

# Fog Prediction from a Multimodel Mesoscale Ensemble Prediction System

BINBIN ZHOU

*NOAA/NCEP/Environmental Modeling Center, and Science Applications International Corporation, Camp Springs, Maryland*

JUN DU

*NOAA/NCEP/Environmental Modeling Center, Camp Springs, Maryland*

(Manuscript received 23 March 2009, in final form 15 September 2009)

## ABSTRACT

A new multivariable-based diagnostic fog-forecasting method has been developed at NCEP. The selection of these variables, their thresholds, and the influences on fog forecasting are discussed. With the inclusion of the algorithm in the model postprocessor, the fog forecast can now be provided centrally as direct NWP model guidance. The method can be easily adapted to other NWP models. Currently, knowledge of how well fog forecasts based on operational NWP models perform is lacking. To verify the new method and assess fog forecast skill, as well as to account for forecast uncertainty, this fog-forecasting algorithm is applied to a multimodel-based Mesoscale Ensemble Prediction System (MEPS). MEPS consists of 10 members using two regional models [the NCEP Nonhydrostatic Mesoscale Model (NMM) version of the Weather Research and Forecasting (WRF) model and the NCAR Advanced Research version of WRF (ARW)] with 15-km horizontal resolution. Each model has five members (one control and four perturbed members) using the breeding technique to perturb the initial conditions and was run once per day out to 36 h over eastern China for seven months (February–September 2008). Both deterministic and probabilistic forecasts were produced based on individual members, a one-model ensemble, and two-model ensembles. A case study and statistical verification, using both deterministic and probabilistic measuring scores, were performed against fog observations from 13 cities in eastern China. The verification was focused on the 12- and 36-h forecasts.

By applying the various approaches, including the new fog detection scheme, ensemble technique, multimodel approach, and the increase in ensemble size, the fog forecast accuracy was steadily and dramatically improved in each of the approaches: from basically no skill at all [equitable threat score (ETS) = 0.063] to a skill level equivalent to that of warm-season precipitation forecasts of the current NWP models (0.334). Specifically, 1) the multivariable-based fog diagnostic method has a much higher detection capability than the liquid water content (LWC)-only based approach. Reasons why the multivariable approach works better than the LWC-only method were also illustrated. 2) The ensemble-based forecasts are, in general, superior to a single control forecast measured both deterministically and probabilistically. The case study also demonstrates that the ensemble approach could provide more societal value than a single forecast to end users, especially for low-probability significant events like fog. Deterministically, a forecast close to the ensemble median is particularly helpful. 3) The reliability of probabilistic forecasts can be effectively improved by using a multimodel ensemble instead of a single-model ensemble. For a small ensemble such as the one in this study, the increase in ensemble size is also important in improving probabilistic forecasts, although this effect is expected to decrease with the increase in ensemble size.

## 1. Introduction

Fog is frequently blamed for traffic disasters and bad air quality in poor-visibility weather and has been extensively studied for more than a century (see the review

by Gultepe et al. 2007). However, progress in the operational forecasting of fog at the National Centers for Environmental Prediction (NCEP) and other numerical weather prediction (NWP) centers has been slow due to the complexity of predicting fog and limited computing resources available for the task. For now, fog is still not a direct model guidance product produced by NWP centers but is diagnosed by local forecasters based either on statistical methods such as model output statistics (MOS; Koziara et al. 1983) and neural network (NN;

---

*Corresponding author address:* Binbin Zhou, NCEP/Environmental Modeling Center, 5200 Auth Rd., Camp Springs, MD 20746.

E-mail: binbin.zhou@noaa.gov

Fabbian et al. 2007; Marzban et al. 2007) or on indirect model output variables (e.g., Baker et al. 2002). The major drawbacks to statistical forecasts are that the models used at NWP centers are frequently upgraded or changed while the statistical approach needs a long period of past forecast data for training and both the MOS and NN approaches are statistical but not flow dependent. The diagnosis of fog from other indirect model output variables strongly depends on the experience of the local forecaster and remains a challenging forecast problem. Thus, there have been growing efforts to numerically predict fog over the last decade, either with local fog models over small areas (e.g., Bott and Trautmann 2002; Bergot et al. 2005) or with NWP models over large domains (e.g., Ballard et al. 1991; Teixeira 1999; Kong 2002; Pagowski et al. 2004; Koracin et al. 2005; Muller 2005; Gao et al. 2007; Toth and Burrows 2008). A recent study done by Roquelaure and Bergot (2008) has shown some promising results in predicting fog using a one-dimensional local ensemble model (not a full NWP model) for the Charles de Gaulle International Airport in Paris, France.

Most of these fog forecasting efforts were deterministic in nature and did not consider forecast uncertainty. Due to the chaotic and highly nonlinear nature of the atmospheric system, initially small differences in either initial conditions (ICs) or the model itself can amplify over time and become large after a certain time period (Lorenz 1965). Since an intrinsic uncertainty always exists in both the ICs and model physics, a forecast predicted by a single model run always has uncertainty. Such forecast uncertainty varies from time to time, from location to location, and from case to case. A dynamical way to quantify such flow-dependent forecast uncertainty is with ensemble forecasting (Leith 1974; Du 2007). Instead of one single integration, multiple model integrations are made, initiated with either multiple slightly different ICs and/or based on different model configurations in an ensemble prediction system. Given the intrinsic uncertainty of the model forecasts and the fact that fog forecasting is believed to be extremely sensitive to the initial conditions and the physics schemes used in a prediction system (Bergot and Guedalia 1994; Gayno 1994; Bergot et al. 2005), it is highly desirable to have fog prediction be part of an ensemble framework. This work is certainly one of the pioneering attempts in the trend toward this new requirement.

To account for the uncertainty in weather forecasts, a global medium-range ensemble prediction system was operationally implemented at NCEP in 1992 (Tracton and Kalnay 1993; Toth and Kalnay 1993, 1997). A regional model-based Short-Range Ensemble Forecast (SREF) system was also developed and operationally

implemented in 2001 (Stensrud et al. 1999; Tracton et al. 1998; Du and Tracton 2001; Du et al. 2003, 2004, 2006). From 2003, with support from the Federal Aviation Administration (FAA), the development of ensemble products based on the NCEP SREF system and particularly tailored to aviation weather forecasting has been carried out. Many aviation weather products have been developed including ceiling, icing, turbulence, visibility, and fog (Zhou et al. 2004). Although subjective evaluations against satellite-detected fog over the continental United States suggested that probabilistic fog forecasts derived from the NCEP SREF system can generally capture fog events (Zhou et al. 2007), no objective and systematic verification has been done yet at NCEP. This work can fill that gap.

During the 2008 Summer Olympic Games, in Beijing, China, a subcomponent of the NCEP SREF system was reconfigured to support daily weather forecasts in China for the event as part of the Research Demonstration Project (hereafter referred as SREF-B08RDP) under the auspices of the World Weather Research Program (WWRP) of the World Meteorological Organization (WMO). Taking advantage of this SREF-B08RDP project, a fog prediction scheme was quantitatively and objectively verified using this mesoscale ensemble data over eastern China to fulfill three goals. The first goal is to examine the effectiveness of a new diagnostic fog-forecasting method compared to a conventional method used in current practice; the second goal is to examine the forecast skill level of current operational NWP models in predicting fog with various approaches, including ensemble technique, multimodel approach, and the increase in ensemble size; and the last goal is to compare the performances of a single-model-based ensemble and multimodel-based ensembles, as well as to examine the impacts of ensemble size on probabilistic forecasts when the ensemble size is small. To the best of our knowledge, this is the first attempt to apply a sophisticated ensemble technique to a state-of-the-art operational NWP model to centrally predict and systematically evaluate this important but difficult and complex low-probability phenomenon although the ensemble technique has been applied to many other weather-related variables such as precipitation, convection, temperature, and cyclones (Du et al. 1997; Hamill and Colucci 1997, 1998; Eckel and Walters 1998; Hou et al. 2001; Stensrud and Yussouf 2003, 2007; Yuan et al. 2005, 2007; Jones et al. 2007; Schwartz et al. 2010; Clark et al. 2009; Charles and Colle 2009). The paper is organized as follows. The ensemble system's configuration is described in section 2, with the fog diagnostic method outlined in section 3. Section 4 summarizes the verification methods and data. Evaluation results and discussions are presented in

section 5, and a summary and plans for future work are given in the last section.

## 2. Configuration of a multimodel mesoscale ensemble prediction system

As part of the WMO/WWRP Research Demonstration Project (on mesoscale ensemble forecasting) for B08RDP (Duan et al. 2009), a subcomponent of the larger NCEP SREF system was reconfigured to the China region and run once per day from 29 January to 7 September 2008. The forecast domain, centered near Beijing (at 40°N, 115°E) and covering 3555 km in the east–west direction (238 grid points) and 2910 km in the north–south direction (195 grid points), covers most of northern and eastern China as shown in Fig. 1. This subsystem, SREF-B08RDP, is a multimodel-based mesoscale ensemble prediction system designed to include physics diversity, which consists of 10 members using two regional models: the Nonhydrostatic Mesoscale Model (NMM; Janjić et al. 2001) component of the National Centers for Environmental Prediction’s (NCEP) Weather Research and Forecasting (WRF) model and the National Center for Atmospheric Research’s (NCAR) Advanced Research version of the WRF (ARW; Skamarock et al. 2005). Each model has five members, one control and four perturbed, to address uncertainty in the initial conditions (ICs). The SREF-B08RDP ran once per day with a forecast length of 36 h initiating at 1200 UTC or 2000 Beijing time (BT, or local time). The control ICs came from the NCEP Global Data Assimilation System (GDAS; information online at <http://www.emc.ncep.noaa.gov/gmb/gdas>). IC perturbations were created using the breeding method (Toth and Kalnay 1993, 1997) and lateral boundary condition (LBC) perturbations were provided by the NCEP Global Ensemble System (GENS; Toth and Kalnay 1993). The horizontal resolution of both models is 15 km. The vertical resolutions are set at 52 sigma levels, with the lowest values at 1.0, 0.9925, 0.9840, 0.9744, etc., for NMM and 51 sigma levels at 1.0, 0.9938, 0.9864, 0.9778, etc., for ARW. The sigma values for the lowest levels indicate that the lowest vertical resolutions (above the surface) for both models are equivalent to about 50 m.

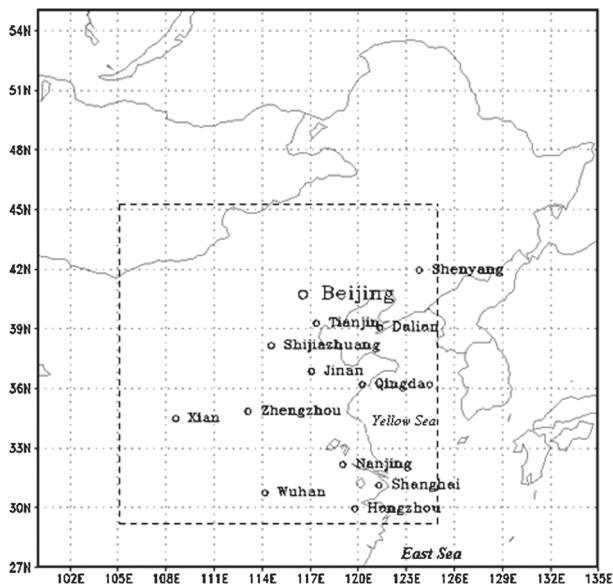


FIG. 1. NCEP SREF-B08RDP ensemble forecast model domain (outer) and the forecast output domain (inner) with the locations of the 13 fog verification cities.

The physics schemes employed by the SREF-B08RDP are listed in Table 1. In addition to the difference in the dynamic cores, the following physics are also different in the two models: convection, planetary boundary layer (PBL), surface boundary layer, and long- and shortwave radiation. The cloud microphysics and land surface schemes are the same in both models. For convection the Betts–Miller–Janjić scheme (BMJ; Janjić 1994) is used in NMM and the Kain–Fritsch scheme (KF, Kain and Fritsch 1990) in ARW. For the PBL, the Mellor–Yamada–Janjić scheme (Janjić 1994) is used in NMM and the Yonsei University scheme (YSU; Hong and Dudhia 2003) in ARW. For the surface boundary layer, the Janjić similarity scheme (Janjić; Janjić 1996) is used in NMM and the classical Monin–Obukhov scheme is employed in ARW. For longwave radiation the Geophysical Fluid Dynamics Laboratory scheme (GFDL; Schwarzkopf and Fels 1991) is used in NMM and the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) in ARW. For shortwave radiation the GFDL scheme (Lacis and Hansen 1974) is used in NMM

TABLE 1. Configuration of the NCEP B08RDP ensemble forecasting system.

Models (members)	Convection physics	Microphysics	Resolution (km) and levels	PBL	Surface boundary layer	Base ICs and BCs	Longwave radiation	Shortwave radiation	Land surface
WRF NMM (five)	NCEP (BMJ)	Ferrier	15 (52)	MYJ	Janjić	GDAS (GENS)	GFDL	GFDL	Noah
WRF ARW (five)	NCAR (KF)	Ferrier	15 (51)	YSU	Monin–Obukhov	GDAS (GENS)	RRTM	Dudhia	Noah

and the Dudhia shortwave transfer model (Dudhia 1989) in ARW. Ferrier's microphysics cloud scheme (Ferrier et al. 2002) and the Noah land surface model (Noah; Ek et al. 2003) are used in both models.

### 3. A multivariable-based diagnostic method for fog detection

Although one hopes that the liquid water content (LWC) at the lowest model level can be explicitly used as fog, experience indicates that an LWC-only approach does not work well with the current NWP models due mainly to two reasons: one is the too coarse model spatial resolution and the other is a lack of sophisticated fog physics. Therefore, it is necessary to find a fog detection scheme based on other model output variables. In such a scheme, one should expect that the performance of a fog forecast depends on how the fog is detected. In previous studies, LWC at the lowest model level was commonly used to represent fog (hereafter referred to as the LWC-only approach). However, these fog predictions or simulations were for case studies in which the model resolutions could be very high. For operational models such as at NCEP, model resolutions are still too coarse to properly resolve important physics for fog near the surface due to computing resource limitations. Additionally, physical schemes or parameterizations in the operational models are not designed for near-ground fog but for precipitation or clouds at higher levels. As a result, the LWC from the models is usually not reliable enough to represent fog and tends to seriously underforecast fog in many cases (G. Toth 2009, personal communication). To better detect fog, other variables besides LWC were considered to enhance the performance of the forecasting.

Fog forecasting is in essence the prediction of visibility in foggy conditions. However, the commonly used visibility computation in fog has a very large error (over 50%) as shown by Gultepe et al. (2006, 2009). The unreliable LWC from the operational models causes an even larger visibility computation error. Due to these two facts, only fog occurrence, and not fog intensity (visibility), is diagnosed in the current SREF system. Below is a description of a new multivariable-based approach we propose in the postprocessor of the SREF system for fog forecasting.

It is well known that fog has different types with different creation mechanisms (Tardif and Rasmussen 2008). Some build from stratus subsidence, some from advection, and some from radiative cooling near the ground. To deal with different fog types, the proposed diagnosis consists of the following criteria or rules with appropriate thresholds:

LWC at the model lowest level  $\geq 0.015 \text{ g kg}^{-1}$  or, (1a)

cloud top  $\leq 400 \text{ m}$  and cloud base  $\leq 50 \text{ m}$ , or (1b)

10-m wind speed  $\leq 2 \text{ m s}^{-1}$  and 2-m RH  $\geq 90\%$ . (1c)

This diagnosis is similar to the conceptual scheme suggested by Croft et al. (1997). The LWC rule in (1a) came from the definition of fog visibility range. With Kunkel's equation (Kunkel 1984),  $\text{LWC} \geq 0.015 \text{ g kg}^{-1}$  is equivalent to visibility  $\leq 1000 \text{ m}$ , the definition of fog from the World Meteorology Organization (WMO 1966).

The cloud-top threshold in (1b) follows the general features of fog. Observations indicated that the depth of most fogs on land is about  $100 \sim 200 \text{ m}$ . Some marine fogs or advection fogs are deeper, but rarely exceed  $400 \text{ m}$ . The cloud-base threshold in (1b) reflects the lowest level of our models. The current NMM and ARW used in the SREF system have a vertical grid spacing of about  $50 \text{ m}$  near the ground (the cloud bases and tops are defined as cloud  $\text{LWC} \sim 10^{-3} \text{ g kg}^{-1}$  in both the NMM and ARW models). At a grid point, if the cloud base touches the lowest model level and the cloud top is less than  $400 \text{ m}$  as well, the cloud at this location is assumed to be fog. An evaluation with satellite-detected fog data over the conterminous United States (CONUS) showed that the full NCEP SREF system using this cloud rule can diagnose large-scale fog events well, particularly marine fog or coastal fog, but not shallow fog or ground fog because this type of fog usually builds upward from the ground and may lie below the lowest model level (Zhou et al. 2007).

To deal with ground fog, the RH-wind rule (1c) was included. Choosing general and centralized thresholds for surface wind and RH over large domains in a model is more difficult than for the LWC and cloud rules because 1) ground fog is more local and 2) different models have different RH and wind biases. In many cases fog was reported while the model RH was less than 100%. On the other hand, ground fog is generally a radiation or radiation-related type of fog. Thus, weak turbulence is a necessary condition for this type of fog. With appropriate thresholds for RH and for turbulence intensity [e.g., those suggested for radiation fog by Zhou and Ferrier (2008)], grid-scale ground fog in a model can be diagnosed. Unfortunately, the turbulence intensity was not output from the B08RDP's postprocessor. An alternative approach is to use a combination of surface RH and wind speed. Surely, since there is no quantitative relationship between wind speed and turbulence intensity, which threshold value should be used for wind

TABLE 2. ETSs of fog forecasts using different combinations of 2-m RH and 10-m wind speed thresholds for (a) the NMM (ARW) control runs in Feb 2008 and (b) the average of the NMM and ARW control runs during Feb–Aug 2008. The highest ETS of all combinations is shown in boldface.

(a)				
Feb	≥85%	≥90%	≥95%	=100%
≤1 m s <sup>-1</sup>	0.036 (0.036)	0.038 (0.032)	0.041 (0.018)	0.041 (0.018)
≤2 m s <sup>-1</sup>	0.102 (0.077)	<b>0.211 (0.109)</b>	0.080 (0.036)	0.041 (0.018)
≤3 m s <sup>-1</sup>	0.084 (0.061)	0.149 (0.065)	0.116 (0.023)	0.041 (0.018)
(b)				
Feb–Aug	≥85%	≥90%	≥95%	=100%
≤1 m s <sup>-1</sup>	0.113	0.095	0.095	0.095
≤2 m s <sup>-1</sup>	0.094	<b>0.192</b>	0.112	0.095
≤3 m s <sup>-1</sup>	0.054	0.133	0.126	0.095

speed in the diagnosis is somewhat empirical and needs to be tuned based on past data. Local forecasters usually use 2-m RH (>90% ~ 100%, some use dewpoint temperature) and 10-m wind speed (<2 ~ 3 m s<sup>-1</sup>) to check for local fog, depending on the location and model employed. For centralized fog forecasting in this study, the optimized RH and wind thresholds of 90% and 2 m s<sup>-1</sup> for both NMM and ARW, shown in Eq. (1c), were obtained through a tuning iteration procedure using the NMM and ARW control forecasts during February 2008. During the process of iteration, the thresholds for the LWC and cloud rules were kept the same but the thresholds for RH and wind speed were tuned around the empirical value ranges. Then, with the tuned thresholds, fog forecasts from both the NMM and ARW control runs in the iteration were compared with the fog data in 13 cities over eastern China to acquire overall forecast skill scores. The forecast skill score computations and the observed fog data will be discussed in the next section. After the tuning procedure was run through all of its iteration steps, the thresholds for RH and wind were selected from the best performance scores (Table 2). Table 2a shows that the combination of 90% RH and 2 m s<sup>-1</sup> 10-m wind speed gave the highest equitable threat score (ETS) for both the NMM (ETS = 0.211) and ARW (0.109) control runs in February. The selected thresholds were then applied to the rest of the verification period (March–August) for fog forecasting. To examine if the February-based thresholds were also valid for other months, ETSs were calculated for other months with the all threshold combinations. The mean ETSs averaged over the entire 7-month period (February–August) are listed in Table 2b, which shows that although the thresholds were tuned using the February data, they seem to work equally well for the other months too with the highest mean ETS (0.192) associated with the combination of the 90% 2-m RH and 2 m s<sup>-1</sup> 10-m wind speed thresholds. Note that since both the NMM and ARW control forecasts had similar patterns of behavior

in the tuning process and they shared the same optimal RH–wind thresholds (see Table 2a), what are listed in Table 2b are the mean scores averaged over the two models.

Now, let us evaluate the influences of the RH–wind rule on fog forecasting since this process should be insightful for other similar works in the future. To be more representative, the results based on the entire time period and the two models (i.e., Table 2b) are discussed here. If the RH threshold is too large (~100%) or the wind threshold too small (~1 m s<sup>-1</sup>), the ETS value generally hits the 0.095 bound, implying that the RH–wind relation has no impact on fog forecasts and only the clouds and LWCs play roles under such circumstances. On the other hand, if the threshold for RH is too low or the wind is too strong (e.g., 85% and 3 m s<sup>-1</sup>), the overall ETS is only 0.054 due to too many false alarms. This is even lower than the value of 0.063, which is the overall ETS of the LWC-only approach (see Fig. 2d). The ETS of the LWC-only indicates that inappropriate RH–wind thresholds may cause a negative contribution to the diagnosis. Since the overall ETS with the LWC and cloud diagnosis (without the impacts from the RH–wind relationship) is 0.095, as shown in Table 2, the RH–wind’s contribution to the overall ETS can range from 0.097 (0.192–0.095) to 0.192. The overall ETS for the cloud rule alone can range from 0.032 (0.095–0.063) to 0.095. Thus, the contributions of the LWC, cloud, and RH–wind rules to the forecast score are 0.063, 0.032–0.095, and 0.097–0.192, respectively. These numbers roughly reveal the relative contributions to a fog forecast by the different rules used in diagnosing fog [Eq. (1)] from the NMM and ARW models in this study. The ETS indicates that the RH and wind variables are critical to a successful fog forecast using a coarse-resolution model with the fog physics missing. By the way, the large variations in ETS either at a fixed column (except for the 100% RH) or row (except for the 1.0 m s<sup>-1</sup> wind) in Table 2 imply that RH-only or wind-only rules would not work well.

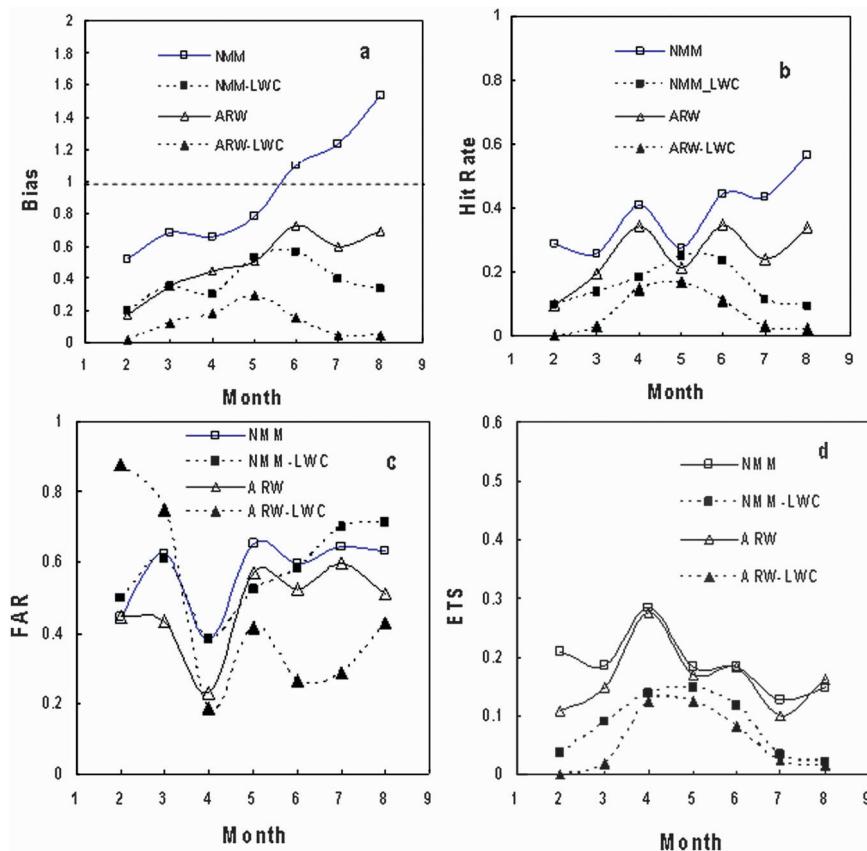


FIG. 2. Monthly scores of (a) bias, (b) HR, (c) FAR, and (d) ETS for the single NMM and ARW control runs using the “multivariable” and “LWC only” methods.

Under the selected thresholds in Eq. (1), fog forecasting from the NMM and ARW control runs and ensembles can then be conducted in the following study. However, can this new method be easily adapted to other NWP models? Although the diagnosis certainly depends on the models and training data, as already discussed above, it does not seem to be that difficult to apply this fog method to other models. Our discussion has mentioned that the LWC criteria [Eq. (1a)] can be generally used (no tuning is needed) and the cloud-base-top criteria [Eq. (1b)] can also be easily determined by the model vertical resolution setting (no tuning either). Although the RH-wind rule [Eq. (1c)] is important and needs a simple tuning based on a particular model and its performance, the fact that the same rule worked quite well for both the NMM and ARW and all months from February to August (Table 2) suggests that the RH-wind rule is not that particularly sensitive to the model or season. Therefore, although the threshold values in Eq. (1) are optimized for NMM and ARW models over eastern China, we believe that they can be used as approximate criteria for other models too (the RH-wind rule can at

least be used to provide reference values for a simple tuning such as the 12 unique combinations involved in our tuning process). As for a very high-resolution model with sophisticated fog physics, the cloud and/or RH-wind rule might be removed (but that remains to be proven). However, without sophisticated fog physics (not currently available), the cloud and RH-wind rule might still well be needed even in a finescale model.

#### 4. Verification methodology and data

##### a. Verification method

The following scores are used to verify a deterministic forecast against an observation of a binary event (yes or no; 1 or 0), where  $F$ ,  $H$ , and  $O$  refer to the numbers of forecast points, correctly forecast points (hits), and observed points, while  $N$  is the total number of points in a verification domain:

$$\text{hit rate [HR, or probability of detection (POD)]} = \frac{H}{O}, \quad (2)$$

false alarm ratio [FAR, or probability of false alarm and

$$(\text{POFA})] = \frac{F - H}{F}, \tag{3}$$

false alarm rate [FARt, or probability of false

$$\text{detection (POFD)]} = \frac{F - H}{N - O}, \tag{4}$$

$$\text{missing rate (MR)} = \frac{O - H}{O}, \tag{5}$$

$$\text{correct rejection rate (CRR)} = \frac{N - O - F + H}{N - O}, \tag{6}$$

$$\text{bias} = \frac{F}{O}, \tag{7}$$

$$\text{threat score (TS)} = \frac{H}{F + O - H}, \tag{8}$$

and

$$\text{ETS} = \frac{H - R}{F + O - H - R}, \tag{9}$$

where  $R = F \times O/N$  is a random hit penalty. Since the bias is defined here as the ratio of total forecast points to total observed points, the best value for the bias is 1.0, while values less than 1.0 indicate underprediction and values greater than 1.0 indicate overprediction. TS or ETS is a positively oriented score; that is, the larger the TS or ETS, the better a forecast will be. From the formula one can see that ETS is a more restricted version of TS. For those who are not familiar with this score, one might keep the current capability in predicting precipitation in mind as a reference when comparing the results described in section 5. Using a current state-of-the-art NWP model, ETS is about 0.3 in predicting warm season precipitation and 0.4 in predicting cold season precipitation. Note that for a particular given probability threshold, a probability forecast can also be viewed as a deterministic forecast. Therefore, probabilistic forecasts could also be verified deterministically, which will be shown in section 5.

For verifying probabilistic forecasts, the following four scores—Brier score (BS), Brier skill score (BSS), relative operating characteristic (ROC), and reliability diagram—are used:

$$\text{BSS} = 1 - \frac{BS}{BS_r}, \tag{10}$$

where

$$\text{BS} = \frac{1}{n} \sum_i \sum_j (p_{ij} - o_{ij})^2, \tag{11}$$

$$\text{BS}_r = \frac{1}{n} \sum_i \sum_j (f_{ij} - o_{ij})^2, \tag{12}$$

where  $i$  and  $j$  denote, respectively, the verifying point (out of the total points) and the forecast number (out of the total number of forecasts);  $n$  is the total number of forecast–observation pairs for all forecasts over all verifying locations;  $p_{ij}$  ( $o_{ij}$ ) is the forecast (observed) probability at location  $i$  from the  $j$ th forecast, where the forecast probability ranges in value from 0 to 1 but the observed probability is either 1 (yes event) or 0 (no event);  $f_{ij}$  is a reference forecast at location  $i$  from the  $j$ th forecast, with a value that either ranges from 0 to 1 (if a probabilistic forecast is used) or is a 1 or 0 (if a single deterministic forecast used). In this study, a single deterministic (control) forecast will be used as a reference. Obviously, the perfect BS is 0.0 and the perfect BSS is 1.0. The smaller the BS or the larger the BSS, the better a probabilistic forecast will be. A positive BSS indicates a skillful probabilistic forecast with respect to its reference forecast; otherwise, the probabilistic forecast has no skill.

The Brier score can be further decomposed into three components to reflect the characteristics of the joint distribution as derived by Murthy (1973):

$$\text{BS} = \text{BS}_{\text{rel}} - \text{BS}_{\text{res}} + \text{UNC}, \tag{13}$$

where  $\text{BS}_{\text{rel}}$  is the reliability,  $\text{BS}_{\text{res}}$  the resolution, and UNC the uncertainty. The meaning of the reliability and resolution will be briefly mentioned in the following discussion about reliability diagrams. The uncertainty reflects the underlying uncertainty of the observations. Since BS is negatively oriented, a good probabilistic forecast (i.e., a smaller value of BS) should have a smaller value of reliability ( $\text{BS}_{\text{rel}}$ ), a larger value of resolution ( $\text{BS}_{\text{res}}$ ), and a smaller value of uncertainty (UNC).

A ROC curve is generated by plotting the HR against the FARt (POFD) in a  $1.0 \times 1.0$  coordinate system [with FARt (POFD) on the  $x$  axis and HR on the  $y$  axis] for all possible decision probability thresholds (Mason 1982; Toth et al. 2003). Each point on the ROC curve represents a pair of (HR, FARt, or POFD). If the curve is above the diagonal line (climatology), it indicates a skillful forecast; otherwise, there is no skill compared to climatology. Normally, forecast accuracy can be quantified by estimating the area enclosed below a ROC curve. Thus, if a ROC area is  $>0.5$ , the forecast is considered skillful. Since the best values for HR and FARt (POFD) are 1.0 and 0.0, respectively, the perfect ROC area is 1.0. In verifying probabilistic forecasts, HR–FARt (POFD)

TABLE 3. Number of fog-observed days for the 13 cities in eastern China from Feb to Aug 2008.

City	Feb	Mar	Apr	May	Jun	Jul	Aug	Total
1) Shenyang (SY)	0	2	1	2	4	5	5	19
2) Beijing (BJ)	0	1	1	2	4	5	4	17
3) Tianjin (TJ)	2	2	1	4	1	12	4	26
4) Shijiazhuang (SJZ)	0	3	3	4	2	4	3	19
5) Zhengzhou (ZZ)	1	1	4	0	0	6	4	18
6) Dalian (DL)	0	2	3	4	4	7	3	23
7) Qingdao (QD)	1	1	3	5	9	8	2	29
8) Jinan (JN)	0	1	0	1	2	2	2	8
9) Xian (XA)	1	2	2	1	2	1	0	9
10) Nanjing (NJ)	4	2	3	0	5	3	2	19
11) Shanghai (SH)	5	1	5	3	4	1	3	22
12) Hangzhou (HZ)	5	2	4	3	4	2	2	22
13) Wuhan (WH)	3	2	4	1	1	1	1	13
Total	22	22	34	30	42	57	35	242

pairs are calculated at all given decision probability thresholds to construct the ROC curve.

The ROC curve, however, does not provide a full depiction of the joint distribution of forecasts and observations (Wilks 2006). Specifically, it does not reveal the reliability aspect although it is a good indicator of the resolution aspect of a forecast system (Toth et al. 2003), where resolution refers to the ability of a forecasting system to distinguish a forecast from averaged observation data or climatology. Reliability (also called statistical consistency) is another important quality of a probabilistic forecast, and measures whether the forecast probability is statistically consistent with the observed frequency of the occurrence of the event considered over a long period of time or over many cases. Therefore, a *reliability diagram* (also called the probability bias) is also verified in this study and can be obtained by plotting an equal-interval predicted probability ( $x$  axis) against the observed frequencies ( $y$  axis). A reliable probabilistic forecast should yield a reliability curve close to the diagonal line, which indicates a perfectly reliable probabilistic forecast. A reliability diagram can provide additional information associated with resolution, data uncertainty, and skill margins (see section 5c).

### b. Verification data

Because of a current lack of fog analysis data, it is impossible to verify fog forecasts grid point by grid point over the entire domain. Instead, 13 big metropolitan areas are chosen for verification and tuning in this study (Fig. 1). Since the forecast is binary, there was no interpolation from the grid points to the observation stations but the forecasts at grid points nearest to the observation stations were verified. Although the number of sampled cities seems small, they are uniformly spread over the eastern part of the forecast domain

where fog occurrence is most frequent. As a matter of fact, there were a total of 242 foggy days during the verification period (Table 3), which should be large enough in temporal space to compensate for an insufficient number of spatial sampling points. Furthermore, fog is a local weather phenomenon that is strongly influenced by factors such as terrain, local flows, and local surface boundary layer conditions and therefore may well be a subgrid-scale event on many occasions. As a result of this, a model with coarser horizontal resolution may not capture all fog events well. Thus, a robust assessment of the systematic performance of fog forecasts can only be obtained from verifications over a long period of time or over a large number of cases. In this study, a total of about 5460 forecasts [ $7 \text{ months} \times 30 \text{ days} \times 2 \text{ forecasts per day (f12h and f36h)} \times 13 \text{ cities}$ ] were used in the verification to make sure that the verification results are representative. Considering that each forecast was actually predicted by 10 ensemble members, the total number of forecasts reached 54 600.

Fog is a relatively frequent weather phenomenon over eastern China, particularly along the coast. The frequency of dense fog events in some coastal cities can reach as high as 50% in the foggy season (Liu et al. 2005). The verification data for this study were daily fog reports issued by local weather services or airports in 13 cities in eastern China from February to September 2008 during the SREF-B08RDP operating period. Since the observational data were reported only for morning fog events, the verification had to be done on both the 12- and 36-h forecasts of a particular cycle, which correspond, respectively, to 0800 BT on the first and the second days after the model initiation time (1200 UTC or 2000 BT). The foggy days in the 13 cities during the 7 months are summarized in Table 3 and include both dense fog (visibility  $< 500 \text{ m}$ ) and light fog (visibility between 500 and 1000 m) events. There were a total of 242 fog events observed with an average monthly fog occurrence frequency of nearly 8.8% (or about 2.7 days per month) at any given city. This percentage was used as the fog climatology level in the reliability diagram (Fig. 10) to determine the *resolution* of fog forecasting in this study.

What is the quality of this observational dataset? Table 3 shows a seasonal and geographic distribution of fog events over eastern China: the east coast has more fog events and the western interior lands fewer; the southeast coast has more fog in the cold season and the northeast coast more fog in the warm season. This seasonal and geographic distribution demonstrated by Table 3 is generally in agreement with 50 yr of fog statistics in China, as reported by Liu et al. (2005). Two cities on the northeast coast, Tianjin and Qingdao, are well recognized as “foggy” cities in China due to their specific locations,

where continuous marine fog events frequently appear under dominant high pressure weather systems during late spring and early summer when warm and moist air flows northward over cold waters or cold air flows southward over warm waters along the coast. This fact is indeed reflected in Table 3. For example, Tianjin reported 12 foggy days in July and Qingdao reported 9 days in June and 8 days in July, which were significantly more than other cities during the same months. All this suggests that the observational fog data collected are reliable. By the way, based on Liu et al. (2005), the highest frequency of fog occurrence over eastern China is in the autumn–winter period (November–January), which is, unfortunately, outside of the SREF-B08RDP operating period. Otherwise, more fog events would have been evaluated.

## 5. Verifications and discussions

### *a. Effectiveness of the new fog detection method*

Since most of the current fog forecasts from NWP models use the model's lowest-level LWC as fog, it will be meaningful to compare how well the multivariable-based diagnosis performs against the LWC-only approach. Figure 2 shows the monthly verification statistics from the NMM and ARW control forecasts (note that given no distinguishing differences in performance between the 12- and 36-h forecasts, all the statistical scores throughout this paper were averaged over both the 12- and 36-h forecasts to increase the sample size). Although all forecasts (with the exception of the multivariable-based NMM forecast in the summer season) underestimate the number of fog events, the multivariable-based method exhibits less bias (is closer to value 1.0) than the LWC-only method for both models (Fig. 2a). This can be further confirmed from the HR plot shown in Fig. 2b where the multivariable-based approach has a consistently higher hit rate than the LWC-only approach, except for the month of May. The hit rate of the LWC-only approach is particularly low (near zero) for the ARW model in the winter and summer seasons, which demonstrates that the LWC-only method is not reliable, although the LWC-only method was only slightly worse than the multivariable-based method in May. Given the fact that more diagnostic parameters are included by the logical operator or in the new method [Eq. (1)], it is expected that the FAR might increase for the new approach. Figure 2c shows, however, that in the new approach such an increase in FAR is not that significant; FAR remained at a similar level for the NMM model in all months, while for the ARW model it became better for February–April and worse for May–August. Combining both HR and FAR, a comprehensive measure is given by ETS (Fig. 2d), which clearly shows that the new

multivariable-based approach is superior to the LWC-only approach in predicting fog events. On average (for the two models and the 7 months), the improvement in ETS reached 205%, increasing from 0.063 to 0.192. Therefore, this new multivariable-based fog detection algorithm will be employed in the rest of this study.

To examine in detail the effectiveness of different diagnostic rules and how well the multivariable diagnosis in fog forecasting does from a single model, a large-scale fog event is presented below. The event occurred over a large area along the east coast of China on the morning of 7 April 2008 (Fig. 3a). Besides the areas of observed fog, the 9-h forecasts of surface wind speeds, 2-m RH, cloud tops and bases, and the LWC from the NMM control are also displayed in Fig. 3. The sea level pressure map (not shown) indicates that the entire coastal region from south to north was controlled by a steady high pressure system centered on the Beijing–Tianjin region. This high pressure system caused a stable planetary boundary layer and a weak surface wind environment along the coast (Fig. 3b). The air over the water was nearly saturated, as shown in Fig. 3c, while the surface temperature gradually decreased from south to north (not shown). The cloud-top and -base forecasts are presented in Figs. 3d and 3e, where the blank areas indicate no cloud, showing that the sky that night was cloud free along the southern coastal region. Figure 3b indicates that the surface wind directions were mostly southeasterly (warm and moist) over land and northwesterly (cold) over the water. Over land, the southeasterly wind brought warm and moist air toward the north (warm advection), which was gradually cooled down during its transport and was further cooled down by strong radiative cooling near the ground along the southern coast during the night of 6 April (clear sky under high pressure). Over the water, on the other hand, the northwesterly cold-air movement (cold advection) cooled the near-saturated air above the water down to its condensation temperature. Under such a favorable combination of wind, temperature, and humidity conditions, a large-scale marine-radiation fog episode developed both on land and over the water along the coast from Hangzhou and Shanghai in the south to Qingdao, Tianjin, Dalian, and Beijing in the north, during the early morning of 7 April as shown in Fig. 3a. The large-scale fog region shown in Fig. 3a is a composite of the observed fog areas derived from the following sources: fog observation mosaics issued by the National Meteorological Center of China, fog reports by local weather stations, and fog images from satellites issued by the National Satellite Meteorological Center of China. The fog mostly dissipated over land after 1000 BT, although it still remained over the water, as can be seen from satellite images (not shown). It was not clear

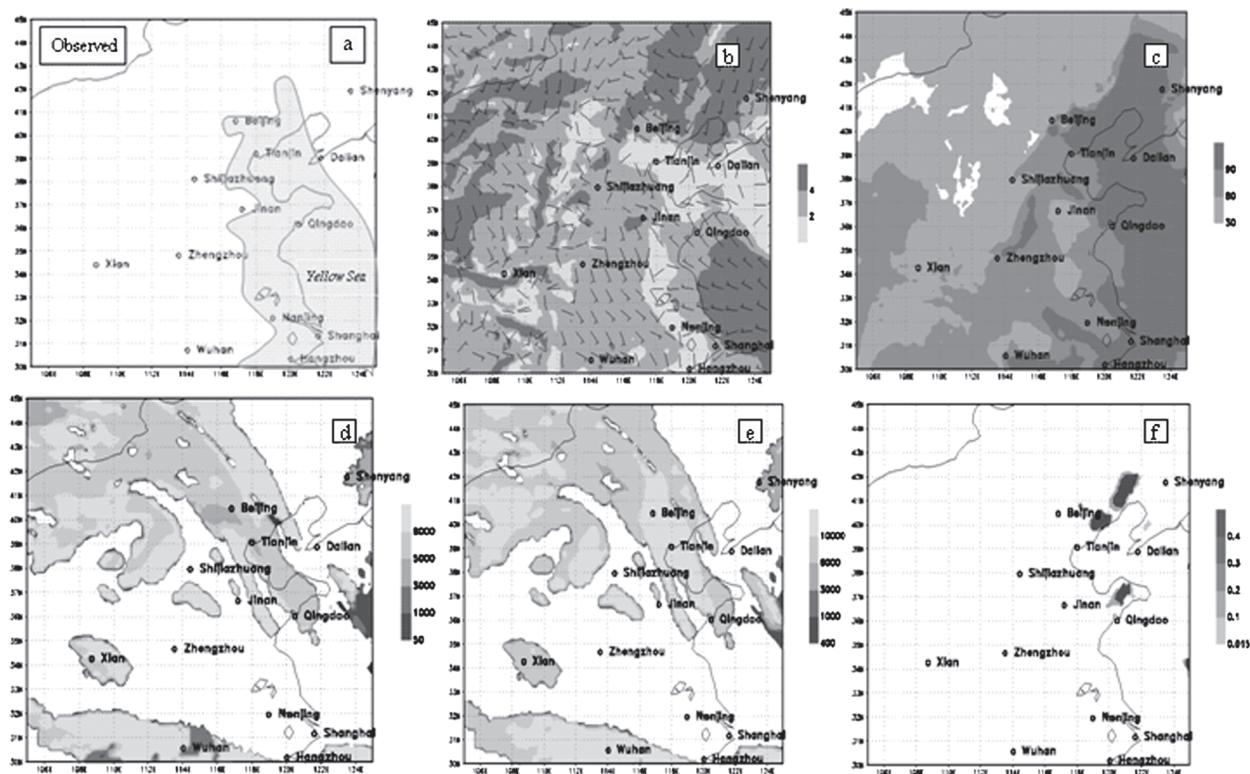


FIG. 3. (a) Observed fog episode at 0500 BT 7 Apr 2008; 9-h forecasts (verified at 0500 BT 7 Apr 2008) of (b) 10-m wind speeds and directions, (c) 2-m RH, (d) cloud base, (e) cloud top, and (f) LWC distributions, from the NMM control run.

whether or not fog developed in the northeast area near Dalian because no fog data were available from there for this fog event. During this large-scale fog episode, numerous traffic interruptions including local traffic tie-ups, shutdowns of several highways, closures of sea harbors, and hours-long delays for many airlines were reported in all the affected cities. Several casualties from a series of fatal car accidents on highways were also reported by local police offices. The societal impacts of accurate fog forecasting are high in such major events.

However, this fog episode was mostly missed by the LWC-only forecasts from the NMM (Fig. 3f) and the ARW (not shown). If the LWC-only approach is used, the fog forecast in this case would be significantly underpredicted. From the above analysis, we can expect that the fog was of the advection type over water and of the radiation type on land along the southern coast. But the LWC-only approach failed to detect both types of fog over most of the region. By using the multivariable diagnosis, however, the fog areas derived from the NMM and the ARW control forecasts were obviously expanded as shown in Figs. 4a and 4b, respectively. By examining the NMM control's forecast distributions of 2-m RH, 10-m wind speeds, and cloud bases and tops in Figs. 3b–e, one can observe that the surface RH and

wind speed on land near Shanghai and over the water to the south of Dalian met the “RH–wind” thresholds while the cloud bases and tops over the water to the east of Qingdao met the “cloud top and base” thresholds. As a result, the NMM with the multivariable diagnosis successfully predicted the fog events in Shanghai and over the water to the south of Dalian and the east of Qingdao, as shown in Fig. 4a (which were missed by LWC-only approach in Fig. 3f). This fog case demonstrates that by using the LWC-only approach the single models in the SREF system would seriously underforecast the fog, while using the multivariable diagnosis would greatly improve the forecast. Although the forecast fog areal coverage was still smaller than the observed, it is much better than the forecast made with the LWC-only approach (cf. Figs. 3f and 4a).

#### *b. Single deterministic forecast versus ensemble forecasts*

As with the two control forecasts, fog occurrence (1 or 0) can also be diagnosed for all perturbed ensemble members. Based on all individual member forecasts, the probability (relative frequency) of fog occurrence can be calculated. A probabilistic forecast can be evaluated both probabilistically and deterministically. For a given

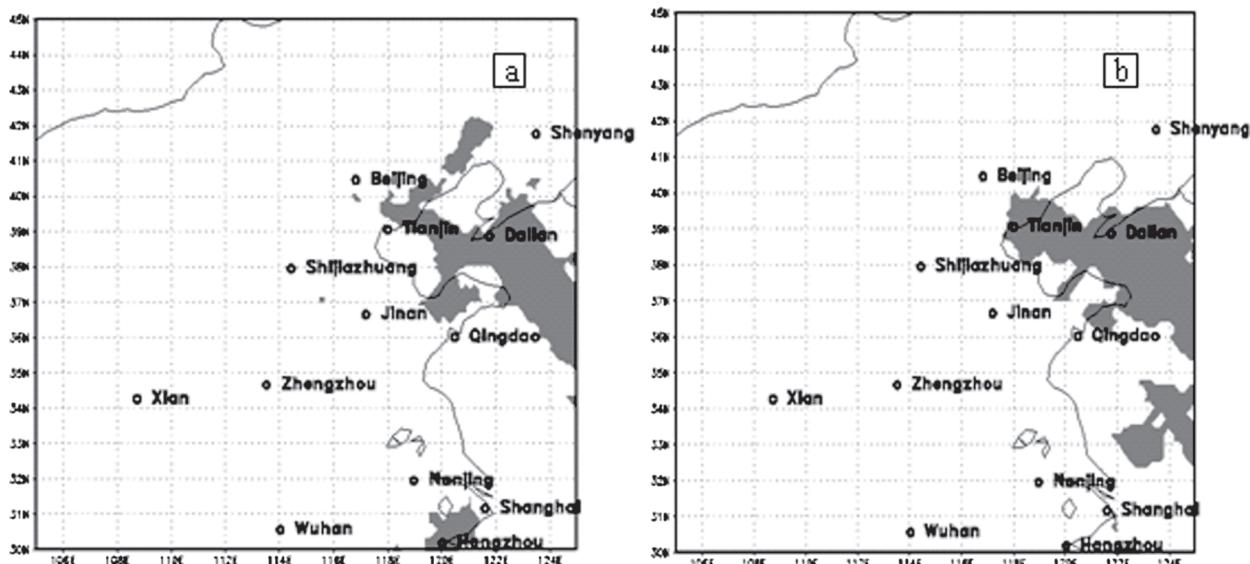


FIG. 4. The 9-h single forecasts of the 7 Apr 2008 fog episode (all verified at 0500 BT 7 Apr 2008) of the (a) NMM and (b) ARW control runs.

specific percentage threshold (such as 10%), a probabilistic forecast can be viewed as a deterministic forecast in the way that an event is expected to occur when the forecast probability is greater than or equal to the selected threshold. Figure 5 shows the fog prediction skill verified with deterministic measures (HR, FAR, MR, CRR, bias, and ETS) of the five subgroups that are the NMM and ARW controls, 5 NMM member-, 5 ARW member-, and the 10 SREF-B08RDP (NMM + ARW) member-based probabilistic forecasts. Here, the probabilistic forecasts were treated as deterministic forecasts with the percentage thresholds fixed at 20%, 40%, 60%, 80%, and 100%. Theoretically, the ensemble mean or median (the 50% probability) forecast should have the most skill on average (Leith 1974), so verifying the median forecast is certainly desirable. Unfortunately, since the prevailing number of ensemble members in this study is five, the 50% probability threshold is not available. Therefore, we have to use either a 40% or a 60% forecast to approximate the ensemble median forecast. For a complete picture, the 20%, 80%, and 100% probability thresholds were also verified. By comparing the single control forecasts and their corresponding (same model) ensemble-based forecasts, the benefits from the ensemble-based forecasts are obvious and are true for both models. For example, the probabilistic forecast HR (MR) is much higher (lower) than for the NMM and ARW control forecasts when the probability threshold is less than 60% (Figs. 5a and 5c). The penalty for this is, however, a slight increase in FAR especially at the 20% threshold (Fig. 5b). But the overall combined score measured by

ETS (Fig. 5f) is a net gain at thresholds of less than 60% for the probabilistic forecasts over the single control forecasts, where the 40% probabilistic forecasts performed the best. Bias scores also seem to suggest that the 40% probabilistic forecast is the best because of an overforecasting tendency when thresholds are less than 40% and an underforecasting tendency when they are greater than 40% (Fig. 5e). Apparently, the bias parallels the behavior of HR (MR). Because of the shrinkage in the forecast area, HR (MR) should be expected to decrease (increase) with the increase in probability thresholds. Therefore, at the high end of the probability thresholds (80% and higher), the HR (MR) became worse for probabilistic forecasts although a reduced false alarm ratio (FAR) was a natural reward in return (Figs. 5a–c). Considering that fog is a relatively rare event, on many occasions there should be no fog in both the forecast and the observation, which implies that the correct rejection rate (CRR) must be quite high for all kinds of fog forecasts. In other words, CRR will be less sensitive to which model groups or probability thresholds are selected. This characteristic is indeed shown in Fig. 5d where CRR, unlike other scores, shows less variation over the different probability thresholds and retains similar values between the single control forecasts and the probabilistic forecasts. The above results demonstrate a clear benefit from the ensemble approach over a single deterministic run in two ways. One is that a much improved deterministic forecast can be achieved by using a forecast close to the ensemble median (50% probability), such as the 40% threshold in this study. The improvement

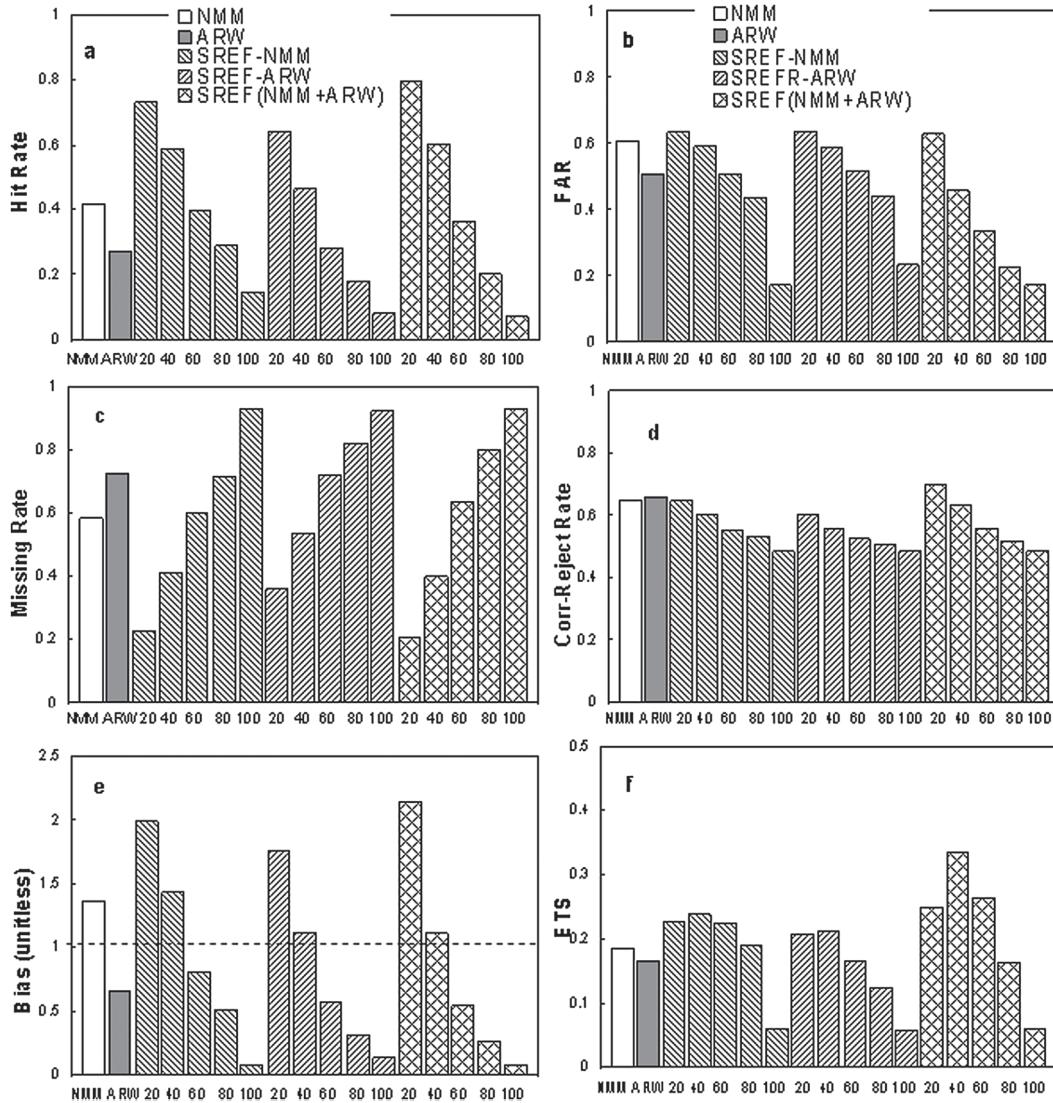


FIG. 5. Averaged scores of (a) HR, (b) FAR, (c) MR, (d) CRR, (e) bias, and (f) ETS of the five subgroups, which are, from left to right, two single control forecasts from NMM and ARW, three groups of probabilistic forecasts (each over the 20%, 40%, 60%, 80%, and 100% thresholds) from the 5-member NMM ensemble, 5-member ARW ensemble, and 10-member multimodel SREF-B08RDP ensemble.

with the 40% forecasts over the single control forecasts in ETS averaged over the two models is a 17.2% increase from 0.192 to 0.225 (see Fig. 5f). Second, the ensemble-based forecast can provide useful information to various types of users with their own unique requirements, objectives, and economic values. For example, some users may prefer a higher hit rate and need not worry about the false alarm rate, while others may be the opposite. This is a situation that ensemble-based forecasts can serve well but a single forecast cannot.

To further demonstrate the value of probabilistic forecasts over a deterministic forecast, probability itself was also evaluated in terms of a probabilistic score. Figure 6a

shows the BSS over each month for both the NMM-ensemble-based and ARW-ensemble-based probabilistic forecasts, where the control run of the corresponding base model was used as the reference forecast for each model. Clearly, both ensembles show skill over their own single control forecast for the entire verification period from February to August. The mean BSS averaged over all 7 months is shown in Fig. 6b. In addition to Fig. 6a, the BSS in Fig. 6b was also calculated using the other model's control forecast as the reference. Since the skill was systematically reduced for all ensembles when switching the reference from the ARW control forecast to the NMM control forecast, this indicates that the single NMM

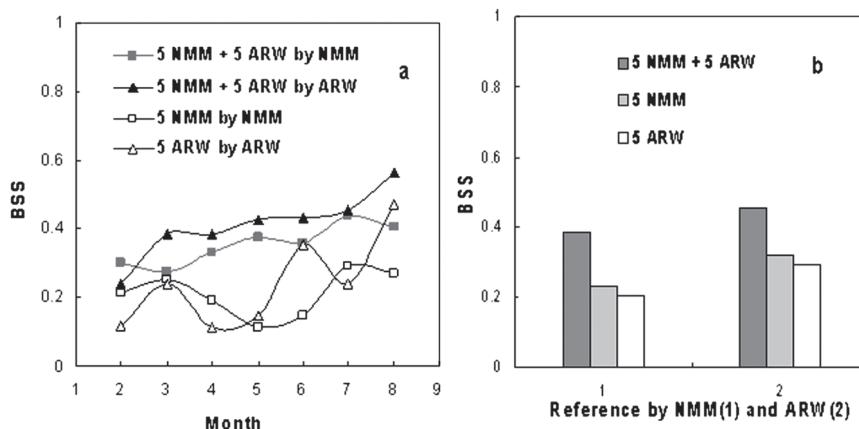


FIG. 6. (a) Monthly BSS scores of probabilistic forecasts based on the 5-member NMM ensemble (using the NMM control as a reference), 5-member ARW ensemble (using the ARW control as a reference), and 10-member multimodel SREF-B08RDP ensemble (using both the NMM and ARW control forecasts as references); (b) BSS averaged over the 7 months for the same three ensembles but with both the NMM and ARW control forecasts used as references.

control forecast might, on average, outperform the single ARW control in predicting fog. This is in agreement with the previous results revealed by scores such as the ETS of Figs. 2 and 5 and Table 2. A possible reason of this is that NMM was found to be slightly more accurate in predicting the RH field than was ARW in this study.

To demonstrate meteorologically why ensemble-based fog forecasts work better than a single forecast, the 6–7 April fog episode is examined again with the 10-member SREF-B08RDP ensemble. Figure 7 shows that the ensemble spreads or forecast variations among ensemble members were quite significant over the fog area: 1.5–2.5 m s<sup>-1</sup> for surface wind speed (Fig. 7a), 10%–20% for 2-m RH (Fig. 7b), 500–5000 m for cloud base (Fig. 7c) and cloud top (Fig. 7d), and 0.2–0.4 g kg<sup>-1</sup> for the LWC (Fig. 7e). From Eq. (1), one can imagine that this large variation in these basic variables among the individual member forecasts used to diagnose fog occurrence would easily translate into an even larger uncertainty in the resulting fog forecasts from one member to another. When taking a closer look at Fig. 7e, one can observe that the spreads for LWC were even larger than the ensemble mean LWC itself at many locations, similar to what is often observed in ensemble precipitation forecasts. Given a large forecast variation in basic fields from one member to another, it is unlikely that a single member could capture the whole picture, but the combined forecast from all members might do a better job. This is exactly the case here. For example, the 9-h control forecasts with multivariable diagnosis failed to predict the whole picture of the fog event (Figs. 4a and 4b). Five out of the seven major fog-reporting

metropolitan cities were missed by each model's control forecast. Only Hangzhou and Dalian (Tianjin and Dalian) were correctly captured by the NMM (ARW) control forecast. Although the ARW control forecast had a tiny indication of fog near Shanghai and Hangzhou, the predicted fog scales were too small to show any confidence (Fig. 4b). Both control forecasts, especially the one from NMM, also failed to predict fog over the vast oceanic area (the Yellow Sea and East Sea between Qingdao and Hangzhou). However, if information from all the individual ensemble members is combined, the situation can be greatly improved. For example, the ensemble-mean forecast of LWC only in Fig. 7e has much better fog coverage than the single control forecasts (Figs. 4a and 4b), compared to the observations (Fig. 3a). The 10-member SREF-B08RDP-based probabilistic forecast in Fig. 7f showed further improvement. For example, the overall predicted areal coverage enclosed within the 10% threshold was in good agreement with the observations, except for a missing northwest-southeast-oriented fog band located north of Nanjing and Shanghai (cf. Figs. 3a and 7f). Note that there was uncertainty in the fog observations in the area northeast of Dalian, as mentioned above. All seven major fog-observed cities were correctly predicted to have fog by at least one or more of the ensemble members, and Hangzhou, Qingdao, Tianjin, and Dalian were predicted by 6–7 members (a 60%–70% probability). Obviously, even if only a few members correctly predicted fog, that would be a valuable piece of information to end users, such as local transportation administrators, to plan ahead and reduce property damage and protect people's lives. This case clearly demonstrates how and why

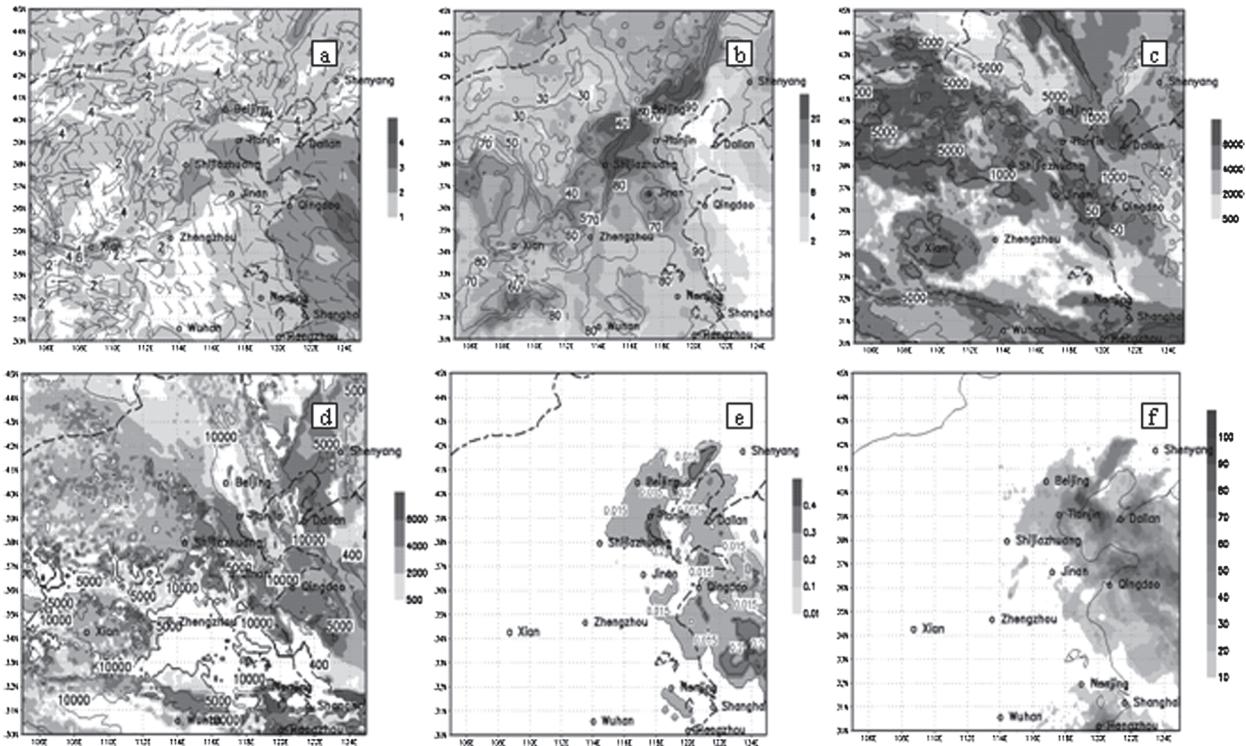


FIG. 7. The 9-h forecasts (verified at 0500 BT 7 Apr 2008) of the ensemble mean (arrows for wind and contours for other variables) and spread (shaded) of (a) 10-m wind ( $\text{m s}^{-1}$ ), (b) 2-m RH (%), (c) cloud base (m), (d) cloud top (m), (e) model lowest-level LWC near the surface ( $\text{g kg}^{-1}$ ), and (f) probabilistic fog forecast based on the multivariable fog diagnostic method, derived from the 10-member multimodel SREF-B08RDP ensemble initiated at 1200 UTC (2000 BT) 6 Apr 2008.

an ensemble-based forecast would be superior to and more beneficial than a single forecast.

### c. Single-model-based ensemble versus multimodel-based ensemble

To examine if model diversity (in both the physics and dynamics) can add extra value to an IC-uncertainty-only based ensemble as suggested by Mullen et al. (1999), the results from a single-model ensemble (either the five-member NMM or five-member ARW ensemble) were compared with those from the multimodel-based SREF-B08RDP ensemble. Two comparison experiments were carried out: first with a combination of the two single-model ensembles, that is, the 10-member SREF-B08RDP system (5 NMM + 5 ARW), and then with the 5-member NMM-ARW system (3 NMM + 2 ARW) to eliminate any ensemble-size effects. Therefore, it is important to keep in mind that the results from the first experiment should reflect a combined impact from both the multimodel approach and increased ensemble size (from 5 to 10 members). When comparing the ETS of the three ensembles shown in Fig. 5f for the first experiment, the improvement in probability-based deterministic forecasts (especially at the 20%, 40%, and

60% thresholds) is quite obvious when an additional model/ensemble is added. For example, for the forecast enclosed by the 40% probability threshold, the averaged ETS was 0.225 for the single-model ensembles and 0.334 for the combined two-model ensemble, which is a 48.4% improvement. It is a big gain with contributions from both the multimodel approach and the membership increase. In light of the fact that forecast skill with a 0.3 ETS is equivalent to the current accuracy level in predicting warm season precipitation, this 40% probability fog forecast is certainly able to provide useful information to users. To exclude the ensemble size effect, a five-member multimodel NMM-ARW ensemble was also constructed by combining three NMM and two ARW members (i.e., the second experiment). The ETS of the 40% probability forecast derived from the second experiment is 0.264 (Fig. 10). Therefore, an improvement of 17.3% in ETS (from 0.225 to 0.264) has been achieved purely by the multimodel approach. On the other hand, comparing the two multimodel NMM-ARW combined ensembles, the improvement in ETS is 26.5%, increasing from 0.264 to 0.334 due to the membership increase from 5 to 10 members. It is inferred that the contribution of the increase in the member size from 5 to 10 members

TABLE 4. Decomposition of BS into reliability, resolution, and uncertainty for the four ensembles. Boldface indicates the lowest values for reliability, uncertainty, and BS, and the highest value for resolution.

BS components	5 NMM + 5 ARW	5 NMM	5 ARW	3 NMM + 2 ARW
Reliability	<b>0.0016</b>	0.0114	0.0099	0.0069
Resolution	<b>0.0338</b>	0.0235	0.0185	0.0217
Uncertainty	<b>0.1078</b>	0.1120	0.1138	0.1134
BS	<b>0.0756</b>	0.0998	0.1052	0.0986

is also significant. More discussion about the impacts of the ensemble size on probabilistic forecasts will be given in the next several paragraphs although no impact purely from the membership increase will be seen in this study since the membership was increased in a “multimodel” environment as mentioned above.

The superiority seen in the deterministic aspect is also true for the probabilistic aspects. For example, the BSS of Fig. 6 from the first experiment shows that the probabilistic forecast based on the 10-member multimodel B08RDP ensemble was much more skillful than those based on the single-model NMM or ARW ensembles. To better understand where the improvement came from, the BS has been decomposed into three components of *reliability*, *resolution*, and *uncertainty* [Eq. (13)]. The result is shown in Table 4. By comparing the multimodel 10-member SREF-B08RDP (second column) ensemble with either the NMM (third column) or ARW (fourth column) ensembles, it can be seen that the main improvement was in the reliability although the resolution was also noticeably improved. The improvement in the uncertainty is, however, very small. After excluding the ensemble size effect, a similar result was also observed from the second experiment; that is, the main improvement between a single-model ensemble (third or fourth columns) and a multimodel ensemble (fifth column) is in the reliability. Since the reliability is mainly attributed to the ensemble technique while the resolution is mainly attributed to the quality of the ICs and base models employed by an ensemble system, this result clearly indicates that a multimodel approach is effective in improving an ensemble technique rather than in improving the “effective quality” of the ICs and model. This is particularly true when the ensemble size is very small, such as in the five-member second experiment, where there was no improvement in resolution but merely a reflection of the original quality of the base models (the resolution of the combined ensemble was somewhere between that of the NMM and ARW ensembles, considering the fact that the NMM performed slightly better than the ARW, as seen previously in this study). Regarding the impacts of ensemble size, it is useful to keep

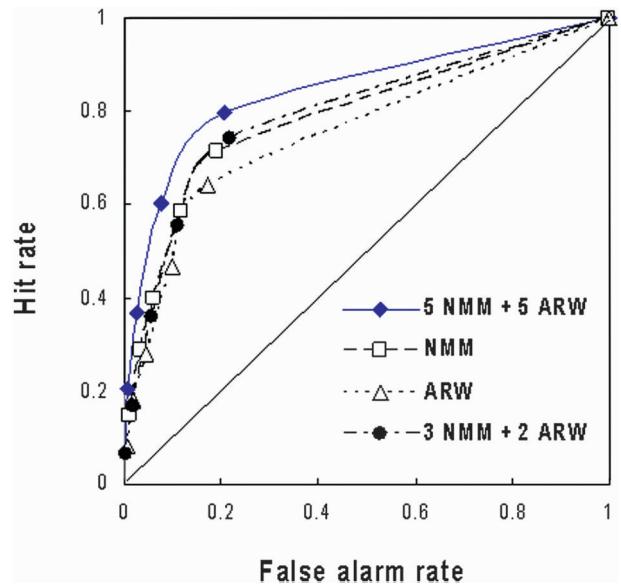


FIG. 8. ROC diagrams of probabilistic forecasts based on the 5-member NMM ensemble, 5-member ARW ensemble, 5-member multimodel NMM–ARW ensemble, and 10-member multimodel SREF-B08RDP ensemble. Note that HR is also known as POD and POFD.

in mind that BS or BSS has a theoretical cap or limit for a given ensemble size (Richardson 2001). The BSS increases rapidly with the increase in ensemble size when the ensemble size is small ( $\leq 10$  members) and becomes nearly saturated when the ensemble size is larger ( $\geq 50$  members), which is particularly so for low-probability events. This same pattern of behavior is also observed using other measuring metrics (Du et al. 1997). This pattern implies that a probabilistic forecast cannot reach its full skill if the ensemble size is too small, especially for low-predictability events like fog. In this sense, increasing the ensemble size from 5 to 10 members should have made a significant contribution (equally as important as the multimodel approach) to the ensemble performance seen in this study, an argument that can be apparently confirmed by comparing the second column with the fifth column in Table 4. However, it is expected that the ensemble size impacts will be much smaller when the ensemble size exceeds 10 members (Du et al. 1997; Richardson 2001). For example, in an experiment combining two 50-member ensembles, one might find that the impacts from the increased ensemble size (from 50 to 100 members) are much less than those from the multimodel effect.

Figure 8 shows the ROC diagram for each of these four ensembles. Obviously, the multimodel ensemble outperformed both single-model ensembles; the 10-member multimodel SREF-B08RDP had the largest ROC areas followed by those of the 5-member

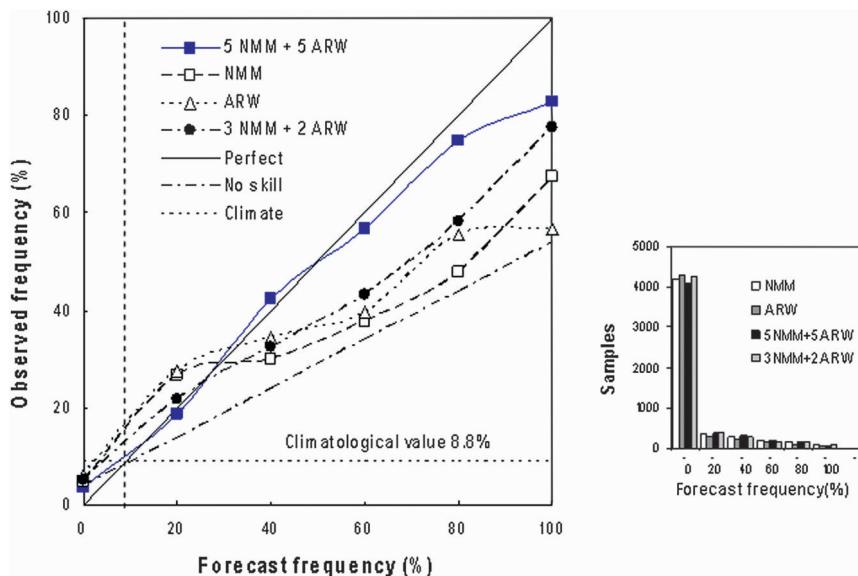


FIG. 9. (left) Reliability diagrams of probabilistic forecasts based on the 5-member NMM ensemble, 5-member ARW ensemble, 5-member multimodel NMM-ARW ensemble, and 10-member multimodel SREF-B08RDP ensemble. (right) Shown is the sharpness diagram for the four ensembles.

multimodel NMM-ARW ensemble, the single-model NMM-based ensemble, and then the single-model ARW-based ensemble. The slightly better performance of the NMM forecasts over that of ARW was seen again as previously in Figs. 2, 5, and 6 and Table 2. Again, given that the improvement in the ROC area of the 5-member NMM-ARW ensemble over the NMM or ARW ensembles is small, the contribution of the ensemble size increase from 5 to 10 members is obviously important to the quality of the ensemble-based probabilistic forecast in this experiment with a small ensemble size, for the same reasons discussed in the last paragraph. The small impacts of a multimodel approach on ROC are due to the nature of the score, which mainly reflects the resolution but not the reliability aspect of a probabilistic forecast. This is consistent with the result revealed in Table 4.

To evaluate the joint distribution of forecasts and observations over various probabilities, the reliability diagrams of the four ensembles are compared in Fig. 9, where the sharpness diagrams, climatology (or resolution limit), and no-skill line are also plotted. The no-skill line is generated in such a way that it is evenly divided between the perfect forecast (diagonal line) and the climatology (Wilks 2006). A probabilistic forecast is considered to be skillful if its reliability curve is above the no-skill line and to have resolution if the curve is above the climatology line. Therefore, Fig. 9 suggests that all four ensembles are skillful as well as having resolution in predicting fog. However, both of the single-model-based

ensembles showed a slight lack of confidence at lower probabilities ( $<20\%$ ) and significant overconfidence at higher probabilities ( $>50\%$ ). This problem has been noticeably corrected by the five-member multimodel NMM-ARW ensemble, which once again shows the positive contribution of the multimodel approach to probabilistic distributions. Further combined with the increase in ensemble size from 5 to 10 members, the 10-member SREF-B08RDP ensemble shows an almost perfect reliability with only slight overconfidence near the high end ( $>80\%$ ). The combined benefits of a multimodel approach and an increase in ensemble size are obvious from this study. Apparently, the results shown by Fig. 9 are very consistent with those of Table 4.

## 6. Summary and future work

A new multivariable-based diagnostic fog-forecasting method has been proposed. Its fog diagnosis is based on the following five basic model variables: model lowest-level liquid water content (LWC), cloud top, cloud base, 10-m wind speed, and 2-m relative humidity. Since all of these base variables are available from a model postprocessor, this fog diagnostic algorithm can also be included as part of a model postprocessor and, therefore, fog forecasts can now be provided conveniently and centrally as direct NWP model guidance to forecasters and end users. The selection of these five variables, their thresholds, and influence on fog forecasting (focusing on 2-m RH and surface wind) were discussed to provide

some insights into similar works in the future. This method can easily be adapted to other NWP models. The practical application of this method is obvious, especially to the transportation community of air, sea, and land as well as navy or marine-related operations. By comparing this new multivariable method to a commonly used method—the LWC-only based approach—it is found that the newly proposed multivariable fog diagnostic method has a much higher detection capability in current operational NWP models. The LWC-only method has a very low detection rate and tends to miss almost 90% of fog events, while the new method can greatly improve the fog detection rate and demonstrates reasonably good forecast accuracy. Reasons why the multivariable approach works better than the LWC-only method were also illustrated in a case study.

To assess fog-forecast skill and account for forecast uncertainty, this fog-forecasting algorithm is then applied to a multimodel-based Mesoscale Ensemble Prediction System. To verify the accuracy of a deterministic forecast, the following six scoring rules were used: hit rate (HR), false alarm ratio (FAR), missing rate (MR), correct rejection rate (CRR), bias score (bias), and equitable threat score (ETS). To verify the performance of a probabilistic forecast, the following four scores were employed: Brier score (BS) and its decomposition, Brier skill score (BSS), relative operating characteristic (ROC), and reliability diagrams (reliability). Verification was focused on the 12- and 36-h forecasts. By comparing the performance between single-value forecasts and ensemble-based forecasts, the benefits of an ensemble approach over a single deterministic approach were clearly shown. The ensemble-based forecasts were, in general, statistically superior to a single-value forecast in fog forecasting. With the aid of an ensemble approach, such as using a forecast close to the ensemble median (50% probability), the current operational NWP models are capable of predicting fog 12–36 h in advance with an accuracy, on average, similar to the level of warm season precipitation forecasts (with an ETS around 0.334). A case was also presented to demonstrate meteorologically why ensemble-based forecasts work better and are socially more beneficial than single-value forecasts.

By further comparing forecasts between those from the single-model ensemble and the two-model ensembles, it was shown that the performance of ensemble-based forecasts could be further improved by using a multimodel approach. The multimodel approach is an effective way in which to enhance the ensemble technique to improve the reliability (but not the resolution and uncertainty aspects) of probabilistic forecasts. For a small-sized ensemble such as the one in this study, the increase in its membership is also important in improv-

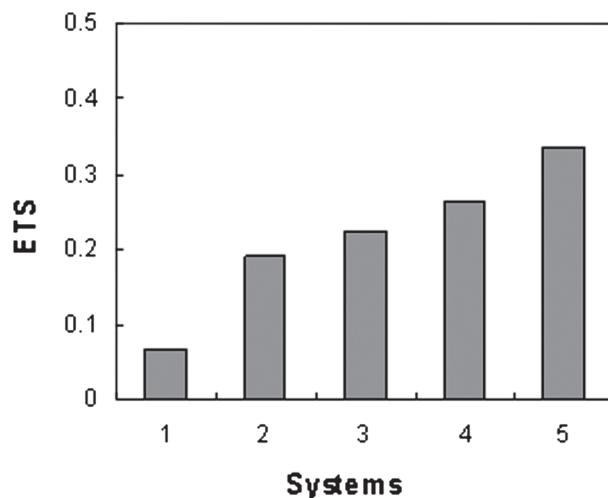


FIG. 10. ETSs (average of the ARW and NMM over the 7-month period at 12- and 36-h forecast lengths) from the various forecast systems: 1) the single control runs based on the LWC-only approach (ETS = 0.063), 2) the single control runs but based on the multivariable fog diagnosis (0.192; a 205% improvement over the previous step), 3) the 40% probability forecasts based on the 5-member single model ensembles (0.225; 17.2%), 4) the 40% probability forecast based on the 5-member multimodel NMM-ARW ensemble (0.264; 17.3%), and 5) the 40% probability forecast based on the 10-member multimodel SREF-B08RDP ensemble (0.334; 26.5%).

ing the quality of the probabilistic forecasts, although this importance is expected to decrease when the ensemble size increases. To summarize and give a quick comparison, Fig. 10 shows all the ETSs from the various approaches used in this study, including the new fog detection method, ensemble technique, multimodel approach, and the increase in ensemble size. We can see that steady improvement was made through each of those steps, with two big jumps, one associated with the use of the new multivariable fog detection method and the other associated with the combining of the two single-model ensembles (a mixed contribution of the multimodel approach and the ensemble size increase). The overall improvement was impressive and dramatic: from basically no skill at all (ETS = 0.063) to a skill level equivalent to that of warm season precipitation forecasts of the current NWP models (0.334).

A problem with this fog diagnostic method is that it can predict only fog occurrence but not fog intensity. In the real world, predicting fog intensity is as important as predicting its occurrence in traffic planning and control of land, air, and sea. This problem might be solved by applying a newer diagnostic method suggested by Zhou and Ferrier (2008), since this newer method can resolve fog liquid water content on the grid scale. Although this newer method was developed for radiation fog, it could

be easily expanded to cover other types of fog by adding advection terms. An extra variable needed for this method is turbulence intensity, which is usually available from a model postprocessor. This method is planned for use in experiments with the next upgrade of the NCEP SREF system (Du et al. 2009). Thus, both fog occurrence and intensity could then be systematically verified over North America within the framework of ensemble prediction. At the same time, it will be also interesting and useful to then compare the ensemble-based fog forecasts to the statistical approaches such as the MOS and neural network-based methods.

*Acknowledgments.* NCEP participation to the WMO/WWRP RDP for the Beijing 2008 Olympic Games was fully supported by the NCEP management team, especially Drs. Geoffrey DiMego, Stephen Lord, and Louis Uccellini. The B08RDP annual workshops were supported by the China Meteorological Administration, especially Drs. Meiyang Jiao, Yihong Duan, Jiandong Gong, Jing Chen, Guo Deng, Hua Tian, Yinglin Li, Xiaoli Li, and Jianjie Wang. The SREF-B08RDP system was routinely run from January to September 2008 by NCEP Central Operations. We also thank Dr. Chune Shi of the Anhui Institute of Meteorological Sciences, Hefei, China, for her help in reviewing the fog observation data. We acknowledge Drs. Garry Toth of the Hydrometeorology and Arctic Laboratory, Environment Canada, and Arnott Justin of the NWS Weather Forecast Office, Binghamton, New York, for their discussions of the fog diagnosis scheme. The fog algorithm development work was partially supported by an FAA project. Our special appreciation goes to Ms. Mary Hart, EMC information officer, for her English editing of the manuscript. Last but not least, the suggestions from Drs. Glenn White and Robert Grumbine of EMC, as well as three anonymous reviewers, gave us opportunities to improve our final version.

#### REFERENCES

- Baker, R., J. Cramer, and J. Peters, 2002: Radiation fog: UPS Airline conceptual models and forecast methods. Preprints, *10th Conf. on Aviation, Range, and Aerospace*, Portland, OR, Amer. Meteor. Soc., 5.11. [Available online at <http://ams.confex.com/ams/pdfpapers/39165.pdf>.]
- Ballard, S. P., B. W. Golding, and R. N. B. Smith, 1991: Mesoscale model experimental forecasts of the haar of northeast Scotland. *Mon. Wea. Rev.*, **119**, 2107–2123.
- Bergot, T., and D. Guedalia, 1994: Numerical forecasting of radiation fog. Part I: Numerical model and sensitivity tests. *Mon. Wea. Rev.*, **122**, 1218–1230.
- , D. Carrer, J. Noilhan, and P. Bougeault, 2005: Improved site-specific numerical prediction of fog and low clouds: A feasibility study. *Wea. Forecasting*, **20**, 627–646.
- Bott, A., and T. Trautmann, 2002: PAFOG—A new efficient forecast model of radiation fog and low-level stratiform clouds. *Atmos. Res.*, **64**, 1–4, 191–203.
- Charles, M. E., and B. A. Colle, 2009: Verification of extratropical cyclones within the NCEP operational models. Part II: The short-range ensemble forecast system. *Wea. Forecasting*, **24**, 1191–1214.
- Clark, A. J., W. A. Gallus, 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.
- Croft, P. J., R. L. Pfof, J. M. Medlin, and G. A. Johnson, 1997: Fog forecasting for the southern region: A conceptual model approach. *Wea. Forecasting*, **12**, 545–556.
- Du, J., 2007: Uncertainty and ensemble forecasting. *Science and Technology Infusion Climate Bulletin*, NOAA/NWS Science and Technology Infusion Lecture Series, 42 pp. [Available online <http://www.nws.noaa.gov/ost/climate/STIP/uncertainty.htm>.]
- , and M. S. Tracton, 2001: Implementation of a real-time short-range ensemble forecasting system at NCEP: An update. Preprints, *Ninth Conf. on Mesoscale Processes*, Fort Lauderdale, FL, Amer. Meteor. Soc., P4.9. [Available online at [http://ams.confex.com/ams/WAF-NWP-MESO/techprogram/paper\\_23074.htm](http://ams.confex.com/ams/WAF-NWP-MESO/techprogram/paper_23074.htm).]
- , S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459.
- , G. DiMego, M. S. Tracton, and B. Zhou, 2003: NCEP Short-Range Ensemble Forecasting (SREF) system: Multi-IC, multi-model and multi-physics approach. *Research Activities in Atmospheric and Oceanic Modelling*, J. Cote, Ed., CAS/JSC Working Group on Numerical Experimentation (WGNE) Rep. 33, WMO/TD-1161, 5.09–5.10. [Available online at <http://www.emc.ncep.noaa.gov/mmb/SREF/reference.html>.]
- , and Coauthors, 2004: The NOAA/NWS/NCEP Short-Range Ensemble Forecast (SREF) system: Evaluation of an initial condition vs. multi-model physics ensemble approach. Preprints, *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 21.3. [Available online at <http://ams.confex.com/ams/pdfpapers/71107.pdf>.]
- , and Coauthors, 2006: New dimension of NCEP Short-Range Ensemble Forecast (SREF) system: Inclusion of WRF members. Rep. to WMO Export Team Meeting on Ensemble Prediction System, Exeter, UK, 5 pp. [Available online at <http://www.emc.ncep.noaa.gov/mmb/SREF/reference.html>.]
- , G. Dimego, Z. Toth, D. Jovic, B. Zhou, J. Zhu, J. Wang, and H. Juang, 2009: Recent upgrade of NCEP Short-Range Ensemble Forecast (SREF) system. Preprints, *19th Conf. on Numerical Weather Prediction/23rd Conf. on Weather Analysis and Forecasting*, Omaha, NE, Amer. Meteor. Soc., 4A.4. [Available online at <http://ams.confex.com/ams/pdfpapers/153264.pdf>.]
- Duan, Y.-H., and Coauthors, 2009: A Report on the WWRP Research and Development Project B08RDP to the WWRP Joint Scientific Committee. WMO/WWRP/CAS/JSC, 63 pp. [Available online at [http://ftp.wmo.int/pages/prog/arep/wwrp/new/documents/Doc3\\_2\\_3\\_B08RDP.doc](http://ftp.wmo.int/pages/prog/arep/wwrp/new/documents/Doc3_2_3_B08RDP.doc).]
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.

- Eckel, F. A., and M. K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting*, **13**, 1132–1147.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Fabbian, D., R. D. Dear, and S. Lelleyett, 2007: Application of artificial neural network forecasts to predict fog at Canberra International Airport. *Wea. Forecasting*, **22**, 372–381.
- Ferrier, B. S., Y. Jin, Y. Lin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model. Preprints, *15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 10.1. [Available online at [http://ams.confex.com/ams/SLS\\_WAF\\_NWP/techprogram/paper\\_47241.htm](http://ams.confex.com/ams/SLS_WAF_NWP/techprogram/paper_47241.htm).]
- Gao, S., H. Lin, B. Shen, and G. Fu, 2007: A heavy sea fog event over Yellow Sea in March 2005: Analysis and numerical modeling. *Adv. Atmos. Sci.*, **24**, 65–81.
- Gayno, G. A., 1994: Development of a higher-order, fog producing boundary layer model suitable for use in numerical weather prediction. M.S. thesis, Dept. of Meteorology, The Pennsylvania State University, 104 pp.
- Gultepe, I., M. D. Muller, and Z. Boybeyi, 2006: A new visibility parameterization for warm-fog application in numerical models. *J. Appl. Meteor. Climatol.*, **45**, 1469–1480.
- , and Coauthors, 2007: Fog research: A review of past achievements and future perspectives. *Pure Appl. Geophys.*, **164**, 1121–1159.
- , and Coauthors, 2009: The fog remote sensing and modeling field project. *Bull. Amer. Meteor. Soc.*, **90**, 341–359.
- Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1322–1327.
- , and —, 1998: Evaluation of Eta–RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.*, **126**, 711–724.
- Hong, S.-Y., and J. Dudhia, 2003: Testing of a new nonlocal boundary layer vertical diffusion scheme in numerical weather prediction applications. Preprints, *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 17.3. [Available online at <http://ams.confex.com/ams/pdfpapers/72744.pdf>.]
- Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX'98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73–91.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **124**, 1225–1242.
- , 1996: The Ssrface layer in the NCEP Eta Model. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 354–355.
- , J. P. Gerrity Jr., and S. Nickovic, 2001: An alternative approach to nonhydrostatic modeling. *Mon. Wea. Rev.*, **129**, 1164–1178.
- 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.
- Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.
- Kong, F., 2002: An experimental simulation of a coastal fog-stratus case using COAMPS model. *Atmos. Res.*, **64**, 205–215.
- Koracin, D., J. A. Businger, C. E. Dorman, and J. M. Lewis, 2005: Formation, evolution, and dissipation of coastal sea fog. *Bound.-Layer Meteor.*, **117**, 447–478.
- Koziara, M. C., J. R. Robert, and W. J. Thompson, 1983: Estimating marine fog probability using a model output statistics scheme. *Mon. Wea. Rev.*, **111**, 2333–2340.
- Kunkel, B. A., 1984: Parameterization of droplet terminal velocity and extinction coefficient in fog models. *J. Climate Appl. Meteor.*, **23**, 34–41.
- Lacis, A. A., and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Liu, X., H. Zhang, O. Li, and Y. Zhu, 2005: Preliminary research on the climatic characteristics and change of fog in China (in Chinese). *J. Appl. Meteor. Sci.*, **16**, 220–271.
- Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333.
- Marzban, C., S. Leyton, and B. Colman, 2007: Ceiling and visibility forecasts via neural networks. *Wea. Forecasting*, **22**, 466–479.
- Mason, I. B., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.*, **102** (D14), 16 663–16 682.
- Mullen, S. L., J. Du, and F. Sanders, 1999: The dependence of ensemble dispersion on analysis-forecast systems: Implications to short-range ensemble forecasting of precipitation. *Mon. Wea. Rev.*, **127**, 1674–1686.
- Muller, M. D., 2005: Numerical simulation of fog and radiation in complex terrain. Ph.D. thesis, University of Basel, Basel, Switzerland, 103 pp.
- Murthy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.
- Pagowski, M., I. Gultepe, and P. King, 2004: Analysis and modeling of an extremely dense fog event in southern Ontario. *J. Appl. Meteor.*, **43**, 3–16.
- Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473–2489.
- Roquelaure, S., and T. Bergot, 2008: A Local Ensemble Prediction System (L-EPS) for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteor. Climatol.*, **47**, 3072–3088.
- Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.
- Schwarzkopf, M. D., and S. B. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res.*, **96**, 9075–9096.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF, version 2. NCAR Tech Note NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307.]
- 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.

- , and —, 2007: Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system. *Wea. Forecasting*, **22**, 3–17.
- , H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.*, **127**, 433–446.
- Tardif, R., and R. M. Rasmussen, 2008: Process-oriented analysis of environmental conditions associated with precipitation fog events in the New York City region. *J. Appl. Meteor. Climatol.*, **47**, 1681–1703.
- Teixeira, J., 1999: Simulation of fog and mist. *Quart. J. Roy. Meteor. Soc.*, **125**, 529–553.
- Toth, G., and W. Burrows, 2008: Automated fog forecasts from an operational NWP model. *Fog Remote Sensing and Modeling (FRAM) Workshop*, Halifax, NS, Environment Canada. [Available online at <http://www.chebucto.ns.ca/Science/AIMET/fog/program.html>.]
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at the NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- , O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. I. T. Jolliffe and D. B. Stephenson, Eds., *Forecast Verification—A Practitioner's Guide to Forecasting*, John Wiley, 137–164.
- Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 378–398.
- , J. Du, Z. Toth, and H. Juang, 1998: Short-range ensemble forecasting (SREF) at NCEP/EMC. Preprints, *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 269–272.
- Wilks, D. S., 2006: *Statistical Methods in Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.
- WMO, 1966: *International Meteorological Vocabulary*. World Meteorological Organization, 276 pp.
- Yuan, H., S. L. Mullen, X. Gao, S. Sorooshian, J. Du, and H. H. Juang, 2005: Verification of probabilistic quantitative precipitation forecasts over the southwest United States during winter 2002–2003 by the RSM ensemble system. *Mon. Wea. Rev.*, **133**, 279–294.
- , —, —, —, —, and —, 2007: Short-range probabilistic quantitative precipitation forecasts over the southwest United States by the RSM ensemble system. *Mon. Wea. Rev.*, **135**, 1685–1698.
- Zhou, B., and B. S. Ferrier, 2008: Asymptotic analysis of equilibrium in radiation fog. *J. Appl. Meteor. and Climatol.*, **47**, 1704–1722.
- , and Coauthors, 2004: An introduction to NCEP SREF Aviation Project. Preprints, *11th Conf. on Aviation, Range, and Aerospace*, Hyannis, MA, Amer. Meteor. Soc., 9.15. [Available online at <http://ams.confex.com/ams/pdfpapers/81314.pdf>.]
- , J. Du, B. S. Ferrier, J. McQueen, and G. DiMego, 2007: Numerical forecast of fog—Central solutions. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 8A.6. [Available online at <http://ams.confex.com/ams/pdfpapers/123669.pdf>.]