J11.4 IN SEARCH OF CI: LESSONS LEARNED FROM APPLYING PROXIES TO CI IN CONVECTION-ALLOWING MODELS

Stuart D. Miller, Jr.[1,2,3], John S. Kain[3]*, Patrick T. Marsh[4], Adam J. Clark[2,3], Michael C. Coniglio[3], James Correia, Jr.[2,4]

[1]University of Oklahoma School of Meteorology

[2]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK

[3]NOAA/OAR/National Severe Storms Laboratory, Norman, OK

[4]NOAA/NWS/NCEP/Storm Prediction Center, Norman, OK

1. INTRODUCTION

Convection-allowing models (CAMs) have proven to be very useful for prediction of convective-scale phenomena, but these models do not have sufficient resolution to simulate with fidelity many of the phenomena of interest. For example, scientists at the National Oceanic and Atmospheric Administration Hazardous Weather Testbed (NOAA/HWT) have shown that CAMs with grid spacing as coarse as 4 km can provide useful guidance for the prediction of supercells (Sobash et al. 2011) and even tornadoes (Clark et al. 2012, 2013), despite the fact that the effective resolution of these models is much too coarse to represent these features on the scales that they naturally occur. The key to success in these applications is finding features that CAMs produce reliably on their convectively active scales and that are reasonable surrogates, or proxies, for the phenomena of interest.

Finding robust proxies can be challenging. During the HWT Spring Forecasting Experiments of (SFE) of 2011 (hereafter SFE2011) simulated lightning flash-rate density (McCaul et al. 2009) and simulated reflectivity were used as proxies for deep convective activity (CA) in CAMs. These proxies worked well for the model configuration used in 2011 but problems were encountered when attempts were made to verify them and then again when they were applied with multiple microphysical parameterizations during SFE2012. These problems complicated efforts to benchmark the performance of CAMs in predicting the timing and location of convection initiation (CI), which was the underlying motivation for this work. However, a number of model sensitivities were uncovered as the work proceeded and these sensitivities have potentially important implications for both model developers and weather forecasters. The purpose of this paper is to highlight these sensitivities and provide a brief summary of the CI results.

* *Corresponding author address*: John S. Kain, NOAA/OAR/National Severe Storms Laboratory, 120 David L. Boren Blvd., Norman, OK 73072; e-mail: Jack.Kain@noaa.gov

2. DATA AND METHODS

2.1 SFE2011

2.1.1 CAM system

The NSSL-WRF (Kain et al. 2010) was the primary modeling system used during 2011. This is a realtime WRF-ARW (Skamarock et al. 2008) system that is operated by the NOAA/National Severe Storms Laboratory (NSSL), in collaboration with the NOAA/Storm Prediction Center (SPC). It is initialized twice daily at 0000 and 1200 UTC throughout the year over a full CONUS domain with forecasts to 36 h, using initial and lateral-boundary conditions from the North American Mesoscale (NAM) model (Rogers et al. 2009). Output has been supplied directly to the SPC since 2006 and in recent years has been disseminated to other National Weather Service offices through the National Centers for Environmental Prediction Central Operations (NCO). It uses the WSM6 microphysics scheme (Hong et al. 2004; Hong and Lim 2006), NOAH land-surface (Chen and Dudhia 2001), and MYJ planetary-boundary-layer parameterizations (Janjic 1994), along with RRTM short-wave (Mlawer et al. 1997) and Dudhia long-wave (Dudhia 1989) radiation. Forecast graphics are available at http://www.nssl.noaa.gov/wrf/. This modeling system is used as the “alpha” testing

framework for many diagnostic tools that are used in annual HWT SFEs.

2.1.2 Objective Identification of Deep Moist Convection

During SFE2011 we considered the viability of three distinctly different sets of criteria to identify grid points in CAM forecasts that had CA:

• *Simulated lightning*. The lightning flash-rate-density (FRD) algorithm, developed by McCaul et al. (2009) was used to infer the presence of cloud-to-ground (CG) lightning, following the work of Miller et al. (2010). This algorithm, based on simulated values of both graupel flux at the -15º C level and vertically integrated graupel, was originally formulated to predict the FRD of total lightning in CAM forecasts, but Miller et al. (2010) showed that it could also be used to provide a useful proxy for the occurrence of CG lightning with selection of an appropriate threshold value of FRD. The threshold value was determined by mapping National Lightning Detection Network (NLDN – see http://gcmd.nasa.gov/records/GCMD_NLDN.html) strike data to the nearest model grid point, then comparing the climatology of model-predicted FRD magnitudes to that of the NLDN data. A threshold value of FRD was determined iteratively to provide approximately the same frequency of grid-point activation (i.e, frequency of threshold exceedance) as given by the NLDN data over the same time period. For the SFE2011, a threshold value of 0.55 km^{-2} (5 min)$^{-1}$ was used, based on calibration in the 4 km NSSL-WRF model from 11 March – 10 June 2010. Specifically, for the simulated lightning (LTG) criteria set, a CA grid point was identified in hourly model output as any point at which the FRD exceeded 0.55 km^{-2} (5 min)$^{-1}$ during the previous hour. For verification purposes, individual NLDN strikes were mapped to the nearest model grid point.

• *Explicit measurement of updraft strength and precipitation content*:. A model grid column was defined as convectively-active given the following conditions: (1) the maximum updraft exceeded a threshold value Wmin; (2) EITHER the maximum graupel mixing ratio exceeded a threshold value QG (g kg^{-1}), OR the maximum rain mixing ratio exceeded a threshold value QR (g kg^{-1}), OR both conditions were met. Preliminary threshold values of W = 5 m s^{-1}, QG = 2 g kg^{-1}, and QR = 1 g kg^{-1} were selected via empirical testing to allow for a range of intensities of surface-based or elevated, warm- or cold-season, and extra-tropical or tropical storms. To prevent shallow, terrain-induced updrafts from being falsely identified as convection, the grid column was scanned from the top of the boundary layer to the equilibrium level (i.e., approximating the maximum probable CAPE-bearing layer depth) to identify the maximum local updraft value. These explicit model diagnostics comprised the "WQQ" set of criteria for CA. There was no corresponding observational dataset for the WQQ diagnostic.

• *Simulated reflectivity:* The 35dBZ threshold for simulated reflectivity (computed as in Kain et al. 2008) was used to identify CA grid points, as in Roberts and Rutledge (2003), Mecikalski and Bedka (2006), and other studies. In order to avoid bright-banding effects, it was required that this threshold be met or exceeded at the -10º C level (see Gremillion and Orville 1999). This defined the simulated minus-10 reflectivity"MTR35" criteria set for CA. The simulated MTR35 dataset was verified against observed reflectivity fields from the national network of WSR-88D radars, available in the National Mosaic and Multi-Sensor QPE (NMQ) dataset (Zhang et al. 2011) on a 1 km CONUS grid at 5-min intervals, with the -10ºC-level determined by RUC (Benjamin et al. 2004) analyses. Specifically, simulated MTR35 grid points were identified in hourly output fields as any grid point at the -10º C level where the 35dBZ threshold was exceeded. In the interest of consistency with the mapping of NLDN data, any point on the model grid that was nearest to an NMQ grid point exceeding 35 dBZ during the previous hour was marked as an observed CA point for verification of the simulated MTR35 field.

2.2 SFE2012

2.2.1 CAM system

The University of Oklahoma Center for Analysis and Prediction of Storms (CAPS) ran a high-resolution ensemble forecast system during SFE2012, with a forecast domain and horizontal resolution very similar to the NSSL-WRF, but with somewhat higher vertical resolution. Specifically, the domain of each member covered the CONUS with 4-km horizontal grid spacing and 51 vertical levels. All members used the 0000 UTC analysis from the North American Mesoscale (NAM) model (Rogers et al. 2009) as the background field in the initialization process and the CAPS 3DVAR/ARPS Cloud analysis scheme (Xue et al. 2003, 2008; Gao et al. 2004; Kong et al. 2008) for assimilation of WSR-88D radar reflectivity and radial velocity data , along with surface and upper air observations.. Results from two subsets of the full ensemble are considered in this paper and all of these operative members used the WRF-ARW dynamic core.

The first subset was comprised of 10 members in which both model physics and initial plus lateral-boundary conditions (IC/LBCs) were perturbed. Physics were varied by changing WRF namelist options from member to member while different IC/LBCs were introduced by extracting perturbations from the operational National Centers for Environmental Prediction Short-Range Ensemble Forecast (SREF) modeling system (Du et al. 2009). This first subset was called the core subset.

In the second subset each member used identical initialization procedures and configurations of the WRF-ARW model, but model physics varied from member to member. The focus here is on the "microphysics members" (4 out of the 8 members) in which only the microphysical parameterizations differed. Specifically all four of these members used the NOAH land-surface and MYJ planetary-boundary-layer parameterizations, along with RRTM short-wave and Goddard long-wave (Chou and Suarez 1994) radiation parameterizations, while individual members used the Thompson et al. (2008; hereafter Thompson), WRF double-moment 6-class (Lim and Hong 2010; hereafter WDM-6;), Milbrandt and Yau (2005; hereafter M-Y), and Morrison et al. (2005; hereafter Morrison) microphysical parameterizations, respectively.

2.2.2 Objective Identification of Deep Moist Convection

Based on favorable indications from NSSL-WRF testing in 2011, MTR35 was the only criterion used to identify CA points during SFE2012, for both model output and observations. However, because multiple double-moment microphysical parameterizations were used, the simulated reflectivity algorithm used in 2011 was abandoned in favor of a more sophisticated algorithm provided by Greg Thompson (G. Thompson, NCAR, personal communication). This algorithm approximates reflectivity by computing Rayleigh scattering due to all microphysical species, and includes an integration over the computed size and number distributions of each species (Smith 1984), with enhancements for melting effects (Blahak 2007). In addition, observed MTR35 was mapped to the 4-km model grid in a different way for 2012. In particular, it was interpolated using the copygb program (http://www.cpc.ncep.noaa.gov/products/wesley/copygb.html) with the budget option. Using this interpolation approach reduced the number of observed CA points compared to the "nearest grid point" mapping technique used in 2011 by effectively requiring the average reflectivity within a 4-km grid box (on the 1-km NMQ grid) to be greater than 35 dBZ instead of just one point on the higher resolution grid.

2.2.3 Objective identification of CI

CI points were identified as a subset of CA points using a 3-dimensional object identification algorithm in which the first two dimensions define horizontal placement and the third dimension is time. Conceptually, Fig. 1 shows how this algorithm works. This figure depicts CA activation points accumulated over a 1-h time period, color-coded by the time interval within the hour that they activated. For individual cells that formed within this hour, the CI point corresponds to the location of the earliest CA point within the contiguous CA object. In practice, this algorithm searches over all 5-minute output intervals within a specified time window, identifying 3-dimensional objects as groups of contiguous (or adjacent) grid-points that were convectively active at the same time or immediately preceding or following times. The CI point is determined by locating the point that activated earliest within an object. Short-lived storms are eliminated from consideration by requiring an intra-object time range of at least 30 minutes.

2.2.4 Experimental forecasts for CI

Forecasts for CI were issued each morning during SFE2012 as a collaborative effort by a team of forecasters and researchers. The first step in this process was to identify a regional domain within which deep convection was anticipated later in the day (e.g., the area outlined by the brown rectangle in Fig. 2a). Next, probability contours were drawn to indicate the likelihood of a CA within 20 km of a point during a specified 4-h time window (e.g., Fig. 2a). Within the regional domain, a limited sub-domain was outlined in the interest of isolating an anticipated CI event. The CI forecast team came to consensus on 1) the most likely time (hourly interval) and location of CI, 2) the overall probability that the targeted CI episode would occur, and 3) an estimate of uncertainty in the timing, given the occurrence of the episode. The consensus location and hour were indicated on the forecast graphic with an "X" and time (labeled as the midpoint of the hour), respectively (Fig. 2a). Uncertainty information was indicated on a separate graphic, using a web-based interactive display (Fig. 2b). This display allowed the CI forecast team to create a temporal probability distribution by simply dragging markers associated with each forecast hour to graphical positions representing the relative likelihood of CI at

individual hours. It included a normalization function that adjusted the total area represented by the hourly probabilities based on the overall likelihood of the CI episode, while maintaining relative hourly amplitudes.

The ten core members of the CAPS ensemble were used to provide specific numerical guidance for the temporal probability forecasts. Specifically, using the automated CI-detection algorithm described above, a histogram of predicted CI timing was generated by searching for the first-CI event in each ensemble member within the forecast sub-domain and within +/- 5 h of the anticipated CI time (e.g., Fig. 2b).

3. RESULTS

3.1 SFE2011

Many results and observations from SFE2011 are summarized in Kain et al. (2013). Here a few additional results are provided, focusing on lessons learned in applying the different proxies for CA.

First, participants in SFE2011 formed a general consensus that the three different CA proxies applied to model output all appeared to be adequate for identifying points where deep convection was occurring. In particular, examination of hourly output during the experiment suggested that, for the most part, all 3 proxies highlighted the same features; any differences did not seem to favor one proxy over another in a systematic way. Temporal trends of CA coverage for the entire SFE2011 period are consistent with this assessment – the trends from the different proxies are highly correlated (Fig. 3). Additional considerations led to the elimination of the LTG and WQQ options for future work. The latter option had no readily available observational counterpart that could be used for verification purposes. The LTG option could be verified using NLDN data but it was found that the NLDN fields had very different spatial characteristics than the diagnostic LTG fields. Specifically, CA entities highlighted by the LTG fields tended to be much more contiguous than the NLDN fields, which often had a “scattershot” appearance (cf. Figs. 4a and b). With the scattershot representation, early CI-detection algorithms too often found multiple individual initiation points in single storms. In short, the geospatial characteristics of the LTG and NLDN fields were different enough that there were concerns about using NLDN as a robust observational counterpart to LTG.

By comparison, the MTR35 fields derived from model output and observations appear to be reasonable analogues. The geospatial characteristics of simulated reflectivity fields from CAMs and observed reflectivity patterns from WSR-88Ds are typically quite similar in appearance, which was part of the appeal of simulated reflectivity output when it first became widely available nearly a decade ago (Koch et al. 2005). Furthermore, they appear to be similar in a climatological sense: Aggregated over all days of SFE2011, the total areal coverage of the simulated MTR35 field from the NSSL-WRF was within 3% of the observed MTR35 coverage derived from the NMQ dataset. For these reasons, MTR35 was used as the sole proxy for CA in SFE2012.

3.2 SFE2012

3.2.1 Sensitivities to microphysical parameterization

During SFE2012 the CAPS ensemble was used for CA/CI studies, building on the lessons learned with the deterministic NSSL-WRF during SFE2011. Since the ensemble used multiple microphysical parameterization (MPs) a concerted effort was made to ensure that simulated reflectivity was computed consistently with all MPs. This paved the way for a systematic assessment of the sensitivities of different reflectivity fields to MP. Simulated satellite imagery (Bikos et al. 2012) was also used to compare the schemes.

3.2.1.1 Impact of New Simulated Reflectivity Algorithm

Before the start of SFE2012, organizers worked with developers of the Thompson (Thompson et al. 2008), Morrison (Morrison et al. 2005), and WDM6 (Lim and Hong 2010) MPs to formulate simulated-reflectivity algorithms that were uniquely designed for these three double-moment MPs[1]. Earlier experiments had used the formulation described in Kain et al. (2008), which is appropriate for the WSM6 MP (Hong et al. 2004) but inadequate in accounting for the unique parameter settings, size distributions, and assumptions that are prescribed in different MPs in the WRF model. Simulated reflectivity fields derived from the “new” and “old” methods were compared to observations on a daily basis and the impact of the change was clearly evident, more so with some MPs than with others. For example, a

[1] Another double-moment MP used in SFE12 was the Milbrandt-Yau scheme (Milbrandt and Yau 2005 – hereafter M-Y). This MP already contains its own unique simulated-reflectivity formulation.

snapshot of composite reflectivity from 27-h forecasts valid 0300 UTC 31 May suggests that the character of differences between the "new" and "old" output fields varies from one MP to another (Fig. 5). One can compare these output fields for different times and dates by clicking here. During SFE2012, the old calculation was used only for quick comparisons like those enabled by Fig. 5. The new MP-specific calculations were used for most purposes and they will be used exclusively from this point forward in this summary.

3.2.1.2 Sensitivity of Simulated Composite Reflectivity to MP

The sensitivity of composite reflectivity (CREF) to MP was quite revealing (Fig. 6). For example, the hourly-maximum (maximum value at each grid point during the preceding hour – hereafter HM) CREF field from the M-Y MP tended to produce the highest coverage of very high reflectivity values (> 60 dBZ), the Morrison scheme often produced relatively large areas of moderately high (35-45 dBZ) values and fewer very high values than the others. The HM-CREF fields associated with the Thompson and WDM6 schemes were not so easy to characterize in terms of systematic biases. These fields can be compared for additional dates and times here.

3.2.1.3 Sensitivity of MTR to MP

Reflectivity fields at the -10º C temperature level (MTR) were compared every day as well. Figure 7 provides an example of HM-MTR fields associated with the different MP schemes. In the context of the CA diagnosis, the HM-MTR field enables a quick visual assessment of points that were CA during the preceding hour (the yellow and "hotter" colors). Although it is not quantitatively evident from figure 7, each of the MPs tended to over-predict the coverage of MTR35, and thus the coverage of CA. Coverage-bias statistics for diagnosed CA points, aggregated for the last 20 days of the experiment, indicated that bias ranged from 1.8 (Thompson scheme) to 4.1 (Morrison scheme). Aside from the coverage of MTR35, participants in SFE2012 noted that the Thompson MP produced typically produced the highest HM-MTR values for a given event. These sensitivities can be explored for other dates and times here.

3.2.1.4 Simulated Reflectivity Histograms

After examining simulated reflectivity fields from different perspectives during the first couple of weeks during SFE12, it became evident that the various MPs have systematic differences in the way that they distribute reflectivity values across the spectrum of possible values. During the remainder of the experiment this aspect of the MP sensitivity was explored by creating histograms of simulated (and observed) reflectivity values each day. These histograms were populated by counting the occurrence of reflectivity values in each 2-dBZ bin over the range of 0 – 80 dBZ, aggregated from grid points over the entire forecast-output[2] domain, the 1200 -1200 UTC period (12-36 h forecast period for the models), and hourly snapshots. They were created for each of the physics members of the CAPS ensemble, four of which used the Thompson MP with unique PBL schemes (i.e., the PBL members) while the other three used the Morrison, M-Y, and WDM6 MPs but were identically configured otherwise. As expected, the distributions of simulated reflectivity varied significantly as a function of MP. For example, for the 24-h period starting1200 UTC on 25 May 2012, all of the PBL/Thompson members produced a distinctive peak in frequency between about 15-20 dBZ, in rough agreement with a peak in observations. This was true from both a low-level (Fig. 8a) and column-maximum (composite - Fig. 8b) perspective. These members also produced a discernible secondary peak between about 40 and 45 dBZ, but with no obvious observational counterpart in that range. In contrast, none of the other MPs seemed to produce this higher-reflectivity secondary peak. The Morrison scheme produced a broader lower-reflectivity peak than did the PBL/Thompson members and there were hints of this peak in the WDM6 distribution, but the M-Y distribution was closer to one of exponential decay as dBZ increased. In terms of distribution amplitude (a measure of areal coverage) the simulated values were significantly higher than observed in almost all bins. This is consistent with the high coverage bias that was noted above for the 35 dBZ threshold of the HM-MTR. Additional reflectivity histograms can be explored by clicking here

3.2.1.5 Simulated Satellite Imagery

Another aspect of the daily assessment of MP sensitivities involved simulated satellite imagery, specifically infrared brightness temperature (see Bikos et al. 2012). As with reflectivity, this assessment indicated a strong sensitivity to MP.

[2] The CAPS ensemble forecast-output domain was a subset of the full computational domain, covering approximately the eastern ¾ of the CONUS. Subsetting was imposed to reduce output-file size and facilitate real-time dataflow.

As is the case in the example shown in Figure 9, the M-Y MP typically produced more extensive, colder cloud shields than the other three MPs, while WDM6 was often associated with the smallest areal coverage of brightness temperature less than 211 K (green color-fill) and with more "spotty" coverage than the other MPs. By comparison the Thompson and Morrison MPs tended to produce brightness-temperature fields with similar coverage and general appearance. During SFE12, it was noted that these implied differences in cloud coverage/opacity were associated with differences in temperature of the land surface and the lower troposphere, apparently due to cloud radiative impacts. These differences could have significant impacts on the evolution of convective systems and subsequent convective development at a given location. These fields can be explored for other dates and times here.

3.2.2 Assessment of CI timing predictions

As discussed at the end of Section 2, the ten core members of the CAPS ensemble were used to generate histograms of CI timing for forecast guidance each day. Unfortunately, after the experiment an error was discovered in the code used to generate IC/LBC perturbations for the ensemble, resulting in considerable degradation of skill for the perturbed members (F. Kong, CAPS, personal communication). Core-ensemble predictions of CI timing were negatively impacted: For SFE2012, the mean timing errors were small, similar to those shown for SFE2011 in Kain et al. (2013), but the distribution about the mean was considerably broader than in SFE2011. The histogram is not shown here, but the poor resolution is implied by a reliability diagram showing that the core ensemble showed little or no skill in predicting higher temporal probabilities of CI (Fig. 10). In contrast, predictions from the forecast teams were quite reliable across a broad range of probabilities.

4. SUMMARY

During SFE2011 and SFE2012 a concerted effort was made to explore methods to detect and predict CI, and to benchmark the CI-prediction characteristics of CAM configurations that are commonly used in SFE. The benchmarking effort was only partially successful, due mainly to the fact that performance metrics are still unrefined. Overall, it was found that the CAMs, initialized at 0000 UTC, showed little systematic bias in predicting the timing of CI for events occurring the next afternoon, but on a given day predictions from the ensemble members were often clustered before or after the observed CI time. Human forecaster teams showed better temporal resolution and more skill in general, but they had the advantage of a much shorter lead time. Forecast teams also benefitted significantly from updated model forecasts and perhaps most of all by monitoring observational trends.

Although some elements of this study were only partially successful, the lessons learned in other areas exceeded expectations. In particular, by monitoring simulated reflectivity and simulated satellite imagery on a daily basis, numerous sensitivities to microphysical parameterizations were identified. At this stage, we can only speculate about the impact of these sensitivities on forecasts of sensible weather. But the nature of the sensitivities provides useful insights for the developers of the different parameterizations and a starting point for future studies that can quantify the impacts on numerical guidance for weather forecasting.

Acknowledgements

This work was enabled by collaborative activities involving numerous partners in the NOAA Hazardous Weather Testbed. We would like to thank our colleagues at the Storm Prediction Center, including Israel Jirak, Chris Melick, Steve Weiss, and Andy Dean, and those at OU/CAPS, including Ming Xue, Fanyou Kong, Kevin Thomas, and Keith Brewster. In addition, we very much appreciate assistance and shared insight from Greg Thompson, Hugh Morrison, Jimy Dudhia, Song-You Hong, Kyo-Sun Lim, and Jason Milbrandt related to the microphysical parameterizations in the WRF model.

References:

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518. doi: http://dx.doi.org/10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2

Bikos, D., D. T. Lindsey, J. Otkin, J. Sieglaff, L. Grasso, C. Siewert, J. Correia Jr., M. Coniglio, R. Rabin, J. S. Kain, S. Dembek, 2012: Synthetic satellite imagery for real-time high-resolution model evaluation. *Wea. Forecasting*, **27**, 784-795.

Blahak, U., 2007: RADAR_MIE_LM and RADAR_MIELIB - Calculation of radar reflectivity from model output. *Internal Rep.*, Institute for Meteorology and Climate Research, University/Research Center Karlsruhe, 150 pp.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model description and implementation. *Mon. Wea. Rev.*, **129**, 569–585.

Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606,85 pp. [Available from NASA Center for Aero Space Information, 800 Elkridge Landing Road, Linthicum Heights, MD 21090-2934.]

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

Clark, A. J., J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia, Jr., M. Xue, and F. Kong, 2013: Tornado path length forecasts from 2010-2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387-407.

Du, J. and Coauthors, 2009: NCEP Short-Range Ensemble Forecast (SREF) system upgrade in 2009. Extended Abstracts, 19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting, Omaha, NE, Amer. Meteor. Soc., 4A.4. [Available online at http://ams.confex.com/ams/23WAF19NWP/techprogram/paper_153264.htm.]

Dudhia, Jimy, 1989: Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model. *J. Atmos. Sci.*, **46**, 3077–3107. doi: http://dx.doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.*, **21**, 457–469.

Gremillion, M. S. and R. E. Orville, 1999: Thunderstorm characteristics of cloud-to-ground lightning at the Kennedy Space Center, Florida: A study of lightning initiation signatures as indicated by the WSR-88D. *Wea. Forecasting*, **14**, 640-649

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.

Hong, S.-Y., and J.-O. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.

Janjic, Z.I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.

Kain, J. S., S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. B. Weber, K. W. Thomas, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931-952.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536-1542.

Kain, J. S., M. C. Coniglio, J. Correia, A. J. Clark, P. T. Marsh, C. L. Ziegler, V. Lakshmanan, S. D. Miller, S. R. Dembek, S. J. Weiss, F. Kong, M. Xue, R. A. Sobash, A. R. Dean, I. L. Jirak, and C. J. Melick, 2013: A Feasibility Study for Probabilistic Convection Initiation Forecasts Based on Explicit Numerical Guidance. *Bull Amer. Meteor. Soc.*, **94**, 1213-1225.

Koch, S. E., B. Ferrier, M. Stolinga, E. Szoke, S. J. Weiss, and J. S. Kain, 2005: The use of simulated radar reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts. *Preprints, 12th Conf. on Mesoscale Processes*, Albuquerque, NM, Amer. Meteor. Soc., J4J.7. [Available online at http://ams.confex.com/ams/pdfpapers/97032.pdf.]

Kong, F., and Coauthors, 2008: Real-time storm-scale ensemble forecast experiment—Analysis of 2008 spring experiment data. *Preprints, 24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.3. [Available online at http://ams.confex.com/ams/pdfpapers/141827.pdf.]

Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*,**138**, 1587–1612.

McCaul, E. W., S. J. Goodman, K. M. LaCasse, D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. *Wea. Forecasting*, **24**, 709–729.

Mecikalski, J. R., K. M. Bedka, 2006: Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES

imagery. *Mon. Wea. Rev*., **134**, 49–78.

Milbrandt, J. A., M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**, 3065–3081. doi: http://dx.doi.org/10.1175/JAS3535.1

Miller, S. D., G. W. Carbin, J. S. Kain, E. W. McCaul, A. R. Dean, C. J. Melick, and S. J. Weiss, 2010: Preliminary investigation into lightning hazard prediction from high resolution model output. *Preprints, 25th Conference on Severe Local Storms*, Amer. Meteor. Soc., Denver, CO. Paper 4B.1

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682.

Morrison, H., J. A. Curry, and V. I. Khvorostyanov, 2005: A new double-moment microphysics parameterization for application in cloud and climate models, Part I: Description. *J. Atmos. Sci.*, **62**, 1665–1677.

Roberts, R. D., S. Rutledge, 2003: Nowcasting storm initiation and growth using GOES-8 and WSR-88D data. *Wea. Forecasting*, **18**, 562–584.

Rogers, E., and Coauthors, 2009: The NCEP North American Mesoscale modeling system: Recent changes and future plans. *Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 2A.4.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y., Huang, W. Wang, and J. G. Powers, 2008: A description of the Advanced Research WRF Version 2, NCAR Tech Note, NCAR/TN-475+STR, 113 pp. [Available at: http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.]

Smith, P. L., 1984: Equivalent radar reflectivity factors for snow and ice particles. *J. Climate Appl. Meteor*., **23**, 1258–1260.

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. *Wea. Forecasting*, **26**, 714-728

Thompson, G., P. R. Field, W. R. Hall, and R. Rasmussen, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115.

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. Phys.*, **82**, 139–170.

Xue, M., and Coauthors, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *Preprints, 24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.2. [Available online at http://ams.confex.com/ams/pdfpapers/142036.pdf.]

Zhang, J., and Coauthors, 2011: National mosaic and multi-sensor QPE (NMQ) system: description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338.

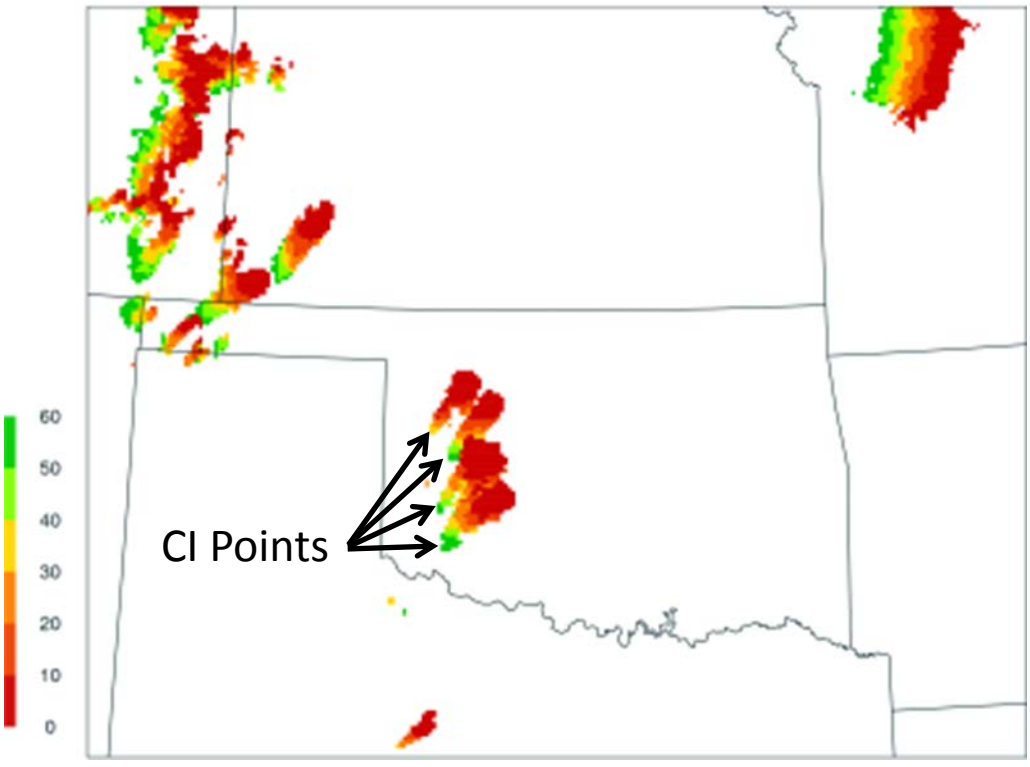


Fig. 1. Convective Activity (CA) points color coded to reflect the 10 minute period within the preceding that they became active. Convection Initiation (CI) points are identified as the earliest instance of CA within each "object" that initiated during the hour.

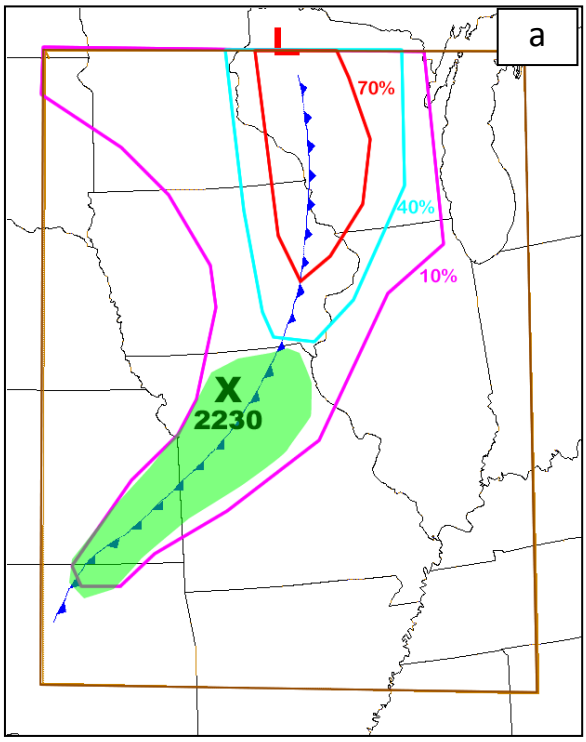
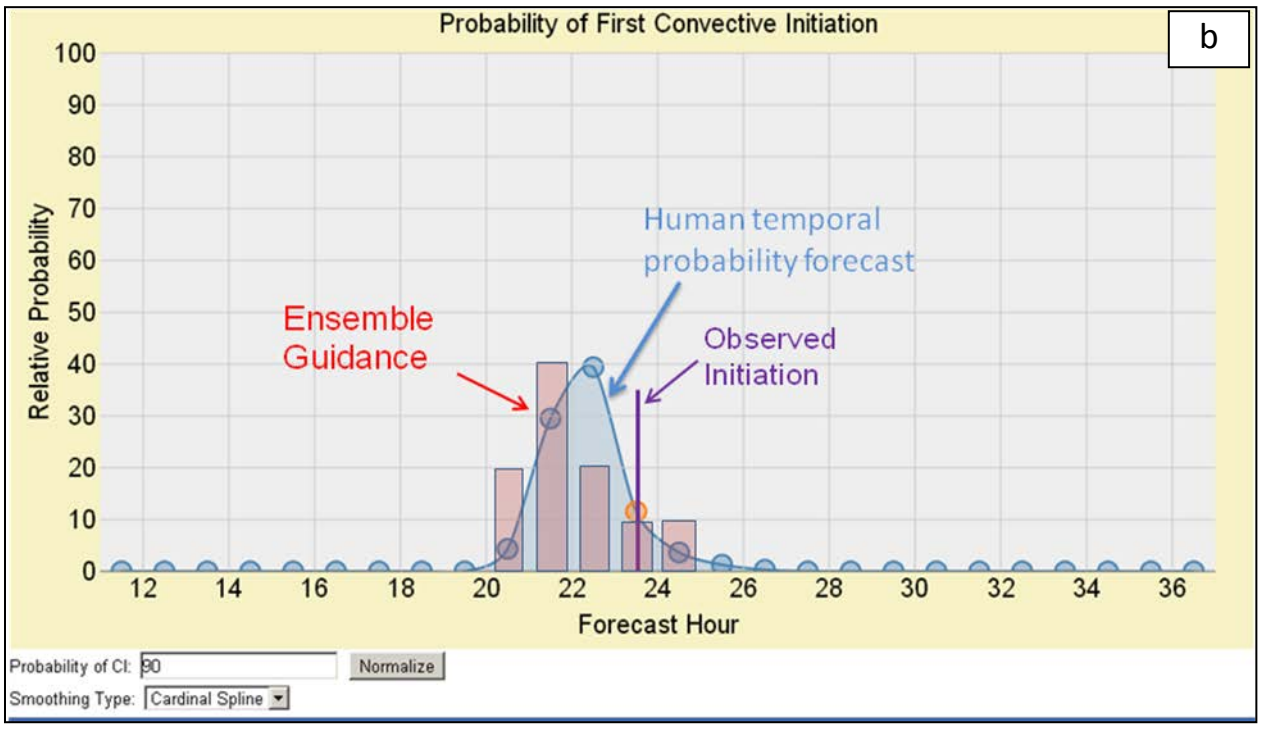


Fig. 2. The two components of the experimental CI forecast product. a) spatial probabilities (probability of convection within 20 km of any point) over a 4-h time period. The green shaded area is the focus area for CI timing forecasts, within which the team-consensus most likely CI time (+/- 0.5 h) and location are indicated; b) the probabilistic CI timing forecast, overlaid with a histogram representing the distribution of timing in CAPS ensemble members (used as guidance for the forecast) and a marker indicating the initiation time that was observed on



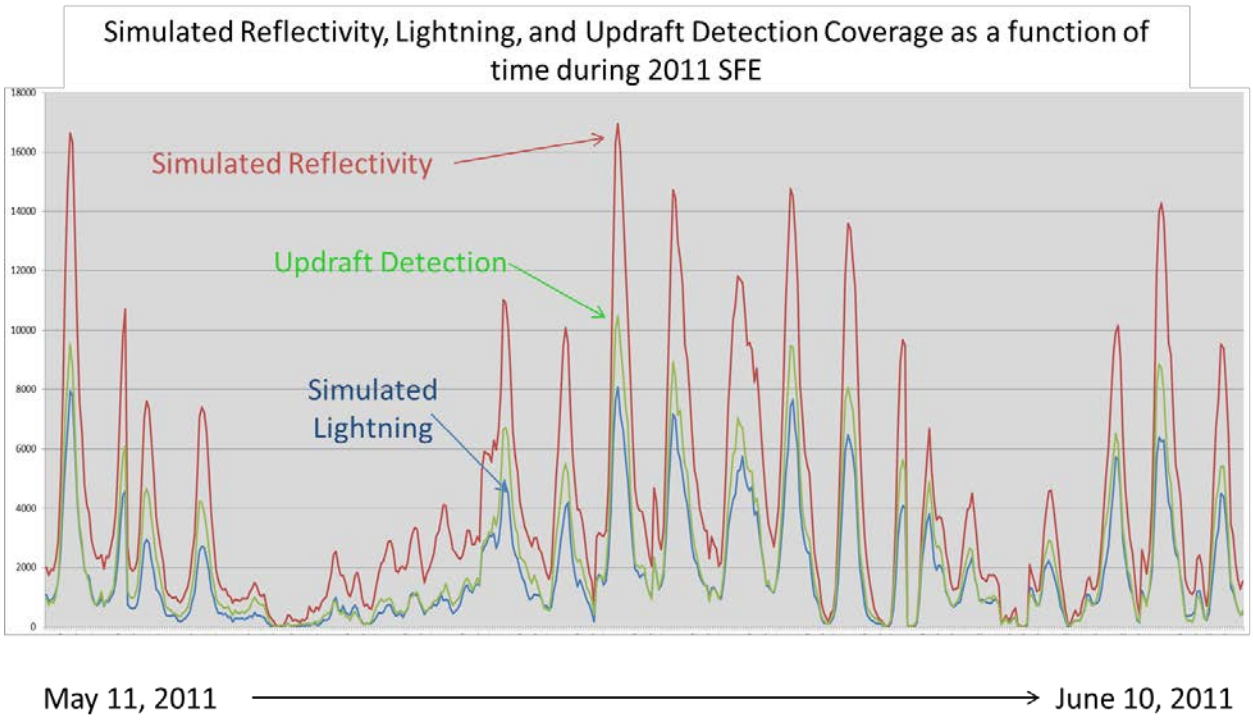


Fig. 3. Time series of the number of CA points activated each hour for each of the CA proxies used during SFE2011.

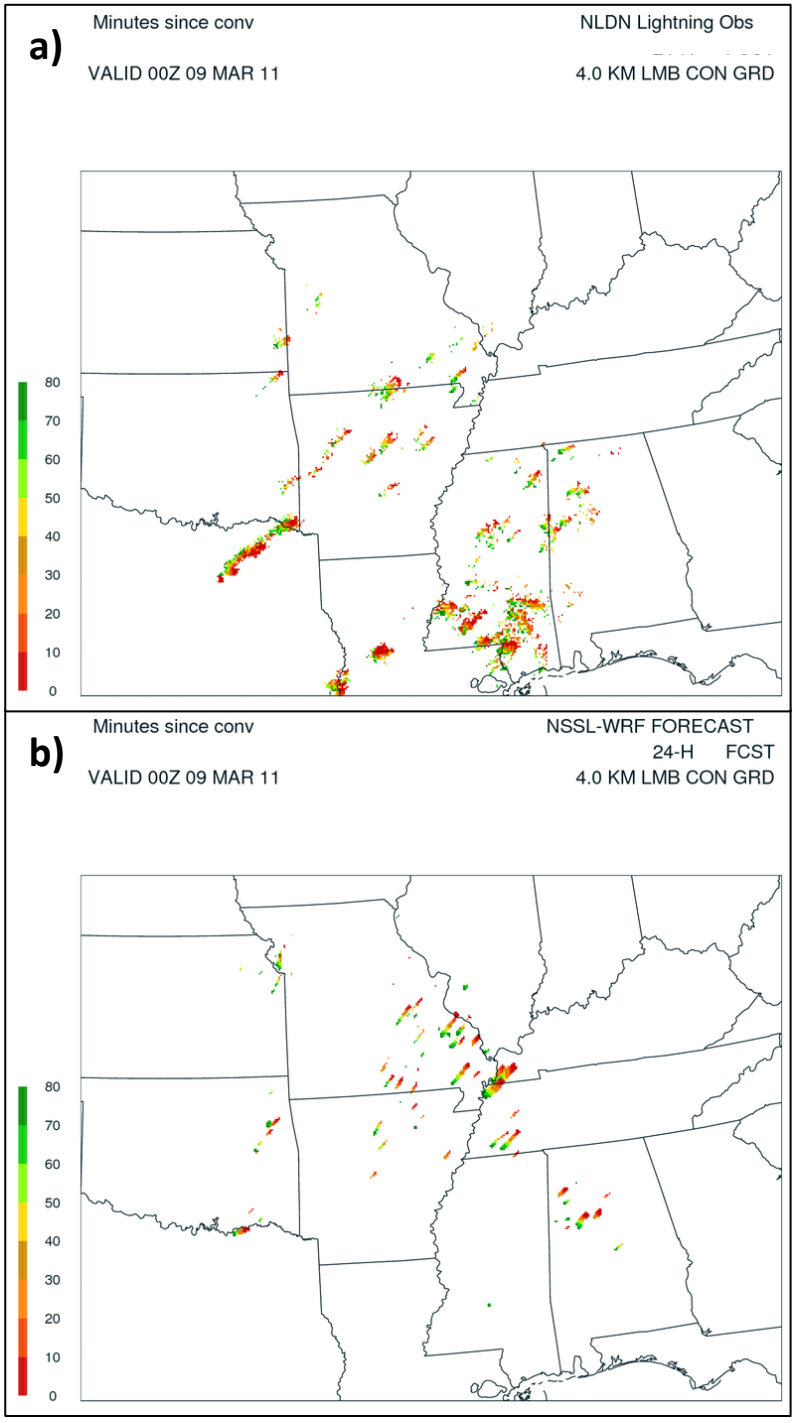


Fig. 4. As in Fig. 1, but for a) NLDN observations and b) the SLTG simulated lightning field. Both (a) and (b) are valid at 0000 UTC on 9 March 2011.

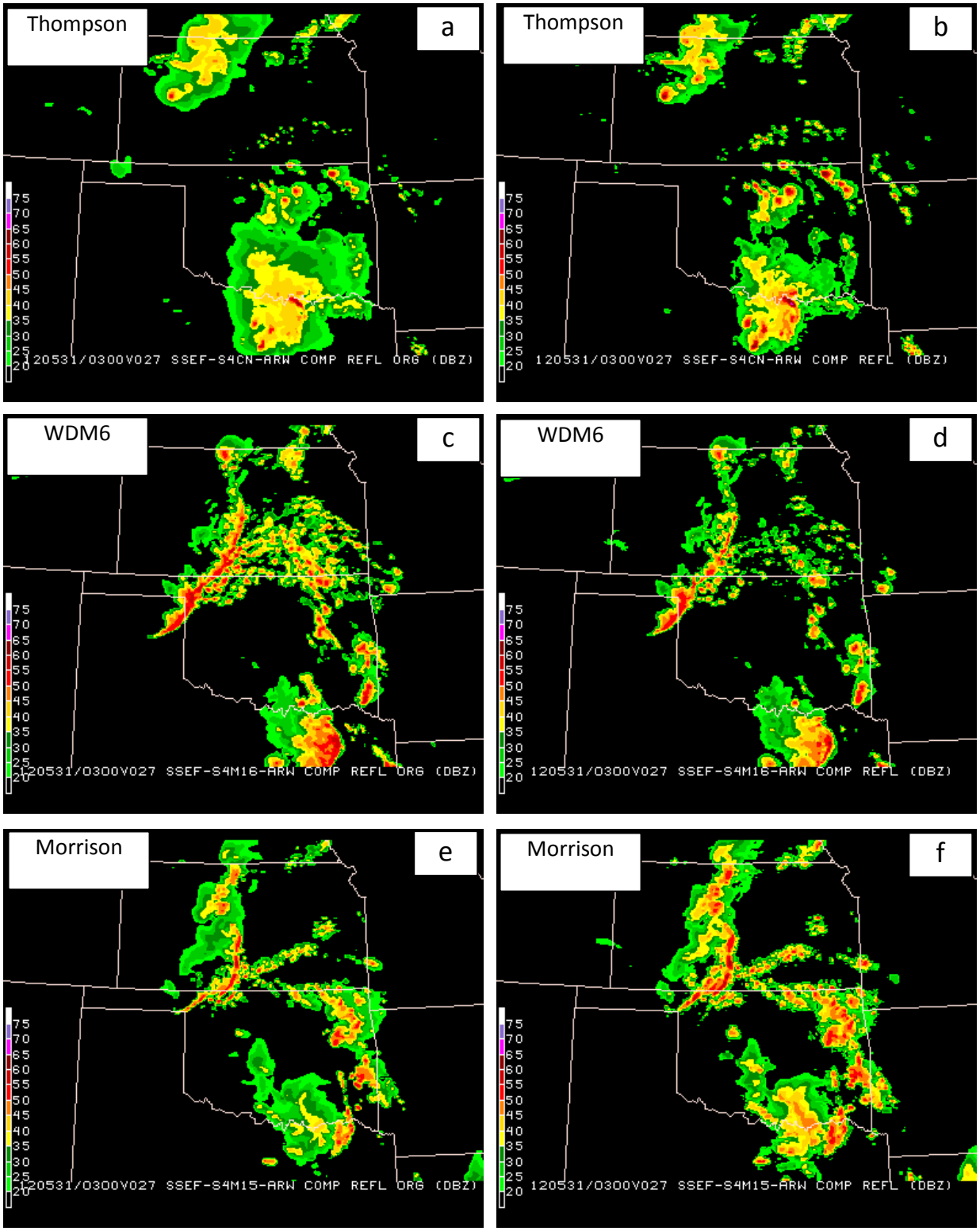


Fig. 5. Comparative example of simulated reflectivity fields derived from the "old" (Kain et al. 2008) reflectivity calculation (left side) and the new, MP-consistent calculation (right side) for the Thompson (a, b), WDM6 (c, d), and Morrison (e, f) MPs.

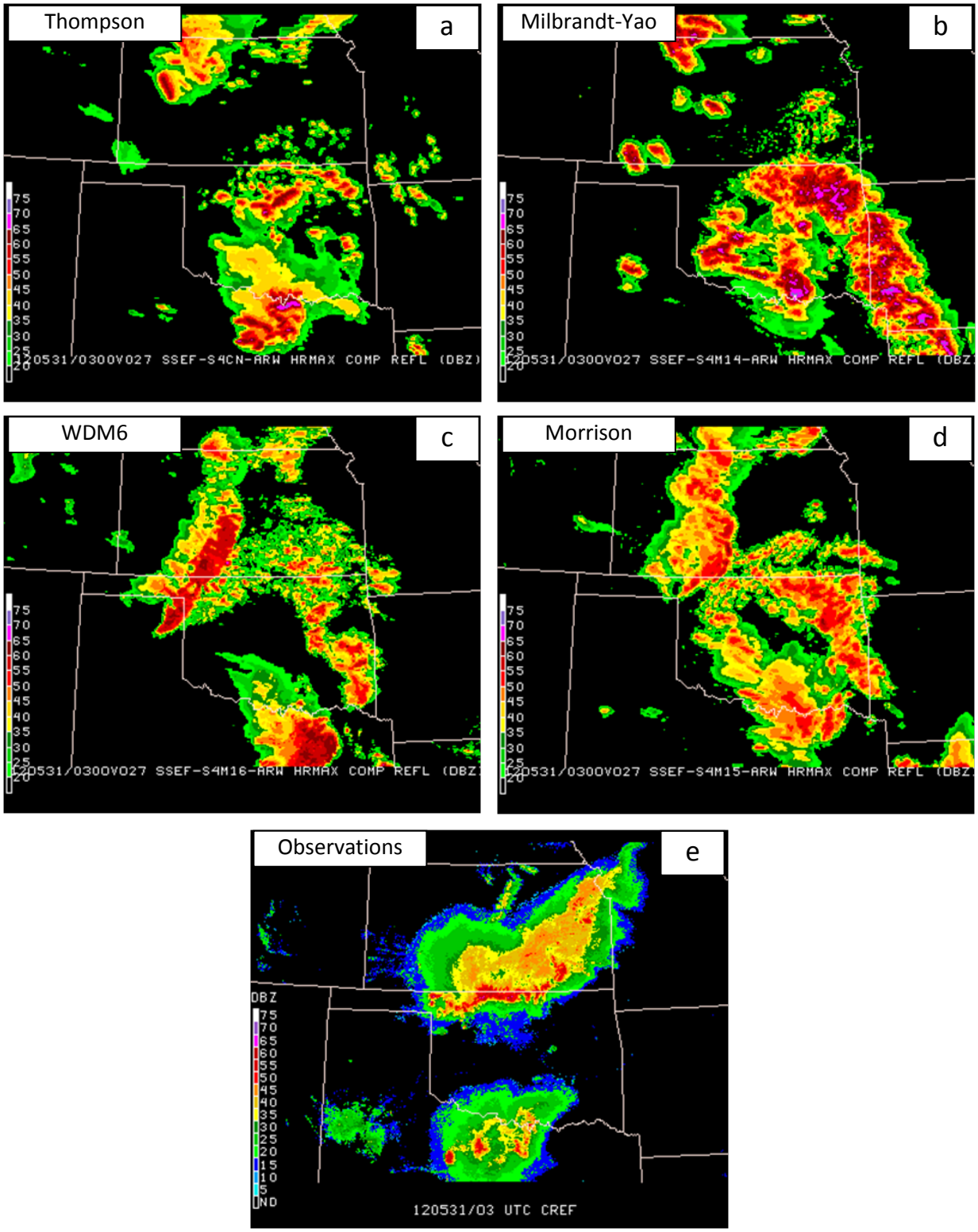


Fig. 6. Example of differences in simulated HM-CREF (see text) as a function of MP, showing a) Thompson, b) Mibrandt-Yau, c) WDM6, d) Morrison, and e) observed HM-CREF.

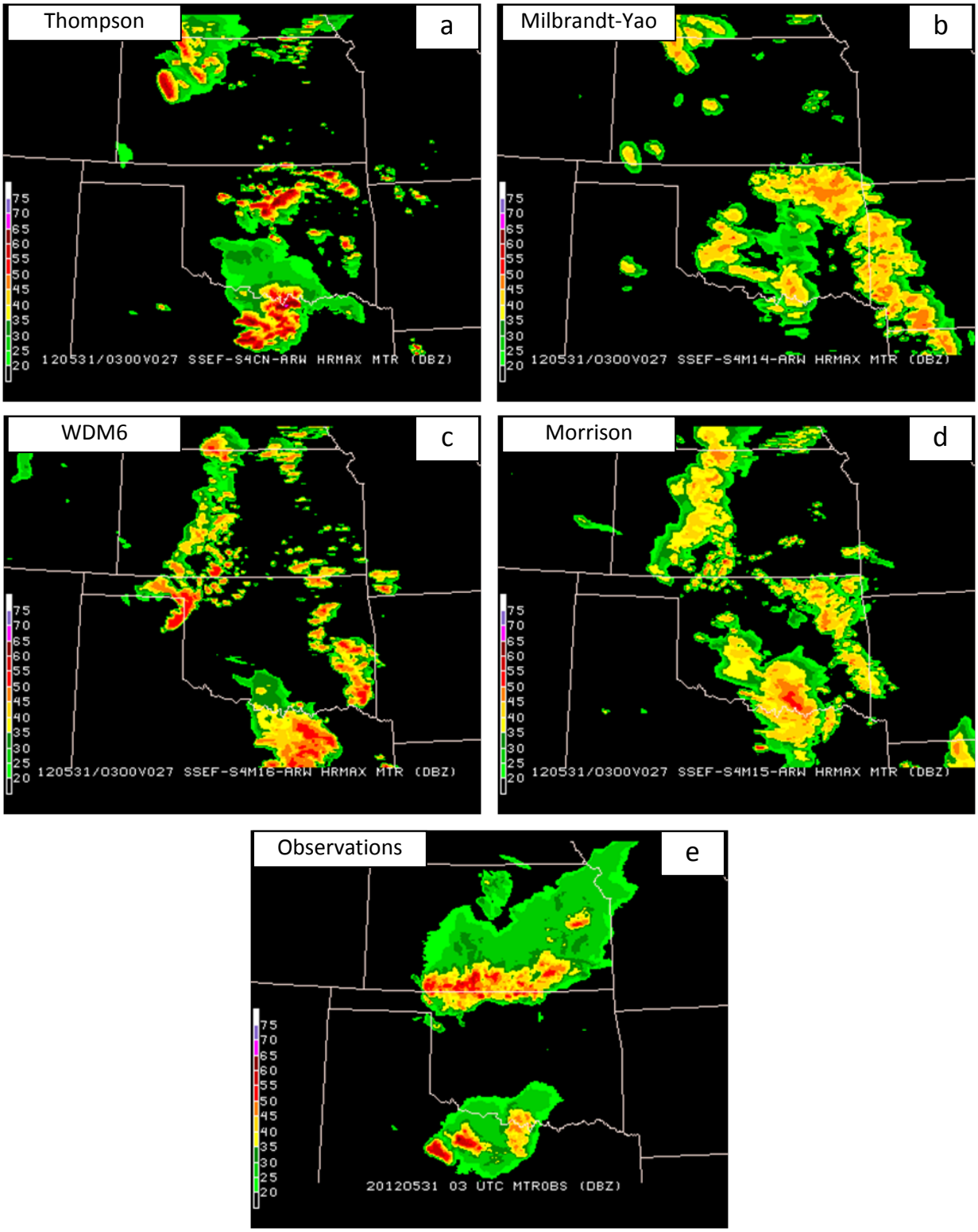


Fig. 7. As in Fig. 6, but for the simulated and observed HM-MTR fields (see text).

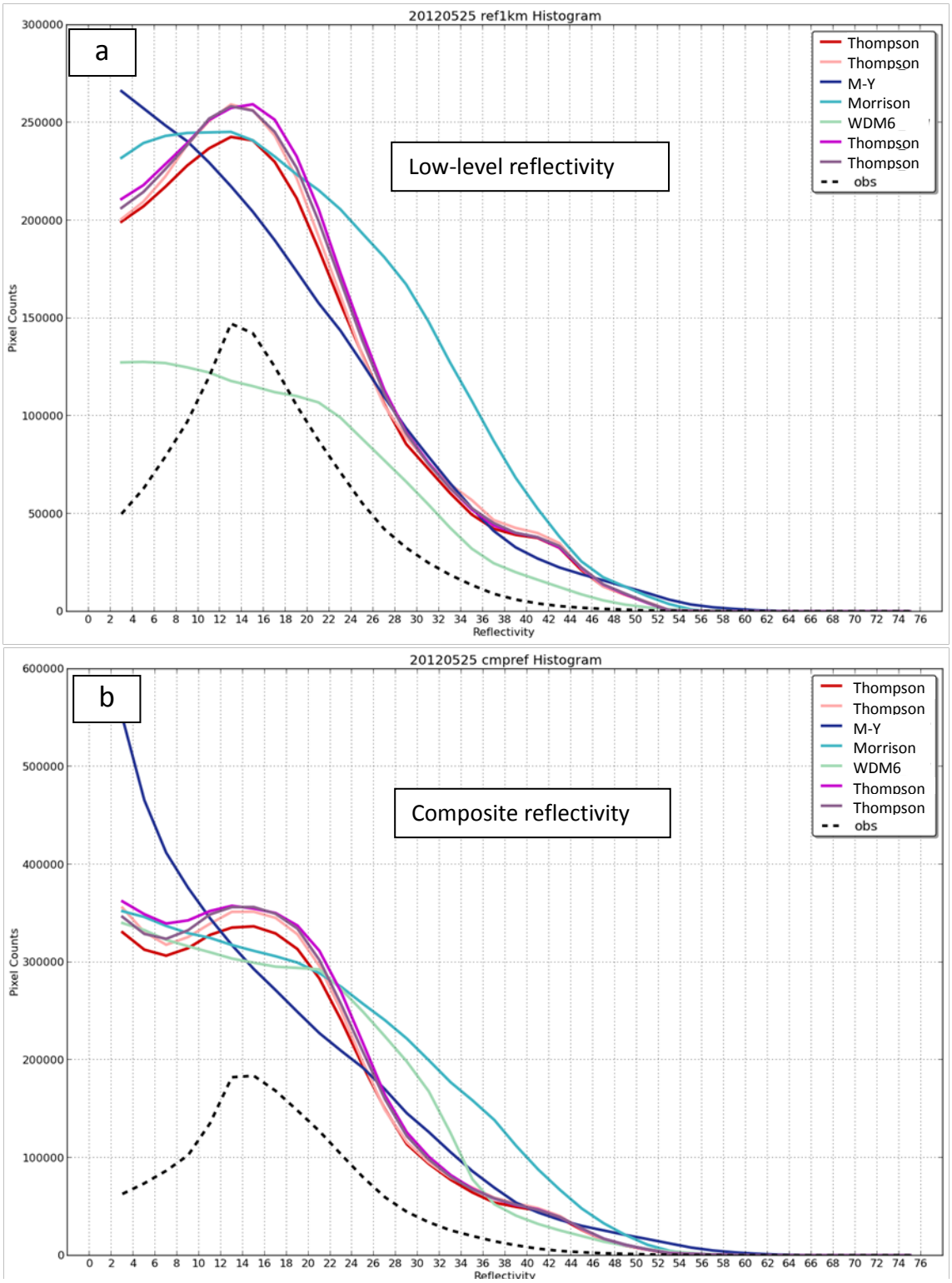


Fig. 8. Examples of reflectivity histograms as a function of microphysical parameterization for the Physics members of the CAPS ensemble. The curves represent the grid-point count in each 2-dBZ reflectivity bin, derived from the 24-h forecast/observation period starting at 1200 UTC on 25 May 2012, based on hourly snapshots and including all grid points in the full CAPS output domain.

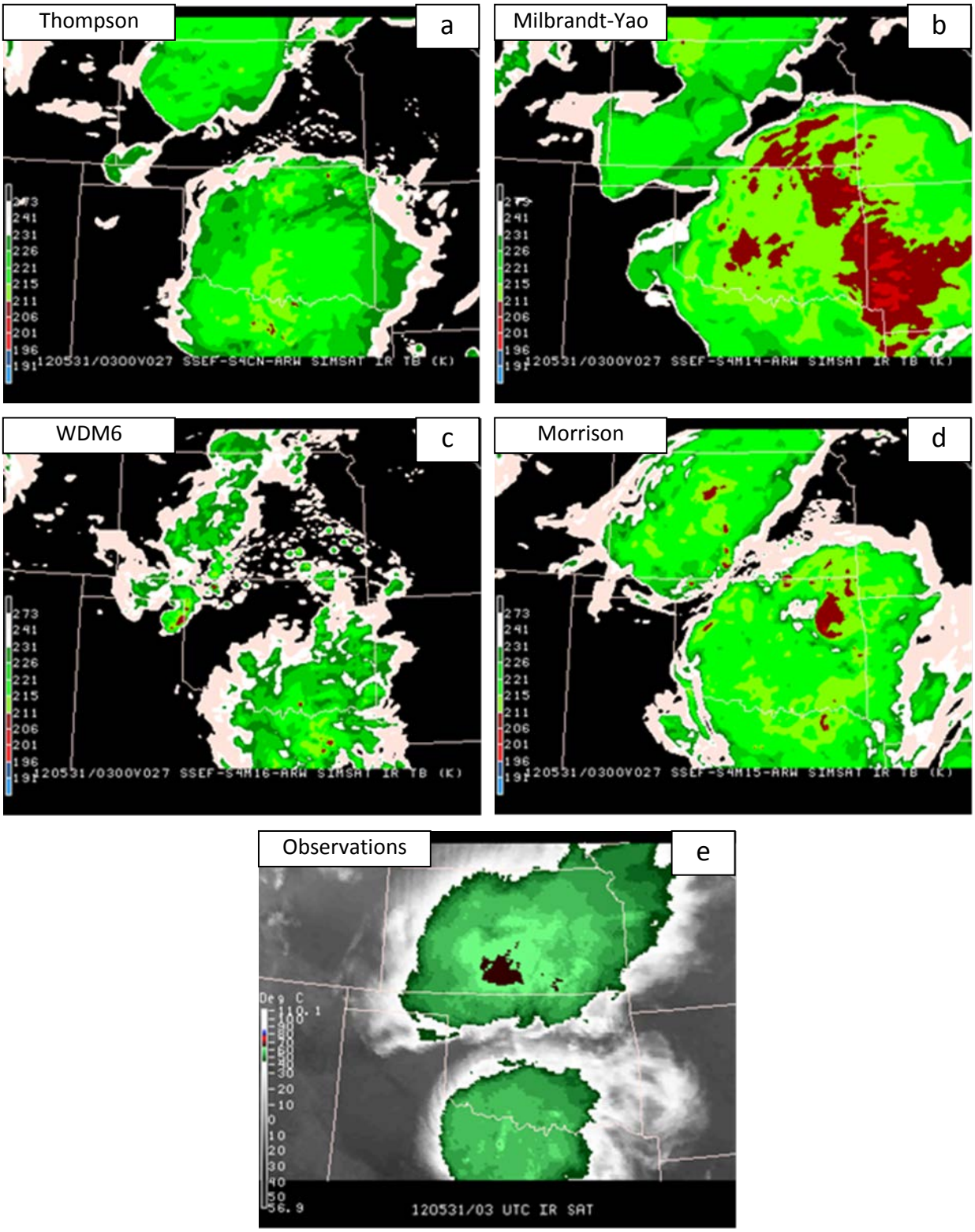


Fig. 9. As in Fig. 6, but for the simulated and observed IR brightness temperature.

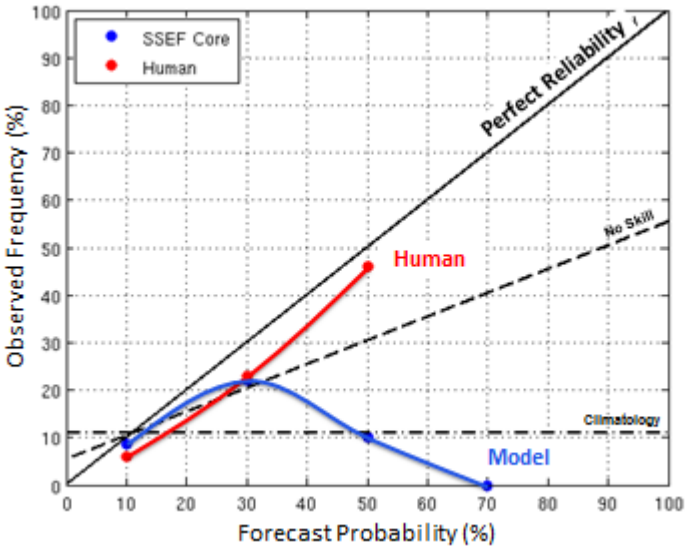


Fig. 10. Reliability diagram for the daily temporal probability forecasts and model guidance.