SPRING FORECASTING EXPERIMENT 2013

Conducted by the

EXPERIMENTAL FORECAST PROGRAM

of the

NOAA/HAZARDOUS WEATHER TESTBED

http://hwt.nssl.noaa.gov/Spring_2013/

HWT Facility – National Weather Center
6 May - 7 June 2013

Program Overview and Operations Plan

Mike Coniglio[2], Israel Jirak1, Adam Clark[2,3], James Correia[1,3], Chris Melick[1,3],
David Imy[1], Dusty Wheatley[2]

(1) NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma
(2) NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma
(3) Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma

1. Introduction

Each spring, the Experimental Forecast Program (EFP) of the NOAA/Hazardous Weather Testbed (HWT), organized by the Storm Prediction Center (SPC) and National Severe Storms Laboratory (NSSL), conducts a collaborative experiment to test emerging concepts and technologies designed to improve the prediction of hazardous convective weather. The primary goals of the HWT are to accelerate the transfer of promising new tools from research to operations, to inspire new initiatives for operationally relevant research, and to identify and document sensitivities and the performance of state-of-the art convection allowing (3 to 4 km grid-spacing) experimental modeling systems (CAMs).

The 2013 Spring Forecasting Experiment (SFE2013), a cornerstone of the EFP, will be conducted 6 May – 7 June. Building upon successful experiments of previous years, a main emphasis of SFE2013 will be the generation of probabilistic forecasts of severe weather valid over shorter time periods than current operational products. This will be an important step toward addressing a strategy within the National Weather Service of providing nearly continuous probabilistic hazard forecasts on increasingly fine spatial and temporal scales. As in previous experiments, a suite of new and improved experimental mesoscale and CAM guidance will be central to the generation of these forecasts.

This operations plan summarizes the core interests of SFE2013 and provides information on the operations of the experiment. Detailed information on the organizational structure of the HWT and information on various forecast tools and diagnostics can also be found in this document. The remainder of the operations plan is organized as follows: Section 2 provides details on a number of new products being introduced during SFE2013 and Section 3 describes the core interests and new concepts being introduced for SFE2013. A list of daily participants, details on the SFE forecasting, and more general information on the HWT are found in appendices.

2. Description of Experimental Models and Products

In addition to incorporating newer model guidance that is operational or nearly operational in the National Centers for Environmental Prediction (NCEP) data stream, the daily forecast activities of SFE2013 are designed to incorporate more rapidly updated experimental model output into the forecast process. A primary tool of the SFE2013 forecasting activities will be a newly developed mesoscale ensemble produced by NSSL (the NSSL Mesoscale Ensemble, or NME) that uses an ensemble-based data assimilation technique to analyze the atmosphere in near-real time. The goal is to evaluate NME analysis products as a supplement or alternative to the Rapid Refresh Model (RAP)-based analysis that is widely used in SPC operations for forecasting severe convective weather. Output from the NME will be available to use as a product similar to how the SFCOA is used in operational forecasting. Likewise, an ensemble of forecasts initialized from select hourly NME analyses will be available to use in a manner similar to the operational NCEP Short-Range Ensemble Forecast (SREF) system.

There are several convection-allowing models new to the experiment this year, including a parallel "hot-start" NSSL WRF (see section 3b) and convection-allowing ensembles initialized at 12Z. Additionally, for the first time in the SFE, the United Kingdom Meteorological (UKMET) Office will provide two CAM forecasts over the CONUS that are based on their Unified Modeling System. The UKMET office's approach to modeling the lower atmospheric boundary layer is quite different than the approaches used in American CAM systems. We are eager to compare these two approaches in convective weather environments. More information on these modeling systems is given below.

a) *NSSL Mesoscale Ensemble (NME)*

A Weather Research and Forecast (WRF)-Advanced Research WRF core (ARW) (v3.4.1) mesoscale data assimilation system is run daily to produce three-dimensional analyses over a CONUS domain with 18-km horizontal gridpoint spacing (278x189) and 51 vertical levels. The 36-member ensemble is constructed from the initial and boundary conditions provided by the 1200 UTC ESRL Rapid Refresh version two (RAPv2) forecast cycle. Random samples of background error are generated by the WRF variational data assimilation (WRF-Var) algorithm and then added to each ensemble member, to account for uncertainties in the initial and boundary conditions of the reference analysis (Torn et al. 2006). The WRF physics options are also varied amongst the ensemble members to address deficiencies in model physics. This diversity is introduced into the physics categories as follows:

- Microphysics: Thompson
- Radiation: RRTM(LW)/Dudhia (SW), RRTMG (LW)/RRTMG (SW), Goddard (LW)/Goddard (SW)
- Noah land-surface model
- Planetary Boundary Layer: YSU, MYJ, MYNN, ACM2
- Convection: Kain-Fritsch, Grell-Devenyi, Tiedtke

Routinely available observations (of altimeter setting, temperature, dewpoint, and horizontal wind components) from land and marine surface stations, rawinsondes, and aircraft—as well as satellite winds—are assimilated using an ensemble Kalman Filter (EnKF) (using the WRF-DART software) at hourly intervals from 1300 UTC to 0300 UTC the following day. At 1400, 1600 and 1800 UTC, the resultant EnKF analyses are used to launch a full ensemble of forecasts, which are run out to 0300 UTC of the following day.

b) NSSL WRF

SPC forecasters have used output from an experimental 4 km WRF-ARW produced by NSSL since the fall of 2006. Currently this WRF model is run twice daily at 00 UTC and 12 UTC throughout the year over a full CONUS domain with forecasts to 36 hours. Output is also available online at http://www.nssl.noaa.gov/wrf/ and http://www.nssl.noaa.gov/wrf/12Z/.

New to the experimental numerical guidance for this year's experiment is a parallel or "hot-start" version of the NSSL-WRF that is initialized from the "best member" of the 0000 UTC NME analysis. The best member is the member with the lowest normalized RMS difference of temperature and horizontal wind components using all 0000 UTC observations. The hot start run is configured identically to the operational NSSL-WRF run so that the impact of the NME analyses in initializing the forecasts can be evaluated. Specifically, both runs use WRF version 3.4.1, NAM forecasts at 3 hourly intervals for lateral boundary conditions, WSM6 microphysics parameterization, and MYJ turbulent-mixing (PBL) parameterization. The hot start runs are available online at http://www.nssl.noaa.gov/wrf/HS/. For comparing the two NSSL-WRF runs, a recently developed interactive web display utilizing Google-maps-like features and GIS will be used. The web display allows zooming, overlaying of chosen fields, and side-by-side comparisons of model and observational fields. An example display is pictured below:

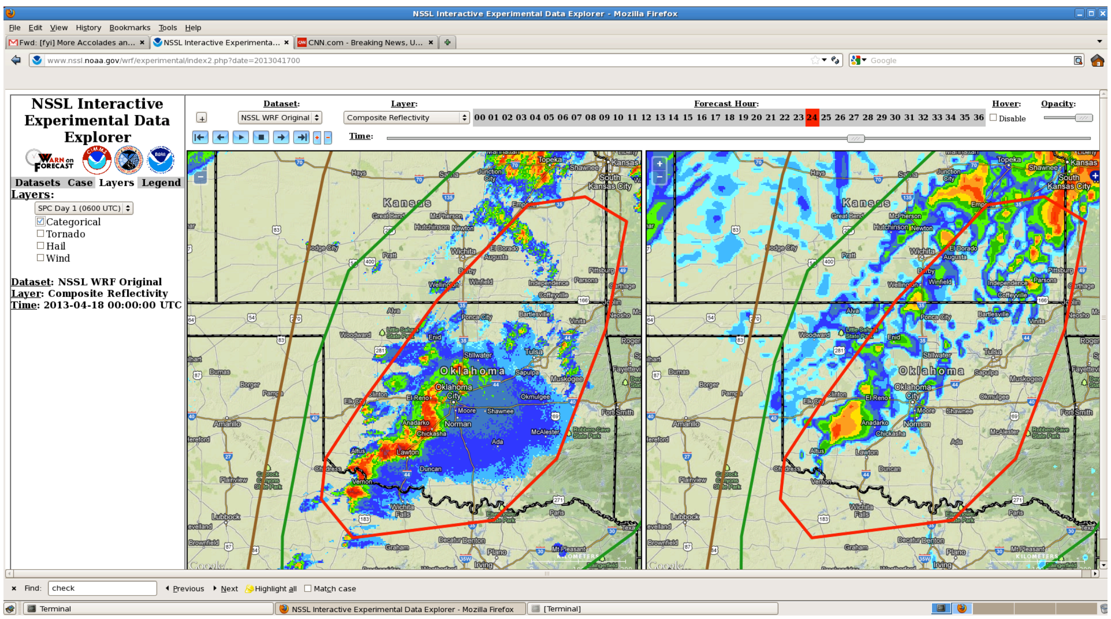

Figure 1. Example display from the NSSL Interactive Experimental Data Explorer (publicly available at http://www.nssl.noaa.gov/wrf/experimental/). In the left panel, observed composite reflectivity at 0000 UTC 18 April 2013 is shown with the SPC categorical severe weather outlooks overlaid. In the right panel, the corresponding 24 h forecast of composite reflectivity from the operational NSSL-WRF is shown.

c) CAPS Storm Scale Ensemble Forecast (SSEF) System

As in previous years, the University of Oklahoma Center for Analysis and Prediction of Storms (CAPS) will provide a 00Z-initialized 4-km grid-spacing Storm Scale Ensemble Forecast (SSEF) system. This year's 00Z SSEF system has 25 members and is run at the University of Tennessee National Institute for Computational Sciences (NICS). The 25-members of the ensemble that will be evaluated in the SFE use the WRF-ARW core and all forecasts use 51 vertical levels. WRF code was modified by CAPS to allow initial hydrometeor fields generated from a 3DVAR/ARPS Cloud analysis of WSR-88D radar reflectivity to pass into WRF initial conditions, and to write out a reflectivity field every 5 min. As in the 2012 season, the 00Z NAM analyses available on the 12 km grid (AWIPS 218) are used for initialization of the control and non-perturbed members, and as a first guess for initialization of perturbed members with the initial condition perturbations coming directly from the operational SREF system. WSR-88D data, along with available surface and upper air observations, are analyzed using ARPS 3DVAR/Cloud-analysis system. Forecast output at hourly intervals (with higher time frequency output for a limited selection of 2D fields) are archived at the NICS mass storage system (HPSS). Specifications for all members are provided in Table 1.

The basic strategy in constructing the SSEF system is to have a set of members accounting for as many error sources as possible that can be used to generate reliable forecast probabilities (non-shaded members in Table 1). These "core" members have IC/LBC perturbations as well as varied physics and model cores. Other sets of members were configured to allow for various sensitivity experiments (shaded members in Table 1). Six members are configured identically except for their microphysics parameterizations and four members are configured identically except for their turbulent-mixing (PBL) parameterizations.

This year, a SSEF system initialized at 12Z will be available for use in the forecasting activities. The basic strategy in constructing the 12Z SSEF system is to have a set of members that have the same configuration as in the 00Z ensemble. Resources for running the 12Z members in real time are more limited than those for the 00Z ensemble, so only 7 members from the 00Z ensemble are run at 12Z (identified by members with an asterisk in Table 1). This will allow for a direct comparison of the change in skill between this set of seven members for forecasts initialized 12 hours apart. Furthermore, the selection of seven members was chosen to match the number of members in the SSEO (see below) so that a comparison of the spread and skill characteristics of these sets of forecasts can be compared fairly.

Table 1. Configuration of the CAPS SSEF system. NAMa and NAMf refer to 12 km NAM analysis and forecast, respectively. ARPSa refers to ARPS 3DVAR and cloud analysis. Members with gray shading indicate physics-only perturbations.

Member	IC	BC	Radar data	Microphy	LSM	PBL
arw_cn*	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	MYJ
arw_c0 (18h)	00Z ARPSa	00Z NAMf	no	Thompson	Noah	MYJ
arw_m3	arw_cn + em-p1_pert	21Z SREF em-p1	yes	Morrison	RUC	YSU
arw_m4*	arw_cn + nmm-n2_pert	21Z SREF nmm-n2	yes	Morrison	Noah	MYJ
arw_m5*	arw_cn + em-n2_pert	21Z SREF em-n2	yes	Thompson	Noah	ACM2

arw_m6	arw_cn + nmmb-p2_pert	21Z SREF nmmb-p2	yes	M-Y	RUC	ACM2
arw_m7*	arw_cn + nmm-p1_pert	21Z SREF nmm-p1	yes	Morrison	Noah	MYNN
arw_m8	arw_cn + nmmb-n1_pert	21Z SREF nmmb-n1	yes	WDM6	RUC	MYJ
arw_m9	arw_cn – nmmb-p1_pert	21Z SREF nmmb-p1	yes	M-Y	Noah	YSU
arw_m10	arw_cn + em-n1_pert	21Z SREF em-n1	yes	WDM6	Noah	QNSE
arw_m11	arw_cn – em-p2_pert	21Z SREF em-p2	yes	M-Y	Noah	MYNN
arw_m12*	arw_cn – nmmb-n3_pert	21Z SREF nmmb-n3	yes	WDM6	Noah	YSU
arw_m13*	arw_cn – nmmb-p3_pert	21Z SREF nmmb-p3	yes	Thompson	Noah	YSU
arw_m14*	arw_cn – em-p3_pert	21Z SREF em-p3	yes	Thompson	Noah	MYNN
arw_m15	arw_cn – nmm-p2_pert	21Z SREF nmm-p2	yes	Morrison	Noah	QNSE
arw_m16	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	MYNNe
arw_m17	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	ACM2
arw_m18	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	YSU
arw_m19	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	QNSE
arw_m20	00Z ARPSa	00Z NAMf	yes	M-Y	Noah	MYJ
arw_m21	00Z ARPSa	00Z NAMf	yes	Morrison	Noah	MYJ
arw_m22	00Z ARPSa	00Z NAMf	yes	WDM6	Noah	MYJ
arw_m23	00Z ARPSa	00Z NAMf	yes	Zed's	Noah	MYJ
arw_m24	00Z ARPSa	00Z NAMf	yes	Thompson	RUC	MYNN
arw_m25	00Z ARPSa	00Z NAMf	yes	Thompson +mod RRTMG	Noah	MYJ

Note 1: For all members: *ra_lw_physics*= RRTMG; *ra_sw_physics*=RRTMG; *cu_physics*=none
Note 2: m16 uses MYNN modified from the official ARW 3.4.1 release.
Note 3: m25 uses Thompson modified RRTMG for both sw and lw radiations.

d) SPC Storm Scale Ensemble of Opportunity (SSEO)

The SSEO is a 7-member, convection-allowing ensemble consisting of deterministic models available operationally to SPC. This "poor man's ensemble" provides a practical alternative to a formal/operational storm-scale ensemble which will not be available in the near-term because of computational/budget limitations. Similar to the SSEF system, hourly maximum storm-attribute fields, such as simulated reflectivity, updraft helicity, and 10-m wind speed are produced from the SSEO. Member specifications are provided in Table 2. Members marked with "-12h" in the Model column are 12h time-lagged members, initialized 12h earlier than the other members. All members are initialized with a "cold-start" from the operational NAM – i.e., no radar data assimilation or cloud model is used to produce ICs.

Table 2. SSEO member specifications.

Member #	Model	Grid-spacing	Agency	PBL	Microphysics	LSM
sseo01	WRF-ARW	4-km	NSSL	MYJ	WSM6	Noah
sseo02	Hi-Res Window WRF-ARW	5.15-km	NCEP/EMC	YSU	WSM3	Noah
sseo03	Hi-Res Window WRF-ARW -12h	5.15-km	NCEP/EMC	YSU	WSM3	Noah
sseo04	CONUS WRF-NMM	4-km	NCEP/EMC	MYJ	Ferrier	Noah
sseo05	Hi-Res Window WRF-NMM	4-km	NCEP/EMC	MYJ	Ferrier	Noah
sseo06	Hi-Res Window WRF-NMM -12h	4-km	NCEP/EMC	MYJ	Ferrier	Noah
sseo07	NMMB Nest	4-km	NCEP/EMC	MYJ	Ferrier+	Noah

e) Air Force Weather Agency (AFWA) 4-km ensemble

The U.S. Air Force Weather Agency (AFWA) has recently implemented a real-time 10-member 4-km WRF-ARW ensemble. Forecasts are initialized at 0000 UTC and 1200 UTC using 6 or 12 hour forecasts from four global models, the UKMET Unified Model, the NCEP Global Forecast System (GFS), the Canadian Meteorological Center Global Environmental Multiscale (GEM) Model, and the U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS). Diversity in the AFWA ensemble is achieved through IC/LBCs from the different global models and varied microphysics and boundary layer parameterizations. SPC is currently ingesting the AFWA grids in their real-time data feed and these forecasts will be available for examination during SFE2013.

Table 3. AFWA ensemble member specifications.

Member #	ICs/LBCs	LSM	Micro-physics	PBL
afwa01	18Z UKMET	Noah	WSM5	YSU
afwa02	18Z GFS	RUC	Goddard	MYJ
afwa03	12Z GEM	Noah	Ferrier	QNSE
afwa04	12Z NOGAPS	Noah	Thompson	MYJ
afwa05	18Z UKMET	RUC	Thompson	MYJ
afwa06	18Z GFS	Noah	Thompson	QNSE
afwa07	12Z GEM	Noah	Goddard	YSU
afwa08	12Z NOGAPS	Noah	WSM5	QNSE
afwa09	18Z UKMET	RUC	Ferrier	QNSE
afwa10	18Z GFS	Noah	WSM5	YSU

f) UK-Met Office convection allowing models

The Unified Model (UM) is the name given to the suite of numerical modeling software used by the UK Met Office. Two fully operational, nested limited-area high-resolution versions of the UM (4.4 km and 2.2 km horizontal resolution) running once per day will be supplied to the 2013 Spring Forecasting Experiment.

The 4.4 km version has 70 vertical levels (spaced between 5m and 40 km) across a domain covering the CONUS and adjacent areas. Taking its initial and lateral boundary conditions from the 00z 25 km horizontal resolution global configuration of the UM, the 4.4 km model initializes without data assimilation and runs out to T+48. This model configuration uses a CAPE limited closure shallow convection parameterization scheme, 2D Smagorinsky boundary layer mixing scheme and single moment microphysics.

Meanwhile, the 2.2 km horizontal resolution version of the UM is nested with the 4.4 km model and runs over a slightly sub-CONUS domain. The 2.2 km model takes its initial and lateral boundary conditions from the T+3 step of the 00z 4.4 km run, thus reducing spin-up time within the 2.2 km model, and runs out to T+45. As with the 4.4 km model, the 2.2 km model initializes without data assimilation and uses the same 70 vertical level spacings as the 4.4 km. The 2.2 km model has identical planetary boundary layer and microphysics schemes as the 4.4 km model but does not have the constrained convection scheme of the 4.4 km and instead has no convection parameterization.

Finally, parallel versions of both the 4.4 km and 2.2 km models are also being run with 3D Smagorinsky boundary layer mixing schemes.

g) ESRL High Resolution Rapid Refresh (HRRR) model

We will continue to include output from the HRRR model developed by the NOAA/Earth Systems Research Laboratory (ESRL). The experimental 3-km grid-spacing HRRR model is nested within the hourly development version of the 13 km RAP model, which provides ICs/LBCs for the HRRR. The HRRR uses a version of the WRF-ARW. A unique aspect of the RAP is the hourly Gridpoint Statistical Interpolation (GSI) data assimilation system that incorporates a wide array of observational datasets including radar reflectivity via the radar-Diabatic Digital Filter Initialization. The HRRR integration is run hourly over a full CONUS domain with forecasts to 15 hours.

In early 2013, ESRL introduced direct 3-km data assimilation using a 1-hour old 13-km ESRL RAP post-DFI (digital filter initialization) analysis as a first guess, followed by assimilation of radar reflectivity-derived latent heating at 15min intervals through a 1-hour period before applying a complete GSI 3D-variational analysis using most observations at 3 km, before applying a non-variational cloud and precipitation hydrometeor analysis. ESRL expects that the change will produce a very noticeable increase in smaller storm-scale structures in the HRRR 00-hr analysis and will greatly impact the forecasts in the first several hours based on their preliminary analysis.

3. SFE2013 Core Interests/Daily Activities

Unlike the previous few experimentss, in which NSSL led a component focused specifically on convective initiation (CI), all of the SFE2013 activities will concentrate on forecasting severe convective weather (but forecasts of CI will be an increasingly important part of the severe weather forecasts as the valid time window of the forecasts decreases, as described later). Two separate teams will generate identical forecast products from the same set of forecast guidance. This design will foster a healthy competition between the two teams, from which we can explore the utility of experimental guidance in a simulated operational severe weather forecasting environment. For the first time forecasts generated by the SFE participants will be used in real time to support decisions made for experimental warning operations during this period.

a. Forecast products and activities

Similar to previous years, the forecasts this year will continue to explore our ability to add temporal specificity to longer-term convective outlooks. The forecasts will be the probability of any severe storm (large hail, damaging wind, and/or tornado) within 25 miles (40 km) of a point ("total severe"), as defined in the SPC operational convective outlooks. We will also predict areas of significant hail and wind (10% or greater probability of hail ≥ 2" in diameter or wind gusts ≥ 65 kt) and if time permits, areas with an enhanced risk of tornadoes. The forecast teams will first create a full-period (16-12Z) total severe outlook (where SPC forecasters have historically shown considerable skill) and then manually stratify that outlook into three periods with higher temporal resolution: 18-21Z, 21-00Z, and 00-03Z. A text product will accompany each daily experimental forecast package that describes the meteorology of the day and the usefulness of the suite of model guidance during the creation of the severe storm forecasts.

During SFE2012, calibrated severe guidance from the storm-scale ensemble of opportunity (SSEO) was used to temporally disaggregate a 16-12Z period human forecast. A scaling factor was formulated by matching the full-period calibrated severe SSEO guidance to the human forecast, then this scaling factor (unique at every grid point) was applied to the SSEO calibrated severe guidance for each individual period, and finally consistency checks were conducted to arrive at the final temporally disaggregated forecasts. These automated forecasts from SFE2012 fared favorably both in terms of objective metrics (e.g., CSI, FSS) and subjective impressions when compared to manually drawn forecasts. Given the encouraging results from SFE2012, the same technique will be applied to forecasts during SFE2013. The 16-12Z human forecasts for each team will be temporally disaggregated into the 3-h periods to provide a first guess for the three forecast periods (18-21Z, 21-00Z and 00-03Z).

The generation of three separate afternoon and evening forecasts valid over 3-h periods is new to the SFE activities. The first set of three forecasts will be issued by 12pm local time. Furthermore, two of the three afternoon and evening forecasts will be updated two times throughout the day, which has not been attempted before in the SFE. In the afternoon, both teams will then use all available observational and numerical model guidance to update the last two forecast periods (21-00Z and 00-03Z) twice. The first afternoon update will be issued at 2:30pm and the second afternoon update will be issued at 4:00pm.

The forecasting activities are designed to explore the process of generating probabilistic forecasts over short time windows as an alternative approach to the more event-driven and categorical approach of issuing Mesoscale Discussions and Convective Watches. As the use of model guidance continues to expand in the generation of first guess forecast products, it is important to explore both the skill of the model guidance at these time scales and the appropriate role of the human in the process of generating nearly continuous probabilistic hazard forecasts on cascading temporal and spatial scales.

Given the climatological preference for CI in the early to mid afternoon and the expansion of severe convective weather in the late afternoon and early evening, it is likely that the forecasts during the first or second periods will focus on issues related to CI, and probabilities that might indicate a Mesoscale Discussion is needed, whereas the second and third periods will be related more to the transition to and continuation of severe weather, and probabilities that might indicate Convective Watches need to be continued. The goal is to determine the skill of our current suite of mesoscale and CAM guidance, the skill of human forecasts, as well as to explore the process of forecasting CI and the continuation of severe convection with probabilistic forecasts over short (3 h) time windows.

To explore the process of sharing digitized forecasts of severe convection over 3 h periods with warning operations, the SFE will be sharing all four of the human generated forecasts from the two teams with the HWT Experimental Warning Program (EWP), which they will use in their preparation for their operations. The SFE and EWP Experiments will overlap for three weeks (6 – 24 May). During these weeks, collaboration activities will occur during one of three periods depending on the timing of convection on that day, either at 12pm, at 2:30pm, or shortly before 4pm, in which the SFE participants will be responsible for briefing the EWP group as to the reasoning of the human forecasts. This is the first such direct interaction between the two forecast and warning components of the HWT and is an early manifestation of the goal of providing probabilistic hazard forecasts on multiple scales from the synoptic scale to the storm scale.

b. Forecast and Model Evaluations

i) Subjective Evaluation of Experimental Forecasts with EWP Products

In the next day evaluations, the individual period team forecasts and the temporally-disaggregated first-guess forecasts will be compared to observed radar reflectivity, reports of severe weather, NWS warnings, experimental warnings produced by the Experimental Warning Program (if possible), and radar-estimated hail sizes and rotation tracks over the same time periods. The SFE participants will provide their subjective evaluations of the strengths and weaknesses of each of the forecasts. The motivation here is to determine the relative skill of the first-guess guidance and the human-generated forecasts over all periods. When possible, the EWP participants will provide feedback on the SFE afternoon total severe probability forecasts as they relate to their evening warning operations.

ii) Objective Evaluation of Experimental Forecasts

Similar to the HWT 2012 Spring Forecast Experiment (SFE), objective forecast verification of the human forecasts will resume in near real-time for the 2013 SFE. Once again, experimental forecasts of total severe thunderstorm probabilities will be evaluated using Critical Success Index (CSI) and Fractions Skill Score (FSS). The utility of verification metrics in assessing forecast skill will continue to be explored by comparing the scores to the subjective impressions of the participants. However, instead of just using one probability threshold for CSI like last year, the maximum CSI value will be calculated as well as a few fixed (5%, 15%, 30%) thresholds used in SPC operational outlooks. In addition, the relative skill score (Hitchens and Brooks 2012, Hitchens et al. 2013) will be introduced to gauge the performance of the experimental forecasts against a baseline reference, namely the practically perfect hindcasts (Brooks et al. 1998).

iii) Evaluation of Deterministic Convection-Allowing Models and Physics Sensitivities

Standard fields from convection-allowing model output (e.g. simulated reflectivity and updraft helicity) and experimental simulated storm attributes will be generated from select ensemble members, the NSSL WRF runs, and the UKMET runs. The focus of these evaluations will be the general skill of the forecasts in predicting severe convection explicitly, as well as the impact of various physics options on the forecasts. These comparisons have shown to be valuable in identifying systematic model deficiencies that can impact convective forecasts significantly (Clark et al. 2012, Coniglio et al. 2013).

iv) Evaluation of 00Z and 12Z Convection-Allowing Ensembles

Although only the SSEO will provide calibrated probabilistic guidance, a variety of ensemble output (spaghetti, exceedance probability, maximum from any member) will be available from all three convection-allowing ensembles (SSEO, CAPS SSEF, and AFWA) based on hourly maximum fields (HMFs) of simulated storm attributes such as 1-km AGL simulated reflectivity, updraft speed, updraft helicity, and 10-m winds will be created. For the first time, 12Z-initialized forecasts will be available for all three convection-allowing ensembles. Subjective and objective evaluations will be performed comparing each ensemble's 00Z forecasts to the 12Z forecasts and comparing the ensembles to one another.

v.) NME compared to SPC Mesoscale Analysis

Evaluation of output from the NSSL mesoscale ensemble will focus on parameters that relate to mesoscale and synoptic-scale environments favorable for severe convective weather and will be made against the ESRL-RAPv2-based version of the mesoscale analysis currently used heavily in SPC operations. The focus will be on how well the 2-m temperature and dewpoint analyses fit the observations and how much these analyses differ from their 1 h prior forecast valid at the same time, with many other environmental fields also available for comparison, including many of the fields available routinely on the SPC mesoscale analysis web page (http://www.spc.noaa.gov/exper/mesoanalysis/).

c. Daily Activities Schedule

Scheduled activities are in local (CDT) time and conducted as one large group unless otherwise indicated. Two separate groups will be generating identical forecast products.

Pre-0800: "Teaser". Because we will not immediately begin evaluating the previous day's forecast, relevant loops (radar, water vapor, visible imagery, storm reports, etc.) will be displayed as participants arrive so they can get a quick look at how the previous' day's forecasts verified.

0800 – 0930: Full-period forecast. Begin activities with hand analyses of 1200 UTC upper-air data and surface charts. Then, large-scale overview and group forecast discussion with consensus selection of a forecast domain. Break into two forecast groups and issue probabilistic forecasts of total severe valid 1600 UTC to 1200 UTC the next day. Products will include probabilities of total severe.

0930 – 0945: Break

0945 – 1015: Evaluation of previous day's human forecasts. Each forecast will be subjectively rated. Each group will rate the forecasts generated by the other group. Also, it will be decided whether the updates continuously improved the forecasts.

1015 – 1100: Model evaluations. Participants will split into two groups. Group 1 will perform evaluations comparing the 0000 UTC initialized storm-scale ensembles to their 1200 UTC initialized counterparts (SSEO, AFWA, and SSEF systems). Group 1 will also compare analyses generated from the NSSL Mesoscale Ensemble (NME) to those generated from the ESRL RAPv2-based SFC-Objective Analyses (SFCOA). Group 2 will examine the impact of microphysics schemes by comparing forecasts from the 5 SSEF system members that differ only by their microphysics parameterizations. Emphases will be placed on comparing two versions of the Thompson scheme as well as the new NSSL double-moment scheme. Group 2 will also conduct comparisons of the operational NSSL-WRF to a parallel version initialized from the 0000 UTC NME analysis using a Google-maps-based interactive comparison interface. Comparisons will also be made to the UKMET's convection-allowing model.

1100 – 1200: Update forecast #1 –Both groups will use 1400 UTC initialized NME forecasts and all other available observations and guidance to issue forecasts for the 18-21, 21-00, and 00-03Z time periods. A first guess for each time period will be generated using temporal disaggregation applied to the full-period forecast issued earlier in the morning. The same products as from the initial forecast will be issued (i.e., probabilities of total and significant severe).

1200 – 1300: Lunch and possible collaboration with the EWP.

1300 – 1330: Weather Briefing – Highlights from yesterday, general overview, discussion of forecast challenges and products. In addition, each group will discuss reasoning for their forecasts.

1330 – 1430: Update forecast #2 – Same as #1, except for just the 21-00 and 00-03Z periods. The 1600 UTC initialized NME and 1200 UTC initialized convection-allowing ensembles will be available.

1430 – 1445: Break and possible collaboration with the EWP.

1445 – 1500: Open time period for discussion and questions of the day.

1500 – 1600: Update forecast #3 – Same as #2. The 1800 UTC initialized NME will be available and possible collaboration with the EWP.

d. Other specialized activities

i) Automated simulated supercell identification using the Method for Object-Based Diagnostic Evaluation Time-Domain (MODE-TD)

The concept of “object-based” forecast verification has become increasingly popular for verifying high-resolution forecasts because, unlike traditional methods, object-based methods provide meaningful diagnostic information on forecast errors like displacement, orientation, and intensity, and they are designed to mimic subjective evaluation approaches. One such object-based verification approach is known as the Method for Object-based Diagnostic Evaluation (MODE; Davis et al. 2006; Davis et al. 2009), which is part of the Developmental Testbed Center’s Model Evaluation Tools (MET; current version available online at http://www.dtcenter.org/met/users/downloads/). In the MODE framework, objects are defined within a spatial field after application of smoothing and thresholding. Then, fuzzy-logic-based algorithms can be applied to match and/or merge objects in the forecast and observed fields.

Until recently, object-based approaches have only considered two-dimensional spatial objects. However, efforts are ongoing to incorporate the time dimension into an extension of MODE known as MODE time-domain (hereafter, MODE-TD) that will eventually become part of DTC’s MET software package (Bullock 2011). In MODE-TD, contiguous regions of grid-points exceeding a specified threshold encompassing both space and time are referred to as time-domain objects. The addition of the time dimension results in a much more powerful diagnostic tool that can provide important information on aspects of phenomena over their entire life cycles (Clark et al. 2012a, Clark et al. 2013). These aspects include longevity, time of initiation and dissipation, translation speed, and evolution (e.g., growth, decay, changes in maximum intensity, etc.), all of which would be difficult to diagnose by considering the spatial dimension alone.

In addition to being a powerful verification tool, MODE-TD also has potential utility as a forecasting tool, following the concept of “feature-specific prediction” proposed by Carley et al. (2011). Feature-specific prediction involves identifying features of potential interest in forecast fields and presenting them as guidance. Carley et al. (2011) summarize feature specific prediction by three simple steps: 1) determination of the feature type of interest, 2) feature identification and tracking, and 3) presentation to forecaster for evaluation. Following this framework, MODE-TD will be applied to the SSEF system to identify and track simulated supercells, and then display this information to forecasters in an efficient and meaningful way. For tracking simulated supercells, we will output model grids to a file at hourly intervals during model integration the grid-coordinates, values, and times (at 5-minute intervals) that a particular field exceeded a specified threshold. With this file, the field can be “reconstructed” at sub-hourly intervals for visualization and calculation of diagnostics without considering all the grid-points that did not exceed the specified threshold, significantly reducing data volume. For the purpose of supercell identification, the maximum value of updraft helicity over all model time-steps within 5-minute periods will be extracted at all grid-points at which UH ≥ 20 m^2s^{-2}. MODE-TD will then be applied to the 5-minute UH grids to identify supercells. Through some initial experimentation the following criteria will be used to identify simulated supercells in the SSEF members: 1) time-domain UH objects are identified using a minimum threshold of 25 m^2s^{-2}, 2) at least one grid-point within each UH object must have UH ≥ 75 m^2s^{-2}, and 3) the UH object must last at least 1 h. The plot below illustrates supercell tracks identified using the aforementioned criteria for a NSSL-WRF forecast initialized 17 April 2013.

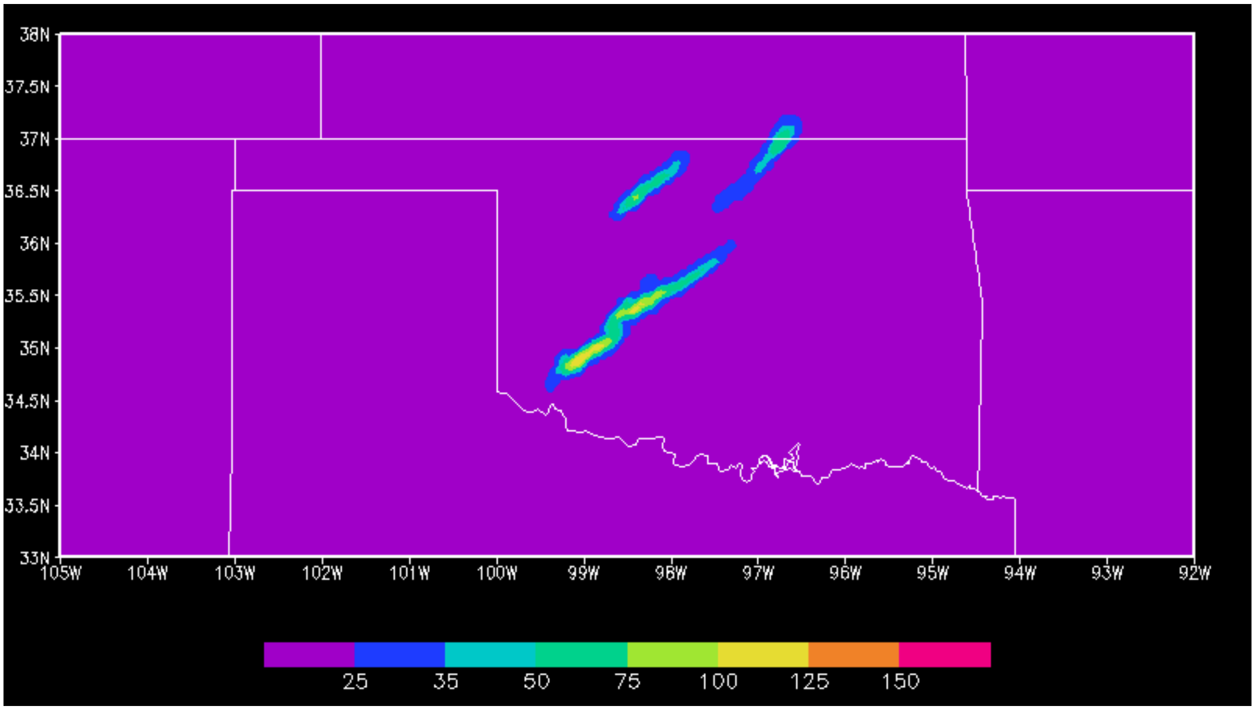

Figure 2. Simulated supercell tracks for NSSL-WRF forecasts initialized 17 April 2013. The shading shows the maximum updraft helicity (m^2s^{-2}) over the lifetime of the simulated supercells.

ii) Model Storm Reports

It is a challenge to verify the severe weather produced by convection allowing model forecasts because the grid spacing is too coarse to predict explicitly most instances of severe phenomenon (particularly large hail and tornadoes). Therefore an effort is underway to explore proxies for severe weather that can be extracted from output on the model grid points. An algorithm to generate probabilities of model proxies for severe weather including tornadoes, hail, and wind will be evaluated.

The algorithm to extract these model storm reports is based on an object algorithm. The algorithm uses a minimum threshold value of the relevant variable to first search for potential objects. The candidate objects must meet an additional three criteria: 1) They must contain at least 4 pixels of the minimum threshold; 2) They must contain at least one pixel of an upper threshold. This algorithm is run multiple times for hourly maximum updraft helicity at four sets of thresholds. The purpose is to identify maxima within individual objects. An example is shown in Figure 3. Hits identified by these algorithm/threshold combinations are then flagged as model storm reports.

For these model storm reports we will generate spatial plots that can be compared to observed storm reports. These plots will be generated for each of the four forecast periods (16-12Z, 18-21Z, 21-00Z, 00-03Z). An example plot of model storm reports (smoothed with a kernel density estimation) overlayed with observed storm reports is shown in Figure 4.

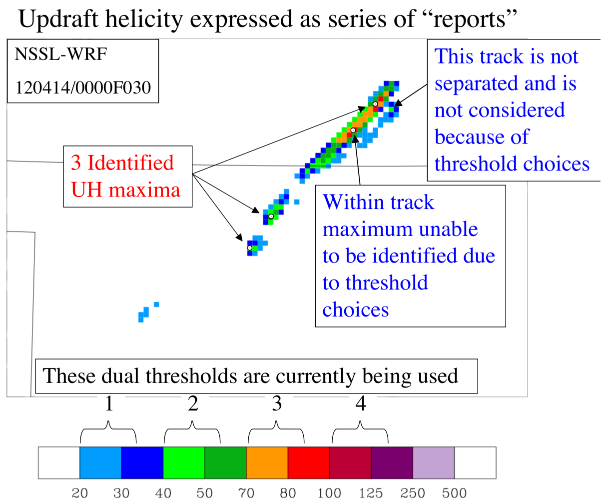

Figure 3. Example UH track annotated with model storm reports identified by the algorithm. Failure modes of the algorithm are also explained.

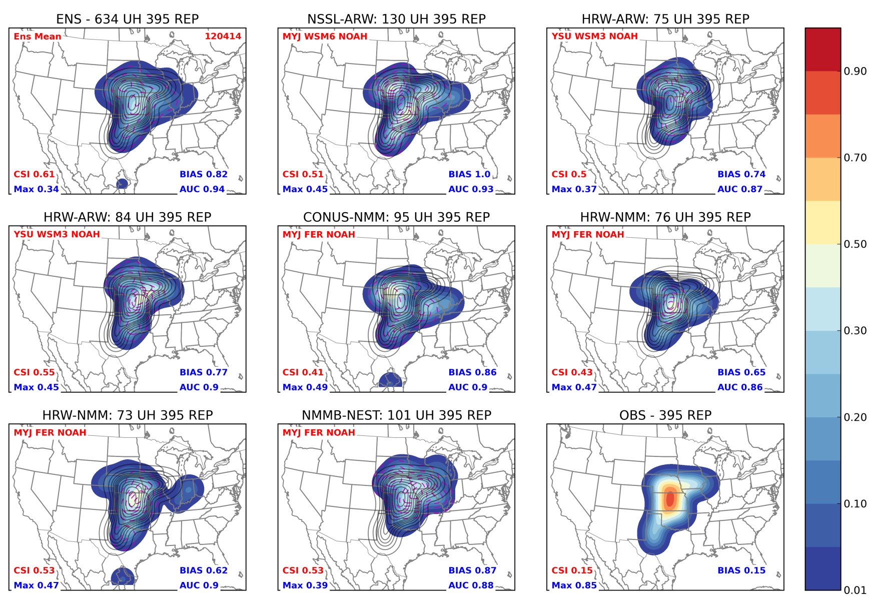

Figure 4. Smoothed probability of model storm reports (using a UH threshold of 75 m^2s^{-2}) (color fill with dashed red contours) and smoothed observed storm reports (color fill in lower right panel, gray contours in other panels) for the 24 h period starting at 1200 UTC 14 April 2012. The seven SSEO members and their mean (upper left panel) are shown.

Appendix A: List of scheduled SFE2013 participants. Facilitators/leaders include: Adam Clark (NSSL), Israel Jirak (SPC), Michael Coniglio (NSSL), James Correia (SPC), Chris Melick (SPC), Kent Knopfmeier (NSSL), Dave Imy (SPC), Steve Willington (UKMET), Dan Suri (UKMET).

6-10 May	13-17 May	20-24 May	28 May – 31 May	3-7 June
John Brown	Lance Bosart	Bill Gallus (T-Fr)	Eric James	Johnathon Wilkinson
Dusty Wheatley	Corey Guastini	Yvette Richardson & student	Thomas Jones	Curtis Alexander
Ted Mansell	Kyle Meier	Clark Evans	Xuguang Wang	Don Giuliano
Steven Cavallo	Noah Lock	Brock Burghardt	Aaron Johnson	Corey Potvin
Louie Grasso	Pat Market	Jason Otkin	Jeff Duda	Ashley Griffin
	Chad Shafer	Fred Carr	Brian Kolts	Jeff Hamilton
	Nusrat Yussouf	Heather Rombough	Conrad Ziegler	Dave Turner
		Helge Tuschy		

Appendix B: Experimental Severe Thunderstorm Forecasts

Probabilistic severe weather forecasts for the 16-12Z outlook period will be issued in the morning **by 15Z**. Sub-period forecasts will then be issued **by 17Z** for the following three-hour periods: 18-21Z, 21-00Z, and 00-03Z. Two additional updates to the latter two forecast periods (i.e., 21-00Z and 00-03Z) should be submitted **by 1930Z and 21Z**.

The severe weather forecast graphics will be similar in format to operational SPC outlooks; except only total severe storm probability contours will be formulated (categorical and general thunderstorm outlooks will not be made). The same probability contours used in the operational outlooks will be used for the severe forecasts (5, 15, 30, 45, and 60 %) **with the option of adding contour lines (every 5%) for localized maxima**. 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.

Day 2 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 C. Organizational structure of 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 are situated between the operations rooms of the SPC and OUN. The proximity to operational facilities, and access to data and workstations replicating those used operationally within the SPC, creates a unique environment supporting collaboration between researchers and operational forecasters on topics of mutual interest.

The HWT organizational structure is composed of three overlapping programs (Fig. 1). The Experimental Forecast Program (EFP) is focused on predicting hazardous mesoscale weather events on time scales ranging from 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. For more information see http://www.nssl.noaa.gov/projects/hwt/efp/.

The Experimental Warning Program (EWP) 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. For more information about the EWP see http://www.nssl.noaa.gov/projects/hwt/ewp/. A key NWS strategic goal is to extend warning lead times through the "Warn-on-Forecast" concept (Stensrud et al. 2009), which involves using frequently updated short-range forecasts (≤ 1h lead time) from convection-resolving ensembles. This provides a natural overlap between the EFP and EWP activities.

The GOES-R Proving Ground (established in 2009) exists to provide pre-operational demonstration of new and innovative products as well as the capabilities available on the next generation GOES-R satellite. The overall goal of the Proving Ground is to provide day-1 readiness once GOES-R launches in late 2015. The PG interacts closely with both product developers and NWS forecasters. More information about GOES-R Proving Ground is found at http://cimss.ssec.wisc.edu/goes_r/proving-ground.html.

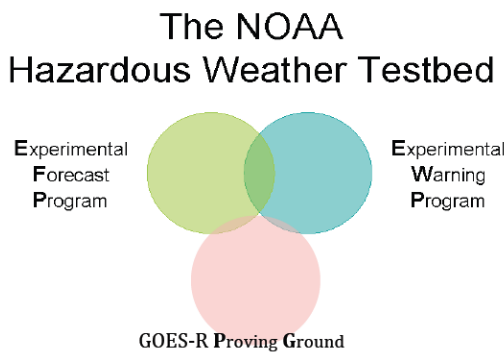

Figure 5: The umbrella of the NOAA Hazardous Weather Testbed (HWT) encompasses two program areas: The Experimental Forecast Program (EFP), the Experimental Warning Program (EWP), and the GOES-R Proving Ground (GOES-R).

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/. Detailed historical background about the EFP Spring Experiments, including scientific and operational motivation for the intensive examination of high resolution NWP model applications for convective weather forecasting, and the unique collaborative interactions that occur within the HWT between the research and operational communities, are found in Weiss et al. (2010 – see http://www.spc.noaa.gov/publications/weiss/hwt-2010.pdf) and Clark et al (2012b).

Appendix D: References

Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. Preprints, *19th Conf. on Severe Local Storms,* Minneapolis, MN, Amer. Meteor. Soc., 552–555.

Bullock, R., 2011: Development and implementation of MODE time domain object-based verification. *24th Conf. on Wea. Forecasting/20th Conf. on Num. Wea.* Pred., Seattle, WA, Amer. Meteor. Soc., 96.

Carley, J. R. B. Schwedler, M. E. Baldwin, R. J. Trapp, J. Kwiatkowski, J. Logsdon, and S. J. Weiss, 2011: A proposed model-based methodology for feature-specific prediction for high-impact weather. *Wea. Forecasting*, **26**, 243-249.

Clark, A. J., J. S. Kain, P. T. Marsh, J. Correia, Jr., M. Xue, and F. Kong, 2012a: Forecasting tornado path lengths using a 3-dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090-1113.

Clark, A. J., and Coauthors, 2012b: An Overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Clark, A.J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia, Jr., M. Xue, and F. Kong, 2013: Tornado path length forecasts from 2010-2011 using ensemble updraft helicity. *Wea. Forecasting*, (In Press).

Coniglio, M.C., J. Correia Jr., P. T. Marsh, F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, (In Press).

Davis, C. A., B. Brown, and R. Bullock, 2006: Object-Based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

Davis, C. A., B. Brown, R. Bullock, and J. Halley-Gotway, 2009: The Method for Object-Based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Wea. Forecasting*, **24**, 1252–1267.

Hitchens, N. M., and H. E. Brooks, 2012: Evaluation of the Storm Prediction Center's Day 1 Convective Outlooks. *Wea. Forecasting*, **27**, 1580–1585.

Hitchens, N. M., Harold E. Brooks, M. P. Kay, 2013: Objective Limits on Forecasting Skill of Rare Events. *Wea. Forecasting*, **28**, 525–534.

Stensrud, D. J., and Coauthors, 2009: Convective-Scale Warn-on-Forecast System. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Torn, R. D., G. J. Hakim, C. Snyder, 2006: Boundary Conditions for Limited-Area Ensemble Kalman Filters. *Mon. Wea. Rev.*, **134**, 2490–2502.