9A.4 The Quantitative Precipitation Forecasting Component of the 2010 NOAA Hazardous Weather Testbed Spring Experiment

Faye E. Barthold[1,2], Michael J. Bodner[1], David R. Novak[1], Richard Bann[1], Robert Oravec[1], Bruce Sullivan[1], Andrew R. Dean[3], Israel L. Jirak[3], Christopher J. Melick[3], Ryan A. Sobash[4], Adam J. Clark[5], Fanyou Kong[4,6], Steven J. Weiss[3], and Ming Xue[4,6]

[1]NOAA/NWS/Hydrometeorological Prediction Center
[2]I.M. Systems Group, Inc.
[3]NOAA/NWS/Storm Prediction Center
[4]University of Oklahoma
[5]NOAA/National Severe Storms Laboratory
[6]Center for Analysis and Prediction of Storms

1. INTRODUCTION

Despite numerous advances in numerical weather prediction, quantitative precipitation forecasts (QPF) remain a challenge, particularly in the warm season (e.g., Fritsch and Carbone 2004). The current operational model guidance available to forecasters typically features grid spacing of about 12-32 km, which is insufficient to resolve mesoscale boundaries and other small scale features that are often important to the development of heavy precipitation. The coarse grid spacing in most operational models also requires the use of convective parameterization schemes, which can lead to problems with erroneous convective feedback and movement. In addition, limited observational data on convective scales often results in model initialization errors that can amplify with time, making even short term forecasts difficult.

Over the past several years, numerous high resolution (1-4 km) convection allowing modeling systems have been developed in an attempt to address some of these problems. These models have been found to improve forecasts of convective system mode (e.g., Fowle and Roebber 2003; Kain et al. 2006; Weisman et al. 2008), diurnal cycle (e.g., Weisman et al. 2008; Schwartz et al. 2009), and system propagation (e.g., Clark et al. 2009) as well as provide more realistic rainfall amplitudes (Clark et al. 2010) than models with coarser grid resolutions. On the other hand, high resolution models tend to have a high precipitation bias (e.g., Schwartz et al. 2009), and model initialization errors can sometimes be significant (Weisman et al. 2008).

Since 2000, the National Oceanic and Atmospheric Administration's (NOAA) Storm Prediction Center (SPC) and National Severe Storms Laboratory (NSSL) have organized annual Spring Experiments to bring the research and operational forecasting communities together to explore new forecasting techniques and evaluate emerging model guidance (see Weiss et al. 2010 for more information). The 2010 Spring Experiment was conducted at the National Weather Center in Norman, Oklahoma over a five week period from 17 May 2010 to 18 June 2010. In addition to its traditional focus on severe weather, the 2010 experiment expanded to feature both an aviation impacts component, led by the National Centers for Environmental Prediction's (NCEP) Aviation Weather Center (AWC), and a QPF component, led by NCEP's Hydrometeorological Prediction Center (HPC). The QPF component explored the use of high resolution convection allowing guidance to improve warm season QPF forecasts by investigating whether the high resolution guidance provides added value over the current operational models.

2. QPF COMPONENT DESCRIPTION

2.1 Data

The datasets used in the QPF component of the HWT Spring Experiment are summarized in Table 1. The deterministic high resolution guidance featured numerous Weather Research and Forecasting (WRF) models with grid spacing of 1-5 km and both Advanced Research WRF (ARW) and Nonhydrostatic Mesoscale Model (NMM) dynamics cores. These models were contributed by NSSL, the NCEP Environmental Modeling Center (EMC), the National Center for Atmospheric Research (NCAR), and the University of Oklahoma's Center for Analysis and Prediction of Storms (CAPS). In addition to the

* *Corresponding author address:* Faye E. Barthold, NOAA/NWS/Hydrometeorological Prediction Center, 5200 Auth Road, Camp Springs, MD 20746. Email: faye.barthold@noaa.gov

Table 1. Experimental convection-allowing models used during the QPF component of the 2010 HWT Spring Experiment. Each model was initialized at 0000 UTC and had forecasts to 30 hrs (SSEF, CAPS-ARW1), 36 hrs (WRF-NSSL4, WRF-NMM4), or 48 hrs (HRWE-ARW4, WRF-NCAR3). The WRF-NCAR3, WRF-NMM4, and HRWE-ARW4 also had 1200 UTC initializations that were not used in the QPF component.

Provider	Model	Delta X	Notes	Label
CAPS	WRF/ARPS 26 member ensemble	4 km	Multi-model, multi-physics, multi-initial condition ensemble system with radar assimilation	SSEF
CAPS	WRF-ARW	1 km	ARPS 3DVAR initial conditions NAM lateral boundary conditions	CAPS-ARW1
NCAR	WRF-ARW	3 km	RUC initial conditions GFS lateral boundary conditions	WRF-NCAR3
NSSL	WRF-ARW	4 km	NAM initial conditions NAM lateral boundary conditions	WRF-NSSL4
NCEP/EMC	WRF-NMM	4 km	NAM initial conditions NAM lateral boundary conditions	WRF-NMM4
NCEP/EMC	WRF-ARW	5.1 km	NAM initial conditions NAM lateral boundary conditions	HRWE-ARW4

deterministic guidance, the experiment also featured a 4 km 26 member multi-model, multi-physics, and multi-initial condition Storm Scale Ensemble Forecast system (SSEF) provided by CAPS that assimilated radar and other observational data. These models were initialized at 0000 UTC and provided forecasts out to 30-48 hours. Finally, data from both the operational 12 km North American Mesoscale model (NAM) and 32 km Short Range Ensemble Forecast system (SREF) were also considered for comparison to the high resolution deterministic and ensemble models, respectively.

A variety of experimental ensemble products were calculated based on the 15 ensemble members with mixed initial conditions/physics perturbations in the CAPS SSEF. In addition to the ensemble mean, the probability matched mean (Ebert 2001) was also calculated. The probability matched mean combines the spatial pattern of the ensemble mean QPF with the frequency distribution of the rainfall rates from the individual members (Ebert 2001). The goal of this technique is to provide a more realistic ensemble rainfall intensity forecast by eliminating the tendency for the ensemble mean to have areal precipitation coverage that is too broad and maximum precipitation amounts that are too low. The ensemble maximum indicates the maximum QPF from any ensemble member. Point exceedance probabilities were calculated to indicate the probability of exceeding a given QPF threshold at a specific grid point. Similarly, neighborhood exceedance probabilities indicate the smoothed probability of exceeding a given QPF threshold within 80 km of each grid point (e.g., Schwartz et al. 2009).

2.2 Daily Activities

Each morning, QPF component participants used 0000 UTC model guidance to issue experimental probabilistic QPF forecasts for two 6 hr time periods, 1800-0000 UTC and 0000-0600 UTC, over a selected forecast domain. These forecasts outlined areas that had a slight (25%), moderate (50%), or high (75%) probability of exceeding 0.50 in and 1.0 in of precipitation during each 6 hr period. When a probability of exceeding 1.0 in was indicated, participants were also asked to predict the most likely areal average maximum value within that area to give an indication of the magnitude of the event they were expecting. After producing the graphical forecast, a forecast discussion was written focusing on model guidance uncertainties and forecast rationale.

In addition to the experimental forecasts, participants also evaluated their forecasts from the previous day and completed a subjective evaluation of the model guidance used in that forecast. This evaluation consisted of a number of different survey questions designed to evaluate whether the high resolution models had provided better forecast guidance than the operational NAM (deterministic guidance) and SREF (ensemble guidance). In the afternoon, participants joined either the aviation impacts or severe weather components before coming back together for a briefing on the day's forecast and evaluation activities.

3. RESULTS

3.1 Experimental Model Performance

Over the course of the five week experiment, there were many cases in which the high resolution models provided a noticeably improved forecast compared to their operational counterparts. Figure 1 shows a comparison of the 6 hr ensemble mean QPF forecasts between the operational SREF and the SSEF for forecasts valid at 0000 UTC 18 May 2010 for precipitation along the east coast of the United States. While the SREF indicated maximum QPF of 0.82 in over southern Maryland and eastern Virginia (Fig. 1b), the SSEF correctly shifted the focus of the precipitation an entire state south into eastern North Carolina, with maximum amounts of close to 2.0 in (Fig. 1c). The SSEF forecast also gave a better indication of the higher precipitation amounts observed along the South Carolina coast.

Figure 2 shows a comparison of the 6 hr QPF forecasts between the operational 12 km NAM and a number of the high resolution deterministic models for forecasts valid at 0000 UTC 21 May 2010 for precipitation across Texas, Louisiana, and Mississippi. In this case, two regions of heavy precipitation were observed, one oriented approximately west-east across northeastern Texas and northern Louisiana and another oriented approximately north-south in eastern Mississippi. While the NAM indicated a broad area of heavier precipitation across central Louisiana and Mississippi with maximum amounts reaching 0.59 in (Fig. 2b), all of the high resolution models correctly indicated the areas of enhanced precipitation, with the WRF-NSSL4 providing the best forecast (Fig. 2c). Although the WRF-NMM4 has a significant high bias, particularly in eastern Texas (Fig. 2d), and the HRWE-ARW misses the heavy precipitation in northern Louisiana (Fig. 2e), both provide a better indication of the potential for

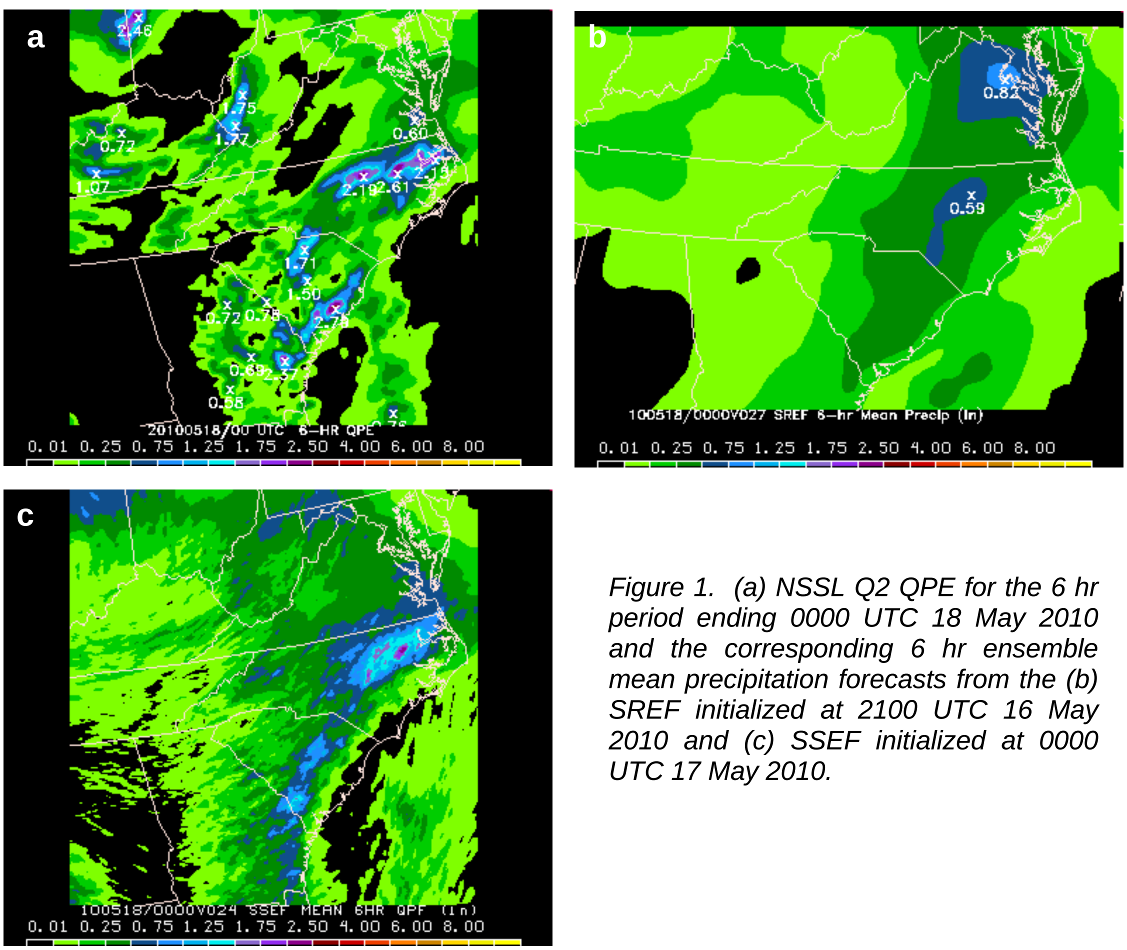


Figure 1. (a) NSSL Q2 QPE for the 6 hr period ending 0000 UTC 18 May 2010 and the corresponding 6 hr ensemble mean precipitation forecasts from the (b) SREF initialized at 2100 UTC 16 May 2010 and (c) SSEF initialized at 0000 UTC 17 May 2010.

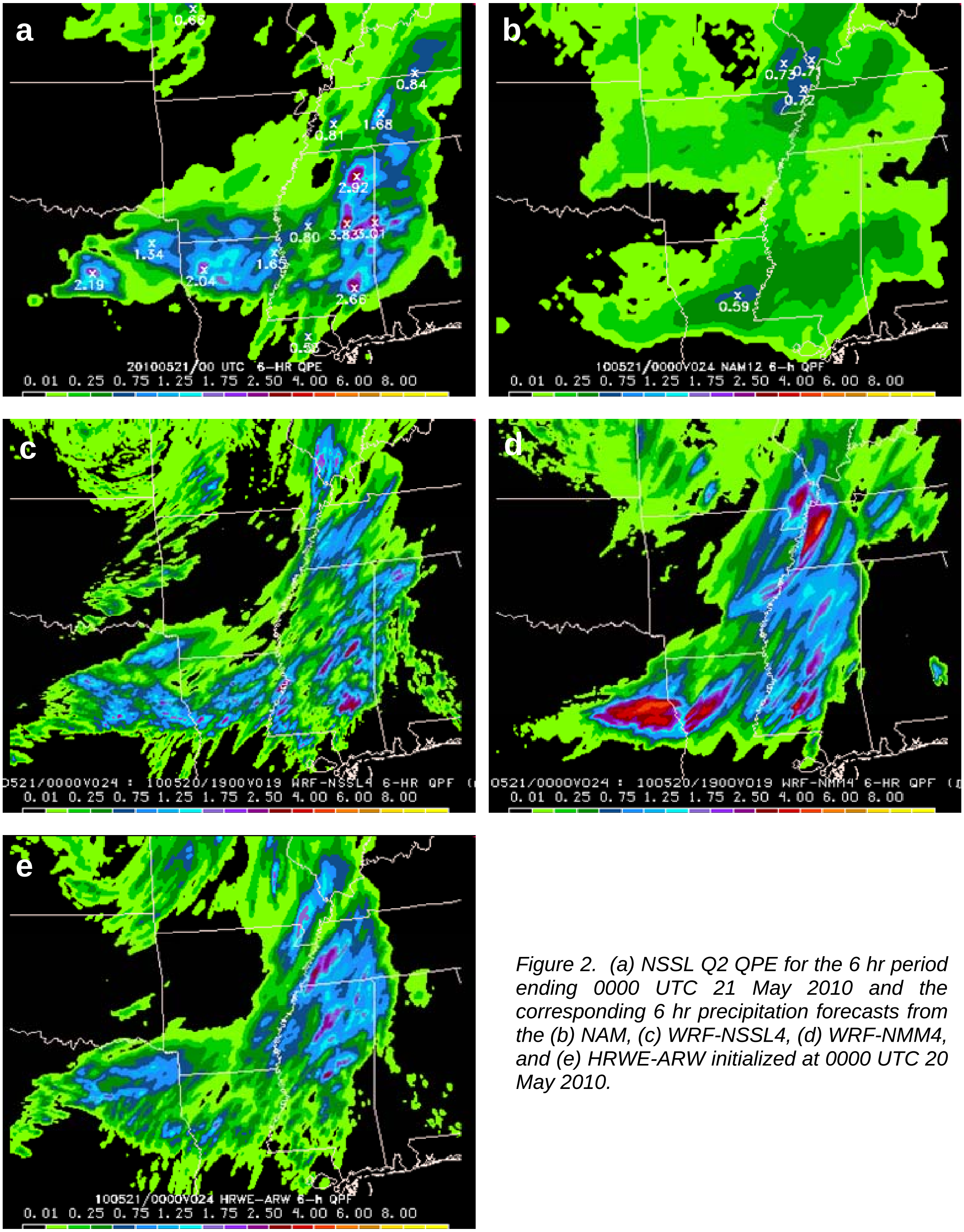
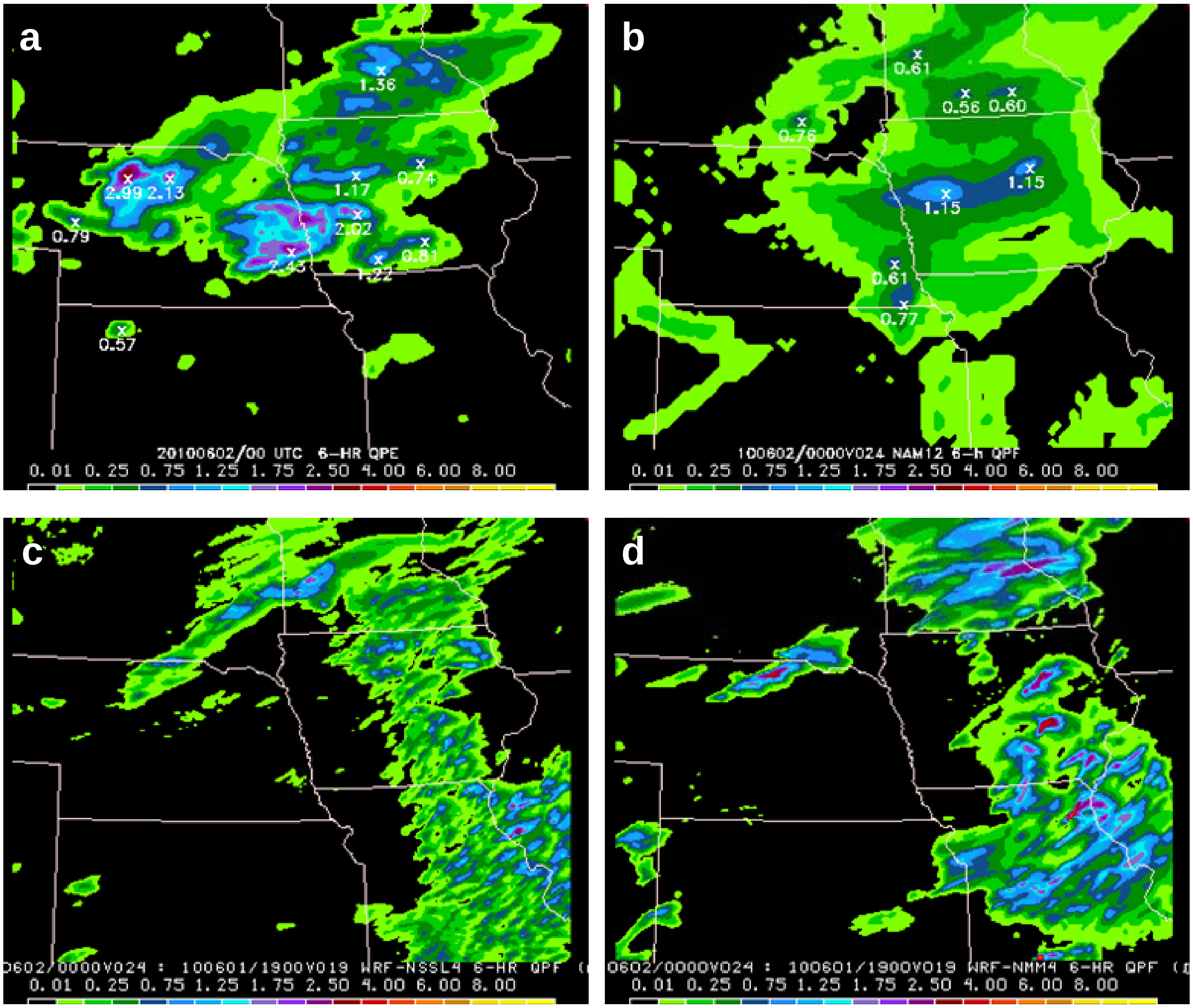


Figure 2. (a) NSSL Q2 QPE for the 6 hr period ending 0000 UTC 21 May 2010 and the corresponding 6 hr precipitation forecasts from the (b) NAM, (c) WRF-NSSL4, (d) WRF-NMM4, and (e) HRWE-ARW initialized at 0000 UTC 20 May 2010.



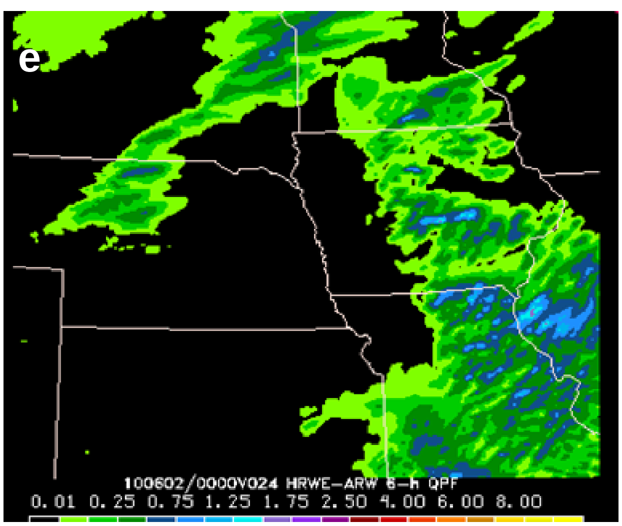


Figure 3. (a) NSSL Q2 QPE for the 6 hr period ending 0000 UTC 2 June 2010 and the corresponding 6 hr precipitation forecasts from the (b) NAM, (c) WRF-NSSL4, (d) WRF-NMM4, and (e) HRWE-ARW initialized at 0000 UTC 1 June 2010.

heavy precipitation in this region than the forecast from the operational NAM.

While there were numerous cases in which both the ensemble and deterministic high resolution models provided improved guidance compared to their operational counterparts, there were also instances where these high resolution models degraded the operational forecast. For example, Fig. 3 shows the 6 hr QPF forecasts valid at 0000 UTC 2 June 2010 from the deterministic high resolution models for heavy precipitation associated with convection observed over Nebraska and Iowa. Although the forecast from the operational NAM doesn't capture the heavy precipitation observed across eastern Nebraska, it does indicate the potential for heavier precipitation across central Iowa, and largely confines the precipitation to Iowa (Fig. 3b). In this case, the deterministic high resolution models focus the bulk of the precipitation too far south and east, giving little to no indication of the potential for heavy precipitation across eastern Nebraska and instead indicating the potential for precipitation across much of Missouri and western Illinois (Figs. 3c-e). In many ways the high-resolution runs look more similar to each other than to the observations in this case.

Model performance was evaluated subjectively through surveys completed each day by the participants. Figure 4 summarizes the results of the subjective evaluation. These results show that the SSEF consistently offered improved forecast guidance compared to that provided by the SREF. In fact, only three out of 40 forecasts from the SSEF were considered worse than the SREF, and the SSEF improved guidance nearly nine times more often than it degraded guidance. Objective verification results of the Relative

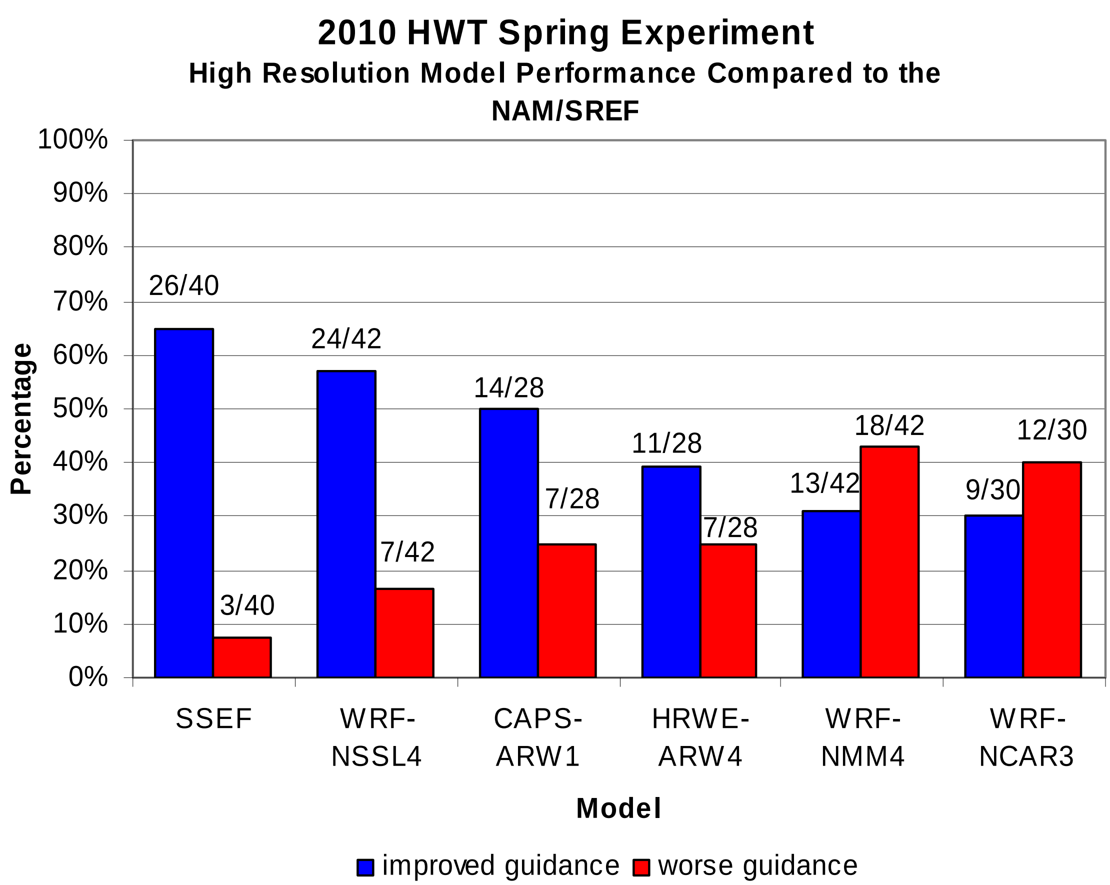


Figure 4. Comparisons of convection-allowing and operational NAM/SREF model performance based on participant feedback from subjective evaluation surveys conducted during the QPF component of the 2010 HWT Spring Experiment. The SSEF is compared to the operational SREF while the other models are compared to the operational 12 km NAM. The CAPS ARW was run at 1 km grid spacing.

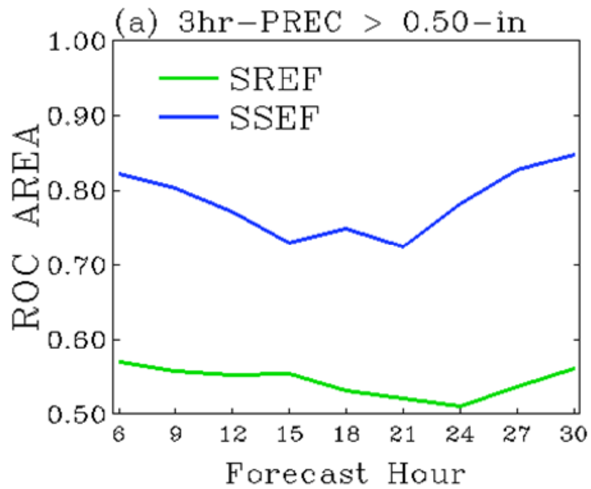


Figure 5. Comparison of Relative Operating Characteristic (ROC) area for 3 hr precipitation greater than 0.5 in for the SREF and the SSEF during the 2010 HWT Spring Experiment. Stage IV multi-sensor precipitation estimates were used for the observations. Courtesy Adam Clark (NSSL).

Operating Characteristic show the dramatic improvement of the SSEF over the SREF (Fig. 5).

Of the deterministic models, the WRF-NSSL4 was consistently rated better than the operational NAM, providing improved forecasts over three times more often than degraded forecasts. The 1 km CAPS-ARW1 also tended to provide improved guidance, exhibiting improved forecasts two times more often than degraded forecasts. However, the fact that the 4 km WRF-NSSL4 was superior to the 1 km CAPS-ARW1 provides further evidence that 4 km grid spacing is sufficient for QPF from a practical standpoint (e.g. Kain et al. 2008). The results for the remainder of the deterministic models were more evenly split, with both the WRF-NMM4 and the WRF-NCAR3 tending to degrade the operational model forecast more often than they improved upon it. The Developmental Testbed Center (DTC) is completing objective verification for select models (Jensen et al. 2011).

3.2 Experimental Ensemble Products

In addition to the evaluation of overall model performance, a subjective evaluation of the experimental ensemble products was also conducted. The ensemble mean was found to be useful, generally providing a realistic depiction of both precipitation amounts and coverage. The probability matched mean was also found to be useful forecast guidance, although there is question whether the technique used is valid on a national scale where, for example, precipitation over Florida can be used to correct forecasts over Iowa. The coverage of the neighborhood exceedance probabilities was considered too broad and the probabilities too high for this guidance to be trusted. This subjective impression is consistent with Hardy et al. (2011) who show that neighborhood guidance is over-biased. Calibration will likely improve this guidance. Finally, while some participants thought that looking at the ensemble maximum precipitation was useful for determining a possible worst case scenario, overall it was not found to be useful given its unrealistically high values. This result highlights the high bias that is often associated with high resolution models.

4. CONCLUSIONS

The QPF component of the 2010 HWT Spring Experiment demonstrated that high resolution convection allowing models can provide skillful QPF guidance. In particular, the SSEF was considered a transformational improvement in warm season QPF forecasting. Participating in the QPF component has been a valuable learning experience for forecasters at the HPC. The experimental forecast and model evaluation process has fostered discussion about the strengths and weaknesses of high resolution model data and the best way to incorporate such data into operational QPFs. It has also placed a renewed emphasis on the verification of HPC's Excessive Rainfall Outlooks. In addition, based on the high ratings received by the WRF-NSSL4 throughout the experiment, this model is now available for daily use at the HPC.

Based on the success of the 2010 experiment, HPC plans to expand its participation in the 2011 Spring Experiment, including adding an afternoon forecasting component that would allow the morning forecasts to be updated with 1200 UTC model guidance and adding the 0600-1200 UTC forecast period (30-36 hr forecast) in order to capture the nocturnal precipitation maximum. Focus will be placed on expanding the available post-processed ensemble guidance including bias-correction. Comparison of the raw and bias-corrected model precipitation forecasts is expected to be an evaluation activity. Finally, experiment participation is expected to grow, allowing more people involved in precipitation forecasting to help shape the development of the next generation of model guidance

Acknowledgements: The authors would like to thank the Hazardous Weather Testbed in Norman, Oklahoma for hosting this experiment and fostering a successful collaborative environment. Thanks are also extended to the 2010 HWT Spring Experiment participants for their helpful feedback and discussions on model performance and forecast uncertainties throughout the course of the experiment. Any opinions, findings, conclusions, or recommendations expressed are those of the authors and do not necessarily represent the views of NOAA/NWS, I.M. Systems Group, Inc., or the University of Oklahoma.

5. REFERENCES

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.

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

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480.

Fowle, M. A., P. J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794.

Fritsch, J. M., R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–965.

Hardy, J. D., J.S. Kain, D. R. Novak, J. J. Gourley, and M. E. Pyle, 2011: Evaluating probabilistic precipitation forecasts generated by deterministic convection-allowing NWP models. Preprints, *24th Conference. on Weather and Forecasting/ 20th Conference on NWP*, Amer. Meteor. Soc., Seattle, WA, 122.

Jensen, T.L., and Coauthors, 2011: The Developmental Testbed Center objective evaluation performed during the 2010 NOAA Hazardous Weather Testbed Spring Experiment. Preprints, *24th Conference on Weather and Forecasting/20th Conference on NWP*, Amer. Meteor. Soc., Seattle, WA, paper 9A.6.

Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181.

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.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.

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

Weiss, S. J., and Coauthors, 2010: An overview of the 2010 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. Preprints, *25th Conference on Severe Local Storms*, Amer. Meteor. Soc., Denver, CO, paper 7B.1.