

Influence of a portable near-surface observing network on experimental ensemble forecasts of deep convection during VORTEX-SE

AARON J. HILL^{*†} AND CHRISTOPHER C. WEISS

Department of Geosciences, Texas Tech University, Lubbock, TX

DAVID C. DOWELL

National Oceanic and Atmospheric Administration, Earth System Research Laboratories, Global Systems Laboratory

ABSTRACT

Ensemble forecasts are generated with and without the assimilation of near-surface observations from a portable, mesoscale network of StickNet platforms during the Verification and Origins of Rotation in Tornadoes EXperiment – Southeast (VORTEX-SE). Four VORTEX-SE intensive observing periods are selected to evaluate the impact of StickNet observations on forecasts and predictability of deep convection within the southeast United States. StickNet observations are assimilated with an experimental version of the High Resolution Rapid Refresh Ensemble (HRRRE) in one experiment, and withheld in a control forecast experiment. Overall, StickNet observations are found to effectively reduce mesoscale analysis and forecast errors of temperature and dewpoint. Differences in ensemble analyses between the two parallel experiments are maximized near the StickNet array and then either propagate away with the mean low-level flow through the forecast period or remain quasi-stationary, reducing local analysis biases. Forecast errors of temperature and dewpoint exhibit periods of improvement and degradation relative to the control forecast, and error increases are largely driven on the storm scale. Convection predictability, measured through subjective evaluation and objective verification of forecast updraft helicity, is driven more by when forecasts are initialized (i.e., more data assimilation cycles with conventional observations) rather than the inclusion of StickNet observations in data assimilation. It is hypothesized that the full impact of assimilating these data is not realized in part due to poor sampling of forecast sensitive regions by the StickNet platforms, as identified through ensemble sensitivity analysis.

Significance statement. In this work, observations from a portable observation network during a large-scale field campaign are incorporated into numerical weather prediction models to improve forecasts of severe storms and their attendant hazards: tornadoes, hail, and severe wind. Observations are gathered from StickNet platforms (developed at Texas Tech University), which were placed throughout northern Alabama and southern Tennessee during the project. Over four cases examined in this manuscript, simulations that include StickNet observations are improved at earlier times, but forecast impacts at later times are varied. The observations improve near-surface temperature and moisture forecasts, but don't routinely influence forecasts of the actual storms, likely because the most sensitive regions that would improve forecasts were not well sampled by the StickNets. Future work should evaluate how more frequent observations could improve forecasts, beyond what was considered here (i.e., one observation per hour).

1. Introduction

The spatial climatology of severe storm hazards and their corresponding environments across the United States (U.S.) generally focuses within the Great Plains and southeast U.S. (e.g., Horgan et al. 2007; Guyer and Dean 2010; Dixon et al. 2011; Gensini and Ashley 2011; Cintineo et al. 2012; Gensini et al. 2020). Whereas particular attention has been given to severe thunderstorms across the Great Plains over the last few decades through various large-scale field programs, including the Verification and Origins of Rotation in Tornadoes EXperiment (VORTEX; Rasmussen et al. 1994), VORTEX2 (Wurman et al. 2012), the Mesoscale Predictability EXperiment (Weisman et al. 2015), and the Plains Elevated Convection At Night project (Geerts et al. 2017), less attention has been provided to the southeast thunderstorm environment. The lack of attention is particularly noteworthy since the Southeast is prone to more frequent and less predictable high-shear, low-CAPE (HSLC) environments preceding severe weather events (Davis and Parker 2014; Sherburn and Parker 2014; Sherburn et al. 2016; King et al. 2017), a disproportionate frequency of nocturnal storms (Ashley 2007; Kis and Straka 2010) and long-track tornadoes (Dixon et al. 2011), forested areas that hinder line-of-sight (Ashley 2007), complicated storm modes (e.g., Smith et al.

^{*}Corresponding author: Aaron J. Hill, aaron.hill@colostate.edu

[†]Current affiliation: Department of Atmospheric Science, Colorado State University, Fort Collins, CO

2012), an ill-defined “storm season” (e.g., Smith et al. 2012), and population vulnerabilities (e.g., mobile homes, poverty, and elderly populations; Ashley 2007; Wallace et al. 2015; Childs et al. 2018). Furthermore, recent work has hinted at the possibility that severe thunderstorm hazards have become more frequent across the Mississippi River Valley and southeast U.S. over the last two decades (Gensini et al. 2020), which suggests improving severe storm hazard forecasts in these regions could be extremely beneficial, greatly reducing the impacts storms and their attendant hazards have on the population and society.

Recently, VORTEX - Southeast (VORTEX-SE) was birthed to improve the understanding of physical processes that contribute to an enhanced risk of severe thunderstorms and associated hazards (e.g., Smith et al. 2012) in the southeast U.S. as well as human behavior that may explain the relatively higher mortality due to severe hazards in this region (e.g., Ashley 2007). A number of meteorological observing assets – e.g., mobile Doppler radars, rawinsondes, instrumented towers, and profilers (e.g., Lee et al. 2019; Markowski et al. 2019; Tanamachi et al. 2019; Lyza et al. 2020) – were deployed during field campaigns in the springs (March and April) of 2016 and 2017, as well as during the cold seasons (November – April) of 2018 and 2019, to address the project’s physical science objectives. In support of VORTEX-SE, a fleet of portable, near-surface in-situ sampling platforms – StickNets (Schroeder and Weiss 2008; Weiss and Schroeder 2008) – were deployed by Texas Tech University to sample near-storm environmental heterogeneities (e.g., McDonald and Weiss 2020) and supplement the existing observational network in northern Alabama and southern Tennessee. Whereas StickNet applications in rapid-deployment scenarios have been well documented – primarily sampling supercell cold pools and outflow wind gusts (e.g., Skinner et al. 2011, 2014; Weiss et al. 2015; Gunter et al. 2017) – their utility as a portable, quickly-deployed, and stationary near-surface sampling network for severe storm environments is relatively new. The placement of a dense, high-frequency in-situ observation network during VORTEX-SE, sampling a relatively data-sparse region, provides a unique and valuable dataset to investigate forecast improvements of severe storms and their hazards in the southeastern U.S. Near-surface observations have immense value for convection forecasts, demonstrated by a number of studies (e.g., Sobash and Stensrud 2015; Chen et al. 2016; Madaus and Hakim 2017), and it is hypothesized that the portable StickNet fleet provided significant value to forecasts of severe hazards during VORTEX-SE. This paper will document and investigate StickNet-observation impacts on severe storm forecasts during VORTEX-SE, and quantify forecast changes when observations are assimilated with a numerical weather prediction model.

In order to properly evaluate these impacts, explicit simulations of deep convection and associated hazards (e.g.,

tornadoes, hail, and wind) are needed, necessitating the use of convection-allowing models (CAMs). CAMs utilize small horizontal grid spacing to explicitly simulate convective processes (Bryan et al. 2003). The last decade has seen a proliferation of CAMs, resulting in markedly improved forecasts of severe hazards across the U.S. (e.g., Done et al. 2004; Kain et al. 2006; Clark et al. 2010). Additionally, coupling CAMs with sophisticated data assimilation techniques has yielded improved representation of the near-storm environment, improving the subsequent forecasts of severe hazards (e.g., Stensrud et al. 2013). Generally, in-situ observations have been found to contribute to this improvement through a more accurate representation of the mesoscale environment (e.g., Stensrud et al. 2009; Wheatley and Stensrud 2010; Wheatley et al. 2012; Knopfmeier and Stensrud 2013; Ha and Snyder 2014; Torn 2014; Coniglio et al. 2016; Hitchcock et al. 2016), motivating the incorporation of StickNet observations with a CAM in this study.

The advent of ensemble prediction systems (EPSs) at convection-allowing resolutions has further improved forecast skill of deep convection and associated hazards (e.g., Clark et al. 2009; Kain et al. 2013; Schwartz et al. 2015; Wheatley et al. 2015; Jones et al. 2016; Sobash 2016). Whereas deterministic CAM solutions provide only one snapshot of a future atmospheric state, EPSs of sufficient size provide a representative measure of forecast uncertainty, which can be extremely valuable to real-time forecasters and a statistical tool for researchers to assess the relative predictability of a particular event (e.g., Nielsen and Schumacher 2016) or series of events. As computing power and capability have risen over the past few decades, real-time convection-allowing ensembles (CAEs) have been developed at universities (e.g., Manser and Ansell 2020), research centers (e.g., Xue et al. 2007; Schwartz et al. 2015), and operational weather centers (e.g., Hagelin et al. 2017; Jirak et al. 2018; Klasa et al. 2018). Operationally, NOAA’s High Resolution Ensemble Forecast (HREF; Jirak et al. 2018) system, which combines a diverse set of deterministic convection-allowing models, produces skillful forecasts of high-impact weather including severe thunderstorms (Jirak et al. 2012; Roberts et al. 2020). Research at NOAA’s Global System Laboratory has focused on developing data-assimilation and forecasting capabilities with a single-model CAE system, based on the High Resolution Rapid Refresh (HRRR) modeling infrastructure (Alexander et al. 2016; Dowell et al. 2021).

For the purposes of this work, the single-model CAE in development at NOAA will be utilized to demonstrate how the portable StickNet array of near-surface observations can improve forecasts of severe storms in the southeast U.S., while simultaneously informing best practices for future development of an operational CAE. The rest of the paper is laid out as follows: the StickNet platforms, modeling system, and analysis methods to quantify forecast

improvements are detailed in Section 2; ensemble analyses and forecasts from each case are subjectively and objectively assessed in Section 3; and discussion on forecast sensitivity is reserved for Section 4. A summary and discussion of the results is presented in Section 5, along with implications for future VORTEX-SE project deployments.

2. Methods

a. StesoNet

During the 2016 and 2017 spring field phases of VORTEX-SE, 16 StickNet (e.g., Fig. 1) platforms were positioned with 40 km spacing across southern Tennessee and northern Alabama (Fig. 2) – hereafter, referred to as the “StesoNet” – in an effort to sample thunderstorm cold pool properties as well as the mesoscale environment preceding these storms. The remaining eight platforms of the StickNet fleet were used for rapid deployments in advance of targeted storms. During the 2018-2019 campaign (Meso18-19), the eight rapid-deployment StickNets were inserted into the stationary network to extend coverage southward towards central Alabama and improve sampling resolution in the central portion of the domain.

Each StickNet probe is equipped to measure pressure, wind speed and direction, temperature, and relative humidity at 10-Hz sampling frequency, approximately 1.5 m above the surface. The computing hardware and pressure sensor are mounted inside a Campbell Scientific data acquisition box with a small battery for < 12 hour deployments. An external battery box can also be deployed that extends the StickNet’s sampling lifetime to > 24 hours (typically less than five days), which has been advantageous for prolonged deployments during landfalling hurricanes (e.g., Zachry et al. 2013; Giannanco et al. 2016; Alford et al. 2019; Fernandez-Caban et al. 2019). During VORTEX-SE, a 100 W solar panel was attached to each probe (see Fig. 1) in an effort to provide continuous power to the StesoNet array and allow uninterrupted sampling of multiple convective events across each season of sampling. Specific details regarding instrument characteristics are provided by Skinner et al. (2011).

Biases for all StickNet probes were determined prior to and after field campaigns through mass tests, in which all probes were deployed in close proximity in an open area for multiple days. An arithmetic average was then taken over all probe observations during a selected period (e.g., clear diurnal pattern) to assess the bias characteristics for each instrument (i.e., temperature, pressure, and relative humidity) on each probe. Mass tests were conducted on either end of the observing periods to determine if any biases changed during transport, deployment, or during the sampling periods. The raw observations from each StickNet platform obtained during the field project were then corrected based on their respective biases determined through the mass tests (to the nearest 0.5 K, 0.5 hPa, and

0.5 % for temperature, pressure, and relative humidity observations, respectively). When biases appeared to change drastically between pre-deployment and post-deployment mass tests, the mass-test biases were averaged together and again rounded. These bias-corrected observations are provided to the ensemble modeling system for data assimilation, which has an additional quality control procedure described in the following subsection, so no subjective or objective quality control is conducted to remove erroneous StesoNet observations.

b. Modeling Configuration and StickNet Assimilation

In an effort to improve forecasts of deep convection and severe hazards across the contiguous United States (CONUS), the NOAA’s Global Systems Laboratory has been developing convection-allowing, ensemble-based data-assimilation and forecasting methods in an experimental system. This modeling system effectively extends the capabilities of the HRRR operational model to an ensemble framework, known as the HRRR Ensemble (HRRRE; Dowell et al. 2021). HRRRE forecasts are made with WRF-ARW version 3.9 on one-way nested domains with 15 and 3-km grid spacing (Fig. 2) and 36 ensemble members. The 3-km grid spacing of the full-CONUS inner domain allows for explicit simulation of deep, moist convection (Bryan et al. 2003). Observations are assimilated hourly into the 36-member ensemble with the Grid-point Statistical Interpolation (GSI) ensemble Kalman filter (EnKF; Evensen 1994) software.

During select VORTEX-SE intensive observing periods (IOPs) analyzed in this work, the HRRRE system is initialized at 0900 UTC the day of an event. The initial atmospheric ensemble-mean state comes from the 0900 UTC Rapid Refresh (RAP) model analysis, interpolated to the 15- and 3-km HRRRE grids (Dowell et al. 2021). The 36 HRRRE atmospheric initial conditions are produced by adding to this initial mean, perturbations from the first 36 members of a 9-h forecast from the NOAA Global Data Assimilation System (GDAS). The land-surface state in all 36 members is initialized from the 0900 UTC HRRR analysis, without perturbations. Boundary conditions come from a Global Forecast System (GFS) forecast, and random perturbations are added to the boundary conditions for each ensemble member. The 15-km outer HRRRE grid exists so that these random perturbations can evolve into more physically-consistent perturbations before reaching the inner 3-km grid.

To forecast an IOP event, data assimilation is performed hourly with the GSI-EnKF beginning at 1000 UTC and ending at 1800 UTC on both domains. Ensemble member backgrounds come from 1-hr forecasts propagated from the previous analysis; the first ensemble member backgrounds are 1-hr forecasts generated from the perturbed initial states created from the RAP and GDAS described

earlier. Conventional observations, such as those from METAR, aircraft, radiosondes, and buoys, are assimilated on both domains, and radar reflectivity from the WSR-88D network is included during assimilation on the 3-km inner domain. Assimilated observations update the model state variables of horizontal wind, temperature, water vapor, and hydrometeor mixing ratios (i.e., cloud water, rain, snow, and graupel). A Gaspari-Cohn localization function (Gaspari and Cohn 1999) is employed to limit observation impacts on state-variable covariances during assimilation. Conventional observations have horizontal and vertical localization radii of 300 km and 0.5 scale height, respectively, which denotes when observation weights reduce to zero. Horizontal and vertical localization radii of 18 km and 0.5 scale height, respectively, are applied to radar observations. Every hour, each member is checked for erroneous clouds that might have developed where satellite observations indicate clear conditions; erroneous clouds are removed by resetting cloud water and cloud ice to zero throughout the column and reducing the relative humidity to 85% at any locations where clouds were removed (Dowell et al. 2021). While ensemble spread is initially generated via random perturbations, it is maintained through a relaxation-to-prior-spread technique after assimilation each hour (Whitaker and Hamill 2012). Stochastic methods for increasing ensemble spread were not included in HRRRE for the current study, thus simplifying interpretations about how forecast perturbations are related to initial-condition perturbations.

At select times during an IOP (e.g., 1200, 1500, and 1800 UTC), ensemble analyses are used to generate 36-member forecasts that encompass the convective event. The aforementioned data assimilation cycling and ensemble forecast procedure serves as the control system (CTRL) for evaluation of forecast skill improvement. A parallel forecast system (EXP) is generated that assimilates bias-corrected, 1-minute averaged StesoNet observations of temperature, pressure, and specific humidity at the top of each hour (1000–1800 UTC) from each station; 16 and 24 stations are included in data assimilation cycles for the 2017 and Meso18-19 IOPs, respectively, as long as each station was operating. The StickNet relative humidity observations are converted to specific humidity for the GDAS assimilation. Only one observation of temperature, pressure, and humidity is used from each StickNet station during each assimilation cycle valid at the assimilation time. GDAS will reject a particular observation if it falls outside of three standard deviations of the background ensemble distribution, which effectively eliminates missing data points and disregards heavily biased observations, regardless of source (i.e., StesoNet or METAR). Forecasts are generated from the EXP analyses and then compared to CTRL forecasts initialized at the same time.

c. Analysis Methods

To evaluate changes in forecast skill due to StesoNet observation assimilation, subjective and objective analyses are employed on the ensemble analyses and forecasts of environmental and storm-specific variables (e.g., 2-m temperature and updraft helicity) from the CTRL and EXP ensemble systems. The analysis fit to observations (hereafter referred to as analysis error) is determined for both EXP and CTRL analyses of temperature and dewpoint by calculating absolute errors (AE) at individual observing sites, which includes the regular automatic surface observing system (ASOS) station network and StesoNet array. Changes in absolute error (ΔAE_i) from the CTRL to EXP simulation at a particular observing location are represented as the CTRL simulation AE ($AE_{i,CTRL}$) subtracted from the EXP simulation AE ($AE_{i,EXP}$),

$$\Delta AE_i = AE_{i,EXP} - AE_{i,CTRL} = |(F_i^{EXP} - O_i)| - |(F_i^{CTRL} - O_i)|, \quad (1)$$

where F_i is the forecast magnitude with respect to the EXP and CTRL simulations (F_i^{EXP} and F_i^{CTRL} , respectively) and O_i is the observation at a station i . Forecast and analysis errors of both simulations are assessed through domain-wide (33–36°N and 83–90°W) root mean squared error (RMSE) statistics computed against the full near-surface observation dataset (i.e., including the StesoNet array). RMSE statistics are calculated using the formula

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (F_i - O_i)^2}, \quad (2)$$

where M represents all stations considered in the domain. Through this quantitative analysis, the spatial and temporal characteristics of analysis and forecast errors across the two prediction systems can be assessed to determine the relative value of StesoNet observations in improving analyses and subsequent forecasts.

Additionally, forecast differences of storm-centric variables are quantified through probabilistic verification techniques. Specifically, changes in ensemble forecast probabilities of 1–6 km updraft helicity (UH) exceeding a threshold are verified against local storm reports (LSRs) of severe hail, wind, and tornadoes via the Brier Score to assess if any adjustments are made to the forecast from StesoNet observations that improve localized prediction of severe-storm hazards. Brier Scores will decrease (i.e., skill increases) in areas where forecast probabilities increase, coincident with a LSR. However, Brier Scores do not capture the underlying skill of CTRL, which may be sufficient given the relative density of LSRs and CTRL forecast probabilities. Moreover, the limited number of cases explored in this work hinders a more robust assessment of forecast skill, thus Brier Scores should be considered just one method to explore forecast skill.

Forecast probabilities are assembled via the neighborhood maximum ensemble probability (NMEP) method (Schwartz and Sobash 2017), whereby member forecasts of hourly-maximum UH are filtered spatially (21 x 21 grid-point box) for UH exceeding $50 \text{ m}^2\text{s}^{-2}$ and each forecast is converted to a binary yes/no UH-exceedance field on the native HRRRE domain. The resulting ensemble is then averaged and the ensemble mean field is smoothed using a Gaussian kernel with a smoothing length scale of 40 km (i.e., 10 grid points). Archived LSRs are obtained from Iowa State University's Iowa Environmental Mesonet (IEM) geographical information system archive (<https://mesonet.agron.iastate.edu/request/gis/>) and gridded such that all grid points within 40 km of an LSR are given a binary magnitude of 1. Brier Scores are calculated in objectively-defined regions defined by the gridded LSRs and NMEP-based ensemble forecast probabilities greater than five percent (approximately two ensemble members) from either forecast simulation. In this way, the Brier Score calculations are restricted to only portions of the analysis domain where hazards were reported or the ensemble systems predicted UH, and the vast portions of the domain where no storms occurred do not bias the Brier Scores. The predictability of each event is also assessed by comparing the overall skill of CTRL forecasts initialized at two lead times prior to the convective event (e.g., 1200 and 1800 UTC initialized forecasts) to forecasts made by the EXP system; the 1500 UTC-initialized forecasts are omitted in this evaluation. The forecasts are evaluated subjectively as well in an effort to describe how StesoNet observations augment the UH forecast probabilities.

Additionally, objective measures of predictability are employed through the application of ensemble sensitivity analysis (ESA; Ancell and Hakim 2007; Hakim and Torn 2008; Torn and Hakim 2008). ESA is a regression tool that takes a vector of scalar forecast metric values within a specified region (e.g., composite reflectivity) and regresses these estimates back to earlier model states to estimate how small changes in the initial conditions will influence the forecast. ESA has been used extensively at synoptic scales (e.g., Torn and Hakim 2009; Garcies and Homar 2009, 2010; Torn 2010; Chang et al. 2013; McMurdie and Ancell 2014; Torn and Cook 2013; Zheng et al. 2013; Xie et al. 2013; Brown and Hakim 2015; Ancell 2016) and has become increasingly popular for convective applications (e.g., Bednarczyk and Ancell 2015; Torn and Romine 2015; Hill et al. 2016; Berman et al. 2017; Torn et al. 2017; Hill et al. 2020). In this work, ESA-based observation targeting methods (e.g., Ancell and Hakim 2007; Hill et al. 2020) are applied to assess where hypothetical temperature and moisture observations would contribute to reducing CTRL forecast metric variance later in the forecast, and these objectively-defined regions are compared to where the StesoNet sampled during convective events. The ESA-based targeting formula for UH forecast variance change

$(\delta\sigma_{i,UH}^2)$ is defined by

$$\delta\sigma_{i,UH}^2 = \frac{\frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}_i - \bar{x}_i)(\mathbf{R} - \bar{\mathbf{R}})}{\frac{1}{N-1} \sum_{n=1}^N (\mathbf{x}_i - \bar{x}_i)^2 + \sigma_{ob}^2}, \quad (3)$$

where the numerator in (3) represents the covariance between a state variable (e.g., 2-m temperature) vector \mathbf{x}_i at grid point i and forecast metric vector \mathbf{R} and the denominator is a summation of the variance in \mathbf{x} and a hypothetical observation variance (σ_{ob}^2), typically defined by instrument error characteristics; the observation variance is set to 1 K for both temperature and dewpoint hereafter.

For the ESA, UH is used as the forecast metric and the forecast region used to calculate ensemble estimates is objectively selected based on UH ensemble variance. UH is particularly useful as a forecast metric since it is generally regarded as a good predictor of strong rotating updrafts and various severe hazards (e.g., Sobash et al. 2011). Following Torn and Romine (2015), the forecast metric region is determined by considering all grid points that equal or exceed 60% of the maximum UH standard deviation over a period of interest that represents large ensemble variability; the objective procedure is described below. This objective region-selection procedure eliminates subjectivity and focuses the ensemble sensitivity analysis on where the forecast is most uncertain, and has been shown to work well for convective regimes (e.g., Torn and Romine 2015; Berman et al. 2017; Torn et al. 2017).

The maximum UH of each ensemble member forecast across a 3-hour time window is smoothed spatially with a Gaussian kernel and a spatial length scale of 42 km. Ensemble standard deviation is calculated from the smoothed UH member fields and a region is demarcated where the standard deviation exceeds 60% of the maximum standard deviation; this is the response region. The response vector \mathbf{R} for ESA calculations is the mean 3-hr maximum UH over the response region of each ensemble member's forecast. Torn et al. (2017) noted in their study that the sensitivity signals were relatively insensitive to the time window and manner of smoothing for the response variable, which may have altered the size of the response region. The ESA-based observation targeting regions are calculated with respect to hypothetical 2-m temperature and dewpoint observations. The ESA predictability analysis identifies target regions within the CTRL forecast system that are deemed critical to forecast error growth. Combined with the knowledge of the StesoNet domain, the target regions can be evaluated to determine potential reasons why EXP forecasts are improved or degraded, and in particular if the StesoNet domain happened to sample a sensitive forecast region that propagated through the stationary network.

3. Case Analysis

Four retrospective cases are evaluated herein – two each from the 2017 and Meso18-19 field programs – in which

StesoNet observations are included and withheld from data assimilation with the HRRRE. The four IOPs are associated with events and forecasts beginning on 22 April 2017 (22Apr), 30 April 2017 (30Apr), 23 February 2019 (23Feb), and 14 March 2019 (14Mar). Both an objective and subjective evaluation is incurred to assess forecast improvement or degradation from assimilating StesoNet observations and their role in augmenting predictability of these severe weather events.

a. Near-surface errors

Averaged RMSE is calculated across the domain (e.g., Fig. 3) for the CTRL and EXP ensemble analyses. Analyses are generated through hourly data assimilation from 1000 – 1800 UTC for each case and the resulting CTRL RMSE is subtracted from the EXP RMSE for temperature (Table 1) and dewpoint (Table 2). Across all analyses and cases, RMSE differences are smaller than 0.5 K for both temperature and dewpoint. However, RMSE differences for a number of assimilation cycles, while small, are statistically significant (bolded magnitudes in Tables 1 and 2). Analysis temperature RMSE is consistently reduced the most in the 23Feb and 14Mar cases, while analysis dewpoint RMSE is reduced most significantly in the 23Feb and 30Apr cases. Both the 23Feb and 14Mar cases exhibit more frequent cycles where the RMSE differences are statistically significant (e.g., 14Mar from 1000 to 1700 UTC, Table 1) than the other two cases.

The spatial distribution of analysis errors as well as analysis differences (EXP - CTRL) are examined through dot plots at select analysis times (Figs. 3 – 6). StesoNet observations are able to adequately adjust the ensemble analyses towards the observations (blue dots in e.g., Figs. 3a, 4d, 5c, 6a), with minimal instances of error increases (red dots in e.g., Fig. 6b), particularly at early analysis cycles. Large swaths of ensemble mean temperature and dewpoint analysis changes are evident across the StesoNet domain, indicating StesoNet observation assimilation is adjusting the mesoscale environment (e.g., Figs. 3a,b,h, 4c-d, 5a-d). At later analysis times, however, the combined effects of multiple assimilation cycles and developing convection contribute to more frequent small-scale analysis errors between the CTRL and EXP simulations (e.g., Figs. 4e-h, 5e-h, 6g,h); the StesoNet is not able to correct small-scale errors due to, for example, developing cold pools. For example, two StesoNet locations and one non-StesoNet location in northern Alabama at 1800 UTC in 30Apr (red rectangle in Fig. 4g) have lower temperature errors within an advancing cold pool due to StesoNet assimilation, while the same stations have both increases and decreases in dewpoint errors at the same time (red rectangle in Fig. 4h). In some instances, small-scale errors are present at early analysis cycles as well (red rectangles in Figs. 3b,d, Figs. 6c,d), which may be more indicative of poor representation

of the near-surface and boundary layer environment in the ensemble simulations before daytime heating commences and the boundary layer deepens.

In the 22Apr case, StesoNet observations reduce analysis dewpoint in southern Tennessee and simultaneously increase dewpoints in north-central Alabama (Fig. 3b). At 1400, 1600, and 1800 UTC, the dewpoint increase is more robust across northern Alabama, and it remains quasi-stationary (Figs. 3d,f,h). In this instance, the StesoNet observations appear to be correcting a persistent low dewpoint bias in the CTRL analyses; interestingly, the persistent adjustment in the background analysis does not necessarily translate to improved AEs at 1400 UTC (Fig. 3b), but does reduce errors at 1800 UTC in northwestern Alabama (Fig. 3h). This result further suggests the assimilation of StesoNet observations adjusts the mesoscale environment more than small scales, which are also heavily influenced by the local topography (Fig. 2). Similar quasi-stationary analysis adjustments are seen in the 23Feb case (Figs. 5a-d), which gradually shrink with later analysis cycles (Figs. 5e-h). In contrast, analysis differences at 1400 UTC in the 30Apr case (Figs. 4c,d) propagate northward into northern TN in later assimilation cycles (Figs. 4e-h); the assimilated StesoNet observations are having a downstream impact in future assimilation cycles.

Select analyses at 1200, 1500, and 1800 UTC are used to initialize ensemble free forecasts (i.e., no data assimilation after forecast begins) for each IOP. As in the above analysis, ensemble mean forecasts of temperature and dewpoint initialized from the CTRL analysis are subtracted from identical forecasts initialized from the EXP analyses; the CTRL AEs are also subtracted from EXP AEs to evaluate how forecast error changes when StesoNet observations are assimilated. Qualitatively, the spatial distribution of forecast errors and environment differences are complex after a few hours of forecast integration, mainly due to convection-induced cold pools (not shown). The domain-wide RMSEs are averaged across each forecast to provide a quantitative assessment of StesoNet observation impacts, which largely averages out small-scale errors (Fig. 7). Forecast RMSEs are reduced for both 2-m temperature and 2-m dewpoint in 22Apr when StesoNet observations are assimilated (Figs. 7a,b) across most forecast hours, and differences remain statistically significant through 2300 UTC for both the 1200 and 1500 UTC-initialized forecasts. On the other hand, only the 1500 and 1800 UTC initialized forecasts for 30Apr have reduced 2-m temperature forecast errors at multiple hours between 1800 - 2300 UTC (Fig. 7c); ensemble mean dewpoint forecast errors remain relatively unchanged in 30Apr, except for a few early hours in the 1200 UTC-initialized forecast (Fig. 7d). Both temperature and dewpoint forecast errors increase beyond 1400 UTC in the 23Feb case (Fig. 7e,f), even though early forecast hours just after forecast initialization have reduced errors.

The 14Mar case features initial reductions in 2-m temperature forecast errors for 1200 and 1500 UTC forecast initializations, which slowly erode later into the forecast (Fig. 7g). In contrast, dewpoint errors increase the first few forecast hours from the 1200 UTC-initialized forecast, but are primarily decreased in the 1800 UTC-initialized forecast (Figs. 7h).

b. Forecast distributions

Changes in ensemble forecast distributions of UH are inspected to evaluate how StesoNet observations are impacting the spatial and temporal placement of severe storms and associated hazards. Only the 1800 UTC-initialized forecasts from each IOP are considered herein for brevity. Initially in 22Apr, StesoNet observations contributed to reducing probabilities of UH in northeast Mississippi at 2000 UTC where one tornado and two instances of severe hail were reported in the previous hour (Fig. 8b). This prominent area of reduced probability propagated through the forecast period (Figs. 8c-f) and seemingly missed all LSRs associated with an individual, prolonged severe storm (not shown). In contrast, at 2200 and 2300 UTC the StesoNet observations contributed to increasing forecast probabilities of UH in northeast Tennessee and along the Alabama-Tennessee border (Fig. 8d,e) which coincided with severe hail and wind reports. Brier Scores are also calculated across relevant portions of the domain where UH is forecast by either the EXP or CTRL simulation or where LSRs were recorded (e.g., purple outline in Fig. 8a). Brier Scores across all forecast times illustrate that CTRL and EXP forecast skill are nearly indistinguishable, and StesoNet observations do not appear to have a significant impact on aggregate UH forecast skill.

The 30Apr case features a number of similarities as the 22Apr case regarding changes to the forecast distributions. At 2000 UTC, LSRs were scattered across northern Alabama and central Tennessee, coincident with areas of reduced and enhanced probabilities, respectively, due to StesoNet assimilation (Fig. 9). The area of enhanced probabilities propagates north-northeastward through the forecast period (Figs. 9b-d) and tracks with a pocket of continuous LSRs; capturing these reports in the forecast distribution has some impact on Brier Scores, but the magnitude is small (0.01 skill difference). The area of reduced UH probabilities in northern Alabama at 2000 UTC (Fig. 9a) also moves northward, as the LSRs wane by 2200 UTC (Figs. 9b,c). At later forecast hours (e.g., 0000 UTC), the assimilation of StesoNet observations drastically reduces probabilities of severe hazards in northeast Alabama and east-central Tennessee, where no LSRs were recorded; the StesoNet observations helped to remove any false alarm in these areas. These forecast improvements are also evident in the surface-based forecast errors (Figs. 7c,d). However, increases in UH probabilities in southern Alabama

and Mississippi likely offsets the probability decreases in Alabama and Tennessee, resulting in minimal changes to forecast skill scores.

23Feb featured fewer LSRs across the StesoNet domain, but StesoNet observation influence is felt across much of the southeast United States in the ensemble forecast distributions. Two hours after forecast initialization (2000 UTC), UH probability changes in northwest Mississippi and southwest Tennessee highlight a shift of convection eastward, induced by the StesoNet observation assimilation (Fig. 10a). This forecast probability dipole change propagates northeastward (Figs. 10b), but only one severe wind report was recorded in this region through the IOP (see Fig. 10c). On the other hand, western-central Mississippi featured a number of severe storms as well as reports (e.g., Fig. 10b). However, LSRs frequently occurred in areas of both increased and reduced forecast probabilities (e.g., Fig. 10c), indicating relatively little skill in delineating areas with increased severe hazard threats. The lack of skill is also manifest in nearly identical Brier Scores for CTRL and EXP across all forecast times. Additionally, forecast probabilities increase across northern Alabama at later forecast hours (e.g., Fig. 10e), where only a few LSRs were recorded (Fig. 10f).

As in the other IOPs, 14Mar features a number of LSRs that collocate with areas of enhanced and reduced forecast UH probabilities (Fig. 11). At 2000 UTC, LSRs in eastern Mississippi and western Alabama exist in between the positive and negative probability differences (Fig. 11a), whereas an hour later LSRs in Alabama exist primarily in areas of increased forecast probabilities (Fig. 11b). As storms move eastward, reports follow suit and forecast UH probabilities do not accurately change via StesoNet data assimilation to highlight the enhanced risks of severe storm hazards (Fig. 11c). Increased UH probabilities in central Alabama at 2300 UTC correspond to a reported tornado (Fig. 11d). By 0000 and 0100 UTC, LSRs have outpaced the forecast probabilities, suggesting storms have moved quicker eastward than either the CTRL or EXP ensemble forecasts suggest (Fig. 11e,f). The Brier Scores at 0000 UTC demonstrate that skill suffers due to StesoNet assimilation (0.1367 to 0.1498) as LSRs occur out ahead of the convection in the EXP forecast.

c. Predictability

For each IOP, separate forecasts are initialized at 1200 and 1800 UTC in both the CTRL and EXP prediction systems and respective forecast probabilities of UH at 2300 UTC are compared (Figs. 12-15). In 22Apr, the 1200 UTC initialized CTRL forecast produces a subjectively poor forecast across Mississippi, Alabama, and Tennessee (Fig. 12a); forecast UH probabilities were generally too far west compared to LSRs. The inclusion of six more data assimilation cycles in the CTRL prediction system helps

to constrain the forecast probabilities across eastern Tennessee and northern Alabama, better aligning the highest forecast probabilities with LSRs (Fig. 12b), and correspondingly the Brier score lowers from 0.1181 to 0.1161. The 1200 and 1800 UTC initialized EXP forecasts at 2300 UTC are qualitatively similar to the CTRL counterparts (c.f. Figs. 12a,b to Figs. 12c,d), with a better forecast resulting when more assimilation cycles are considered (Fig. 12d). Despite the CTRL and EXP forecast similarities, the greatest skill comes from the 1800-UTC EXP forecast initialization (0.105 skill score).

LSRs in northern Tennessee and southern Kentucky in the 30Apr case are completely missed by the 1200 UTC-initialized CTRL and EXP forecasts (Fig. 13a,c), resulting in similar forecast skill (Brier Scores of 0.1127 and 0.1125 for CTRL and EXP forecasts, respectively). Both forecasts erroneously place a large swath of probabilities in southern Mississippi and Alabama where only one report was recorded (Fig. 13a,c). CTRL and EXP forecasts initialized six hours later consolidate probabilities in southern Alabama while increasing the maximum probability, and add UH probabilities in south-central Tennessee, just south of the Tennessee wind reports (Fig. 13b,d). The 1800 UTC-initialized EXP forecast skill is improved over the CTRL forecast (compare 0.086 to 0.0935) in part due to reduced false alarm and lower probabilities in northeast and southern Alabama.

The distributions of forecast UH probabilities in 23Feb are again more similar across CTRL and EXP simulations compared to early and later-initialized forecasts (Fig. 14). CTRL and EXP forecasts placed UH probabilities too far west in Mississippi (Fig. 14a,c), and had a relative probability minimum near two tornado reports in northeast Mississippi. Additionally, both forecast simulations issued probabilities in central Tennessee where no LSRs were reported. The Brier Score of the EXP forecast is slightly smaller than the CTRL forecast, likely due to less erroneous probabilities in western Tennessee and slightly lower probabilities overall. The 1800 UTC-initialized CTRL and EXP forecasts have reduced skill, as the forecasts reduce probabilities in MS but increase UH probabilities across Tennessee in areas where no LSRs were reported. In fact, more LSRs were missed by the 1800-UTC initialized CTRL and EXP forecasts compared to those forecasts initialized 6 hours prior when considering where the five percent probability contours lie.

Qualitatively, 1800 UTC-initialized forecasts for the 14Mar case appear better than their 1200-UTC counterparts (c.f. Figs. 15b,d vs. 15a,c), but statistically they verify worse; Brier Scores increase as a result of more assimilation cycles. The decrease in skill is likely due to increased probabilities over areas that had no LSRs, and an expansion of lower probabilities into southern Alabama (Fig. 15b). Furthermore, it is still apparent that

the greatest forecast changes occur between forecast initializations and not the inclusion of StesoNet observations (c.f. Figs. 15a,b vs. 15c,d). The inclusion of StesoNet observations for 1200 UTC-initialized forecasts has minimal impact on the forecast distribution (Fig. 15c), slightly decreasing probabilities in southern Mississippi and northern Tennessee. On the other hand, adding more assimilation cycles and running new forecasts at 1800 UTC drastically reduces the areal coverage of forecast UH probabilities but increases the maximum probabilities across Alabama in both the CTRL and EXP forecasts (Figs. 15b,d).

Across the four cases, despite the inconsistent changes to forecast skill between different forecast initializations, the best skill came from the EXP simulations in each case. Moreover, the intermixed small-scale forecast successes and failures between CTRL and EXP further suggest that the StesoNet observations cannot consistently improve small-scale errors. On the other hand, StesoNet assimilation always contributed to better forecast skill, reducing Brier Scores in all cases and forecasts initialized at 1200 and 1800 UTCi.

4. Ensemble Sensitivity Analysis

As discussed previously, ESA statistically and dynamically relates changes to a response metric (e.g., $UH > 50 m^2 s^{-2}$) to prior model states, such that a prescribed perturbation in a forecast variable (e.g., 2-m temperature) can be used to estimate the change in response variance (i.e., observation targeting). In the four IOPs examined herein, ESA-based observation targeting fields are produced and examined to determine where the UH forecasts are sensitive to low-level thermodynamics.

In the 22Apr case, the primary targeting regions for both 2-m temperature and dewpoint exist in central Tennessee and southern Kentucky as much as eight hours prior to the UH forecast response, valid at 2200 UTC (Fig. 16). These target regions appear tied to earlier convection (gray shading in Fig. 16a) as well as the pre-convective environment for storms later in the afternoon (e.g., Fig. 12a). ESA does identify localized regions in northwest Alabama in conjunction with the StesoNet domain where dewpoint observations at early forecast times could provide significant value towards reducing UH forecast uncertainty later in the forecast period (Fig. 16d-f), but the regions are temporally weakly correlated and generally small scale. 1800 UTC-initialized forecasts, which incorporate observations that may have sampled these sensitive regions, are generally improved in this region of eastern Tennessee (Fig. 12). A few pockets of target regions develop later in the forecast period (2-4 hours prior to the response) in northern Alabama and Georgia (not shown). In contrast to the near-surface targeting fields, ESA identifies coherent regions for hypothetical 850-hPa temperature observations along a northeast-southwest oriented axis extending from central

Tennessee to southwest Mississippi (Fig. 16g-i) in advance of the afternoon convection. The 850-hPa temperature targeting region is identified as early as 1300 UTC (not shown) and propagates through the southeast U.S. during the full forecast period, indicating perhaps that near-surface environment is not the most important aspect of the forecast to sample in order to reduce UH forecast uncertainty for this case.

The evolution of convection in 30Apr offers a different perspective from the 22Apr case in regards to ESA identified target regions. The majority of forecast variability exists along the southern periphery of a QLCS in eastern Mississippi and western Alabama at 1900 UTC (green contour in Fig. 17), and ESA suggests the most important areas to sample 2-m temperature and dewpoint are well behind the convective line at all prior forecast times (Fig. 17). This large region is likely tied to the cold pool produced by the QLCS, which is dynamically related to the strength of convection. As in the 22Apr case, localized target regions exist within the StesoNet domain early in the forecast period (Fig. 17a,b), but they are not sustained at later hours, propagating away from the observing domain. It is likely that improved UH forecasts along the QLCS (Fig. 13) are a result of the conventional observational network sampling the sensitive regions. ESA-identified targets of 850-hPa temperature observations again feature coherent signals for propagating regions out ahead of convection across Alabama (e.g., Fig. 17g,h) and tied to the back side of the QLCS with a well-defined region propagating from western Mississippi at 1500 UTC (Fig. 17g) to western Tennessee by 1900 UTC (not shown).

As in 22Apr and 30Apr, low-level observation targeting regions for 2-m temperature and dewpoint are relatively scarce and small-scale in the 23Feb case (Fig. 18). Identified areas for targeting across the StesoNet domain are temporally and spatially incoherent; a spatially coherent target region in Louisiana at 1800 UTC (Fig. 18c) persists throughout the forecast period and propagates northeastward (not shown), but is never sampled by the StesoNet array. A variety of target areas for both near-surface variables exist, again, behind the developing convection in northwestern Mississippi (e.g., Fig. 18c,f), with few areas identified in the pre-convective environment across Alabama and Georgia (e.g., Fig. 18a,b,d,e). ESA-identified regions to target 850-hPa temperature observations again are anchored to the developing convection (Fig. 18g-i), translating with the storms in central Mississippi at 1800 UTC (Fig. 18i) to northwest Alabama by 0000 UTC (not shown). Noisy 850-hPa temperature target regions exist out ahead of the convection in central and southern Alabama as well (Fig. 18h,i), which originate over the warm waters of the Gulf of Mexico (not shown).

Even fewer near-surface observation targeting regions exist for 14Mar across Alabama, Tennessee, and the StesoNet VORTEX-SE domain (Fig. 19) compared to the

other IOPs. 2-m temperature and dewpoint target areas at 1400 and 1600 UTC (Fig. 19a,b,d,e) in western Mississippi exist primarily along and behind the developing convective line. Pre-convective target regions develop by 1800 UTC (Fig. 19c,f) west and southwest of the StesoNet domain. Later in the forecast, 2-m temperature targets are enhanced across the entire response region, while relatively no target areas are defined for 2-m dewpoint temperature (not shown). Aloft, ESA identifies ribbons of targeting regions for 850-hPa temperature across Arkansas over a number of forecast hours (Fig. 19g-i). A number of localized targeting regions for 850-hPa temperature appear tied to individual storms across Alabama (Fig. 19h,i), which suggests sampling thermodynamics aloft to improve UH forecasts would be particularly difficult. Similarly small target areas exist in southern Mississippi and eastern Arkansas (Fig. 19i), and a different region develops in northeastern Alabama concurrent with the response time at 2100 UTC (not shown).

5. Summary and Discussion

Four IOPs from the VORTEX-SE 2017 and Meso18-19 field campaigns were selected to evaluate the impact of near-surface observations on convection forecasts. Texas Tech University deployed StickNet platforms in a stationary, mesoscale network across northern Alabama and southern Tennessee, sampling temperature, humidity, and pressure at 10Hz frequency during each IOP (i.e., the StesoNet). Observations across 16 and 24 stations in 2017 and Meso18-19, respectively, are retrospectively assimilated with the experimental HRRRE at hourly frequency, with forecasts initialized from these analyses and compared against forecasts with no StesoNet data assimilation.

Across nearly all analysis times and IOPs, StesoNet observations reduce analysis errors of 2-m temperature and dewpoint across the domain, with error reductions generally less than 0.25K. A number of the analysis RMSE reductions are statistically significant despite their small magnitudes. Spatially, the StesoNet observations are routinely adjusting ensemble analyses around the StesoNet domain in early data assimilation cycles. Generally, these mesoscale adjustments are positive, in that absolute errors are also reduced. However, small-scale errors are present even within some of the mesoscale environmental adjustments. In a number of IOPs, these localized areas of improved errors remain stationary, indicating the propensity for the HRRRE system to be routinely biased and the StesoNet data are consistently trying to correct these biases. In other cases, the error improvements propagate downstream with later analyses, and subsequently impact other portions of the domain. EXP forecasts of temperature and dewpoint are improved in some instances, and

degraded in others. The 22Apr case saw the most consistent forecast improvements, while the 23Feb case exhibited forecast degradations that were statistically significant.

When evaluating UH forecasts, EXP simulations do not consistently shift ensemble forecast probabilities where severe hazards were observed; the collocation of increased probabilities with storm reports at some forecast hours is replaced by increased probabilities in areas where no reports existed in following hours. In many instances, EXP forecasts reduce probabilities of UH in areas where severe reports were recorded. UH probability changes are tied to reflectivity objects, which are highly sensitive to changing initial conditions induced by new observations. Furthermore, forecast probability distributions of UH are more sensitive to initialization time than the assimilation of StesoNet observations, with the most drastic changes in forecast distributions arising from extra assimilation cycles and additional conventional observations. However, StesoNet assimilation always had a positive impact on forecast skill (i.e., reducing Brier Scores) regardless of initialization time. It should be noted as well that despite the largest changes in forecast distributions coming from different initializations, later-initialized forecasts did not necessarily always have improved UH forecasts. 23Feb and 14Mar featured poorer forecasts from 1800 UTC-initialized forecasts, as higher probabilities were issued by the forecast systems in areas of no reports. In general, StesoNet observations are able to improve the mesoscale environment preceding and during deep convection, but these improvements do not translate to storm-anchored UH probabilities.

Ensemble sensitivity analysis-based observation targeting methods applied to the CTRL ensemble forecasts initialized at 1200 UTC reveal that targeting regions of 2-m temperature and dewpoint across the StesoNet domain are infrequent and short-lived. At most, target regions propagate through the StesoNet domain in less than three hours for these cases. Weak impact of StesoNet observations on forecast UH distributions may be partially explained by ESA-identified target regions that were infrequent over the StesoNet domain. Generally, there was a lack of ESA-identified targeting regions in the pre-convective environment, and when pre-convective target areas were identified, they rarely occurred across the StesoNet domain. Furthermore, the strongest signal of near-surface target areas occurred behind the convection, dynamically tied to the cold pools being produced by the convection. ESA was also applied to evaluate targeting regions aloft for 850-hPa temperature, which identified spatially and temporally coherent regions directly tied to the developing convection. These results indicate that near-surface sampling may not have been the most important facet of the environment to improve the convection forecasts, and sampling the environment aloft may have yielded additional forecast improvements. Additionally, the assimilation of conventional

observations in later assimilation cycles was likely sufficient to improve forecasts since these observations sampled sensitive regions.

The inconsistent storm-scale improvements in both near-surface environmental fields as well as storm-centric UH forecasts could likely be addressed by considering improvements to the data assimilation configuration. Planned future work will investigate to what extent additional StesoNet observations assimilated per cycle (e.g., ten observations over the assimilation window) could enhance storm-scale forecasts as well as the impact of sub-hourly data assimilation. Alternative UH-forecast verification methods should also be explored, since it is difficult to accurately project hourly forecast skill with limited LSRs. Future efforts may benefit from utilizing sub-daily or hourly practically perfect (PPER) analyses (Gensini et al. 2020) to verify forecasts. Even though PPER fields are similarly generated from LSRs, they provide a probabilistic baseline to verify probabilistic forecasts, which may be a more appropriate measure of forecast skill compared to Brier Scores.

Through the analysis of these four IOPs, it is evident that a mesoscale, portable array of near-surface sampling has the potential to impact and improve convection forecasts within the southeast United States. However, even though there were reductions in near-surface analysis and forecast errors due to assimilating StesoNet observations, it is plausible that the near-surface environment was not the most sensitive area for these particular cases. The evolution of the planetary boundary layer (PBL), including the transportation of heat, moisture, and momentum, is a critical factor in the development and intensification of convection, and it is often insufficiently modeled due to the small-scale processes inherent within (e.g., Cohen et al. 2015). It makes sense that in situ observations above the surface and within the PBL, where the regular observing network is particularly lacking, would provide substantial benefit to convection forecasts within this geographical region. Additionally, these IOP case studies represent only a subsample of events, and clearly do not fully determine the value of StesoNet observations for southeast United States convection events. Furthermore, a larger set of convective events and forecast cases would allow for a more robust assessment of forecast reliability and resolution, which can only be gleaned from a sufficiently large set of forecasts. Continued deployment of portable mesoscale networks in future VORTEX-SE field campaigns (e.g., the Propagation, Evolution, and Rotation in Linear Systems (PERiLS) project) will allow for a more exhaustive evaluation of the utility of such a network in the southeast as well as other areas of the country.

Acknowledgments. We would like to acknowledge the individuals that contributed to the deployment, testing, and

maintenance of the StickNet platforms during VORTEX-SE: Dr. Eric Bruning, Dr. Vanna Chmielewski, Abby Hutson, Alex Schueth, and Jessica McDonald. We are also indebted to the many landowners who supported our science objectives for over nine cumulative months. We greatly appreciate the sharing of analysis and plotting code by Dr. Patrick Skinner. Computing support was generously provided by the National Oceanic Atmospheric Administration (NOAA). This work is supported by NOAA awards NA17OAR4590206 and NA18OAR4590318.

Data availability statement. Forecast simulation output for all cases described in the manuscript are archived on NOAA computers as netCDF files and available upon request. Raw and pre-processed StickNet observations from VORTEX-SE are stored on external hard drives at Texas Tech University and are available upon request from Dr. Christopher Weiss (chris.weiss@ttu.edu). Local storm reports are available from the web at <https://mesonet.agron.iastate.edu/request/gis>. Analysis scripts are housed in private repositories on GitHub, and can be shared upon request.

References

- Alexander, C., and Coauthors, 2016: The hourly updated US High-Resolution Rapid Refresh (HRRR) storm-scale forecast model. *EGU General Assembly Conference Abstracts*, EPSC2016-11 044, EGU General Assembly Conference Abstracts.
- Alford, A. A., M. I. Biggerstaff, G. D. Carrie, J. L. Schroeder, B. D. Hirth, and S. M. Waugh, 2019: Near-surface maximum winds during the landfall of hurricane harvey. *Geophysical Research Letters*, **46**, 973–982, doi:<https://doi.org/10.1029/2018GL080013>.
- Ancell, B. C., 2016: Improving High-Impact Forecasts through Sensitivity-Based Ensemble Subsets: Demonstration and Initial Tests. *Weather and Forecasting*, WAF-D-15-0121.1, doi:10.1175/WAF-D-15-0121.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/WAF-D-15-0121.1?af=R>.
- Ancell, B. C., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Monthly Weather Review*, **135**, 4117–4134.
- Ashley, W. S., 2007: Spatial and temporal analysis of tornado fatalities in the United States: 1880–2005. *Wea. Forecasting*, **22**, 1214–1228, doi:10.1175/2007WAF2007004.1.
- Bednarczyk, C. N., and B. C. Ancell, 2015: Ensemble sensitivity analysis applied to a southern plains convective event. *Monthly Weather Review*, **143**, 230–249.
- Berman, J. D., R. D. Torn, G. S. Romine, and M. L. Weisman, 2017: Sensitivity of Northern Great Plains Convection Forecasts to Upstream and Downstream Forecast Errors. *Monthly Weather Review*, **145** (6), 2141–2163, doi:10.1175/mwr-d-16-0353.1.
- Brown, B. R., and G. J. Hakim, 2015: Sensitivity of intensifying Atlantic hurricanes to vortex structure. *Quarterly Journal of the Royal Meteorological Society*, **141**, 2538–2551, doi:10.1002/qj.2540.
- Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, **131**, 2394–2416.
- Chang, E. K. M., M. Zheng, and K. Raeder, 2013: Medium-range ensemble sensitivity analysis of two extreme pacific extratropical cyclones. *Monthly Weather Review*, **141**, 211–231.
- Chen, X., K. Zhao, J. Sun, B. Zhou, and W. C. Lee, 2016: Assimilating surface observations in a four-dimensional variational Doppler radar data assimilation system to improve the analysis and forecast of a squall line case. *Advances in Atmospheric Sciences*, **33**, 1106–1119, doi:10.1007/s00376-016-5290-0.
- Childs, S. J., R. S. Schumacher, and J. T. Allen, 2018: Cold-season tornadoes: Climatological and meteorological insights. *Weather and Forecasting*, **33**, 671–691, doi:10.1175/WAF-D-17-0120.1.
- Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous united states. *Wea. Forecasting*, **27**, 1235–1248.
- Clark, A. J., W. a. Gallus, and M. L. Weisman, 2010: Neighborhood-Based Verification of Precipitation Forecasts from Convection-Allowing NCAR WRF Model Simulations and the Operational NAM. *Weather and Forecasting*, **25**, 1495–1509, doi:10.1175/2010waf2222404.1.
- Clark, A. J., W. a. Gallus, M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Weather and Forecasting*, **24**, 1121–1140.
- Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A Review of Planetary Boundary Layer Parameterization Schemes and Their Sensitivity in Simulating a Southeast U.S. Cold Season Severe Weather Environment. *Weather and Forecasting*, 150224120634008, doi:10.1175/WAF-D-14-00105.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/WAF-D-14-00105.1>.
- Coniglio, M. C., S. M. Hitchcock, and K. H. Knopfmeier, 2016: Impact of assimilating pre-convective upsonde observations on short-term forecasts of convection observed during MPEX. *Monthly Weather Review*, **in press**, doi:10.1175/MWR-D-16-0091.1.
- Davis, J. M., and M. D. Parker, 2014: Radar climatology of tornadic and nontornadic vortices in high-shear, low-CAPE environments in the mid-atlantic and southeastern united states. *Weather and Forecasting*, **29**, 828–853, doi:10.1175/WAF-D-13-00127.1.
- Dixon, P. G., A. E. Mercer, J. Choi, and J. S. Allen, 2011: Tornado risk analysis: Is dixie alley an extension of Tornado alley. *Bull. Amer. Meteor. Soc.*, **92**, 433–441, doi:10.1175/2010BAMS3102.1.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of nwp: Explicit forecasts of convection using the weather research and forecasting (wrf) model. *Atmospheric Science Letters*, **5**, 110–117.
- Dowell, D. C., and Coauthors, 2021: The high-resolution rapid refresh (hrrr): An hourly updating convection permitting forecast model. part 1: Motivation and system description. *Wea. Forecasting*, **in preparation**.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10 143–10 162.
- Fernandez-Caban, P. L., and Coauthors, 2019: Observing Hurricane Harvey’s Eyewall at Landfall. *Bulletin of the American Meteorological Society*, **100**, 759–775, doi:10.1175/BAMS-D-17-0237.1.
- Garcies, L., and V. Homar, 2009: Ensemble sensitivities of the real atmosphere: application to Mediterranean intense cyclones. *Tellus A*, **61**, 394–406.

- Garcies, L., and V. Homar, 2010: An optimized ensemble sensitivity climatology of Mediterranean intense cyclones. *Natural Hazards and Earth System Science*, **10**, 2441–2450.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125**, 723–757.
- Geerts, B., and Coauthors, 2017: The 2015 plains elevated convection at night field project. *Bulletin of the American Meteorological Society*, **98**, 767–786.
- Gensini, V., and W. Ashley, 2011: Climatology of potentially severe convective environments from the north american regional reanalysis. *Electronic J. Severe Storms Meteor.*, **6**.
- Gensini, V. A., A. M. Haberlie, and P. T. Marsh, 2020: Practically Perfect Hindcasts of Severe Convective Storms. *Bulletin of the American Meteorological Society*, **101**, E1259–E1278, doi: 10.1175/BAMS-D-19-0321.1.
- Giammanco, I. M., J. L. Schroeder, F. J. Masters, P. J. Vickery, I. Krupar, Richard J., and J.-A. Balderrama, 2016: Influences on Observed Near-Surface Gust Factors in Landfalling U.S. Gulf Coast Hurricanes: 2004–2008. *Journal of Applied Meteorology and Climatology*, **55**, 2587–2611, doi:10.1175/JAMC-D-16-0053.1.
- Gunter, W. S., J. L. Schroeder, C. C. Weiss, and E. C. Bruning, 2017: Surface measurements of the 5 june 2013 damaging thunderstorm wind event near pep, texas. *Wind and Structures*, **24**, 185–204.
- Guyer, J. L., and A. R. Dean, 2010: Tornadoes Within Weak CAPE Environments Across the Continental United States. *25th Conference on Severe Local Storms*, 1.5.
- Ha, S.-Y., and C. Snyder, 2014: Influence of surface observations in mesoscale data assimilation using an ensemble Kalman filter. *Monthly Weather Review*, **142**, 1489–1508.
- Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tenant, 2017: The met office convective-scale ensemble, mogreps-uk. *Quarterly Journal of the Royal Meteorological Society*, **143**, 2846–2861.
- Hakim, G. J., and R. D. Torn, 2008: Ensemble synoptic analysis. *Synoptic-Dynamic Meteorology and Weather Analysis and Forecasting: A Tribute to Fred Sanders*, Vol. No. 55, American Meteorological Society, No. 55, Amer. Meteor. Soc., 147–161.
- Hill, A. J., C. C. Weiss, and B. C. Ancell, 2016: Ensemble Sensitivity Analysis for Mesoscale Forecasts of Dryline Convection Initiation. *Monthly Weather Review*, **144**, 4161–4182, doi:10.1175/MWR-D-15-0338.1.
- Hill, A. J., C. C. Weiss, and B. C. Ancell, 2020: Factors influencing ensemble sensitivity-based targeted observing prediction at convection-allowing resolutions. *Mon. Wea. Rev.*, **148**, 4497–4517.
- Hitchcock, S. M., M. C. Coniglio, and K. H. Knopfmeier, 2016: Impact of MPEX upsonde observations on ensemble analyses and forecasts of the 31 may 2013 convective event over Oklahoma. *Monthly Weather Review*, **144**, 2889–2913, doi:10.1175/MWR-D-15-0344.1.
- Horgan, K. L., D. M. Schultz, J. E. Hales Jr., S. F. Corfidi, and R. H. Johns, 2007: A five-year climatology of elevated severe convective storms in the united states east of the rocky mountains. *Wea. Forecasting*, **22**, 1031 – 1044.
- Jirak, I., A. Clark, B. Roberts, B. Gallo, and S. Weiss, 2018: Exploring the optimal configuration of the high resolution ensemble forecast system. *25th Conf. on Numerical Weather Prediction*, Amer. Meteor. Soc. Denver, CO.
- Jirak, I., S. Weiss, and C. Melick, 2012: The spc storm-scale ensemble of opportunity: Overview and results from the 2012 hazardous weather testbed spring forecasting experiment. *26th Conf. on Severe Local Storms*.
- Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. *Weather and Forecasting*, **31**, 297–327, doi:10.1175/WAF-D-15-0107.1.
- Kain, J. S., S. J. Weiss, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the wrf model for the prediction of severe convective weather: The spc/nssl spring program 2004. *Weather and Forecasting*, **21**, 167–181.
- Kain, J. S., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bulletin of the American Meteorological Society*, **94**, 1213–1225, doi:10.1175/BAMS-D-11-00264.1.
- King, J. R., M. D. Parker, K. D. Sherburn, and G. M. Lackmann, 2017: Rapid evolution of cool season, low-cape severe thunderstorm environments. *Weather and Forecasting*, **32**, 763–779, doi: 10.1175/WAF-D-16-0141.1.
- Kis, A. K., and J. M. Straka, 2010: Nocturnal tornado climatology. *Wea. Forecasting*, **25**, 545–561, doi:10.1175/2009WAF2222294.1.
- Klasa, C., M. Arpagaus, A. Walser, and H. Wernli, 2018: An evaluation of the convection-permitting ensemble cosmo-e for three contrasting precipitation events in switzerland. *Quarterly Journal of the Royal Meteorological Society*, **144**, 744–764.
- Knopfmeier, K. H., and D. J. Stensrud, 2013: Influence of mesonet observations on the accuracy of surface analyses generated by an ensemble Kalman filter. *Weather and Forecasting*, **28**, 815–841.
- Lee, T. R., M. Buban, D. D. Turner, T. P. Meyers, and C. B. Baker, 2019: Evaluation of the high-resolution rapid refresh (hrrr) model using near-surface meteorological and flux observations from northern alabama. *Wea. Forecasting*, **34**, 635–663.
- Lyza, A. W., T. A. Murphy, B. T. Goudeau, P. T. Pangle, K. R. Knupp, and R. A. Wade, 2020: Observed near-storm environment variations across the southern cumberland plateau system in northeastern alabama. *Mon. Wea. Rev.*, **148**, 1465–1482.
- Madaus, L. E., and G. J. Hakim, 2017: Constraining Ensemble Forecasts of Discrete Convective Initiation with Surface Observations. *Monthly Weather Review*, **145**, 2597–2610, doi:10.1175/MWR-D-16-0395.1.
- Manser, R. P., and B. C. Ancell, 2020: Verification of convection-allowing initial condition ensemble modeling systems with wrf-arw. *30th Conference on Weather Analysis and Forecasting (WAF)/26th Conference on Numerical Weather Prediction (NWP)*, Amer. Meteor. Soc. Boston, MA, poster 642.
- Markowski, P. M., N. T. Lis, D. D. Turner, T. R. Lee, and M. S. Buban, 2019: Observations of Near-Surface Vertical Wind Profiles and Vertical Momentum Fluxes from VORTEX-SE 2017: Comparisons to Monin–Obukhov Similarity Theory. *Mon. Wea. Rev.*, **147**, 3811–3824, doi:10.1175/MWR-D-19-0091.1.
- McDonald, J. M., and C. C. Weiss, 2020: Cold pool characteristics from tornadic quasi-linear convective systems and other convective modes

- observed during the vortex-se project. *Monthly Weather Review*, in prep.
- McMurdie, L. A., and B. Ancell, 2014: Predictability characteristics of land-falling cyclones along the North American west coast. *Monthly Weather Review*, **142**, 301–319.
- Nielsen, E. R., and R. S. Schumacher, 2016: Using convection-allowing ensembles to understand the predictability of an extreme rainfall event. *Monthly Weather Review*, **144**, 3651–3676, doi: 10.1175/MWR-D-16-0083.1.
- Rasmussen, E. N., J. M. Straka, R. Davies-Jones, C. A. Doswell, F. H. Carr, M. D. Elts, and D. R. MacGorman, 1994: Verification of the origins of rotation in tornadoes experiment: Vortex. *Bulletin of the American Meteorological Society*, **75**, 995–1006, doi:10.1175/1520-0477(1994)075<0995:VOTOOR>2.0.CO;2.
- Roberts, B., B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What Does a Convection-Allowing Ensemble of Opportunity Buy Us in Forecasting Thunderstorms? *Wea. Forecasting*, **35**, 2293–2316, doi:10.1175/WAF-D-20-0069.1.
- Schroeder, J. L., and C. C. Weiss, 2008: Integrating research and education through measurement and analysis. *Bull. Amer. Meteor. Soc.*, **89**, 793.
- Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Weather and Forecasting*, **30**, 1645–1654, doi:10.1175/WAF-D-15-0103.1.
- Schwartz, C. S., and R. A. Sobash, 2017: Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations. *Monthly Weather Review*, **145**, 3397–3418, doi:10.1175/MWR-D-16-0400.1.
- Sherburn, K. D., and M. D. Parker, 2014: Climatology and ingredients of significant severe convection in high-shear, low-CAPE environments. *Weather and Forecasting*, **29**, 854–877, doi:10.1175/WAF-D-13-00041.1.
- Sherburn, K. D., M. D. Parker, J. R. King, and G. M. Lackmann, 2016: Composite environments of severe and nonsevere high-shear, low-CAPE convective events. *Weather and Forecasting*, **31**, 1899–1927, doi:10.1175/WAF-D-16-0086.1.
- Skinner, P. S., C. C. Weiss, M. M. French, H. B. Bluestein, P. M. Markowski, and Y. P. Richardson, 2014: VORTEX2 observations of a low-level mesocyclone with multiple internal rear-flank downdraft momentum surges in the 18 May 2010 Dumas, Texas, supercell. *Mon. Wea. Rev.*, **142**, 2935–2960, doi:10.1175/MWR-D-13-00240.1.
- Skinner, P. S., C. C. Weiss, J. L. Schroeder, L. J. Wicker, and M. I. Biggerstaff, 2011: Observations of the surface boundary structure within the 23 May 2007 Perryton, Texas, supercell. *Monthly Weather Review*, **139** (12), 3730–3749.
- Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. part i: Storm classification and climatology. *Weather and Forecasting*, **27**, 1114–1135.
- Sobash, R. A., 2016: Severe Weather Prediction Using Storm Surrogates from an Ensemble Forecasting System. *Weather and Forecasting*, 255–271, doi:10.1175/WAF-D-15-0138.1.
- Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic Forecast Guidance for Severe Thunderstorms Based on the Identification of Extreme Phenomena in Convection-Allowing Model Forecasts. *Weather and Forecasting*, **26**, 714–728, doi:10.1175/WAF-D-10-05046.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/WAF-D-10-05046.1>.
- Sobash, R. A., and D. J. Stensrud, 2015: Assimilating surface mesonet observations with the EnKF to improve ensemble forecasts of convection initiation on 29 May 2012. *Monthly Weather Review*, 150529123037005, doi:10.1175/MWR-D-14-00126.1.
- Stensrud, D. J., N. Yussouf, D. C. Dowell, and M. C. Coniglio, 2009: Assimilating surface data into a mesoscale model ensemble: Cold pool analyses from spring 2007. *Atmospheric Research*, **93**, 207–220.
- Stensrud, D. J., and Coauthors, 2013: Progress and challenges with warn-on-forecast. *Atmospheric Research*, **123**, 2 – 16, doi:<https://doi.org/10.1016/j.atmosres.2012.04.004>. 6th European Conference on Severe Storms 2011. Palma de Mallorca, Spain.
- Tanamachi, R. L., S. J. Frasier, J. Waldinger, A. LaFleur, D. D. Turner, and F. Rocadenbosch, 2019: Progress toward characterization of the atmospheric boundary layer over northern Alabama using observations by a vertically pointing, s-band profiling radar during vortex-southeast. *Journal of Atmospheric and Oceanic Technology*, **36**, 2221–2246.
- Torn, R. D., 2010: Ensemble-based sensitivity analysis applied to African easterly waves. *Weather and Forecasting*, **25**, 61–78.
- Torn, R. D., 2014: The impact of targeted dropwindsonde observations on tropical cyclone intensity forecasts of four weak systems during PREDICT. *Monthly Weather Review*, **142**, 2860–2878.
- Torn, R. D., and D. Cook, 2013: The role of vortex and environment errors in genesis forecasts of Hurricanes Danielle and Karl (2010). *Monthly Weather Review*, **141**, 232–251.
- Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Monthly Weather Review*, **136**, 663–677.
- Torn, R. D., and G. J. Hakim, 2009: Initial condition sensitivity of western Pacific extratropical transitions determined using ensemble-based sensitivity analysis. *Monthly Weather Review*, **137**, 3388–3406.
- Torn, R. D., and G. S. Romine, 2015: Sensitivity of central Oklahoma convection forecasts to upstream potential vorticity anomalies during two strongly forced cases during MPEX. *Monthly Weather Review*, **143**, 4064–4087.
- Torn, R. D., G. S. Romine, and T. J. Galarneau, 2017: Sensitivity of Dryline Convection Forecasts to Upstream Forecast Errors for Two Weakly Forced MPEX Cases. *Monthly Weather Review*, **145**, 1831–1852.
- Wallace, Z. C., L. Keys-Mathews, and A. A. Hill, 2015: The role of experience in defining tornado risk perceptions: A case from the 27 April 2011 outbreak in rural Alabama. *Southeastern Geographer*, **55**, 400–416.
- Weisman, M. L., and Coauthors, 2015: The mesoscale predictability experiment (mpex). *Bulletin of the American Meteorological Society*, **96**, 2127–2149, doi:10.1175/BAMS-D-13-00281.1.
- Weiss, C. C., D. C. Dowell, J. L. Schroeder, P. S. Skinner, A. E. Reinhardt, P. M. Markowski, and Y. P. Richardson, 2015: A comparison of near-surface buoyancy and baroclinity across three VORTEX2 supercell intercepts. *Mon. Wea. Rev.*, **143**, 2736–2753, doi: 10.1175/MWR-D-14-00307.1.

Weiss, C. C., and J. L. Schroeder, 2008: Sticknet-a new portable, rapidly-deployable, surface observation system. *24th Conference on IIPS*.

Wheatley, D. M., K. H. Knopfmeier, T. a. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part 1: Radar data experiments. *Weather and Forecasting*, 151030091332000, doi:10.1175/WAF-D-15-0043.1, URL <http://journals.ametsoc.org/doi/abs/10.1175/WAF-D-15-0043.1>.

Wheatley, D. M., and D. J. Stensrud, 2010: The impact of assimilating surface pressure observations on severe weather events in a WRF mesoscale ensemble system. *Monthly Weather Review*, **138**, 1673–1694.

Wheatley, D. M., D. J. Stensrud, D. C. Dowell, and N. Yussouf, 2012: Application of a WRF mesoscale data assimilation system to spring-time severe weather events 2007–09. *Monthly Weather Review*, **140**, 1539–1557.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089.

Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second verification of the origins of rotation in tornadoes experiment: Vortex2. *Bulletin of the American Meteorological Society*, **93**, 1147–1170, doi:10.1175/BAMS-D-11-00010.1.

Xie, B., F. Zhang, Q. Zhang, J. Poterjoy, and Y. Weng, 2013: Observing strategy and observation targeting for tropical cyclones using ensemble-based sensitivity analysis and data assimilation. *Monthly Weather Review*, **141**, 1437–1453.

Xue, M., and Coauthors, 2007: Caps realtime storm-scale ensemble and high-resolution forecasts as part of the noaa hazardous weather testbed 2007 spring experiment. *22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 3B.1.

Zachry, B. C., J. L. Schroeder, A. B. Kennedy, J. J. Westerink, C. W. Letchford, and M. E. Hope, 2013: A Case Study of Nearshore Drag Coefficient Behavior during Hurricane Ike (2008). *Journal of Applied Meteorology and Climatology*, **52**, 2139–2146, doi:10.1175/JAMC-D-12-0321.1.

Zheng, M., E. K. M. Chang, and B. A. Colle, 2013: Ensemble Sensitivity Tools for Assessing Extratropical Cyclone Intensity and Track Predictability. *Weather and Forecasting*, **28**, 1133–1156, doi:10.1175/WAF-D-12-00132.1.

TABLE 1. Root mean squared error differences between CTRL and EXP ensemble analyses of 2-m temperature from 1000 to 1800 UTC for each case calculated across 33–36° N and 83–90° W. Bolded differences are statistically significant at the 95% confidence level using bootstrapping with 1,000 re-samples of each analysis' sample error distribution. Negative values represent lower errors in EXP simulations.

Cases	Hour (UTC)								
	10	11	12	13	14	15	16	17	18
22Apr	-0.03	-0.05	-0.05	-0.05	-0.03	-0.02	-0.05	0.01	-0.08
30Apr	-0.02	-0.03	-0.02	-0.02	-0.03	-0.03	-0.03	-0.13	-0.33
23Feb	0.01	-0.05	-0.11	-0.13	-0.19	-0.18	-0.15	-0.05	0.0
14Mar	-0.03	-0.04	-0.08	-0.08	-0.08	-0.07	-0.05	-0.07	0.01

TABLE 2. Root mean squared error differences between CTRL and EXP ensemble analyses of 2-m dewpoint temperature from 1000 to 1800 UTC for each case calculated across 33–36° N and 83–90° W. Bolded differences are statistically significant at the 95% confidence level using bootstrapping with 1,000 re-samples of each analysis' sample error distribution. Negative values represent lower errors in EXP simulations.

Cases	Hour (UTC)								
	10	11	12	13	14	15	16	17	18
22Apr	0.01	-0.06	-0.04	-0.05	-0.03	-0.06	-0.1	-0.05	-0.2
30Apr	0.02	0.01	-0.03	-0.11	-0.14	-0.07	0.06	-0.12	-0.07
23Feb	-0.01	-0.01	-0.12	-0.16	-0.23	-0.15	-0.05	-0.02	0.05
14Mar	0.0	-0.03	0.01	0.02	-0.01	0.0	-0.1	-0.12	-0.1

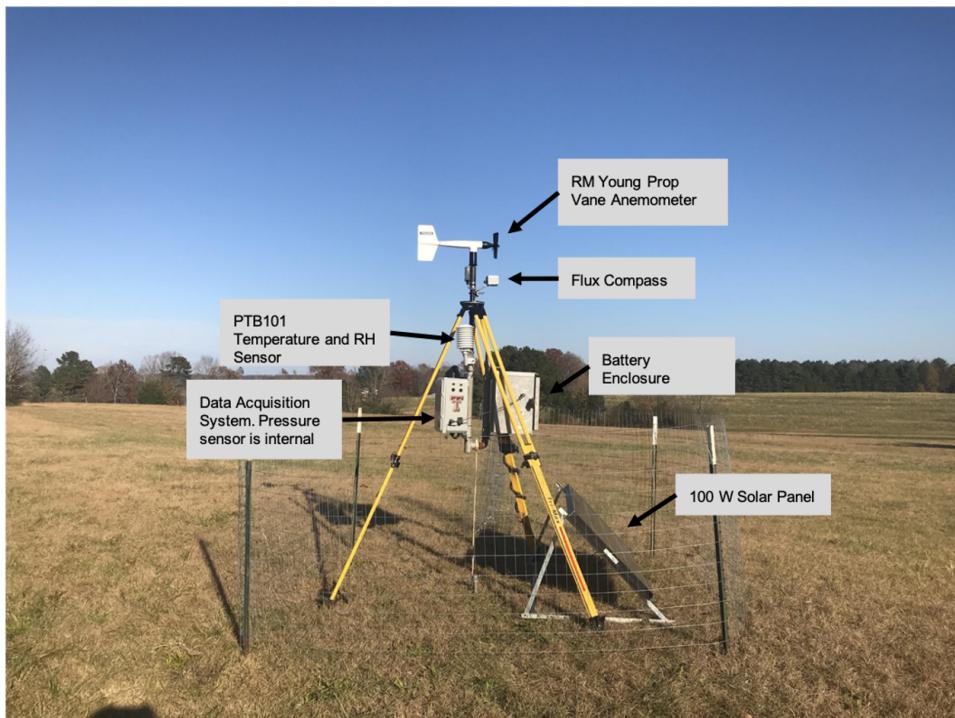


FIG. 1. An example deployment of a StesoNet station with accompanying instruments, data acquisition system, batteries, and solar panel labeled.

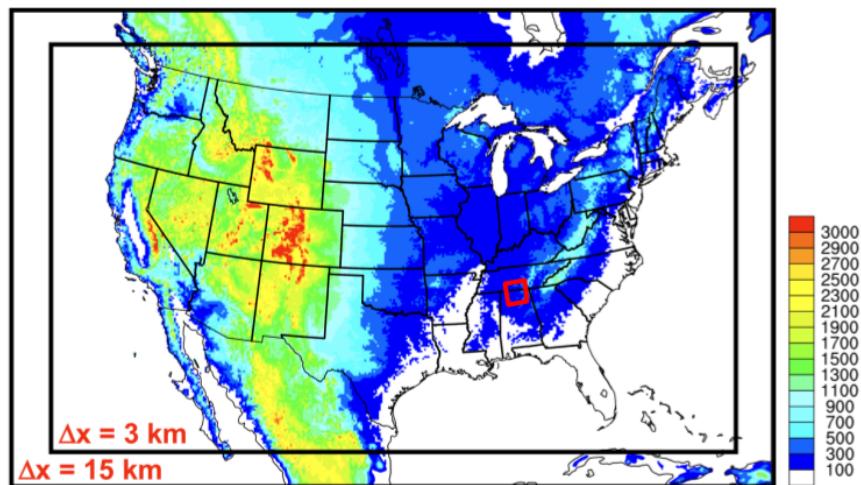


FIG. 2. Model domains of the HRRRE prediction system across the contiguous United States and model terrain (m, filled contours). The red box approximates the location of the StesoNet domain in northern Alabama and southern Tennessee.

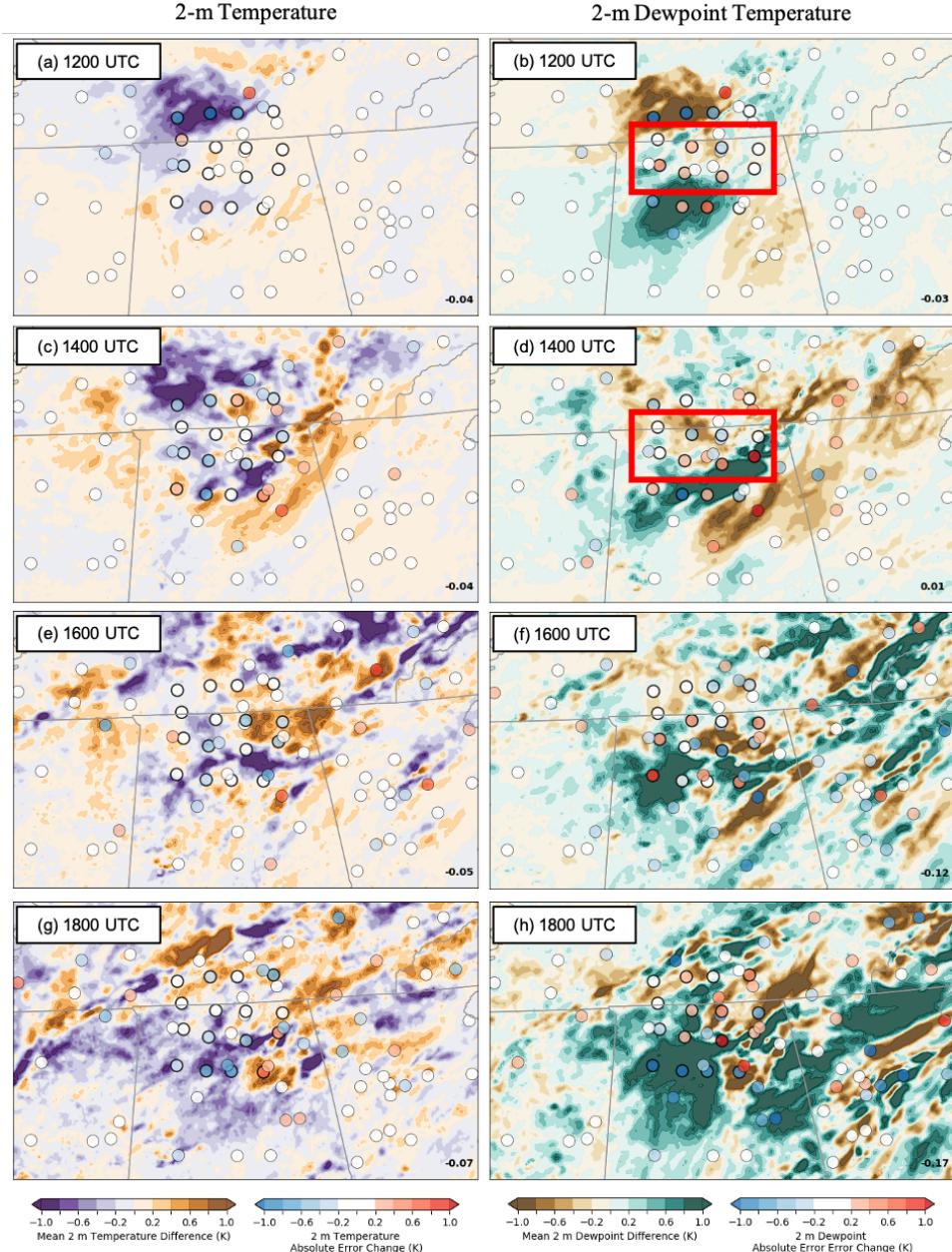


FIG. 3. Ensemble mean analysis differences calculated as EXP - CTRL of (left) 2-m temperature and (right) 2-m dewpoint temperature (K, shaded) at (a)-(b) 1200, (c)-(d) 1400, (e)-(f) 1600, and (g)-(h) 1800 UTC, respectively, for the 22Apr case. Colored dots represent the change in ensemble mean absolute error (K) between EXP and CTRL at surface observing stations, with negative values indicating a reduction in error in the EXP analysis and positive values an increase in error. The average change in absolute error at an analysis time across the mapped domain is denoted in the lower right corner. Circles with bolded outlines depict StesOnet observing sites.

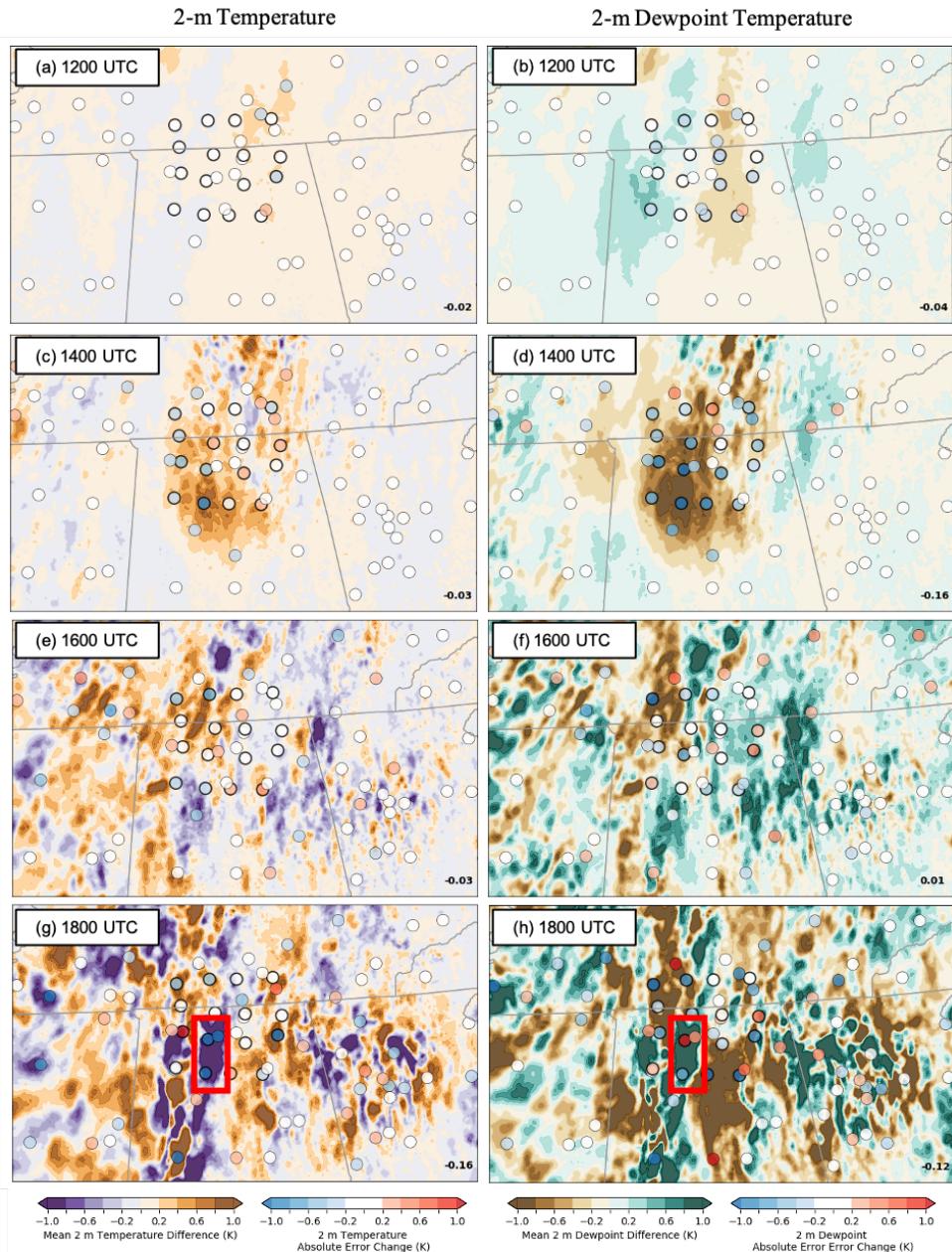


FIG. 4. As in Fig. 3, but for the 30Apr case.

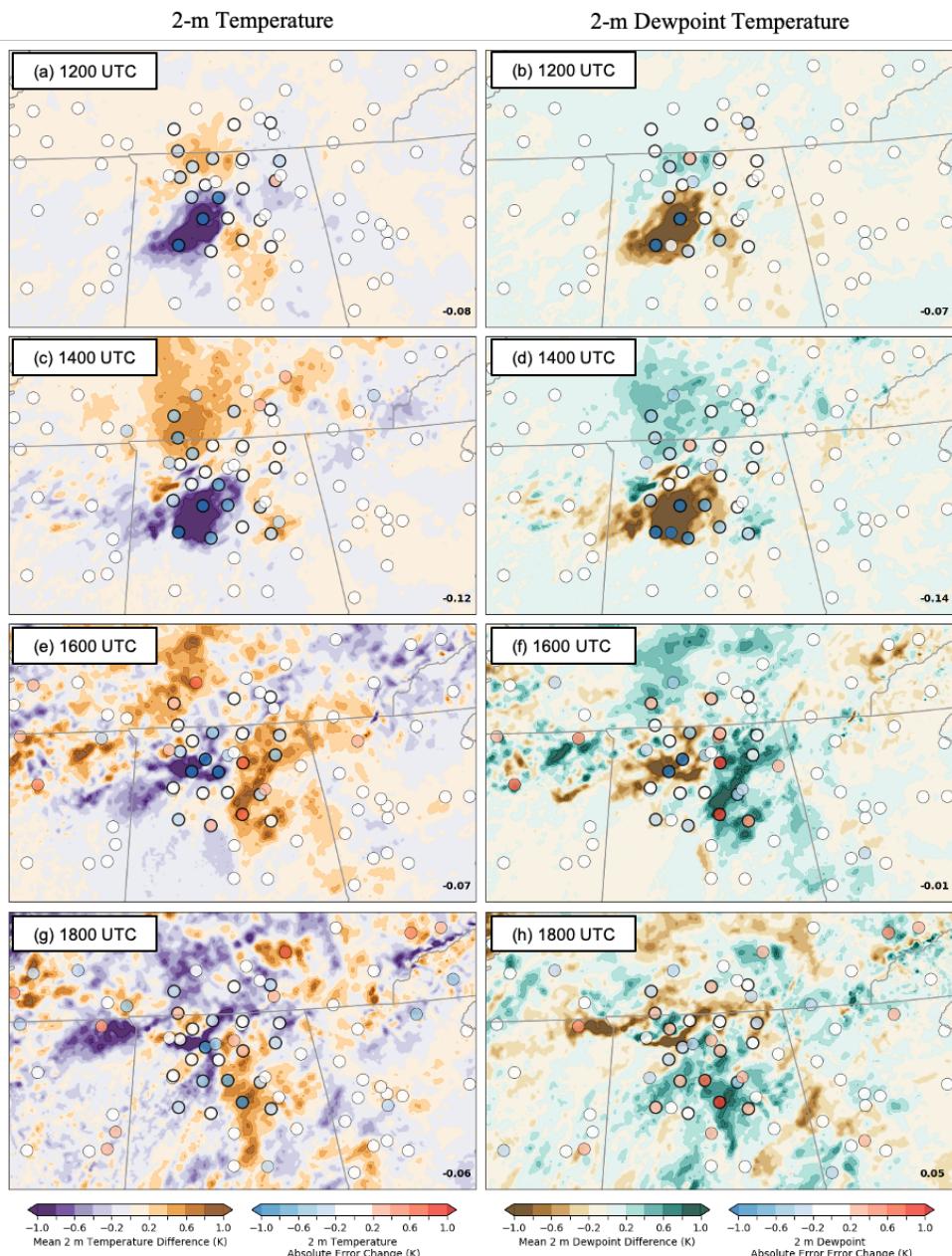


FIG. 5. As in Fig. 3, but for the 23Feb case.

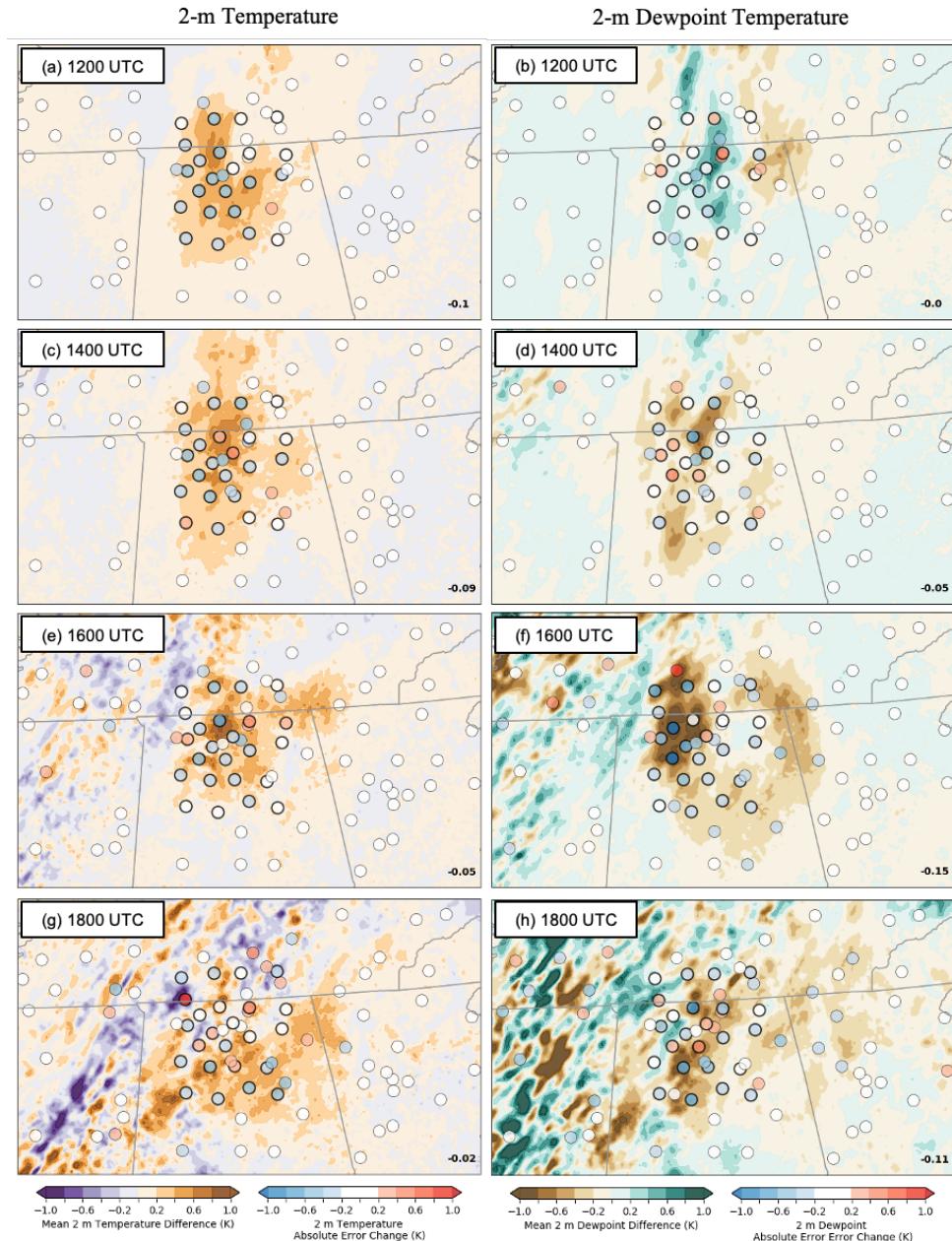


FIG. 6. As in Fig. 3, but for the 14Mar case.

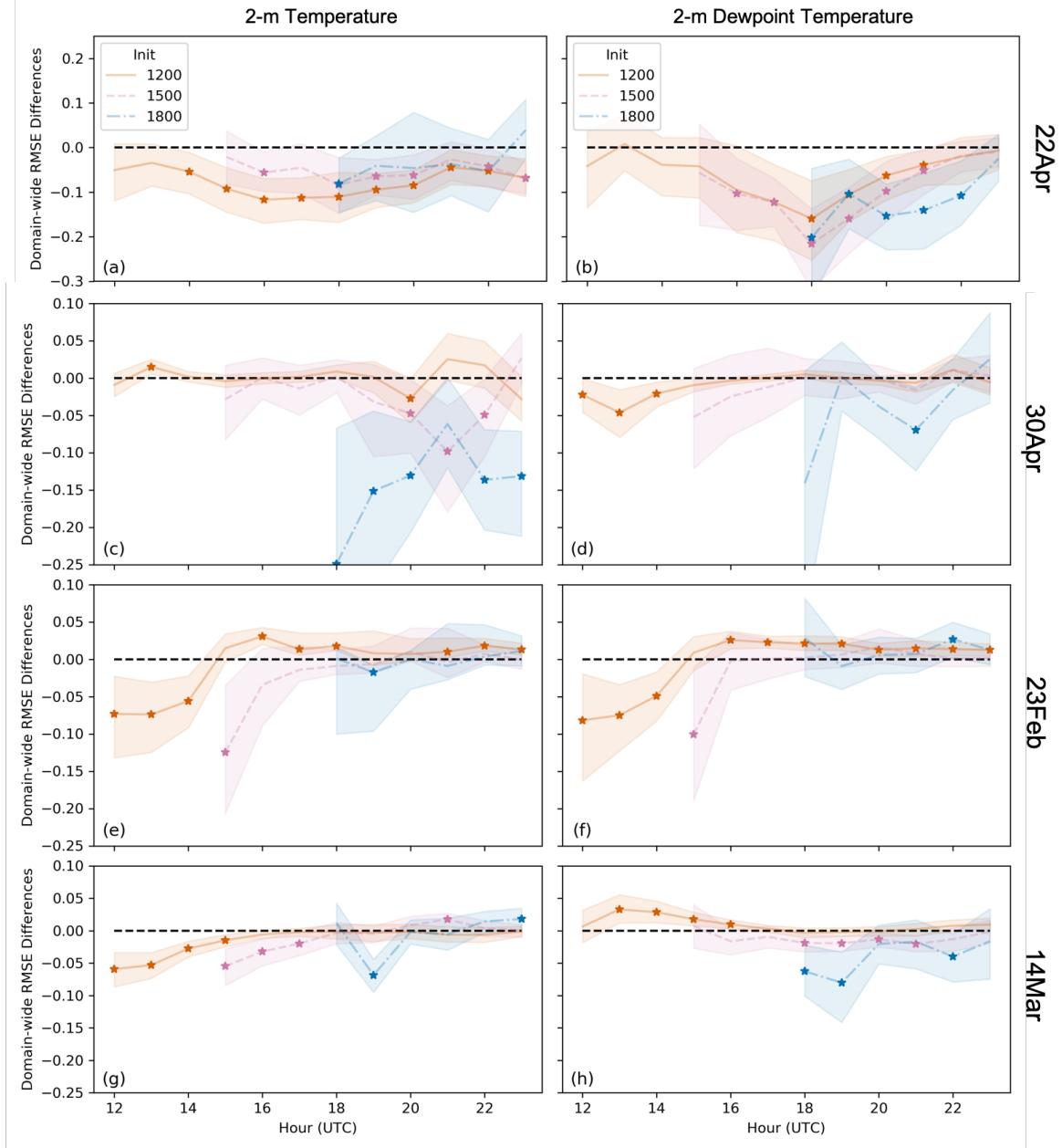


FIG. 7. Domain-wide ensemble mean analysis and forecast RMSE differences (EXP-CTRL) of (left) 2-m temperature (K) and (right) 2-m dewpoint temperature (K) for (a)-(b) 22Apr, (c)-(d) 30Apr, (e)-(f) 23Feb, and (g)-(h) 14Mar cases calculated across $33\text{--}36^\circ\text{N}$ and $83\text{--}90^\circ\text{W}$. Shading around each line represents a 95% confidence interval obtained via bootstrapping with 1,000 resamples of the paired sample error distributions from each forecast. Stars spanning each colored line delineate when the RMSE differences between EXP and CTRL are statistically significant.

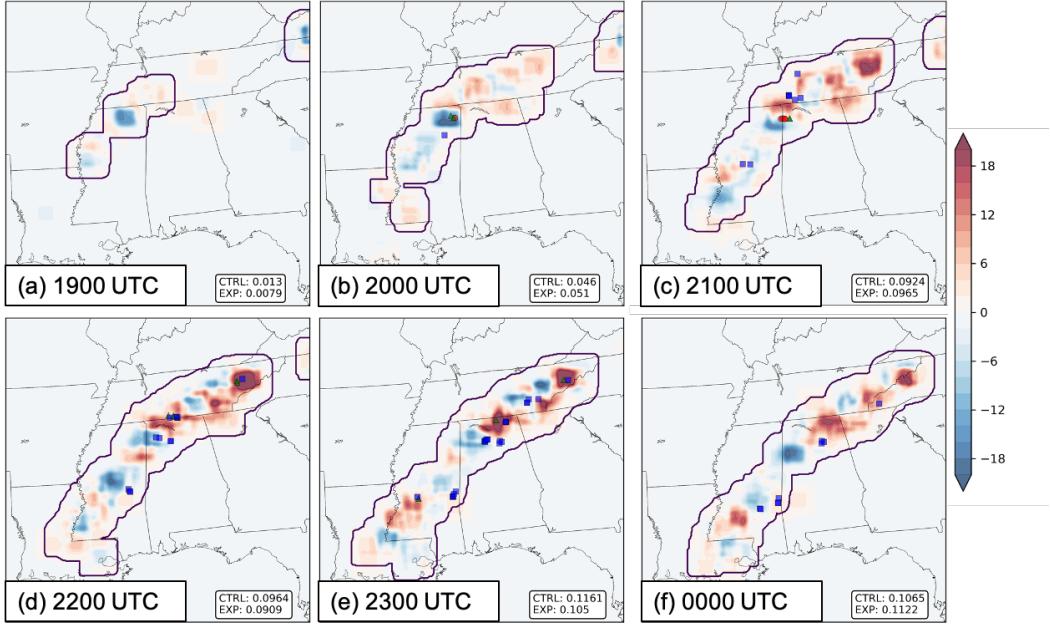


FIG. 8. Differences between 22Apr EXP and CTRL ensemble forecast probabilities ($\text{EXP} - \text{CTRL}$) of 1–6 km maximum UH exceeding $50 \text{ m}^2 \text{s}^{-2}$ over the previous hour initialized at 1800 UTC and valid at (a)–(f) 1900, 2000, 2100, 2200, 2300, and 0000 UTC (forecast hours 1–6). Denoted in the lower right corner are the Brier Scores for each respective forecast. Local storm reports of severe wind, hail, and tornadoes within the last hour are denoted with blue squares, green triangles, and red circles, respectively. Storm reports are obtained from an archive on the Iowa State University Iowa Environmental Mesonet website (<https://mesonet.agron.iastate.edu/request/gis/lsts.phtml>, accessed 28 April 2020).

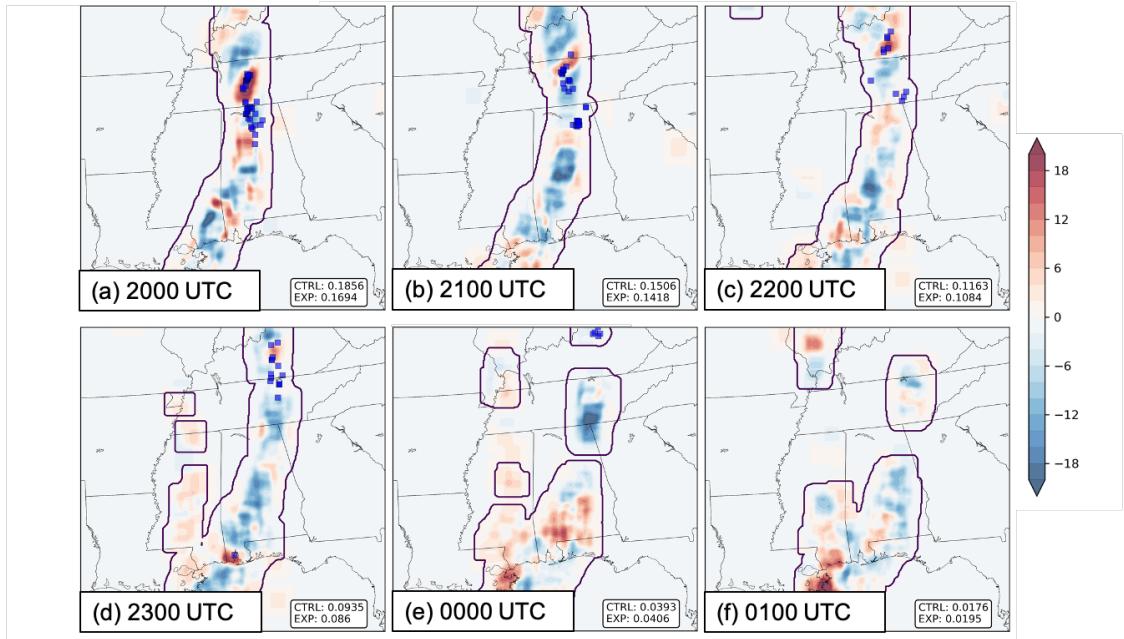


FIG. 9. As in Fig. 8, but for the 30Apr case at hours (a)–(f) 2000, 2100, 2200, 2300, 0000, and 0100 UTC.

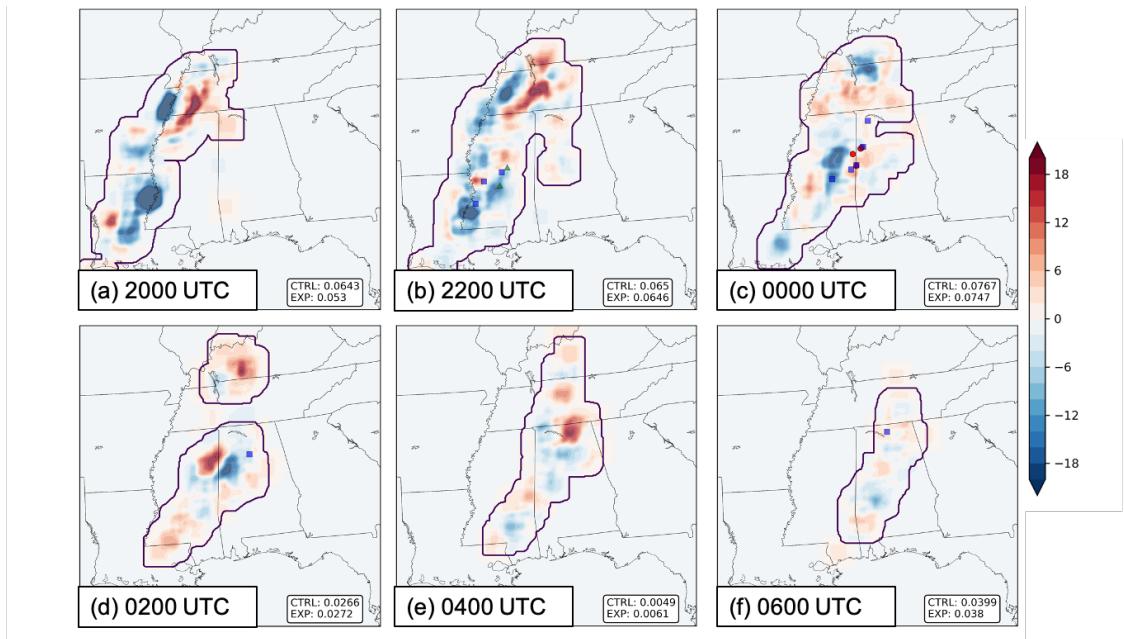


FIG. 10. As in Fig. 8, but for the 23Feb case at hours (a)-(f) 2000, 2200, 0000, 0200, 0400, and 0600 UTC.

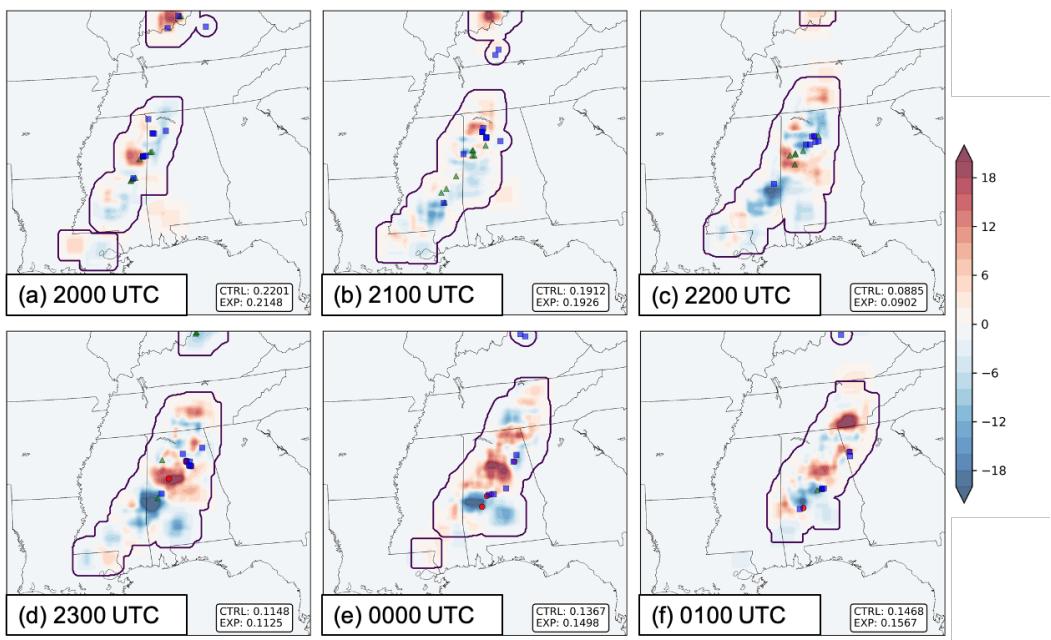


FIG. 11. As in Fig. 9, but for the 14Mar case.

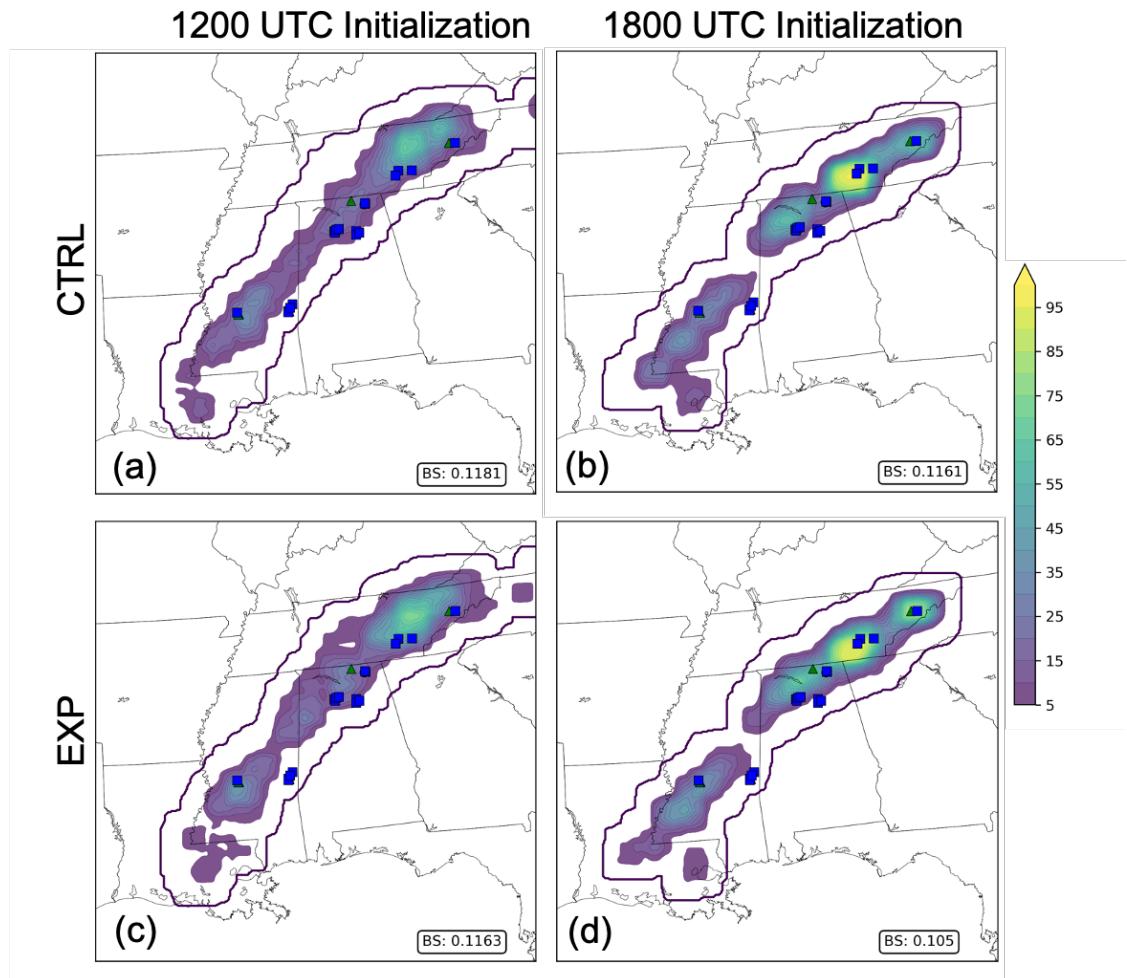


FIG. 12. Probabilistic ensemble forecasts of 1–6 km UH exceeding $50 \text{ m}^2 \text{s}^{-2}$ over the previous hour from (a)–(b) the CTRL and (c)–(d) EXP ensemble systems. Forecasts are initialized at (left) 1200 UTC and (right) 1800 UTC valid at 2300 UTC for the 22Apr case. Local storm reports within the last hour of severe wind, hail, and tornadoes are denoted with blue squares, green triangles, and red circles, respectively. Storm reports are obtained from an archive on the Iowa State University Iowa Environmental Mesonet website (<https://mesonet.agron.iastate.edu/request/gis/ltrs.php>, accessed 28 April 2020).

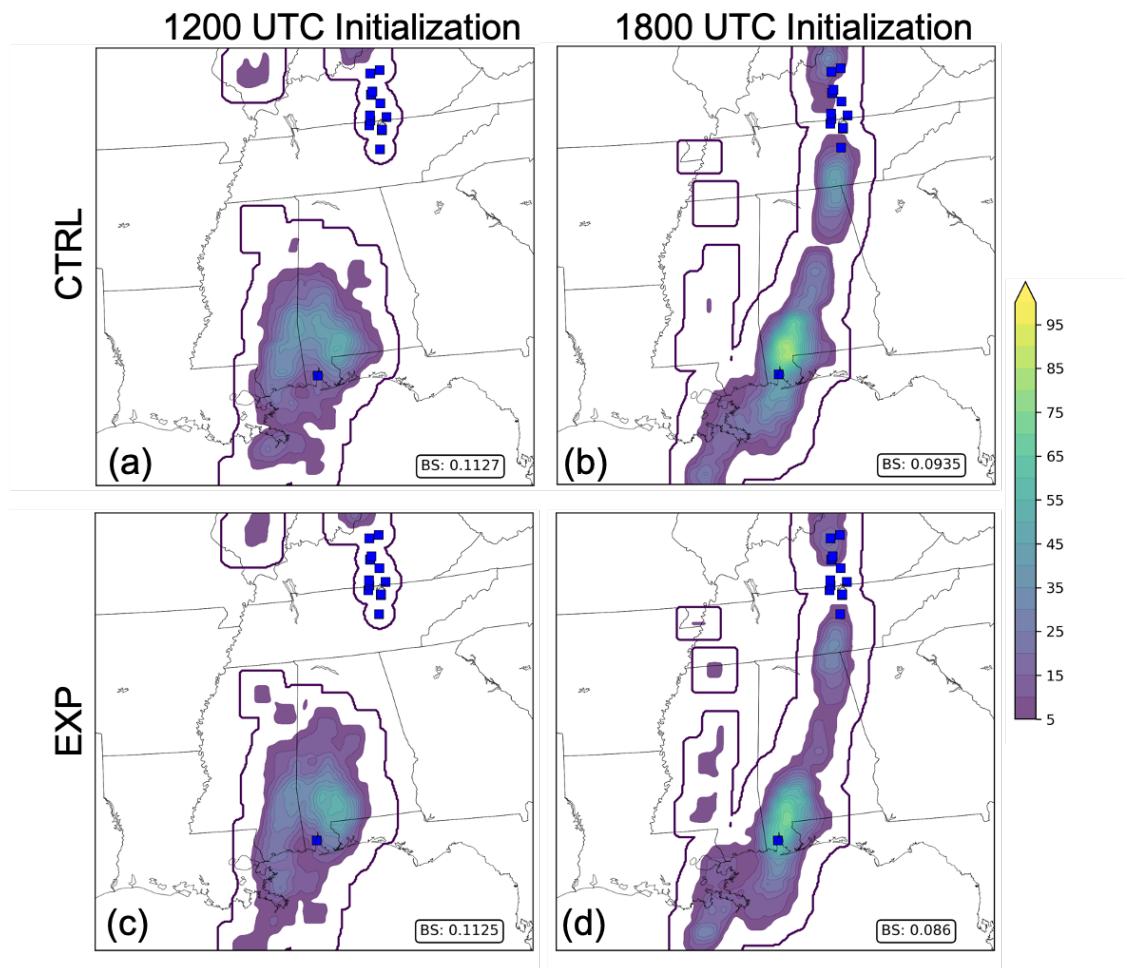


FIG. 13. As in Fig. 12, but for the 30Apr case.

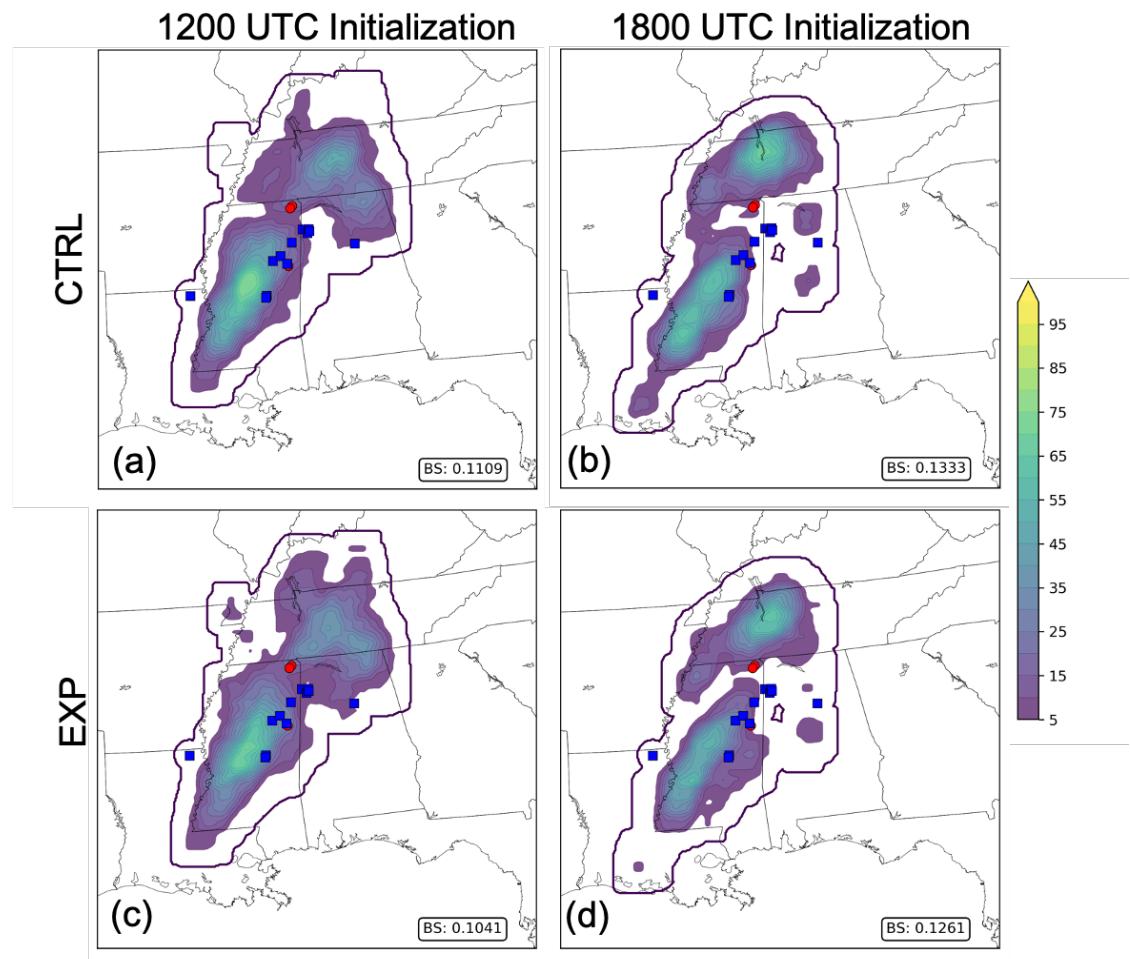


FIG. 14. As in Fig. 12, but for the 23Feb case.

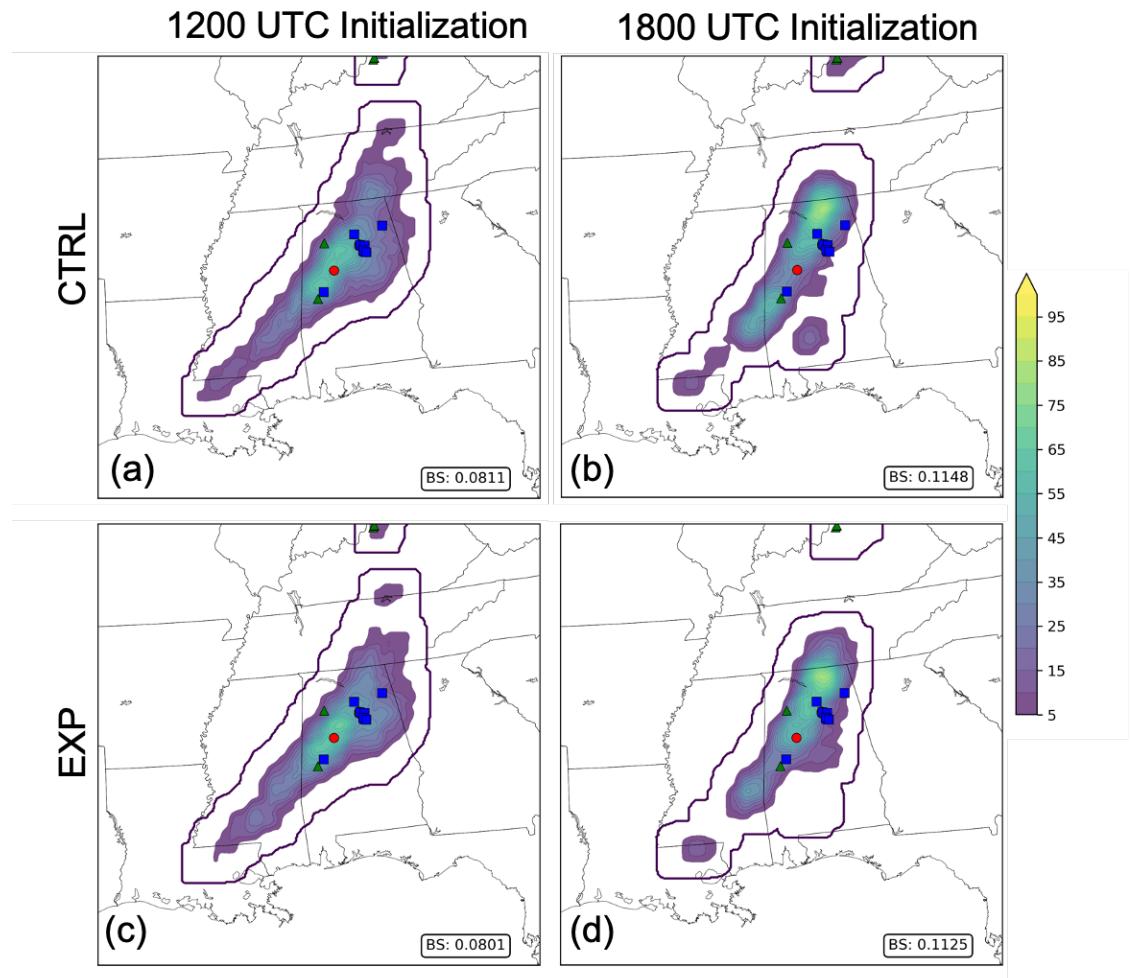


FIG. 15. As in Fig. 12, but for the 14Mar case.

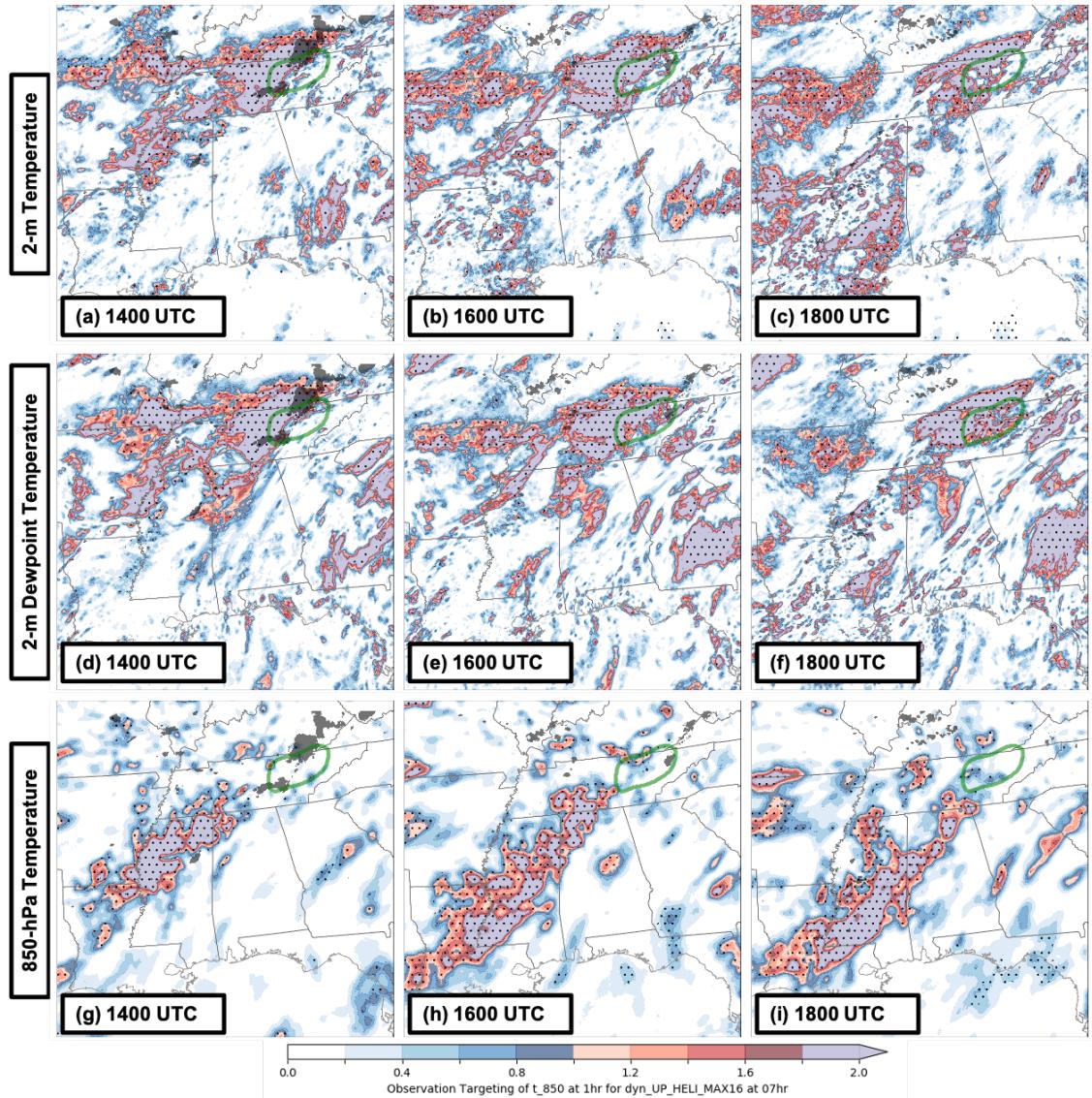


FIG. 16. Expected ensemble variance reduction (i.e., ESA targets) of UH response ($m^4 s^{-4}$, color shading) due to hypothetical assimilation of (a)-(c) 2-m temperature, (d)-(f) 2-m dewpoint temperature, and (g)-(i) 850-hPa temperature at (a),(d),(g) 1400, (b),(e),(h) 1600, and (c),(f),(i) 1800 UTC 22 April 2017 calculated from the 1200-UTC initialized CTRL ensemble forecast. Gray shading is the ensemble probability matched mean composite reflectivity greater than 40 dBZ. The green contour delineates the UH response region valid at 2200 UTC. Black dots denote geographical areas where the ensemble sensitivity passes a statistical significance test demonstrating the regression slope between initial condition and response variables is greater than zero with 95% confidence.

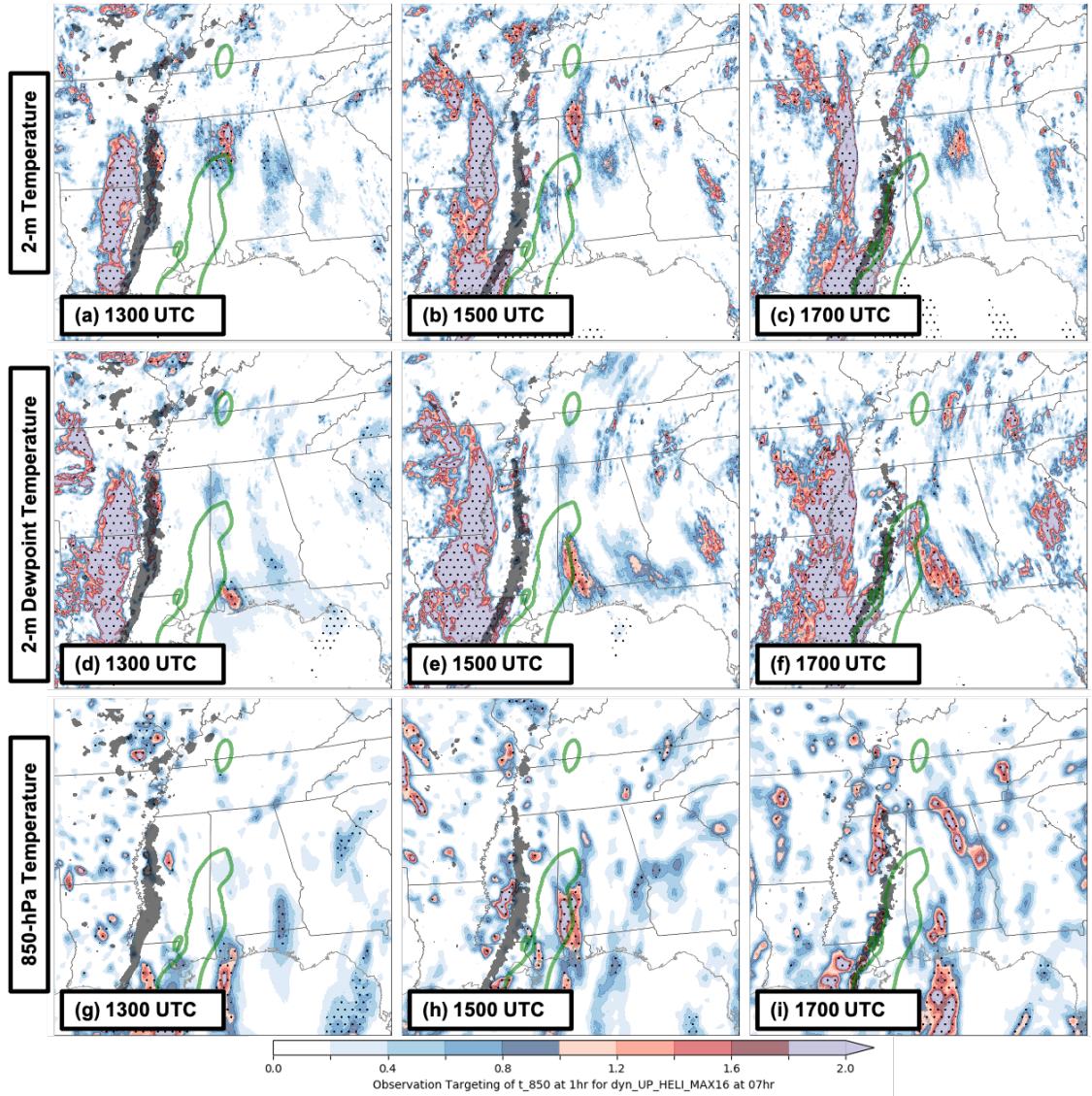


FIG. 17. As in Fig. 16, but for the 30Apr case. The response is valid at 1900 UTC and ESA-based targets are valid at (a),(c),(g) 1300, (b),(e),(h) 1500, and (c),(f),(i) 1700 UTC.

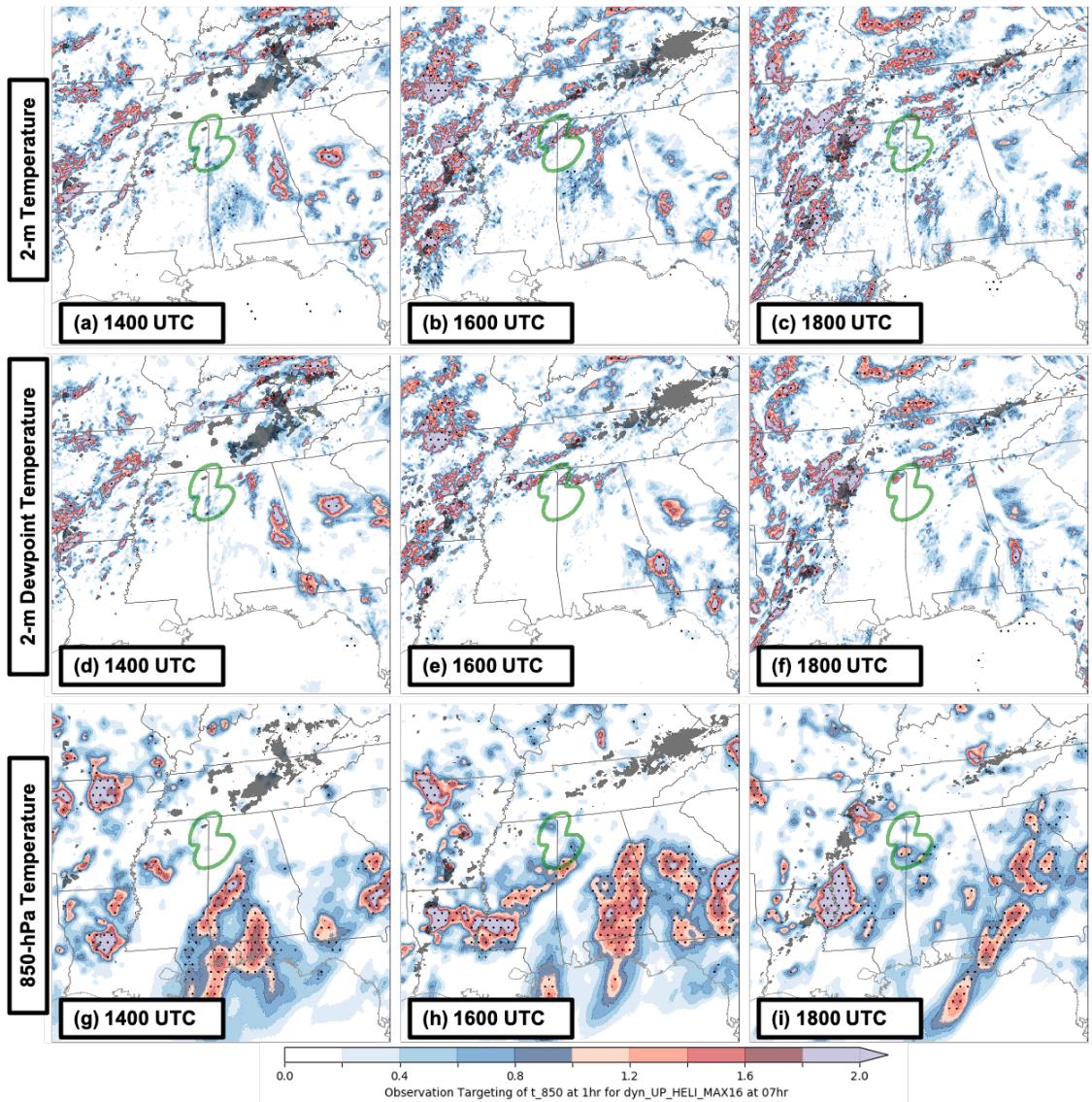


FIG. 18. As in Fig. 16, but for the 23Feb case. The response is valid at 0100 UTC 24 February 2019 and ESA-based targets are valid at (a),(c),(g) 1400, (b),(e),(h) 1600, and (c),(f),(i) 1800 UTC.

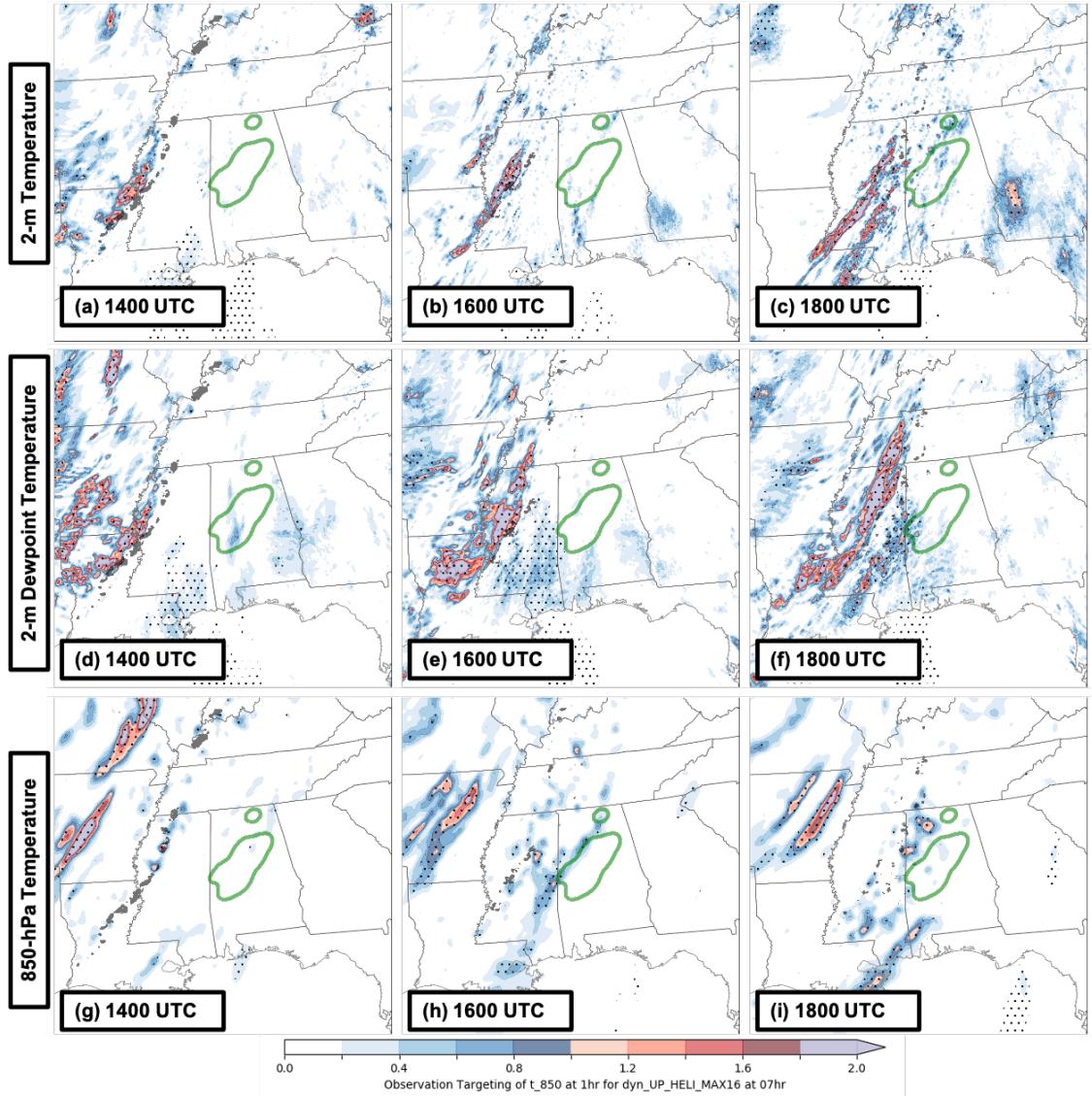


FIG. 19. As in Fig. 16, but for the 14Mar case. The response is valid at 2100 UTC and ESA-based targets are valid at (a),(c),(g) 1400, (b),(e),(h) 1600, and (c),(f),(i) 1800 UTC.