Sources of Uncertainty in Precipitation-Type Forecasting

HEATHER DAWN REEVES, KIMBERLY L. ELMORE, ALEXANDER RYZHKOV, TERRY SCHUUR, AND JOHN KRAUSE

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

(Manuscript received 9 January 2014, in final form 4 April 2014)

ABSTRACT

Five implicit precipitation-type algorithms are assessed using observed and model-forecast sounding data in order to measure their accuracy and to gauge the effects of model uncertainty on algorithm performance. When applied to observed soundings, all algorithms provide very reliable guidance on snow and rain (SN and RA). However, their skills for ice pellets and freezing rain (IP and FZRA) are comparatively low. Most misclassifications of IP are for FZRA and vice versa. Deeper investigation reveals that no method used in any of the algorithms to differentiate between IP and FZRA allows for clear discrimination between the two forms. The effects of model uncertainty are also considered. For SN and RA, these effects are minimal and each algorithm performs reliably. Conversely, IP and FZRA are strongly impacted. When the range of uncertainty is fully accounted for, their resulting wet-bulb temperature profiles are nearly indistinguishable, leading to very poor skill for all algorithms. Although currently available data do not allow for a thorough investigation, comparison of the statistics from only those soundings that are associated with long-duration, horizontally uniform regions of FZRA shows there are significant differences between these profiles and those that are from more transient, highly variable environments. Hence, a five-category (SN, RA, IP, FZRA, and IP–FZRA mix) approach is advocated to differentiate between sustained regions of horizontally uniform FZRA (or IP) from more mixed environments.

1. Introduction

One of the greatest difficulties facing forecasters during the cool season is the accurate determination of precipitation type (Ralph et al. 2005). The importance of this is obvious as some forms of winter precipitation can pose a threat to human health and safety and/or disrupt travel and commerce. Freezing rain is particularly dangerous because of its ice-loading effects on power wires and its threat to travel safety. Correct specification of precipitation type is also critical for reliable streamflow forecasts (e.g., Fassnacht et al. 2006) and for the accurate prediction of low-level temperature and dewpoint following the onset of precipitation (Kain et al. 2000; Lackmann et al. 2002). There are multiple potential causes for inaccurate predictions of precipitation type. Herein, deficiencies in the choice of algorithm and effects from uncertainty in the analysis–forecast system are addressed.

There are four primary types of precipitation observed at ground level in winter storms: snow,[1] rain, ice pellets, and freezing rain (SN, RA, IP, and FZRA, respectively). These forms can occur in isolation from one another or in mixes. Mixes are thought to be the consequence of larger hydrometeors only partially melting while smaller hydrometeors completely melt in the melting layer (Stewart and King 1987; Zerr 1997; Thériault et al. 2010). Partial melting or freezing may also affect whether the precipitation type at the ground is IP or FZRA. To aid in more efficient prediction, most modeling systems employ one or more algorithms that provide the user with a two-dimensional gridded analysis of precipitation type. These algorithms take one of

Corresponding author address: Heather Dawn Reeves, NOAA/National Severe Storms Laboratory, Ste. 2401, 120 David L. Boren Blvd., Norman, OK 73072-7319.
E-mail: heather.reeves@noaa.gov

DOI: 10.1175/WAF-D-14-00007.1

[1] Snow can be subdivided into multiple categories including snow pellets, dry snow, and wet snow. Since neither the observing network nor the algorithms used herein distinguish between these forms, we consider only one broad category of snow.

two forms. Explicit algorithms, such as that used in the Rapid Refresh model (Ikeda et al. 2013), employ forecasts of hydrometeor mixing ratios and temperature to determine the precipitation type. These algorithms have the potential for more accuracy, given that the precipitation rate (e.g., Thériault et al. 2010), downstream transport of hydrometeors (Petch and Dudhia 1998), and hydrometeor interactions (Thériault et al. 2012) are accounted for. However, causes of systematic errors are difficult to diagnose as observations of hydrometeor mixing ratios are not routinely taken.

Implicit algorithms use environmental profiles of temperature and/or humidity to infer the precipitation type. Both the North American Mesoscale Model and Global Forecast System (Moorthi et al. 2001; Janjić et al. 2005) use implicit algorithms. There are only a handful of publications assessing implicit algorithm performance. Manikin et al. (2004) and Manikin (2005) found that some algorithms have substantial biases toward one or another form of precipitation. Wandishin et al. (2005) found that different algorithms perform better in different situations and that ''none of the...algorithms can be discarded as universally inferior.'' Yet, these studies use model forecast, rather than observed, temperature and relative humidity (biases in model predictions of either could affect algorithm performance) and none identify the causes of errors. Perhaps a better route for determining the efficacy of a given algorithm is to use observed data. In this way, causes of systematic errors can be isolated from model biases. Such a study was performed in Bourgouin (2000). The algorithms they considered gave fairly similar results, but only 46 soundings were examined and they did not explain the causes for errors.

Another limiting factor to the accurate prediction of precipitation type is the uncertainty in the model analysis/forecast system. Several authors find the uncertainty in short-range forecasts of temperature and dewpoint ranges from 0.5 to 4 K in the lower troposphere (e.g., Colle et al. 2003; Stensrud and Yussouf 2003; Jones et al. 2007; Coniglio et al. 2010). Given that changes as little as 0.5 K are known to alter the precipitation type (Thériault et al. 2010), it is logical to question whether the above degree of model uncertainty precludes the meaningful use of implicit algorithms. Indeed, Wandishin et al. (2005) found some variations in precipitation type as model initial conditions and physical assumptions are changed. But the degree to which precipitation type is dependent on model uncertainty remains an open question.

The purpose of this paper is twofold. First, we assess the accuracy of precipitation type from five different implicit algorithms using observed soundings (section 2). Second, we gauge the effects of model uncertainty using the Rapid Update Cycle (RUC; Benjamin 1989) analyses and forecasts (section 3). Concluding thoughts are provided in section 4.

2. Precipitation-type algorithms

There are five precipitation-type algorithms considered in this study. Four of these are used operationally by the National Centers for Environmental Prediction. These are the Baldwin1, Baldwin2, Bourgouin, and Ramer algorithms (Ramer 1993; Baldwin et al. 1994; Bourgouin 2000). The last is the algorithm presented in Schuur et al. [2012; hereafter referred to as the (National Severe Storms Laboratory) NSSL algorithm]. In this section, we provide brief descriptions of the algorithms, perform a statistical assessment of how individual algorithms behave, consider two case-study examples, and attempt to isolate sustained, horizontally uniform FZRA events from the average.

a. Description of algorithms

The Baldwin algorithms (hereafter B1 and B2) were derived using numerical model data and observed sounding data (where available) from select storms in Oklahoma and Alabama. In either algorithm, a precipitation-generation layer is first diagnosed. Areas between the wet-bulb temperature T_w profile and the 273- and 269-K isotherms below this layer are then calculated. The algorithms use the size of these areas to assume whether the environment is conducive to melting aloft and refreezing in the surface-based subfreezing layer, if one exists. Algorithms B1 and B2 differ only in the discrimination between SN and IP. In B1 (B2), the area between T_w and the 269- (273-) K isotherm in the warm layer is used. If this area is large enough, the schemes assume hydrometeors start as liquid and a SN classification cannot be made. The categories that B1 and B2 diagnose are SN, RA, IP, and FZRA.

The Bourgouin (BG) algorithm was devised using statistical analyses of 173 observed soundings for different precipitation types (the 46 soundings mentioned in section 1 were used for validation). This algorithm classifies soundings according to the vertical profile of temperature T. Sounding profiles that are completely subfreezing are diagnosed as SN. Sounding profiles that have only one crossing of the 273-K isotherm and a sufficiently deep and warm surface-based warm layer are classified as RA. For other sounding types, the mean T of the elevated warm and surface-based cold layers and their depths are the means of distinguishing between IP, FZRA, and RA. There are three key details of the BG algorithm that are believed to limit its accuracy.

First, it uses T rather than T_w. Hence, any modification of precipitation type that may come about as a result of cooling from evaporation or melting will not be anticipated. Second, no precipitation-generation layer is calculated. Rather, all hydrometeors are assumed to start as frozen. There is evidence to suggest this can lead to the incorrect specification of the precipitation type for FZRA events wherein the cloud-top temperatures are too warm to support the formation of ice crystals (Rauber et al. 2000, 2001). Third, in cases where no clear statistical separation between classes exists, the precipitation type is randomly selected. In this study, the random assignment was discarded in favor of stating that an uncertainty exists and between what forms of precipitation. For the dataset used herein, uncertainty only exists between IP and FZRA.

In the NSSL algorithm, four sounding types are diagnosed using the T_w profile below a precipitation-generation layer. These are 1) a completely subfreezing profile in which SN is diagnosed, 2) a profile with only one crossing of the 273-K isotherm wherein only SN or RA are diagnosed (depending on the height of the crossing), 3) a profile with three 273-K crossings in which only IP or RA are diagnosed, and 4) a profile with two crossings of the 273-K isotherm. In the latter sounding type, IP, FZRA, or an IP–FZRA mix may be diagnosed. For sounding types 3 and 4, the maximum T_w in the elevated warm layer ($T_{w\text{—max}}$) and minimum T_w in the lowest subfreezing layer ($T_{w\text{—min}}$) are used to distinguish between the precipitation types.

The Ramer algorithm was statistically derived using data from observed soundings and numerical model output. It uses a top-down methodology, tracing hydrometeors from a precipitation-generation layer to the surface. The ice fraction I of hydrometeors is calculated using T_w and relative humidity (RH) in each model layer. If $T_w > 266.55$ K, hydrometeors are assumed to start as all liquid and, as with B1 and B2, a SN classification cannot be made. The value of I at the surface and at the precipitation-generation layer are the primary means of distinguishing between SN, RA, IP, and FZRA. For intermediate ranges of surface I, an IP–FZRA mix is diagnosed.

b. Algorithm performance using observed soundings

To assess how the algorithms perform, they are applied to observed soundings for various winter precipitation events. The surface observations come from the Automated Surface Observing Stations (ASOS) at sites located within 35 km of the radiosonde launch locations. The 35-km distance is rather large, especially for highly transient systems wherein the precipitation type undergoes rapid changes. To mitigate this, only soundings

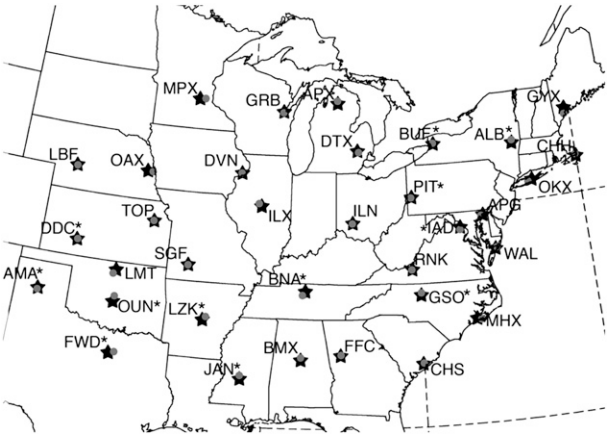

FIG. 1. Locations of ASOS sites (gray circles) and soundings (black stars) used to select events. The asterisks denote those stations that are routinely augmented with human observations.

are used where the reported form of precipitation does not change over the 40-min window surrounding the launch time. Surface observations can be recorded as frequently (infrequently) as 1 min (1 h) during winter precipitation events. Observations are made more frequently when a change in one or more weather types is observed. The 40-km distance is consistent with that used in Ramer (1993) and the time window allows enough time for the radiosonde to ascend through the lower troposphere (Seidel et al. 2011).

All soundings for winter months (October–March) from 2002 to 2013 are considered. An event is defined as any time IP or FZRA are observed coincident with a radiosonde launch for the sites shown in Fig. 1. For each event, all soundings in Fig. 1 for which *any* form of precipitation occurs are collected so that some RA and SN profiles are included. This gives 649 SN, 545 RA, 125 IP, and 422 FZRA soundings. There are fewer IP soundings than for the rest of the categories because IP cannot be automatically detected and, hence, these soundings are only reported at those sites that are augmented with human observations. These sites are indicated in Fig. 1. It is also quite rare for pure IP showers to last more than an hour or two, which further limits the size of this category. No soundings that are associated with mixes are included.

The data were manually quality controlled to remove soundings associated with obvious errors in the precipitation-type report. The most common errors were classifications of RA that occurred in the presence of a deep surface-based subfreezing layer. This is a known deficiency in the ASOS system, which uses the terminal velocity of hydrometeors to distinguish between forms (RA and IP have similar terminal velocities; NOAA 1998). We could find no statistics on typical error rates

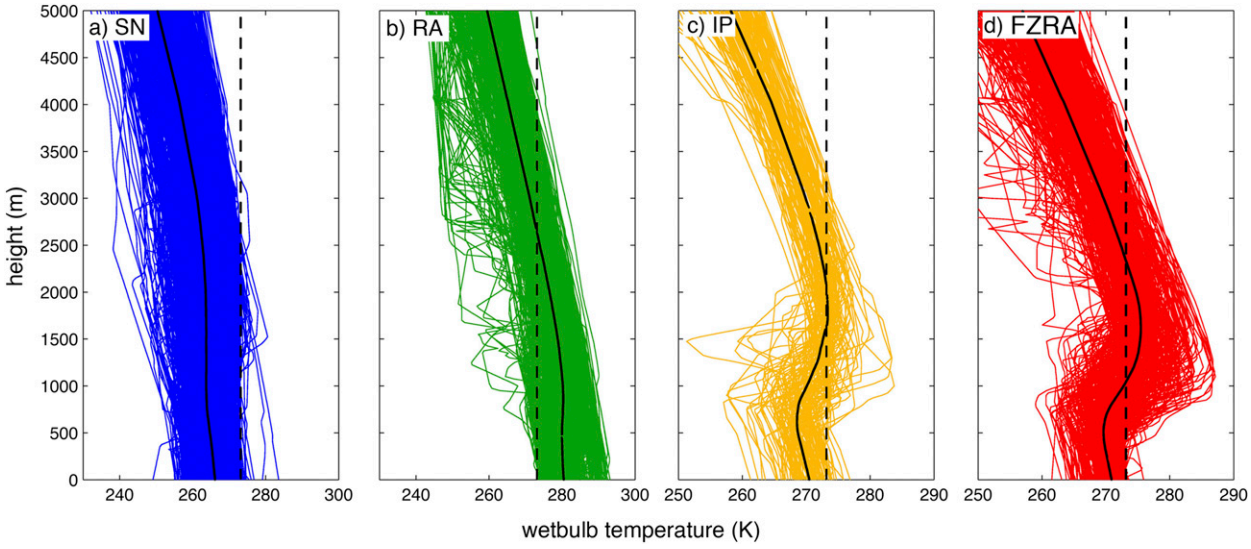

FIG. 2. Observed wet-bulb temperature profiles (colored curves) for (a) SN, (b) RA, (c) IP, and (d) FZRA events. The mean for each category is given as the thick, black contour and the 273-K isotherm is indicated by the dashed, black line. Note the *x* axis in (a),(b) is larger than in (c),(d).

in ASOS for this type of error or any other. It is possible for FZRA to occur for a purely subfreezing profile if the temperatures are not cold enough for ice nuclei to be active (Rauber et al. 2000, 2001). However, neither BG nor NSSL are able to diagnose FZRA in these situations. Therefore, all FZRA soundings that do not have an elevated warm layer are discarded. Also removed are any soundings that failed to ascend to at least 5000 m above ground level.

The T_w profiles for the SN observations have a few soundings (29) with an elevated warm layer and a few cases (32) with surface T_w greater than 273 K, but most are well below subfreezing throughout the troposphere (Fig. 2a). The RA soundings cross the 273-K isotherm anywhere from 50 to 4500 m above ground level and have surface T_w ranging from 273 to 293 K (Fig. 2b). Most of these have lower-tropospheric temperatures well above freezing. The IP and FZRA soundings have an elevated warm layer capping a surface-based subfreezing layer (although 23 of the IP profiles have surface T_w above 273 K; Figs. 2c,d). The primary distinction between the IP and FZRA profiles is the depth and $T_{w\text{—max}}$ of the elevated warm layer; the FZRA mean is warmer and deeper. This is consistent with previous studies that suggest the primary distinction between an IP and FZRA environment is the ability of hydrometeors to completely melt aloft (Brooks 1920; Hanesiak and Stewart 1995; Zerr 1997; Bernstein 2000). The surface-based cold layer also appears to be deeper in the IP profiles. However, bootstrapping tests show these layers are not significantly different (not shown).

The Heidke skill scores (HSSs; Heidke 1926), binned by algorithm and category of precipitation, show that all algorithms do very well with SN and RA, but poorly for IP (Table 1). The scores for FZRA are mixed. B1 and B2 have rather low scores, but BG, NSSL, and Ramer have scores between 0.5 and 0.8. These scores are still lower than those for SN and RA. A comparatively poor performance for IP and FZRA has long been recognized in operational forecasting, thus motivating the use of an ensemble-like approach wherein the precipitation type from all algorithms is computed and the dominant category provided as output to the forecaster (Manikin et al. 2004). The results using the ensemble approach are included in Table 1. It is questionable to state that this approach is helpful for any category of precipitation. The SN, RA, and IP scores are comparable to those from individual algorithms. For FZRA, both BG and Ramer have significantly higher scores. BG, NSSL, and

TABLE 1. HSSs for the different algorithms using observed soundings. In BG, NSSL, Ramer, and Ensemble, the second value corresponds to the score if one assumes the IP–FZ mix (or IP–FZ uncertainty in BG) is a hit.

	SN	RA	IP	FZRA	SFZR
B1	0.864	0.896	0.309	0.369	0.449
B2	0.846	0.896	0.254	0.369	0.449
BG	0.929	0.839	0.332/0.344	0.553/0.592	0.602/0.556
NSSL	0.941	0.876	0.242/0.312	0.478/0.723	0.499/0.463
Ramer	0.846	0.817	0.291/0.296	0.797/0.797	0.821/0.785
Ensemble	0.940	0.917	0.351/0.321	0.577/0.436	

TABLE 2. PODs (%) for the different algorithms using observed soundings. In BG, NSSL, and Ramer, the second value corresponds to the score if one assumes the IP–FZ mix (or IP–FZ uncertainty in BG) is a hit.

	SN	RA	IP	FZRA	IP/FZRA combined	SFZR
B1	86.7	96.1	89.6	28.4	42.4	48.0
B2	97.1	96.1	56.0	28.4	34.7	48.0
BG	92.6	96.1	50.4/60.0	48.8/55.7	56.7	68.0/76.0
NSSL	94.1	96.4	26.4/70.4	40.3/78.9	77.0	56.0/90.0
Ramer	94.9	99.6	25.6/25.6	65.4/66.1	68.7	90.0/90.0

Ramer include an IP–FZRA mix and the HSS is calculated assuming these are hits. (No mixes were used in this study, so these classifications are technically not correct.) Though the number of hits increases when this assumption is made, the number of correct nulls decreases by a larger amount, leading to a decreased HSS (Table 1). Other statistical measures, such as the false alarm ratio, critical success index, etc., yield similar conclusions.

Perhaps more illuminating is the percent of soundings that are correctly identified, or probability of detection (POD). The PODs from each algorithm are provided in Table 2. Again, all do quite well at diagnosing SN and RA. But for IP and FZRA, the PODs are comparatively low. The PODs reveal that, excepting BG, all algorithms favor IP over FZRA or vice versa. One example is B1. It performs remarkably well with IP [a consequence of that algorithm's known IP bias (Manikin et al. 2004)], but has very low PODs for FZRA. If one assumes a mix is a hit, the PODs for BG, NSSL, and Ramer are improved. But, as noted above, these are not strictly correct diagnoses.

c. *Causes for biases in the different algorithms*

The biases in IP and FZRA noted for some schemes lead one to posit whether some improvement can be found by tuning internal thresholds in individual algorithms. Let us consider the B1 and B2 algorithms, which favor IP over FZRA. To distinguish between these forms, both algorithms use the area of the surface-based subfreezing layer. This is defined as the area bounded by the ground, 150 hPa above ground level, the 273-K isotherm on the right, and the environmental T_w on the left. Kernel density distributions (Silverman 1992) of these areas for IP and FZRA reveal that not only do they have significant overlap, both have maximum densities in the region designated as IP, thus explaining the strong IP bias in these algorithms (Fig. 3a). Though IP may form when completely melted hydrometeors refreeze in a sufficiently cold subfreezing layer, conventional understanding indicates that the key distinction between an IP and FZRA environment is the ability of hydrometeors to completely melt in the elevated warm layer (Brooks 1920; Hanesiak and Stewart 1995; Zerr 1997; Bernstein 2000). This secondary process is likely more prevalent because active ice nuclei are already present, whereas in the first process, the ambient temperature must be cold enough for the activation of ice nuclei. B1 and B2 do calculate the area of the elevated warm layer, and kernel density distributions show that FZRA profiles are more heavily weighted toward higher values (Fig. 3b). Nevertheless, there is a great deal of overlap for IP and FZRA; no one threshold can be used to differentiate between the two.

In the BG algorithm, the function $A_{\text{cold}} - 0.66A_{\text{warm}}$, where A_{cold} (A_{warm}) is the area of the surface-based cold (elevated warm) layer, is used to distinguish between IP and FZRA. Kernel density distributions of this function show that the IP and FZRA curves straddle the zone separating IP and FZRA, thus explaining the near 50% PODs in both categories (Fig. 3c). This is a different result than in Bourgouin (2000), but we note that our sounding collection covers a much longer time span and includes many more soundings than in that study.

In NSSL, an IP/FZRA mix is favored over either IP or FZRA. In this algorithm, three conditions are assessed:

$$\begin{aligned} &\text{If} \quad T_{w_{\min}} < 268\,\text{K} \quad \text{and} \quad T_{w_{\max}} < 275\,\text{K} \Rightarrow \text{IP} \\ &\text{else} \quad \text{if} \quad T_{w_{\min}} \geq 268\,\text{K} \quad \text{and} \quad T_{w_{\max}} > 275\,\text{K} \Rightarrow \text{FZRA} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{else} \Rightarrow \text{IP/FZRA mix}. \end{aligned} \tag{1}$$

Kernel density distributions show that most IP soundings fail the first two conditions, leading to a bias toward an IP–FZRA mix (Figs. 3d,e). Most of the FZRA profiles also fail the first two conditions. Hence, most of FZRA soundings also fit into the IP–FZRA-mix category.

The Ramer algorithm does not use bulk properties from the elevated warm or surface-based cold layers to distinguish between IP and FZRA. This algorithm is sensitive to the assumed temperature at which ice nuclei activate ($T_{\text{ice}} = 266.55$ K). Yet, there is disagreement in the literature as to the value of T_{ice} (e.g., DeMott et al. 2011 and references therein). Tests are performed with T_{ice} ranging from 263 to 269 K. The PODs for these experiments show that when T_{ice} is increased, FZRA PODs decrease while IP PODs increase (Fig. 3f). The T_{ice} for which the PODs for both categories are

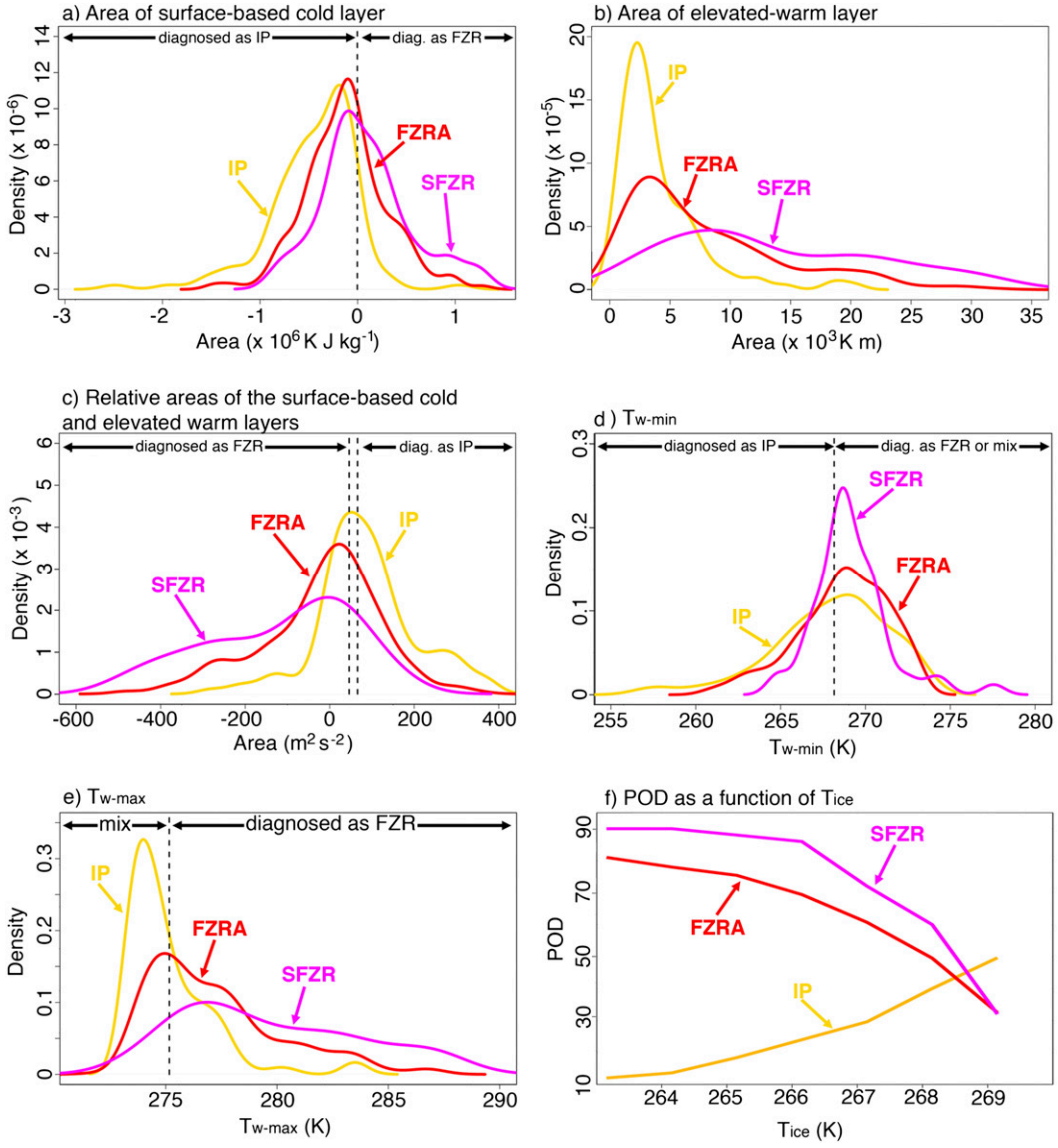

FIG. 3. (a),(b) The kernel density distributions of the areas of the surface-based cold and elevated warm layers as calculated in B1 and B2; (c) the relative areas of the same as calculated by BG (see section 2c); (d),(e) $T_{w\text{—min}}$ and $T_{w\text{—max}}$ as in NSSL; and (f) the POD as a function of T_{ice} from Ramer.

maximized is 268 K. However, the POD for IP never exceeds 50%. Ramer is also sensitive to whether the hydrometeors are assumed to start as all liquid or all frozen. This can affect the precipitation type at the surface [see discussions in Rauber et al. (2000, 2001) for more on this topic]. As a simple test of whether or not it is sensitive to the initial state of hydrometeors, Ramer is rerun for IP first, assuming all hydrometeors start as liquid, then as ice. When the initial I is assumed to be zero for SN, RA, and FZRA, the PODs are nearly identical to those in Table 2, but the IP PODs increase to 49.2%. These results seem counterintuitive for SN. One would ordinarily expect the PODs for SN to decrease with this assumption. However, Ramer assumes all profiles whose maximum $T_w < 266.55$ K are SN, regardless of I. Hence, it is possible to improve this algorithm, albeit through questionable methods.

d. *Precipitation-type distribution according to the different algorithms*

Inspection of plan views of precipitation type sheds light on how the biases of the different algorithms affect an analysis of precipitation type (Fig. 4). These plots are made using the Rapid Refresh (RAP; Brown et al. 2011) 13-km analyses. Only areas where the reflectivity exceeds 0 dBZ are shaded. We also show observations collected as

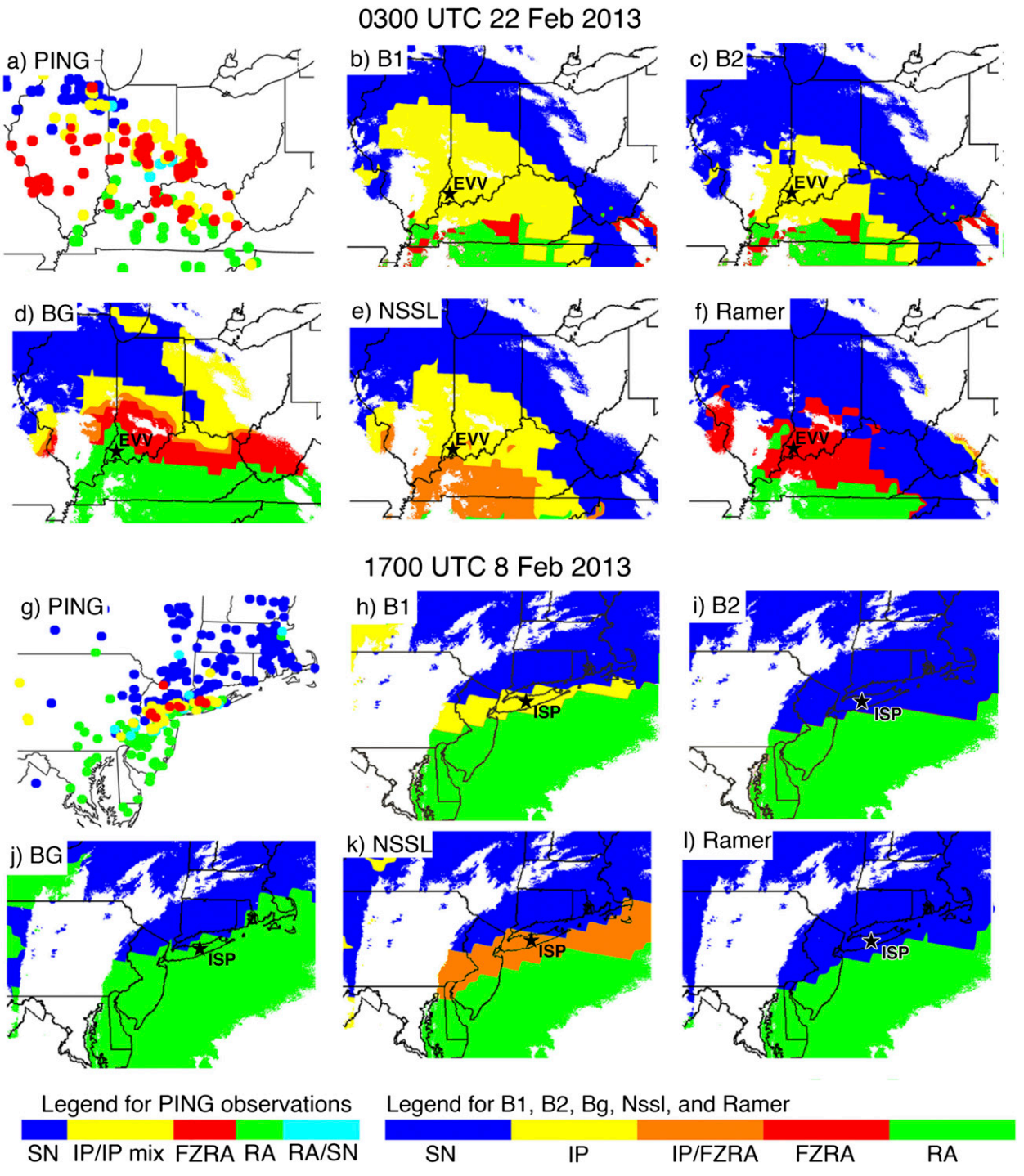

FIG. 4. (a),(g) The mPING observations for a 2-h window surrounding the time indicated for each example and (b)–(f),(h)–(l) the diagnosed precipitation type from each algorithm.

a part of the meteorological Phenomena Identification near the Ground (mPING; Elmore et al. 2014) project. This is a crowd-sourced observation-gathering project wherein observations of precipitation type are collected from anonymous citizens via mobile devices. At present, there are no quality-control protocols for the mPING data, but all mPING data shown in herein were inspected to ensure they are reasonable.

The mPING observations show that for the first event there is a broad region of IP and FZRA over Illinois,

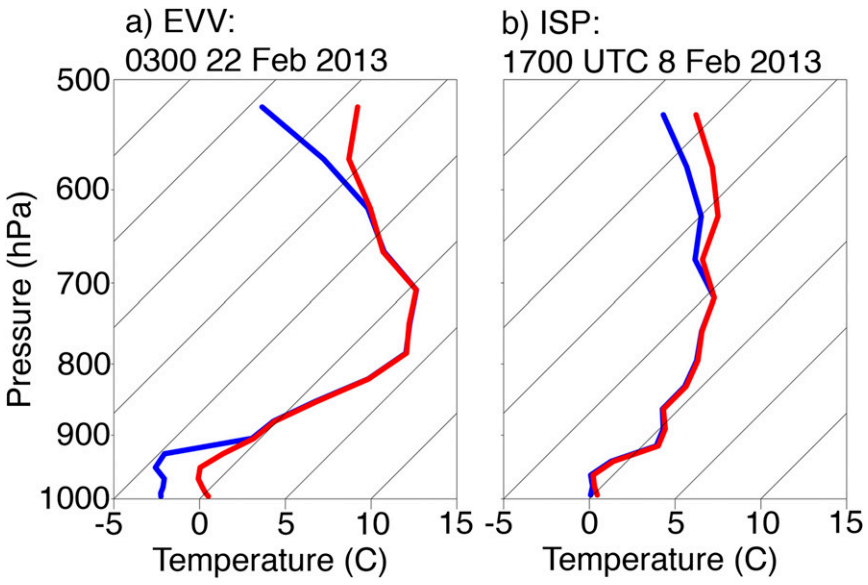

FIG. 5. Skew T–log$p$ diagrams of temperature (red) and dewpoint (blue) at the times and stations indicated at the top of each panel. See Fig. 4 for locations of stations.

Indiana, Ohio, and northern Kentucky (Fig. 4a) while the second has a narrow transition zone of mostly IP and a few FZRA reports over Long Island, New York (Fig. 4g). For both events, B1 has broader regions of IP—a consequence of its strong IP bias (Figs. 4b,h). It also fails to adequately capture the area of FZRA in the first event. The B2 analysis is quite similar, excepting for a smaller (or nonexistent) IP region (Figs. 4c,i). The BG algorithm does a nice job of capturing the existence of an SN–IP–FZRA–RA transition in the first event, but the region over which the transition occurs is too narrow (Fig. 4d). In the second event, BG does not capture the IP–FZRA region at all (Fig. 4j). The NSSL algorithm fails to capture the region of FZRA in the first event, instead diagnosing a broad region of IP and IP–FZRA mix that extends too far south (Fig. 4e). It reasonably captures the IP–FZRA zone in the second event, but has too broad a region (Fig. 4k). Last, Ramer fails to produce any IP for the first event, although one could argue it provides the most accurate depiction of FZRA (Fig. 4f). It also fails to produce IP in the second event (Fig. 4l).

Inspection of RAP soundings at select locations sheds further light on the biases in different algorithms. Consider the sounding at Evansville, Indiana (EVV), at 0300 UTC 22 February 2013 (Fig. 5a). The low-level cold layers in B1 and B2 have an area of -0.997×10^5 K J kg^{-1}, which is sufficient for this sounding to be classified as IP. Though the surface-based warm layer is very shallow, the BG algorithm considers the surface-based cold layer to be too warm with respect to the elevated warm layer to allow for IP. Hence, the only category it can diagnose in this circumstance is RA. This sounding falls into the third sounding class for NSSL. Since its $T_{w\text{—min}}$ is 270.4 K and $T_{w\text{—max}}$ is 276.5 K, it is classified as an IP–FZRA mix. Last, in Ramer, the top of the precipitation generation layer has a $T_w < 266.55$ K, implying that the hydrometeors start as all ice. The elevated warm layer is assumed to be warm enough to melt the hydrometeors completely. However, the surface-based cold layer is insufficiently cold for the activation of ice nuclei (i.e., $T_w > 266.55$ K). Hence, the diagnosis is FZRA.

Now consider the sounding at Islip, New York (ISP), at 1700 UTC 8 February 2013 (Fig. 5b). Recall that the primary distinction between B1 and B2 is the discrimination between SN and IP. In B1, the area of the elevated warm layer that is between the environmental T_w and the 269-K isotherm is used. For this sounding, that area is sufficiently large to merit an IP classification. In B2, the area between T_w and the 273-K isotherm is used. It is clear from Fig. 5b that this area is quite small. Hence, the resulting classification is SN. According to BG, the elevated warm layer is sufficiently large and the low-level cold layer sufficiently small for the algorithm to classify the precipitation type as RA. In NSSL, the sounding fails the first two conditions described in (1), resulting in an IP–FZRA classification. Last, in Ramer, the T_w at the top of the precipitation generation layer is 256 K, indicating that the hydrometeors are assumed to start as frozen. Though the hydrometeors fall through two layers with $T_w > 273$ K, these layers are not sufficiently

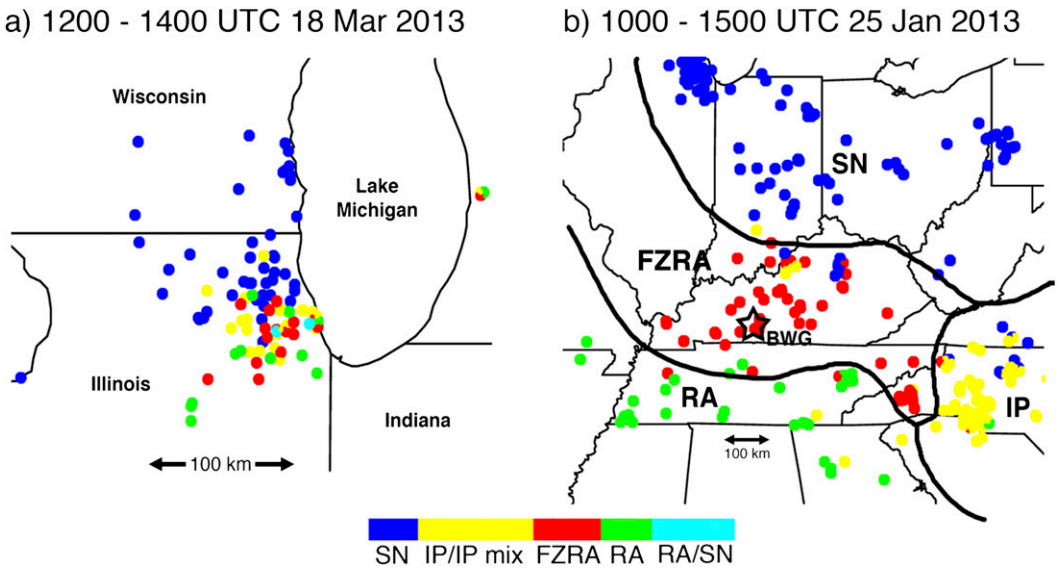

FIG. 6. The mPING observations from (a) 1200 to 1400 UTC 18 Mar and from (b) 1000 to 1500 UTC 25 Jan 2013.

cold or deep to allow for complete melting. This leads to an I at the surface of 0.97. The incomplete melting, initial T_w, and final I combine to yield a classification of SN.

Overall, these results are consistent with Wandishin et al. (2005) in that no one algorithm outperforms the rest. This is particularly true near transition zones. While each algorithm appears capable of detecting IP and/or FZRA transition zones, if they are broad enough, no algorithm performs consistently well at placing the boundaries of these zones leading to the rather low PODs noted in Table 2.

e. *Effects of horizontal nonuniformity in FZRA and IP*

The mPING observations suggest some transitional winter precipitation events have a high degree of horizontal variability. Such an event occurred over Chicago, Illinois, on 18 March 2013 (Fig. 6a). During the 2 h shown, the greater metropolitan area had near-equal numbers of intermingled SN, RA, IP, and FZRA reports. It would be very difficult to draw isopleths on this map. Given that the average radiosonde drift in the lowest half of the atmosphere is 5 km (Seidel et al. 2011), it is unreasonable to expect the algorithms to perform well in this kind of situation. While the mPING observations are made by untrained observers and are not quality controlled, there is sufficient evidence in the literature to suggest that many transitional events have a high degree of variability (Crawford and Stewart 1995; Robbins and Cortinas 2002; Cortinas et al. 2004), consistent with what the mPING observations indicate. Such variations may be due to any one of a variety of smaller-scale forcings, such as heat transfer from unfrozen water bodies, urban heat island effects, varying precipitation rate and drop-size distributions, and whether the hydrometeors start as frozen or unfrozen (Crawford and Stewart 1995; Bernstein 2000; Cortinas 2000; Rauber et al. 2000, 2001; Robbins and Cortinas 2002; Changnon 2003; Cortinas et al. 2004; Thériault et al. 2010). In such circumstances, it may be more desirable to either allow for an IP–FZRA mix or use a three-category algorithm (SN, RA, and IP–FZRA mix). If one were to combine the IP and FZRA profiles from Figs. 2c,d into a single category, this would yield PODs ranging from 35% to 77% (Table 2). (In this exercise, B1 and B2 come out as distinctly inferior to the other algorithms.)

These findings notwithstanding, we are hard pressed to advocate a three-category approach as it does not differentiate between IP and FZRA. This may impede a forecaster's ability to anticipate long-duration, high-ice accumulation FZRA events as readily. It is possible that such events have thermal profiles that are statistically different from the average FZRA profile (Fig. 2d). Consider the event that occurred on 25 January 2013 (Fig. 6b). Over the 5 h shown, the categories of precipitation remain distinct so that clear transitions from one form to the next exist. The precipitation type over Kentucky is mostly FZRA. Ice accumulations reached 0.5 in. (1.3 cm) in parts of the state, according to the National Climatic Data Center's Storm Events Database (NCDC-SED). There is no observed sounding data from central Kentucky for this event, but the profile from the RAP analysis at 1200 UTC 25 January at Bowling Green (BWG) has a deeper and warmer elevated warm

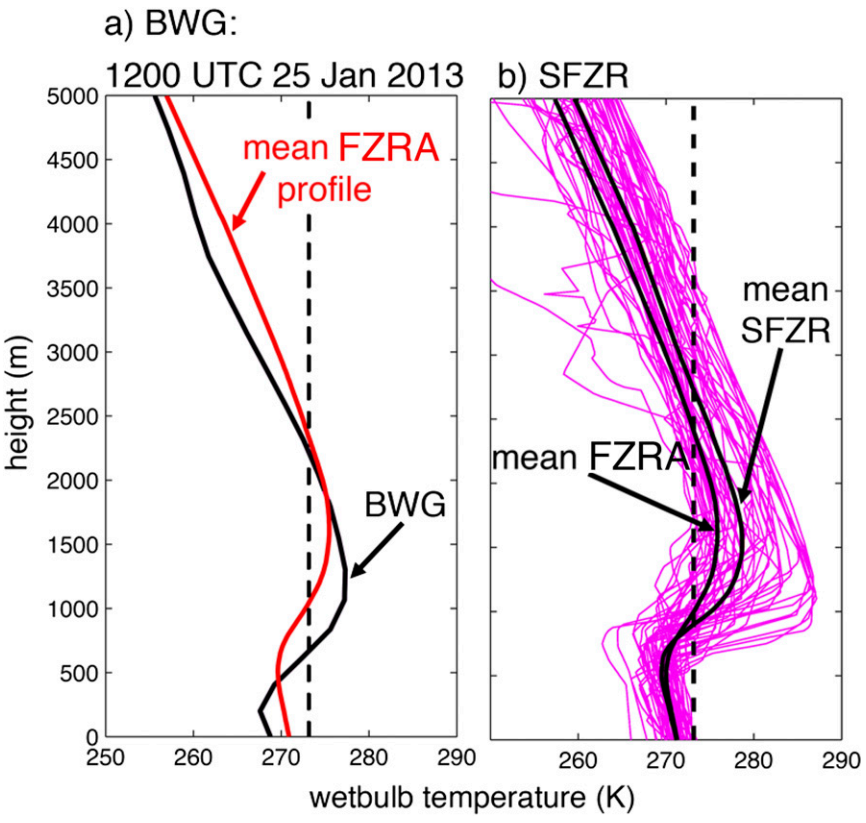

FIG. 7. The (a) RAP T_w profile from BWG (see Fig. 5b for location) at 1200 UTC 25 Jan 2013 and the mean FZRA profile (as in Fig. 2d) and the (b) SFZR profiles (colored curves), their mean, and the mean FZRA profile (indicated in panel).

layer and a shallower surface-based cold layer than the average FZRA profile (Fig. 7a).

Although every effort was made to ensure the ASOS observations used in this study correlate well with the sounding data, it is possible that some or many of the IP and FZRA events occur during conditions more similar to Fig. 6a, where there is a high degree of horizontal variability. Surface observations from the 2002–13 winter seasons are not sufficiently dense for one to isolate soundings that occur in the midst of long-duration, homogeneous areas of FZRA. So, the NCDC-SED was queried and soundings occurring coincident with storms described as having significant [>0.5 in. (1.3 cm)] ice accumulation without rapid transitions of precipitation type identified. There are 51 such soundings, referred to as SFZR. These soundings have slightly warmer and deeper elevated warm layers and shallower cold layers than the FZRA soundings (cf. Figs. 2d and 7b). The mean $T_{w\text{—max}}$ in this collection is more than 3 K higher than in FZRA.

Kernel density distributions for SFZR show the properties of the surface-based cold layer are still not significantly different from IP (Figs. 3a,d). The relative size of the warm and cold layers, as in BG, is slightly different in SFZR (Fig. 3c). This category is also sensitive to the choice of T_{ice} in Ramer (Fig. 3f). However, the greatest difference between SFZR and FZRA is in the characteristics of the elevated warm layer (Figs. 3b,e). Namely, in SFZR, the mean $T_{w\text{—max}}$ is higher than in FZRA. One could more easily assign a threshold that separates the SFZR events from the IP events without significant loss of skill for either. Nevertheless, all algorithms enjoy higher PODs for the SFZR category than for FZRA (Table 2). One can also consider the performance of the algorithms using the HSS (Table 1). As with the POD, the scores increase for the SFZR relative to FZRA, except when one assumes a mix is a hit. As before, this decrease is due to a decreased number of correct nulls. Given the uncertainty in the choice of soundings for the SFZR category (the NCDC-SED is more descriptive than quantitative), we proceed no further in attempting to improve existing classification algorithms or build a new one in this work. However, future work that isolates long-duration IP and FZRA events from more mixed environments is highly recommended.

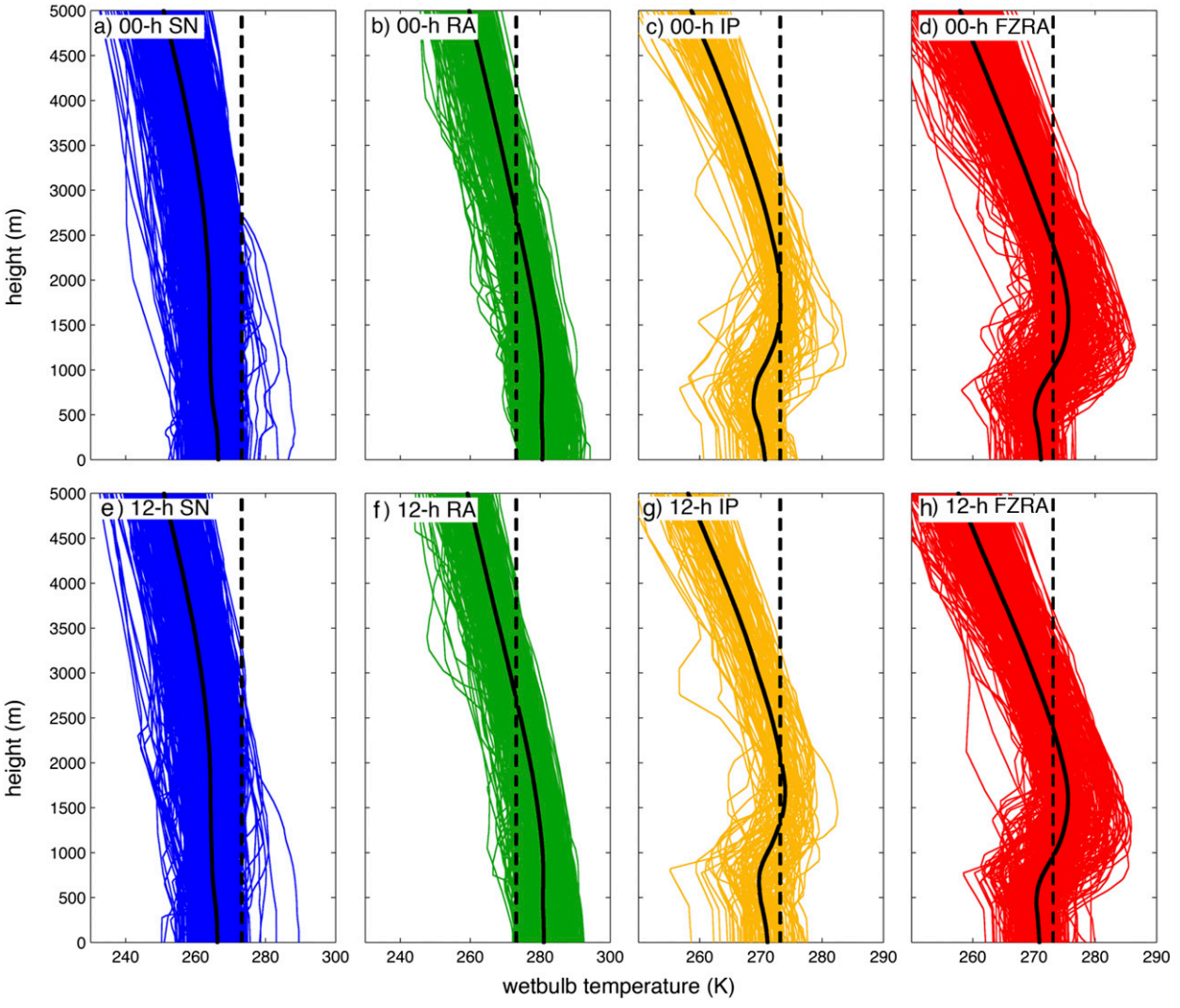

FIG. 8. The (top) analyzed and (bottom) forecast wet-bulb temperature profiles (colored curves) corresponding to the (a),(e) SN; (b),(f) RA; (c),(g) IP; and (d),(h) FZRA observations shown in Fig. 2. Solid, black curves are the means for each panel and dashed, black curves indicate the 273-K contour.

3. Uncertainty in the analysis–forecast dataset

We have seen how the different algorithms are sensitive to the character in the vertical profiles of T or T_w. But, given that model uncertainty leads to temperature variations on the order of 5 K, to what extent is the classification of precipitation type affected by uncertainty in the analysis or forecast data? To address this question, soundings for the events described in section 2 are collected from the 20-km RUC model from October 2003 to December 2012 in order to first establish a representative range of uncertainty for T and RH and then to use those ranges to address how well each algorithm performs after the effects of uncertainty are accounted for.

a. Uncertainty of T and RH *as a function of lead time*

Forecasts and analyses from the 0-, 1-, 6-, and 12-h lead times are considered. Comparisons of the T_w profiles from the 0-h lead time to the observations have similar means and ranges of T_w (cf. Figs. 2 and 8). The 0-h SN profiles have 25 soundings with surface $T_w >$ 273 K and 37 soundings with an elevated warm layer (Fig. 8a). The 0-h RA profiles cross the 273-K isotherm between 25 and 4400 m above ground level and have surface T_w ranging from 273 to 294 K (Figs. 8b). The 0-h IP and FZRA profile means are primarily distinguished by $T_{w\text{—max}}$ and the depth of the elevated warm layer (Figs. 8c,d). The 12-h profiles share many of these similarities. There are 30 SN profiles with an elevated warm

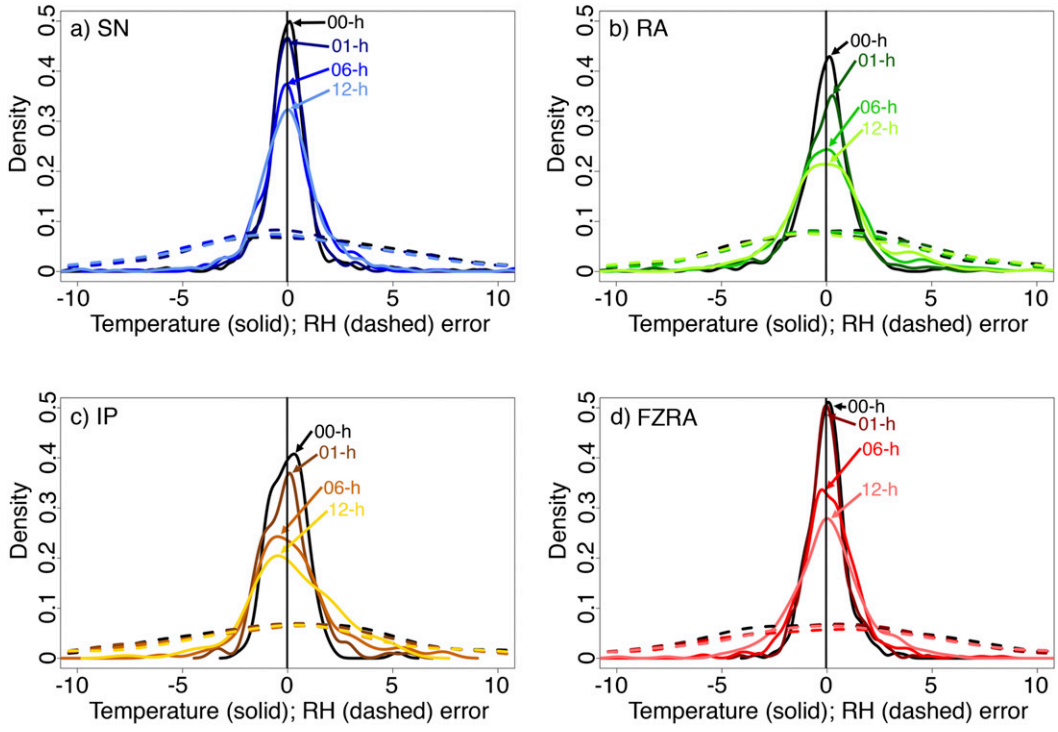

FIG. 9. The kernel density distributions of the difference between the observed and forecast temperature (K; solid) and relative humidity (%; dashed) for the different precipitation types and lead times.

layer and 29 with surface $T_w > 273$ K (Fig. 8e). The 12-h RA profiles cross the 273-K isotherm between 30 and 4300 m above ground and have surface T_w ranging from 273 to 292 K (Fig. 8f). The 12-h mean IP and FZRA profiles are also primarily distinguished by the depth and $T_{w\text{—max}}$ of the elevated warm layer (Figs. 8g,h). Though not shown, the 1- and 6-h profiles are similar to the above.

There are, of course, differences between individual observed profiles and their corresponding forecasts. The character of these differences is shown using unbiased kernel density distributions of the difference between the observed and model-forecast T and RH at 0 m above ground level (Fig. 9). The maximum density and spread are different for each category, with IP and RA having the lowest maximum densities and greatest spread (Figs. 9b,c). This suggests there is a greater degree of uncertainty in the temperature forecasts and analyses for these precipitation types. For IP, this may be due to model biases in frontal location and how well the slope of the frontal interface is captured. Many of the RA profiles come from deep within the warm sector of midlatitude cyclones. Consideration of the observed radar reflectivity for all RA events shows that many of these soundings are from a convective environment, where large T and RH errors can occur (Cintineo and Stensrud 2013). Analyses from higher altitudes leads to the same conclusions, but the maximum densities increase with height (not shown), suggesting a greater degree of certainty with increasing altitude.

b. *Uncertainty effects on algorithm performance*

To gauge whether or how the above uncertainty affects algorithm performance, each observed sounding is perturbed 1000 times. In this exercise, we roughly follow the methodology of Cintineo and Stensrud (2013). Namely, perturbations are applied to the observed soundings in accordance with the depth over which the vertical correlations in the unbiased errors approach zero (the vertical distance over which the autocorrelation becomes insignificant) and a polynomial spline is used to interpolate between those levels. These depths vary by category and lead time. The perturbation distributions match the distributions in the unbiased kernel density distributions (e.g., Fig. 9), having a higher concentration of errors near 0 K (or 0% in the case of RH). The resulting sets of perturbed soundings represent the range of possible soundings that could have been produced by the RUC model, excepting any mean biases. The mean biases are removed in order to have a perturbed sounding collection that may be more representative of the uncertainty that exists for any mesoscale forecast system.

The five algorithms are applied to the perturbed soundings and their precipitation types compared to the ASOS observations. The resulting PODs are provided graphically as a function of lead time in Fig. 10. All of the SN profiles have some decrease in the POD between the observed and 0-h perturbed soundings (Fig. 10a). The decrease ranges from 3.1% to 8.0%, with the Ramer scheme having the greatest drop. However, there is very little difference in POD between the 0-, 1-, 6-, and 12-h perturbations. A similar pattern exists for the FZRA profiles (Fig. 10d); there is a drop in POD between the observed and the 0-h perturbed soundings ranging from 4.5% to 13.1% and the PODs show little variation between the 0- and 12-h perturbations. The RA and IP results are a little different. In the RA category, PODs decrease as lead time increases throughout the lead times considered (Fig. 10b). However, the decrease over the entire set of lead times is quite modest, leaving 12-h PODs that range from 91.7% to 95.4%. In the IP category (Fig. 10c), most of the algorithms have a slight decrease in POD over the range of lead times considered, except for the Ramer algorithm, which has a slight increase between the observed and 0-h soundings. It is not obvious why this happens, but since the increase is statistically insignificant, we investigate it no further. Notice the performance of the algorithms relative to each other changes only a little as one switches from observed to perturbed soundings. For example, B2 has the highest PODs for SN for both the observed and perturbed soundings. But, no one algorithm has higher PODs than the others for all categories, again confirming previous results—that no one algorithm is universally superior to the others. One can repeat the above exercise using any one of a number of skill scores. But, the results remain the same: that no one algorithm is superior to the others; that as one switches from observed to perturbed soundings, the SN, RA, and FZRA scores decrease; and that the IP and FZRA scores are much lower than the SN and RA scores (not shown).

To better isolate whether some algorithms are more vulnerable to uncertainty effects, the above exercise is repeated using a perfect algorithm assumption. To do this, only those sets of perturbed soundings whose observed counterparts were correctly classified by the algorithm in question are considered. The PODs as a function of lead time for these soundings show that Ramer is the most vulnerable to uncertainty for SN, with a 12-h POD of 90.4% (Fig. 10e). The B1 and B2 algorithms are most vulnerable for RA, with a 12-h POD of 93.6% (Fig. 10f). Both of these PODs are quite high in comparison those from IP and FZRA. In these classes, the 12-h PODs range from 38.8% to 78.3% for IP and 44.6% to 63.9% for FZRA (Figs. 10g,h). In either, NSSL shows the greatest POD decreases. Therefore, the uncertainty is greater for IP and FZRA than for SN or RA. The causes for this are investigated in the following section.

c. *Causes for the low PODs for the perturbed IP and FZRA profiles*

The lower PODs for the perturbed IP and FZRA profiles lie in their proximity to the 273-K isotherm and to each other. Consider the observed and perturbed soundings in Fig. 11. Each observed sounding was correctly classified by all algorithms. While the SN profile has some perturbations that are above 273 K in the lower troposphere, because the observed sounding is sufficiently cold, most of the perturbations are completely subfreezing (Fig. 11a). Likewise, the RA example has several perturbations with surface-based cold layers, but most are above freezing in the lower troposphere (Fig. 11b). Because they are sufficiently close to 273 K, both the IP and FZRA examples have several perturbations with no elevated warm layers (Figs. 11c,d). Both also have several perturbations with elevated warm layers that are much deeper and warmer than observed and surface-based cold layers that are shallower and warmer than observed. Hence, one can easily imagine that many of these profiles are misclassified by one or more algorithms while most of the SN and RA profiles are correctly classified.

A more universal perspective is provided in Fig. 12, which shows the mean, 2.5th, and 97.5th percentile profiles of T_w for *all* of the 0-h perturbed soundings (not just those shown in Fig. 11). Most SN soundings are below 273 K and most RA profiles are above 273 K in the lower troposphere (Figs. 12a,b). This explains the comparatively high PODs for these classes. However, the 2.5th (97.5th) percentile for IP and FZRA is completely below (above) 273 K in the lower troposphere (Figs. 12c,d). The only notable difference between the 97.5th percentiles is that in FZRA, the elevated warm layer is slightly deeper and warmer. Such strong overlap is quite significant as it indicates that once the model uncertainty is accounted for, IP and FZRA T_w profiles are not easily distinguished. Indeed, the kernel density distributions of the areas of the elevated warm and surface-based cold layers (as in Fig. 3) show that the perturbed IP and FZRA profiles cannot be distinguished (not shown).

Let us now consider the SFZR category from section 2e. The mean and percentiles from the perturbed SFZR T_w profiles show that almost all have an elevated warm layer, as opposed to the perturbed FZRA profiles in which many have no elevated warm layer (cf. Figs. 13a and 12d). The elevated warm layers are,

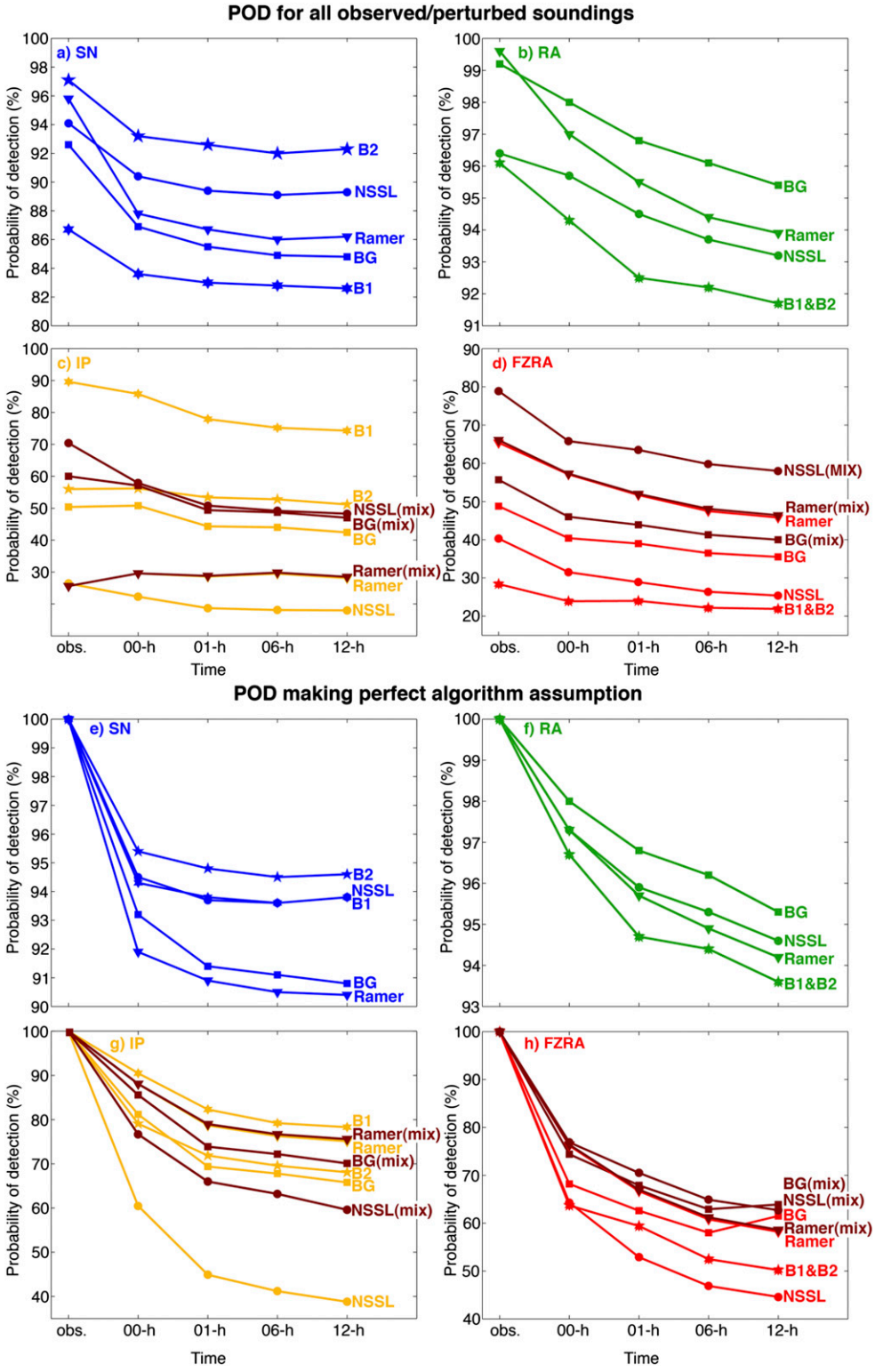

FIG. 10. The POD for (a)–(d) all sets of observed and perturbed soundings and (e)–(h) the observed and perturbed soundings making a perfect algorithm assumption. Note that the *y* axis is different in each panel.

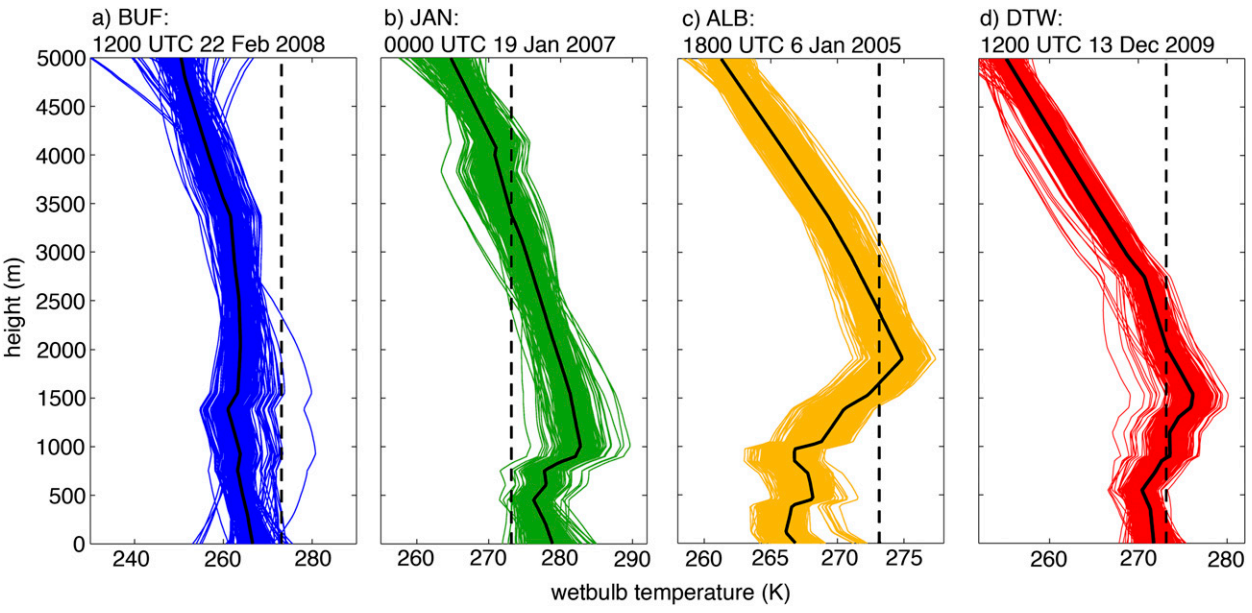

FIG. 11. The observed wet-bulb temperature profile (black curves) for select events and their corresponding 0-h perturbed profiles (colored curves). The black, dashed line is the 273-K contour. Locations of the profiles are indicated in Fig. 1.

on average, much deeper and warmer than in FZRA. The effect of this is to improve the PODs for these soundings. While all algorithms show some drop in POD between the observed and 0-h perturbed soundings, the 12-h PODs range from 34.2% to 68.9% (Fig. 13b). These PODs are between 10.9% and 17.1% higher than in FZRA (Fig. 10d). The PODs making a perfect algorithm assumption are between 4.6% and 13.4% higher than in FZRA (not shown). These differences in PODs are statistically significant at the 95% confidence interval (not shown), which suggests that taking an approach that isolates sustained, horizontally uniform regions of FZRA from more mixed environments may allow for higher PODs in model forecasts and, consequently, higher confidence on the part of forecasters.

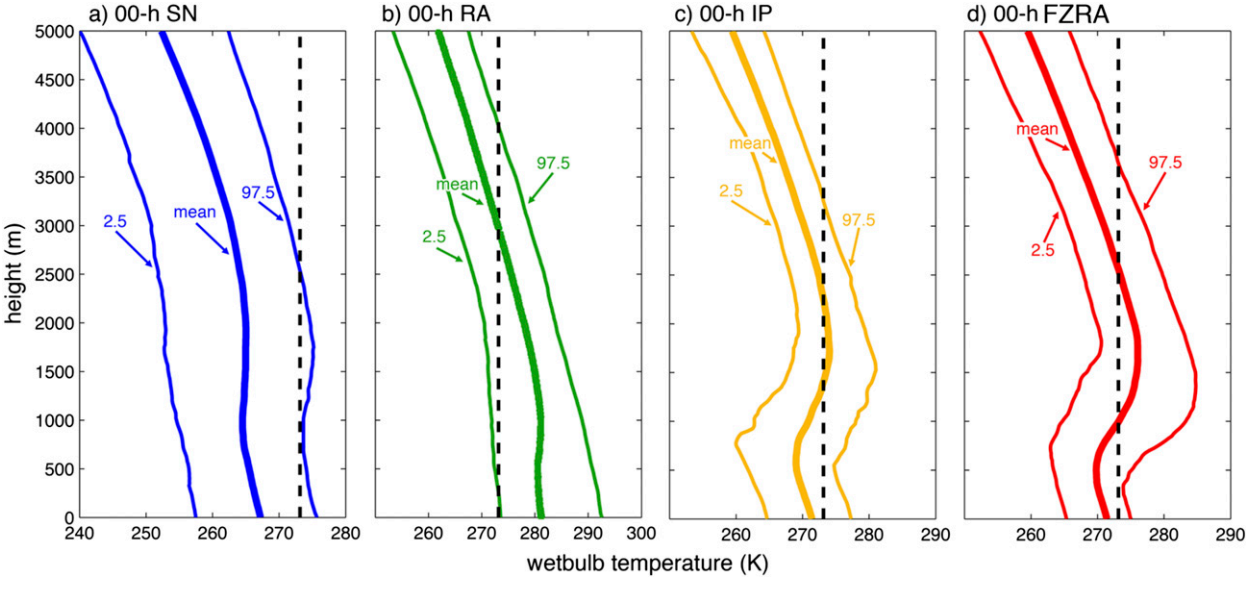

FIG. 12. The mean (a) SN, (b) RA, (c) IP, and (d) FZRA T_w profiles along with the 2.5th and 97.5th percentiles.

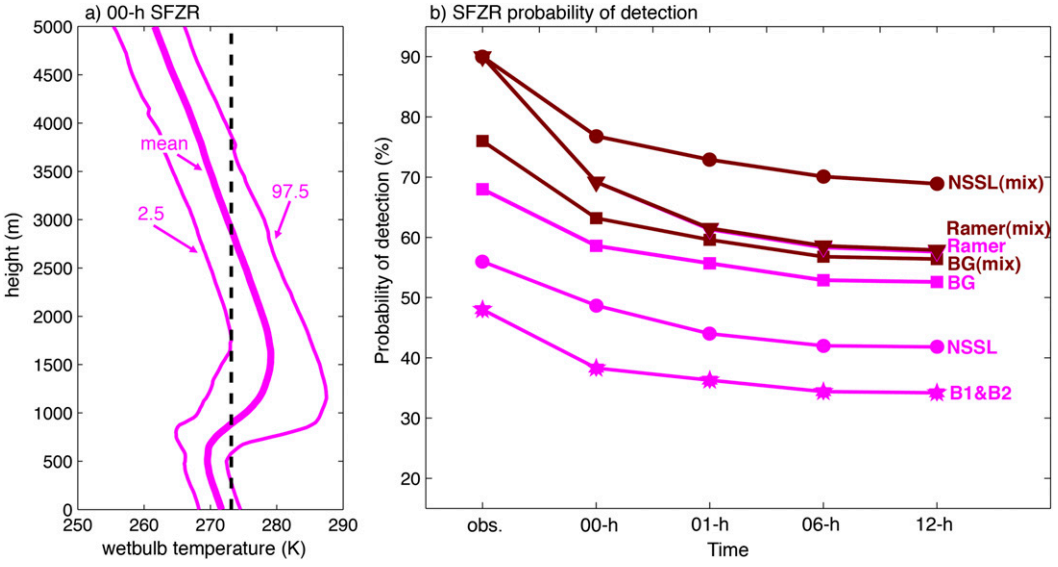

FIG. 13. The (a) mean, 2.5th, and 97.5th percentile profiles of T_w for the 0-h perturbed SFZR soundings and (b) the POD for all sets of perturbed SFZR soundings.

4. Conclusions

The effects of two sources of uncertainty on the correct determination of precipitation type were considered. The first is the choice of algorithm. Five different implicit precipitation-type algorithms were applied to a collection of observed soundings that correspond to surface observations of SN, RA, IP, and FZRA. All algorithms do quite well at detecting SN and RA, which is expected given that in the lower troposphere most of these profiles are well below or well above 273 K. The algorithms are much less successful with IP and FZRA. All algorithms except one favor one form over the other. The various methods used by each algorithm were investigated to understand these biases. No one threshold in any of the algorithms considered could be used to differentiate between IP and FZRA without introducing substantial errors for one or both forms.

The second source of uncertainty is the model error. In this part of the paper, the Rapid Update Cycle (RUC) model was used to first estimate the range of errors present in the vertical profiles of T and RH and then gauge how these errors affect the performance of the various algorithms. The mean model errors were removed from all analyses in order to better represent the range of errors one might expect from any mesoscale modeling system. The unbiased model errors show that the uncertainty is greatest for IP and RA, but that all forms have increased uncertainty as lead time is increased. The spread in the unbiased errors was used to generate a set of 1000 perturbed soundings for each observed sounding that represents the range of soundings that could have been diagnosed/forecast by the model. These soundings were applied to the five algorithms tested herein. In almost all cases, the PODs for the perturbed soundings are less than for the observed. The degree of degradation, however, is strongly linked to the type of precipitation. The SN and RA classes have quite high PODs even when model uncertainty is accounted for. This, again, is due to the fact that most of these profiles are either well above or well below freezing in the lower troposphere. The IP and FZRA classes, however, are much more sensitive to model uncertainty. Further investigation shows that the primary reason IP and FZRA suffer more from the effects of uncertainty is because their ranges of T_w profiles are so similar. Once the uncertainty is accounted for, these profiles are nearly indistinguishable.

The above findings may present a good argument for using an explicit approach. However, we reference the work of Ikeda et al. (2013), who assess the performance of the explicit algorithm used by the High-Resolution Rapid Refresh (HRRR) model. Their PODs for three classes of events (binned by the aerial extent of precipitation) range from 39% to 61%. Combining our IP and FZRA classes into a single class, as was done in Ikeda et al. (2013), yields a very similar range (from 35% to 77% for the observed and 31.5% to 63.8% for the perturbed soundings). But, there are a sufficient enough number of differences between their approach and ours to merit further study.

The above findings are rather sobering, as they suggest a clear distinction between IP and FZRA cannot be achieved using algorithms that primarily rely on vertical

profiles of T_w. Yet, such a distinction is highly desirable, as some FZRA events can result in significant ice accumulations that can cause loss of power and increase travel hazards. High-resolution observations collected as a part of the mPING project suggest that many events may be characterized by rapidly changing precipitation type and strong horizontal variability. Using a five-category algorithm that includes an IP–FZRA mix may be more appropriate in such situations. Promise was found when soundings that are associated with long-duration, horizontally homogeneous regions of FZRA, according to the National Climatic Data Center's Storm Events Database, were considered separately. These soundings were better diagnosed by all algorithms both for the observed and perturbed soundings. Future work that isolates these types of events and their corresponding thermal characteristics is highly recommended.

Acknowledgments. This study was made possible in part due to the data made available by the governmental agencies, commercial firms, and educational institutions participating in MesoWest. Special thanks to H. Brooks, H. Burcham, R. Cintineo, S. Koch, D. Stensrud, and the anonymous reviewers. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce.

REFERENCES

Baldwin, M., R. Treadon, and S. Contorno, 1994: Precipitation type prediction using a decision tree approach with NMCs mesoscale eta model. *10th Conf. on Numerical Weather Prediction,* Portland, OR, Amer. Meteor. Soc., 30–31.

Benjamin, S. G., 1989: An isentropic mesoα-scale analysis system and its sensitivity to aircraft and surface observations. *Mon. Wea. Rev.,* **117,** 1586–1603, doi:10.1175/1520-0493(1989)117<1586:AIMSAS>2.0.CO;2.

Bernstein, B. C., 2000: Regional and local influences on freezing drizzle, freezing rain, and ice pellet events. *Wea. Forecasting,* **15,** 485–508, doi:10.1175/1520-0434(2000)015<0485:RALIOF>2.0.CO;2.

Bourgouin, P., 2000: A method to determine precipitation type. *Wea. Forecasting,* **15,** 583–592, doi:10.1175/1520-0434(2000)015<0583:AMTDPT>2.0.CO;2.

Brooks, C. F., 1920: The nature of sleet and how it is formed. *Mon. Wea. Rev.,* **48,** 69–72, doi:10.1175/1520-0493(1920)48<69b:TNOSAH>2.0.CO;2.

Brown, J. M., and Coauthors, 2011: Improvement and testing of WRF physics options for application to Rapid Refresh and High Resolution Rapid Refresh. *14th Conf. on Mesoscale Processes/15th Conf. on Aviation, Range, and Aerospace Meteorology,* Los Angeles, CA, Amer. Meteor. Soc., 5.5. [Available online at https://ams.confex.com/ams/14Meso15ARAM/webprogram/Paper191234.html.]

Changnon, S. A., 2003: Urban modification of freezing-rain events. *J. Appl. Meteor.,* **42,** 863–870, doi:10.1175/1520-0450(2003)042<0863:UMOFE>2.0.CO;2.

Cintineo, R. M., and D. J. Stensrud, 2013: On the predictability of supercell thunderstorm evolution. *J. Atmos. Sci.,* **70,** 1993–2011, doi:10.1175/JAS-D-12-0166.1.

Colle, B. A., J. B. Olson, and J. S. Tongue, 2003: Multiseason verification of the MM5. Part I: Comparison with the Eta Model over the central and eastern United States and impact of MM5 resolution. *Wea. Forecasting,* **18,** 431–457, doi:10.1175/1520-0434(2003)18<431:MVOTMP>2.0.CO;2.

Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF model output for severe weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting,* **25,** 408–427, doi:10.1175/2009WAF2222258.1.

Cortinas, J. V., Jr., 2000: A climatology of freezing rain in the Great Lakes region of North America. *Mon. Wea. Rev.,* **128,** 3574–3588, doi:10.1175/1520-0493(2001)129<3574:ACOFRI>2.0.CO;2.

——, B. C. Bernstein, C. C. Robbins, and J. W. Strapp, 2004: An analysis of freezing rain, freezing drizzle, and ice pellets across the United States and Canada: 1976–1990. *Wea. Forecasting,* **19,** 377–390, doi:10.1175/1520-0434(2004)019<0377:AAOFRF>2.0.CO;2.

Crawford, R. W., and R. E. Stewart, 1995: Precipitation type characteristics at the surface in winter storms. *Cold Reg. Sci. Technol.,* **23,** 215–229, doi:10.1016/0165-232X(94)00014-O.

DeMott, P. J., and Coauthors, 2011: Resurgence in ice nuclei measurement research. *Bull. Amer. Meteor. Soc.,* **92,** 1623–1635, doi:10.1175/2011BAMS3119.1.

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

Fassnacht, S. R., Z.-L. Yang, K. R. Snelgrove, E. D. Soulis, and N. Kouwen, 2006: Effects of averaging and separating soil moisture and temperature in the presence of snow cover in SVAT and hydrological model for a southern Ontario, Canada, watershed. *J. Hydrometeor.,* **7,** 298–304, doi:10.1175/JHM489.1.

Hanesiak, J. M., and R. E. Stewart, 1995: The mesoscale and microscale structure of a severe ice pellet storm. *Mon. Wea. Rev.,* **123,** 3144–3162, doi:10.1175/1520-0493(1995)123<3144:TMAMSO>2.0.CO;2.

Heidke, P., 1926: Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungdienst. *Geogr. Ann.,* **8,** 301–349.

Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution Rapid Refresh Model. *Wea. Forecasting,* **28,** 921–939, doi:10.1175/WAF-D-12-00085.1.

Janjić, Z. I., T. L. Black, M. E. Pyle, H.-Y. Chuang, E. Rogers, and G. I. DeMego, 2005: The NCEP WRF-NMM core. Preprints, *Joint WRF/MM5 User's Workshop,* Boulder, CO, NCAR, 2.9. [Available online at http://www2.mmm.ucar.edu/wrf/users/workshops/WS2005/abstracts/Session2/9-Janjic.pdf.]

Jones, M. S., B. A. Colle, and J. S. Tongue, 2007: Evaluation of a mesoscale short-range ensemble forecast system over the northeast United States. *Wea. Forecasting,* **22,** 36–55, doi:10.1175/WAF973.1.

Kain, J. S., S. M. Goss, and M. E. Baldwin, 2000: The melting effect as a factor in precipitation-type forecasting. *Wea. Forecasting,* **15,** 700–714, doi:10.1175/1520-0434(2000)015<0700:TMEAAF>2.0.CO;2.

Lackmann, G. M., K. Keeter, L. G. Lee, and M. B. Ek, 2002: Model representation of freezing and melting precipitation: Implications for winter weather forecasting. *Wea. Forecasting,*

17, 1016–1033, doi:10.1175/1520-0434(2003)017<1016:MROFAM>2.0.CO;2.

Manikin, G. S., 2005: An overview of precipitation type forecasting using NAM and SREF data. *24th Conf. on Broadcast Meteorology/21st Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction,* Washington, DC, Amer. Meteor. Soc., 8A.6. [Available online at https://ams.confex.com/ams/pdfpapers/94838.pdf.]

——, K. F. Brill, and B. Ferrier, 2004: An Eta Model precipitation type mini-ensemble for winter weather forecasting. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction,* Seattle, WA, Amer. Meteor. Soc., 23.1. [Available online at https://ams.confex.com/ams/pdfpapers/73517.pdf.]

Moorthi, S., H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. NWS Tech. Procedures Bull. 484, 14 pp. [Available online at http://www.nws.noaa.gov/om/tpb/484.htm.]

NOAA, 1998: Automated Surface Observing System user's guide. NOAA/NWS, 58 pp. [Available from ASOS Program Office, NWS, 1325 East–West Hwy., Silver Spring, MD 20910.]

Petch, J. C., and J. Dudhia, 1998: The importance of horizontal advection of hydrometeors in a single-column model. *J. Climate,* **11,** 2437–2452, doi:10.1175/1520-0442(1998)011<2437:TIOTHA>2.0.CO;2.

Ralph, F. M., and Coauthors, 2005: Improving short-term (0–48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP workshop. *Bull. Amer. Meteor. Soc.,* **86,** 1619–1632, doi:10.1175/BAMS-86-11-1619.

Ramer, J., 1993: An empirical technique for diagnosing precipitation type from model output. *Fifth Int. Conf. on Aviation Weather Systems,* Vienna, VA, Amer. Meteor.Soc., 227–230.

Rauber, R. M., L. S. Olthoff, and M. K. Ramamurthy, 2000: The relative importance of warm rain and melting processes in freezing precipitation events. *J. Appl. Meteor.,* **39,** 1185–1195, doi:10.1175/1520-0450(2000)039<1185:TRIOWR>2.0.CO;2.

——, ——, ——, D. Miller, and K. E. Kunkel, 2001: A synoptic weather pattern and sounding-based climatology of freezing precipitation in the United States east of the Rocky Mountains. *J. Appl. Meteor.,* **40,** 1724–1747, doi:10.1175/1520-0450(2001)040<1724:ASWPAS>2.0.CO;2.

Robbins, C. C., and J. V. Cortinas Jr., 2002: Local and synoptic environments associated with freezing rain in the contiguous United States. *Wea. Forecasting,* **17,** 47–65, doi:10.1175/1520-0434(2002)017<0047:LASEAW>2.0.CO;2.

Schuur, T. J., H.-S. Park, A. V. Ryzhkov, and H. D. Reeves, 2012: Classification of precipitation types during transitional winter weather using the RUC model and polarimetric radar retrievals. *J. Appl. Meteor. Climatol.,* **51,** 763–779, doi:10.1175/JAMC-D-11-091.1.

Seidel, D. J., B. Sun, M. Pettey, and A. Reale, 2011: Global radiosonde balloon drift statistics. *J. Geophys. Res.,* **116,** D08102, doi:10.1029/2010JD014891.

Silverman, B. W., 1992: *Density Estimation for Statistics and Data Analysis.* Chapman and Happ, 175 pp.

Stensrud, D. J., and N. Yussouf, 2003: Short-range ensemble predictions of 2-m temperature and dewpoint temperature over New England. *Mon. Wea. Rev.,* **131,** 2510–2524, doi:10.1175/1520-0493(2003)131<2510:SEPOMT>2.0.CO;2.

Stewart, R. E., and P. King, 1987: Freezing precipitation in winter storms. *Mon. Wea. Rev.,* **115,** 1270–1279, doi:10.1175/1520-0493(1987)115<1270:FPIWS>2.0.CO;2.

Thériault, J. M., R. E. Stewart, and W. Henson, 2010: On the dependence of winter precipitation types and temperature, precipitation rate, and associated features. *J. Appl. Meteor. Climatol.,* **49,** 1429–1442, doi:10.1175/2010JAMC2321.1.

——, and Coauthors, 2012: A case study of processes impacting precipitation phase and intensity during the Vancouver 2010 Winter Olympics. *Wea. Forecasting,* **27,** 1301–1325, doi:10.1175/WAF-D-11-00114.1.

Wandishin, M. S., M. E. Baldwin, S. L. Mullen, and J. V. Cortinas Jr., 2005: Short-range ensemble forecasts of precipitation type. *Wea. Forecasting,* **20,** 609–626, doi:10.1175/WAF871.1.

Zerr, R. J., 1997: Freezing rain: An observational and theoretical study. *J. Appl. Meteor.,* **36,** 1647–1661, doi:10.1175/1520-0450(1997)036<1647:FRAOAT>2.0.CO;2.