A Method for Calibrating Deterministic Forecasts of Rare Events

PATRICK T. MARSH

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

JOHN S. KAIN

National Severe Storms Laboratory, Norman, Oklahoma

VALLIAPPA LAKSHMANAN AND ADAM J. CLARK

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

NATHAN M. HITCHENS

National Severe Storms Laboratory, Norman, Oklahoma

JILL HARDY

School of Meteorology, University of Oklahoma, Norman, Oklahoma

(Manuscript received 17 June 2011, in final form 6 October 2011)

ABSTRACT

Convection-allowing models offer forecasters unique insight into convective hazards relative to numerical models using parameterized convection. However, methods to best characterize the uncertainty of guidance derived from convection-allowing models are still unrefined. This paper proposes a method of deriving calibrated probabilistic forecasts of rare events from deterministic forecasts by fitting a parametric kernel density function to the model's historical spatial error characteristics. This kernel density function is then applied to individual forecast fields to produce probabilistic forecasts.

1. Introduction

Rare meteorological events[1] that occur on small spatial and short temporal scales pose significant challenges to forecasters. This is related to the limited predictability of phenomena occurring on short time–space scales. However, these events compose a substantial portion of meteorological phenomena that negatively impact society, such as heavy rain, large hail, and tornadoes. Thus, accurate numerical guidance of these events would provide large societal benefits.

Convection-allowing models (CAMs) have shown improved skill, compared to parameterized convection models, in identifying regions where rare meteorological events associated with convection (hereafter rare convective events, RCEs) may occur (Clark et al. 2010b). Furthermore, CAMs are able to do this by explicitly representing deep-convective storms and their unique attributes—not just storm environments (Kain et al. 2010). Yet, quantifying the uncertainty associated with explicit numerical predictions of RCEs is particularly challenging (Sobash et al. 2011). Of course, ensembles are powerful tools for quantifying uncertainty, but when convection-allowing ensemble prediction systems are used to provide guidance for forecasting storm attributes, they are subject to the same fundamental limitation that

[1] Murphy (1991) defined a rare meteorological event as one that occurs on less than 5% of forecasting occasions.

Corresponding author address: Patrick T. Marsh, National Severe Storms Laboratory, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: patrick.marsh@noaa.gov

DOI: 10.1175/WAF-D-11-00074.1

handicaps single-member CAM forecast systems: too little is known about the performance characteristics of CAMs in predicting RCEs explicitly.

There are three main reasons for this deficiency. First, routine, explicit, contiguous, or near-contiguous, United States (CONUS or near-CONUS) scale forecasts of RCEs have been available for only 6–7 yr in the United States, so there is still much to learn about which phenomena can be skillfully predicted with CAMs (Kain et al. 2008, 2010). Second, most real-time forecasting efforts with CAMs have been short-term initiatives, focusing on specific tasks (e.g., Done et al. 2004; Weisman et al. 2008). Third, there is a limited database of forecasts for RCEs, making robust statistical techniques difficult (e.g., Hamill and Whitaker 2006). In short, there is a limited track record of the use of CAMs as guidance for prediction of RCEs.

This paper presents a strategy for calibrating, or quantifying the uncertainty of, forecasts of RCEs based on the idea of generating probabilistic forecasts from a single underlying deterministic model. It uses a conceptual approach similar to that described by Theis et al. (2005) and refined by Sobash et al. (2011). As in those two studies, this strategy differs from other methods for both deterministic models (e.g., Glahn and Lowry 1972) and ensemble modeling systems (e.g., Hamill and Colucci 1998; Raftery et al. 2005; Clark et al. 2009; Glahn et al. 2009) by including a neighborhood around each model grid point as a fundamental component of the calibration process.

The strategy is loosely based on kernel density estimation (KDE), which can be used to retrieve spatial probability distributions from point observations or, in this case, forecasts. In other words, if a model forecasts an event at point A, KDE can be utilized to gain insight into the probability that the event might occur at a nearby point. This is achieved by utilizing a statistical distribution to redistribute a total of 100% probability over multiple (typically nearby) grid points. The result is a probability forecast, the character of which is determined by one's choice of statistical distribution and the number of grid points over which this distribution is applied. The smoothing effect is similar to that obtained with ensemble output by Wilks (2002), but calibration efforts herein focus on output from a single deterministic model. Sobash et al. (2011) demonstrated with a two-dimensional, isotropic Gaussian function that calibration of the probability forecasts derived using this technique is most easily done by changing the number of grid points over which nonzero probabilities are distributed. In this study, however, an objective calibration method, based on past model performance, is presented.

The method is presented in following sections of this paper. Section 2 describes the datasets used to develop and test the approach. Section 3 describes how the method is applied and section 4 provides initial results. The paper concludes with a brief summary and discussion.

2. Data

Model forecasts and observations of precipitation were obtained for the 48-month time period 1 April 2007–31 March 2011 and subdivided into two classifications: training and verification. Forecasts and observations during the time period 1 April 2007–31 March 2010 (36 months) were used in the training dataset and the remaining 12 months were used to test and verify the proposed method.

Model forecasts were taken from the 4-km grid-length Weather Research and Forecasting Model (WRF) configuration (Skamarock et al. 2008) run daily at the National Oceanic and Atmospheric Administration/National Severe Storms Laboratory (NOAA/NSSL). The NSSL produces numerical weather prediction forecasts from the WRF as part of an ongoing collaborative effort with the NOAA/Storm Prediction Center (SPC). Model forecasts are produced daily out to 36 h, using 0000 UTC initial and lateral boundary conditions from the operational North American Mesoscale Model (Rogers et al. 2009), over a CONUS domain. Information on the configuration is provided in Kain et al. (2010). (Images of output from the WRF forecasts generated at the NSSL, hereafter NSSLWRF, can be found online at http://www.nssl.noaa.gov/wrf.)

Observations were taken from the NOAA/National Centers for Environmental Prediction (NCEP) stage IV national quantitative precipitation estimate analyses. The stage IV analyses are based on the multisensor hourly–6-hourly "stage III" analyses (on local 4.7-km polar-stereographic grids) produced by the 12 River Forecast Centers in the CONUS. NCEP mosaics the stage III into a national product (the stage IV analyses) available at hourly, 6-hourly, and 24-hourly (accumulated from the 6-hourly) intervals. Lin and Mitchell (2005) describe further details of these analyses. (Archives of the stage IV dataset can be found online at http://data.eol.ucar.edu/codiac/dss/id=21.093.)

Diagnostic analyses were conducted on the stage IV grid, requiring interpolation of the NSSLWRF output. The program copygb (http://www.cpc.ncep.noaa.gov/products/wesley/copygb.html) was used for the interpolation and domain-wide total liquid volume was conserved. Six-hour accumulation periods were used, taken from the 12–36-h forecasts ending at 1800, 0000, 0600, and 1200 UTC. A mask was applied to both the

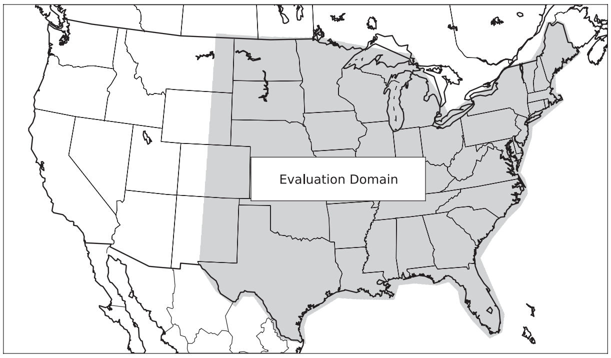

FIG. 1. The subset of the stage IV grid used in the analysis (shaded).

NSSLWRF forecasts and stage IV observations to limit the region studied to CONUS and near-CONUS areas east of the Rocky Mountains (Fig. 1).

3. Proposed method

The method proposed in this study goes beyond Theis et al. (2005) and Sobash et al. (2011) by employing a compositing technique for calibration of forecast probabilities. The technique determines the two-dimensional spatial histogram of observations of a phenomenon relative to forecasts of the same phenomenon. Once this histogram is ascertained, a two-dimensional analytic function can be fitted to it. In this approach, the fitted statistical distribution determines the character of the probability forecasts and corrects for systematic displacement errors. Several analytical distributions might be good candidates for this purpose, but a two-dimensional Gaussian function is applied here, fitted using methods similar to those in Lakshmanan and Kain (2010).

For this study, the rare convective event of choice was defined to be 6-h precipitation accumulation of greater than or equal to 25.4 mm, which occurred on less than approximately 0.5% of all stage IV and NSSLWRF grid points in the training dataset. The NSSLWRF and stage IV training datasets were converted from forecasts and observations of precipitation amounts into binary grids of ones (RCE criteria was met) and zeros (RCE criteria was not met). Next, a two-dimensional frequency distribution, representing the location of stage IV RCEs occurring within 400 km (85 grid points) relative to corresponding forecasts of RCEs by the NSSLWRF, was created using the compositing technique described by Clark et al. (2010a). The two-dimensional, anisotropic Gaussian function was then fitted to the distribution. For this function, the parameters necessary to describe the distribution are the area under the Gaussian curve, the center of the fitted distribution relative to the forecast point (h, k), the standard deviation in the x direction (σ_x), the standard deviation in the y direction (σ_y), and the rotation angle of the x axis (xrot).

4. Results

The frequency distribution of the locations of the observed events relative to the forecast events for the training period (1 April 2007–31 March 2010) is shown in Fig. 2. It is clear from Fig. 2 that the maximum observed frequency is observed to the north-northeast of the forecast location and the observed distribution has an elliptical shape. When the anisotropic Gaussian function is fitted to this distribution, the resulting parameters, determined using the methods described in Lakshmanan and Kain (2010), are $(h, k) = (4.7, 23.5)$ km, indicating that the NSSLWRF forecasts were, on average, approximately 4.7 km too far west and 23.5 km too far south, and $\sigma_x \approx 180$ km, $\sigma_y \approx 160$ km, and xrot $\approx 60°$ in the counterclockwise direction, revealing the anisotropy of the distribution. To some extent, the shape and anisotropy are closely related to the mean shape and orientation of individual precipitation objects, as revealed by comparing the average size-weighted orientation of the precipitation objects, determined by the Baldwin object identification algorithm (Baldwin et al. 2005), to the orientation angle of the fitted distribution (not shown).

Using this fitted functional distribution, probabilistic forecasts for each 6-h time period from 1 April 2010 to 31 March 2011 were generated in a manner similar to what was done by Sobash et al. (2011), except that the shifted, fitted anisotropic distribution was used instead of the simple isotropic Gaussian used by Sobash et al. (2011). In essence, the fitted distribution was applied to every grid point exceeding 25.4 mm in 6 h, and the resulting individual distributions were then linearly combined to create the forecast probability. Four sample forecasts (all of differing lead times) are shown in Figs. 3 and 4 and are now discussed. However, one must be cautious about assessing the skill of a probabilistic forecasting system on the basis of individual events.

Figures 3a, 3c, and 3e depict observations and model forecasts of precipitation for the 6 h ending at 1800 UTC 2 May 2010 (a 12–18-h forecast). During this 6-h period, heavy rain fell over an elongated area stretching from central Mississippi north-northeastward into southeastern Ohio and western West Virginia, with an area exceeding 200 mm in north-central Tennessee (Fig. 3a). South and east of this axis of heaviest rainfall, areas in eastern Mississippi had precipitation totals around the 25.4-mm threshold. The NSSLWRF forecast of this event was generally good, cluing forecasters on the general area of concern. However, the NSSLWRF forecast had three

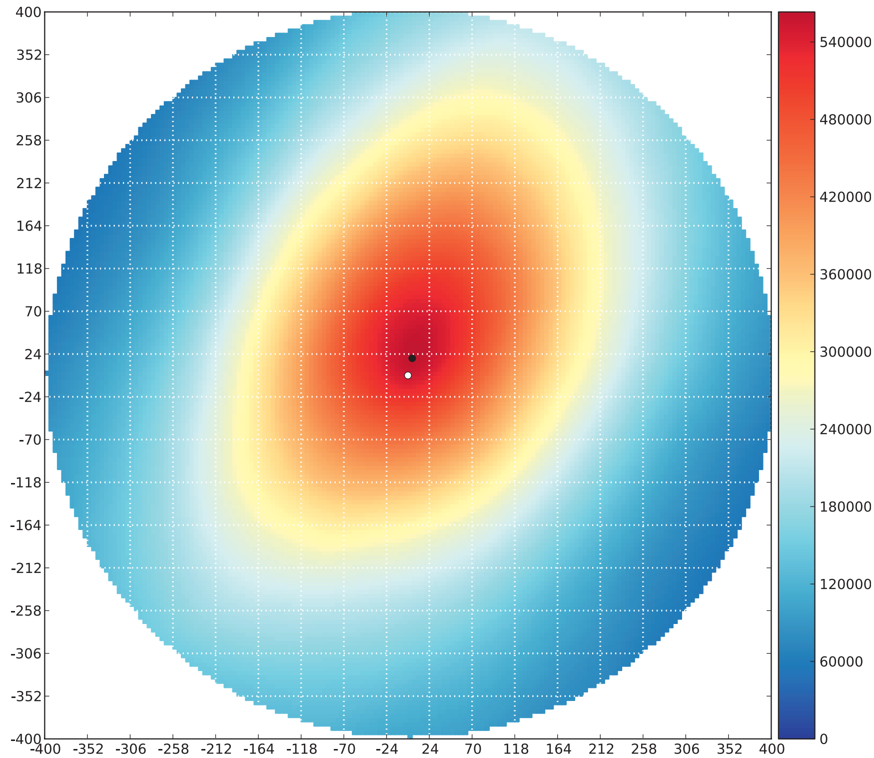

FIG. 2. The two-dimensional, frequency distribution of stage IV observations > 25.4 mm relative to NSSLWRF forecasts of the same events for the training dataset (1 Apr 2007–31 Mar 2010). The representative NSSLWRF forecast grid point is marked by a white dot in the middle of the domain and the stage IV observation frequency is color filled. To illustrate the displacement between forecasts and observations, the center of the fitted two-dimensional, anisotropic Gaussian is denoted by the black dot.

distinct areas of heavy rain compared to the single large band that was observed: one northwest of the observed axis of heavy rain, one southeast, and one along the northeastern most observed area exceeding 25.4 mm (Fig. 3c). Applying the proposed probabilistic method resulted in the area of highest probabilities of reaching or exceeding 25.4 mm (between 25% and 30%) occurring very near the area of maximum rainfall (Fig. 3e). Additionally, the axis of highest probabilities extending northeast of the maximum probabilities aligned very well with the observed area equal to or exceeding 25.4 mm. The axis of highest probabilities also extends to the south and southwest of the maximum forecast probabilities, capturing the southwestward extent of the observed heavy rain, and at the same time highlighting areas in eastern Mississippi (Fig. 3e).

Figures 3b, 3d, and 3f depict observations and model forecasts of precipitation for the 6 h ending at 0000 UTC 27 September 2010 (an 18–24-h forecast). Observations depict a large area of precipitation greater than or equal to 25.4 mm stretching from southeastern Alabama northeastward into far northwestern South Carolina with scattered areas reaching this threshold across southern Mississippi and eastern North and South Carolina (Fig. 3b). The NSSLWRF forecast of this event depicted two areas exceeding 25.4 mm of precipitation, essentially capturing both observed areas (cf. Figs. 3b and 3d). The corridor of observations greater than 25.4 mm is generally contained within 5%–10% probabilities (Fig. 3f). In this case, much of the area covered by the highest probabilities of 15%–20% did not receive heavy rainfall during this period.

A 24–30-h forecast and observations of precipitation for the 6 h ending at 0600 UTC 6 June 2010 are presented in Figs. 4a, 4c, and 4e. Observations depict two areas over Michigan that reach the 25.4-mm threshold. The first extends from the southeastern portion of Lake Michigan eastward to the western portions of Lake Erie. The second area extends from the northern portion of Lake Michigan eastward to the western portions of Lake Huron. A third area reaching the 25.4-mm threshold is found across Illinois and stretches into Indiana (Fig. 4a). Although slightly farther west, the NSSLWRF deterministic forecast does a reasonable job depicting the

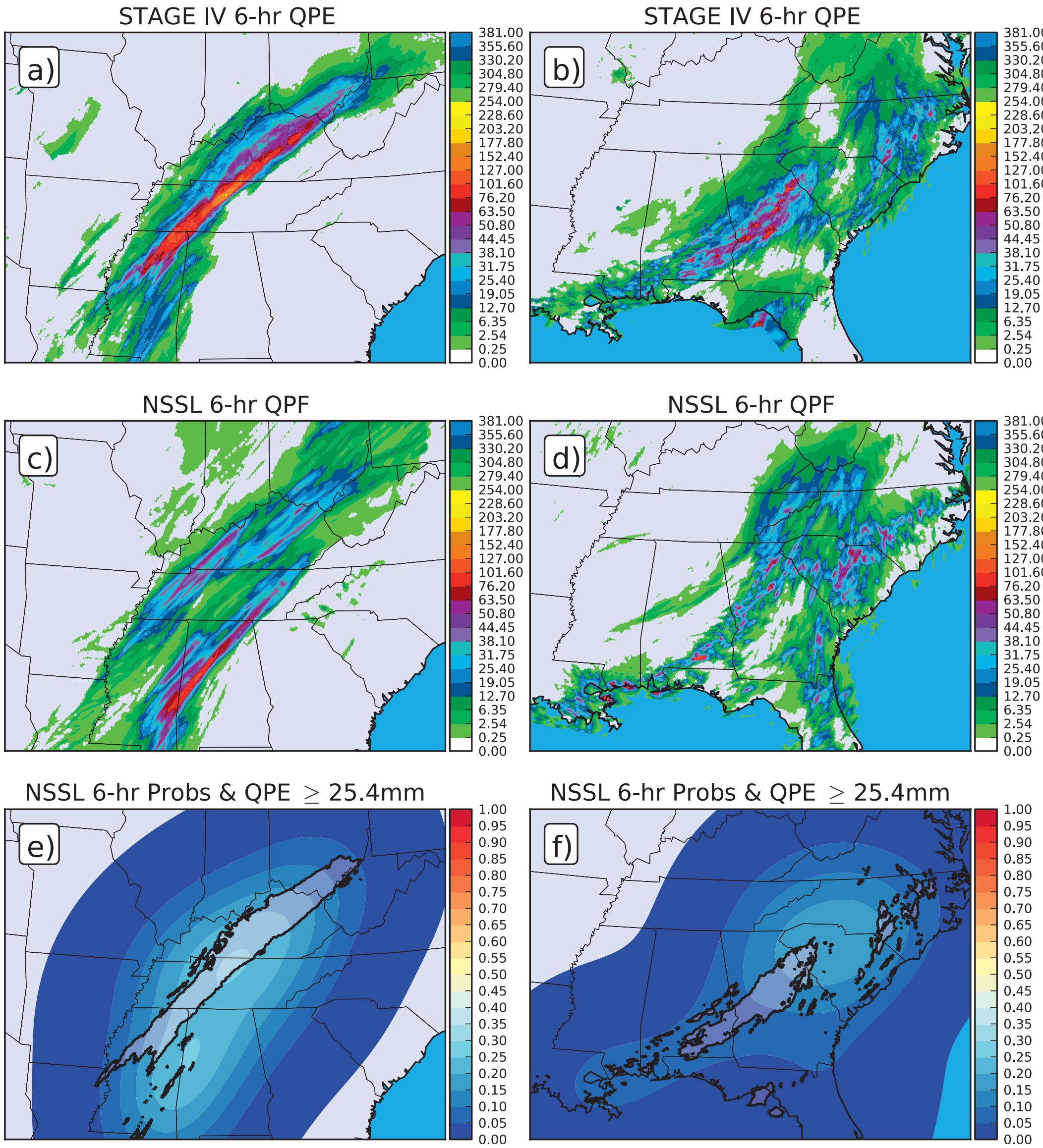

FIG. 3. Example forecasts and observations from two separate days and differing forecast lengths. The forecasts and observations for the 6 h ending at (a),(c),(e) 1800 UTC 2 May 2010 (12–18-h forecast) and (b),(d),(f) 0000 UTC 27 Sep 2010 (18–24-h forecast). (a),(b) The stage IV 6-h quantitative precipitation estimates (QPEs), (c),(d) the 6-h NSSLWRF 6-h quantitative precipitation forecasts (QPFs), and (e),(f) the stage IV QPEs $>$ 25.4 mm contoured on top of the NSSLWRF probability of exceeding 25.4 mm in 6 h. The minimum shaded probability is 0.001 (0.1%).

general location of the heaviest precipitation across Illinois and southern Michigan. However, it underpredicts the heavy precipitation across northern Michigan (Fig. 4c). The probabilistic forecast derived from the NSSLWRF captures most, if not all, observed areas that reached the 25.4-mm threshold with a probability of at least 5%—including the area across northern Michigan that was not explicitly forecast to exceed 25.4 mm by the deterministic forecast. Furthermore, the highest probabilities are located in southwestern Michigan (30%–35%), conjoined with the western portion of the southern Michigan heavy-rain axis (Fig. 4e).

A 30–36-h forecast and observations of precipitation for the 6 h ending at 1200 UTC 30 September 2010 are presented in Figs. 4b, 4d, and 4f. Observations depict a large area exceeding the 25.4-mm threshold extending from eastern South Carolina northward into far southeastern New York (Fig. 4b). Additionally, a small region of precipitation reaching the 25.4-mm threshold is found across northeastern Georgia. The NSSLWRF deterministic

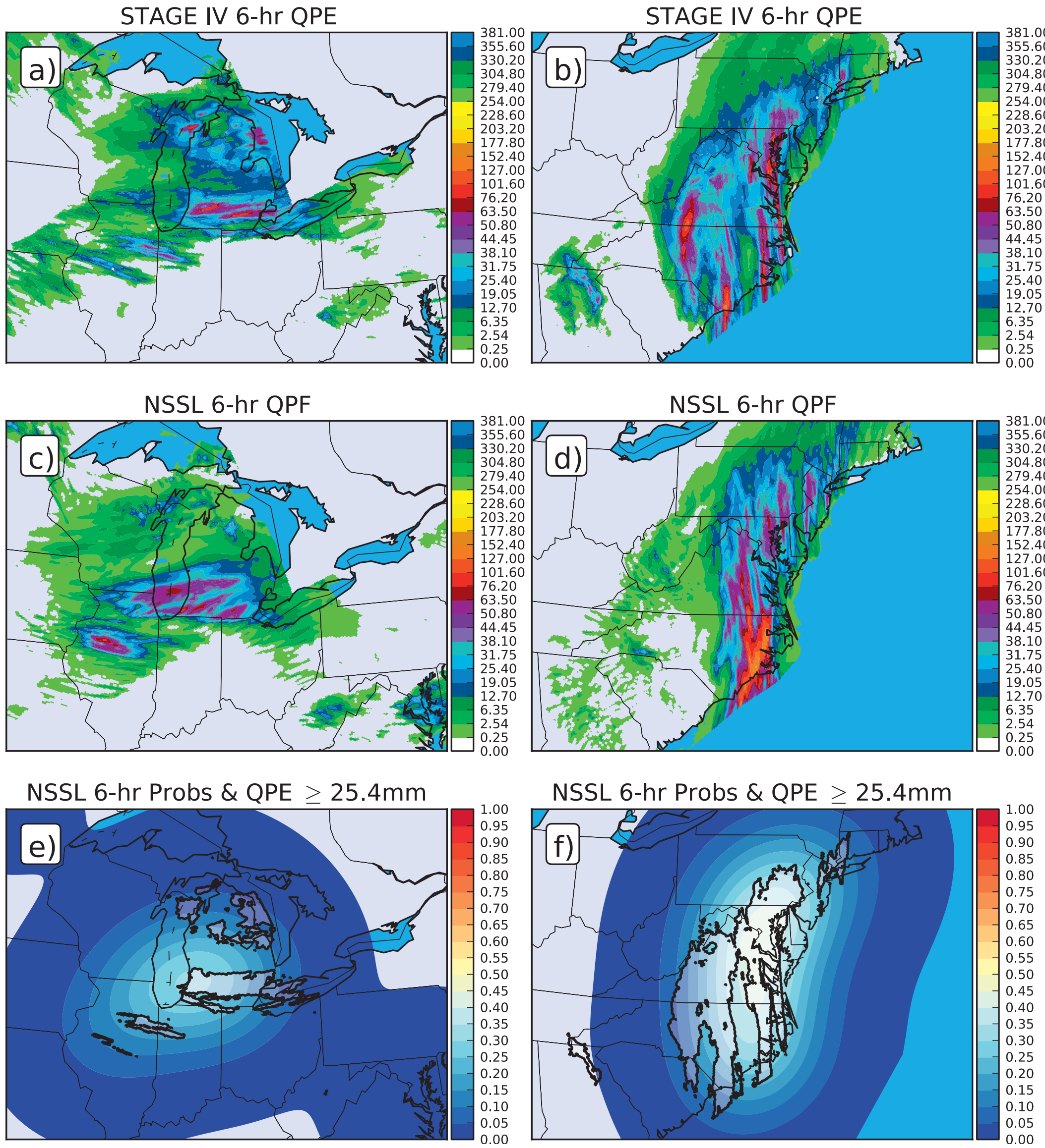

FIG. 4. As in Fig. 3, but here for the forecasts and observations for the 6 h ending at (a),(c),(e) 0600 UTC 6 Jun 2010 (24–30-h forecast) and (b),(d),(f) 1200 UTC 30 Sep 2010 (30–36-h forecast).

forecast is slightly narrower and farther east with its forecast, misplacing the axis of heaviest precipitation across North Carolina and Virginia (Fig. 4d). However, the NSSLWRF-generated probabilities encompass the area exceeding the 25.4-mm threshold, with the maximum probabilities of 40%–45% near Washington D.C. (Fig. 4f). The NSSLWRF deterministic forecast completely missed the heavy precipitation across northeastern Georgia, and this area is sufficiently far from the area to the east that it falls outside the 0.1% contour of the probabilistic forecast.

These examples are illuminating but many events are required to assess the skill of probabilistic forecast systems. A more objective verification is provided here by applying well-known verification metrics to the entire 12 months' worth of forecasts generated in this manner. First, a relative operating characteristic (ROC) curve (Mason 1982) is computed from all of the forecasts and observations (Fig. 5a). The resulting curve yields an area under the curve (AUC) of 0.94, indicating that the probabilistic forecasts have considerable skill in discriminating between events and nonevents. To visualize the reliability of the generated probabilistic forecasts, a reliability diagram was constructed (Fig. 5b). The resulting diagram indicates that the forecasts are quite reliable over a broad range of probabilities.

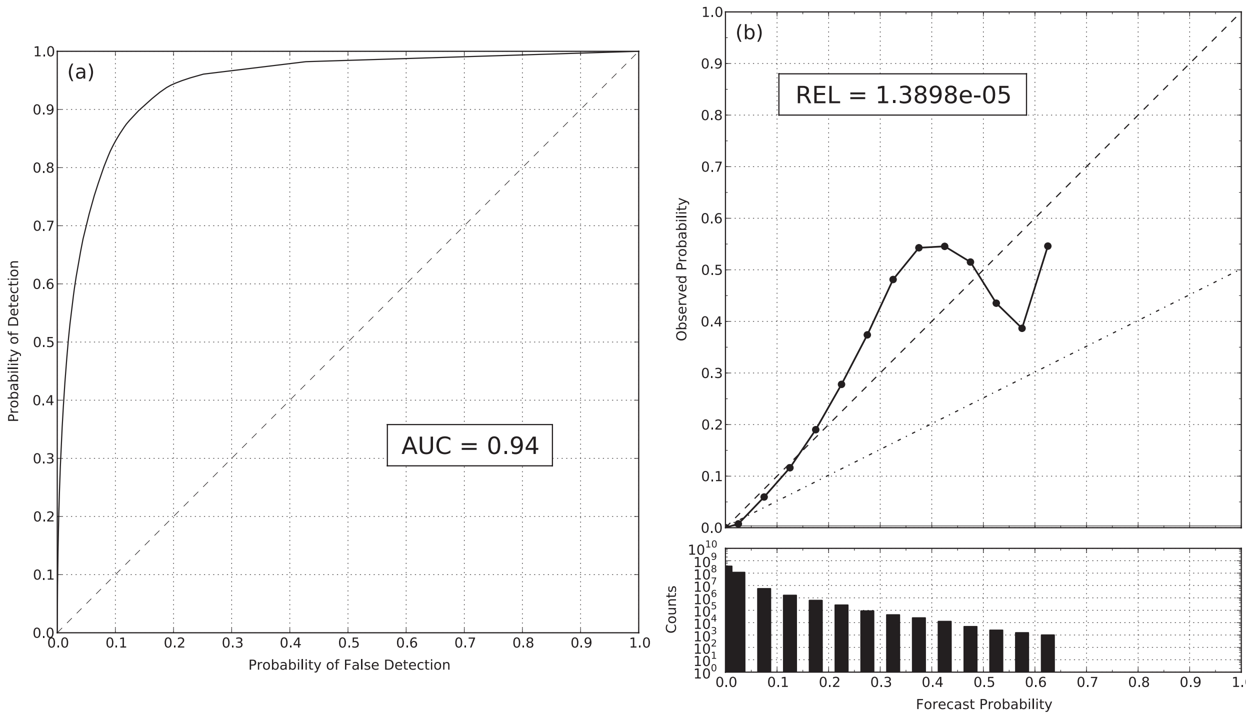

FIG. 5. (a) The ROC curve and (b) reliability diagram with corresponding forecast counts, both computed over the 1 Apr 2010–31 Mar 2011 time period. On the ROC curve, AUC = 0.94 and the line of no skill (diagonal, dashed) are also plotted. The reliability component of the Brier Score [REL = 1.3898×10^{-5}; Murphy (1973)], line of perfect reliability (diagonal, dashed), and line of no skill (dot–dash line) are also plotted in the reliability diagram. The climatology line is plotted, but because it is <0.005 it cannot be distinguished from the *x* axis. The forecast counts associated with the reliability diagram are plotted on a log scale below the reliability diagram.

5. Discussion

This paper offers a method of objectively generating calibrated probabilistic forecasts of RCEs from a deterministic model. This is achieved by computing a two-dimensional frequency distribution of observed event locations relative to the forecast event location. This frequency distribution is then used to determine the necessary parameters of an analytical function, which, in turn, can be used to convert a deterministic (1/0) forecast into a probabilistic forecast. As a proof of concept, this study uses a training dataset containing 36 months' worth of high-resolution output from the real-time NSSLWRF model and a verification dataset consisting of 12 months of forecasts from the same modeling system. Early results demonstrate this technique has the potential to produce very skillful probabilistic forecasts.

The technique is successful because it objectively represents the spatial uncertainty associated with the underlying deterministic forecast system. Preliminary assessments suggest that this uncertainty varies systematically as a function of numerous factors, such as forecast lead time, geographic location, meteorological season and regime, etc. Further refinements to the technique could include dependencies on these factors. For example, since cool-season precipitation forecasts tend to be more accurate than those for the warm season, Gaussian fits to the position-error fields could vary as a function of season, with sharper, higher-amplitude distributions in the cool season and broader, lower-amplitude distributions in the warm season.

This technique could also be used to improve probabilistic forecasts from ensemble prediction systems. Well-crafted ensemble prediction systems are likely to be more effective at sampling the range and character of possible solutions and yielding skillful probabilistic forecasts than a KDE-based approach that uses a single underlying deterministic model. But the two approaches are complementary. For example, consider that Wilks (2002) showed that smoothing of ensemble-generated probabilities tends to improve the forecast skill of the ensemble, much like adding additional members. We propose that this impact could be enhanced and better targeted if smoothing parameters were based on the historical performance characteristics of the individual members of the ensemble. This strategy is currently being investigated by the authors. This method relies heavily on accurate observations of the phenomenon being predicted. This poses significant limitations when attempting to apply this technique to other RCEs, including, but not limited to, damaging localized wind gusts, large hail, and tornadoes. This is due to the lack of quality observations of these phenomena, not to mention the inability of operational numerical models to predict these phenomena explicitly. Numerical guidance of severe thunderstorms has improved in recent years with the advent of convection-allowing models and high temporal resolution storm-attribute parameters [e.g., updraft helicity, downdraft intensity, graupel loading, etc; Kain et al. (2010)]; however, corresponding observational datasets with spatial and temporal coherences comparable to the model data are not available. It is our hope that as robust observational datasets of radar-derived convective fields become readily available, calibration of KDE-based approaches utilizing historical model performance will become more viable. One such application is the generation of probabilistic hazard information of rare convective events in a warn-on-forecast (Stensrud et al. 2009) type environment.

Acknowledgments. This work is a portion of the first author's (PTM) Ph.D. dissertation research under the supervision of the second author (JSK) and was funded by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA17RJ1227, U.S. Department of Commerce. The authors would like to acknowledge two anonymous reviewers for their insightful comments that strengthened the paper.

REFERENCES

Baldwin, M. E., J. S. Kain, and S. Lakshmivarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.,* **133,** 844–862.

Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting,* **24,** 1121–1140.

——, ——, and M. L. Weisman, 2010a: Neighborhood-based verification of precipitation forecasts from convection-allowing NCAR WRF model simulations and the operational NAM. *Wea. Forecasting,* **25,** 1495–1509.

——, ——, M. Xue, and F. Kong, 2010b: Convection-allowing and convection-parameterizing ensemble forecasts of a mesoscale convective vortex and associated severe weather environment. *Wea. Forecasting,* **25,** 1052–1081.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.,* **5,** 110–117, doi:10.1002/asl.72.

Glahn, B., M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev.,* **137,** 246–268.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Hamill, T. M., and S. J. Colucci, 1998: Evaluation of Eta–RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.,* **126,** 711–724.

——, and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.,* **134,** 3209–3229.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting,* **23,** 931–952.

——, S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting,* **25,** 1536–1542.

Lakshmanan, V., and J. S. Kain, 2010: A Gaussian mixture model approach to forecast verification. *Wea. Forecasting,* **25,** 908–920.

Lin, Y., and K. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology,* San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at http://ams.confex.com/ams/pdfpapers/83847.pdf.]

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.,* **30,** 291–303.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.,* **12,** 595–600.

——, 1991: Probabilities, odds, and forecasts of rare events. *Wea. Forecasting,* **6,** 302–307.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174.

Rogers, E., and Coauthors, 2009: The NCEP North American Mesoscale modeling system: Recent changes and future plans. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction,* Omaha, NE, Amer. Meteor. Soc., 2A.4. [Available online at http://ams.confex.com/ams/pdfpapers/154114.pdf.]

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]

Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting,* **26,** 714–728.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system. *Bull. Amer. Meteor. Soc.,* **90,** 1487–1499.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.,* **12,** 257–268, doi:10.1017/S1350482705001763.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting,* **23,** 407–437.

Wilks, D. S., 2002: Smoothing forecast ensembles with fitted probability distributions. *Quart. J. Roy. Meteor. Soc.,* **128,** 2821–2836, doi:10.1256/qj.01.215.