NOAA HAZARDOUS WEATHER TESTBED

EXPERIMENTAL FORECAST PROGRAM SPRING EXPERIMENT 2009

http://hwt.nssl.noaa.gov/Spring_2009/

HWT Facility – National Weather Center
4 May - 5 June 2009

Program Overview and Operations Plan

5/18/09 Version

Steven Weiss[1], Jack Kain[2], Michael Coniglio[2], David Bright[1], Jason Levit[1], Gregory Carbin[1], Ryan Sobash[3], John Hart[1], and Russell Schneider[1]

(1) NOAA/NWS/Storm Prediction Center, Norman, Oklahoma
(2) NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma
(3) School of Meteorology, University of Oklahoma, Norman, Oklahoma

I. The NOAA Hazardous Weather Testbed

NOAA's Hazardous Weather Testbed (HWT) is a facility jointly managed by the National Severe Storms Laboratory (NSSL), the Storm Prediction Center (SPC), and the NWS Oklahoma City/Norman Weather Forecast Office (OUN) within the National Weather Center building on the University of Oklahoma South Research Campus. The HWT is designed to accelerate the transition of promising new meteorological insights and technologies into advances in forecasting and warning for hazardous mesoscale weather events throughout the United States. The HWT facilities include a combined forecast and research area situated between the operations rooms of the SPC and OUN, and a nearby development laboratory. The facilities support enhanced collaboration between research scientists and operational weather forecasters on specific topics that are of mutual interest.

The HWT organizational structure is composed of two primary overlapping program areas (Fig. 1). The first program area focuses on application of cutting edge numerical weather prediction models to improve severe weather forecasts under the auspices of the Experimental Forecast Program (EFP), and the second program tests research concepts and technology specifically aimed at short-fused warnings of severe convective weather under auspices of the Experimental Warning Program (EWP). A key NWS strategic goal is to extend warning lead times under the concept of "Warn-on-Forecast" through the development and application of convection-allowing numerical models to extend short-term predictability of hazardous convective weather. This provides a natural overlap between the EFP and EWP activities. As the distinction between warnings and short-term forecasts of convective weather gradually diminishes, the degree of overlap is expected to increase. Both programs reside beneath the overarching HWT organization with a focus on national hazardous weather needs.

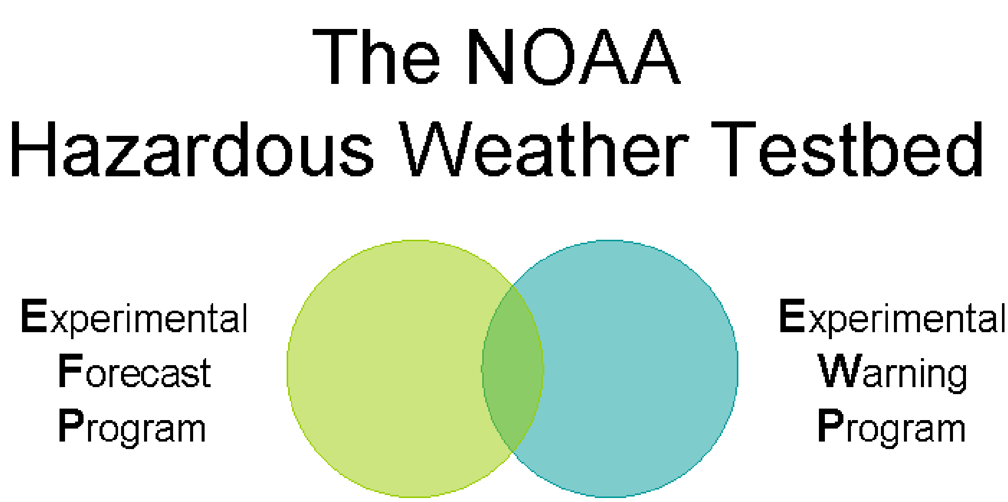


Figure 1: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP) and the Experimental Warning Program (EWP).

The specific mission of each HWT program branch is:

The Experimental Forecast Program - EFP

The EFP branch of the HWT is focused on predicting hazardous mesoscale weather events on time scales ranging from a few hours to a week in advance, and on spatial domains ranging from several counties to the CONUS. The EFP embodies the collaborative experiments and activities previously undertaken by the annual SPC/NSSL Spring Experiments. See http://hwt.nssl.noaa.gov/ for more information about the EFP.

The Experimental Warning Program – EWP

The EWP branch of the HWT is concerned with detecting and predicting mesoscale and smaller weather hazards on time scales of minutes to a few hours, and on spatial domains from several counties to fractions of counties. The EWP embodies the collaborative warning-scale experiments and technology activities previously undertaken by the OUN and NSSL. See http://ewp.nssl.noaa.gov/ for more information about the EWP.

Rapid science and technology infusion for the advancement of operational forecasting requires direct, focused interactions between research scientists, numerical model developers, information technology specialists, and operational forecasters. The HWT provides a unique setting to facilitate such interactions and allows participants to better understand the scientific, technical, and operational challenges associated with the prediction and detection of hazardous weather events. The HWT allows participating organizations to:

- Refine and optimize emerging operational forecast and warning tools for rapid integration into operations
- Educate forecasters on the scientifically correct use of newly emerging tools and to familiarize them with the latest research related to forecasting and warning operations
- Educate research scientists on the operational needs and constraints that must be met by any new tools (e.g., robustness, timeliness, accuracy, and universality)
- Motivate other collaborative and individual research projects that are directly relevant to forecast and warning improvement

For more information about the HWT, see http://www.nssl.noaa.gov/hwt/

II. Historical Perspective

Co-location of the Storm Prediction Center (SPC) with the National Severe Storms Laboratory (NSSL), the Oklahoma City/Norman Weather Forecast Office, and many University of Oklahoma meteorological organizations in the National Weather Center in

Norman provides a unique opportunity to enhance long-standing community interactions and collaboration on a variety of operationally relevant research and experimental forecast programs. Since the re-location of the SPC to the previous NSSL facility Norman in early 1997, a wide cross section of local and visiting forecasters, research scientists, and model developers has participated in a number of experimental programs since the late 1990s. These include forecasting support for field programs such as the International H2O Project (IHOP), establishing the SPC winter weather mesoscale discussion product, evaluating operational and experimental NWP models for application in convective forecasting, including Short Range Ensemble Forecast (SREF) systems and convection-allowing Weather Research and Forecasting (WRF) models, and integrating new observational data, objective analyses, and display tools into forecast operations. A key goal of these programs is to improve forecasts of hazardous meteorological phenomena by: 1) accelerating the transfer of new technology and research ideas into forecast operations at the SPC and other NWS offices, and 2) sharing new techniques, skills, and applied research results more freely with others in the operational forecasting community. Typical issues addressed in these activities include, but are not limited to: optimizing use of vast and ever increasing quantities of observational and model data in operational forecasting, testing and evaluation of new NWP models, better understanding of operational forecast problems, development and evaluation of diagnostic conceptual models, and new product development and display strategies utilizing operational workstations.

Each spring during the climatologically most intense severe weather period, annual multi-agency collaborative forecasting experiments known as the HWT EFP Spring Experiment (formerly called the SPC/NSSL Spring Program) have occurred since 2000. The only exception was in 2006 when the physical move to the new National Weather Center building precluded a large collaborative experiment. During that spring SPC conducted a focused internal pre-implementation evaluation of the NCEP NAM-WRF model.

Summaries about earlier Spring Experiments are available at:

http://www.nssl.noaa.gov/projects/hwt/sp2000.html
http://www.nssl.noaa.gov/projects/hwt/sp2001.html
http://www.nssl.noaa.gov/projects/hwt/sp2002.html
http://www.nssl.noaa.gov/projects/hwt/sp2003.html
http://www.nssl.noaa.gov/projects/hwt/sp2004.html
http://www.nssl.noaa.gov/projects/hwt/sp2005.html
http://www.nssl.noaa.gov/projects/hwt/sp2007.html
http://www.nssl.noaa.gov/projects/hwt/sp2008.html

The following sections provide additional background information about the motivation for the Spring Experiments, the SPC national severe weather forecasting mission and associated scientific and service challenges, an overview of the scientific goals of the 2009 Spring Experiment and its relevance to operational forecasting, the schedule of daily forecasting and evaluation activities, and a list of weekly participants for the 2009 Spring Experiment.

III. Spring Experiment Background and Motivation

Operational Forecasting of Severe Convective Storms – Current State and Challenges

The prediction of convective weather is important from both meteorological and public service/societal impact perspectives. A primary mission of the National Weather Service is the protection of life and property from hazardous weather phenomena, and applied research aimed at improving the prediction of high impact weather such as severe thunderstorms and tornadoes is a critical activity at the NSSL, SPC, OUN, and other NWS offices.

The SPC is responsible for the prediction of severe convective weather over the contiguous United States on time scales ranging from several hours to eight days. To meet these responsibilities, the SPC issues Convective Outlooks for the Day 1, Day 2, Day 3, and Day 4-8 periods to highlight regions with enhanced potential for severe local storms (defined as thunderstorms producing hail $\geq$ 3/4 inch in diameter, wind gusts $\geq$ 50 kt or thunderstorm induced wind damage, and/or tornadoes). These Outlooks are issued in both categorical (slight, moderate, or high risk) and probabilistic formats, using graphical and text products, and are issued with increasing frequency as the severe weather time frame draws nearer. In addition to the scheduled Outlooks, Severe Thunderstorm and Tornado Watches are issued as needed to provide a higher level of alert over smaller regions in time and space when atmospheric conditions are favorable for severe thunderstorms and/or tornadoes to develop. The SPC also issues short-term Mesoscale Discussion products that emphasize hazardous weather on the mesoscale and often serve to fill the gap between the larger scale Outlooks and smaller scale Watches. The suite of specialized hazardous weather forecast products depends on the ability of SPC forecasters to assess the current state and evolution of the environment over varied time frames, and to synthesize a wide variety of observational and numerical model data sources. In general, observational data play a dominant role in diagnostic assessment for short-term forecasting, however, the development of more accurate and higher resolution models in recent years has allowed model information to influence the short-term prediction of convection as well. This is especially evident in the use of the hourly Rapid Update Cycle model, which forms a foundation for the SPC Mesoscale Analysis fields.

An effective NWS severe weather forecast and warning program should provide the public and other specialized users with sufficient advance notice of impending hazardous weather. Human response studies have shown that when a severe thunderstorm or tornado warning is issued, people are more likely to seek safe shelter if they have been made aware of the severe weather threat prior to the issuance of the warning. However, if they have not been pre-conditioned to the threat prior to hearing a warning, their first response is often to seek confirmation of the threat, rather than to seek shelter. This can result in the loss of critical reaction time when life and property are at immediate risk. Thus, there is a substantial need for the SPC to issue severe weather watches prior to the issuance of warnings by local NWS Weather Forecast Offices (WFOs), in order to allow WFO staffs, emergency managers, broadcast media, etc. sufficient time to implement contingency plans prior to the onset of severe weather.

This goal places additional requirements on SPC forecasters to determine in advance the characteristics of potential severe thunderstorm activity. Operational experience and research studies suggest that the type of severe weather that occurs (tornadoes, hail, or damaging winds) is often closely related to the convective mode (or morphology) exhibited by storms, such as discrete cells, squall lines (or quasi-linear convective systems (QLCS)), and multi-cellular convective systems. A disproportionate number of intense tornado and widespread straight-line wind damage events appear to be associated with two dynamically unique classes of thunderstorms: supercells and bow echoes. Thus, accurate severe weather watches are dependent on forecasters being able to predict properly not only where and when severe thunderstorms will develop and how they will evolve over the next 2 – 8 hours, but also the convective mode(s) that are most likely to occur.

There is also an increasing requirement to provide higher temporal resolution forecast information on thunderstorms and a variety of associated hazardous weather phenomena, including severe local storms, heavy rain/flash flooding, lightning strike potential, and aviation-related hazards of turbulence, icing, and low-level wind shear. Users such as emergency managers and other first responders, air traffic flow managers and others in transportation, power companies, etc., need greater time/space specificity in thunderstorm forecasts, and the SPC is now positioned to begin examining ways to provide higher temporal resolution convective forecasts.

Given the SPC's primary mission of mesoscale forecast responsibility, we continue to place a strong emphasis on assessing the current state of the atmosphere by using real-time observational data and derived diagnostic parameters for short-term thunderstorm prediction. However, owing to insufficient sampling of the mesoscale environment (especially when the horizontal and vertical distribution of water vapor is considered) coupled with limited scientific knowledge of important mesoscale and storm-scale processes, considerable uncertainty exists in the prediction of convection. While traditional operational models such as the NAM and GFS often can predict broader regions of precipitation utilizing parameterized convection, they are not capable of resolving important details of the smaller scale convective structure that are critical to severe weather forecasters. Furthermore, various proximity sounding studies using observed radiosondes and RUC model analyses indicate that the relationship between environmental characteristics (such as CAPE and vertical shear) and storm mode is not unique; rather it is found that similar storm types occur within different parts of the CAPE-shear parameter space, and different storm types occur within similar parts of parameter space. Therefore, in recent years the Spring Experiment has been focusing on the testing and evaluating cutting edge high resolution numerical weather prediction (NWP) models to determine potential contributions to operational severe weather forecasting.

Evaluation of High Resolution NWP in the Spring Experiment

Earlier research studies using idealized cloud resolving models to simulate deep convective storms at the National Center for Atmospheric Research (NCAR) and the University of Oklahoma Center for Analysis and Prediction of Storms (CAPS), among others, indicated that in some cases the models could replicate severe storm structures including supercells and bow echo systems. However, it was not until recently that sufficient computer resources, communications bandwidth, and advanced workstations became available to facilitate the testing of convection-allowing WRF models over large domains in a semi-operational forecasting environment, and to assess their potential utility for severe weather forecasting. It has been demonstrated over the last six years through Spring Experiments, field programs such as BAMEX, and daily use by SPC forecasters of experimental 4 km WRF models from the NCEP Environmental Modeling Center (EMC) and NSSL, that near-cloud resolving configurations of the WRF model can predict convective storms that, at times, appear remarkably similar to actual storms as seen on radar.

Progress has also been made in developing output fields such as simulated reflectivity that displays model-generated precipitation systems and storms that are visually similar to radar-derived images of actual storms. This allows forecasters to apply their knowledge of storm structure, intensity, and associated severe weather threats gained through observation of radar detected storms to aid in their interpretation of model generated storms. Furthermore, extraction of new parameters such as updraft helicity (a marker for a rotating updraft) has benefited forecasters by identifying explicit storm attributes that indicate enhanced severe potential. This is in contrast to traditional approaches where forecasters utilize mesoscale model output to provide information about evolution of the pre-convective environment, and then they use their knowledge of model biases and thunderstorm physical processes to determine the spectrum of storms that are possible. The first generation of operationally applied convection-allowing models takes this one step further, as they provide explicit information about the types of storms that may develop within predicted mesoscale environments.

Experiments with different WRF model configurations also indicate that it is not uncommon for each of the models to produce a variety of convective solutions for initiation, mode, and evolution, especially within more weakly forced environments. Thus, the model forecasts appear to reflect various uncertainties associated with real-world convective forecasting. These uncertainties arise primarily from: 1) the need to better sample and predict the pre-convective and near-storm environments, as deep convection can be sensitive to small variations in the mesoscale environment, and 2) limits in our understanding of smaller scale physical processes relevant to convection, which are modulated by mesoscale and stormscale forcing that are difficult to assess in the actual atmosphere.

Several years of experience with 00 UTC "cold start" WRF models using NAM model initial conditions and lateral boundary conditions (ICs/LBCs) have also revealed that it takes several hours of "spin up" time before the models can generate coherent, stable precipitation systems. These "cold start" runs are typically unable to provide substantial short-term guidance in the 0-6 hr time frame, but they have often demonstrated value in providing useful guidance for next day's diurnal heating cycle during the 18-30 hr forecast period. It has also been seen that the larger scale forcing provided by the "parent" NAM ICs/LBCs modulates the areas of convective storm development in the WRF models. This is particularly evident within strongly forced environments where the WRF convective storms have a tendency to occur in regions where the NAM generates larger scale areas of precipitation.

If WRF models initialized at 00 UTC are to provide useful forecast guidance for the next day's diurnal heating cycle, they must correctly spin up deep convection during the evening, then predict properly the evolution of the storms and their impact on the environment during the overnight hours. If this sequence of events is poorly represented, the pre-convective environment in the model during the subsequent afternoon may not replicate the actual environment, and the model prediction of storms may reflect errors in the environment specification. For example, if the 00 UTC model forecast erroneously maintains convective storm systems too late into the morning, the effects of precipitation, clouds, and an expanding low-level cold pool/convective outflow may maintain a stable environment that is unfavorable for later storm development. When this type of error occurred during the 2008 Experiment, the model(s) typically underpredicted afternoon storm development in areas where the spurious cold pool was located. On the other hand, when the 00 UTC models predicted correctly the evolution of nocturnal storms, they were much more likely to produce skillful forecasts of storms for the next afternoon and evening.

This indicates the critical importance of predicting correctly the evolution of the mesoscale environment, and suggests that the ability to run "update" models at later times with new ICs/LBCs can be of value to forecasters. In 2008, the EMC High Resolution Window WRF-NMM runs at 12 UTC were often compared with 00 UTC WRF runs on days when the earlier runs were determined to have predicted inaccurate environmental conditions by late morning (e.g., misplaced surface boundaries and errors in thermodynamic fields). In many of these cases, the 12 UTC update run predicted the afternoon environment more accurately and this translated into improved convective forecasts.

To fully capitalize on high resolution models to provide short-term forecast guidance on convective scales, advanced data assimilation techniques that include 3D radar reflectivity and velocity fields are necessary in order for the models to "know" where storms are located at the start of the model run. This very challenging task was introduced into the Spring Experiment in 2008, as CAPS used a real-time 3DVAR system to assimilate radar data over a three-fourths CONUS domain for the first time. Although the impact of the radar assimilation on the model forecasts typically appeared to diminish after several hours, this experimental arena will be a focus of activity in coming years.

Our experience has also shown that variations in WRF model convective storm predictions are at times difficult for operational forecasters to reconcile, in part because all solutions may appear to be plausible for a given mesoscale environment. Thus, the forecaster must determine how much confidence to place in specific model solutions, which is often difficult to assess because very high resolution models will attempt to predict phenomena (such as thunderstorms) on scales that are inherently unpredictable. The uncertainty in thunderstorm prediction suggests at least several possible research approaches to explore: 1) development of appropriate data assimilation systems for convection-allowing models to better resolve the initial conditions, and 2) improvement in the model itself with more realistic physics and increased resolution. However, inherent limits to the predictability of thunderstorms further suggest that application of ensemble forecasting concepts, currently used operationally for synoptic scale and mesoscale forecasting, may also be applicable to address challenges of convective-scale forecasting.

A Storm-Scale Ensemble Forecast (SSEF) system has been tested in Spring Experiments since 2007 to systematically explore aspects of uncertainty in thunderstorm prediction. Although questions remain concerning what are appropriate perturbation strategies for a convection-allowing ensemble system, experiments with a 10 ARW-member SSEF with mixed physics and ICs/LBCs have shown promising results. Development of new display tools for probabilistic assessment of thunderstorm potential and model-generated storm characteristics, utilizing "neighborhood" approaches that more properly reflect limits to grid scale predictability, have enhanced our ability to utilize SSEF output.

Finally, a key component of the annual experiments is the participation of operational forecasters from the SPC, other NCEP Centers, NWS WFOs, Environment Canada, and several private sector companies. Their insights and experience provide a real-world severe weather forecasting perspective when assessing the usefulness of convection-allowing WRF modeling systems, and provide them with opportunities to become familiar with cutting-edge science and technology applications before they are implemented operationally. This operational-research link increases the likelihood that HWT activities will result in improved severe weather forecasts and better public service. Forecaster interactions with model developers, research scientists, university faculty, and graduate students create a unique forum where a diverse mix of scientific backgrounds and insights work together to advance operationally relevant research and improve forecasts of hazardous convective weather.

IV. Experimental Models

The 2009 Spring Experiment will benefit from the continued participation and key contributions from CAPS, EMC, and NCAR, and new contributions from NOAA/Global System Division (GSD) and the Air Force Weather Agency (AFWA). Each of these collaborators (along with NSSL) will generate high resolution, convection-allowing WRF-based model guidance initialized at 00 UTC, and some will provide additional model runs at 12 UTC. In addition, GSD will provide an hourly update convection-allowing WRF model. Model domains will vary somewhat, but nearly all will cover at least the eastern two-thirds to three-fourths of the CONUS and all forecasts will extend to at least 30 hrs. The primary exceptions will be the 12 UTC CAPS WRF runs, which will focus on the VORTEX2 field program domain centered over the plains states and run to 18 hrs, and the hourly GSD update runs which will go out to 12 hrs.

CAPS Models – 4 km Storm Scale Ensemble Forecast, 1 km WRF-ARW, and 12 UTC 4 km WRF-ARW Runs

A major CAPS contribution will be a 20 member *Storm Scale Ensemble Forecast (SSEF) system* with grid spacing of 4 km and forecasts to 30 hrs, utilizing the resources of the Pittsburgh Supercomputing Center (PSC) and the National Institute for Computational Sciences (NICS)/University of Tennessee located at Oak Ridge National Laboratory. The SSEF will include additional model diversity this year with the inclusion of 8 WRF-NMM members and 2 ARPS members to complement the 10 WRF-ARW members utilized in the two previous years. The SSEF will draw additional initial condition (IC) and physics diversity from mixed IC/physics perturbations.

In all members, the background initial condition will come from interpolation of the 12 km NAM analysis. Mesoscale atmospheric perturbations will be introduced in the initial and lateral-boundary conditions of 14 members (8 ARW and 6 NMM) by extracting four pairs of positive/negative perturbations from EMC's operational Short Range Ensemble Forecast (SREF) system and applying them to the 14 members. In addition, convective-scale perturbations will be introduced in the initial conditions of 17 members by assimilating reflectivity and velocity data from radar and a cloud analysis as part of a CAPS 3DVAR system. For the remaining two ARW, two NMM, and two ARPS members, identical model configurations will be used for each pair and there will be no SREF-based perturbations. Radar data will be assimilated into one of the two ARW, NMM, and ARPS members (the control member), but not the other. Comparison of output from these two pairs of ARW, NMM, and ARPS members will allow us to isolate the impact of the radar data from other sensitivities at 4 km grid spacing.

Overall, the SSEF configuration builds upon lessons learned from the earlier SSEF systems tested during the 2007 and 2008 Spring Experiments, and the development this year of a larger multi-model, multi-physics, multi-IC SSEF is expected to be more robust and contain improved statistical performance. For operational forecasting applications, it is anticipated the SSEF will provide improved probabilistic guidance on high impact convective weather events by quantifying aspects of uncertainty and offering further insights about a possible range of solutions.

CAPS will also provide a *single WRF-ARW forecast at 00 UTC with a 1-km grid length* integrated to 30 hrs that is run at the NICS. Radar data will also be assimilated into the 1 km ARW but there will be no SREF-based perturbations. This will allow a direct comparison with the SSEF ARW control member and a clean measure of sensitivity to 1 versus 4 km grid spacing when radar data is assimilated. Statistical verification measures have indicated similar forecast results from the 2 and 4 km ARW forecasts that were produced for previous Spring Experiments, suggesting that the benefit gained by increasing horizontal resolution was not sufficient to justify the approximate eight-fold increase in computational resources to produce the 2 km run. Other very high resolution modeling studies have found that more realistic convective storms in terms of structure, size, and number of storms begin to appear when the grid spacing approaches 1 km, suggesting that the sensitivity to resolution may become more apparent this spring.

Finally, CAPS will produce two *4 km WRF-ARW runs initialized at 12 UTC* over a smaller domain centered on the VORTEX2 field program in the plains states. These two runs will be integrated to 18 hrs, and they will have physics configurations identical to the two SSEF ARW members without IC/LBC perturbations (one with radar assimilation and one without radar). These 12 UTC runs are designed to provide updated convective-scale guidance for afternoon and evening storms that is based on later initial conditions, and utilize resources at the University of Oklahoma Supercomputing Center for Education and Research.

The CAPS computational domain for the 00 UTC SSEF and 1 km WRF is in Fig. 1, and the 12 UTC WRF-ARW "VORTEX2" domain is in Fig. 2. The SSEF member configuration is provided in Tables 1, 2, and 3.

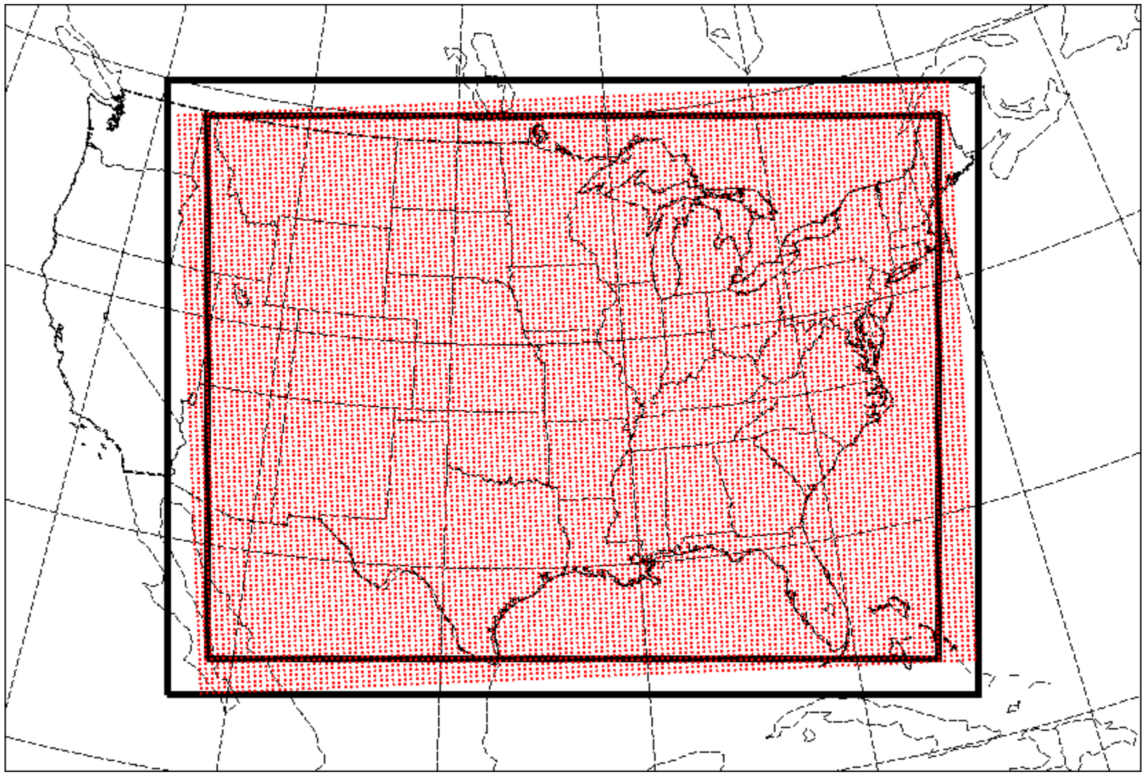


*Figure 1. CAPS 00 UTC domains for the 2009 Season. The outer thick rectangular box represents the domain for performing 3DVAR (**Grid 1** – 1000×760). The red dot area represents the WRF-NMM domain (**Grid 2** – 979x650). The inner thick box is the domain for WRF-ARW and ARPS, and also for common verification (**Grid3** - 900×672).*

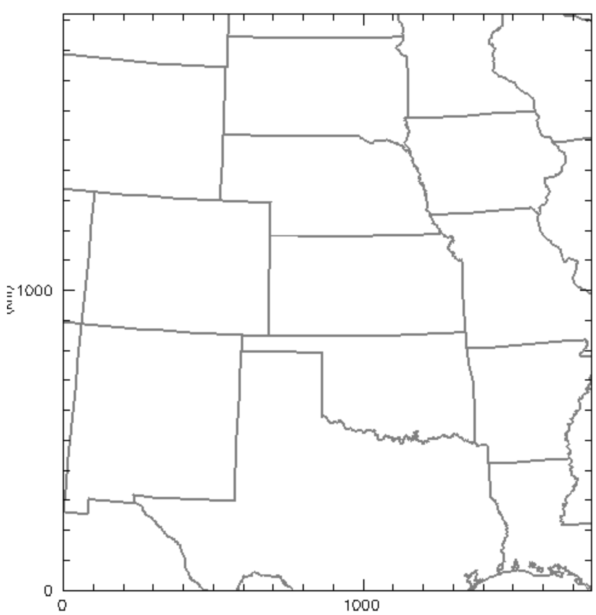


Figure 2. CAPS 12 UTC 4 km WRF-ARW update model domain with 444x480 horizontal grid points.

Table 1. Configurations for SSEF members with the ARW core. NAMa and NAMf refer to 12 km NAM analysis and forecast. ARPSa refers to ARPS 3DVAR. All ARW members use RRTM long-wave radiation. Member arw_n2 (shaded) is configured with physics used in the GSD-HRRR run.

member	IC	BC	Radar data	mp_phy	sw-phy	sf_phy	pbl_phy
arw_cn	00Z ARPSa	00Z NAMf	yes	Thompson	Goddard	Noah	MYJ
arw_c0	00Z NAMa	00Z NAMf	no	Thompson	Goddard	Noah	MYJ
arw_n1	arw_cn – em_pert	21Z SREF em-n1	yes	Ferrier	Goddard	Noah	YSU
arw_p1	arw_cn + em_pert	21Z SREF em-p1	yes	WSM 6-class	Dudhia	Noah	MYJ
arw_n2	arw_cn – nmm_pert	21Z SREF nmm-n1	yes	Thompson	Dudhia	RUC	MYJ
arw_p2	arw_cn + nmm_pert	21Z SREF nmm-p1	yes	WSM 6-class	Dudhia	Noah	YSU
arw_n3	arw_cn – etaKF_pert	21Z SREF etaKF-n1	yes	Thompson	Dudhia	Noah	YSU
arw_p3	arw_cn + etaKF_pert	21Z SREF etaKF-p1	yes	Ferrier	Dudhia	Noah	MYJ
arw_n4	arw_cn – etaBMJ_pert	21Z SREF etaBMJ-n1	yes	WSM 6-class	Goddard	Noah	MYJ
arw_p4	arw_cn + etaBMJ_pert	21Z SREF etaBMJ-p1	yes	Thompson	Goddard	RUC	YSU

Table 2. As in Table 1, except configurations for each SSEF member with the NMM core. Note that there are additional physics perturbations for long-wave radiation that are not included in the ARW members. (Note: the two gray shaded members have been removed from the current SSEF configuration owing to computing resource limitations at the PSC. This results in a total of 8 NMM members in the SSEF.)

Member	IC	BC	Radar data	mp_phy	lw_phy	sw-phy	sf_phy	pbl_phy
nmm_cn	00Z ARPSa	00Z NAMf	yes	Ferrier	GFDL	GFDL	Noah	MYJ
nmm_c0	00Z NAMa	00Z NAMf	no	Ferrier	GFDL	GFDL	Noah	MYJ
nmm_n1	nmm_cn – em_pert	21Z SREF em-n1	yes	Thompson	RRTM	Dudhia	Noah	MYJ
nmm_p1	nmm_cn + em_pert	21Z SREF em-p1	yes	WSM 6-class	GFDL	GFDL	RUC	MYJ
nmm_n2	nmm_cn – nmm_pert	21Z SREF nmm-n1	yes	Ferrier	RRTM	Dudhia	Noah	YSU
nmm_p2	nmm_cn + nmm_pert	21Z SREF nmm-p1	yes	Thompson	GFDL	GFDL	RUC	YSU
nmm_n3	nmm_cn – etaKF_pert	21Z SREF etaKF-n1	yes	WSM 6-class	RRTM	Dudhia	Noah	YSU
nmm_p3	nmm_cn + etaKF_pert	21Z SREF etaKF-p1	yes	Thompson	RRTM	Dudhia	RUC	MYJ
nmm_n4	nmm_cn – etaBMJ_pert	21Z SREF etaBMJ-n1	yes	WSM 6-class	RRTM	Dudhia	RUC	MYJ
nmm_p4	nmm_cn + etaBMJ_pert	21Z SREF etaBMJ-p1	yes	Ferrier	RRTM	Dudhia	RUC	YSU

Table 3. Configurations for each individual member with ARPS

member	IC	BC	Radar data	Microphy.	radiation	sf_phy
arps_cn	00Z ARPSa	00Z NAMf	yes	Lin	Chou/Suarez	Force-restore
arps_c0	00Z NAMa	00Z NAMf	no	Lin	Chou/Suarez	Force-restore

* For all members: no cumulus parameterization

EMC 4 km WRF-NMM Model

SPC forecasters have used output from earlier versions of the EMC WRF-NMM model since the spring of 2004. Several new attributes of the EMC 4 km WRF-NMM are available for this spring. It will be the first model run over a full CONUS domain at convection-allowing resolution (Fig. 3), and is now run twice daily at 00 and 12 UTC with forecasts to 36 hrs. The latter run time will provide a morning update to provide later guidance for afternoon and evening thunderstorms. It will continue to be nested within the 12 km NAM and incorporates NAM ICs/LBCs. In addition, five new parameter fields containing the maximum parameter value during the previous hour are available for the following fields (threshold values are scaled to a 4 km grid):

1. Simulated reflectivity (dBZ) at 1 km AGL
2. Updraft helicity (m2/s2) integrated through the 2-5 km AGL layer (Values associated with model supercells start ~ 50 m2/s2)
3. Updraft speed (m/s) in the lowest 400 mb (Values associated with strong model updrafts start ~ 10-15 m/s)
4. Downdraft speed (m/s) in the lowest 400 mb
5. 10 m wind speed (kt)

These can be considered "history variables" that track parameter values during each time step of the model integration, saving the maximum value that occurs at each grid point during each hour of the model run. They are intended fill in the temporal gaps between the standard top of the hour model output and provide unique information about the most intense storm attributes, which are unlikely to occur only at the hourly output times. For faster moving model storms, the hourly maximum parameters can also reveal storm attribute tracks. These special fields are nearly identical to the hourly maximum fields available in the WRF-ARW produced by NSSL (see below).

Note - if the EMC experimental CONUS WRF-NMM is not available, the operational High Resolution Window 4 km WRF-NMM will be used. This model is also nested within the 12 km NAM and is initialized twice daily at 00 and 12 UTC over a domain covering the eastern three-fourths of the CONUS, producing forecasts to 48 hrs. The hourly maximum parameters and model forecast soundings are not currently available from the HiResWindow runs.

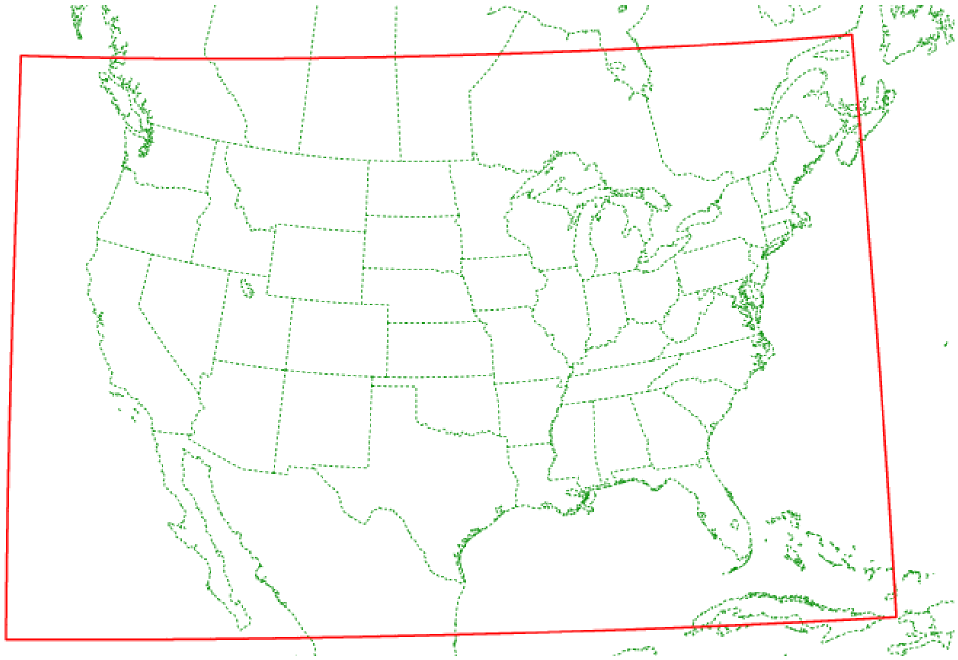


Figure 3. EMC 4 km WRF-NMM domain with 1239x920 horizontal grid points.

NSSL 4 km WRF-ARW Model

SPC forecasters have used output from a 4 km WRF-ARW produced by NSSL since the fall of 2006. This WRF model is run over a three-fourths CONUS domain (Fig. 4) once daily at 00 UTC, with forecasts to 36 hrs. The NSSL WRF pioneered the creation of hourly history variables that provide the maximum parameter value during the previous hour. These parameters are identical to those produced by the EMC WRF-NMM, except the maximum simulated reflectivity comes from the lowest model level (rather than 1 km AGL).

Since the NSSL and EMC WRF models are utilized heavily as "operational" models by SPC forecasters on a year-round basis, it is especially important to study and document their performance characteristics. The availability of the five hourly maximum parameter fields from both models have the potential to provide new insights about model generated storm intensity and temporal continuity, and these parameters will be compared during the experiment to assess their potential added value as forecast guidance for severe storms.

Figure 4. NSSL 4 km WRF-ARW domain with 980x750 horizontal grid points.

AFWA 4 km WRF-ARW Model

AFWA is contributing a WRF-ARW model run at 00 and 12 UTC that is nearly identical to the NSSL 4 km WRF-ARW, except it incorporates the NASA Land Information System (LIS) for the initial analysis. The LIS provides land and soil property information at high resolutions appropriate for convection-allowing models. Numerous studies have shown the importance of land-surface conditions to PBL evolution and subsequent development of thunderstorms, and the AFWA model will permit testing of sensitivity to the land-surface conditions represented by the NSSL (Noah) and AFWA (NASA LIS) configurations. The five maximum hourly parameter values available from the NSSL and EMC models are also produced by the AFWA model.

GSD 3 km High Resolution Rapid Refresh (HRRR) Model

The 3 km HRRR model is nested within the hourly 13 km RUC model, which provides ICs/LBCs for the HRRR. The HRRR uses a version of the WRF-ARW with generally "RUC-like" physics. A unique aspect of the RUC is the hourly data assimilation incorporates a wide array of observational datasets including radar reflectivity via the radar-Digital Filter Initialization. The HRRR integration is run over a two-thirds CONUS domain (Fig. 5) with forecasts to 12 hrs. The maximum parameter values during each forecast hour generated by the NSSL-ARW4 are also produced by the HRRR. It is being developed to serve users needing frequently updated short-range weather forecasts, including those in the US aviation and severe weather forecasting communities.

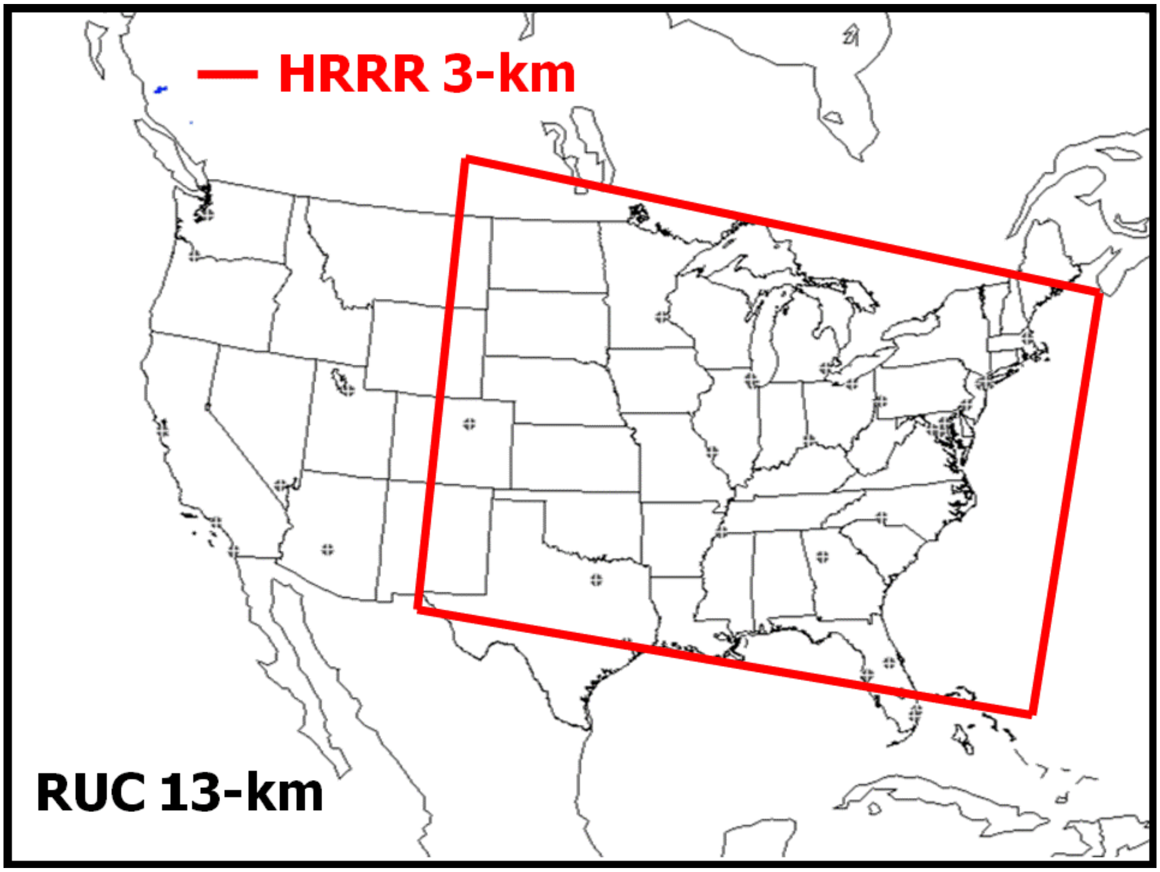


Figure 5. RUC 13 km domain (black) and HRRR 3 km domain (red) with 1000x700 horizontal grid points.

NCAR 3 km WRF-ARW Model

NCAR will focus this spring on running a 3 km WRF-ARW that utilizes initial conditions from the 13 km RUC that includes radar reflectivity via the radar-Digital Filter Initialization. This will use the same ICs that are used by the HRRR, but the LBCs for the NCAR WRF will be provided by the GFS model. The NCAR 3 km WRF-ARW will be run twice daily at 00 and 12 UTC with forecasts out to 48 hrs over a two-thirds CONUS domain (Fig. 6), so a comparison of guidance from sequential model forecasts at 12 hour intervals can be made.

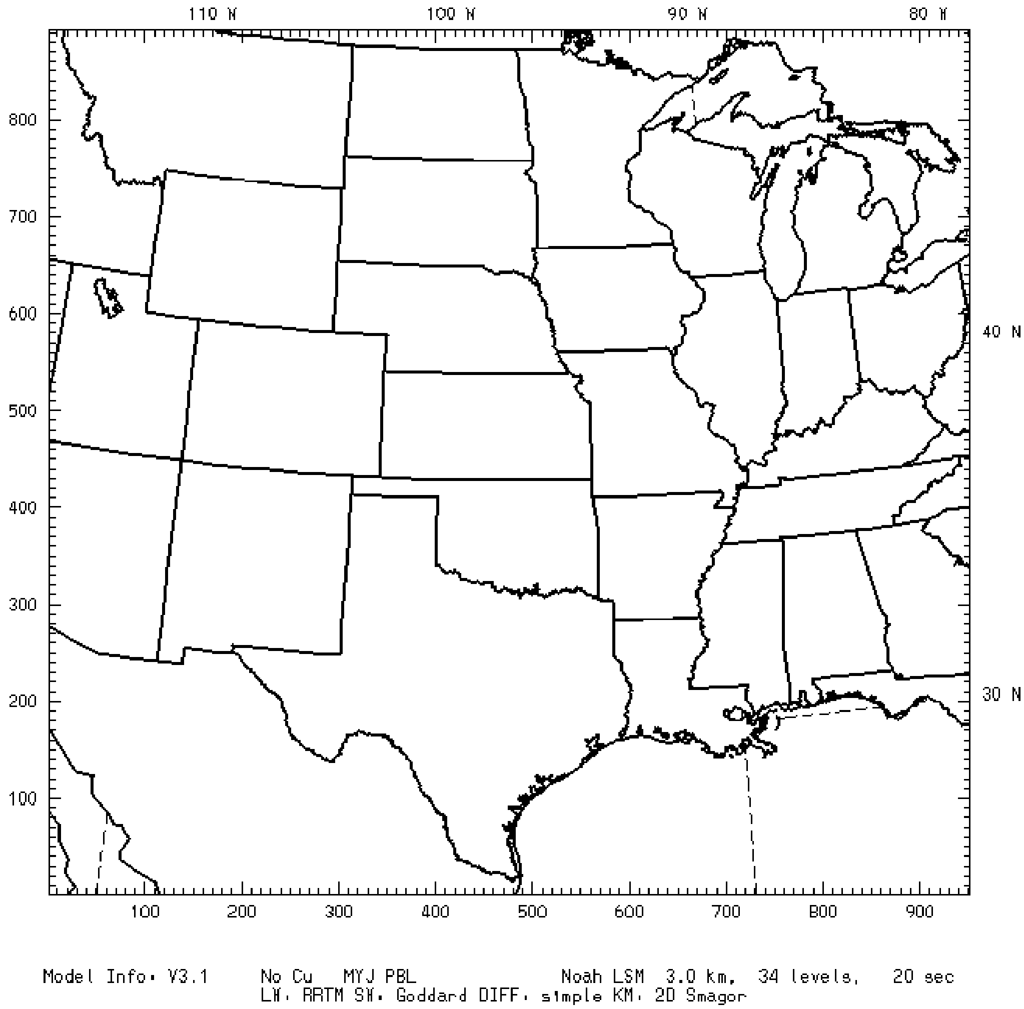


Fig. 6. NCAR 3 km WRF-ARW domain.

Table 3. Configurations of deterministic WRF models. The GSD-HRRR3 is initialized hourly with forecasts to 12 hrs; the EMC-NMM4 is initialized at 00 and 12 UTC with forecasts to 36 hrs; the NCAR-ARW3 is initialized at 00 and 12 UTC with forecasts to 48 hrs; the NSSL-ARW4 is initialized at 00 UTC with forecasts to 36 hrs; and the CAPS-ARW1 is initialized at 00 UTC with forecasts to 30 hrs.

	GSD-HRRR3 (ARW)	**EMC-NMM4**	**NCAR-ARW3**	**NSSL-ARW4**	**CAPS-ARW1**
Horiz. Grid (km)	3.0	4.0	3.0	4.0	1.0
Vertical Levels	50	35	34	35	51
PBL/Turb. Parameterization	MYJ	MYJ	MYJ	MYJ	MYJ
Microphysical Parameterization	Thompson	Ferrier	Thompson	WSM6	Thompson
Radiation (SW/LW)	Dudhia/RRTM	GFDL/GFDL	Goddard/RRTM	Dudhia/RRTM	Goddard/RRTM
Initial Conditions	13 km RUC	32 km NAM	13 km RUC	40 km NAM	CAPS-3DVAR

V. New Objective Verification Approaches

Subjective verification of model forecasts has been a cornerstone to HWT activities in previous years. This approach provides valuable insights into how forecasters use numerical models, and facilitates the gathering of information about the value of new guidance tools from the perspective of a forecaster. In addition, traditional verification measures (e.g., Equitable Threat Score or ETS) used for synoptic scale and mesoscale model forecasts of discontinuous variables such as precipitation typically provide less useful information (and even misleading information) about forecast accuracy as the scale of the phenomena being evaluated decreases. This is because the ETS is proportional to the degree of grid scale overlap in space and time between the forecasts and observations, and there is typically low predictability on convective scales. Despite these limits, operational severe weather forecasters have often found value in WRF forecasts of thunderstorms and convective systems, since they can provide unique information about convective mode, coverage, and evolution that is not resolved by mesoscale models using parameterized convection. In recent years, we have found that subjective evaluation has great potential to serve as a comparative benchmark for assessing new objective verification techniques designed for high resolution NWP, and has had a significant positive impact on model development strategies.

In order to better utilize subjective and objective verification techniques in a complementary manner, simulated composite reflectivity and 1-hr QPF output from several model runs will be evaluated using subjective visual comparisons and objective statistical measures produced by the WRF Developmental Testbed Center's Meteorological Evaluation Toolkit (MET). This evaluation will focus on comparing forecast performance during the first 12 hrs of the model integrations to assess the impact of radar data assimilation on short-term convective forecasts. The models include the two CAPS 00 UTC SSEF 4 km

ARW "control" runs (arw_cn member with radar assimilation and arw_c0 member without radar), the two CAPS 12 UTC 4 km ARW model runs over the VORTEX2 domain, and the GSD 3 km HRRR (with radar assimilation).

MET is designed to be a highly-configurable, state-of-the-art suite of verification tools. We will focus on the use of the object-based verification called Method for Object-based Diagnostic Evaluation (MODE) that compares gridded model data to gridded observations for the QPF and simulated reflectivity forecasts. MODE output will be tested to evaluate its ability to diagnose different types of convective modes considered important in forecasts and observations of convective weather, such as linear systems, discrete cells, and MCS's. Traditional verification statistics will also be computed. More information about DTC verification measures is found in Appendix F, and details about the DTC MET system is at http://www.dtcenter.org/met/users/ .

Verification "truth" will be provided by NSSL National Mosaic and Multi-Sensor QPE (NMQ) project's Quantitative Verification System (QVS). The QVS produces state-of-the-science high resolution multi-sensor Quantitative Precipitation Estimates (QPE) and three-dimensional radar reflectivity data bases. See http://www.nssl.noaa.gov/projects/q2/ for more information about the NMQ.

VI. Objectives and Expected Outcomes

The primary objectives of Spring Experiment 2009 are to:

- Test and evaluate an improved real-time, large domain, multi-model, multi-analysis, multi-physics Storm Scale Ensemble Forecast (SSEF) system during the prime severe weather season to gauge high performance computing, networking, data transfer and processing, product creation, and workstation display requirements for future high impact weather forecasting initiatives associated with the Warn-on-Forecast concept.

- Explore the relative impacts of assimilating radar reflectivity and velocity data into convection-allowing WRF models on short-term forecasts of hazardous convective weather, through comparison of forecasts from two SSEF ARW "control" members and the HRRR model during the first 12 hrs of the forecast period.

- From real-time and post analyses of the SSEF, determine strengths and limitations of the latest ensemble configuration, focusing on the statistical impact of adding 10 new NMM members to the existing 10 ARW members, and the incorporation of additional physics diversity in all members.

- Identify and test innovative ways to extract useful information from the SSEF and deterministic WRF models, and develop new products and display techniques that provide forecasters with meaningful probabilistic guidance on high impact convective weather events, including severe convective weather, heavy rain, and aviation applications.

- Explore the value of "update" WRF model runs using later initial conditions by comparing forecasts from the EMC and NCAR WRF models initialized at 00z with forecasts from the same models run at 12z. Furthermore, examine the utility of providing forecasters with convection-allowing HRRR model forecasts that are updated on an hourly basis.

- Expand previous subjective model evaluation approaches to include traditional and new objective verification measures produced by the DTC and test their utility to provide unique and meaningful information about convection-allowing model performance.

- Provide focused feedback to model developers on the performance of the experimental SSEF and deterministic models during severe thunderstorm episodes.

The experiment expected outcomes include:

- Documentation of statistical verification properties of the SSEF, leading to improvements in subsequent SSEF formulations.

- Documentation of the utility of a SSEF to quantify uncertainty and provide probabilistic guidance for high impact convective weather events, including severe storms, heavy rain, and applications for aviation.

- Documentation of the impacts of incorporating data assimilation systems to initialize convection-allowing WRF models and the SSEF, including radar reflectivity and velocity data.

- Confirmation and further documentation of the ability of convection-allowing WRF models to provide unique information on convective mode, intensity, and evolution, and how operational severe weather forecasters can better utilize this guidance in daily forecasting.

- Documentation of the evolving complimentary relationship between operational mesoscale deterministic models, the current mesoscale SREF, and convection-allowing WRF models including the SSEF in quantifying uncertainty in high impact convective weather forecasts.

- Documentation of the ability of traditional and new objective verification approaches to provide meaningful quantitative information about high resolution convection-allowing model performance.

- Internal NWS documentation of challenges to the real-time display and utilization of very high resolution NWP output in an operational forecast setting.

- Enhanced communication and collaboration between forecasters and model developers leading to enhancements in the transfer of research to operations.

- Continued effective collaboration between research scientists, model developers, and forecasters during the Spring Experiment with high participant satisfaction as measured by responses to a survey form given to all participants.

VII. Spring Experiment Web Site

A full description of all program objectives, types of model output, forecast products, evaluation and verification forms, a data archive, and other related links are available at the Spring Experiment web site:

http://hwt.nssl.noaa.gov/Spring_2009/

This web site is intended to support real time activities as well as additional research and reference after the conclusion of the program.

VIII. Dates of the Spring Experiment

The 2009 Spring Experiment will run Monday-Thursday 8 am – 4 pm from May 4 through June 5, 2008. Friday sessions will end by 10-11 am as no forecast activities take place on Friday. **On each Monday, participants are asked to arrive by 7:30 am for a brief orientation session.** During the last week of April, final spin-up activities will be tested with in-house participants only. Beginning May 4, a full range of in-house and external participants will staff the program. Full time participants will work for periods of one week, with part-time visiting scientists and forecasters participating on a 2-3 day basis (schedule permitting). Program operations will be conducted in the Hazardous Weather Testbed facility (Room 2380) located on the second floor of the NWC between the SPC and WFO Norman operations areas. Each full time weekly team will complete daily experimental forecasts and participate in evaluation and verification activities; part-time visitors can participate in daily activities at levels appropriate with their interest and expertise. Staffing typically will include one SPC forecaster, one or more NSSL scientists, and a number of visiting scientists, model developers, forecasters, university faculty, and graduate students. A list of weekly participants is found in Appendix A.

IX. Daily Operations Schedule

Participants in the experiment will create experimental forecast products and conduct evaluation activities in the HWT from 8 am - 4 pm on Monday-Thursday, with a short wrap-up session Friday mornings when final evaluation activities and a weekly summary are conducted.

Participants are expected to perform forecast and evaluation activities in a collaborative manner, such that results reflect a consensus decision. Participants may eat lunch while conducting program activities or at their discretion any time during the day. Visitors may purchase lunch at a food court located on the south side of the first floor of the NWC. Below is an outline of the daily schedule for activities during the experiment; a more detailed description is found in Appendix B.

Daily Operations Plan (Monday only)
Italics denotes Monday-only activities

7:30 am-8:00 am: Weekly Orientation

8:00 am-8:30 am: HWT Coordinators will review and explain Friday's evaluation and results of the previous day's initial and updated experimental forecasts (A briefing/demo of what was done Friday and what the new team will do Tuesday morning)

8:30-10:45 am: Break into forecast teams; chart analysis; select forecast and evaluation domain; prepare and issue initial experimental two-period severe storm probability forecast graphics for 20-00z and 00-04z (all severe; significant severe)

10:45-11:00 am: Team briefings on their respective experimental forecasts

11:00 am-12:15 pm: HWT Coordinators and DTC participant will review and explain Friday's model evaluation activities and results (This will be a briefing/demo of what was done Friday and what the new team will do Tuesday morning)

- NMM-NSSL comparison (reflectivity and hourly maximum fields)
- NSSL-AFWA comparison (reflectivity and 2m temperature/dewpoint)
- CAPS SSEF ARW CN-CAPS 1 km comparison (reflectivity)
- 00z CAPS ARW C0-CN and HRRR comparison (0-12 hr subjective evaluation and objective DTC verification)
- NCAR 00z-12z comparison

12:15-1:00 pm: Lunch, informal discussions, start preparation of afternoon forecast updates

1:00-2:30 pm: Break into forecast teams again; prepare and issue updated experimental two-period severe storm probability forecast graphics for 20-00z and 00-04z (all severe; significant severe)

2:30-2:45 pm: Team briefings on their respective experimental forecasts

2:45-3:00 pm: Break time, prepare for Spring Experiment End-of-Day Briefing and Discussion

3:00-4:00 pm: Daily Wrap-up Briefing and Discussion including contributions from DTC, GOES-R PG, and VORTEX Operations Center. (SSEF discussion including possible aviation applications will be part of briefing.)

4:00 pm: Select Preliminary Day 2 forecast and evaluation domain for overnight 00z model graphics processing.

Daily Operations Plan (Tuesday-Thursday)

8:00 am-8:30 am: Review previous day severe weather and evaluate initial and updated experimental forecasts

8:30-10:45 am: Break into forecast teams; chart analysis; select forecast and evaluation domain; prepare and issue initial experimental two-period severe storm probability forecast graphics for 20-00z and 00-04z (all severe; significant severe)

10:45-11:00 am: Team briefings on their respective experimental forecasts

11:00 am-12:15 pm: Conduct previous day model evaluation activities

- NMM-NSSL comparison (reflectivity and hourly maximum fields)
- NSSL-AFWA comparison (reflectivity and 2m temperature/dew point)
- CAPS SSEF ARW CN-CAPS 1 km comparison (reflectivity)
- 00z CAPS ARW C0-CN and HRRR comparison (0-12 hr subjective evaluation and objective DTC verification)
- NCAR 00z-12z comparison

12:15-1:00 pm: Lunch, informal discussions, start preparation of afternoon forecast updates

1:00-2:30 pm: Break into forecast teams again; prepare and issue updated experimental two-period severe storm probability forecast graphics for 20-00z and 00-04z (all severe; significant severe)

2:30-2:45 pm: Team briefings on their respective experimental forecasts

2:45-3:00 pm: Break time, prepare for Spring Experiment End-of-Day Briefing and Discussion

3:00-4:00 pm: Daily Wrap-up Briefing and Discussion including contributions from DTC, GOES-R PG, and VORTEX Operations Center. (SSEF discussion including possible aviation applications will be part of briefing.)

4:00 pm: Select Preliminary Day 2 forecast and evaluation domain for overnight 00z model graphics processing.

Daily Operations Plan (Friday only)
No experimental forecast activities are conducted on Friday

8:00-8:30 am: Review previous day severe weather and evaluate initial and updated experimental forecasts

8:30-10:00 am: Conduct previous day model evaluation activities

- NMM-NSSL comparison (reflectivity and hourly maximum fields)
- NSSL-AFWA comparison (reflectivity and 2m temperature/dewpoint)
- CAPS SSEF ARW CN-CAPS 1 km comparison (reflectivity)
- 00z CAPS ARW C0-CN and HRRR comparison (0-12 hr subjective evaluation and objective DTC verification)
- NCAR 00z-12z comparison

10:00-10:30 am: Weekly wrap-up discussion including lessons learned and key questions that arose during the week, and topics to focus on in coming weeks

X. Acknowledgments

Special thanks and appreciation is extended to many people for their creative insights and assistance in Spring Experiment preparations, planning, and execution of numerous complex and ground-breaking technical and scientific activities. Without the combined efforts of many SPC and NSSL staff, the Spring Program could not be conducted. In particular, special thanks go to SPC's Gregg Grosshans for helping to establish model data flow and configuring the experimental forecasts for transmission and archival, and for helping to organize model display files, Andy Dean for data processing and graphics display support, and Jay Liang and Joe Byerly for expertise in configuring and upgrading hardware/software, network and workstations in the HWT. Expert NSSL technical support was provided by Brett Morrow, Jeff Horn, Steve Fletcher and Brad Sagowitz to address networking, data flow, hardware, and archive requirements. Linda Crank (SPC), Peggy Stogsdill (SPC), and Tracy Reinke (CIMMS) ably assisted with logistical and budget support activities. J. J. Gourley's assistance in organizing the NSSL Seminar Series for visiting scientists during the experiment is also greatly appreciated.

The experimental activities could not take place without the dedicated collaborative efforts of many people at CAPS, EMC, and NCAR who are working to enhance community efforts to improve severe weather forecasting. We acknowledge the expertise of CAPS scientists Ming Xue, Kelvin Droegemeier, Fanyou Kong, Kevin Thomas, Yunheng Wang, Jidong Gao, and Keith Brewster for outstanding efforts to develop and run the 20 member SSEF, and the 1 km and 4 km WRF models; the Pittsburgh Supercomputing Center, the National Institute for Computational Sciences/University of Tennessee, and the University of Oklahoma Supercomputing Center for Education and Research for providing technical support and computer facilities for the CAPS model runs; NCAR scientists Morris Weisman, Wei Wang, Greg Thompson, and Jimy Dudhia for developing the NCAR WRF runs and offering technical support and advice on the ARW system and configuration of the SSEF; EMC scientists Matt Pyle, Zavisa Janjic, Brad Ferrier, Jun Du, Zoltan Toth, and Geoff DiMego for developing and contributing the WRF-NMM models and for scientific input and infrastructure support for the SSEF; Stan Benjamin, Steve Weygandt, Curtis Alexander, and John Brown at ESRL/GSD for development of the large domain HRRR model for use in the Spring Experiment; Jon Case, Sujay V. Kumar, and colleagues at NASA/Huntsville for providing the Land Information System for use at AFWA; Evan Kuchera at AFWA for his expert collaborative efforts to configure and make available a WRF model incorporating the NASA LIS; and Barb Brown, Tara Jensen, and Steve Sullivan from the DTC for their expert development of the MET verification system and special efforts to tailor it for application during the Spring Experiment.

We further wish to recognize the full support of SPC and NSSL management; and the numerous contributions and insights provided by the many participants who clearly demonstrated the value of collaborative experiments involving the research, academic, and forecasting communities, and whose presence and enthusiasm resulted in a positive learning experience for everyone.

Appendix A: Spring Experiment Participant Schedule

Weekly Calendar of EFP Spring Experiment Participants – 2009

Mon April 27	Tue April 28	Wed April 29	Thu April 30	Fri May1
(Spin-Up Week) Mike Coniglio Jack Kain Jason Levit David Bright	*(Spin-Up Week)* Mike Coniglio Jason Levit David Bright	*(Spin-Up Week)* Mike Coniglio Jason Levit David Bright	*(Spin-Up Week)* Mike Coniglio Jack Kain Steve Weiss Jason Levit David Bright	*(Spin-Up Week)* Mike Coniglio Jack Kain Steve Weiss Jason Levit David Bright
Mon May 4	**Tue May 5**	**Wed May 6**	**Thu May 7**	**Fri May 8**
John Brown Mike Hardiman Tom Hultquist Evan Kuchera Craig Schwartz Mike Coniglio Jack Kain Steve Weiss Tara Jensen GOES-R PG	John Brown Mike Hardiman Tom Hultquist Evan Kuchera Craig Schwartz Mike Coniglio Jack Kain Steve Weiss Tara Jensen GOES-R PG	John Brown Mike Hardiman Tom Hultquist Evan Kuchera Craig Schwartz Mike Coniglio Jack Kain Steve Weiss Tara Jensen Rebecca Schneider GOES-R PG	John Brown Bryan Smith Mike Hardiman Tom Hultquist Evan Kuchera Craig Schwartz Mike Coniglio Jack Kain Steve Weiss Tara Jensen GOES-R PG	John Brown Bryan Smith Mike Hardiman Tom Hultquist Evan Kuchera Craig Schwartz Mike Coniglio Jack Kain Steve Weiss Tara Jensen GOES-R PG
Mon May 11	**Tue May 12**	**Wed May 13**	**Thu May 14**	**Fri May 15**
Jon Racy Bruce Entwistle Andy Fischer Chris Gitro John Huhn Nigel Roberts Steve Weygandt Mike Coniglio Jack Kain Steve Weiss Jamie Wolff GOES-R PG	Jon Racy Bruce Entwistle Andy Fischer Chris Gitro John Huhn Nigel Roberts Steve Weygandt Mike Coniglio Jack Kain Steve Weiss Sarah Wong Jamie Wolff GOES-R PG	Jon Racy Bruce Entwistle Andy Fischer Chris Gitro John Huhn Nigel Roberts Steve Weygandt Mike Coniglio Jack Kain Steve Weiss Jamie Wolff GOES-R PG	Jeff Evans Bruce Entwistle Andy Fischer Chris Gitro John Huhn Nigel Roberts Steve Weygandt Mike Coniglio Jack Kain James Cummine Jamie Wolff Steve Koch GOES-R PG	Bruce Entwistle Andy Fischer Chris Gitro John Huhn Nigel Roberts Steve Weygandt Mike Coniglio Jack Kain Jamie Wolff Steve Koch GOES-R PG
Mon May 18	**Tue May 19**	**Wed May 20**	**Thu May 21**	**Fri May 22**
John Hart Lance Bosart Tom Galarneau Ryan Sobash Morris Weisman Jon Zeitler Mike Coniglio Jack Kain Steve Weiss Dave Ahijevych GOES-R PG	John Hart Lance Bosart Tom Galarneau Ryan Sobash Morris Weisman Jon Zeitler Mike Coniglio Jack Kain Steve Weiss Dave Ahijevych GOES-R PG	John Hart Lance Bosart Tom Galarneau Ryan Sobash Morris Weisman Jon Zeitler Mike Coniglio Jack Kain Steve Weiss Dave Ahijevych GOES-R PG	Jon Racy Lance Bosart Tom Galarneau Ryan Sobash Morris Weisman Jon Zeitler Mike Coniglio Jack Kain Steve Weiss Dave Ahijevych GOES-R PG	Lance Bosart Tom Galarneau Ryan Sobash Morris Weisman Jon Zeitler Mike Coniglio Jack Kain Steve Weiss GOES-R PG

Mon May 25	Tue May 26	Wed May 27	Thu May 28	Fri May 29
Holiday-No Operations	Jared Guyer Faye Barthold Jon Case Adam Clark Matt Eastin Mike Fries Bill Martin John Mejia Mike Coniglio Jack Kain Steve Weiss Barb Brown GOES-R PG	Jared Guyer Faye Barthold Jon Case Adam Clark Matt Eastin Mike Fries Bill Martin John Mejia Jack Kain Matt Wandishin Steve Weiss Barb Brown GOES-R PG	Jared Guyer Faye Barthold Jon Case Adam Clark Matt Eastin Mike Fries Bill Martin John Mejia Jack Kain Matt Wandishin Steve Weiss GOES-R PG	Jared Guyer Faye Barthold Jon Case Adam Clark Matt Eastin Mike Fries Bill Martin John Mejia Jack Kain Matt Wandishin GOES-R PG
Mon Jun 1	**Tue Jun 2**	**Wed Jun 3**	**Thu Jun 4**	**Fri Jun 5**
Steve Goss Mike Bodner Jim Clark Mike Fowle Amy Harless Reid Hawkins Matt Pyle Derek Stratman Jack Kain Matt Wandishin Steve Weiss Tressa Fowler GOES-R PG	Steve Goss Mike Bodner Jim Clark Mike Fowle Amy Harless Reid Hawkins Matt Pyle Derek Stratman Jack Kain Matt Wandishin Steve Weiss Tressa Fowler GOES-R PG	Steve Goss Mike Bodner Jim Clark Mike Fowle Amy Harless Reid Hawkins Matt Pyle Derek Stratman Jack Kain Matt Wandishin Steve Weiss GOES-R PG	Steve Goss Mike Bodner Jim Clark Mike Fowle Amy Harless Reid Hawkins Matt Pyle Derek Stratman Jack Kain Matt Wandishin Steve Weiss GOES-R PG	Steve Goss Mike Bodner Jim Clark Mike Fowle Amy Harless Reid Hawkins Matt Pyle Derek Stratman Jack Kain Matt Wandishin Steve Weiss GOES-R PG

EFP Spring Experiment 2009 Participants and Affiliations

Week of April 27 (Internal Spin-Up Week)
Mike Coniglio (NOAA/OAR NSSL)
Jack Kain (NOAA/OAR NSSL)
Steve Weiss (NOAA/NWS/NCEP SPC)
Jason Levit (NOAA/NWS/NCEP SPC)
David Bright (NOAA/NWS/NCEP SPC)

Week of May 4
Bryan Smith (NOAA/NWS/NCEP SPC)
John Brown (NOAA/OAR/ESRL/GSD Boulder CO)
Mike Hardiman (NOAA/NWS El Paso, TX)
Tom Hultquist (NOAA/NWS Chanhassen-Minneapolis, MN)
Evan Kuchera (AFWA Offutt AFB, Bellevue NE)
Rebecca Schneider (Environment Canada, Montreal)
Craig Schwartz (University of Oklahoma, Norman and NSSL)
Mike Coniglio (NOAA/OAR NSSL)
Jack Kain (NOAA/OAR NSSL)
Steve Weiss (NOAA/NWS/NCEP SPC)
Tara Jensen (NCAR/Developmental Testbed Center, Boulder, CO)
1-2 Participants from the GOES-R Proving Ground, Norman, OK

Week of May 11
Jon Racy (NOAA/NWS/NCEP SPC)
Jeff Evans (NOAA/NWS/NCEP SPC)
James Cummine (Environment Canada, Winnipeg)
Bruce Entwistle (NOAA/NWS/NCEP AWC)
Andy Fischer (NOAA/NWS/NCEP AWC)
Chris Gitro (NOAA/NWS Midland, TX)
John Huhn (Mitre Corp/FAA, McLean VA)
Nigel Roberts (United Kingdom Meteorological Office/JCMM, Reading)
Steve Weygandt (NOAA/OAR/ESRL/GSD Boulder, CO)
Sarah Wong (Environment Canada, Toronto)
Mike Coniglio (NOAA/OAR NSSL)
Jack Kain (NOAA/OAR NSSL)
Steve Weiss (NOAA/NWS/NCEP SPC)
Jamie Wolff (NCAR/Developmental Testbed Center, Boulder, CO)
Steve Koch (NOAA/ESRL/GSD and Developmental Testbed Center, Boulder, CO)
1-2 Participants from the GOES-R Proving Ground, Norman, OK

Week of May 18
John Hart (NOAA/NWS/NCEP SPC)
Jon Racy (NOAA/NWS/NCEP SPC)
Lance Bosart (University at Albany-SUNY)
Tom Galarneau (University at Albany-SUNY)
Ryan Sobash (University of Oklahoma, Norman and NSSL)
Morris Weisman (NCAR, Boulder CO)
Jon Zeitler (NOAA/NWS San Antonio/Austin, TX)
Mike Coniglio (NOAA/OAR NSSL)
Jack Kain (NOAA/OAR NSSL)
Steve Weiss (NOAA/NWS/NCEP SPC)
Dave Ahijevych (NCAR/Developmental Testbed Center, Boulder, CO)
1-2 Participants from the GOES-R Proving Ground, Norman, OK

Week of May 26
Jared Guyer (NOAA/NWS/NCEP SPC)
Faye Barthold (NOAA/NWS/NCEP HPC)
Jon Case (ENSCO Inc./SPoRT Center, Huntsville, AL)
Adam Clark (Iowa State University, Ames)
Matt Eastin (University of North Carolina at Charlotte)
Mike Fries (NOAA/NWS Spokane, WA)
Bill Martin (NOAA/NWS Glasgow, MT)
John Mejia (NOAA/OAR NSSL)
Mike Coniglio (NOAA/OAR NSSL)
Jack Kain (NOAA/OAR NSSL)
Matt Wandishin (NOAA/OAR NSSL)
Steve Weiss (NOAA/NWS/NCEP SPC)
Barb Brown (NCAR/Developmental Testbed Center, Boulder, CO)
1-2 Participants from the GOES-R Proving Ground, Norman, OK

Week of June 1
Steve Goss (NOAA/NWS/NCEP SPC)
Mike Bodner (NOAA/NWS/NCEP HPC)
Jim Clark (NOAA/NWS/NCEP OPC)
Mike Fowle (NOAA/NWS Aberdeen, SD)
Amy Harless (University of Oklahoma, Norman and SPC)
Derek Stratman (University of Oklahoma, Norman)
Reid Hawkins (NOAA/NWS Wilmington, NC)
Matt Pyle (NOAA/NWS/NCEP EMC)
Derek Stratman (University of Oklahoma, Norman)
Jack Kain (NOAA/OAR NSSL)
Matt Wandishin (NOAA/OAR NSSL)
Steve Weiss (NOAA/NWS/NCEP SPC)
Tressa Fowler (NCAR/Developmental Testbed Center, Boulder, CO)
1-2 Participants from the GOES-R Proving Ground, Norman, OK

Appendix B: Detailed Daily Schedule and Order of Activities (Subject to Modification)

Monday only
7:30-8:00 am: Weekly orientation. Introduction to SPC and NSSL, followed by individual participant introductions, experience and interests; purpose and goals of HWT and Spring Experiments; collaborative daily activities

Monday-Thursday
Morning activities primarily focus on evaluation of previous day human and model forecasts

1. 8:00 – 8:30 am: Human forecast evaluation
(Note – on Mondays the HWT coordinators will review will review and explain Friday's evaluation and results of the previous day's initial and updated experimental forecasts, This is a briefing/demo of what was done Friday and what the new team will do Tuesday morning)

Subjective verification of the previous day preliminary and final two-period probabilistic severe weather forecasts issued by both forecast teams for the 20-00 and 00-04 UTC periods over the selected regional domain. Formulate an overall rating by averaging the accuracy of different forecast areas when necessary. Areas with greater severe storm occurrence, higher forecast probabilities, and the forecast or occurrence of significant reports should be given more weight in the rating process. Numerical ratings of 0-10 and descriptive text information are entered into internal web page survey form.

Primary data sources: Web page displays of probability forecast graphics and corresponding plot of severe reports with "practically perfect" forecast contour overlays. (Information about the "practically perfect" forecast concept is found in Appendix D.) Forecasts display contours of standard SPC probability values for all severe storms (less than 5%, 5, 15, 30, 45, 60%) and 10% or greater probability of significant severe (hatched area) if needed.

2. 8:30 – 10:45 am: Selection of current day regional forecast domain; chart analysis and formulation of preliminary two-period probabilistic severe weather forecasts.

2a. Selection of current day regional forecast domain.
The two-period experimental severe weather forecasts will be valid for the 20-00 and 00-04 UTC periods corresponding to the afternoon and evening climatological peak in severe storms, and a regional domain will be selected for the forecasts. The forecast area will be restricted to fall within the common domain of all 00 UTC WRF model guidance, and is typically placed where the most intense severe storms are expected to occur. However, areas where the forecast is considered to be particularly difficult can also be considered in the domain selection process. The domain is selected by entering a three letter METAR station ID centered on the domain into an internal web page.

Primary data sources: 13 UTC SPC Day 1 Convective Outlook available on SPC web page, brief discussion with SPC forecasters, and early examination of observational and model data.

2b. Chart Analysis
Standard SPC 12 UTC upper air (250 mb, 500 mb, 700 mb, and 850 mb) and the latest surface charts will be provided for subjective hand analysis to determine the synoptic and mesoscale observational background setting for the afternoon experimental forecasting activity.

A brief discussion of the analysis findings will take place focusing on the relationship of specific features (jet streaks, short-wave troughs, surface boundaries) and kinematic/thermodynamic ingredients (vertical shear, moisture), and relevant observed soundings.

2c. Formulation of preliminary experimental severe weather forecasts (Forecast leaders: SPC Forecaster and Team 2 Coordinator (Weiss, Kain, Coniglio, Wandishin, or Carbin); SSEF resource: Bright; WRF resource: Kain)
The participants will be assigned to "East" and "West" forecast teams, which will be randomly chosen each day. One team will be lead by the weekly SPC forecaster, and one by a Spring Experiment coordinator with severe weather forecasting background and familiarity with N-AWIPS data and functionality. The forecast teams will use N-AWIPS to examine a wide variety of observational data (e.g., METAR, satellite, radar including wind profilers, sfcoa fields (hourly SPC Mesoscale Analysis), and both operational (e.g., NAM, RUC, SREF, WRF-NMM) and experimental (various convection-allowing WRF models and CAPS SSEF) numerical model output. Two separate forecasts valid 20-00 and 00-04 UTC will be created in N-AWIPS. Each forecast consists of probability contours of all severe events (combined tornado, hail, and wind events) and significant severe events over the regional domain, using standard SPC outlook product probability conventions (<5%, 5%, 15%, 30%, 45%, ≥60%, and ≥10% significant severe).

Brief descriptive text information is entered into internal web page survey form describing the role of model data, especially the convection-allowing WRF models and SSEF, in the forecast decision making process, including identification of specific specialized products considered to be useful for impact weather forecasting. ***A lengthy "SPC-style" outlook synopsis and forecast discussion will not be done***. Instructions for electronically creating and submitting the experimental forecast products, and a table showing the relationship between the probabilistic forecasts and SPC categorical outlooks (Slight, Moderate, High Risk) is contained in Appendix C.

3. 10:45-11:00 am: Team Discussions of Their Preliminary Forecasts
Each forecast team will briefly discuss their severe weather forecast emphasizing what steps they took to assess the current severe weather environment and ongoing convection, and key inputs to determine the evolution of ongoing storms as well as expected development of new convection. A focus will be placed on the use and interpretation of

WRF and SSEF model data as they relate to convective initiation, evolution, and mode, and the likelihood of specific convection details.

4. 11:00 am – 12:15 pm: Evaluation of Previous Day WRF Model Forecasts

4a: Evaluation of 1 km AGL simulated reflectivity forecasts

Subjective verification of 1 km AGL simulated reflectivity forecasts from deterministic WRF models during 20-04z forecast period over the same regional domain. Assessment includes how well model reflectivity forecasts corresponded to observed reflectivity, including convective initiation, direction and speed of system movement, areal coverage, configuration and orientation of mesoscale features, and convective mode. The evaluations compare forecast performance of model pairs that address specific needs of forecasters and/or model developers. Subjective comparisons and descriptive text information are entered into internal web page survey forms.

The following model reflectivity comparative evaluations will be made:

- 00z NMM-NSSL comparison (and hourly maximum fields with severe reports)
- 00z NSSL-AFWA comparison
- 00z CAPS SSEF ARW CN-CAPS 1 km comparison
- NCAR 00z-12z comparison

Primary data sources: Web page 3-panel displays showing hourly model reflectivity forecasts and observed BREF images from 18-06z. Additional fields such as the maximum parameter value during the previous hour may be examined from the WRF-NSSL4 and EMC-NMM4. *(Time period covers 2 hrs before and after the forecast periods to assess possible timing errors.)*

4b: Comparison of 00z CAPS SSEF 4 km control members with/without radar and GSD 3 km HRRR

Subjective verification focusing on impacts of CAPS 3DVAR radar data assimilation and GSD radar-DFI assimilation on model analyses of 00hr composite reflectivity and short term model spin-up and maintenance of coherent reflectivity fields during the first 12 hrs of the model runs. Descriptive text information is entered into internal web page survey form.

Primary data sources: Web page 4-panel display showing hourly model composite reflectivity forecasts and observed CREF images from 00z through 12z.

4c: Objective Verification of CAPS SSEF 4 km control members with/without radar and GSD 3 km HRRR model predictions using DTC-MET system (Leader: DTC Expert)

This will focus on objective verification of model 1-hr QPF and reflectivity forecasts over selected regional domain focusing on precipitation/convective structure and object-oriented analysis results produced by MET-Mode. Assessment includes how well the MET-Mode results compare to subjective evaluation. Descriptive text information is entered into internal web page form.

Primary data sources: Special DTC webpage displaying model forecasts and MET-Mode object-oriented structure analysis and traditional verification measures from MET.

5. 12:15 – 1:00 pm: Lunch, discussions, and resumption of weather monitoring as preparation for afternoon forecast update

Participants will break to obtain lunch, and they are encouraged to bring food back and eat in the HWT to continue discussions and resume weather monitoring including surface chart analysis.

6. 1:00 – 2:30 pm: Formulation of final experimental severe weather forecasts (Forecast leaders: SPC Forecaster and Team 2 Coordinator Team 2 Coordinator

Participants will again break into their "East" and "West" forecast teams, and using latest observational and model datasets, update their morning two-period experimental severe weather forecasts. Teams should assess the value of the 12z WRF model guidance from CAPS, EMC, and NCAR plus hourly HRRR runs from GSD relative to the earlier 00z WRF guidance, using current observational data as a comparison benchmark. The preliminary two-period probabilistic severe weather forecasts will be modified as necessary through the consideration of more recent model and observational data. New descriptive text information is entered into a internal web page survey form describing the impact of new model and observational data on the update process. Again, a lengthy "SPC-style" outlook synopsis and forecast discussion will not be done. Instructions for electronically creating and submitting the experimental forecast products, and a table showing the relationship between the probabilistic forecasts and SPC categorical outlooks (Slight, Moderate, High Risk) is contained in Appendix C.

7. 2:30-2:45 pm: Team Discussions of Their Final Forecasts

Each forecast team will briefly discuss their severe weather forecast emphasizing what steps they took to assess the current severe weather environment and ongoing convection, and key inputs to determine the evolution of ongoing storms as well as expected development of new convection. A focus will be placed on the use and interpretation of WRF and SSEF model data as they relate to convective initiation, evolution, and mode, and the likelihood of specific convection details.

8. 3:00 – 4:00 pm: Daily Wrap-up Briefing and Discussion including Contributions from DTC, GOES-R PG, and VORTEX Operations Center.

Participants will briefly summarize the team experimental severe weather forecasts, including changes made in the final forecast. The daily wrap-up will solicit input from all participants to identify key results from the forecast and evaluation activities, plus new questions we should explore on subsequent days. In short, what do we think we learned today, what questions came up that we can't answer at this time, and what specific topics should we explore further in the coming days. Key findings will be documented and saved in the online data archive.

In addition, there will be time available to share information and results of new objective verification tools from the DTC that are being utilized in the Spring Experiment, and several satellite-based analysis and short-term prediction algorithms that are being tested in the GOES-R Proving Ground. Finally, updates from the VORTEX Operations Center on current field program activities will be presented when time permits.

(SSEF discussion including possible aviation applications will be part of briefing.)

9. 4:00 pm: Select Preliminary Day 2 forecast and evaluation domain for overnight 00z model graphics processing.

Appendix C: Instructions for Creating and Submitting Experimental Forecasts

1. Experimental Forecast Graphics
Preliminary two-period severe weather forecasts will be issued in the morning and **final** two-period forecasts will be issued in the afternoon. The forecast valid time periods will be 20-00z and 00-04z. The severe weather forecast graphics will be very similar to operational SPC outlooks, except only total severe storm probability contours will be formulated (no categorical outlook, and no general thunderstorms will be forecast). The same probability contours used in the operational outlooks will be used for the severe forecasts (5, 15, 30, 45, and 60 %); an area delineating potential for significant severe storms will be included when the probability for significant severe is 10% or greater. The Probability-to-Categorical conversion for total severe is identical to that used for the SPC Day 2 Outlook, and is shown below in tabular form.

2. Drawing and Saving the Experimental Forecasts in NMAP
a. For the preliminary and final forecasts, the forecaster will draw in NMAP separate probability contours for each valid period, and will save each forecast as a separate graphic product. The process will utilize NMAP software that is used in SPC operations. When saving each experimental forecast graphic, the following modifications are required:

1) In the format outlook box, ***change valid time to 2000z to 0000z (or 0000z to 0400z)***
2) In the product save box, ***replace "outlook" with "west_prelim (final)" or "east_prelim (final)"***

b. Enter command in xterm window: sp09bg STN team-name forecast # (such as sp09bg OKC east final 2)

STN is METAR centerpoint ID**, team-name** is "east" or "west", **forecast** is "prelim" or "final", and # is NAWIPS workstation number (1-6) where the graphic is created**.** This script archives the severe weather forecast, attaches date/time to the graphics file, and sends graphics to the web page.

3. Completing Model Discussion Section on Internal Web Page
a. On HWT Spring Experiment web page click on Experimental Forecast Generation (East or West)
b. Click on "Preliminary" or "Final" and the two-period forecast graphics will appear
c. Complete Discussion Text Box and when finalized, click on Submit.

Probability to Categorical Outlook Conversion
(SIGNIFICANT SEVERE area needed where denoted by hatching - otherwise default to next lower category)

Outlook Probability	Combined TORN, WIND, and HAIL
5%	SEE TEXT
15%	SLGT
30%	SLGT
45%	MDT
60%	HIGH

Appendix D: Practically Perfect Forecasts

(From Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. *Preprints*, 19th Conference on Severe Local Storms, Minneapolis, Minnesota, American Meteorological Society, 552-555.)

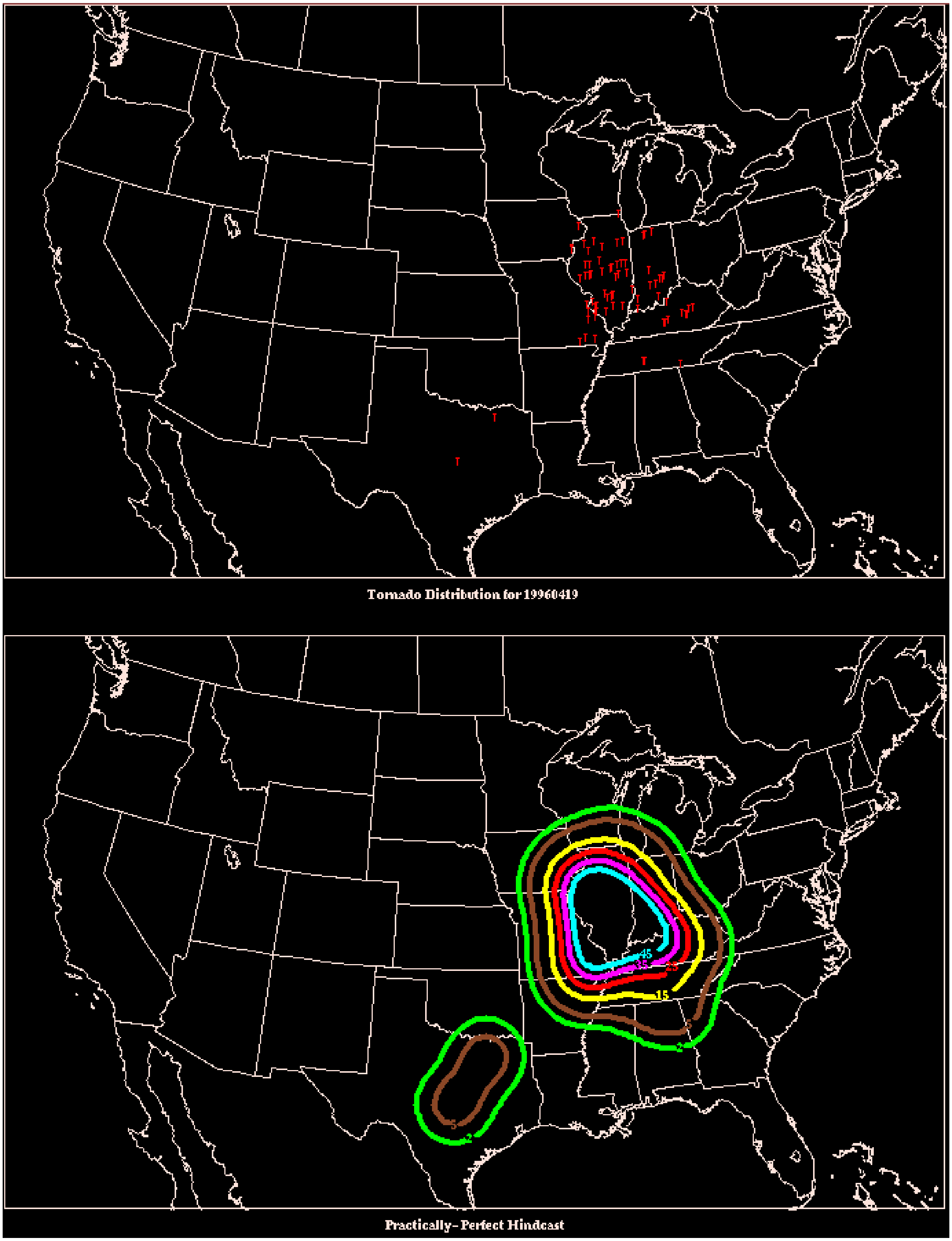
Severe weather forecasts such as SPC outlook and watch products are issued with the explicit expectation that there will be "false alarms" (parts of the forecast for which there are no events) and "missed detections" (events which are not included in the forecast). Thus, the expected range of values of the probability of detection (POD) or false alarm rate (FAR), for example, does not run from 0 to 1 in practice. The concept of a "practically" perfect (PP) forecast can then be used to estimate the minimum and maximum scores that a forecaster could reasonably be expected obtain given real world distributions of severe weather reports and the low predictability of specific severe convective storms in advance. In general, that range will be much smaller than the absolute minimum and maximum, but will provide a range over which meaningful forecast performance can be judged.

To compute the PP forecast, reports of severe weather are recorded on a grid with each grid box representing an area 80 x 80 km. (This grid corresponds to SPC Outlook products where probability values correspond to a probability within 25 miles of each grid point.) All severe weather reports are considered equal and the computation considers only whether a box has had an event or not. The PP forecast is then created by smoothing the events using nonparametric density estimation with a two dimensional Gaussian kernel. Specifically, at each grid point in the domain, the PP forecast value, *f*, is given by

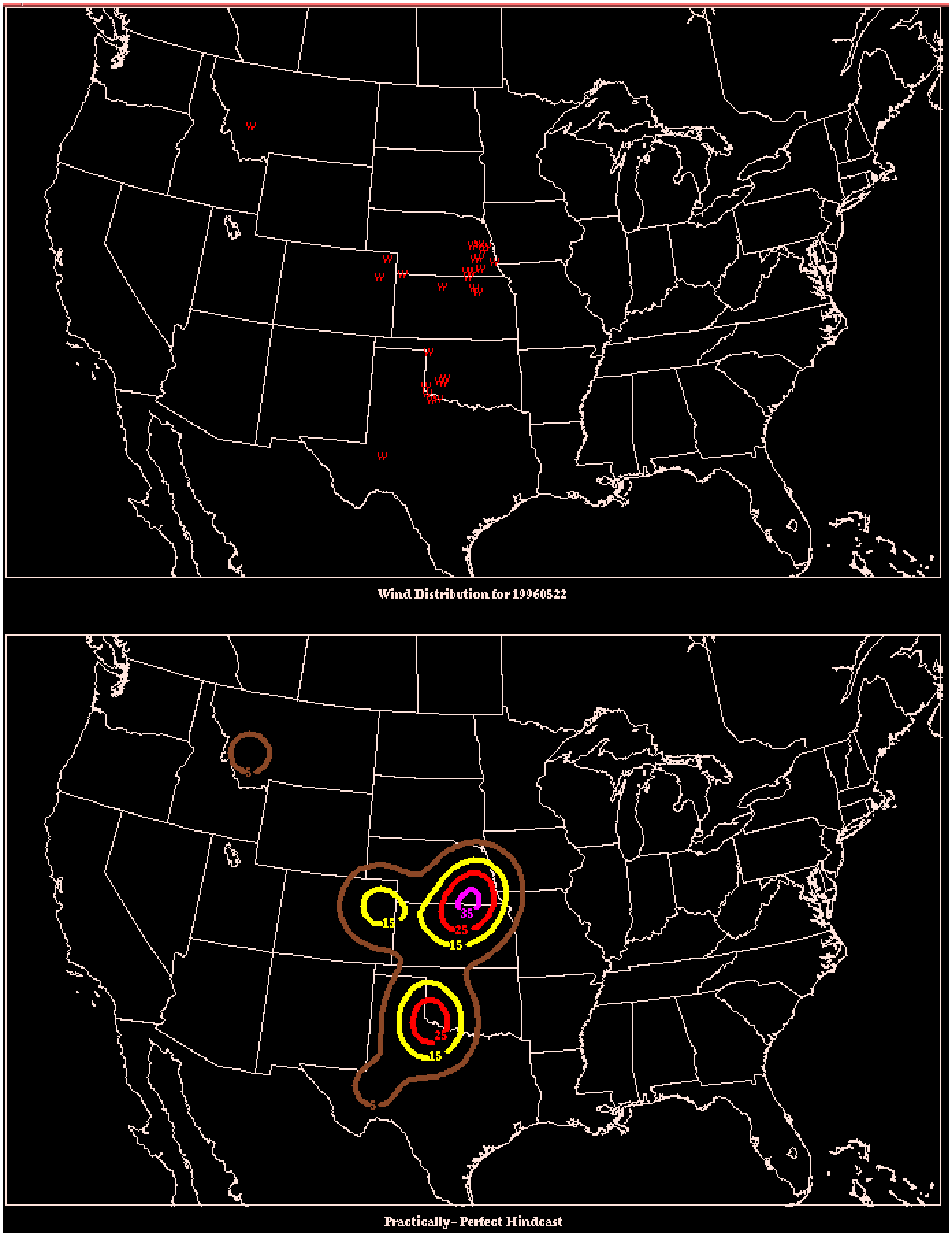
$$f = \sum_{n=1}^{N} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2} \left[\frac{d_n}{\sigma} \right]^2 \right)$$

where *dn* is the distance from the forecast grid point to the *n*-th location that had an event occur, *N* is the total number of grid points with events, and σ is a weighting function that can be interpreted as the confidence one has in the location of the forecast event. Increasing σ is
is equivalent to increasing the uncertainty associated with the forecast as one would do with increasing lead time of the forecast. That is, in the context of severe weather forecasting, very small σ can be thought of as being associated with the warning stage, while larger σ is associated with the watch or convective outlook stages. For SPC forecasts a value of 3 used.

Examples of practically perfect forecasts based on actual severe weather reports are shown on the next two pages.



19 April 1996 tornado reports (top) and PP forecast (bottom) based on tornado reports



22 May 1996 wind damage reports (top) and PP forecast (bottom) based on the wind reports only

Appendix E: WRF Model Identification of Convective Storms with Rotating Updrafts – Computation of Updraft Helicity

1. Storm Relative Environmental Helicity

Helicity, H, is a scalar measure of the potential for helical flow (i.e., the pattern of a corkscrew) to develop in a moving fluid defined by

$$H = \vec{V} \bullet \nabla \times \vec{V} .$$

Expressed in its component form,

$$H = u(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}) + v(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}) + w(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}) .$$

The portion of helicity associated with the *storm relative streamwise component* is that along the ambient horizontal velocity vector, or

$$H_s = -(u - c_u)(\frac{\partial v}{\partial z}) + (v - c_v)(\frac{\partial u}{\partial z}) ,$$

where is the storm motion and terms involving w neglected. Integrating H_s vertically through the thunderstorm inflow layer, z, yields the *storm relative environmental helicity*, SREH,

$$SREH = -\int_{z_o}^{z} [(u - c_u)(\frac{\partial v}{\partial z}) - (v - c_v)(\frac{\partial u}{\partial z})] dz .$$

SREH is a commonly used parameter to assess the severe thunderstorm potential of the environment and is often integrated from the surface to 1 - 3 km AGL. Order of magnitude values of SREH are ~ O(50) to O(300) m^2/s^2 in environments that tornadic storms.

2. Updraft Helicity

With the availability of numerical models containing sufficient resolution to resolve convective processes explicitly, it is now possible to calculate a *vertical component of helicity* associated with the convective updraft. This is the vertical integral of the third term in equation (2) and referred to as *updraft helicity*, U_H. Thus,

$$U_H = \int_{z_o}^{z} [w(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})] dz = \int_{z_o}^{z} [w\zeta] dz$$

where is the vertical component of the relative vorticity at grid points where w > 0. In post processing the WRF members for the SPC/NSSL Spring Program, equation (5) is integrated vertically from z_o = 2 km to z = 5 km AGL using a midpoint approximation. Data are available every 1000 meters AGL, so equation (5) is computed as

$$U_H = \int_{z_o}^{z} [w\zeta]dz \approx \sum_{z=2000m}^{z=5000m} (\overline{w\zeta}\Delta Z) = (\overline{w\zeta}_{2,3} + \overline{w\zeta}_{3,4} + \overline{w\zeta}_{4,5}) \times 1000 \text{ ,}$$

where the over bar indicates a layer average and the subscripts indicate the bottom and top of the layer in kilometers. Early experience indicates that typical values of U_H associated with WRF predicted supercell thunderstorms are have U_H of at least ~O (50) m^2/s^2 and that significant supercells have U_H ~O (150) m^2/s^2.

Appendix F: DTC Verification Metrics Summary

Traditional Verification Metrics

Statistics for dichotomous (2-category) variables

For dichotomous variables (e.g., precipitation amount above or below a threshold) on a grid, typically the forecasts are evaluated using a diagram like the one shown in Fig. 1. In this diagram, the area "**H**" represents the intersection between the forecast and observed areas, or the area of **Hits**; "**M**" represents the observed area that was missed by the forecast area, or the "**Misses**"; and "**F**" represents the part of the forecast that did not overlap an area of observed precipitation, or the "**False Alarm**" area. A fourth area is the area outside both the forecast and observed regions, which is often called the area of "**Correct Nulls**" or "**Correct Rejections**".

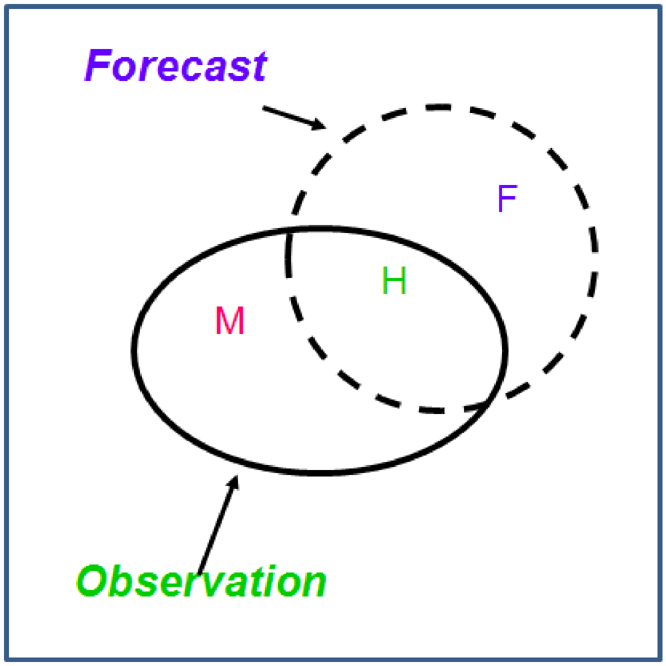


Figure 1. Diagram showing hits, misses, and false alarms for dichotomous forecast/observations.

This situation can also be represented in a "contingency table" like the one shown in Table 1. In this table the entries in each "cell" represent the counts of hit, misses, false alarms, and correct rejections. The counts in this table can be used to compute a variety of traditional verification measures, described in the following sub-sections.

Table 1. Contingency table illustrating the counts used in verification statistics for dichotomous (e.g., Yes/No) forecasts and observations. The values in parentheses illustrate the combination of forecast value (first digit) and observed value. For example, YN signifies a Yes forecast and and a No observation.

Forecast	Observed		
	Yes	**No**	
Yes	Hits (YY)	False alarms (YN)	YY + YN
No	Misses (NY)	Correct rejections (NN)	NY + NN
	YY + NY	YN + NN	Total = YY + YN + NY + NN

Base rate

$$\text{Base rate} = \frac{\text{Hits} + \text{Misses}}{\text{Total}} = \frac{\text{YY+NY}}{\text{Total}}$$

Also known as **sample climatology** or **observed relative frequency of the event.**

Answers the question: *What is the relative frequency of occurrence of the Yes event?*

Range: 0 to 1.

Characteristics: Only depends on the observations. For convective weather can give an indication of how "active" a day is.

Probability of detection (POD)

$$\text{POD} = \frac{\text{Hits}}{\text{Hits} + \text{Misses}} = \frac{\text{YY}}{\text{YY} + \text{NY}}$$

Also known as **Hit Rate.**

Answers the question: *What fraction of the observed Yes events was correctly forecasted?*

Range: 0 to 1. **Perfect score:** 1.

Characteristics: Sensitive to hits, but ignores false alarms. Good for rare events. Can be artificially improved by issuing more Yes forecasts to increase the number of hits. Should be used in conjunction with the false alarm ratio (below) or at least one other dichotomous verification measure. POD also is an important component of the Relative Operating Characteristic (ROC) used widely for evaluation of probabilistic forecasts.

False alarm ratio (FAR)

$$\text{FAR} = \frac{\text{False alarms}}{\text{Hits} + \text{False alarms}} = \frac{\text{YN}}{\text{YY} + \text{YN}}$$

Answers the question: *What fraction of the predicted "yes" events did* ***not*** *occur (i.e., were false alarms)?*

Range: 0 to 1. **Perfect score:** 0.

Characteristics: Sensitive to false alarms, but ignores misses. Very sensitive to the climatological frequency of the event. Should be used in conjunction with the probability of detection (above). Relative Operating Characteristic (ROC) used widely for evaluation of probabilistic forecasts.

Bias

$$\text{Bias} = \frac{\text{Hits} + \text{False alarms}}{\text{Hits} + \text{Misses}} = \frac{\text{YY} + \text{YN}}{\text{YY} + \text{NY}}$$

Also known as **Frequency Bias.**

Answers the question: *How similar were the frequencies of Yes forecasts and Yes observations?*

Range: 0 to infinity. **Perfect score:** 1.

Characteristics: Measures the ratio of the frequency of forecast events to the frequency of observed events. Indicates whether the forecast system has a tendency to underforecast (Bias < 1) or overforecast (Bias > 1) events. Does not measure how well the forecast gridpoints correspond to the observed gridpoints, only measures overall relative frequencies. Can be difficult to interpret when number of Yes forecasts is much larger than number of Yes observations.

Critical Success Index (CSI)

Also known as **Threat Score (TS).**

$$\text{CSI} = \text{TS} = \frac{\text{Hits}}{\text{Hits} + \text{Misses} + \text{False alarms}} = \frac{\text{YY}}{\text{YY} + \text{NY} + \text{YN}}$$

Answers the question: *How well did the forecast "yes" events correspond to the observed "yes" events?*

Range: 0 to 1, 0 indicates no skill. **Perfect score:** 1.

Characteristics: Measures the fraction of observed and/or forecast events that were correctly predicted. It can be thought of as the *accuracy* when correct negatives have been removed from consideration. That is, CSI is only concerned with forecasts that are important (i.e., assuming that the correct rejections are not important). Sensitive to hits, penalizes both misses and false alarms. Does not distinguish the source of forecast error. Depends on climatological frequency of events (poorer scores for rarer events) since some hits can occur purely due to random chance. Non-linear function of POD and FAR. Should be used in combination with other contingency table statistics (e.g., Bias, POD, FAR).

<u>Gilbert Skill Score (GSS)</u>

Also commonly known as **Equitable Threat Score (ETS).**

$$GSS = ETS = \frac{\text{Hits} - \text{Hits}_{\text{random}}}{\text{Hits} + \text{Misses} + \text{False alarms} - \text{Hits}_{\text{random}}} = \frac{YY - YY_{\text{random}}}{YY + NY + YN - YY_{\text{random}}}$$

where

$$\text{Hits}_{\text{random}} = YY_{\text{random}} = \frac{(\text{Hits} + \text{False alarms})(\text{Hits} + \text{Misses})}{\text{Total}} = \frac{(YY + YN)(YY + NY)}{\text{Total}}$$

Answers the question: *How well did the forecast "yes" events correspond to the observed "yes" events (accounting for hits that would be expected by chance)?*

Range: -1/3 to 1; 0 indicates no skill.   **Perfect score:** 1.

Characteristics: Measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for the frequency of hits that would be expected to occur simply by random chance (for example, it is easier to correctly forecast rain occurrence in a wet climate than in a dry climate). The GSS (ETS) is often used in the verification of rainfall in NWP models because its "equitability" allows scores to be compared more fairly across different regimes; however it is not truly equitable. Sensitive to hits. Because it penalizes both misses and false alarms in the same way, it does not distinguish the source of forecast error.   Should be used in combination with at least one other contingency table statistic (e.g., Bias).

Statistics for continuous forecasts and observations

For this category of statistical measures, the grids of forecast and observed values – such as precipitation or reflectivity – are overlain on each other, and error values are computed. The grid of error values is summarized by accumulating values at all of the grid points and used to compute measures such as mean error and root mean squared error.

These statistics are defined in the sub-sections below. In the equations in these sections, $\boldsymbol{f_i}$ signifies the **forecast value** at gridpoint *i*, $\boldsymbol{o_i}$ represents the observed value at gridpoint *i*, and ***N*** is the **total number** of gridpoints.

Mean error (ME)

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)$$

Also called the **(additive) Bias.**

Answers the question: *What is the average forecast error?*

Range: minus infinity to infinity. **Perfect score:** 0.

Characteristics: Simple, familiar. Measures *systematic* error. Does not measure the magnitude of the errors. Does not measure the correspondence between forecasts and observations; it is possible to get a **perfect ME score for a bad** forecast if there are compensating errors.

Pearson Correlation Coefficient (r)

$$r = \frac{\sum_{i=1}^{n}(f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{n}(f_i - \bar{f})^2}\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}}$$

where $\bar{f}$ is the average forecast value and $\bar{o}$ is the average observed value.

Also called the **linear correlation coefficient.**

Answers the question: *What is the linear association between the forecasts and observations?*

Range: -1 to 1. **Perfect score**: 1

Characteristics: *r* can range between -1 and 1; a value of 1 indicates perfect correlation and a value of -1 indicates perfect negative correlation. A value of 0 indicates that the forecasts and observations are not correlated. The correlation does not take into account the mean error, or additive bias; it only considers linear association.

Mean squared error (MSE) and root-mean squared error (RMSE)

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

$$\text{RMSE} = \sqrt{\text{MSE}}$$

MSE can be re-written as

$$\text{MSE} = (\bar{f} - \bar{o})^2 + s_f^2 + s_o^2 - 2 s_f s_o r_{fo},$$

where $\bar{f}$ is the average forecast value, $\bar{o}$ is the average observed value, s_f is the standard deviation of the forecast values, s_o is the standard deviation of the observed values, and r_{fo} is the correlation between the forecast and observed values. Note that $\bar{f} - \bar{o} = \text{ME}$ and $s_f^2 + s_o^2 - 2 s_f s_o r_{fo}$ is the estimated variance of the error, s_{f-o}^2. Thus, $\text{MSE} = \text{ME}^2 + s_{f-o}^2$. To understand the behavior of **MSE**, it is important to examine *both* of these terms of **MSE**, rather than examining **MSE** alone. Moreover, **MSE** can be strongly influenced by **ME**, as shown by this decomposition.

The standard deviation of the error, s_{f-o}, is simply $s_{f-o} = \sqrt{s_{f-o}^2} = \sqrt{s_f^2 + s_o^2 - 2 s_f s_o r_{fo}}$.

Note that the standard deviation of the error (ESTDEV) is sometimes called the "Bias-corrected MSE" (BCMSE) because it removes the effect of overall bias from the forecast-observation squared differences.

Answers the question: *What is the average magnitude of the forecast errors?*

Range: 0 to infinity. **Perfect score:** 0.

Characteristics: Simple, familiar. Measures "average" error, weighted according to the square of the error. Does not indicate the direction of the deviations. The RMSE puts greater influence on large errors than smaller errors, which may be a good thing if large errors are especially undesirable, but may also encourage conservative forecasting.

MODE summary metrics

The Method for Object-based Diagnostic Evaluation (MODE) identifies and matches spatial objects in the forecast and observed fields. A convolution radius (*r*) and a precipitation/reflectivity threshold (*t*) are used to identify objects; different combinations of these parameters lead to objects with different characteristics, and can be used to evaluate forecasts as a function of threshold and scale.

In the object matching and merging[1] process, all possible pairs of forecast and observed objects are assigned a total "interest" value. This value is formulated from the weighted sum of specific interest values that are associated with differences in particular attributes between the forecast and observed objects. According to the current weighting scheme, the total interest value is large when objects are located close to each other and are about the same size, and is smaller for pairs of objects that are further apart and have different sizes. Note that users can specify other components of interest, and their relative weights, in the configuration file for running MODE, according to what is most relevant for their particular application.

Figure 2 illustrates a scenario in which three forecast objects and two observed objects have been identified in the two fields. The total interest values for all of the pairs of forecast and observed objects are shown in the associated table. In previous work an interest threshold of 0.70 has been found to be a reasonable indicator of a good match. Thus, in this case, forecast object 1 is a good match with both observed objects 1 and 2, and forecast object 3 matches well with observed object 2. Forecast object 3 does not match well with either of the observed objects, mostly because of its small size. Because both forecast objects 1 and 2 match observed object 2, and forecast object 1 also matches observed object 1, these objects form a matched "cluster" in the forecast and observed fields.

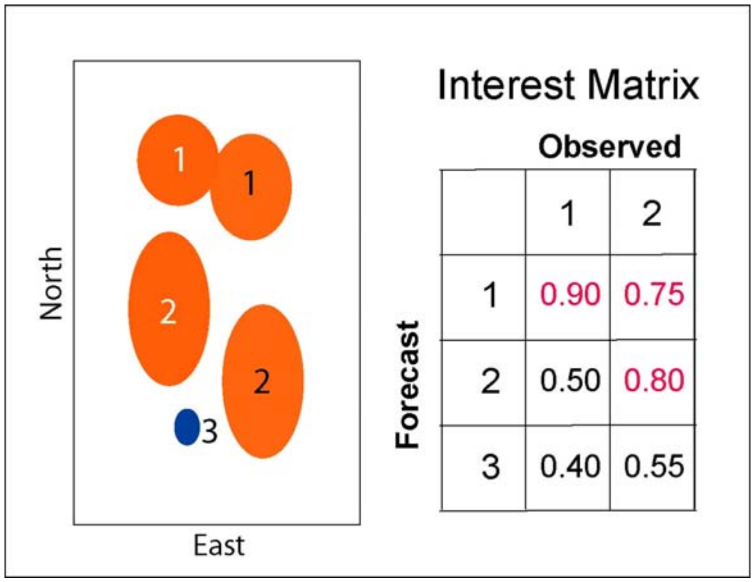


Figure 2. Schematic showing hypothetical forecast rain objects (black numerical labels) and observed rain objects (white numerical labels) with the corresponding interest matrix at right. Orange-shaded objects are matched whereas blue shading denotes no match. Total interest values greater than 0.7 are shown in red numbers in matrix. From Davis et al. (2009).

Some of the forecast attributes that are (or can be considered) in determining matches between forecast and observed objects include object size, distribution of intensity values, orientation angle, and location. Comparisons of these attributes, along with the total interest values, also can be used to help measure the quality of the forecast performance. Some of the measures that can be used to summarize performance using MODE are described in the following subsections.

[1] "Merging" refers to the connection of objects in the same field, while "matching" refers to the connection between objects in the forecast and observed field.

Median of Maximum Interest (MMI)

This measure is computed using the total interest values for all of the pairs of objects. It considers the maximum total interest values associated with each forecast object and each observed object. From this set, the median value is computed and is the MMI.

Example: Forecast and observed objects in Fig. 2
Maximum interest values for all of the forecast and observed objects are as follows:

For forecast object 1, the maximum total interest is 0.90.
For forecast object 2, the maximum total interest is 0.80.
For forecast object 3, the maximum total interest is 0.55.
For observed object 1, the maximum total interest is 0.90.
For observed object 2, the maximum total interest is 0.80.

The median of those 5 numbers is 0.80, so MMI = 0.80.

This number can be small because no objects match well, or because there are many extra objects that don't match well.

Larger MMI values imply a better match between forecast and observed objects.

Area-Weighted CSI

Area Weighted Critical Success Index (AWCSI)

AWCSI = [(hit area weight) * #hits] / [(hit area weight * # hits) + (miss area weight * # misses) + (false alarm area weight * # false alarms)]

Where each area weight is the ratio of size of the (hit, miss, or false alarm) objects to the total area of all objects and # hits = number of matched objects; # misses = # unmatched observed objects; and # false alarms = # unmatched forecast objects.

Answers the question: How well did the forecast "yes" objects correspond to the observed "yes" objects?

Range: 0 to 1, 0 indicates no skill. Perfect score: 1.

Characteristics: Measures the area-weighted fraction of observed and/or forecast events that were correctly predicted. It can be thought of as the /accuracy/ when correct negatives have been removed from consideration, that is, /TS/ is only concerned with forecasts that count. Sensitive to hits, penalizes both misses and false alarms. Does not distinguish source of forecast error.
In a grid-based CSI each gridpoint that is counted in computing the CSI contributes represents an area with the same size but with MODE objects, the various objects can have a wide variety of sizes. Thus, area weighting makes sense. and observed objects.

Mean Intersection over Area

Ratio of intersection area to union area (unitless). Ranges from zero to one: One is perfect, smaller implies less overlap. This measure is the mean for all clusters of objects with interest values greater than 0.7.

Area Ratio

Ratio of the areas of two objects defined as the lesser of the forecast area divided by the observation area or its reciprocal (unitless). The ideal value is 1, since this means that the forecast and observed objects are exactly the same size. Smaller implies that the forecast was either too small or too large. This measure is the mean for all clusters of objects with interest values greater than 0.7.

Centroid Distance

Distance between two objects centroids (in grid units). Smaller is better, since this means the objects are closer. This measure is the mean for all clusters of objects with interest values greater than 0.7.

Angle Difference

Difference between the axis angles of two objects (in degrees). This is only meaningful if objects seem to be more linear than circular, e.g. lines of thunderstorms. When they are linear, this measure tells you how well the angle of the forecast line matches the angle of the observed line. Smaller differences are better. This measure is the mean for all clusters of objects with interest values greater than 0.7.

Intensity with confidence intervals

10th, 25th, 50th, 75th, and 90th percentiles of intensity of the filtered field within the object (various units). This tells you the distribution of values within an object (think of this as the numeric equivalent of a boxplot). There are no ideal values. However, if you compare the distribution of values within a forecast object and an observed object, you would like them to match up. I would check to see how close the median and 90^{th} percentile values are. This will tell you if you forecast is too intense or not intense enough. This measure is the mean for all clusters of objects with interest values greater than 0.7.
