Convective Initiation Component of the 2011 HWT Spring Forecasting Experiment

Preliminary Report

John S. Kain[1], Michael C. Coniglio[1], James Correia[2,3], Adam J. Clark[1,3], Patrick T. Marsh[1,3,4] Conrad L. Ziegler[1], Valliappa Lakshmanan[1,3], Stuart D. Miller[4], Scott R. Dembek[3], Steven J. Weiss[2], Fanyou Kong[4,5], and Ming Xue[4,5], Ryan A. Sobash[1,4], Andrew R. Dean[2], Israel L. Jirak[2], Christopher J. Melick[2,3]

[1]NOAA/National Severe Storms Laboratory, Norman, OK

[2]NOAA/Storm Prediction Center, Norman, OK

[3]Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK

[4]University of Oklahoma School of Meteorology, Norman, OK

[5]Center for Analysis and Prediction of Storms, Norman, OK

1. Introduction

The annual Spring Forecasting Experiment (SFE) is conducted in the NOAA Hazardous Weather Testbed during the climatological peak of severe convective weather in the U.S. (Clark et al. 2011). This experiment is uniquely designed to bring together meteorological scientists and practitioners to work on emerging problems of mutual interest. In 2011, a major component of the SFE (hereafter SFE2011) was a pilot study on convective initiation (CI). The focus on CI was motivated by a growing awareness that its prediction is one of the weak links in forecasts of thunderstorm activity, presenting a significant challenge for forecasters, including those who specialize in prediction of severe convective weather, flash flooding, hazards to aviation, and other specific threats.

Although many thunderstorm forecasts implicitly include information about when and where storms are likely to form, the NWS does not issue explicit forecasts for CI. In fact, to our knowledge, there is no generally accepted, specific definition of CI in the meteorological community. From the perspective of a research meteorologist, CI might be defined as a process that occurs in the context of a single cloud as buoyant parcels begin to accelerate rapidly upward and produce precipitation. If this process occurs in the absence of any other deep moist convection, it can be labeled unambiguously as a CI event, but observations show that most thunderstorms are associated in some way with previous storms. This association could come in the form of isolated initiation along the downdraft outflow from a pre-existing storm, generation along the periphery of an organized mesoscale area of convection, or a development associated

with a multitude of other interactions. "Clean slate" CI events are relatively rare and oftentimes it is not clear whether a particular convective updraft should be labeled as a separate CI event or as a new element within ongoing activity. These details, while interesting, are likely to be secondary considerations for operational forecasters, who may be more concerned about what new convection will lead to rather than refining the definition of CI.

Because of these ambiguities, an overarching goal of the CI component of SFE2011 was to define the challenges of CI prediction and establish a framework for additional studies and possible routine forecasting of CI. It was decided that this framework should be applicable to both convection-allowing models (CAMs), which have become a fundamental part of the SFE in recent years, and to observations, so that meaningful verification statistics could be derived.

The first challenge was to come up with objective diagnostics to determine whether deep moist convection was active at any point on a grid. The second was to identify a subset of these points that should be labeled as initiation (CI) points. A third major challenge was to investigate ways to generate guidance for CI forecasts and use this guidance to issue experimental probabilistic forecasts for CI. A corollary objective was to develop and examine new diagnostic tools that would provide insight into the CI process within CAMs and provide new ways of manipulating high-resolution ensemble output. More information on each of these topics is provided in the next section.

2. Data and Methodology

a. Convection-Allowing Models (CAMs)

Following is a brief summary of the modeling systems in which CI diagnostics were embedded during SFE2011. Additional details about these model configurations can be found online in this document:
http://hwt.nssl.noaa.gov/Spring_2011/Spring_Experiment_2011_ops_plan_13May_v5.pdf

1) *NSSL-WRF*: SPC forecasters have used output from this experimental 4 km WRF-ARW produced by NSSL since the fall of 2006. This WRF model is run once daily at 00 UTC throughout the year over a full CONUS domain with forecasts to 36 hrs. Initial and lateral-boundary conditions (IC/LBCs) are provided by the operational NAM. Output is also available on the internet 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 SFEs, as was the case for the CI diagnostics used in SFE2011.

2) *CAPS ensemble*: CAPS produced a 50 member storm-scale ensemble Storm Scale Ensemble Forecast (SSEF) system with grid spacing of 4 km and forecasts to 36 hrs covering a CONUS domain. A total of 28 of these members (all using the WRF-ARW dynamic core) contained the CI diagnostic code. Ten of the 28 used identical initial conditions but different

parameterizations of the planetary boundary layer (PBL), while the remaining 18 were initialized with both IC/LBC and physics perturbations.

3) *High Resolution Rapid Refresh (HRRR)*: The experimental 3 km HRRR model is nested within the hourly development version of the 13 km Rapid Refresh (RR) model, which provides IC/LBCs. The HRRR uses a version of the WRF-ARW with generally RUC-like physics. A unique aspect of the RR is the hourly GSI data assimilation system that incorporates a wide array of observational datasets including radar reflectivity via the radar-Diabatic Digital Filter Initialization. The HRRR integration is run over a full CONUS domain with forecasts to 15 hrs.

b. Objective Identification of Deep Moist Convection in CAM Forecasts

There is no universally valid definition for the existence of deep moist convection, either in model or observational data. During SFE2011 we considered the viability of three distinctly different sets of criteria to identify "convectively active" (CA) grid points in CAM forecasts:

1) 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 by selecting an appropriate threshold value of FRD. The threshold value was determined by mapping National Lightning Detection Network (NLDN) strike data to the model framework, then comparing the climatology of model-predicted FRD 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 2011 Spring Experiment, 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. Thus, for the simulated lightning (SLTG) criteria set, a CA grid point was identified as any point at which the FRD exceeded 0.55 km^{-2} $(5\ min)^{-1}$

2) 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), OR the maximum rain mixing ratio exceeded a threshold value QR (g/kg), OR both conditions were met. Preliminary threshold values of W = 5 m/s, QG = 2 g/kg, and QR = 1 g/kg were selected 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 "SWQQ" set of criteria for CA.

3) Simulated reflectivity: The 35dBZ threshold for simulated reflectivity (computed as in Kain et al. 2008) was used to identify a CA grid point, 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 exceeded at the -10° C level (see Gremillion and Orville 1999). This defined the simulated minus-10 reflectivity"SMTR35" criteria set for CA.

CA points were identified and tracked independently for each of these sets of criteria. They were updated at the end of every model-integration time step (~20 - 25s) for the SLTG and SWQQ definitions. For the SMTR35 definition, CA points were identified approximately every 5 minutes during integration, to emulate the frequency of available updates for observed reflectivity data from the nationwide network of WSR-88D radars. Thus, every CA event, for all three criteria sets, was assigned a grid box and a time-step bin (time window).

c. Default Method for Identifying CI Points as a Subset of CA Points in CAM Forecasts (and Corresponding Observations)

A preliminary set of rules was developed to allow for identification of convective initiation (CI) points as a subset of the CA points. These rules were applied independently to each of the SLTG, SWQQ, and SMTR35 datasets, yielding three corresponding SLTG, SWQQ, and SMTR35 datasets for CI. The basic idea of this algorithm was to identify CI grid points as those CA points that could be characterized by "new" convective development, i.e., convection *not* associated with ongoing activity, whether local or moving in from surrounding areas. As implemented within the WRF model, a grid point was labeled a CI point at a given time if 1) it was convectively active (i.e., a CA point), 2) it had not been labeled a CI point within the preceding hour, and 3) none of the adjacent grid points were convectively active (CA) during the preceding time-step bin. The time and space thresholds used in steps 2 and 3 are somewhat arbitrary. In step 2, one hour was chosen because output grids are saved hourly. In step 3, a single time step and single grid-point neighborhood were selected for computational efficiency – to avoid saving two-dimensional model grids every time step during integration. Additional studies are being conducted to assess the sensitivity to these parameters. Hereafter, this algorithm for identifying CI points as a subset of CA points will be referred to as the CI_1 algorithm

d. Observational Datasets for Verifying CAM Forecasts of CA and CI

Observational datasets were collected for the purpose of verifying/validating the convective activity diagnosed in the CAMs. There was no directly applicable verification dataset for the SWQQ-derived data, but NLDN data and observed radar-reflectivity data may be useful analogs for the SLTG and SMTR35 datasets, respectively.

1) NLDN data. These data, which consist almost exclusively of observed cloud-to-ground (CG) lightning strikes (see http://gcmd.nasa.gov/records/GCMD_NLDN.html), were collected in the form of a value for time, latitude, and longitude for each lightning strike. Each lightning point was considered a CA point.

2) Observed WSR-88D Reflectivity: As part of their National Mosaic and Multi-Sensor QPE initiative (NMQ), NSSL has implemented a system that produces a national (CONUS) 3D radar mosaic grid with a 1-km horizontal resolution over 31 vertical levels and a 5-minute update cycle (see http://www.nssl.noaa.gov/projects/q2/nmq.php for additional details). Using these data, combined with analyzed temperature fields from the RUC, any point at which observed reflectivity exceeded 35 dBZ at the -10º C level was marked as a CA point and each point was saved in the same format as the NLDN data: time, latitude, and longitude. These points comprised the raw observed minus-10 reflectivity (OMTR35) dataset

Both the NLDN and OMTR35 data points were mapped to the spatial and temporal coordinates of the model data as the first step in the verification process. Specifically, the observed data was mapped to the nearest model grid point and appropriate time-step bin. The size of the latter was a single model-integration time step for NLDN data (~20 – 25s, depending on the model) and ~ 5 mins for the OMTR35 data. CI points were generated from the NLDN and OMTR35 datasets by applying the same algorithm used for the model data to extract CI points as a subset of the corresponding CA points – the CI_1 algorithm.

e. Output format for raw CA and CI fields.

Two-dimensional grids of CA and CI fields were generated every time step during WRF-model integration (using the CI_1 algorithm), but it was prohibitively expensive to archive these 2-D fields every time step. Thus, as a compromise, CA and CI activity was denoted by a time stamp at every grid point and 2-D fields containing this information were saved at regularly-scheduled model-output intervals, i.e., hourly. Specifically, CA fields for each set of criteria contained the number of minutes since the last convective activity at each point. Likewise, CI fields denoted the number of minutes since each grid point had been labeled a CI point. This format lacked explicit information about convective activity at every grid point every time step, but it nonetheless provided very useful sub-hourly information about CI and CA. Separate CI and CA grids were available for each set of criteria for both model output and observations.

f. Display and Post-processing of raw CA and CI output fields

Displays showing the raw CA fields are useful for visualizing the coverage of convective activity as a function of time, and when storms are sufficiently isolated, clear storm tracks are evident. For example, the left-side panels of Fig. 1 show an example of the raw CA field from a 24 h forecast of the NSSL-WRF, valid 00 UTC 25 May 2011. Individual storm tracks and storm mergers are easily discernible in some areas, e.g., northeastern Texas. These panels also indicate that the three different definitions of CA lead to some differences in identification of

convectively active grid points. For this particular output period, the SLTG definition yields the smallest number of individual active points, the SWQQ definitions indicates somewhat higher coverage, while the SMTR35 definition flags still more points. Generally speaking, however, the different definitions highlight the same general areas and, for the most part, the same specific simulated storms. The differences are mostly a matter of porosity and size of individual features. It appears that SMTR35 has the lowest threshold for activation, followed by SWQQ, then SLTG.

The panels on the right side of Fig. 1 depict different ways of viewing the CI field. The small '+' signs in these panels indicate the specific locations of diagnosed CI points associated with each definition of CA. The exact locations of the individual points and the detailed patterns formed by nearby points are quite different, but the general areas of CI activity are similar. This similarity is underscored by the similar color-fill patterns in these images. The color fill is derived using kernel density estimation (KDE) as a technique for creating spatial probability fields from deterministic prediction of individual CI events. It is appropriate to use a technique like this because it is well known that CAM forecasts such as those used in SFE2011 have little skill in pinpointing the location of individual storms.

The specific KDE methodology is similar to that used by Sobash et al. (2011). The first step in this method is to search a neighborhood around each grid point for a "hit", i.e., a diagnosed CI point as indicated by a '+' in Fig. 1. In this case, a neighborhood radius of 40 km (~ 25 miles) is used. If a grid point has one or more hits in its neighborhood, the grid point itself is assigned a preliminary probability of 100% - specifically, 100% probability of CI within 40 km during the time period in question. Grid points with no hits in their neighborhoods are assigned a preliminary probability of 0%. The final probability field, as shown in Fig. 1, is derived by applying a Gaussian smoother to all of the preliminary probability values. This smoothing process is depicted in a one-dimensional framework for a single grid point and different values of Gaussian standard deviation (σ) in Fig. 2. It has the effect of conserving domain-wide probability while distributing the influence of each deterministically-predicted CI event over a user-specified neighborhood. A range of different σ values was applied for sensitivity testing during SFE2011, but a fixed value of 30 grid points was used to generate the probability fields in Fig. 1. For a given raw CI field, smaller (larger) σ values yield higher (lower) amplitude, lower (higher) coverage probability fields. Regardless of the σ value, the formal definition of the probability field remains "the probability of CI with 40 km of any point within the output time period". With appropriate verifying observations, the optimal σ value could be estimated on the basis of past model performance, as in Marsh et al. (2012), but such observations are not available at this time.

As guidance for human forecasts of CI, these raw-CA and probabilistic-CI plots formed the basis for all guidance products that were examined. For example, probabilistic CI forecasts for the CAPS ensemble were produced by averaging the KDE-derived probability fields from individual members.

g. Alternative methods for diagnosing CI

1) Warping and filtering of observed radar reflectivity

With this technique, the raw OMTR35 dataset was used to identify convectively active points on a 1 km CONUS grid. Radar images 5 minutes apart were searched for grid points where reflectivity threshold (35 dBZ) was exceeded at current time, but not at the previous sample time. To prevent detection of small, noisy pixels, the images were median filtered using a 5x5 median filter. The effect of this median filtering is that only storm cells that are at least 13 km^2 are considered to represent convective initiation. To account for movement of storm cells, the image at one frame is warped (by a maximum movement of 60 km/hr) to align it with the next image before the images are compared. To account for new updrafts within the same storm, a cell is considered new only if there are no convective pixels within X km of it. We experimented with distance thresholds ranging from 3km to 25km and from forecaster feedback, finally settled on 15km. Based on the comparison, pixels are classified into one of four categories: (a) ongoing convection (b) decayed convection (c) new convection and (d) none. (Fig. 3).

Finally, three impossible situations are identified and removed. These were: (a) new convection connected to decayed convection is changed to "ongoing" (b) new convection connected to ongoing convection is changed to "new" and (c) decayed convection connected to ongoing is changed to "none" (Fig. 4). Hereafter, this method for identifying CI points is referred to as algorithm CI_2

2) A 3D object based algorithm

The "object-based" CI algorithm utilizes a 3-dimensional (in space and time) object identification algorithm. Objects are defined using the convective activity (CA) field, which is based on the 35 dbz reflectivity threshold at -10º C. The value of the raw CA field is the number of model time steps since the top of the hour being considered. For use in the CI algorithm, the raw CA values are converted to the number of time steps since forecast hour 36, which is the latest forecast hour available from the SSEF system forecasts. The 3-dimensional objects are defined as groups of contiguous (or adjacent) grid-points in space and time with non-zero CA values. In other words, for each grid-point with a non-zero CA value, if there are adjacent grid-points at the same time, next time, or previous time with non-zero CA values, these grid-points are part of the same object. The only other criteria for classification as a CA object is that the object encompass a time period of at least 30 minutes. Computing the time period encompassed by the object simply involves using the maximum and minimum CA value within each object. The threshold of 30 minutes is an adjustable parameter and was chosen somewhat arbitrarily to avoid classifying "CI failures" (i.e., short-lived moist updrafts that are suppressed immediately after initiating for whatever reason) as actual CI events.

CI points are defined using grid-points of local time minima within CA objects. These local minima are found by searching within a specified 3-d radius (in space and time) of each grid-

point within an object to find whether that grid-point has the largest (i.e., earliest) CA value within the radius. Using local time minima allows not only the grid-points with the earliest CA values in CA objects to be used to define CI points, but also grid-points that represent formation of storms that initiate later than the first convection in the object, and then merge with the object. Thus, multiple CI events can be associated with a single 3-d CA object. Usually, the object-based CI algorithm identifies a group of adjacent CI points that represent the initiation of a single storm. When this occurs, the algorithm finds the centroid grid-point of the CI points and this point is defined as a "CI event". Otherwise, if a CI point has no other CI points adjacent to it, then this CI point is classified as a "CI event".

Hereafter, this method for identifying CI points is referred to as algorithm CI_3. The CI values determined using this algorithm have a systematic bias towards later times because they were actually derived from the CA field. Recall that each value in the CA field represents the *last* (most recent) time step that a given grid point was convectively active. The CA field was used for this algorithm because it provides contiguous coverage of the track of convectively-active objects; the default CI points (from the CI-1 algorithm) represent the first time step that a grid point was convectively active, but by design they provide an incomplete indication of total coverage of convection. Thus, the magnitude of the systematic bias toward later CI times is approximately equal to the average time that points remain convectively active. However, for comparisons with observations, this bias should disappear if CI_3 is also used to determine CI points in the observations, as done in the results shown below.

h. Human Forecasts for CI

Human forecasts of CI consisted of a spatial categorical outlook for a 3 hour period over which the forecasters had high confidence that CI was likely. Categorical risk was defined using a slight, moderate and high ranking where slight was essentially equivalent to a 10% chance of having CI via any of the CI definitions we investigated (reflectivity or lightning). The other categories had no specific percentages attached as we had very little guidance or evidence to assign probabilities. The slight definition followed that of SPCs definition of 10% chance of thunder occurring with 25 miles of a point via their experimental thunder outlook, the closest operational product we had available to our own forecast (with the exception that SPCs product is for a set 4 hour window versus our roaming 3 hour window).

In addition each forecaster chose a point on this 3 hour product where they expected the first storm to initiate. Each forecaster was responsible for entering a time, time window, and confidence (percent likely a storm will occur within your time window within 25 miles of your point).

i. Model diagnostics for CI-related physical processes

Model soundings were generated for 1100+ locations for the control and select physics members of the CAPS ensemble, totaling 18 members. Each of these soundings was then processed

through a custom built version of the Storm Prediction Center's National Sounding and Hodograph Analysis and Research Program (N-SHARP). Output for each forecast location, from each of the 18 members, was then post-processed into an ensemble-data format readable by the Warning Decision Training Branch's BUFKIT sounding analysis tool. This tool enabled very efficient visualization of large quantities of sounding output with powerful user controls. For example, it allowed users to view forecast soundings from 1 to 18 members simultaneously (overlaid, color coded) at any time and location.

A number of new fields were extracted from the CAPS ensemble ARW members to aid in our interpretation of the CA and CI fields. These variables were located specifically at model level 12 in the neighborhood of 1.1km AGL. The standard variables were vertical velocity, mixing ratio, temperature, pressure, and u and v component winds. Following Kain et al (2010) we also extracted unique variables such as the number of time steps where vertical velocity exceeded 0.25 m/s in the last hour, and the maximum vertical velocity in the last hour. These latter variables attempted to measure the persistence of updrafts as well as the maximum magnitude of updrafts respectively. In addition, we used standard surface variables and extracted fields such as dew point temperature gradients to identify mesoscale features such as drylines and outflow boundaries.

All of these variables were used to construct displays depicting relevant boundaries, CA timing, CI points, and storm locations via reflectivity. In this way we could trace the storms back to the initiating mechanism and make reasonable assessments as to their likelihood. In pursuing this we could also examine the nearby ensemble soundings to see the vertical profile of temperature and mixing ratio, and if these profiles were what we expected of the near storm initiation environment. A key component of understanding CI in the model is to confirm that the model correctly represents actual physical processes. The imposed time constraints did not allow for complete understanding but it did help to identify mechanisms that led to CI during the course of the experiment. These included horizontal convective roll circulations intersecting the dryline and/or sea breeze fronts as well as areas of convergence and moisture upwelling.

In the latter portion of the experiment, we attempted to automate the detection of the first storm for each ARW member in our selected domain of interest. We used this as a comparison for the timing that each forecaster submitted, treating the forecasters timing as a human ensemble to compare with that of the model ensemble. The algorithm for first storm simply recorded the time in the nearest 15 min bin for each member similar to what the human forecaster was asked to do. However, this approach was not without its flaws, since this algorithm considered the whole domain not just where forecasters were focused. This led to issues where spurious convection either in the form of peripheral convection (on the domain edge), or convection that lingered too long or convection that developed early could be detected. This resulted in a less than favorable comparison and highlighted the difficulty in automating this task.

3. Results

SFE2011 was conducted from 9 May through 10 June 2011. The primary objectives were to 1) evaluate the utility of different criteria used for automatic detection of CA in CAMs, 2) evaluate different methods to determine CI points as a subset of total convective points, and 3) quantify the skill of currently available CAMs for predicting CI. Secondary objectives included development and evaluation of new model diagnostic tools and visualization software.

The default CI algorithm (CI_1) was the only one that was used in real-time within the WRF model and it was applied extensively to observational data as well. These applications allowed us to identify deficiencies in CI_1. The alternative algorithms, CI_2 and CI_3 both show promise for ameliorating these deficiencies. This preliminary report focuses primarily on the results obtained using CI_1. Thus, unless otherwise stated below, CI data are derived using CI_1.

a. Sensitivity to different criteria for automatic detection of CA and compatibility with CI_1

1) Model Output

Probabilistic model guidance similar to that shown in Fig. 1 was examined on a daily basis during SFE2011 and used as guidance for the preparation of human forecasts for CI. Subjective assessments from this process indicated that each of the 3 sets of criteria used to identify CA points in model output was adequate for the purpose of detection. Specifically, they all seemed to highlight the same major features, with few or no systematic failures in detection. As occurred in the specific case highlighted in Fig. 1, the SMTR35 definition consistently identified the largest number of CA points at a given time, followed in order by the SWQQ and SLTG definitions (Fig. 5a). But again (based on subjective assessment of many cases), these differences in numbers of points did not appear reflect differences in numbers of identified convective storms. Rather, the differences in active-point totals appeared to be related to corresponding differences in porosity and size of individual features that were mutually identified using all CA definitions.

The relative numbers of simulated CI points associated with the different CA definitions were roughly proportionate to those for CA. In fact, the mean ratios of CI to CA were similar (Table 1, SLTG, SWQQ, and SMTR35 columns), but on some days the number of SMTR35 points appeared to be disproportionately high and on others unexpectedly low (Fig. 5b). Occasionally the number of CI points derived using the SWQQ definition for CA exceeded those derived from the SMTR35 definition. The reason for this behavior is not completely understood and will require more investigation, although some factors are worth noting. For example, the 5 minute sample interval used for SMTR35 (as opposed to 24 s for SLTG and SWQQ within the NSSL-WRF) can lead to relatively high ratios of CI to CA points with the default CI algorithm (CI_1).

The reason for this is that detection of nearby ongoing convective activity is less effective with the 5 minute sampling interval, causing more points to be labeled as “new” convection (CI). On the other hand, the SMTR35 CA field tends to be more contiguous (have lower porosity) than the

SWQQ- and SLTG-derived CA fields. With fewer gaps in its depiction of individual convective features, the SMTR35 criteria can allow for more effective discrimination between new and ongoing convection and a lower number of false CI detections. These factors can offset each other and one or the other may dominate depending on other (poorly understood) influences.

In summary, any one of the three definitions appears to be adequate for detecting convectively active grid points in the 4 km NSSL-WRF model. However, the default algorithm for identifying CI points as a subset of CA points (CI_1) shows some sensitivity to the sampling interval, the detailed character (small scale structure) of the CA field, and perhaps other factors.

1) Lightning and Radar-Reflectivity Observations

The SLTG and SMTR35 data were validated using NLDN and OMTR35 data, respectively. These datasets are described in the Data and Methodology section. An appropriate observational analog for the SWQQ data is under development, but not yet available. The observational data were mapped to the time steps and spatial grid of the model and, to make the comparison as robust as possible, the CI_1 algorithm was applied to observations just as it was for model data. Subjective comparisons were conducted on a daily basis during SFE2011. Processed observational data valid 00 UTC 25 May 2011 are presented here for direct comparison with Fig. 1.

In a mean sense, the CA field associated with the SLTG definition compares favorably with the observed NLDN coverage, with a coverage bias of about 0.7 (Table 2), but even a cursory comparison of the CA/CI fields derived from 25 May NLDN and SLTG data reveals significant differences in the character of the fields (cf. Figs 1a,b and Figs. 6a,b). The CA field associated with NLDN data has a much more of a "scattershot" appearance than the SLTG-CA field and individual features are relatively porous (cf. Figs. 1a and 6a). This discontinuous representation appears to have a negative impact on the CI_1 algorithm, causing too many of the CA points to be identified as independent CI points (cf. Figs 1a,b and Figs. 6a,b and the first two columns in

Table 1).

The OMTR35 fields seem to be better suited for verifying the corresponding model data (i.e., SMTR35 fields). For example, the OMTR35-derived CA field has a character similar to the SMTR35-derived CA field in this one case (cf. Figs. 1e and 6c) and the mean coverage of the SMTR35 and OMTR35 fields is quite similar (Table 2). Visually, the relationship between the SMTR35 and OMTR35-derived CI and CA fields seem consistent for the case shown here (cf. Figs. 1f and 6d and their corresponding CA fields), but with the CI_1 algorithm a higher fraction of OMTR35 CA points are labeled CI points, compared to the CI points derived from the SMTR35 CA fields (cf. last 2 columns in Table 1). This difference in the CI to CA ratio is likely due to the relatively contiguous character of the simulated reflectivity field compared to observed reflectivity.

In general, this analysis suggests that the reflectivity-based datasets (SMTR35 and OMTR35) have characteristics that will allow for meaningful preliminary assessments of skill in the prediction of CI. These fields have been used for additional analyses described below.

b) Detecting systematic errors in CI timing using the CAPS ensemble

Hourly probabilistic guidance for CI, derived using the CI_1 algorithm, was available from numerous modeling systems during the SFE2011. Corresponding NLDN and WSR-88D observations were available for validation. But it became obvious early in the experiment that, after the first convective cells developed, both the model and observational output from the CI_1 algorithm showed a proliferation of individual CI points on many days. In other words, there appeared to be a gross over-prediction of CI points, because the CI_1 algorithm was not effective at discriminating between truly independent CI and CI associated with ongoing mesoscale convective activity. This problem severely limited the utility of quantitative, spatially-probabilistic guidance for CI as it was presented during the experiment (and shown, e.g., in Figs 1 and 5). However, because the criteria used to detect convective activity (including the first point) were quite effective, the dataset still provides an excellent resource to evaluate systematic errors in the model guidance for CI timing and CA coverage.

1) CI timing for individual events

Systematic errors in timing can be evaluated by comparing simulated CI events to corresponding observed CI events. However, automating this comparison, which is necessary to get enhanced statistical significance, is difficult. Automation is challenging because, even within a limited region, convective activity often involves multiple modes of initiation and organization, and frequently it includes multiple waves of activity. Convective features invariably "look" different and are characterized by spatial and temporal offsets in observations as compared to simulations, so it can be difficult to ensure that features of interest are isolated within appropriate temporal and spatial windows in corresponding datasets. In short, automated detection and comparison of "matching" features in observations and simulations involves considerable uncertainty.

The strategy that we followed for this preliminary work was to step through observational data hour-by-hour to identify manually a subset of all the CI events that were the focus of daily experimental forecasts for CI during the experiment. An event was included in this subset if it occurred in a regional spatial domain (e.g., an area encompassed by a "slight", "moderate", or "high" risk for CI in an experimental forecast) in which no CI points were detected in the 5-h period leading up to the observed event. This manual selection process yielded a total of 14 CI events on different days and at the locations marked in Fig. 7.

For each of these 14 events, and for all 11 "PBL members" of the CAPS ensemble (members with perturbations in PBL parameterizations, but identical initial conditions), the differences in the timing of the first CI event were diagnosed. A frequency histogram of these differences (Fig. 8) shows a somewhat uneven distribution, perhaps due to the small sample size. However, some

of the noise in this distribution is probably due to spurious convection in the model forecasts, i.e., diagnosed CI points that are not necessarily associated with the same meteorological feature that was associated with the observed CI event.

One particular event is selected to demonstrate some of the challenges involved in this approach (Fig. 9). In realtime forecasting for this event, CI was expected to occur in the region from northern Arkansas into south-central Missouri and, for the purposes of demonstration, the verification domain is placed over this region. The first CI in the verification domain clearly occurs between 15 and 16 UTC. The timing of CI in the model forecasts is not so obvious. The simulated reflectivity fields from both forecasts suggested the presence of weak convective showers in the verification domain at 12 UTC (top row in Fig. 9). The automated CI detection algorithm flagged this activity as the first CI event in the YSU run (middle panel), but not in the MYJ forecast (right panel). In fact, the algorithm indicated that the first CI event in the MYJ run occurred between 15 and 16 UTC, as with observations. In effect, the automated procedure assigned a significant timing error to the YSU run and minimal error to the MYJ forecast of CI, although the forecasts were quite similar and by 18 UTC (last row in Fig. 9) they have about the same degree of resemblance to reality. This one example illustrates that, like all other attempts to verify detailed convective processes in model output, automated verification of individual CI episodes is subject to considerable uncertainty.

In general, these results show that there is no clear evidence of a systematic bias in the timing of CI for the CAPS PBL sub-ensemble. At the same time, they suggest that considerable uncertainty is associated with individual model forecasts and with methods to verify forecasts. Clearly, further work is needed to unravel this complex relationships involved in both simulated and observed CI.

2) CI timing in a bulk sense

Another way to assess model skill for CI timing is to compare the model climatology for CI to that of observations. This was done by plotting the frequency of simulated and observed CI as a function of time-of-day. For this assessment, CI points/events were detected using the 3D object based algorithm, CI_3, with the SMTR35 definition of CA and climatology was estimated over various spatial sub-domains. Preliminary results are shown for observations, selected CAPS-ensemble PBL members, and selected sub-domains in Fig. 10.

There is considerable sensitivity to the PBL schemes used in individual members but, in general, this ensemble subset seems to capture the timing and amplitude of peak CI frequency corresponding to the diurnal heating maximum. The plots suggest that these ensemble members under-predict CI frequency overnight (24-36 h forecast time) and during the morning hours (prior to about 18 UTC). As with the event-based assessment, there is no unambiguous evidence of a systematic bias in the timing of CI events, but insufficient information to inspire much confidence in specific model guidance for CI timing.

c) New diagnostic tools developed for SFE2011

Boundary-layer based CI is strongly modulated by a multitude of PBL processes and constrained by thermodynamic profiles within and near the top of the PBL. As described in the Methodology section, numerous diagnostic tools were developed for SFE2011 to help visualize relevant processes in model output. As described in the Methodology section one set of diagnostic fields was based on model-predicted mass and momentum fields at model level 12, which was about 1.1 km AGL – typically within or near the top of the PBL during the afternoon heating cycle. Examination of these fields often allowed one to infer important information about PBL processes related to CI. For example, judicious presentation of vertical velocity and moisture fields can indicate the presence of horizontal convective rolls within the boundary layer, transverse rolls near the top of the boundary layer, and the correlation between the upward-motion branch of these rolls and the moisture field (Fig. 11). The association between these features and specific CI events was often discernible.

Examination of model-predicted sounding structures was also very revealing, particularly the sensitivity of these structures to PBL parameterizations. For example, forecast-sounding structures from the individual PBL members often clustered in association with the specific closure assumptions used in each of the PBL parameterizations (Fig. 12). CI is undoubtedly sensitive to these details in sounding structure, but much more work is needed to gain complete insight into how sounding structure and dynamic PBL processes modulate the CI process in CAMs.

d) Human forecasts for CI

Human forecasts were prepared as a group activity, typically involving 4-8 people. After carefully reviewing and discussing all available observations and model guidance, the first decision for the group involved selection of a regional domain within which "clean slate" CI (i.e., CI preceded by at least several hours of no convective activity) was deemed possible. On most days, diurnal forcing was a significant factor, so the focus was typically on afternoon-evening CI, although overnight events were occasionally selected. Next, the time frame for the forecast was narrowed down to a 3h period. The last task for the group was to draw contours highlighting areas within the forecast domain having "slight", "moderate", and "high" probability for CI during the 3-h period. On many days, categorical forecasts were no higher than "slight", and on at least one day the consensus decision was to leave out even a "slight" area (consensus of less than 10% probability of CI with 40 km of any point during the 3 h period). As a final step, each member of the forecast team was asked to select the most likely time and location for the first storm, along with an assessment of confidence in the timing and overall likelihood of CI during the 3-h period. This information was stored for later analysis and rough outlines were drawn on the forecast product to indicate the areas where individual "first storm" points were clustered.

A sample forecast is shown in Fig. 13. In this case the first storms occurred within the region of highest categorical risk and close to a cluster of "first storm" prediction points, near the southern shores of Lake Ontario. The categorical forecasts were broadly consistent with probabilistic CI guidance from the CAPS ensemble.

Preparing human forecasts for the probability of CI within a fixed spatio-temporal domain was a very useful exercise. It focused the attention of participants on the working definitions of CA and CI, numerical guidance for CI and related fields like convective coverage, post-processing strategies for producing probabilistic information, and physical processes associated with CI in the models and observations. Requiring individual forecasts for the location, timing, and likelihood of the first storm required additional consideration of how precise one can be in this type of forecast with any sort of confidence and skill. Furthermore, preparation of the human forecasts prompted numerous discussions regarding the utility of CI/CA model guidance for various user groups, such as SPC and HPC forecasters and the aviation industry.

The process was very thought-provoking and appropriate for the SFE. However, the human forecasts are not likely to be useful for assessing how much skill trained NWS forecasters would have in predicting CI.

4. Summary

This CI component of SFE2011 was viewed as a pilot project designed to investigate the utility and feasibility of routine forecasting for CI. This project yielded several important results:

- Several different set of criteria were found to be adequate for detecting convective activity in gridded model fields and observations, but the reflectivity field (simulated and observed) has key advantages that appear to make it the best surrogate (among those tested) for automation of both detection of convection and discrimination between new and ongoing activity (i.e., reliable identification of CI events).
- The default algorithm for identifying CI points as a subset of total convective points had a propensity to significantly over-predict the frequency of distinct CI events. New algorithms that utilize object-based tracking methodologies and warping and filtering strategies show promise for better performance.
- Preliminary results indicate no systematic bias, but considerable variance in the timing of CI events in 4-km CAMs.
- Simple diagnostic tools reveal that simulated CI events rooted in the convective boundary layer are often associated with localized perturbations in convergence, vertical velocity, and moisture associated with mesoscale boundaries and boundary-layer rolls. These simulated phenomena appear to be directly analogous to observed features such as horizontal convective rolls, although they are sometimes realized on inappropriately large scales in current CAMs due to coarse spatial resolution of the models.
- Routine probabilistic forecasts for CI are certainly possible and probably desirable. However, it is still not clear exactly what the parameters of a successful and useful CI forecast should be. Further discussions with potential user communities are needed to clarify this issue.

5. Plans for SFE2012 [Comments encouraged]

CI studies will be a significant part of the SFE again in 2012. Based on promising results from SFE2011, it is anticipated that convective activity (CA) will be identified in both model and observational data primarily using the general concepts of the MTR35 set of criteria, i.e., reflectivity exceeding a specified threshold at some level well above the melting layer. Although this criteria-set may not be the best indicator of detailed CI processes, it is preferred because it appears to represent the formation, expansion, merger, movement, splitting, etc. of convective storms and larger systems quite well. Furthermore, both researchers and forecasters have considerable experience examining reflectivity fields from both observations and simulations, which is desirable. CI points will be identified as a subset of CA points using algorithms based on tracking, warping, and filtering of convective objects, rather than the simple default algorithm used in 2011.

Experimental probabilistic forecasts for CI-related processes will be generated in SFE2012 as well. These forecasts are likely to have multiple elements. For example, forecasts should attempt to answer some or all of the following questions, given a mesoscale time/space window:

- What is the overall likelihood that CI will occur (any time or location in the window)?
- If storms occur, what is the *relative* likelihood of occurrence of first storms at different locations? (If this is presented as a spatial probability of CI, the area under the 2-D probability field should be directly related to the overall likelihood of CI in the region, but this scaling could be done after the fact)
- What is the *relative* temporal probability distribution of first storms? (Again, the area under this distribution should be directly related to the overall likelihood of CI, but this scaling could be done after the fact. How sharp is the distribution? Is it bimodal? Trimodal?)
- If CI occurs, how is the bulk coverage of storms within the region expected to change with time (i.e., rapid expansion of individual cells, mergers, new growth/initiation, etc?)
- If CI occurs, how is the convective mode expected to change with time?
- If CI occurs, how is the "local coverage' (porosity?) expected to change with time, perhaps as a function of location within the domain?

References:

Clark, A. J., S. J. Weiss, J. S. Kain, and co-authors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. In press, *Bull. Amer. Meteor. Soc.*

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

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.

Marsh, P. T., J. S. Kain, A. J. Clark, V. Lakshmanan, N. M. Hitchens, and J. Hardy, 2012: A Method for Calibrating Deterministic Forecasts of Rare Events. *Wea. Forecasting*, in press.

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

Roberts, R. D., S. Rutledge, 2003: Nowcasting Storm Initiation and Growth Using GOES-8 and WSR-88D Data. *Wea. Forecasting*, **18**, 562–584.

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**, in press.

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.

Average CI/CA (%)				
SLTG	**NLDN**	**SWQQ**	**SMTR35**	**OMTR35**
11.1	*77.3*	*13.5*	*13.8*	*24.5*

Table 1. Percentage of CA points that were identified as CI points over all days during the Spring Forecasting Experiment

	SMTR35/OMTR35	SLTG/NLDN
CA	1.03	0.717
CI	0.580	0.102

Table 2. Mean coverage bias of simulated fields compared to observational counterpart.

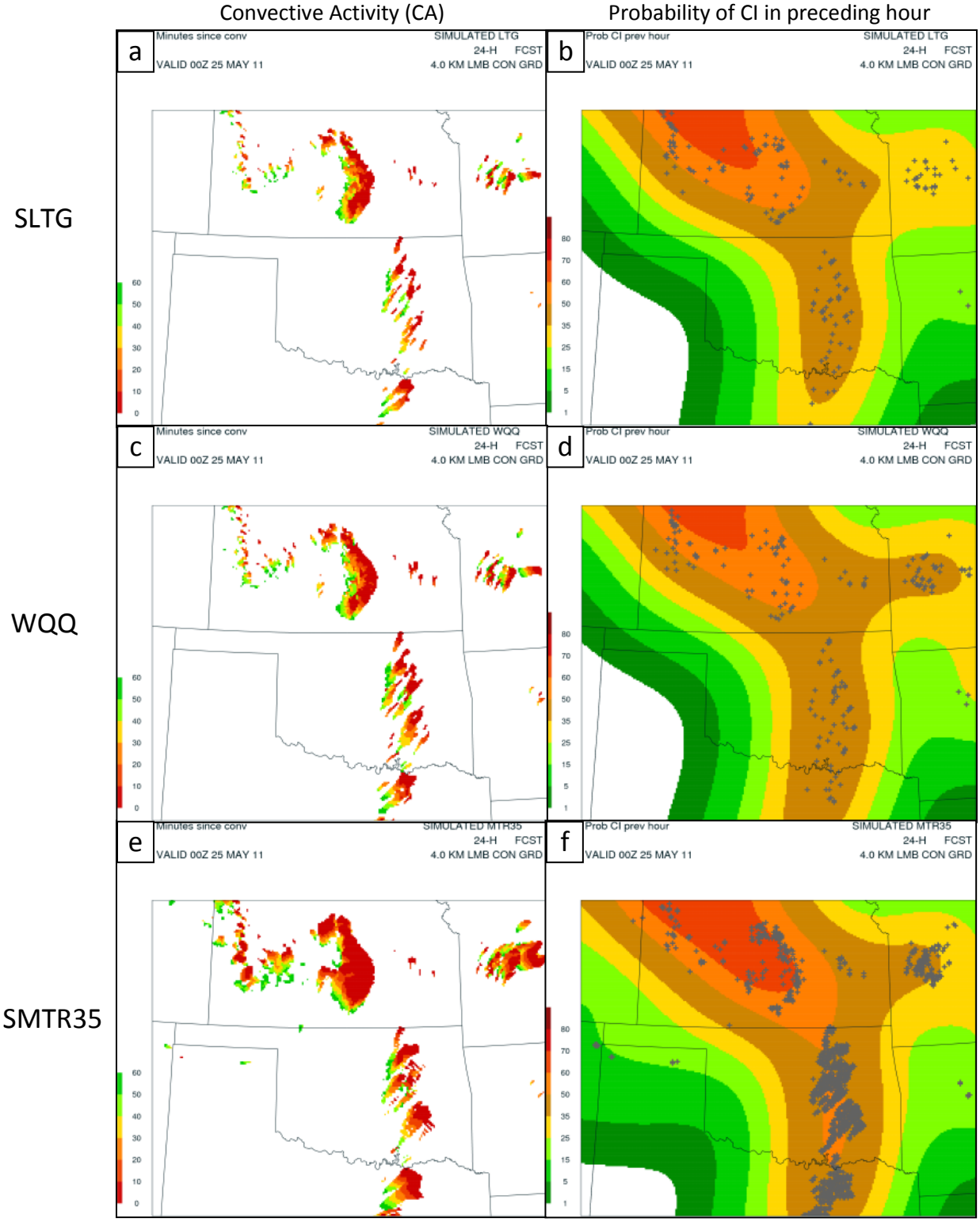


Fig. 1 Examples of raw CA field (left column) and derived 1h probabilities of CI (right column) for each of the 3 sets of CA criteria, derived from a 24 h NSSL-WRF forecast valid 00 UTC 25 May 2011. The small '+' signs on the right indicate the specific locations of CI points within the last hour, identified using the CI_1 algorithm.

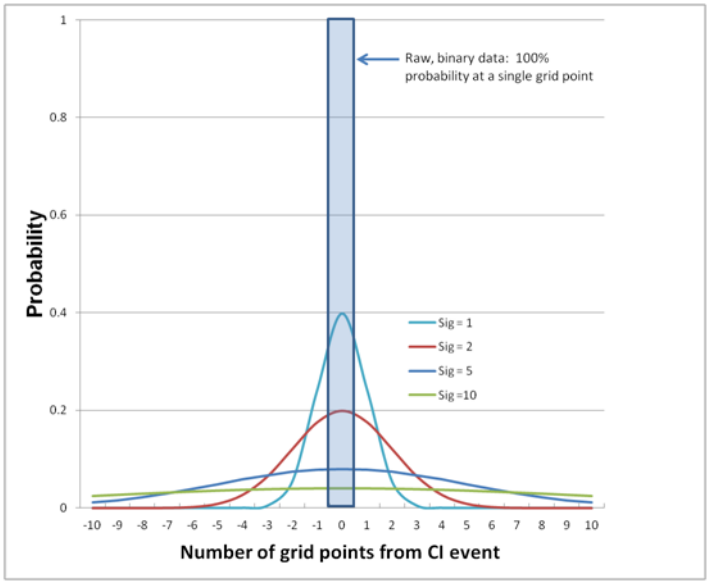


Fig. 2. Conceptual 1-D diagram of the Gaussian smoother used to generate probability fields from a deterministic forecast, depicting derived probability values as a function of distance for different values of standard deviation (sigma)

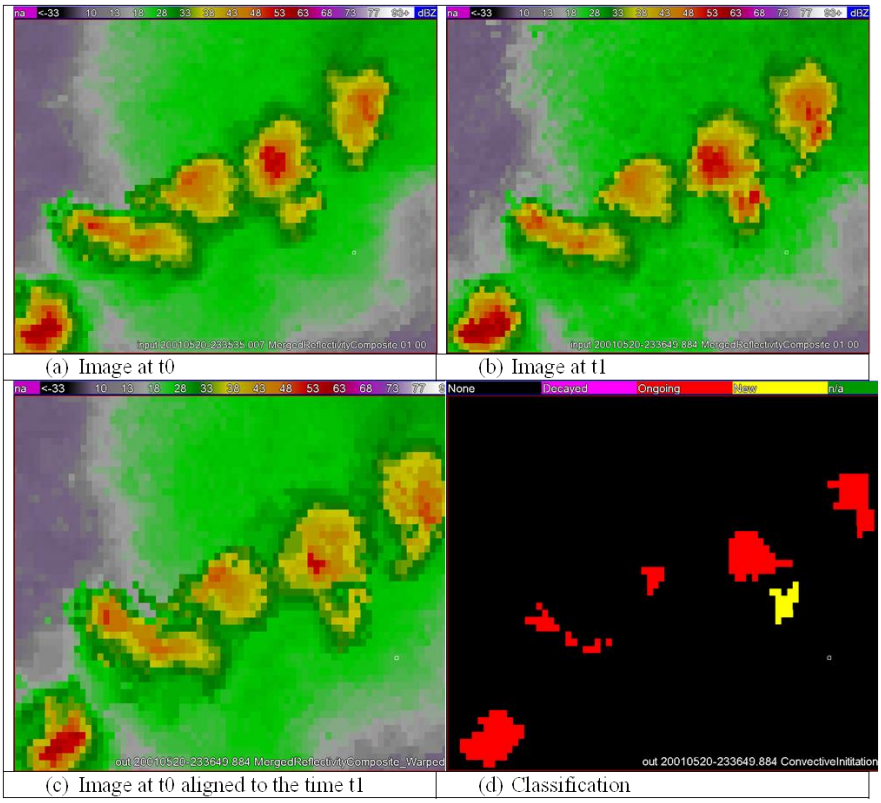
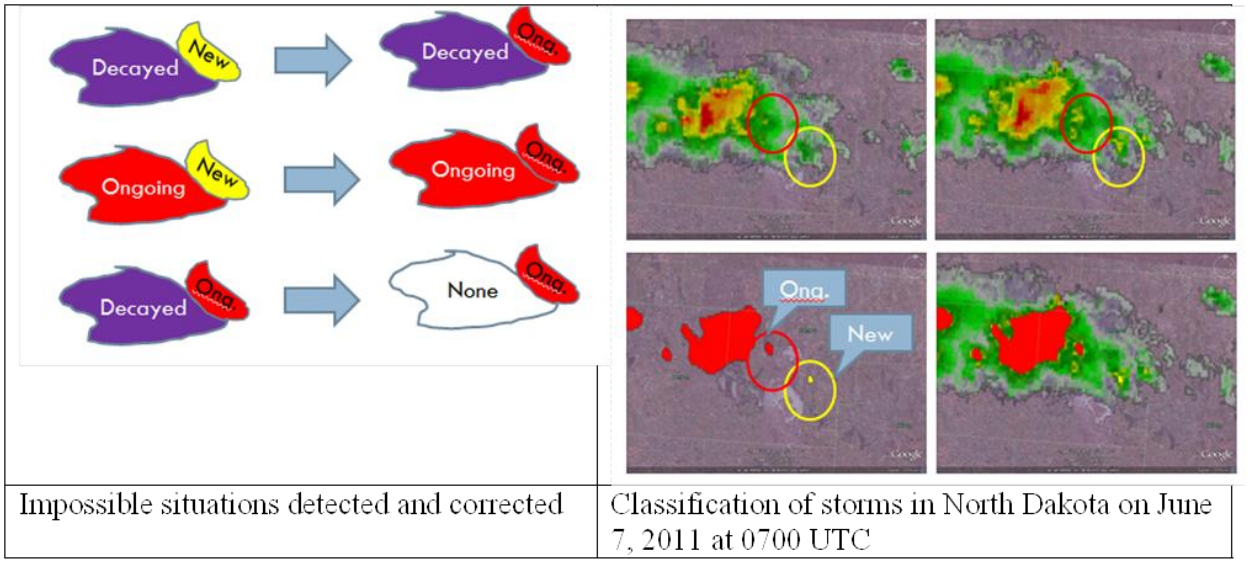


Fig. 3. Example showing how 2 images (a, b) are compared (c) and the resulting identification of CI points.



Impossible situations detected and corrected	Classification of storms in North Dakota on June 7, 2011 at 0700 UTC

Fig. 4. Example showing corrections for impossible situations

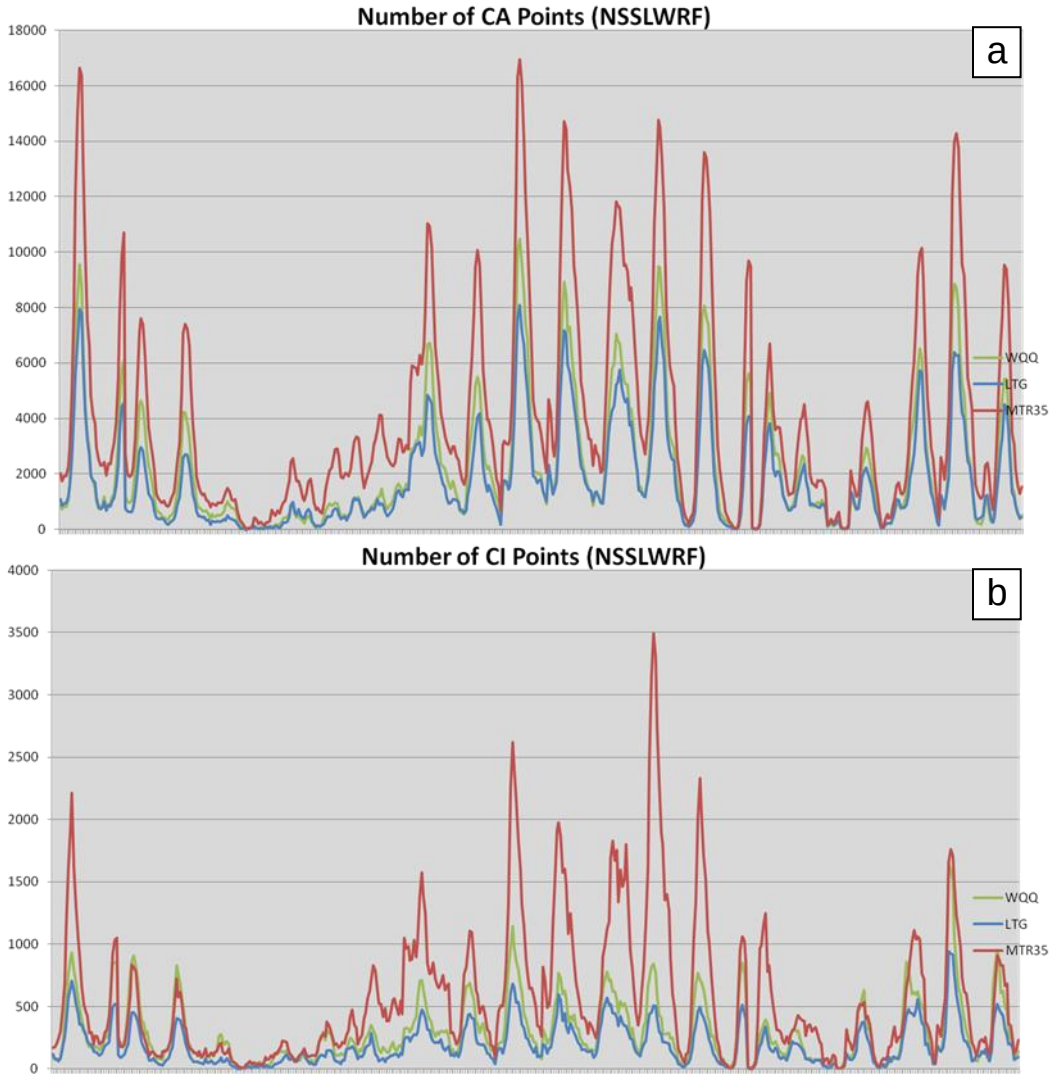


Fig. 5. Time series of numbers of CA (top) and CI (bottom) grid points for each hour from 12 UTC 11 May through 12 UTC 11June 2011, for the different CA definitions. Data are from a subdomain of the daily 12-36 h forecasts from the NSSL-WRF, covering the CONUS east of the Rockies.

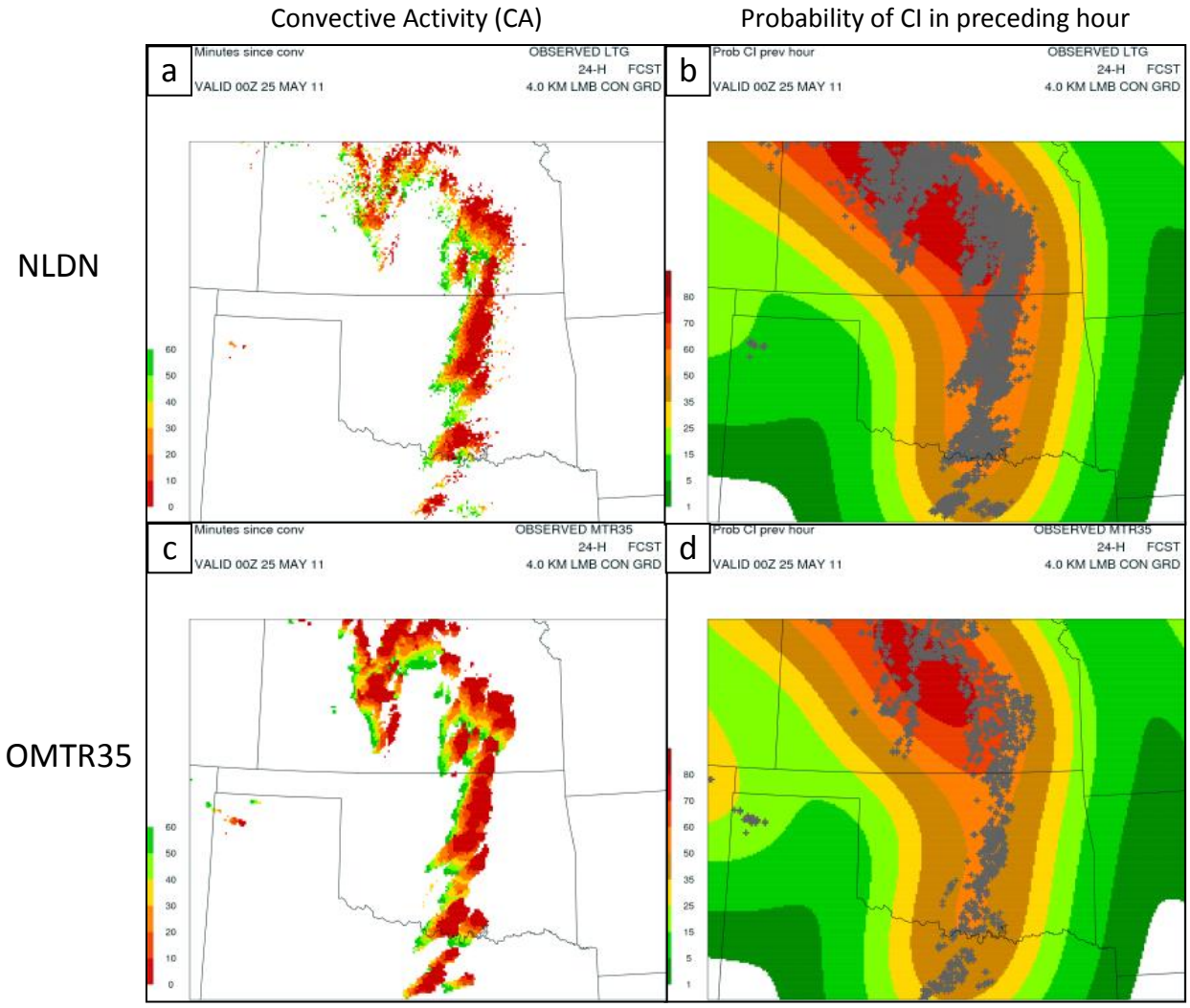


Fig. 6 Examples of raw CA field (left column) and derived 1h probabilities of CI (right column) for each of the 3 sets of CA criteria derived from 1-h observations of NLDN and WSR-88D data ending 00 UTC 25 May 2011. The small '+' signs on the right indicate the specific locations of CI points within the last hour.

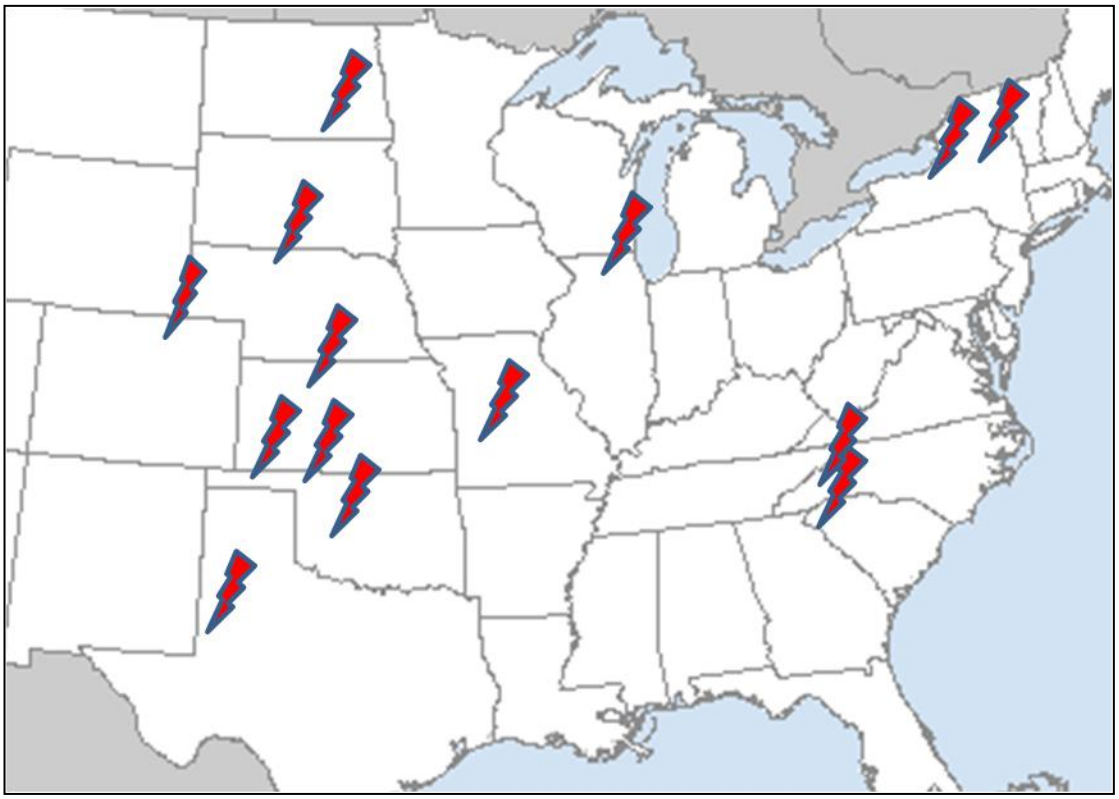


Fig. 7. Locations of 14 distinct CI events used for preliminary assessment of CI timing.

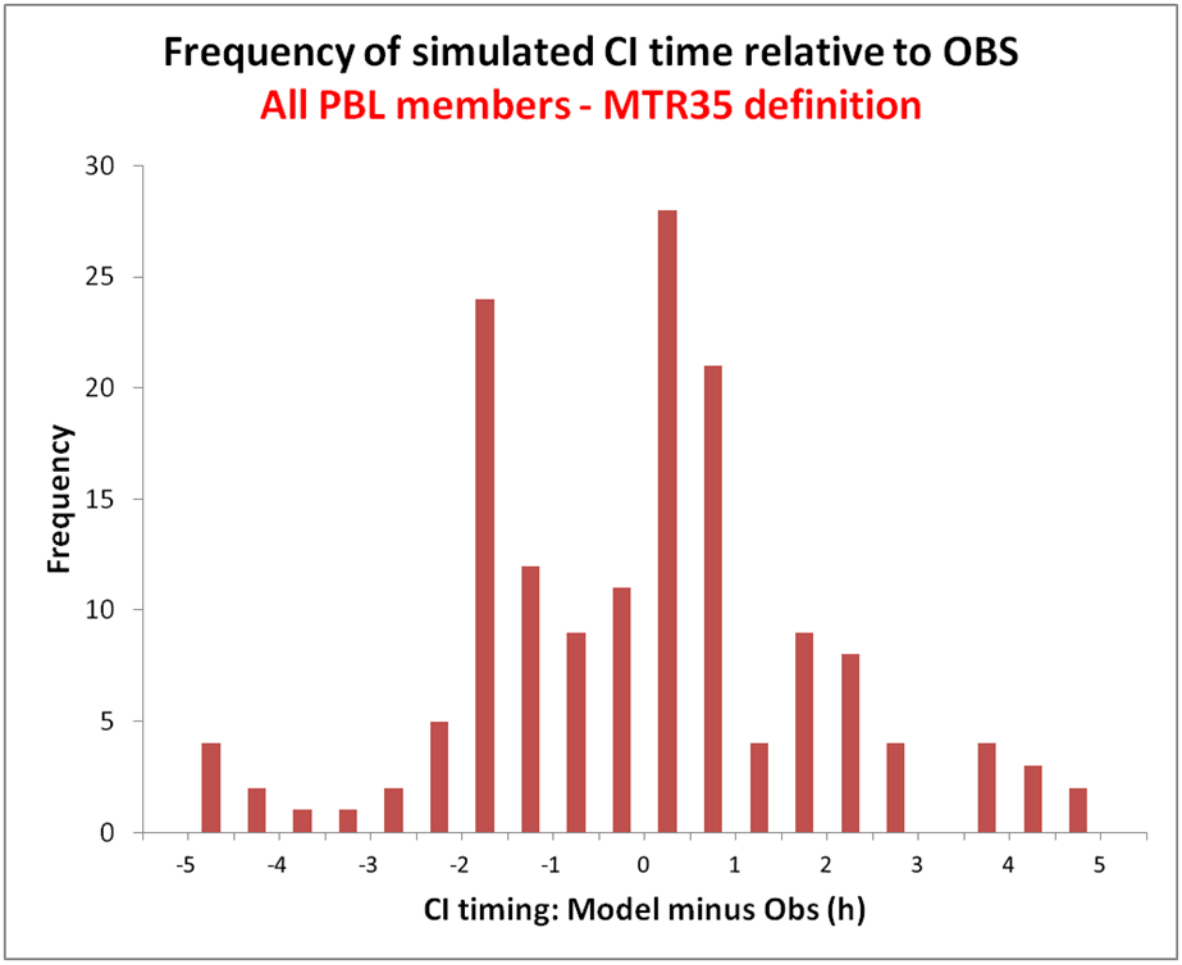


Fig. 8. Frequency histogram of differences between predicted and observed convective initiation times within small regional domains centered at the locations marked in Fig. 7. The events occurred on different days during the period 17 May – 10 June 2011. The predictions came from the "PBL members" of the CAPS ensemble.

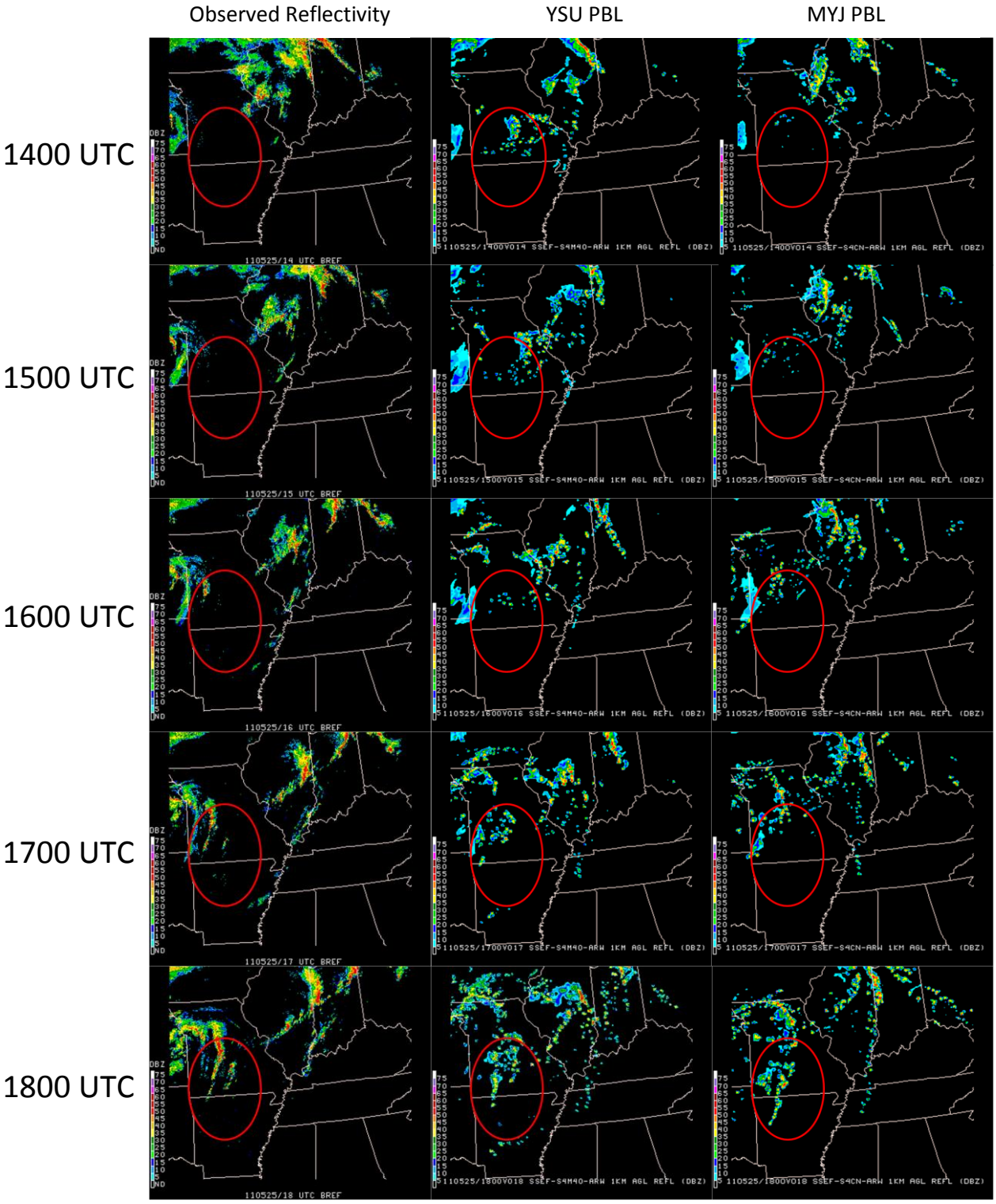


Fig. 9. Hourly sequence of images showing observed reflectivity (left column) and simulated reflectivity from two different physics members from the CAPS ensemble (middle and right columns). Red ellipse indicates the CI verification domain.

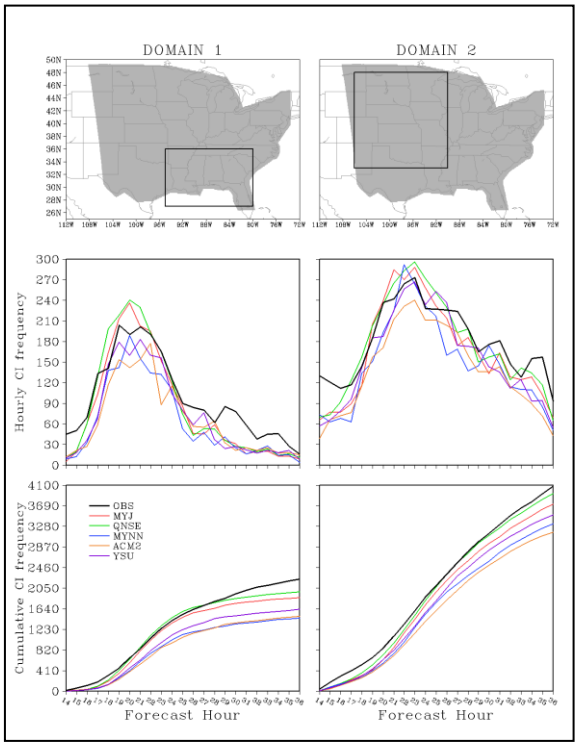


Fig. 10. Hourly (middle row) and cumulative (bottom row) CI frequency as a function of forecast hour for observations and five selected members CAPS PBL ensemble. Two different analysis domains were used, as indicated in the top row.

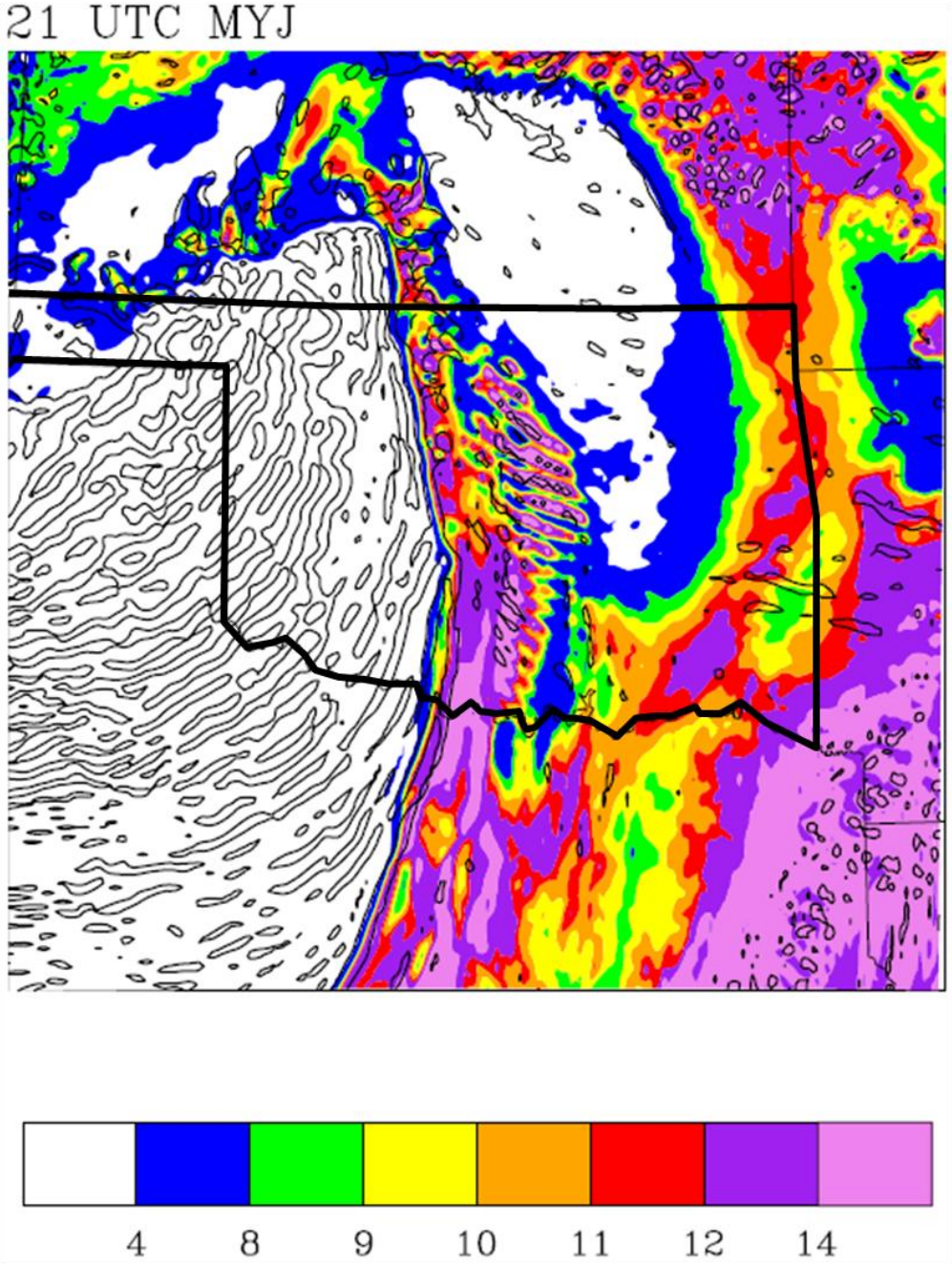


Fig. 11. Sample diagnostic plot showing vertical velocity (contour interval 0.25 ms^{-1}) and water vapor mixing ratio (color fill, g kg^{-1}) at model level 12 (approximately 1.1 km AGL). Note the horizontal-convective-roll-like features in the drier air west of the dry line (indicated by sharp east-west moisture gradient) and the transverse rolls, apparently in stable air above the PBL, in central Oklahoma.

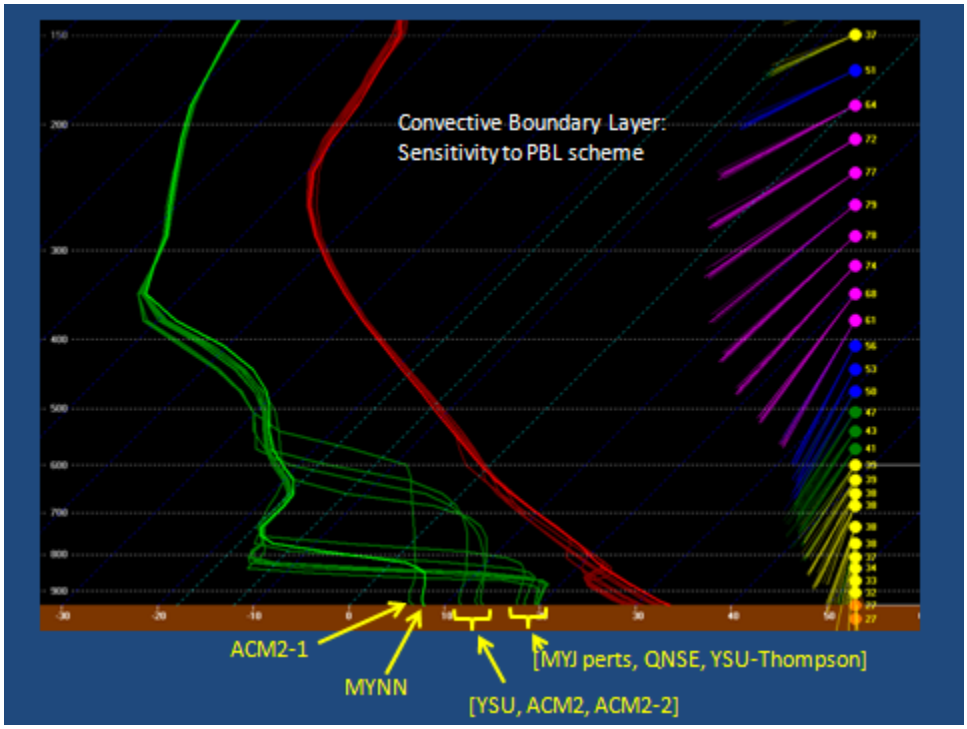


Fig. 12. Forecast soundings valid at a single time and location from each of the PBL members from the CAPS ensemble. Note the sensitivities to PBL parameterization and parameter values within a given parameterization.

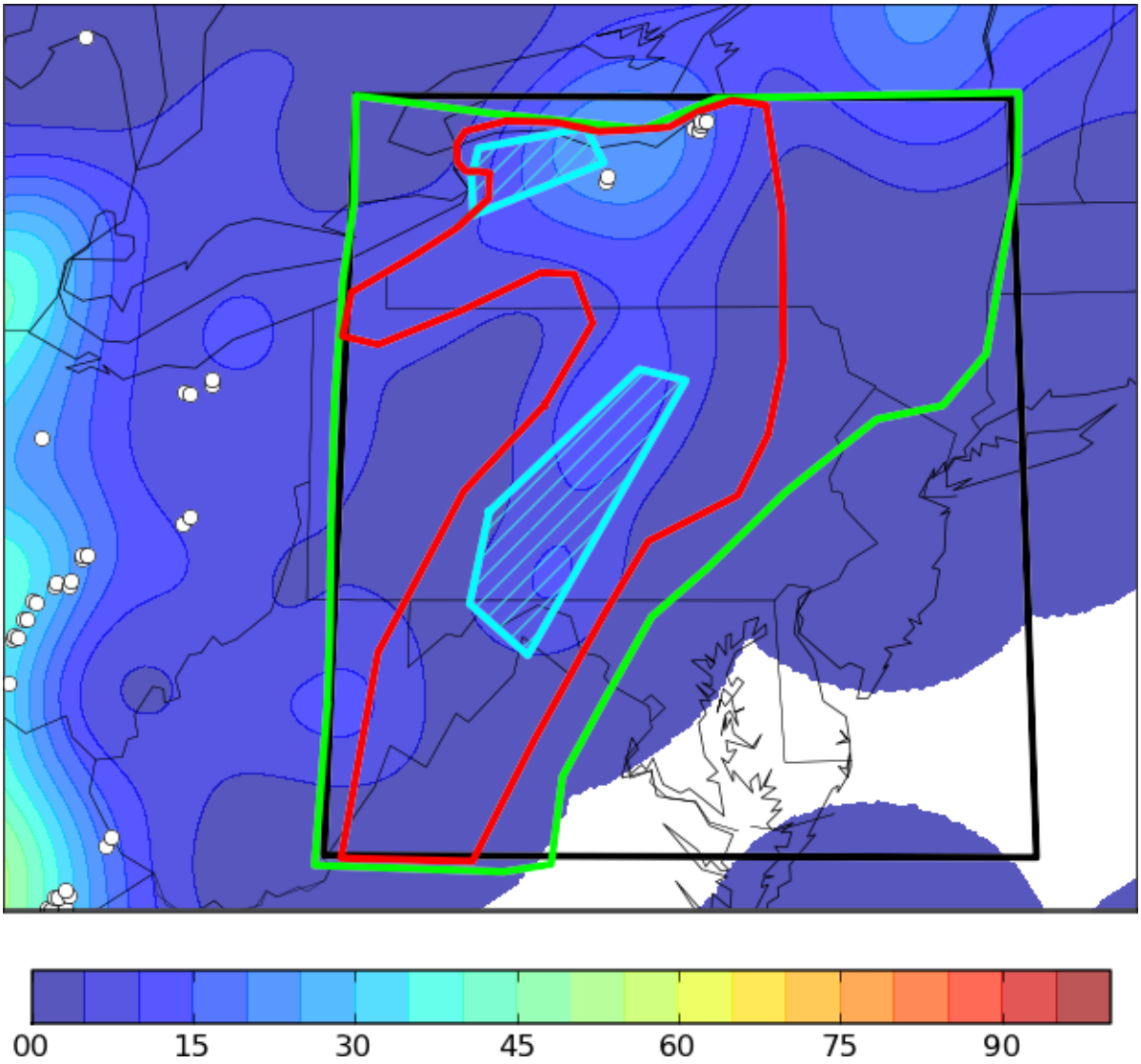


Fig. 13. A sample experimental human forecast for CI, valid for the 3h period ending 1900 UTC 26 May 2011. Thick black line outlines the forecast domain, green and red lines outline the slight and moderate risk areas, respectively, while the hatched blue areas roughly outline the regions where individual forecasters predicted the first CI event within the forecast domain. The color fill shows the CI guidance provided by the CAPS ensemble, expressed as the probability of CI within 40 km of any point during the preceding hour. The white dots indicate the locations of observed CI points between 1700 and 1800 UTC. Both observed and model-predicted CI events were detected using the CI_1 algorithm and the MTR35 definition for CA.