ASSESSING THE PREDICTABILITY OF CONVECTION INITIATION USING AN

OBJECT-BASED APPROACH

by

Brock Burghardt

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the degree of

Master of Science

in Mathematics

at

The University of Wisconsin-Milwaukee

May 2013

ABSTRACT
ASSESSING THE PREDICTABILITY OF CONVECTION INITIATION USING AN
OBJECT-BASED APPROACH

by

Brock Burghardt

The University of Wisconsin-Milwaukee, 2013

Under the Supervision of Professors Clark Evans and Paul Roebber

Improvements in numerical forecasts of deep, moist convection have been notable in

recent years and are in large part due to increased computational power allowing for the

explicit numerical representation of convection. Accurately forecasting the timing and

location of convection initiation (CI), however, remains a substantial forecast challenge.

This is attributed to the inherently limited intrinsic predictability of CI due to its

dependence on highly non-linear moist physics and fine-scale atmospheric processes that

are poorly represented in observations. Because CI is the starting point of deep, moist

convection that grows upscale, even small errors in initial convective development can

rapidly spread to larger scales, having potentially significant impacts on downstream

forecasts. This study investigates the practical predictability of CI using the Advanced

Research Weather Research and Forecasting (WRF-ARW) model with a horizontal grid

spacing of 429 meters. A unique object-based method is used to evaluate very high-

resolution model performance for twenty-five cases of CI across the west-central High

Plains of the United States from the 2010 convective season. CI objects are defined as

areas of higher observed or model simulated radar reflectivity that develop and remain

sustained for a sufficient period of time. Model simulations demonstrate an average probability of detection of 0.835, but due to significant overproduction of CI, an average false alarm ratio of 0.664 and bias ratio of 2.49. The average critical success index through all simulations is 0.315. Model CI objects that are matched with observed CI objects show, on average, an early bias of about 7 minutes and distance errors of around 62 kilometers. The operational utility and inherent biases of such high-resolution simulations are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Introduction

The development of deep, moist convection (DMC), known as convection initiation, hereafter CI, refers to the process in which convectively unstable air reaches and exceeds a critical thermodynamic level through a series of cumulus plumes to buoyantly accelerate and develop a mature deep, moist updraft. Conversely, failed CI is the process in which CI is not fully achieved (shallow or non-sustained deep updraft). Failures of forecast CI (false positive), or the less frequent unexpected CI (false negative), are two of the most perplexing occurrences in operational meteorology and their impacts on later forecasts can be significant. Once deep, moist convection has initiated in a favorable atmospheric environment, it may become severe, producing flash flooding, damaging straight-line winds, large hail, and tornadoes. The societal and economic impacts of severe convection are enormous, resulting in more than $4 billion in damage annually across the United States (Weiss et al. 2011). Still, forecasting CI and the intensity and evolution of convection remains difficult, therefore serving as motivation for further research on the topic.

Deep, moist convection can be initiated by a large array of different atmospheric circulations, oftentimes referred to as triggering mechanisms. This is one of three basic elements to develop convection, in addition to the presence of convective available potential energy (CAPE) and a sufficient moist layer at low- to mid-levels of the troposphere (Doswell 1987; Johns and Doswell 1992). The most common triggering mechanisms, as summarized by Jorgensen and Weckwerth (2003) and Weckwerth and Parsons (2006), include frontal boundaries, drylines, elevated convergence zones, gust fronts, sea breezes, orographic circulations, undular bores, and horizontal convective

roles. Further, inhomogeneities in land-surface characteristics can produce sufficient vertical motion to force CI, particularly under strong surface heating (Kang and Bryan 2011).

The complexity of CI occurrence, however, is far greater than just the presence of a triggering mechanism. Rather, a favorable interaction between phenomena on multiple atmospheric scales is required to establish an environment favorable for deep, moist convection and then force CI. In general, the three scales most influential on CI are the synoptic-, meso- and micro- scales. The synoptic- to meso-alpha scales, on the spatial order of 10-100 km, are responsible for the creation of an environment favorable for convection (e.g., Doswell 1987; Schumann and Roebber 2010). This is accomplished through the advection of higher potential energy air at low levels and ascent that acts to increase CAPE and reduce CIN. The meso-gamma to micro- scales, on the spatial order of 1 km and smaller, are responsible for the local deepening and moistening of the convective boundary layer (CBL) through ascent that can ultimately eliminate CIN (e.g., Wilson et al. 1992; Doswell 1987; Markowski et al. 2006; Weckwerth et al. 2008, among others). This is accomplished through local moisture convergence, connected to upward vertical motion associated with triggering mechanisms, which results in localized preferred areas of CI where parcels are most likely to reach their level of free convection (LFC).

The sensitivity and predictability of CI has been exclusively studied in three past field campaigns. The first, the Severe Environmental Storms and Mesoscale Experiment (SESAME; e.g., Hill et al. 1979), investigated the influence of meso- to synoptic- scale features on CI. The second, the Convection Initiation and Downburst Experiment

(CINDE; e.g., Wilson et al. 1988), observed mesoscale features and their influence on storm environment and initiation. Most recently, the International H_2O Project (IHOP; e.g., Weckwerth et al. 2004), investigated meso- to micro- scale features and their influences on CI. Fine-scale data obtained during IHOP has permitted investigations of several observed CI events and failed CI events that have greatly helped to illuminate the crucial finer-scale processes that largely influence CI. For instance, confirming what Ziegler and Rasmussen (1998) found from a CI case along a dryline in Oklahoma, Markowski et al. (2006) found that CI is highly dependent upon initial low-level potentially unstable parcels remaining within a zone of mesoscale ascent with higher relative humidity until they have reached their LFC. This suggests that CI is most likely when deep CBL plumes become superposed with regions of mesoscale ascent associated with triggering mechanisms. This, however, still does not fully explain the occurrence of CI as cloud-scale processes such as dry air entrainment (Houston and Niyogi 2007) and momentum flux remain crucial.

Furthermore, the occurrence of CI is remarkably sensitive to the environment in which it occurs (e.g., Lee et al. 1991; Crook 1996; Weckwerth 2000; Houston and Niyogi 2007). This is especially true regarding the low-level thermodynamic state as demonstrated by Crook (1996), who showed that a temperature change of 1°C in a mixed surface layer determined whether CI did or did not occur. Furthermore, if CI did occur, a 1 $g\,kg^{-1}$ change in low-level moisture content resulted in considerable variability in storm intensity. This is one of the leading factors that limit the practical predictability of CI since such variability falls into the range of observational uncertainty and local variability (e.g., Zhang et al. 2002, 2003; Weckwerth and Parsons 2006).

With the occurrence of CI being so sensitive to the environmental state, including small scale details that most observational networks cannot resolve, it can be difficult to numerically model our atmosphere with sufficient accuracy to predict CI. This is related to the fact that forecast skill is strongly dependent upon the accuracy of the initial and lateral boundary conditions used in a numerical weather model (e.g. Lorenz 1969; Liu and Xue 2008; Hsiao et al. 2009). Although our current observational network can resolve synoptic- and, to a degree, meso- scale features, we are unable to properly resolve (if at all) smaller-scale details that are known to be crucial to CI (Weckwerth et al. 1999; Weckwerth and Parsons 2006). Several recent studies have demonstrated the importance of fine-scale observations on the skill of convection-allowing model (CAM) forecasts. Liu and Xue (2008) demonstrated this sensitivity through the use of a 3 km grid spacing CAM to simulate an observed CI case that occurred during the IHOP field program. They showed that the best short-range (0-7 h) forecasts were achieved when the most amount of fine-scale observations were included in model initialization, and also demonstrated strong sensitivity to the placement of the lateral boundaries. Further supporting this finding, Sun et al. (2012) analyzed multiple 0-12 h CAM forecasts of convection during the IHOP field program with varying amounts of fine-scale observations incorporated during model initialization. They found that convective precipitation forecasts were most skillful when the most fine-scale observations were incorporated into the model. This is attributed to the fine spatial resolution observations aiding in the faster spin up of smaller-scale processes that strongly control CI and DMC evolution. Both studies found that the inclusion of the most fine-scale observations had a negligible impact on boosting

forecast performance beyond 7-9 h from model start time, when larger meso-alpha to synoptic- scale features become more influential upon DMC evolution.

Notable improvements in numerical forecasts of DMC have been seen in recent years due in large part to increased computational power and advances in model initialization and ensemble techniques. Such improvements have especially helped to improve our practical predictability on the synoptic scale (e.g., Roebber and Bosart 1998). Predictability on the meso-gamma to convective- scales intrinsically remains lower, however, because of its greater sensitivity to non-linear processes (Zhang et al. 2003; Melhauser and Zhang 2012). For example, although the evolution and upscale growth of deep, moist convection in a convective outbreak is often strongly dependent on synoptic-scale forcing (e.g., Roebber et al. 2002; Fowle and Roebber 2003; Schumann and Roebber 2009; Houston and Wilhelmson 2011), the exact location and timing of CI is strongly tied to meso-alpha (referring to scales of order 10 km) and finer-scale physical processes (e.g. Weckwerth and Parsons 2006; Weisman et al. 2008) that we do not fully understand and that are highly non-linear in nature (e.g., moist processes).

It is well documented that CAM forecasts typically demonstrate better short-range (0-24 h) prediction of DMC compared to their convection-parameterizing counterparts regarding rainfall amounts, storm mode, and mesoscale circulations (e.g., Weisman et al. 1997; Mass et al. 2002; Fowle and Roebber 2003; Done et al. 2004; Roebber et al. 2004; Wilson and Roberts 2006; Kain et al. 2008; Weisman et al. 2008; Clark et al. 2010). However, beyond 24 h, and sometimes sooner, error growth and spread tend to increase faster in CAMs than in coarser convection-parameterizing models (e.g. Zhang et al. 2003; Clark et al. 2010). This issue was well described by the works of Zhang et al. (2002,

2003) that demonstrated how small-scale errors, in the form of misplaced and misrepresented DMC, quickly spread and amplified through the convective- to meso-scales before growing up scale to the synoptic-scale, playing a strong role in the poor forecast of an east coast snowstorm in 2000. This places a stricter intrinsic time limit of predictability in CAMs than in coarser convection-parameterizing models. One way of visualizing this error growth is to envision a forecast scenario where one or a series of convective cells are incorrectly placed in space. Errors in representing the convectively-impacted environment increase through time as the volume of space impacted grows, but further, these errors are essentially doubled, given that the model incorrectly depicts the location of both convection activity and inactivity. The upscale growth of this error then results from differences in the geostrophic adjustment to diabatic forcing from DMC in the simulated and observed environments.

Although grid spacings in operational CAMs allow for the resolution of at least larger storm-scale features, they remain too coarse to truly resolve all scales of motion in DMC. Bryan et al. (2003) demonstrated, through theoretical scale analysis, that a spatial grid spacing of 0.1 mm is needed to fully resolve all turbulent components of a geophysical fluid. They also show that horizontal grid spacing on the order of 100 m is necessary to adequately resolve the structure and dynamics of squall lines. However, better resolving DMC may not be as simple as decreasing grid spacing to order 100 m. This was demonstrated by Adlerman and Droegemeier (2002), who found that the dynamics of an idealized modeled supercell change drastically as horizontal grid spacing is decreased from 2 km to 0.5 km, but, that results are very sensitive to only minor changes in physical parameter constants such that results become nearly interchangeable.

Although the implementation of horizontal grid spacings suggested to adequately resolve DMC remain far from operationally practical for regional model applications, current CAM grid spacings (1-4 km) still provide added value for short-range convective forecasts in conjunction with coarser resolution operational models. Deterministic CAM forecasts are most useful in providing insight into storm mode and structure (e.g. Done et al. 2004; Roebber et al. 2004; Weisman et al. 2008) that coarser operational models cannot. For instance, Fowle and Roebber (2003) found that warm season model forecasts in the Lake Michigan region using 6 km grid spacing scored quite well for the occurrence of CI and even more so for the forecasted storm mode. Recently, Duda and Gallus (2013) showed that 3 km grid spacing WRF simulations for convective events across the U.S. demonstrated good skill in matching observed CI objects prior to upscale growth of convection. They found that on average model simulations showed no discernible timing error in forecasted CI with a mean spatial displacement error of 105 km. In addition, higher resolution models can provide guidance on the types and intensity of severe weather phenomena (i.e. hail, damaging winds, tornadoes, etc.) with rather simple post processing techniques, as demonstrated by Sobash et al. (2011).

An additional tool that has proved invaluable to forecasting is the ensemble model system, a suite of different model configurations run in parallel with changes in at least one of the following: initial and/or lateral boundary conditions, different physical parameterizations, and different model cores. This is especially true for convective-scale ensembles that utilize an ensemble of CAM simulations to give a probabilistic perspective upon forecast occurrence and represent (to first order) the sensitivity and confidence in a forecast (Roebber et al. 2004; Kong et al. 2006; Schwartz et al. 2009;

Schwartz et al. 2010; Kain et al. 2013; Coniglio et al. 2013). Because of their greater computational demand, such forecasts tend to use horizontal grid spacings that can explicitly resolve convection but are greater than those used in some deterministic CAM applications. As already mentioned, higher resolution models with grid spacings of 4-12 km typically yield better forecast skill at short ranges than coarser model configurations; however, it is uncertain if there is added skill for even higher resolution models with horizontal grid spacings of less than 4 km (e.g., Kain et al. 2008; Schwartz et al. 2009; Coniglio et al. 2010; Clark et al. 2011; Clark et al. 2012a). Because CAM ensemble means tend to outperform deterministic CAM models (e.g. Coniglio et al. 2010; Clark et al. 2011), further improvement in convective model forecasts will require the optimization of adequate grid resolution and computational resources, as illustrated by Clark et al. (2011).

Advances in computing power have resulted in model grid spacings that have decreased at a rate faster than updates to many of the utilized physical parameterizations. Resultantly, physical parameterizations that have been tuned to best compliment models with horizontal grid spacings of 4-10 km or greater are not best fit for higher resolution CAMs. This has been particularly apparent with planetary boundary layer (PBL) physical parameterizations that emulate turbulent heat, moisture and momentum transports. The parameterization of turbulent energy in the PBL can be approximated via two methods. One is a "local" closure solution that allows for the growth of the PBL and turbulent energy transport through adjacent vertical model levels (Stull 1991; Hong et al. 2006). A second method is a "non-local" closure solution where the depth of the PBL is pre-determined to assess the layer that turbulent energy transport is applied (Hong et al.

2006). At horizontal grid spacings that are near or smaller than the depth of the PBL, larger overturning circulations often are beginning to be explicitly resolved by the model while still being parameterized by a PBL scheme. This interaction occurs in a "grey area" where model horizontal grid spacings of approximately 1-3 km crudely resolve large wavelength PBL structures yet still require the parameterization of eddy circulations on scales of tens to thousands of meters (e.g., Agee and Gluhovsky 1999; Stevens et al. 2005; Stensrud 2007). In such cases, the crudely-resolved large-scale PBL features begin to blend and possibly superpose with features handled by the PBL scheme, analogous to convection-parameterizing mesoscale models at horizontal grid spacings of 4-10 km that can begin to resolve large convective systems. Breaking such processes into components of model resolved and physically parameterized is difficult (Stensrud 2007), but will be necessary when updating or developing PBL parameterizations for higher resolution applications.

In addition to the issue of PBL parameterizations being improperly tuned for convection resolving grid spacings, they also contain noticeable and unique errors and biases. In turn, these errors and biases are likely to amplify through model integration time. Demonstrating this, Romine et al. (2013) found that in a cycled ensemble model system, biases and errors with different physical parameterization schemes that develop by a given model simulation's initialization time subsequently amplify while typically retaining initial bias characteristics. This issue was also recently investigated by Coniglio et al. (2013), wherein they found noticeable errors and biases in the low-level thermodynamic fields downstream of DMC episodes with both "local" and "non-local" PBL schemes. Specifically, "local" schemes tend to be too shallow in PBL depth and too

moist near the surface, which tends to overproduce CAPE. The "non-local" schemes tend to be too dry near the surface with slightly under biased CAPE, but have little bias in PBL depth. Because of CI's strong dependence on the low-level thermodynamic state, this undoubtedly results in biases and amplified error in CI prediction within CAMs, as has been alluded to in past work (e.g., Weisman et al. 2008; Coniglio et al. 2013).

The central issue of the CI problem is thus two-fold. It is an issue that, on one hand, is the result of inadequate observational density and frequency and, on the other, is the result of inadequate knowledge, leading to the inadequate modeling of finer-scale processes that dictate CI occurrence. Future research on the topic of CI is necessary in order to acquire better understanding of the convective-scale processes influencing its occurrence and to aid the improvement of the physical parameterization packages, particularly PBL schemes that play a large role in bias and error development in numerical forecasts of DMC. Also warranted is the continued advancement in ensemble modeling techniques that can help quantify uncertainty in the smaller-scale (meso-gamma-scale and finer) details that strongly control CI occurrence.

Given the realization that error growth is typically greater in higher resolution, convection-allowing models than in coarser convection-parameterizing models beyond day 1 forecasts (e.g. Clark et al. 2010; Zhang et al. 2002, 2003), this study will focus on examining and quantifying the short-range (0-12 h) forecast skill of a very high-resolution model in predicting CI across the west-central High Plains of the United States.

Methodology

In this research, we investigate CI events that occurred during the 2010 convective season across the west-central High Plains of the United States with very high-resolution numerical simulations (Figure 1). This region is chosen for two primary reasons. First, the west-central High Plains contain a diverse set of topographic features that, through synoptic flow interaction and sensible heating, develop mesoscale circulations that are capable of priming the environment for and, in some cases, initiating DMC (Figure 2). The most notable topographic features in this region are the Black Hills in western South Dakota and the Front Range of the Rocky Mountains across central and eastern Colorado and Wyoming. Two additional smaller-scale topographic features adjacent to the Front Range include the Palmer Divide and Cheyenne Ridge that are responsible for developing the Denver convergence vorticity zone (e.g., DCVZ; Denver cyclone) and Chugwater circulation, respectively (Abbs and Pielke 1986; Wilczak and Christian 1990; Schreiber-Abshire and Rodi 1991; Davis 1997). Second, the 2010 convective season from mid-May through mid-August was anomalously active across the selected region (Blunden et al. 2011; personal accounts), with numerous cases of CI forced by a large variety of triggering mechanisms.

We focus on quantifying the ability of a high-resolution numerical model to accurately forecast CI when only synoptic- to meso-alpha-scale features are resolved within its initial and lateral boundary conditions. Both observed and simulated CI objects are objectively defined through radar reflectivity-based methods (e.g., Gremillion and

Orville 1999). Model performance is evaluated by the ability to replicate CI occurrence within realistic time and distance constraints.

a. Model description and data

A sample set of CI cases is created by selecting twenty-five days between May 15 and August 15, 2010 that exhibited a noticeable CI event across the west-central High Plains. Most cases of CI are essentially "clean slate" events with CI occurring either during or after the late morning hours without prior ongoing convection. Several cases however, involve the observed development of isolated convection before a more pronounced CI event later in the day. Cases are chosen such that a variety of triggering mechanisms and synoptic-scale forcing is incorporated in the sample set. Two "null" cases of CI, in which CI was expected in operational forecasts 6-12 hours prior but was not observed, are included in the sample of 25 cases. Table 1 shows the selected dates of CI events with the approximate time of the first observed CI event as well as the suspected triggering mechanisms. DMC development was assessed utilizing the Storm Prediction Center's (SPC) storm events archive (http://www.spc.noaa.gov/exper/archive/events) and the University Corporation for Atmospheric Research's (UCAR) archive (http://locust.mmm.ucar.edu) of composite radar reflectivity, satellite and surface observations.

Hindcasts of CI events are simulated using version 3.3.1 of the Advanced Research version of the Weather Research and Forecasting model (WRF-ARW; Skamarock et al. 2008). To resolve CI triggering mechanisms down to spatial scales on the order of 1 km, simulations use 429 m horizontal grid spacing in a nested inner grid

across the western High Plains. This inner domain is embedded within an outer domain with 3 km horizontal grid spacing that extends across the western two thirds of North America (Figure 1). Model performance of CI is only assessed in the very high-resolution inner domain with the outer domain serving the purpose of providing lateral boundary conditions with mesoscale resolved details. Two-way feedback is permitted between the outer and inner model domains. Both domains use a hybrid sigma-pressure vertical coordinate with 35 vertical levels. Each simulation is initialized at 1200 UTC on the day of a CI event to specifically investigate the short-range (0-12 h) forecast performance of the high-resolution model configuration. Model simulations are initialized with a cold start using NAM (North American Mesoscale) model 0-h analysis. Lateral boundary conditions on the outer domain are provided by subsequent NAM 0-h analyses at 6 hour intervals. This is not believed to give an unfair advantage to model performance in hindcast simulations as these lateral boundary updates are only utilized at 1800 UTC and 0000 UTC and because the inner domain's lateral boundaries are sufficiently removed (approximately 1,000-1,500 km) from the outer domain's lateral boundaries.

Because of the grid spacing chosen for both model domains, no cumulus parameterization is used. The RRTM (Mlawer et al. 1997) and Dudhia (Dudhia 1989) schemes are used to parameterize longwave and shortwave radiation, respectively. The Noah land surface model (Ek et al. 2003) is used with four soil levels. Microphysical species are handled using the Thompson et al. (2008) bulk microphysical parameterization which includes the mixing ratio of six different microphysical species and number concentration for cloud water and ice. Planetary boundary layer (PBL) circulations are parameterized using the local mixing Mellor-Yamada-Janjic (MYJ)

scheme (Janjic 2001). While a horizontal grid spacing of 429 m is able to crudely resolve the largest eddies within the PBL, a PBL parameterization is nevertheless necessary to reasonably represent the presence and impacts of smaller-scale PBL eddies. The high spatial resolution of the model configuration is complimented with five minute interval output on the inner domain to analyze the evolution of convective-scale features and verify simulated reflectivity at the sampling frequency of observed radar reflectivity.

Analysis and verification of the model is primarily based upon radar reflectivity at the -10°C level. Radar reflectivity at the -10°C level is chosen to remove bright branding effects that may falsely amplify the occurrence and areal coverage of DMC (e.g., Gremillion and Orville 1999; Kain et al. 2008). For the numerical simulations, this data was obtained using version 4 of the Read/Interpolate/Plot (RIP; http://www.mmm.ucar.edu/wrf/users/docs/ripug.htm) post-processing software. It should be noted that the radar reflectivity calculation used in RIP does not utilize the second-moment number concentration provided by model output, although this calculation is now experimentally included in the WRF-ARW model since the release of version 3.4.1 in Fall 2012 (http://www.mmm.ucar.edu/wrf/users/wrfv3.4/updates-3.4.1.html). Although such differences in calculation method will impact the post-processed representation of radar reflectivity to some degree, the general characteristics of and the ability to produce CI objects are expected to be meaningfully represented by the chosen simulated reflectivity calculation algorithm.

Calculating observed reflectivity at an approximate -10°C height required several steps. First, three-dimensional radar data on a 0.01° latitude by 0.01° longitude grid is obtained from the National Mosaic & QPE (NMQ) project at the National Severe Storms

Laboratory (Zhang et al. 2005; Langston et al. 2007) and trimmed to fit the area of the

inner model domain. Second, hourly Rapid Update Cycle (e.g., RUC; Benjamin et al.

2004) model analysis data is obtained and interpolated to the NMQ grid. Third, linear

interpolation is used to find the approximate height of the -10°C isotherm at each

observed grid point in the RUC data, to which the observed reflectivity data is

subsequently interpolated. Note that both observed and model -10°C radar reflectivity

data is up-interpolated to a slightly coarser 1.29 km grid spacing common grid. Using this

common grid simplifies the comparison of characteristics of CI between observations and

model simulations, as described below.

b. CI object algorithm

In order to examine the holistic performance of CI in our model configuration

both in time and space, we use an object-based approach to verify CI. The horizontal grid

spacing used here invalidates the use of standard grid-to-grid verification techniques as

they will likely yield very poor model performance. Following the work of Clark et al.

(2012a, b, 2013) and Kain et al. (2013), where an algorithm was created to track objects

in post-processed model fields such as updraft helicity and simulated reflectivity in time

and space, we develop an algorithm to identify convectively-active (CA) objects based

upon radar reflectivity and track them through time. This idea stems from initial work by

Davis et al. (2006a, b; 2009), who developed a method to verify precipitation forecasts

through the characteristics of observed and forecast rainfall objects. Our algorithm starts

by interpolating the observed and the simulated reflectivity fields to a common 1.29 km

grid. Then, a Gaussian smoothing operator is applied to the interpolated model and

observed fields. A sigma value (smoothing factor) of 1.25 is chosen based upon

sensitivity testing to ensure enough smoothing to eliminate spurious grid cells with high

reflectivity but to still preserve the overall shape and intensity of actual convective

objects, even if small. Any grid cells that do not have both an interpolated reflectivity $\geq$

35 dBz and smoothed reflectivity $\geq$ 25 dBz are filtered out, further helping to remove

very small, borderline CA objects (i.e. localized, transient $\geq$ 35 dBz cells in stratiform

precipitation). From here, an expanding radial search function is used to cluster CA grid

cells that are within two grid points (~ 2.58 km) from the last cell clustered into an object.

This radial distance threshold of two grid points is found to group cells into an object that

may contain a small discontinuity in 35 dBz reflectivity but still properly separates close

neighboring objects. To ensure these groupings of CA grid cells (objects) are resolvable

features, they must have an area equal to or greater than the length scale of resolvability

(approximately eight model grid point lengths; Skamarock 2004) squared, where an 8×8

model grid cell area is ~ 12 km^2. Although lesser values, say of 4-6 grid point lengths on

the common analysis grid, can be sufficient to resolve features of interest, we choose 8

grid point lengths to create a proxy length scale of resolvability between the model and

coarser observed reflectivity. Objects with areas less than the size threshold are

considered poorly resolved or altogether unresolved by the model and/or by radar

observations and thus not indicative of true, sustained DMC features. This series of steps

is performed on every five minutes of observed and model simulated -10°C radar

reflectivity data. Figure 3 shows the beginning and end results of the CI algorithm for a

representative example from 1 August 2010.

      After each 5 minute period, the CI algorithm attempts to match each CA object to

CA objects present at previous times. CA objects are matched to previous objects when

they are within 10 minutes of each other and their centroids are within ten interpolated grid points (~ 13 km) distance. If there are multiple matches at some time to one object from the previous time, the spatially-closest object is matched. The distance matching requirement helps to properly match objects through time by constraining how far the algorithm should search, but is within the realistic bounds of distance covered by very fast storm movement in five to ten minutes (i.e., less than 78 km hr^{-1}). CI is declared once an object has been matched for 30 minutes, demonstrating a mature and sustained deep, moist convective updraft that is precipitating. The timing and location (centroid) of CI is then set from when the object was first identified as CA with the CI object area size averaged over the 30 minute interval.

c. Verifying CI objects

Once a list of observed and modeled CI objects is created for a case, the set of objects are compared to each other for model verification purposes. A verification algorithm goes through each observed CI object in a case, creating a sorted list of model CI objects that are closest in time and space. The closeness of modeled and observed CI objects is ordered based upon a flow-dependent error metric. We use an error metric that depends on the spatial and temporal errors of a modeled CI object from an observed CI object, as well as the characteristic flow, or velocity, of that object. This C^2 error metric is given as:

$$C^2 = (\text{Error}_{\text{distance}})^2 + (\text{Velocity}_{\text{characteristic}} * \text{Error}_{\text{time}})^2 \quad (1)$$

For simplicity, we represent this spatio-temporal error as C with units of km. The characteristic velocity to first order, approximates the storm object motion due to advection from the magnitude of the mean flow. We approximate this characteristic

velocity by taking the difference of the horizontal wind between the 4th and 16th vertical levels in the model at hourly intervals, representing roughly the magnitude of shear in the 850-450 hPa layer. Distance and time error thresholds are applied to matching as done in similar studies (e.g. Fowle and Roebber 2003; Weisman et al. 2008; Duda and Gallus 2013), albeit more strictly so in this study, such that if a modeled CI object exceeds either a timing error of three hours or a distance error of 200 km, it is unmatched and classified as a false negative. In such a case, this means that the closest model CI object cannot be matched to an observed CI object and demonstrates the model's inability to reasonably produce a CI object in time and space to an observed one. These error thresholds correspond to an approximate C error of about 294 km for a characteristic velocity average from all cases of around 20 m/s.

Because we are specifically focusing on the spatio-temporal character of CI objects, rather than just its occurrence alone, a potential issue arises near the lateral boundaries of the common verification grid. That is, the possibility that CI objects that develop outside of the common verification grid and move into this grid may incorrectly be flagged as CI. Given the climatological mean westerly flow across our domain area, this potential issue is greatest near the western lateral boundary of the common grid domain. Since there are inherently cases with a meridional component of the mean flow due to synoptic-scale wave disturbances, this possibility also exists near the southern and northern common grid lateral boundaries. For consistency, we also consider the less likely possibility of storms propagating westward through the eastern domain boundary (i.e. case of a cut-off low or negatively-titled synoptic short-wave trough). To mitigate this problem, we apply a "buffer zone" near all lateral boundaries of the common grid

domain in which all CI objects occurring within the buffer zone are discarded. Accounting for potentially fast storm motions (up to 43 m/s), a buffer zone of ten interpolated-grid points (~13 km) inward from the lateral boundaries is applied. Say for example, a very fast traveling storm initially located outside of the common grid domain moves 10 km into the common grid domain in a five minute interval. Further say, that this storm is considered convectively active in the common grid domain for sufficient time and resultantly flagged CI. Applying this buffer zone would discard this CI object because its true start time and location (defined when first flagged CA) is not captured within our high-resolution domain area of interest (common grid). This buffer zone is applied in all cases for deriving model CI performance results.

Table 1. Selected dates, start times and apparent triggering mechanisms of CI events.

Date	Approx. CI start	Initiating Mechanism (s)
5/21	20:30 UTC	Frontal boundary, orographic
5/26	19:00 UTC	Outflow boundary, orographic
5/29	20:30 UTC	Frontal boundary
6/3	"null" event	N/A
6/4	22:00 UTC	Frontal boundary
6/9	19:00 UTC	Frontal boundary, orographic
6/10	19:00 UTC	Frontal boundary, orographic
6/13	21:00 UTC	Frontal boundary
6/16	19:30 UTC	Dryline, orographic
6/21	19:00 UTC	Frontal boundary, orographic
6/24	"null" event	N/A
6/26	18:30 UTC	Frontal boundary, orographic
7/2	18:00 UTC	Orographic
7/3	17:30 UTC	Frontal boundary, orographic
7/4	14:00 UTC	Frontal boundary, orographic
7/6	19:30 UTC	Frontal boundary, orographic
7/9	19:00 UTC	Orographic
7/10	18:00 UTC	Frontal boundary, orographic
7/15	21:00 UTC	Orographic, wind shift line
7/17	21:00 UTC	Frontal boundary
7/21	19:00 UTC	Frontal boundary, orographic
7/27	20:00 UTC	Frontal boundary, orographic
7/29	18:30 UTC	Frontal boundary, orographic
8/1	19:00 UTC	Frontal boundary, orographic
8/9	14:30 UTC	Orographic, outflow boundary
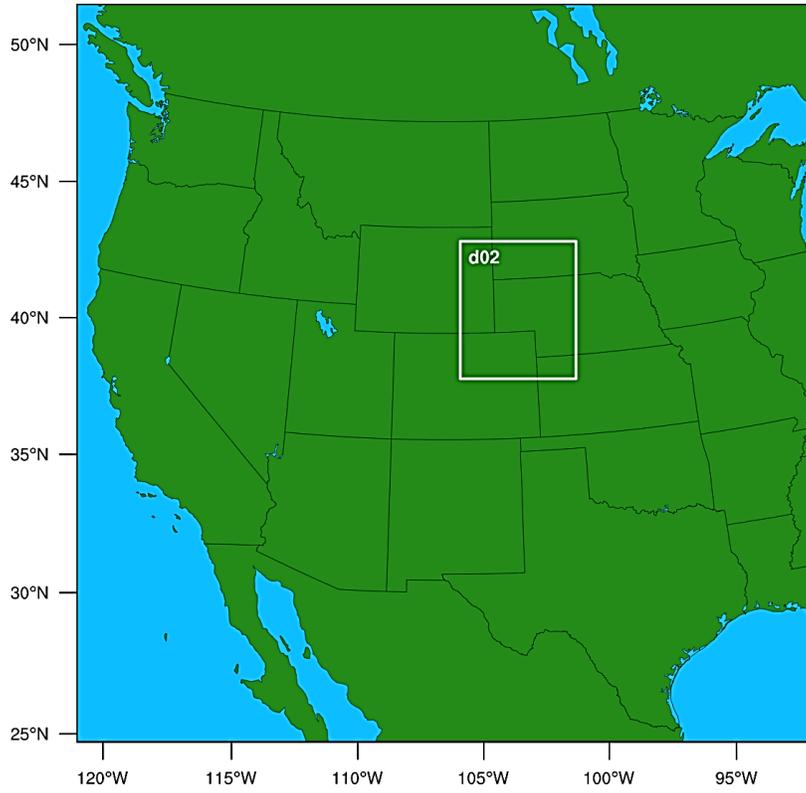
Figure 1. Outer domain with inner domain (d02) shown in white box.
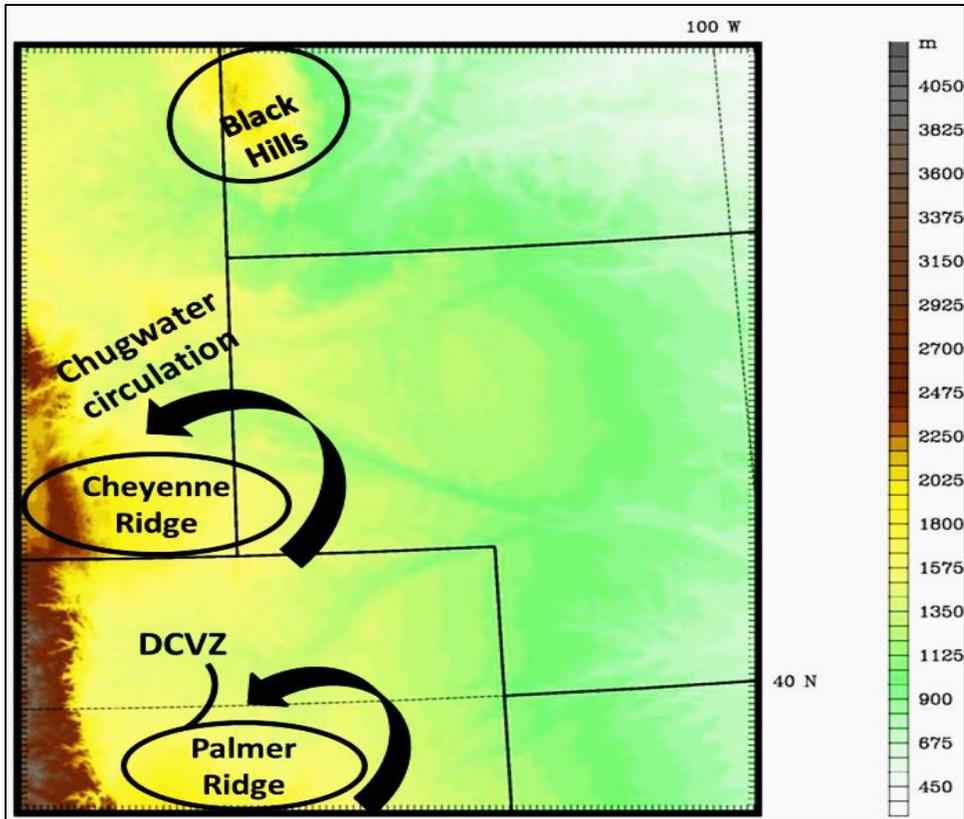
Figure 2. Surface elevation (m) above sea level in inner model domain. The Front Range of the Rocky Mountains is not labeled but encompasses the western third of the domain. Arrows denote typical low-level flow direction with notable terrain-induced circulations.
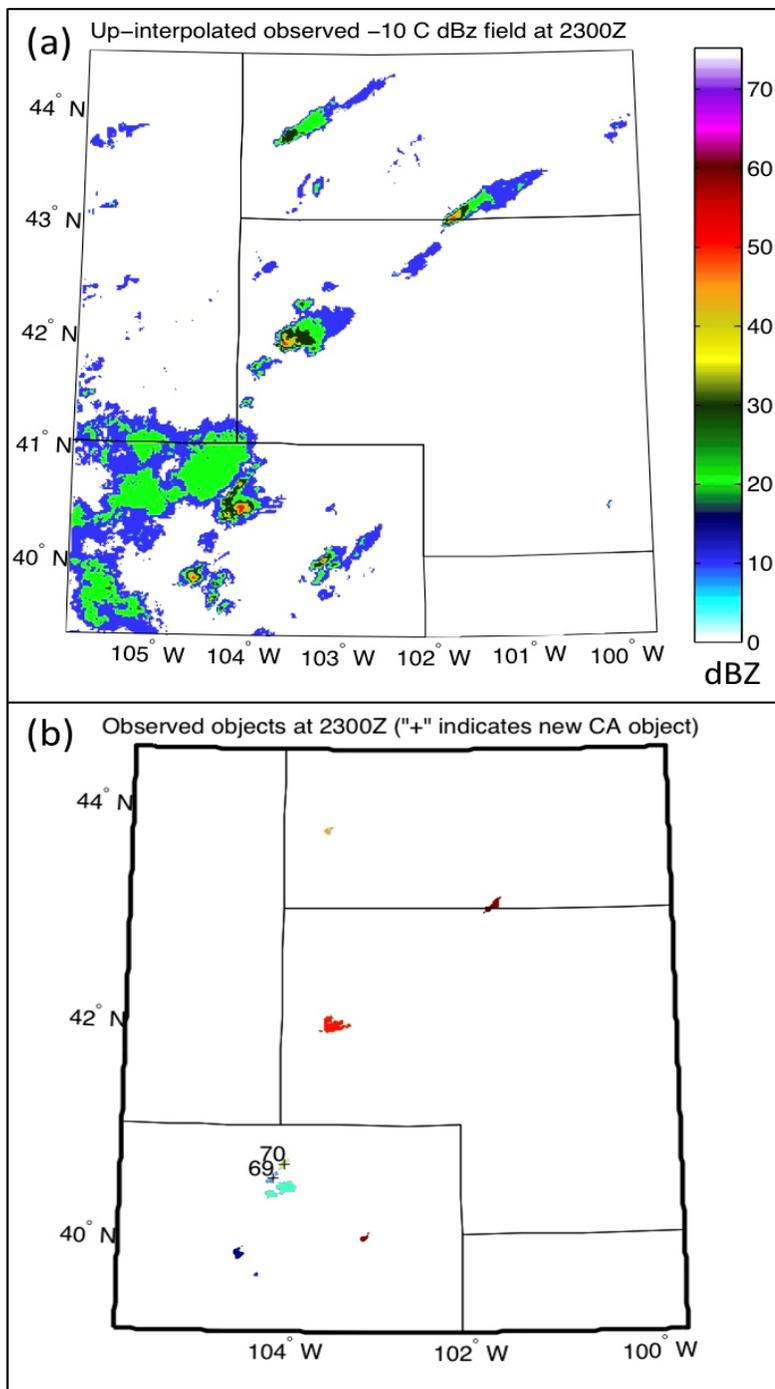
Figure 3. An example of CI algorithm output from 2300 UTC
1 August 2010 showing (a) interpolated observed reflectivity
and (b) identified CA objects.

Results

This section will analyze the performance of all simulations of our model configuration in three ways. First, we use some standard error metrics on the occurrence of CI based upon a contingency-based classification of CI objects. In conjunction, model performance is assessed via an evaluation of temporal and spatial errors associated with model forecast CI objects. Second, notable model biases and tendencies that limit model performance are addressed. Third, results of sensitivity testing accomplished by varying physical parameterizations are discussed for a small sample of CI events.

a. Model verification

Although we cannot realistically verify all aspects of CI by standard skill scores, as has been done with coarser grid spacing CAM studies (e.g. Fowle and Roebber 2003; Weisman et al. 2008; Stratman et al. 2013), we are able to verify the occurrence of CI with some standard skill scores through the use of a contingency table (Wilks 1995; Fowle and Roebber 2003). Through all 25 cases, model simulations produce a total of 717 CI objects compared to only 272 observed CI objects. A total of 208 model CI objects are successfully matched with observed CI objects that are within the spatial and temporal error thresholds mentioned above, and are considered true positives (Figure 4, "a"). Another 411 modeled CI objects are considered false positives that the model overproduces in excess of actual CI objects observed (Figure 4, "b"). This overproduction of modeled CI objects will be further discussed later in this paper. An additional 41 model CI objects are considered for matching, but fail the temporal and spatial error thresholds described in the Methodology section, and thus are not successful matches.

These objects are considered false negatives, in which CI objects are observed, but the model does not produce CI objects that are physically representative in time and/or space (Figure 4, "c"). A fourth classification in the contingency table is the true negative, which on a grid-to-grid verification, is when forecast occurrence is correctly negative (Figure 4, "d"). In an object-based study, however, this classification is inapplicable as there is no way of forecasting a non-existent object.

The collection of classified objects on a contingency table can then be used to calculate summary measures, quantifying the overall model performance in forecasting the occurrence of some feature (Murphy 1993; Roebber 2008), here being CI. One metric is the probability of detection (POD), a ratio that ranges from 0 to 1 representing the ratio of correctly forecasted objects to the total number of observed CI objects. A POD of 1 indicates all observed CI objects are forecast (Eq. 2). A second metric is the false alarm ratio (FAR), a ratio of overproduced forecast CI objects to the total number of CI objects forecasted that also ranges from 0 to 1. A value of 1 indicates that CI objects are forecasted but none are observed (Eq. 3). A third metric is the bias score, which is the ratio of the number of overforecast CI objects to the number of underforecast CI objects (Eq. 4). The bias score can be any real number, with a value of 1 indicating equal POD and FAR scores. The final metric used is the critical success index (CSI), which is the ratio of correctly forecasted objects to the total number of CI objects correctly and incorrectly forecasted (Eq. 5).

$$POD = \frac{a}{a+c} , \qquad\qquad (2)$$

$$FAR = \frac{b}{a+b} , \qquad\qquad (3)$$

$$\text{bias} = \frac{a+b}{a+c} \, , \qquad\qquad\qquad (4)$$

$$\text{CSI} = \frac{a}{a+b+c} \, , \qquad\qquad\qquad (5)$$

The model's ability to develop CI objects through all cases, as constrained by the spatio-temporal thresholds mentioned earlier, reveals a probability of detection (POD) score of 0.835. This however, comes at the expense of a false alarm ratio (FAR) of 0.664, a result of the significant overproduction of modeled CI, quantitatively given by a bias of 2.49 model CI objects per 1 observed CI object. Further, the model produces a critical success index (CSI) score of 0.315. Although these performance measures generally fall short of the performance of both operational and non-operational models alike (e.g., Fowle and Roebber 2003; Roebber 2009; Duda and Gallus 2013; Kain et al. 2013), we emphasize that they should be viewed in light of the rather strict spatio-temporal thresholds used to match forecast to observed CI objects and determine the dichotomous outcome of model CI objects. To put in context the scores of model CI occurrence, the duality of error, as mentioned in Doswell (2004), demonstrates that in order to, say, decrease the FAR of model CI occurrence, the POD consequently will also need to decrease as false positive occurrences are converted to false negatives.

It should be noted that, for the two observed "null" cases, model simulations developed at least one CI object, thus giving a POD of 0 and FAR of 1 for those cases. Results from the 23 cases in which CI was observed are plotted (Figure 5) on a Roebber performance diagram (Roebber 2009), showing the performance of CI occurrence as well as average object C errors. This diagram visualizes multiple measures of forecast quality, including POD, success ratio (SR; 1-FAR), bias, and CSI. For this study, we also add in coloring to denote the average object C error for each case. In general, better performing

forecasts will fall towards the top-right portion of the diagram near the y = x line, indicating high POD and SR and near unity bias. Most points on this diagram are skewed to the top-left, indicating higher values of POD and bias, coinciding with the model overproduction of CI objects. Several points are seen along the right y-axis, indicating a FAR of 0, but relatively lower values in POD. These points represent four cases in which model simulations actually underforecast the number of observed CI objects.

As mentioned earlier, we use a flow-dependent error metric to quantify the spatial and temporal errors of a modeled CI object. The flow dependence is represented in the temporal error term, where a characteristic velocity that represents an approximate deep-layer shear or storm motion vector is multiplied to the timing error. The derivation of this characteristic velocity is further defined in the Methodology section. This flow dependence serves two purposes. One is to allow for consistency in dimensional units between the spatial and temporal errors, and two, to correctly characterize timing errors associated differences in storm motion. For the 208 model CI objects matched, an average early error of 7.48 minutes is seen, with the distribution of timing errors near normally distributed about zero time error. The average distance error for modeled CI objects is 62.1 km. The average C error, incorporating the distance and timing errors of matched CI objects, is approximately 91.5 km per object. CI occurrence and spatial-temporal error results are shown in Table 2. Both the average object timing and distance errors are similar to results of Duda and Gallus (2013), who found average model object errors of -0.56 min (earlier than observed) and 105 km for CI occurring prior to the subsequent upscale growth of DMC. Average object C errors for all cases range from about 55 km to 150 km. Also of note, we find for all cases the matched model CI objects

on average are larger than their observed counterparts in areal coverage by about 16 km^2. These results coincide well with those of Davis et al. (2006b), who found convective rainfall objects were too numerous and large in areal coverage in a 4 km CAM study. It is interesting to see such common tendencies within CAM studies with horizontal grid spacing differences of one full order of magnitude, suggesting that such biases simply do not go away with the use of finer horizontal grid spacings and are likely an artifact of a more general issue with the explicit handling of DMC.

Cases of CI are broken into categories of weak and strong synoptic-scale forcing to investigate any possible correlation between large-scale forcing and model performance. The strength of synoptic forcing is subjectively determined from the presence of any synoptic-scale frontal boundaries and/or notable mid-level geostrophic vorticity advection during a CI event. Moderate-to-strong event classification required at least one of the two mentioned synoptic factors to be strong or both to be moderately strong. Not surprisingly, a majority (17) of the CI cases and matched CI objects (145/208) are characterized by moderate to strong synoptic-scale forcing, a notable characteristic of the active convective season across this region in 2010. Cases involving stronger synoptic forcing did show slightly lower average object errors in distance (55.5 km) and time (-6.52 min) compared to cases of weaker synoptic forcing, which had average errors of 77.2 km and -9.68 min in distance and time, respectively. Note, that negative timing errors indicate model CI objects occurring earlier than those observed. Similar results are seen in the average spatio-temporal C errors, with the stronger synoptic forcing cases scoring an average C error of 85.6 km and the weaker synoptic cases scoring 103.8 km. These C error differences occur despite stronger synoptic forcing

cases having a slightly stronger average characteristic velocity magnitude (19.4 m/s versus 15.5 m/s), which amplifies the temporal error term in the C^2 error equation. Interestingly, the weaker synoptic forcing cases on average scored better for the occurrence of CI, with a CSI score of 0.360 compared to 0.299 with the stronger synoptically-forced cases. To test the significance of these results we use the bootstrap method to simulate a large sample set of the difference in average object C error and average CSI scores in all cases of strong versus weak synoptic forcing. Results show that neither the improved average object C error in stronger synoptic forcing cases nor the improved average CSI score in weaker synoptic forcing cases are statistically significant at the 75% confidence level.

These results are similar to those of Duda and Gallus (2013), who found no identifiable relationship between large-scale forcing and forecast model skill of CI in their 3 km CAM simulations. However, the weak relationship we find between model performance (average C error) and synoptic forcing compared to Duda and Gallus (2013), who use different model domain placements across the U.S. to sample CI events, may be because our inner domain comparatively includes more topographically-induced circulations. These circulations, which are capable of promoting CI, are partially driven by the more predictable large-scale background flow, as noted by Mass et al. (2002). This suggests that although the occurrence of CI is more strongly influenced by meso- to micro- scale features, the predictability of CI may be enhanced if these features are more dependent upon larger-scale forcing. Additionally, intuition would suggest that stronger synoptic forcing would aid in the predictability of convection in a CAM, or any numerical weather model, since CI is expected to be more dependent upon larger-scales

when larger-scale forcing is stronger. Although synoptic forcing plays a large role in modulating DMC type and evolution (Schumann and Roebber 2010, among others), its role in supporting CI appears mainly to be in producing a more thermodynamically favorable environment, as the precise occurrence of CI is still ultimately dictated by the smaller, meso-gamma- to micro- scale features (Weckwerth et al. 2004, Markowski et al. 2006). However, if synoptic-scale forcing can help to reduce CIN and lower the height of a parcel's LCL and LFC, then theoretically the ascent with smaller-scale features would not need to be as strong.

b. Model tendencies and biases of CI

As previously mentioned and evident in the forecast bias, the model overproduces CI objects by a factor of approximately 2.5. Although matched CI objects on average demonstrate good model performance, they are 'buried', in a sense, by numerous falsely produced model CI objects. From an operational forecasting standpoint, having a model that produces CI objects that closely matches what is observed is significantly degraded in utility if those objects are surrounded by many other false modeled CI objects both in time and space. Upon further investigation, these overproduced model CI objects show certain tendencies that highlight, in time and space, where overproduction is more likely, and also shed light on potential reasons for model CI overproduction.

To illustrate the bias and tendencies of all model CI objects, we analyze their temporal and spatial distributions in comparison to observations. For the temporal distribution, the time of all – and not just matched – CI objects are binned in 5 minute intervals spanning the range of all CI times from 1330 to 0100 UTC (Figure 6, top). We

fit a high order polynomial to the CI time distributions to better illustrate the difference, or spread, in CI frequency per time between model simulations and observations (Figure 6, middle). We also fit a high order polynomial to the ratio of model CI objects to observed CI objects to visually demonstrate the times of greatest overproduction relative to the observed CI count (Figure 6, bottom). This shows that the greatest model overproduction of CI occurs during the peak times of CI occurrence roughly between 1800 UTC and 2200 UTC. On average, model forecast CI objects (matched and unmatched) form 43 minutes too early compared to observed CI objects (Figure 6, top).

To best illustrate the spatial distribution of all CI objects, a weighted CI density field is created. We first sum the CI count at each grid point from all 25 cases on the 1.29 km common verification grid. Then, similar to Davis et al. (2006a, b), we apply a Gaussian-like weighting function that is summed at each grid cell to smooth the total CI count,

$$\text{Weighting function} = e^{-c\left(\frac{d^2}{10^2}\right)}, \qquad\qquad (6)$$

Where "c" is the number of CI events in any grid cell and "d" is the distance from any grid cell with 'count' CI events to the cell in which the weighting is being applied. The numerator in the exponential function represents the decaying scale factor, which we choose as 10 interpolated grid lengths. This smoothing factor of 10 grid cell lengths, a similar value to that used in Davis et al. (2006a), is subjectively chosen by testing varying values and determining that it best visually demonstrates areas of local maxima and minima without entirely eliminating small-scale maxima and minima. The differenced

spatial CI density between the model and observations helps to show areas in which the model produces too many, and too few, CI objects (Figure 7).

At first sight, two widespread biases are apparent in the model minus observed spatial density image (Figure 7). The most noticeable difference is the spatial coverage of overproduction (red shading) of model CI across much of the inner domain area. In general, a slight model underproduction of 1-2 CI objects (blue shading) is seen across much of western Nebraska with several other smaller-scale areas of model CI underproduction across the inner model domain.

The overproduction of CI objects in the model relative to observations is most notable near the western inner domain boundary extending east to the Front Range of central and eastern Wyoming and Colorado. Several reasons for CI object overproduction in the model are possible. One is that the observed radar sampling is more limited across far west portions of the inner model domain near the peaks of the Front Range Mountains. In these areas, radar data from the Denver and Cheyenne WSR-88D sites is partially obscured by terrain, thus likely limiting to (some extent) the number of observed CI objects across western portions of the domain. See Figure 8 for an illustrative example. Another potential reason for this overproduction is the location of our western inner domain boundary near steep gradients in elevation height. We attempt to mitigate this issue by placing the western lateral boundary away from the steepest gradients of topography within the Rocky Mountains (e.g., along the Continental Divide rather than to its east or west). Further, outer-inner domain interface blending of orography in the WRF-ARW model that helps to smoothly ramp up or down topography height within the first 5 grid points of the inner domain from domain interface (Skamarock et al. 2008) is

hypothesized to mitigate most, if not all, possible spurious, short-wavelength energy development from strong topography gradients near the domain interface. The third potential reason is the propagation of explicitly-resolved DMC with strong perturbations to the thermodynamic and kinematic fields from the outer domain into the inner domain. Although CI objects that move into the inner domain area, whether modeled or observed, are discounted, it is speculated that such convective objects may promote low-frequency spurious wave growth near the lateral boundary that may aid in forcing additional CI events. However, we speculate that this should be mitigated with the use of an odd outer-inner domain grid spacing ratio (7:1) on the WRF-ARW default Arakawa-C grid, which allows for variables to have the same physical point at the interface between the outer and inner domain (Skamarock et al. 2008). Conditions from the outer domain moving into the inner domain then only need to be down interpolated to fit the first row or column of inner domain grid cells. This simple downscaling does not allow for gradients in meteorological variables in the outer domain, say associated with DMC, to be amplified (at least initially) as they move into the inner domain. Additionally, we note that there are no significant visual signs of such wave activity influencing DMC near the lateral boundaries through the analysis of derived simulated radar reflectivity.

To first order, the possible lack of radar observed CI over higher terrain would appear to be the more likely reason for model overproduction of CI relative to observed CI. All of the mentioned potential reasons for overproduction near the western lateral boundary, however, should be at least partially mitigated with the use of a ten grid point (common grid) buffer zone along each directional lateral boundary.

c. Physical parameterization scheme sensitivity and case analyses

Many recent high-resolution modeling studies have suggested that physical parameterizations are not best fit for horizontal grid spacings that are capable of explicitly resolving DMC (e.g., Mass et al. 2002; Done et al. 2004, Skamarock 2004; Davis et al. 2006b; Stensrud 2007; Weisman et al. 2008; Schwartz et al. 2009; Coniglio et al. 2010 ; Coniglio et al. 2013; Kain et al. 2013) . This has been most suggested for PBL schemes, where CAMs are capable of resolving some of the larger-scale PBL circulations that may superpose with those parameterized in a PBL scheme (Stensrud 2007). This may be especially true in this study with the use of 429 m horizontal grid spacing. High-resolution CAM studies have also shown a large forecast sensitivity and ensemble spread to different microphysical parameterization schemes (Clark et al. 2010; Coniglio et al. 2010; Schwartz et al. 2010), related particularly to the strength of cold pools produced by simulated DMC. These findings motivate the testing forecast of forecast performance for different PBL and microphysics schemes.

In addition to the overall early and overproduction bias of CI in model simulations, many simulations tend to show numerous small, pulsing convective objects in radar reflectivity, including at the -10°C level. We refer to these small spatial and temporal objects as plumes, seeming to be the model's representation of deepening cumulus clouds associated with local maxima of ascent extending from the PBL. These pluming objects often develop collectively in fields, sometimes as early as 1-2 hours before CI, like that of an observed *cumulus congestus* field aiding in the reduction of CIN prior to CI. It is uncertain if such objects directly impact the occurrence and tendency of modeled CI. However, because their spatial scale is typically on the order of several model grid point lengths (429 m), most if not all pluming objects are discarded as CA by

the area size threshold in our CI algorithm. One possible reason for these pluming objects is that they represent *cumulus congestus* towers that may be too fine-scale and/or weak to be detected and resolved by operational radars (WSR-88D). In observations, localized plumes, manifest as horizontal convective rolls, are frequently observed in pre-convective environments and can be identified via visible satellite imagery. However, they are often too weak in vertical height and cloud water content, to be meaningfully represented by land-based radars. Consequently, the presence of plumes within the numerical simulations presented herein is not unexpected, but the intensity of such plumes, particularly in the context of their simulated reflectivity structure, is unexpectedly high (35-45 dBz) compared to observations. Given the lack of literature on model performance at the horizontal grid spacing utilized here, and the noting of improper tuning of physical parameterizations for CAM applications, we speculate that shortcomings in physical parameterizations could be responsible for these plume objects.

To investigate how different physical parameterizations impact model evolution and CI performance, we choose three cases and run several series of model simulations with different physical parameterizations. We specifically choose to investigate cases of CI on 26 May, 1 August, and 9 August 2010 that demonstrate numerous pluming objects in simulated radar reflectivity to see if the development of plumes is impacted by the choice of physical parameterization. Microphysical sensitivity is examined by utilizing the Morrison 5-class double-moment scheme (Morrison et al. 2005) in lieu of the Thompson et al. (2008) microphysical scheme. The Morrison scheme is specifically chosen to investigate how the added double-moment representation of more microphysical species (among differences in algorithm structure, empirical values,

physical assumptions, etc.) impacts model performance compared to the hybrid single-moment control Thompson scheme, which is a double-moment scheme only for cloud water and ice.

Planetary boundary layer sensitivity is examined by utilizing the Yonsei University (YSU; Hong et al. 2006) "non-local" scheme in lieu of the MYJ PBL scheme. The YSU PBL scheme is specifically chosen to compare its "non-local" representation of turbulent mixing to the MYJ's "local" representation of turbulent mixing, as the two tend to have a comparative difference in PBL structure in CAM applications. Notably, "local" schemes such as the MYJ scheme tend to produce cooler, moister PBL structures, while "non-local" schemes such as the YSU scheme tend to produce warmer, drier PBL structures in DMC environments (Weisman et al. 2008; Coniglio et al. 2013). These differences may strongly impact the occurrence, timing, and location of CI, all of which can significantly alter the evolution of DMC.

Although this sample of physical parameterization sensitivity testing is small, there are some weakly supported results that merit further investigation in later studies. In two out of the three simulations, the YSU PBL scheme has a reduced overproduction tendency as compared to the MYJ PBL, as exhibited by slightly lower bias scores. In all three cases, the Morrison microphysics scheme produces slightly lower CSI scores than the Thompson scheme, mainly due to an amplified overproduction of CI objects. Through all three cases, however, the control simulations produce slightly higher CSI scores than both the YSU PBL and Morrison microphysics runs. Statistics from the sensitivity runs can be seen in Table 3. Through all three cases, these pluming objects are noticeably weaker and fewer in number in YSU simulations compared to simulations using the MYJ

PBL scheme (Figure 9). Of note, the Morrison microphysics scheme tended to have larger areas of higher reflectivity in the strongest convective cells in comparison to the Thompson scheme, but visually appeared to be a closer representation to reflectivity values in observed convection. It should again be noted, however, that the simulated reflectivity calculation algorithm used here did not utilize the second-moment number concentration of the microphysical species.

Analyzing vertical profile soundings at approximate locations of CI about 1 h prior to average CI timing between different physical parameterization shows no discernible difference in low-level temperature or PBL depth (not shown). However, visually comparing the near-surface distribution of dewpoint temperature values between simulations with different PBL schemes shows that in all three cases, the YSU scheme tends to produce dewpoint values that are consistently around 1° C lower than in the MYJ. This matches well with other studies showing the YSU scheme tends to be drier, while the MYJ scheme tends to be moister in the PBL (Coniglio et al. 2013). Although these differences in dewpoint are not drastic, slight differences in low level moisture can significantly impact CAPE, where a 1 g/kg change in surface mixing ratio can change surface based CAPE by up to 20% (Bluestein 1993). Again, however, given the small sample size, these results and their implications toward the use of widely-used physical parameterization packages for high-resolution CAM applications require further investigation.

Observed

	Yes	No
Yes	a	b
No	c	d

Forecast

Figure 4. A 2 x 2 contingency table, with shaded boxes denoting correct forecasts.
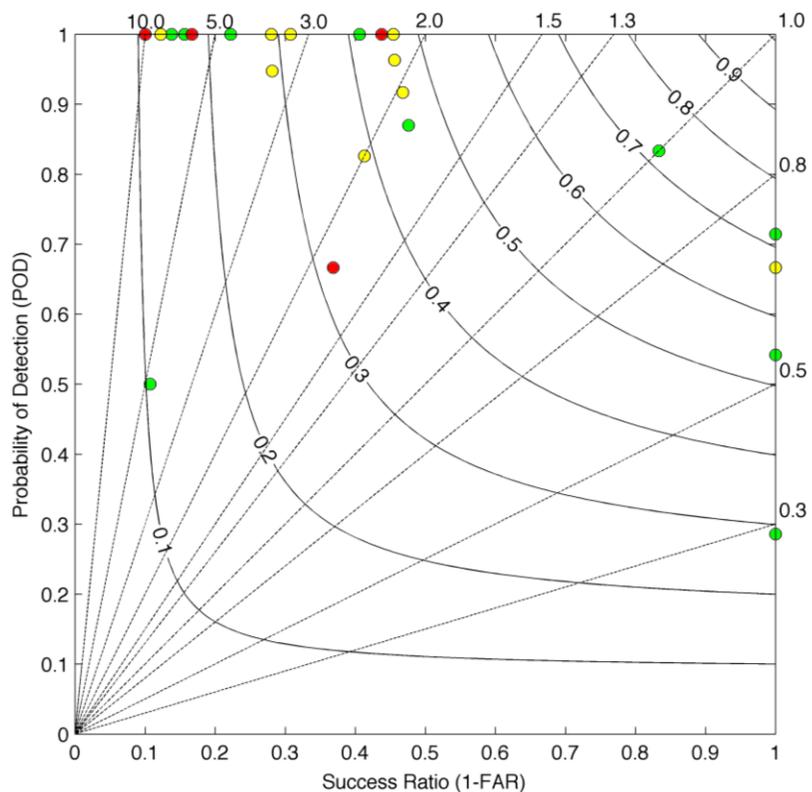
Figure 5. Performance (Roebber) diagram for each CI case showing SR on the x-axis and POD on the y-axis. Bias is shown through dashed lines and CSI through curved, solid lines. Circles are filled based upon the case average object *C* error. Green denotes *C* errors less than 80 km, yellow 80-110 km and red greater than 110 km.

Table 2. Case average model performance ordered by CSI score. Average object *C* errors are color coded as mentioned in Figure 5.

Date	CSI	C error(km)
6/16	0.7143	63.66
7/9	0.7143	72.41
7/15	0.6667	84.45
7/6	0.5417	72.62
7/3	0.4545	90.69
7/21	0.4490	97.03
8/1	0.4483	91.60
7/10	0.4444	71.70
5/26	0.4375	147.44
7/29	0.4063	71.05
6/10	0.3800	82.62
8/9	0.3111	125.00
6/9	0.3077	82.82
7/17	0.2857	67.03
5/29	0.2800	96.09
6/26	0.2769	96.28
6/21	0.2222	68.88
5/21	0.1667	118.23
7/2	0.1563	66.64
7/27	0.1379	55.20
7/4	0.1224	107.36
6/4	0.1000	132.63
6/13	0.0968	55.69
6/3	0.0000	NaN
6/24	0.0000	NaN
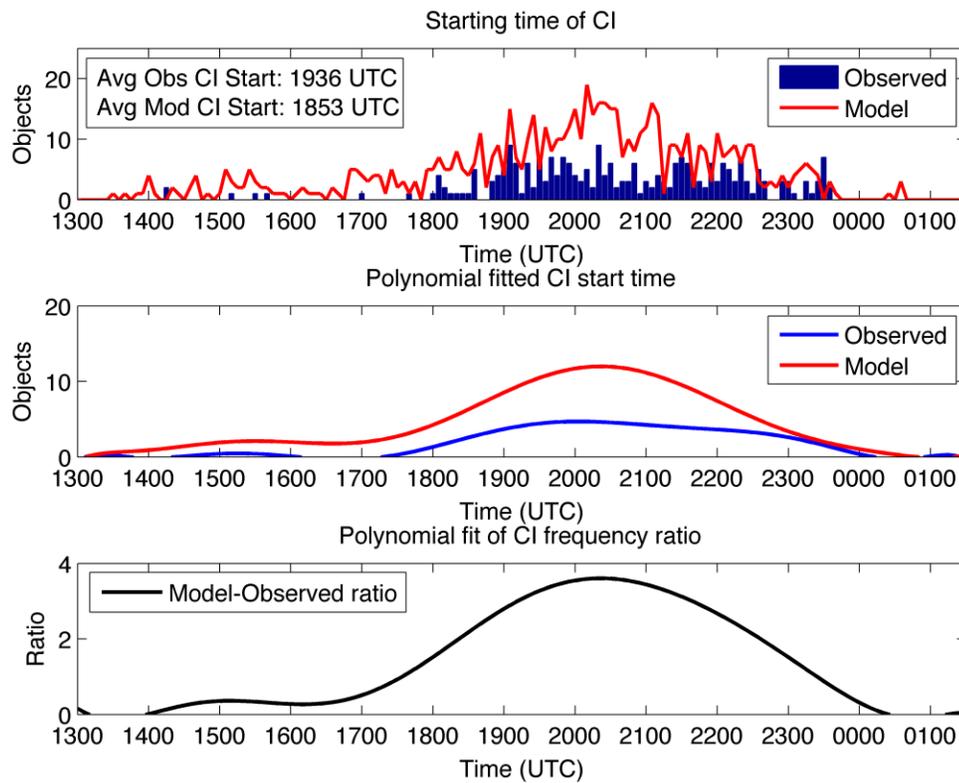
Figure 6. (top) Temporal distribution of all CI objects in 5 minute bin intervals, (middle) its fitted polynomial distribution and (bottom) fitted polynomial of the ratio of modeled to observed object count.

Model-Observed CI spatial density difference
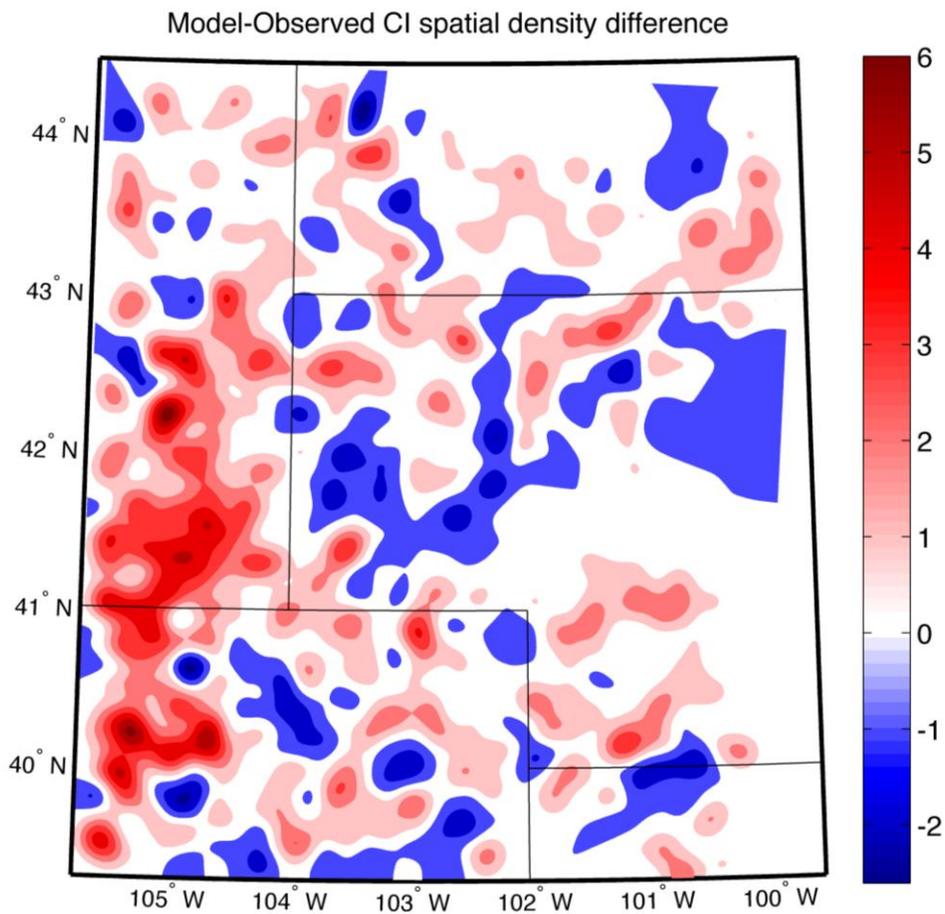
Figure 7. Weighted spatial density difference in CI occurrence between model and observed, where positive (negative) values denote more modeled (observed) CI objects.
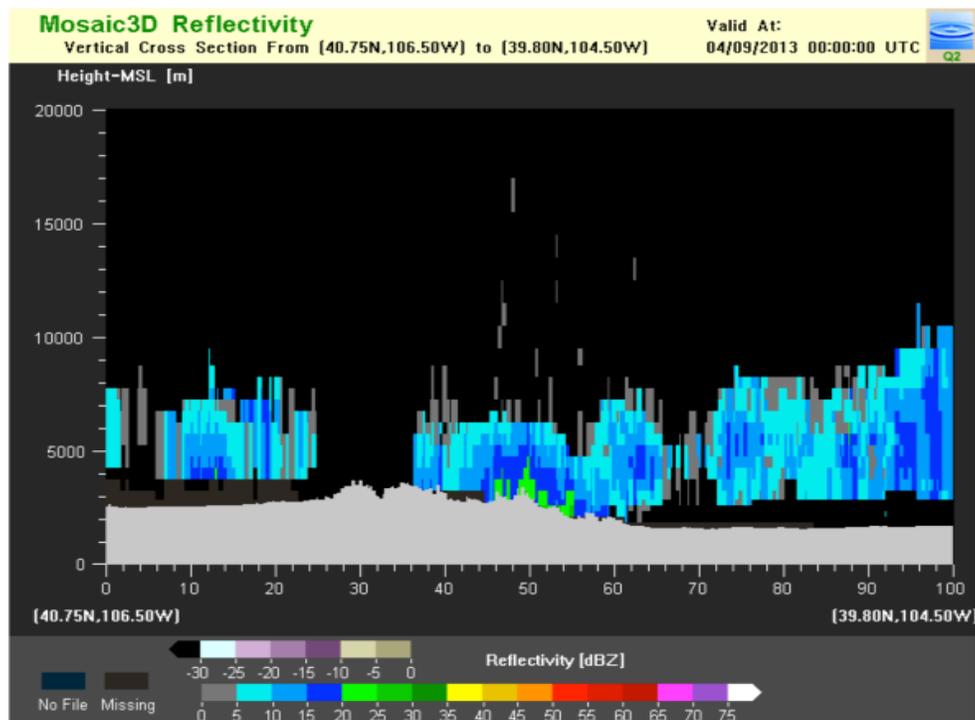
Figure 8. An example cross section of 3D data from the NSSL NMQ product. From right to left is the approximate Denver/Boulder, CO WSR-88D location extending radially into north-central Colorado along the Front Range of the Rocky Mountains. Note the "Missing" dark grey shading beneath areas of reflectivity west of the highest terrain (light grey shading). The NMQ product accordingly weights reflectivity values from the neighboring Cheyenne, WY WSR-88D site for north-central Colorado (left portion of cross section).

Table 3. Physical parameterization scheme sensitivity results of CI occurrence. Following Table 1, "A" denotes the number of matched observed and model CI objects, "B" denotes the number of overproduced CI objects, and "C" denotes the number of observed objects that model simulations did not produce or match within specific thresholds.

| Date	Scheme	A	B	C	FAR	POD	bias	CSI
26-May	CONTROL	7	9	0	0.563	1	2.29	0.438
	YSU	4	7	3	0.636	0.571	1.57	0.286
	MORRISON	4	11	3	0.733	0.571	2.14	0.222
1-Aug	CONTROL	26	31	1	0.544	0.963	2.11	0.448
	YSU	18	24	9	0.571	0.667	1.56	0.353
	MORRISON	26	71	1	0.732	0.963	3.59	0.265
9-Aug	CONTROL	14	24	7	0.632	0.667	1.81	0.311
	YSU	16	32	5	0.667	0.762	2.29	0.302
	MORRISON	15	28	6	0.651	0.714	2.05	0.306
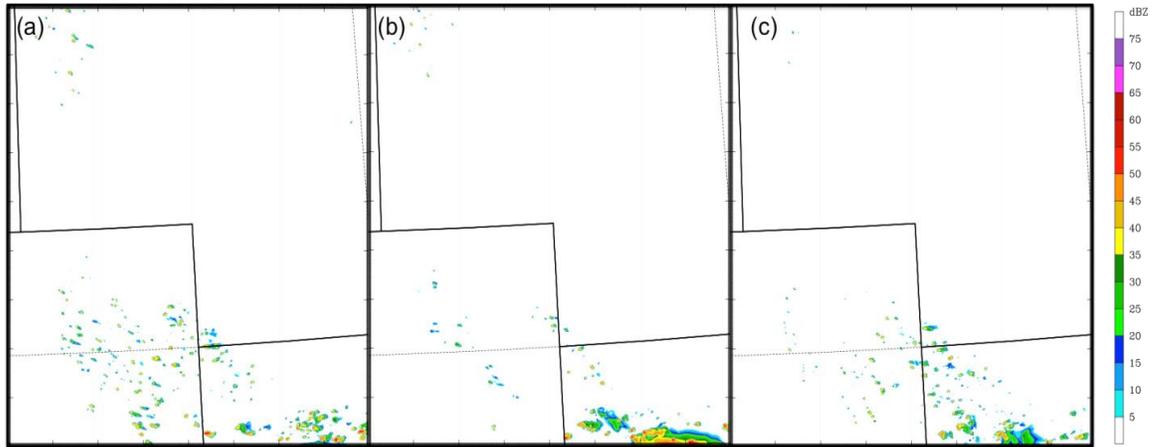
Figure 9. Simulated reflectivity at the -10°C level at 2000 UTC 26 May with the (a) MYJ PBL, Thompson microphysics control simulation, (b) Morrison microphysics simulation and (c) YSU PBL simulation. These images are zoomed in across NE CO, W NE and NW KS to better show small-scale details.

Discussion

Two topics, highlighting potential issues in high-resolution forecasting, are discussed: the use of PBL physical parameterization in CAM applications and the impact of incorporating non-standard observations into high-resolution simulations.

A "grey zone" (Stensrud 2007) has been identified for PBL parameterization use in CAMs, where at horizontal grid spacings <4 km, some of the larger-scale PBL eddies are beginning to be resolved, a process that is traditionally handled in a PBL parameterization scheme. The horizontal grid spacing we use in our model simulations (429 m) would appear to be well into this "grey zone", as we are resolving more finer-scale details than in most CAMs with horizontal grid spacings of 2-4 km but are still not properly resolving most features in the PBL as in large-eddy simulations with horizontal grid spacings of order 1-10 m. This is analogous to the issue of whether to use a convective parameterization with horizontal grid spacings between 4 km and 10 km (Molinari and Dudek 1992) that are beginning to resolve the larger-scale structure of DMC. This complicates the decision of choosing explicit or parameterization of turbulent energy transport in the PBL. This, though, creates the concern of resolving turbulent energy at small wavelengths; i.e., too much if a PBL parameterization is utilized and too little if one is not utilized. Since it has been suggested that larger-scale PBL eddies are beginning to be crudely resolved in CAMs of grid spacing 2-4 km, therefore likely superposing with output from a PBL parameterization (Stensrud 2007), this issue is likely amplified to an extent in our model simulations. This highlights two critical points to be addressed regarding CAMs and PBL parameterizations. A threshold in horizontal and vertical grid spacing to adequately and accurately resolve all eddies within the PBL

responsible for turbulent energy transport needs to be identified. In other words, at what grid spacing should a PBL parameterization no longer be used? The second and more practical point is the need for the tuning or development of PBL schemes that are better suited for CAM applications, accounting for the ability of CAM simulations to resolve, at least crudely, larger-scale PBL circulations.

Incorporating more observations into ICs and LBCs for model simulations has been shown in many studies to improve model performance. This is especially true in CAMs for very short-range forecasts, as demonstrated in several studies incorporating non-standard observations through the utilization of advanced data assimilation techniques (e.g., Xue and Martin 2006; Liu and Xue 2008; Sun et al. 2012). These studies, however, incorporate numerous observations from the IHOP field program that both exceed the resolution of standard observational systems and focus on their impact upon numerical forecasts within nine hours of the forecast initialization time. Less studied is the impact of more observations that would slightly exceed the number of observations currently provided by observational systems, but could be realistically obtained without the means of a field program or similar construct. This is the one of the objectives for the Mesoscale Predictability EXperiment (MPEX; Weisman et al. 2012), which will investigate whether additional sub-synoptic-scale observations gathered in environments immediately upstream of where DMC is forecast to develop can improve day 1 (6-24 h) forecast skill of the formation, evolution, and impacts of DMC.

Summary

Very high resolution WRF-ARW simulations with horizontal grid spacings of 429 m are run for twenty-five CI case events across the west-central High Plains from the 2010 convective season. Specifically, we investigate the period from 15 May to 15 August 2010, which demonstrated numerous CI events as triggered by a diverse set of atmospheric and orographic triggering mechanisms. We investigate the ability of very high resolution model simulations to replicate observed CI by evaluating forecast skill in the context of that documented in previous CI-focused modeling studies. Model performance is assessed using an object-based approach, classifying CI as spatially- and temporally-sustained radar reflectivity-derived objects. The occurrence of CI is verified with traditional statistical scores.

Objective matching of modeled and observed CI objects demonstrates an average early bias of 7.48 minutes per matched model object. The average distance error is 62.1 km, with model objects tending to be slightly too large in areal coverage. The average spatio-temporal, flow-dependent C error is 91.5 km per object. Matching is only considered for model objects that are within three hours and 200 km of an observed object. The model's ability to produce matched CI objects is given by a POD of 0.835. However, a significant overproduction of CI objects compared to observations (717:272) produces a FAR of 0.664 and bias ratio of 2.49 modeled CI objects to 1 observed CI object. The overall CSI score for all cases is 0.315. For all CI objects, matched and unmatched, between model simulations and observations, model simulations show a 43 minute early bias due to CI overproduction. Model overproduction in CI is most spatially

prevalent along and west of the Front Range of the Rocky Mountains and most temporally prevalent in the 1800-2200 UTC time range, during peak heating and peak observed CI frequency. This constrains the present operational utility of such a model configuration with positive performing matched CI objects being 'buried' in time and space by the model's overproduced CI objects.

Model performance is found to be slightly higher in regards to average spatio-temporal C errors, albeit statistically-insignificantly, in CI cases with stronger synoptic forcing. The opposite is true for the occurrence scores of CI, with weaker synoptic forcing cases performing better. The average CSI score for stronger synoptic cases is 0.299 versus 0.360 for weaker synoptic cases. In essence, model simulations with stronger synoptic forcing tend to produce more CI objects, and thus lower CSI scores due to amplified CI overproduction, but, the greater number of model CI objects yields slightly closer matches in time and space to observed CI objects. The average object distance and timing errors are improved by 21.7 km and 3.17 minutes in cases of stronger synoptic forcing, respectively. We hypothesize that this very slight added spatio-temporal CI skill in cases of stronger synoptic forcing may be due to the dependence of mesoscale topographic circulations (Mass et al. 2002) along the Front Range of the Rocky Mountains that can locally favor CI upon synoptic-scale details.

Although a limited return in forecast skill improvement is expected for model simulations utilizing 429 m horizontal grid spacing compared to the more typical 3-4 km spacing in CAMs (Kain et al. 2008; Schwartz et al. 2009; Clark et al. 2011), since both can sufficiently resolve convective-scale details, there are a couple of specific reasons as to why such improvements are inherently limited in nature. One is the interaction and

superposing of explicitly resolved large-scale PBL turbulence with physically-parameterized turbulent energy. Although the existence of this issue has been speculated by many (e.g., Skamarock 2004, Strensund 2007) in regards to CAMs with grid spacings of 3-4 km, how much this may be further amplified in simulations with grid spacing one full order of magnitude smaller is unknown. A second potential limiting factor is insufficient observations in the model initial conditions. Although mesoscale-resolving models are capable of developing mesoscale circulations with only large-scale atmospheric details provided, a "spin-up" time is required to resolve such features, limiting forecast performance in the first few hours of model integration. How the inclusion of non-standard observations, although not to the detail obtained in specialized field campaigns, impacts regional CAMs beyond very short-range forecast times (0-9 h) warrants further study.

References

Abbs, Deborah J., Roger A. Pielke, 1986: Thermally forced surface flow and convergence patterns over northeast colorado. *Mon. Wea. Rev.*, **114**, 2281–2296.

Agee, Ernest, Alexander Gluhovsky, 1999: Les model sensitivities to domains, grids, and large-eddy timescales. *J. Atmos. Sci.*, **56**, 599–604.

Adlerman, Edwin J., Kelvin K. Droegemeier, 2002: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters. *Mon. Wea. Rev.*, **130**, 2671–2691.

Benjamin, Stanley G., and Coauthors, 2004: An hourly assimilation–forecast cycle: the ruc. *Mon. Wea. Rev.*, **132**, 495–518.

Bluestein, H. B., 1993: *Synoptic–Dynamic Meteorology in Midlatitudes.* Vol. II, *Observations and Theory of Weather Systems.* Oxford University Press, 446 pp.

Blunden, J., D. S. Arndt, M. O. Baringer, 2011: State of the climate in 2010. *Bull. Amer. Meteor. Soc.*, **92**, S1–S236.

Bryan, George H., John C. Wyngaard, J. Michael Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416.

Clark, Adam J., William A. Gallus, Ming Xue, Fanyou Kong, 2010: Growth of spread in convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **25**, 594–612.

Clark, Adam J., and Coauthors, 2011: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.*, **139**, 1410–1418.

Clark, Adam J., and Coauthors, 2012a: An overview of the 2010 hazardous weather testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Clark, Adam J., John S. Kain, Patrick T. Marsh, James Correia, Ming Xue, Fanyou Kong, 2012b: Forecasting tornado pathlengths using a three-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.

Coniglio, Michael C., Kimberly L. Elmore, John S. Kain, Steven J. Weiss, Ming Xue, Morris L. Weisman, 2010: Evaluation of wrf model output for severe weather forecasting from the 2008 noaa hazardous weather testbed spring experiment. *Wea. Forecasting*, **25**, 408–427.

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

Davis, Christopher A., Barbara G. Brown, Randy Bullock, John Halley-Gotway, 2009: The method for object-based diagnostic evaluation (mode) applied to numerical forecasts from the 2005 nssl/spc spring program. *Wea. Forecasting*, **24**, 1252–1267.

Crook, N. Andrew, 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767–1785.

Davis, Christopher A., 1997: Mesoscale anticyclonic circulations in the lee of the central rocky mountains. *Mon. Wea. Rev.*, **125**, 2838–2855.

Davis, Christopher, Barbara Brown, Randy Bullock, 2006a: Object-based verification of precipitation forecasts. part i: methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

Davis, Christopher, Barbara Brown, Randy Bullock, 2006b: Object-based verification of precipitation forecasts. part ii: application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795.

Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecast (WRF) model. *Atmos. Sci. Lett.*, **5,** 110–117.

Doswell, Charles A., 1987: The distinction between large-scale and mesoscale contribution to severe convection: a case study example. *Wea. Forecasting*, **2**, 3–16.

Doswell, Charles A., 2004: Weather forecasting by humans—heuristics and decision making. *Wea. Forecasting*, **19**, 1115–1126.

Duda, Jeffrey D., William A. Gallus, Jr., 2013: The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in the WRF. *Wea. Forecasting*, (In Press).

Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J. Atmos. Sci., **46**, 3077–3107.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarplay, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res*., **108**, 8851, doi:10.1029/2002JD003296.

Fowle, Michael A., Paul J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794.

Gremillion, Michael S., Richard 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.

Hill, K., G. S. Wilson, and R. E. Turner, 1979: NASA's participation in the AVE-SESAME '79 program. *Bull. Amer. Meteor. Soc*, **60,** 1323–1329.

Hong, Song-You, Yign Noh, Jimy Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341.

Houston, Adam L., Dev Niyogi, 2007: The sensitivity of convective initiation to the lapse rate of the active cloud-bearing layer. *Mon. Wea. Rev.*, **135**, 3013–3032.

Houston, Adam L., Robert B. Wilhelmson, 2011: The dependence of storm longevity on the pattern of deep convection initiation in a low-shear environment. *Mon. Wea. Rev.*, **139**, 3125–3138.

Hsiao, Ling-Feng, Melinda S. Peng, Der-Song Chen, Kang-Ning Huang, Tien-Chiang Yeh, 2009: Sensitivity of typhoon track predictions in a regional prediction system to initial and lateral boundary conditions. *J. Appl. Meteor. Climatol.*, **48**, 1913–1928.

Janjic, Z. I., 2001: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso Model. NOAA/NWS/NCEP Office Note 437, 61 pp.

Johns, Robert H., Charles A. Doswell, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612.

Jorgensen, David P., Tammy M. Weckwerth, 2003: Forcing and organization of convective systems. *Meteorological Monographs*, **30**, 75–75.

Kain, John S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing nwp. *Wea. Forecasting*, **23**, 931–952.

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., (In Press).

Kang, Song-Lak, George H. Bryan, 2011: A large-eddy simulation study of moist convection initiation over heterogeneous surface fluxes. *Mon. Wea. Rev.*, **139**, 2901–2917.

Kong, Fanyou, Kelvin K. Droegemeier, Nicki L. Hickmon, 2006: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. part i: comparsion of coarse- and fine-grid experiments. *Mon. Wea. Rev.*, **134**, 807–833.

Langston, Carrie, Jian Zhang, Kenneth Howard, 2007: Four-dimensional dynamic radar mosaic. *J. Atmos. Oceanic Technol.*, **24**, 776–790.

Lee, Bruce D., Richard D. Farley, Mark R. Hjelmfelt, 1991: A numerical case study of convection initiation along colliding convergence boundaries in northeast colorado. *J. Atmos. Sci.*, **48**, 2350–2366.

Liu, Haixia, Ming Xue, 2008: Prediction of convective initiation and storm evolution on 12 june 2002 during ihop_2002. part i: control simulation and sensitivity experiments. *Mon. Wea. Rev.*, **136**, 2261–2282.

Lorenz, Edward N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646.

Markowski, Paul, Christina Hannon, Erik Rasmussen, 2006: Observations of convection initiation "failure" from the 12 june 2002 ihop deployment. *Mon. Wea. Rev.*, **134**, 375–405.

Martin, William J., Ming Xue, 2006: Sensitivity analysis of convection of the 24 may 2002 ihop case using very large ensembles. *Mon. Wea. Rev.*, **134**, 192–207.

Mass, Clifford F., David Ovens, Ken Westrick, Brian A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts?. *Bull. Amer. Meteor. Soc.*, **83**, 407–430.

Melhauser, Christopher, Fuqing Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-

k model for the longwave. J. Geophys. Res., **102**, 16 663–16 682.

Molinari, John, Michael Dudek, 1992: Parameterization of convective precipitation in mesoscale numerical models: a critical review. *Mon. Wea. Rev.*, **120**, 326–344.

Morrison, H., J. A. Curry, 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.

Murphy, Allan H., 1993: What is a good forecast? an essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.

Roebber, Paul J., Lance F. Bosart, 1998: The sensitivity of precipitation to circulation details. part i: an analysis of regional analogs. *Mon. Wea. Rev.*, **126**, 437–455.

Roebber, Paul J., David M. Schultz, Romualdo Romero, 2002: Synoptic regulation of the 3 may 1999 tornado outbreak. *Wea. Forecasting*, **17**, 399–429.

Roebber, Paul J., David M. Schultz, Brian A. Colle, David J. Stensrud, 2004: Toward improved prediction: high-resolution and ensemble modeling systems in operations. *Wea. Forecasting*, **19**, 936–949.

Roebber, Paul J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.

Romine, Glen S., Craig S. Schwartz, Chris Snyder, Jeff L. Anderson, Morris L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284.

Schwartz, Craig S., and Coauthors, 2009: Next-day convection-allowing wrf model guidance: a second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372.

Schwartz, Craig S., and Coauthors, 2010: Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.

Schumann, Melissa R., Paul J. Roebber, 2010: The influence of upper-tropospheric potential vorticity on convective morphology. *Mon. Wea. Rev.*, **138**, 463–474.

Schreiber-Abshire, Wendy, Alfred R. Rodi, 1991: Mesoscale convergence zone development in northeastern colorado under southwest flow. *Mon. Wea. Rev.*, **119**, 2956–2977.

Skamarock, William C., 2004: Evaluating mesoscale nwp models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.

Skamarock, W. C., et al., 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, 61 pp., available at http://www.mmm.ucar.edu/wrf/users/docs/arw v3.pdf.

Sobash, Ryan A., John S. Kain, David R. Bright, Andrew R. Dean, Michael C. Coniglio, Steven 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.

Stensrud, D. J. 2007: Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models. Cambridge Press, 459 pp.

Stevens, Bjorn, and Coauthors, 2005: Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Mon. Wea. Rev.*, **133**, 1443–1462.

Stratman, Derek R., Michael C. Coniglio, Steven E. Koch, Ming Xue, 2013: Use of multiple verification methods to evaluate forecasts of convection from hot- and cold-start convection-allowing models. *Wea. Forecasting*, **28**, 119–138.

Stull, Roland B., 1991: Static stability—an update. *Bull. Amer. Meteor. Soc.*, **72**, 1521–1529.

Sun, Juanzhen, Stanley B. Trier, Qingnong Xiao, Morris L. Weisman, Hongli Wang, Zhuming Ying, Mei Xu, Ying Zhang, 2012: Sensitivity of 0–12-h warm-season precipitation forecasts over the central united states to model initialization. *Wea. Forecasting*, **27**, 832–855.

Thompson, Gregory, Paul R. Field, Roy M. Rasmussen, William D. Hall, 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.

Weckwerth, Tammy M., Thomas W. Horst, James W. Wilson, 1999: An observational study of the evolution of horizontal convective rolls. *Mon. Wea. Rev.*, **127**, 2160–2179.

Weckwerth, Tammy M., 2000: The effect of small-scale moisture variability on thunderstorm initiation. *Mon. Wea. Rev.*, **128**, 4017–4030.

Weckwerth, Tammy M., and Coauthors, 2004: An overview of the international h2o project (ihop_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.*,

85, 253–277.

Weckwerth, Tammy M., David B. Parsons, 2006: A review of convection initiation and motivation for ihop_2002. *Mon. Wea. Rev.*, **134**, 5–22.

Weckwerth, Tammy M., Hanne V. Murphey, Cyrille Flamant, Janine Goldstein, Crystalyne R. Pettet, 2008: An observational study of convection initiation on 12 june 2002 during ihop_2002. *Mon. Wea. Rev.*, **136**, 2283–2304.

Weisman, Morris L., William C. Skamarock, Joseph B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548.

Weisman, Morris L., Christopher Davis, Wei Wang, Kevin W. Manning, Joseph B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the wrf-arw model. *Wea. Forecasting*, **23**, 407–437.

Weisman, Morris L., Chris Davis, Glen Romine, Robert J. Trapp, 2012: The Mesoscale Predictability Experiment (MPEX). [available online at http://www.eol.ucar.edu/projects/mpex/documents/MPEX_Proposal_041613.pdf].

Weiss, Steven, Jack Kain, Mike Coniglio, James Correia, Israel Jirak, Chris Melick, Patrick Marsh, Adam Clark, Chris Siewert, Conrad Ziegler, Ryan Sobash, Andy Dean, Tara Jensen, David Novak, Mike Bodner, Faye Barthold, 2011: EXPERIMENTAL FORECAST PROGRAM SPRING EXPERIMENT 2011. [available online at http://hwt.nssl.noaa.gov/Spring_2011/Spring_Experiment_2011_ops_plan_13May_v5.pdf].

Wilczak, J. M., T. W. Christian, 1990: Case study of an orographically induced mesoscale vortex (denver cyclone). *Mon. Wea. Rev.*, **118**, 1082–1102.

Wilks, D. S., 1995: *Statistical Methods in Atmospheric Sciences: An Introduction.* Academic Press, 500 pp.

Wilson, James W., James A. Moore, G. Brant Foote, Brooks Martner, Taneil Uttal, James M. Wilczak, Alfred R. Rodi, 1988: Convection initiation and downburst experiment (cinde). *Bull. Amer. Meteor. Soc.*, **69**, 1328–1348.

Wilson, James W., G. Brant Foote, N. Andrew Crook, James C. Fankhauser, Charles G. Wade, John D. Tuttle, Cynthia K. Mueller, Steven K. Krueger, 1992: The role of boundary-layer convergence zones and horizontal rolls in the initiation of thunderstorms: a case study. *Mon. Wea. Rev.*, **120**, 1785–1815.

Wilson, James W., Rita D. Roberts, 2006: Summary of convective storm initiation and evolution during ihop: observational and modeling perspective. *Mon. Wea. Rev.*,

134, 23–47.

Zhang, F., Chris Snyder, Richard Rotunno, 2002: Mesoscale predictability of the "surprise" snowstorm of 24–25 january 2000. *Mon. Wea. Rev.*, **130**, 1617–1632.

Zhang, F., Chris Snyder, Richard Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185.

Zhang, Jian, Kenneth Howard, J. J. Gourley, 2005: Constructing three-dimensional multiple-radar reflectivity mosaics: examples of convective storms and stratiform rain echoes. *J. Atmos. Oceanic Technol.*, **22**, 30–42.

Ziegler, Conrad L., Erik N. Rasmussen, 1998: The initiation of moist convection at the dryline: forecasting issues from acase study perspective. *Wea. Forecasting*, **13**, 1106–1131.