13B.5 An environmental climatology of the CAPS Storm-Scale Ensemble Forecast system during the 2010 HWT Spring Experiment

Christopher J. Melick*[,1,5], Israel L. Jirak[1], Steven J. Weiss[1], Adam J. Clark[2], Patrick T. Marsh[2,3], John S. Kain[2], Ming Xue[3,4], Fanyou Kong[4], and Kevin W. Thomas[4]

[1]*NOAA/NWS/NCEP Storm Prediction Center*
[2]*NOAA/OAR/National Severe Storms Laboratory*
[3]*School of Meteorology, University of Oklahoma*
[4]*Center for Analysis and Prediction of Storms, University of Oklahoma*
[5]*CIMMS, University of Oklahoma*
Norman, OK

1. Introduction

Since 2007, the University of Oklahoma (OU) Center for Analysis and Prediction of Storms (CAPS) has provided the NOAA Hazardous Weather Testbed (HWT) Spring Experiment with data from their 4-km grid length storm scale ensemble forecast (SSEF) system. In 2010, the SSEF was composed of several configurations of the Weather and Research Forecasting (WRF) model as well as the Advanced Regional Prediction System (ARPS; Xue et al. 2003). To better understand the temporal and domain-wide behavior of the 26-member SSEF, data from the 5-week period encompassing the 2010 Spring Experiment (Weiss et al. 2010) were examined to create an ensemble climatology for selected meteorological parameters related to convective development, including 2-m temperature, 2-m dew point, surface-based CAPE (SBCAPE), and 0-6 km AGL bulk wind shear. Results from the statistics will be illustrated through 2D field analyses (e.g., domain wide geographic displays at different forecast hours) to help augment the standard graphical results from the forecast climatology. Finally, some skill-related comparisons with hourly objective mesoscale analyses are presented.

**Corresponding author address*: Christopher J. Melick, NOAA/NWS/NCEP Storm Prediction Center, 120 David L. Boren Blvd, Norman, OK 73072; E-mail: chris.melick@noaa.gov

2. Background and Motivation

Ensemble guidance in numerical weather prediction has provided a means to display a range of plausible forecast solutions and to quantify forecast uncertainty, which is generally not possible by relying on just one deterministic model configuration. Unlike medium-range and short-range systems which utilize grid lengths on the synoptic to upper-mesoscale spectrum, the advent of storm-scale ensembles occurred recently in an experimental setting, partly due to the high computational costs involved. By removing convective parameterizations within convection-allowing models (CAMs) with horizontal grid lengths on the order of 4-km, more realistic structures and circulations associated with convective systems have been noted in forecast fields (e.g., Kain et al. 2005; Weisman et al. 2008). However, Coniglio et al. (2010) recently evaluated environmental variables in CAMs from the 2008 SSEF system and found that the CAMs did not predict the environment as well as a lower-resolution operational model. Their results, in combination with subjective assessments in prior Spring Experiments, led the authors to speculate that the high amplitude errors away from boundaries were associated with the WRF model's physical parameterizations. Accordingly, CAPS continued to make adjustments to the ensemble in 2009 with the inclusion of two more CAMs (NMM and ARPS) to supplement the existing members using the ARW version of the WRF. Consequently, the number of members doubled to 20 and errors arising from using a single dynamic core were mitigated (Xue et al. 2009).

The advantage of implementing a multi-model design also persisted into 2010, when the SSEF increased to 26 members as a more diverse selection of perturbations were incorporated (see Xue et al. 2010 for details). One of the goals of this investigation is to further test the modifications in the SSEF and gauge sensitivities to the variations amongst the individual members. To accomplish this objective, a statistical diagnosis of the aforementioned variables (2-m temperature, 2-m dew-point, SBCAPE, and 0-6 km AGL bulk wind shear) was conducted. From this climatology, it would be possible to identify typical magnitudes and spatial/temporal patterns of environmental features often exhibited during the 30-hr forecast period of the SSEF. Consequently, a goal is to provide valuable guidance to ensemble model developers and experimental forecasters in the HWT on the general behavior and tendencies in the ensemble. Additionally, verification datasets from the coarser resolution SPC hourly mesoscale analyses (Bothwell et al. 2002) allowed a qualitative measure of performance for each of the selected variables.

3. Data and Methodology

3.1. Data

A thorough description of the 2010 CAPS SSEF can be found in Xue et al. (2010). Briefly, the SSEF is a multi-initial condition (IC), multi-lateral boundary condition (LBC), and multi-physics system incorporating three dynamical cores (19 ARW, 5 NMM, 2 ARPS). From the listing of the 26 member configurations found in Tables 1-3, it should be noted that several new planetary boundary layer and double moment microphysics schemes were introduced this year. As with the last two experiments, reflectivity and velocity data from the national NEXRAD radar network and a cloud analysis were assimilated into 23 members as part of a CAPS 3DVAR system. Initialized daily at 00 UTC, each forecast cycle had hourly model output extending to 30-hours on a larger domain covering the full, continental United States (CONUS). This study uses data from the 25 weekdays during which the Spring Experiment was conducted (May 17th-June 18th). In order to focus on the convectively active areas east of the Rocky Mountains and minimize data flow issues into the HWT, CAPS provided the SPC GEMPAK (GEneral Meteorological PAcKage; desJardins et al. 1991) fields on a 4-km sub-domain (863x693; Fig. 1) over the central and eastern CONUS.

Table 1. Configurations of initial conditions (IC), lateral boundary conditions (LBC), microphysics schemes (MP), land surface models (LSM), and planetary boundary layer (PBL) schemes for ARW members. NAMa and NAMf refer to 12 km NAM analysis and forecast, respectively. ARPSa refers to ARPS 3DVAR and cloud analysis. All members assimilate radar data except for arw_c0. All members utilize RRTM (Goddard) scheme for long-wave (short-wave) radiation physics.
No convective parameterization is utilized.

member	IC	LBC	MP	LSM	PBL
arw_cn	00Z ARPSa	00Z NAMf	Thomps	Noah	MYJ
arw_c0	00Z NAMa	00Z NAMf	Thomps	Noah	MYJ
arw_m3	arw_cn + random pert	00Z NAMf	Thomps	Noah	MYJ
arw_m4	arw_cn + recursive pert	00Z NAMf	Thomps	Noah	MYJ
arw_m5	arw_cn + em-p1 + recur pert	21Z SREF em-p1	Morrison	RUC	YSU
arw_m6	arw_cn + em-p1_pert	21Z SREF em-p1	Morrison	RUC	YSU
arw_m7	arw_cn + em-p2_pert	21Z SREF em-p2	Thomps	Noah	QNSE
arw_m8	arw_cn – nmm-p1_pert	21Z SREF nmm-p1	WSM6	RUC	QNSE
arw_m9	arw_cn + nmm-p2_pert	21Z SREF nmm-p2	WDM6	Noah	MYNN
arw_m10	arw_cn + rsmSAS-n1_pert	21Z SREF rsmSAS-n1	Ferrier	RUC	YSU
arw_m11	arw_cn – etaKF-n1_pert	21Z SREF etaKF-n1	Ferrier	Noah	YSU
arw_m12	arw_cn + etaKF-p1_pert	21Z SREF etaKF-p1	WDM6	RUC	QNSE
arw_m13	arw_cn – etaBMJ-n1_pert	21Z SREF etaBMJ-n1	WSM6	Noah	MYNN
arw_m14	arw_cn + etaBMJ-p1_pert	21Z SREF etaBMJ-p1	Thomps	RUC	MYNN
arw_m15	arw_cn	00Z NAMf	WDM6	Noah	MYJ
arw_m16	arw_cn	00Z NAMf	WSM	Noah	MYJ
arw_m17	arw_cn	00Z NAMf	Morrison	Noah	MYJ
arw_m18	arw_cn	00Z NAMf	Thomps	Noah	QNSE
arw_m19	arw_cn	00Z NAMf	Thomps	Noah	MYNN

Table 2. Configurations of IC, LBC, MP, long-wave physics (LWP), short-wave physics (SWP), and LSM for NMM core members. All members utilize MYJ scheme for PBL. Again, c0 configuration does not assimilate radar data. No convective parameterization is utilized.

member	IC	LBC	MP	LWP/SWP	LSM
nmm_cn	00Z ARPSa	00Z NAMf	Ferrier	GFDL/GFDL	Noah
nmm_c0	00Z NAMa	00Z NAMf	Ferrier	GFDL/GFDL	Noah
nmm_m3	nmm_cn + nmm-n1_pert	21Z SREF nmm-n1	Thomps	RRTM/Dudhia	Noah
nmm_m4	nmm_cn + nmm-n2_pert	21Z SREF nmm-n2	WSM 6-class	RRTM/Dudhia	RUC
nmm_m5	nmm_cn + em-n1_pert	21Z SREF em-n1	Ferrier	GFDL/GFDL	RUC

Table 3. Configurations of IC, LBC, MP, radiation, and LSM for ARPS members. All members utilize MYJ scheme for PBL. Again, c0 configuration does not assimilate radar data. No convective parameterization is utilized.

member	IC	LBC	MP	radiation	LSM
arps_cn	00Z ARPSa	00Z NAMf	Lin	Chou/Suarez	Force-restore
arps_c0	00Z NAMa	00Z NAMf	Lin	Chou/Suarez	Force-restore

The SPC mesoscale analyses merge surface observations with a short-term analysis from the Rapid Update Cycle model (RUC; Benjamin et al. 2004). The hourly updated mesoscale analysis grids provided an accuracy benchmark for the SSEF results as an estimate of the three-dimensional "depiction" of the atmosphere on a frequent basis. An effective comparison of the statistics required the construction of a common geographical area so that coincident, non-missing data points existed for both the SSEF forecasts and verifying mesoscale analyses, which roughly covered the central and eastern CONUS (Fig. 1).

3.2. Methodology

The examination of SSEF environmental variables was conducted using three methods. First, domain-wide ranked values were obtained by sorting all grid points from highest to lowest for each dataset. This technique was accomplished for each of the 26 members at each forecast hour by time-averaging (for select percentiles) over all of the initialization days, from which a mean (ENS-MEAN) was calculated to characterize the entire SSEF. For completeness, an analogous procedure was applied to the mesoscale analyses to acquire a reference state at each verifying, nominal hour. The resulting ENS-MEAN statistics were then compiled to produce time-series box-and-whisker diagrams. The second method incorporated variations within the SSEF via trend plots of just the median. This examination also determined how often each ensemble member produced the top three and lowest three contributions at each forecast hour to establish clear distinctions across the SSEF composition. With respect to the box-and-whisker analyses, the 10th and 90th percentile positions were utilized instead of the grid maximum and minimum since some of the variables, especially 2-m dew point, contained physically unrealistic magnitudes on the extremes of the distribution. A cursory examination of lowest model level dew point forecasts from the CAPS web page (http://www.caps.ou.edu/wx/hwt/) and several sounding profiles (not shown) indicated the extreme dew point values were restricted to the 2-m fields and not the raw model data. This suggests that the unrealistic dew point values were introduced into the dataset through the algorithm in the WRF post used to create the 2-m dew point field. Unfortunately, lowest model level data were not available for this study, and the source of these apparent poor forecasts is under investigation.

To complete the three approaches, spatial composite plots for all four atmospheric variables were created to provide common patterns and temporal/spatial evolution over the course of a 30-hour forecast. Similar to the prior diagnosis, this process required a ranked distribution, but the sorting was restructured to organize all possible results for each grid point at one forecast hour. This utilized a 25-day (5-week) collection of outcomes for the mesoscale analysis fields, and when multiplied over all the configurations available in the SSEF, the maximum sample increased to about 650 solutions. The median was once again selected for analysis of the central tendency to diminish the influence of anomalous outliers. Considering the larger sample for the ensemble, the inter-quartile range (IQR) was computed to quantify ensemble spread. Only the 19-hour forecasts and corresponding 19 UTC mesoscale analyses are presented to highlight features and differences while remaining concise. This particular choice in the afternoon was near the time of peak diurnal values of SBCAPE. Animations of each hourly snapshot time over the entire forecast cycle for each of the four parameters can be obtained from these website links:

(http://www.spc.noaa.gov/publications/melick/shr06_ssef_sfcoa_fhrloop.gif; http://www.spc.noaa.gov/publications/melick/cape_ssef_sfcoa_fhrloop.gif; http://www.spc.noaa.gov/publications/melick/dwpf_ssef_sfcoa_fhrloop.gif; http://www.spc.noaa.gov/publications/melick/tmpf_ssef_sfcoa_fhrloop.gif).

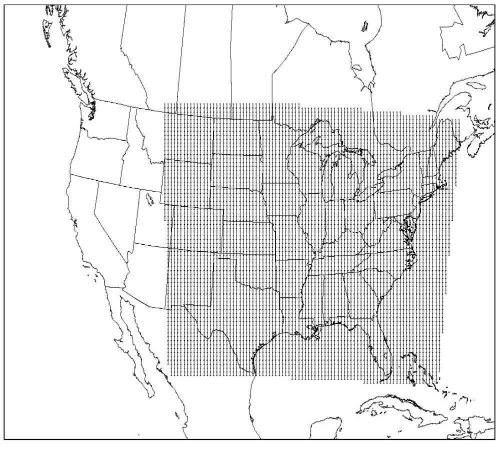

Fig. 1. Geographical map domain. Region of North America within box indicates extent of the SPC hourly mesoscale analysis datasets. Grid points highlighted over the central and eastern United States represent common area with SSEF system for statistical analysis and comparison.

4. Results

4.1. Graphical Domain-wide Statistics

An inspection of hourly box-and-whisker diagrams for 2-m temperature (Fig. 2) emphasized a recurring diurnal signal in both the ENS-MEAN of the SSEF and mesoscale analyses. Consistently, the ENS-MEAN showed the coldest readings occurring at 10 UTC with the warmest ones around 21 UTC, which was consistent with the surface analyses. (More specifically, the minimum value resided in the upper 40 ^{o}F range for the 10th percentile on the gridded dataset and the 90th percentile maximum reached to near 91 ^{o}F.) The overlay method utilized in Fig. 2 effectively illustrates the nearly coincidental temperature distributions regardless of percentile and hour as the majority of departures were small (peaking around 1-2 ^{o}F). In the same manner, average 2-m dew point distributions are presented in Fig. 3 and revealed a nearly steady range of about 35 ^{o}F between the 10th and 90th percentiles on any one dataset. In comparison to temperature, more substantial differences between the SSEF and mesoscale analyses are evident in dew points that were generally double in magnitude and always positive. Thus, the preference toward overly moist conditions appeared to indicate systematic problems in the SSEF. However, given the previous indications of errors in the 2-m dew point values from the WRF-post algorithm, these results should be viewed with caution. In the future, evaluations of the raw data instead of post-processed data will be required to get a better idea of the true environment.

Figures 4 and 5 present box-and-whisker graphs for SBCAPE and 0-6 km AGL bulk wind shear, respectively. An assessment of the former shows values greater than zero over half the domain early in the morning increasing to well over three-quarters of the domain later into the afternoon. The ENS-MEAN in the SSEF typically exhibited a minimum value close to 50 J/Kg for the median and a broad time period (forecast hour 17-24) when the 90th percentile CAPE was in excess of 2500 J/Kg. The diurnal trends show a more (less) unstable atmosphere earlier (later) into the forecast cycle compared to the mesoscale analysis (Fig. 4). The mesoscale analyses display a more pronounced peak in the upper tail of the distribution around 19 UTC. However, there are differences in how CAPE is calculated amongst the two frameworks, and this likely impacts the statistical results. In particular, the mesoscale analyses incorporate virtual temperature in their CAPE calculations whereas the SSEF calculations do not. This typically results in higher CAPE values for the former, especially when the surface mixing ratio is large. Finally, all of the ranked values for vertical shear in Fig. 5 remained very uniform with time with the 10th and 90th percentiles ranging from near 10 knots to greater than 45 knots. More importantly, this parameter tended to produce accurate forecasts and small biases for the SSEF compared to the mesoscale analyses.

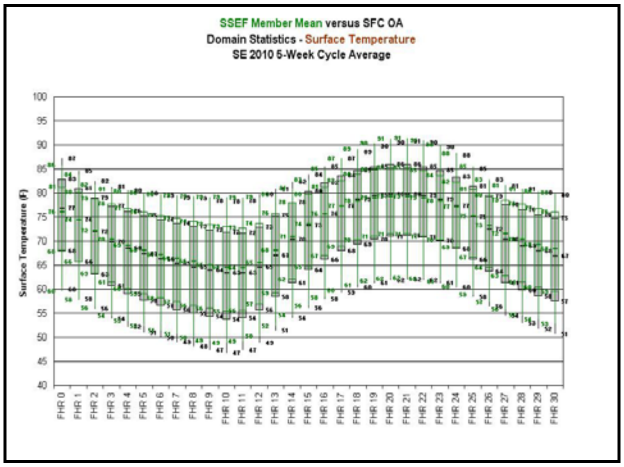

Fig. 2. Box-and-whisker overlay plots of 2-m (surface) temperature (F) for all 31 forecast hours (including analysis) from the 26-ensemble member mean of the SSEF with the verifying observations obtained from the SPC mesoscale analysis scheme. Results from the former (latter) are displayed in green (black) with open (gray-filled) boxes and data labels offset to the left (right).

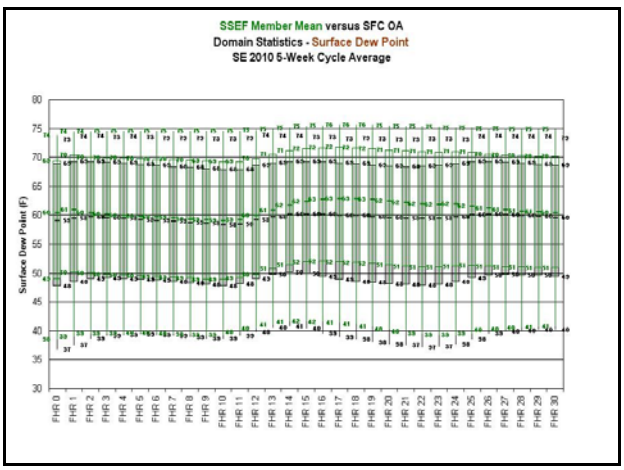

Fig. 3. As in Fig. 2 except for 2-m (surface) dew points (F).

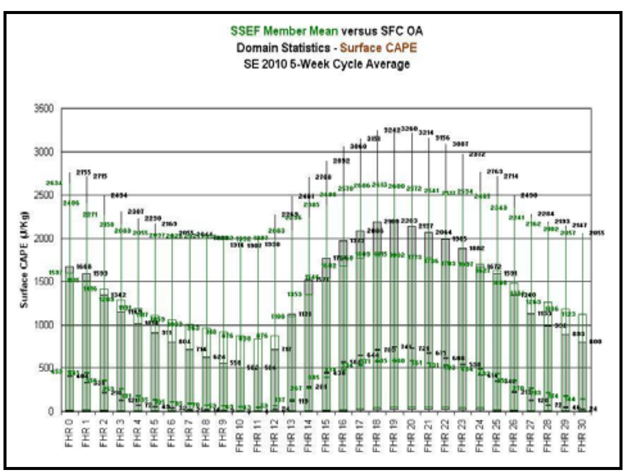

Fig. 4. As in Fig. 2 except for SBCAPE (J/Kg).

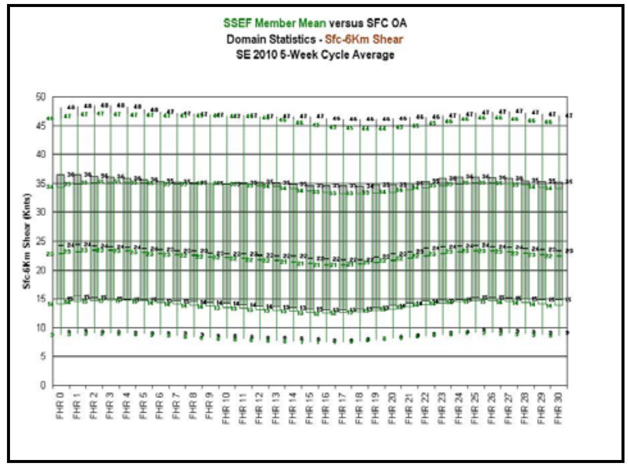

Fig. 5. As in Fig. 2 except for 0-6 km AGL bulk wind shear (knots).

4.2 Ensemble Member Contributions

To assess the contribution of individual SSEF members to overall ensemble performance, the median from each member during the 30-hr forecast period is displayed for each of the four variables (Figs. 6, 7, 8, and 9). Figure 6 shows a clustering of temperatures in the middle 70s at SSEF initialization with most solutions within 0.5 °F of the ENS-MEAN and mesoscale analysis values. The range in the median increased to ~8 °F halfway through the forecast before decreasing somewhat thereafter. Additionally, a considerable portion of the ensemble members demonstrated a nocturnal warm bias, followed by members being more evenly distributed about the mean during the daylight hours (Fig. 6). When individual members are considered, fifteen different members contributed at least once to the top three rankings in the ensemble median, whereas 23 of the members did so for the bottom three positions. For the highest and lowest member rankings, only two WRF-ARW (s4m10_arw and s4m6_arw) members were associated with the highest outcomes for more than fifty percent of the forecast hours. On the other hand, one other ARW (s4m13_arw) and both ARPS solutions surpassed the same condition for the lowest results.

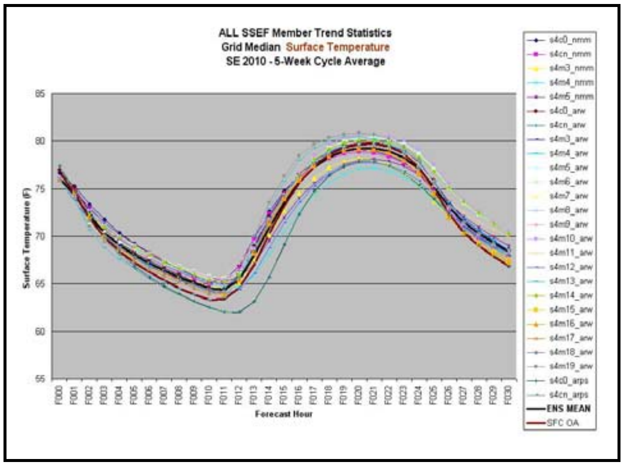

Fig. 6. Domain-wide median trend lines of surface (2-m) temperatures (F) for all 26-members of the SSEF. The ensemble mean and verifying SPC mesoscale analysis statistics are also provided for the 0-30 hour forecast cycle.

The variations in median dew point forecasts with time (Fig. 7) portrayed inconsistency and inaccuracy for the SSEF with little similarity in the behavior for all traces. The statistics exhibited standard deviations often on the order of 1-2 ^{o}F and even more pronounced differences between the maximum and minimum values. The one recurring feature in the diagnosis was that an overwhelming majority of the individual members displayed positive errors, with some times showing an excessively high, moist bias (e.g., 67.3 ^{o}F from s4m14_arw versus 59.9 ^{o}F from mesoscale analysis at forecast-hour 18). However, these results appear to reflect computational problems resulting from the 2-m algorithm, and should not necessarily be considered representative of actual SSEF performance until more analysis using raw model data is performed. The results from the s4m14_arw and s4m8_arw members each ranked more than 25 different times in the highest three magnitudes, and more significantly, combined to be the maximum in all but one instance. By looking at the lower end of the probability distribution, the s4m11_arw contribution frequently occurred in the last or second-to-last standing and was by far the statistic closest to actual conditions (Fig. 7).

The distribution of SBCAPE is seen in Fig. 8. Most differences of the median between members were typically a few hundred J/Kg but reached to near 1000 J/Kg when comparing the extremes during the afternoon hours. Comparisons with the mesoscale analyses showed the SSEF distribution having a high bias (e.g., greater instability) between 00 UTC and 12 UTC, but having a low bias during the daytime hours. Figure 8 reveals that all members displayed a diurnal periodic pattern but the daily median peak had a lower amplitude and was not as sharp as the mesoscale analysis. One of the NMM (s4m5_nmm) members accounted for a substantial portion of the number one rankings, such that the estimates during the heat of the day were 50-100 J/Kg higher than verification results. As was mentioned earlier, though, most of the members were less unstable than the mesoscale analysis, with two ARW (s4m13_arw and s4m5_arw) members and one ARPS (s4cn_arps) member usually representing the lowest two contributions. The 0-6 km AGL bulk wind shear in Fig. 9 displayed very small differences (1-2 knots) among the SSEF members over the entire forecast cycle and corresponded well to the verifying data, although a slight low bias was apparent.

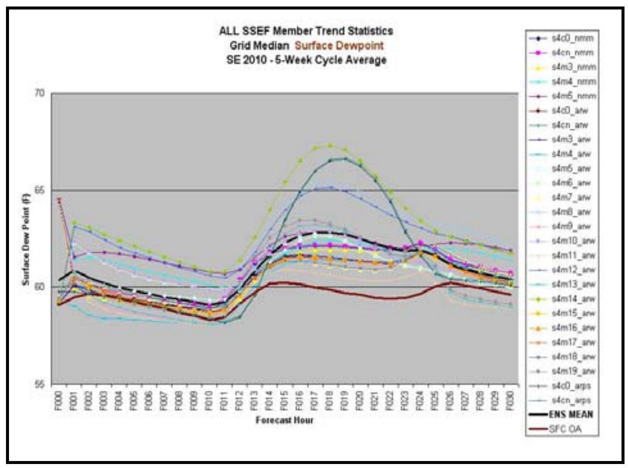

Fig. 7. As in Fig. 6 except for surface (2-m) dew points (F).

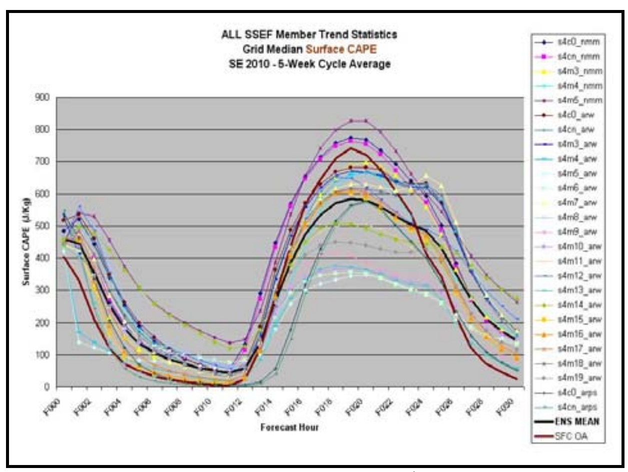

Fig. 8. As in Fig. 6 except for SBCAPE (J/Kg).

4.3 Side-by-Side Composite Maps

Composite plots for the 5-week period facilitated a way to present a comprehensive 2D examination of the SSEF and verifying SPC mesoscale analyses. Using this technique, typical characteristics for 2-m temperatures at forecast hour 19 are shown in Fig. 10. The side-by-side comparison presents striking similarities in the strength, placement, and pattern of features. Nonetheless, there are small regions where the SSEF had slightly elevated temperatures, such as over Oklahoma, Kansas, and North Dakota, which coincided with the largest values for the IQR (e.g., greater than 12 °F). Alternatively, while a general match in the 2-m dew point pattern existed, Fig. 11 indicated a tendency for the SSEF to keep the atmosphere too moist near the surface over a substantial part of the eastern and central United States. Again, however, the quantitative values should be viewed with caution given concerns about the 2-m dew point values. One of the more interesting components of this plot was the distinct dry line oriented north-to-south across the western domain on both the ensemble and the objective analysis. The forecasted placement of the feature corresponded quite well with observations but the magnitude of the moisture gradient, especially behind the system in New Mexico, was too weak. It is not uncommon for ensemble output to portray relatively weaker gradients given the spread inherent in member solutions. Once again, though, the placement and magnitude would frequently fluctuate at 19 UTC given the strong variations in the SSEF sample in this area (Fig. 11).

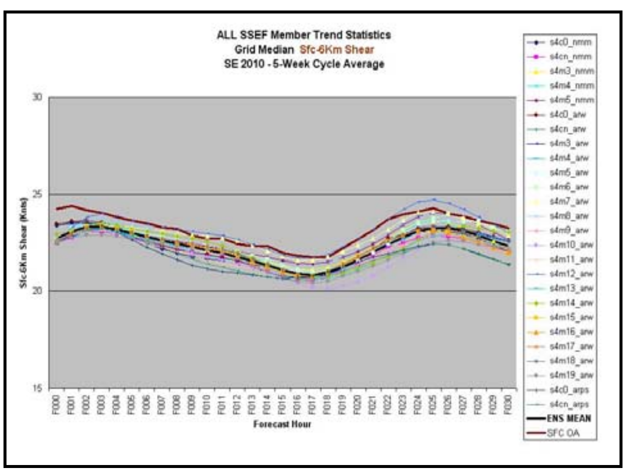

Fig. 9. As in Fig. 6 except for 0-6 km AGL bulk wind shear (knots).

SBCAPE plots representing median values during the 2010 Spring Experiment are shown in Fig. 12. The geographical extent for minimal levels of instability (e.g., 250 J/Kg) was practically identical in both panels with near-zero SBCAPE values confined to the Northern Plains, Upper Midwest, Great Lakes, New England, and west of the dry line. The IQR, however, suggests that a much broader region of the country can experience positive CAPE estimates at this forecast hour on any one day. Further, the large results for the 25^{th} to 75^{th} percentile ranges, sometimes of 1000-3000 J/Kg and pronounced spatial differences therein, was indicative of strong sensitivity to model configuration and/or initialization. Nonetheless, the SSEF exhibited values well below the mesoscale analysis, especially (500-1000 J/Kg) along the Gulf Coast. The low bias is likely an artifact of the different CAPE calculation methods mentioned earlier. Consistent with earlier results, the evaluation of vertical shear (Fig. 13) indicated minimal errors overall and the lower IQR implied a clustering of solutions from the SSEF. A majority of what small, negative biases (i.e. on the order of 5 knots) were present appeared to occur near the general location of the jet stream, which by the spring and early summer is over higher latitudes and the northern periphery of the United States.

5. Conclusions

A climatology and initial assessment of four select variables (2-m temperature, 2-m dew point, SBCAPE, and 0-6 km AGL bulk wind shear) from the SSEF during the 2010 HWT Spring Experiment has been presented. Using three different analysis techniques, average domain-wide statistics for the ensemble as a whole, ensemble member contributions, and composite plots representing environmental conditions are used to explore SSEF performance. In evaluating the performance of the ENS-MEAN and individual member configurations, verification for the entire 30-hour forecast run was performed using the SPC hourly mesoscale analysis fields. One of the main findings was that SSEF forecasts of 2-m temperatures and 0-6 AGL bulk wind shear values were the most accurate in terms of diurnal cycle phase, magnitude, and distribution of geographical features. Moreover, the spread amongst the various members was relatively small and evenly distributed about the ENS-MEAN. However, there was a tendency for the temperatures for the individual members to exhibit larger variability beyond the 10-hour forecast time, thus emphasizing the importance of examining spread in an ensemble (e.g., ENS-MEAN).

For the CAPE field, pronounced sensitivities existed with respect to member configuration, grid point location, and synoptic setup. The opportunity for quantitative comparisons with the mesoscale analysis SBCAPE fields was impacted, however, because different formulas were used to calculate CAPE from the two sources. Nonetheless, it is interesting to note that the ensemble members generally under-predicted the diurnal range and sharpness of the afternoon peak of CAPE compared to the verifying data. The median of the SSEF SBCAPE showed one NMM (s4m5_nmm) member to consistently contribute to the maximum. As for the other extreme, two ARW (s4m13_arw and s4m5_arw) members and one ARPS (s4cn_arps) member usually represented the lowest two rankings. Finally, 2-m dew point data revealed a wide range of values, but the presence of some extreme (and physically unrealistic) values suggested that further investigation should be conducted before meaningful conclusions are drawn. Thus, a separate analysis will be undertaken on the raw CAPS SSEF data in an attempt to explain the apparent systematic, moist bias observed in this study.

A major goal of this work was to provide forecasters and model developers with a better understanding of spatial and temporal characteristics of several environmental fields form the 2010 SSEF, and performance characteristics of individual ensemble members. More work would still be needed, however, to more thoroughly investigate SSEF performance, especially with respect to dew point. Future research will be mainly focused on a more direct comparison between SSEF raw model output from the variables discussed in this work and observations by using an objective analysis with comparable, storm-scale grid-lengths (e.g., the NWS Real-Time Mesoscale Analysis).

Acknowledgements. This research was supported by an allocation of advanced computing resources provided by the National Science Foundation. The computations were performed on Athena (a Cray XT4) at the National Institute for Computational Science (NICS; http://www.nics.tennessee.edu/).

References

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Wea. Rev.*, **32**, 495-518.

Bothwell, P. D., J. A. Hart, and R. L. Thompson, 2002: An integrated three-dimensional objective analysis scheme in use at the Storm Prediction Center. *Preprints, 21st Conf. on Severe Local Storms*, Amer. Meteor. Soc., San Antonio, TX, J117-120.

Coniglio, M. C., K. L. Elmore, J. S. Kain, S. J. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF model output for severe-weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, **25**, 408-427.

desJardins, M.L., K.F. Brill, and S.S. Schotz, 1991: Use of GEMPAK on Unix workstations, *Proc. 7th International Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, New Orleans, LA, Amer. Meteor. Soc., 449-453.

Kain, J. S., S. J. Weiss, M. E. Baldwin, G. W. Carbin, D. R. Bright, J. J. Levit, and J. A. Hart, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL Spring Program. *Preprints, 21st Conf. on Weather Analysis and Forecasting/17thConf. on Numerical Weather Prediction,* Washington, DC, Amer. Meteor. Soc., 2A.5. [Available online at http://ams.confex.com/ams/pdfpapers/94843.pdf.]

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

Weiss, S. J., A. J. Clark, I. L. Jirak, C. J. Melick, C. W. Siewert, R. Sobash, P. T. Marsh, A. R. Dean, M. Xue, F. Kong, K. W. Thomas, J. Du, D. R. Novak, F. E. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, T. Jensen, J. S. Kain, M. C. Coniglio, and R. S. Schneider, 2010: An overview of the 2010 NOAA Hazardous Weather Testbed spring forecasting experiment. *25th Conf. Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., Paper 7B.1

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82**, 139-170.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, X. Wang, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: CAPS realtime 4-km multi-model convection-allowing ensemble and 1-km convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2009 Spring Experiment. *23rd Conf. Wea. Analysis and Forecasting/19th Conf. Numerical Wea. Prediction*, Omaha, NE, Amer. Meteor. Soc., Paper 16A.2.

Xue, M., F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S.J. Weiss, A.J. Clark, J.S. Kain, M.C. Coniglio, J. Du, T.L. Jensen, and Y.H. Kuo, 2010: CAPS Realtime Storm Scale Ensemble and High Resolution Forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. *25th Conf. Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., Paper 7B.3.

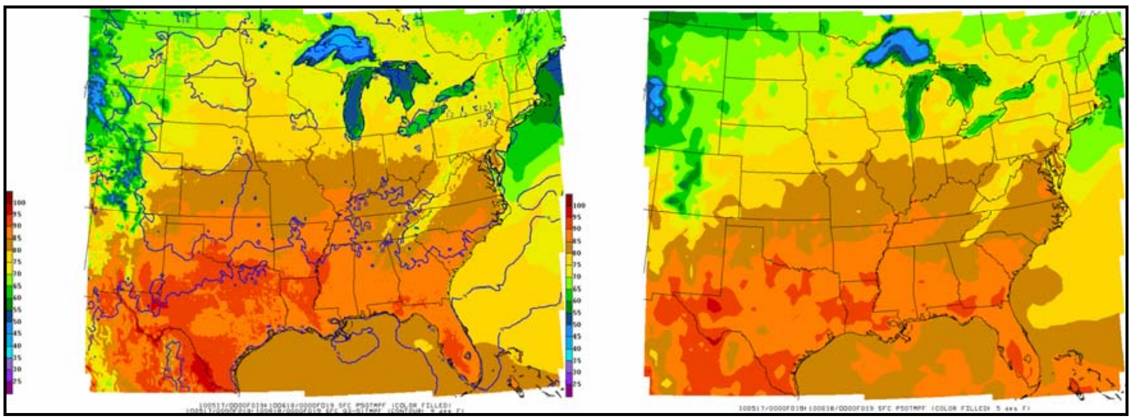

Fig. 10. Spatial composite surface (2-m) temperature plots from the SSEF (left) at forecast hour 19 matched with the verifying SPC mesoscale analysis (right). Each panel displays median grid-point values (color-shading; 5 deg F) determined by sorting all times in the five-week period examined, and in the case of the SSEF, all available datasets for each of the 26-members. The IQR (contour lines; 4 deg F) is also overlaid on the left panel to give a measure of variation in the SSEF sample.

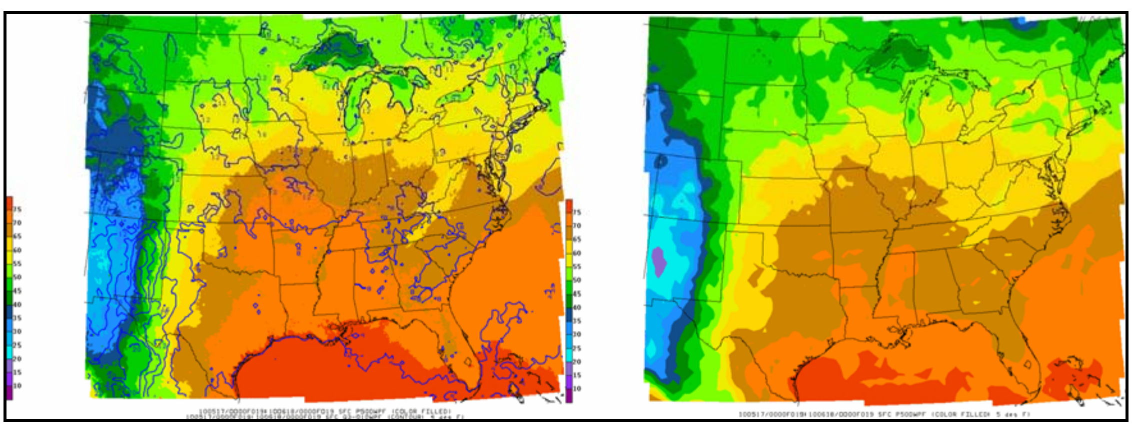

Fig. 11. As in Fig. 10 except for surface (2-m) dew points (F).

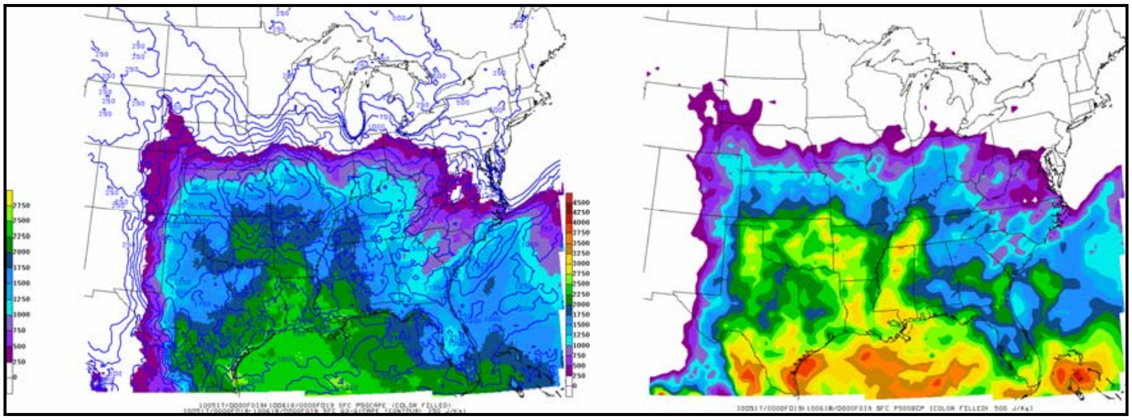

Fig. 12. As in Fig. 10 except for SBCAPE with contour (IQR) and color-filled (median) interval every 250 J/Kg.

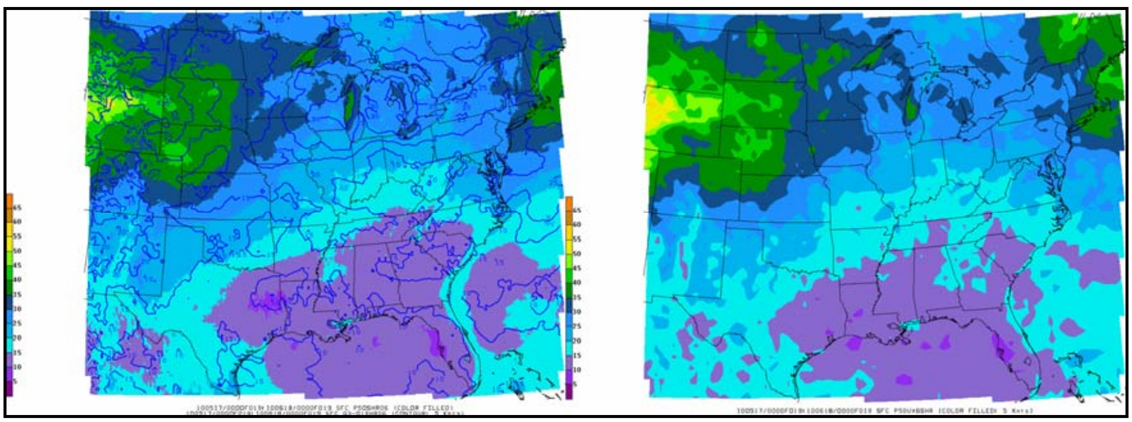

Fig. 13. As in Fig. 10 except for 0-6 km AGL bulk wind shear with contour (IQR) and color-filled (median) interval every 5 knots.