90% for nonbanded times over

the southern half of the domain, but only ~50%–70% over the same area for banded times.

Conditional and potential instability is so scarce in all four cases that it proves a poor delineator between banded and nonbanded times. No portion of the domain reaches 10% conditional instability for either the banded or nonbanded group of member hours (not shown). Only the 26–27 November case has 10%–20% for potential instability during the nonbanded times over the southeastern corner of the domain. Probabilities of CSI (calculated using full wind and saturated equivalent potential temperature as in Ganetis et al. 2018) are <10% everywhere for both banded and nonbanded times for both 15–16 February 2014 (not shown) and 7–8 January 2017 (not shown). Probabilities are only slightly greater for 26–27 November 2014 because there are a few scattered points of >10% probability for nonbanded times.

Figure 15 shows the probability of the vertical change in saturated potential temperature $< 2 \text{ K km}^{-1}$ for a 50-hPa layer centered around 600 hPa. The 15–16 February 2014 case has the greatest probability (least stable), and there is little difference between the banded versus nonbanded members (Figs. 15a,b). The other three cases more stable (lower probabilities) and the banded cases have slightly more probability (10%–30%) than the nonbanded members (0%–10%).

Figure 16 shows the probability of vertical shear in a 50-hPa layer around 900 hPa (typical low-level frontal

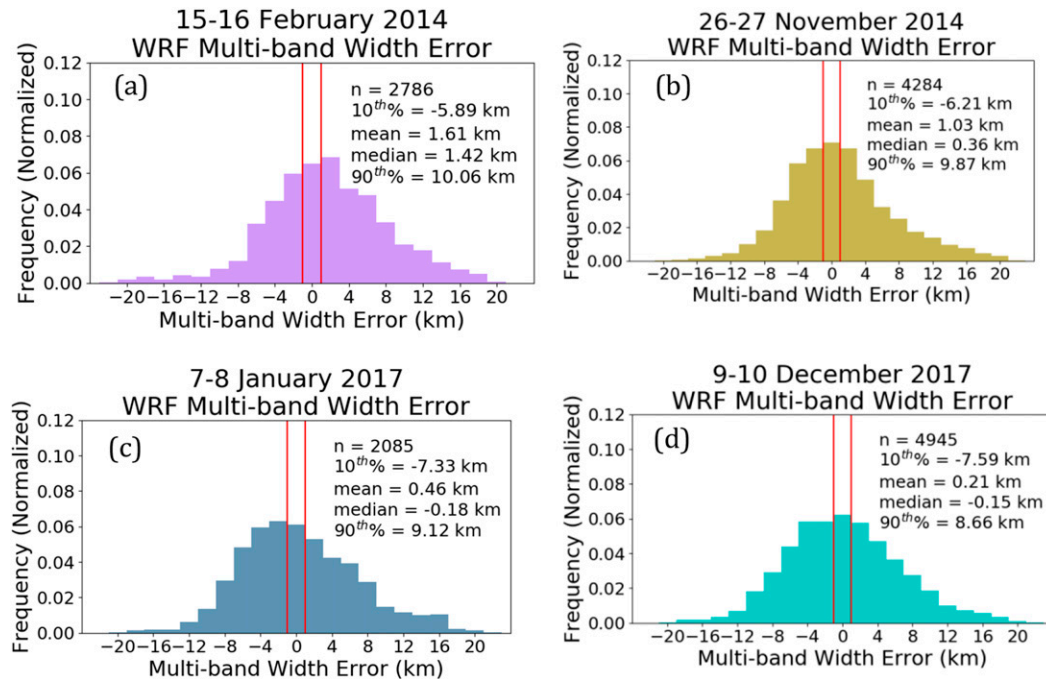


FIG. 12. As in Figs. 9c and 10, but bandwidth errors (km) for the four individual cases.

zone layer) greater than 10^{-3} s^{-1} . The 7–8 January 2017 has the largest shear and the banded members have somewhat greater probabilities. The other cases have little shear that meet the threshold around the band centroid location.

c. Vertical error profiles

The WRF vertical profiles of the ambient conditions surrounding the bands were compared with the observed. The RAP analysis was used to validate because of

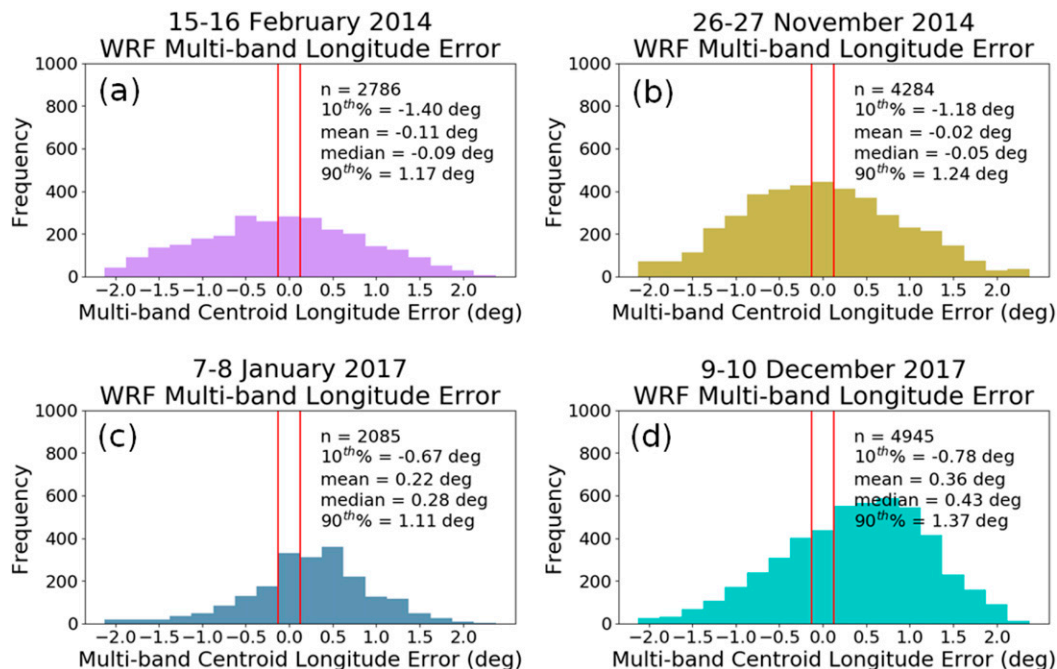


FIG. 13. As in Figs. 9d and 10, but band longitude errors ($^{\circ}$) for the four individual cases.

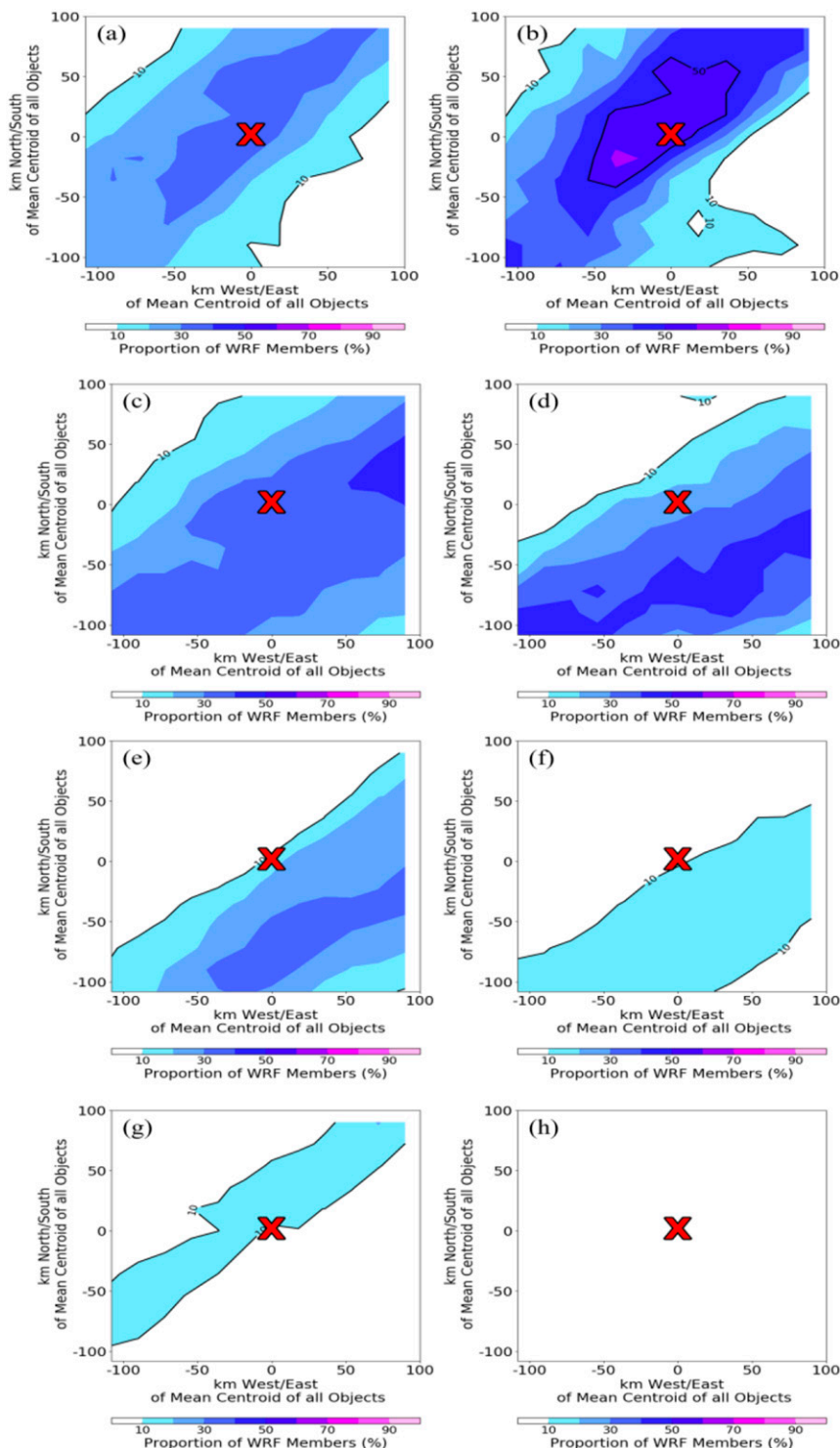


FIG. 14. Probability of 700-hPa WRF frontogenesis from the 18-km domain centered around the band centroid (km) exceeding $5.0 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$ for the (a) banded and (b) nonbanded member hours for 15–16 Feb 2014. (c),(d) As in (a) and (b), but for 26–27 Nov 2014 and 700-hPa frontogenesis exceeding $10.0 \text{ K (100 km)}^{-1} (3 \text{ h})^{-1}$. (e),(f) As in (a) and (b), but for 7–8 Jan 2017. (g),(h) As in (a) and (b), but for 9–10 Dec 2017.

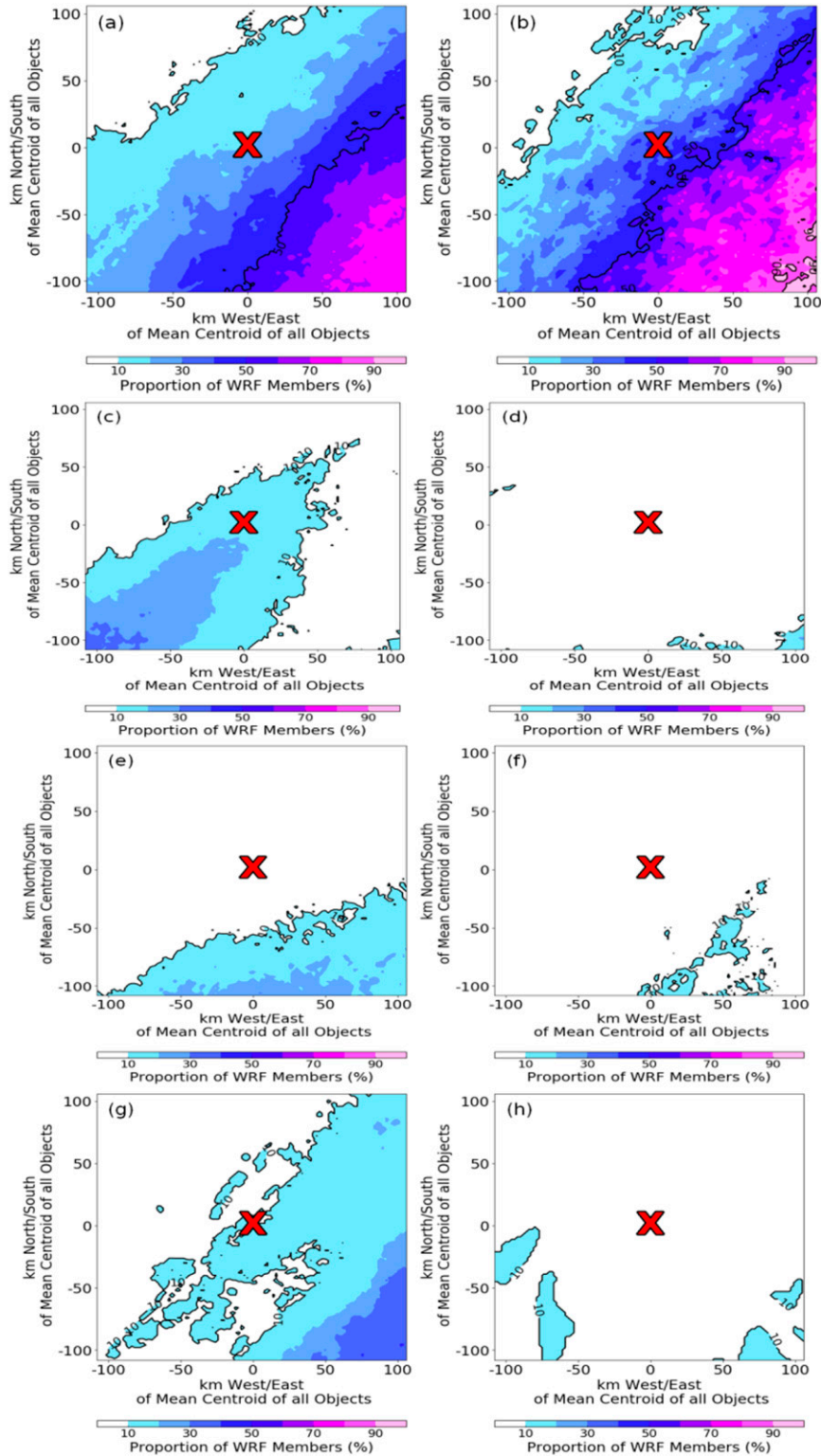


FIG. 15. Probability of d/dz of saturated equivalent potential temperature in a 50-hPa-thick layer centered on 600 hPa less than 2.0 K km^{-1} for (a) banded and (b) nonbanded member hours for 15–16 Feb 2014. (c),(d) As in (a) and (b), but for 26–27 Nov 2014. (e),(f) As in (a) and (b), but for 7–8 Jan 2017. (g),(h) As in (a) and (b), but for 9–10 Dec 2017.

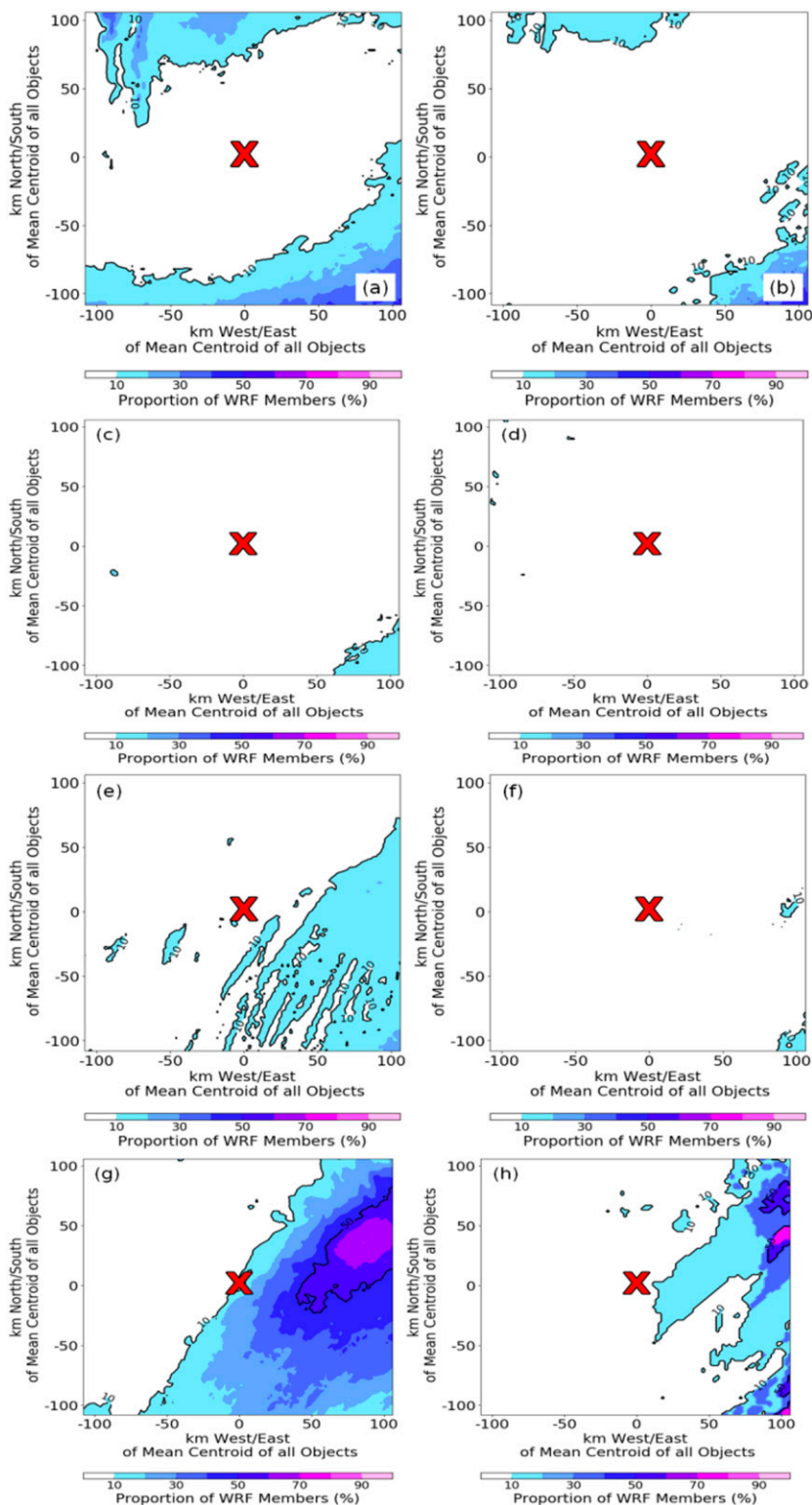


FIG. 16. Probability of du/dz in a 50-hPa-thick layer centered on 900 hPa being greater than 10^{-3} s^{-1} for (a) banded and (b) nonbanded member hours for 15–16 February 2014. (c),(d) As in (a) and (b), but for 26–27 Nov 2014. (e),(f) As in (a) and (b), but for 7–8 Jan 2017. (g),(h) As in (a) and (b), but for 9–10 Dec 2017.

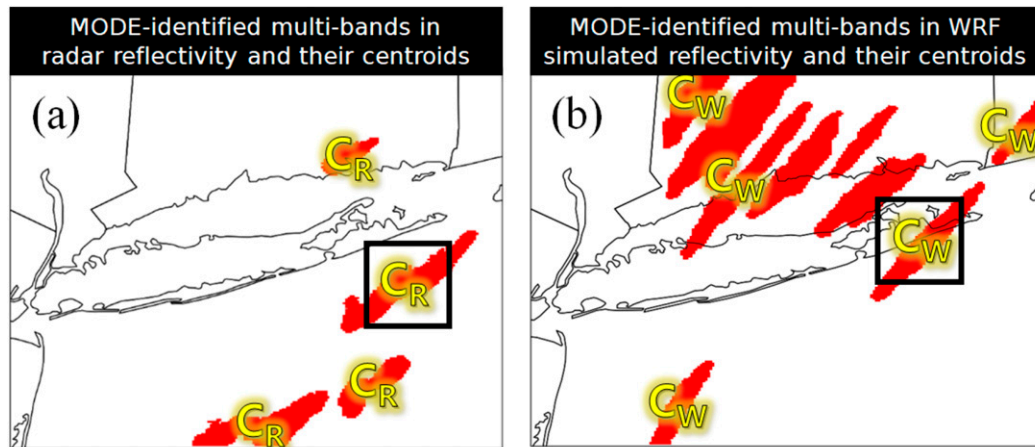


FIG. 17. MODE-identified multibands in (a) observed radar reflectivity and (b) WRF-simulated reflectivity for an arbitrary time. The label C_R in (a) denotes multiband centroids (in this example, four) in observed radar reflectivity. The label C_W in (b) denotes multiband centroids (in this example, five of eight are denoted) in WRF-simulated reflectivity. Each C_R in (a) is used as the center of 27.09-km-side boxes used for averaging values from the RAP field at those locations at the given time (one example box shown). Each C_W in (b) is the center of 28-km-side boxes used for averaging values from the WRF field at those locations at the given time (one example box shown). The average of the four averages in (a) is used as the single value to represent the magnitude of a given physical quantity in the observations at the given time. This is also done for the eight averages in (b). The single RAP “average of averages” value is subtracted from the single WRF “average of averages” value to give one single WRF error value for the given WRF member for the given time.

its finer-scale grid spacing (13.545 km) than other analyses such as the North American Regional Reanalysis (NARR, 32 km; Mesinger et al. 2006) or the Climate Forecast System Reanalysis (CFSR, ~50 km, Saha et al. 2010).

To assess how errors in WRF correspond to multiband count error, WRF member hours were grouped by their multiband count error into two groups: 1) those member hours producing two or more too *few* multibands, and 2) those member hours producing one or more too *many* multibands. For each of these groups, and for each case, vertical profiles are constructed for frontogenesis, the vertical gradient of saturation equivalent potential temperature, and the vertical gradient of the u component of the wind. Error is defined as WRF minus RAP, where the values used for both WRF and RAP is as follows. For a given time, all of the multiband centroids are identified, indicated as C_R for those in the observed field and C_W in the WRF modeled field in Figs. 17a and 17b, respectively. A box is drawn around each centroid that is 9 grid points (27.09 km \times 27.09 km) for the observed field, based on the RAP’s 13.545-km grid spacing, and 225 (15 each side) grid points (28 km \times 28 km) for the WRF field, based on WRF’s innermost nest’s grid spacing of 2 km. The 2-km grid is used for stability and vertical shear, while for frontogenesis the grid point from the 18-km domain on either side of the band centroid are used. Values are calculated for a given variable and averaged within that box for all of the boxes. These

averages are then again averaged together to produce a single representative value for a given variable at a given time in each of the RAP field and WRF field. This single value for RAP is subtracted from this single value for WRF, giving the WRF error for that member hour. These errors are then appended to lists for each of the overbanded and underbanded groups of WRF member hours, and plotted together on a vertical box-and-whisker plot, with overbanded on top (“over”) and underbanded on bottom (“under”) for each level, every 50 hPa, for each case. A red star indicates that the difference between the over- and underbanded lists of errors is statistically significant at that pressure level at a 95% confidence interval using a bootstrap resample 10 000 times.

Figure 18a shows the frontogenesis error plotted for all four cases combined. Difference in frontogenesis error distributions for over and underbanded WRF member hours are statistically significant at nearly every pressure level from 950 to 650 hPa. In general, the underbanded member hours have a positive frontogenesis error while the overbanded member hours have an error near zero. This result suggests that a stronger frontogenesis than observed may favor consolidation of bands with the deformation and thus the underprediction.

Stability errors are determined using the vertical gradient in saturation equivalent potential temperature

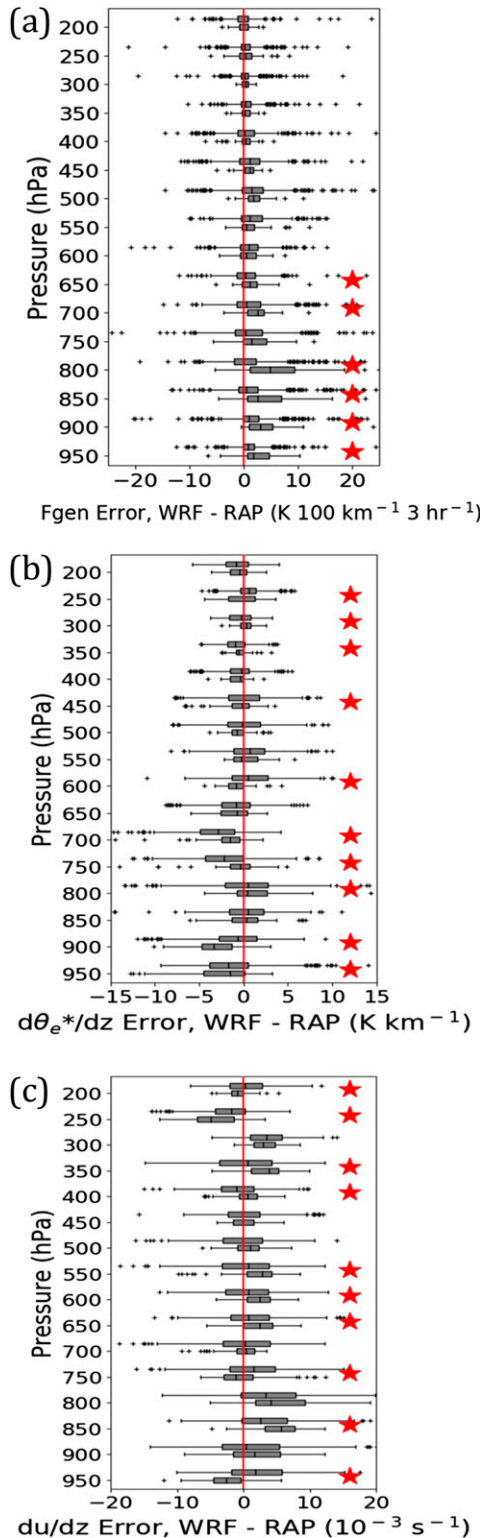


FIG. 18. Vertical profile of error for all four cases combined in (a) frontogenesis (from the 18-km WRF domain), (b) change in saturated equivalent potential temperature with height ($d\theta_e^*/dz$) error centered across a 50-hPa-thick layer in the 2-km grid, and (c) du/dz error from the 2-km grid. Overbanded (underbanded)

(Fig. 18b). Differences between over- and underbanded distributions are statistically significant at 950 and 900 hPa, again at 800–700 hPa, and at 600 hPa. However, the sign of the difference is not consistent at all of these pressure surfaces. For example, at 900 hPa, underbanded member hours have a distinct negative stability error (WRF too unstable) and overbanded member hours have a stability error that is near zero. Conversely, at 750 and 700 hPa, the overbanded member hours have more negative stability error and the underbanded member hours have stability error closer to zero. This inconsistency among pressure surfaces would be consistent with these bands forced by parcels lifting below 750–700 hPa, and not below 900 hPa.

Shear errors were calculated using the vertical gradient in the u component of the wind. Vertical shear is examined because it has been hypothesized that shear-induced waves in the boundary layer may grow upscale and help organize precipitation into multibands (Ganetis 2017). Difference in du/dz error distributions for over- and underbanded WRF member hours are statistically significant at 950, 850, 750, and 650–550 hPa. At 950 hPa the overbanded member hours have a notable positive error in du/dz , while the underbanded member hours exhibit a notable negative error at the same pressure surface. This suggests that vertical shear in the boundary layer may be an important feature in the multiband formation process. However, just like for stability errors, other pressure surfaces have the opposite signal. At 850 hPa, underbanded member hours have the more positive shear error. The signal at 750 hPa is the same as at 950 hPa, except not as pronounced. Boundary layer shear may be worth investigating further, but its limited vertical extent leaves the overall role of vertical shear unclear.

5. Conclusions

This study uses a 40-member WRF ensemble on four selected snowstorm cases between 2014 and 2017 to explore the ability of high-resolution models in simulating small-scale precipitation bands (multibands) within the extratropical cyclone comma head along the Northeast U.S. coast. Half of the ensemble varied PBL

← error is displayed as the upper (lower) box-and-whisker plot for each pressure level. The box-and-whisker plots are constructed the same as in Fig. 8, with plus signs indicating statistically determined outliers. Red stars at the right indicate that the difference between the over- and underbanded distributions of WRF error is statistically significant at that pressure level at a 95% confidence interval performed after a bootstrap resample with $n = 10000$.

and microphysical parameterization schemes, while the other half introduced a stochastic kinetic energy backscatter scheme (SKEBS) and stochastic physics perturbations (STTP). Both halves of the ensemble were applied to initial and boundary conditions from five members of the GEFS reforecast. For the 26–27 November 2014 event over western Maine there is little benefit using a horizontal grid spacing less than 2 km, so 2 km is used in this study. There is sensitivity to the band number and intensity to the predictions the outer domain resolutions, which illustrates that the ambient conditions modified by the parent domain are just as important as decreasing the grid spacing.

A feature-based verification is applied to hourly WRF reflectivity fields from each ensemble member and the WSR-88D radar reflectivity at 2-km height above sea level. The Method for Object-Based Diagnostic Evaluation (MODE) tool is used for identifying multibands in both observed radar reflectivity and WRF-simulated reflectivity fields. Multibands are defined as two or more bands that are 5–20 km in width and that also exhibit a $>2:1$ aspect ratio. Given the uncertainty in identifying bands 12 MODE permutations were run for each forecast hour, each WRF member, and the observed radar reflectivity field. These permutations were comprised of these three convolving disc radii and four reflectivity thresholds.

The 2-km WRF systematically underpredicts the occurrence of multibands for all four cases, with the POD ranging from 0.69 at a two-band threshold to 0.32 at a four-band threshold for all four cases combined when defining hits and misses using a ± 1 -h time window. The median multiband count error is shown to be either -1 or -2 for each case analyzed, and -2 for all four cases combined. The observed median length for these bands is 37.48 km, with a 90th percentile of 78.72 km and 10th percentile of 21.53 km. For all four cases combined using all 12 MODE permutations, WRF exhibits a slight positive length bias, with a mean and median error of 5.31 and 2.54 km, respectively. The observed median width is 13.23 km, and there is little overall bias. There is little position error in the north–south direction for the centroid band position, while there is a slight positive (eastward) bias for all four cases combined ($+0.22^\circ$ longitude).

Some of the physical processes in the model are explored to understand some of the band predictions using RAP analyses. Although there is some suggestion that banded member times have slightly more frontogenesis, it is not true for all events and it is not statistically significant. Differences in the spatial patterns of relative humidity for banded versus nonbanded members are small, and stability also proves to be a poor delineator

between banded and nonbanded times. Vertical profiles of the errors illustrate that the underbanded members have slightly more frontogenesis, which suggests that additional deformation may be consolidating some of the bands, and there may be more boundary layer shear for the overbanded members. However, the source for the band underprediction requires additional hypotheses, and it is a motivation for future work using field data. For example, Hoban (2016) and Hoban et al. (2017) showed that gravity waves often coexist with these multibands, suggesting that this could be a potential forcing mechanism for these bands.

Acknowledgments. This work was supported by the National Science Foundation (AGS-1347499) and NOAA NGGPS program (NA15NWS4680003). We thank the two anonymous reviewers for their comments that helped improve this manuscript.

REFERENCES

- Barnes, L. R., D. M. Schultz, E. C. Grunfest, M. H. Hayden, and C. C. Benight, 2009: CORRIGENDUM: False alarm rate or false alarm ratio? *Wea. Forecasting*, **24**, 1452–1454, <https://doi.org/10.1175/2009WAF2222300.1>.
- 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>.
- Bergeron, T., 1950: Über der mechanismus der ausgeibigen Niederschläge. *Ber. Dtsch. Wetterdienstes*, **12**, 225–232.
- Berner, J., S. Ha, J. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, <https://doi.org/10.1175/2010MWR3595.1>.
- Coniglio, M. C., J. Correia, P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, <https://doi.org/10.1175/WAF-D-12-00103.1>.
- Davis, C., B. 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>.
- Gallus, W., and J. Bresch, 2006: Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts. *Mon. Wea. Rev.*, **134**, 2632–2641, <https://doi.org/10.1175/MWR3198.1>.
- Ganetis, S. A., 2017: The role of thermodynamic processes in the evolution of single and multibanding within winter storms. Ph.D. thesis, State University of New York at Stony Brook, 259 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>.

- Grell, G. A., and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14**, 5233–5250, <https://doi.org/10.5194/acp-14-5233-2014>.
- Greybush, S. J., S. Saslo, and R. Grumm, 2017: Assessing the ensemble predictability of precipitation forecasts for the January 2015 and 2016 East Coast winter storms. *Wea. Forecasting*, **32**, 1057–1078, <https://doi.org/10.1175/WAF-D-16-0153.1>.
- Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, <https://doi.org/10.1175/BAMS-D-12-00014.1>.
- Hoban, N. P., 2016: Observed characteristics of mesoscale banding in Coastal Northeast U.S. snow storms. M.S. thesis, Dept. of Marine, Earth, Atmospheric Sciences, North Carolina State University, 66 pp.
- Hoban, N., and Coauthors, 2017: Observed characteristics of mesoscale banding in coastal Northeast U.S. snow storms. *28th Conf. on Weather Analysis and Forecasting/24th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 133, <https://ams.confex.com/ams/97Annual/webprogram/Paper308387.html>.
- Hobbs, P. V., and J. D. Locatelli, 1978: Rainbands, precipitation cores and generating cells in a cyclonic storm. *J. Atmos. Sci.*, **35**, 230–241, [https://doi.org/10.1175/1520-0469\(1978\)035<0230:RPCAGC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<0230:RPCAGC>2.0.CO;2).
- Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, <https://doi.org/10.1175/MWR3199.1>.
- Jankov, I., W. Gallus, M. Segal, and S. Koch, 2007: Influence of initial conditions on the WRF-ARW model QPF response to physical parameterization changes. *Wea. Forecasting*, **22**, 501–519, <https://doi.org/10.1175/WAF998.1>.
- Jones, M., B. Colle, and J. Tongue, 2007: Evaluation of a mesoscale short-range ensemble forecast system over the Northeast United States. *Wea. Forecasting*, **22**, 36–55, <https://doi.org/10.1175/WAF973.1>.
- Martin, J., 1998: The structure and evolution of a continental winter cyclone. Part II: Frontal forcing of an extreme snow event. *Mon. Wea. Rev.*, **126**, 329–348, [https://doi.org/10.1175/1520-0493\(1998\)126<0329:TSAEOA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0329:TSAEOA>2.0.CO;2).
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.
- Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, <https://doi.org/10.1175/2008MWR2556.1>.
- Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada Level-3 Model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, <https://doi.org/10.1007/s10546-005-9030-8>.
- Nicosia, D., and R. Grumm, 1999: Mesoscale band formation in three major northeastern United States snowstorms. *Wea. Forecasting*, **14**, 346–368, [https://doi.org/10.1175/1520-0434\(1999\)014<0346:MBFITM>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0346:MBFITM>2.0.CO;2).
- Norris, J., G. Vaughan, and D. Schultz, 2014: Precipitation banding in idealized baroclinic waves. *Mon. Wea. Rev.*, **142**, 3081–3099, <https://doi.org/10.1175/MWR-D-13-00343.1>.
- 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. Aiyer, 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>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp.
- Petterssen, S., 1936: Contribution to the theory of frontogenesis. *Geophys. Publ.*, **11** (6), 1–27.
- Rauber, R. M., and Coauthors, 2017: Finescale structure of a snowstorm over the northeastern United States: A first look at high-resolution HAIPER cloud radar observations. *Bull. Amer. Meteor. Soc.*, **98**, 253–269, <https://doi.org/10.1175/BAMS-D-15-00180.1>.
- Rosenow, A. A., D. M. Plummer, R. M. Rauber, G. M. McFarquhar, B. F. Jewett, and D. Leon, 2014: Vertical velocity and physical structure of generating cells and convection in the comma head region of continental winter cyclones. *J. Atmos. Sci.*, **71**, 1538–1558, <https://doi.org/10.1175/JAS-D-13-0249.1>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, <https://doi.org/10.1175/2010BAMS3001.1>.
- Schumacher, R. S., and A. J. Clark, 2014: Evaluation of ensemble configurations for the analysis and prediction of heavy-rain-producing mesoscale convective systems. *Mon. Wea. Rev.*, **142**, 4108–4138, <https://doi.org/10.1175/MWR-D-13-00357.1>.
- Shields, M. T., R. M. Rauber, and M. K. Ramamurthy, 1991: Dynamical forcing and mesoscale organization of precipitation bands in a Midwest winter cyclonic storm. *Mon. Wea. Rev.*, **119**, 936–964, [https://doi.org/10.1175/1520-0493\(1991\)119<0936:DFAMOO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<0936:DFAMOO>2.0.CO;2).
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- Stark, D., S. E. Yuter, and B. A. Colle, 2013: Observed microphysical evolution for two East Coast winter storms and the associated snow bands. *Mon. Wea. Rev.*, **141**, 2037–2057, <https://doi.org/10.1175/MWR-D-12-00276.1>.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, <https://doi.org/10.1175/2008MWR2387.1>.
- Xu, Q., 1992: Formation and evolution of frontal rainbands and geostrophic potential vorticity anomalies. *J. Atmos. Sci.*, **49**, 629–648, [https://doi.org/10.1175/1520-0469\(1992\)049<0629:FAEOFR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<0629:FAEOFR>2.0.CO;2).