



January 2014

Forecast Bias Analysis Using Object-Based Verification Of Regional WRF Summertime Convective Forecasts

Mariusz Starzec

Follow this and additional works at: <https://commons.und.edu/theses>

Recommended Citation

Starzec, Mariusz, "Forecast Bias Analysis Using Object-Based Verification Of Regional WRF Summertime Convective Forecasts" (2014). *Theses and Dissertations*. 1714.
<https://commons.und.edu/theses/1714>

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.

FORECAST BIAS ANALYSIS USING OBJECT-BASED
VERIFICATION OF REGIONAL WRF SUMMERTIME
CONVECTIVE FORECASTS

by

Mariusz Starzec
Bachelor of Science, University of North Dakota, 2012

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota

December

2014

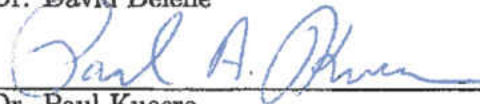
This thesis, submitted by Mariusz Starzec in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.



Dr. Gretchen Mullendore



Dr. David Delene



Dr. Paul Kucera

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.



Wayne Spisher

Dean of the School of Graduate Studies



DATE

PERMISSION

Title	Forecast Bias Analysis Using Object-Based Verification Of Regional WRF Summertime Convective Forecasts
Department	Atmospheric Sciences
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in her absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Mariusz Starzec
December 11, 2014

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	xi
ACKNOWLEDGMENTS	xiii
ABSTRACT	xiv
CHAPTER	
1 INTRODUCTION	1
2 BACKGROUND	4
2.1 Convective Verification	4
2.2 Model Sensitivity	9
2.3 Microphysical Properties	12
3 DATA	16
4 METHODOLOGY	20
4.1 Preprocessing for Forecast Verification	20
4.2 Convective Object Verification	21
5 RESULTS	29
5.1 Campaign Overview	29
5.2 Object Verification and Bias Summary of Forecasted Con- vection	31
5.3 Comparison of 3 km and 1 km Forecasts	41
5.4 Uncertainty Discussion of Objective Verification	44
5.5 Case Study	48
5.5.1 Selection of Case Day	48
5.5.2 Setup of Sensitivity Study	54
5.6 Case Study Sensitivity Analysis and Findings	62

5.6.1	Discussion of Observed Cloud Droplet Concentration Sensitivities	69
5.6.2	Discussion of Microphysical Parameterization Sensitivities	79
5.6.3	Discussion of Horizontal Resolution Sensitivities	82
5.7	Robustness of Sensitivity Analysis	84
6	DISCUSSION	88
6.1	Implications of the Objective Verification	88
6.2	Implications of the Sensitivity Study	93
7	CONCLUSIONS	101
	REFERENCES	106

LIST OF FIGURES

Figure	Page	
1	<p>Simulated reflectivity (top), observed reflectivity (bottom), and corresponding traditional skill scores (right) where a perfect forecast is a 1. The traditional skill scores are the Critical Success Index (CSI), Equitable Threat Score (ETS), Hanssen Kuipers (HK) and Heidke Skill Score (HSS).</p>	6
2	<p>Domains used for a) POLCAST3 3 km, b) POLCAST4 3 km, and c) & d) POLCAST4 1 km simulations where the blue boxes indicate nested domains within the parent grid (outer black box).</p>	17
3	<p>An example of the masking procedure used when comparing forecasts to observations, where the white denotes areas that contains data used and gray denotes a masked region, where no data is available. Panel a) denotes the original model domain and b) denotes the model domain after being masked to the radar range.</p>	21
4	<p>The following figure details how the Method for Object-based Diagnostic Evaluation (MODE) tool analyzes data. MODE begins with the (a) raw data field, (b) applies a circular convolution for object merging, (c) masks data and applies any thresholds, and (d) declares objects [Image from Davis et al., 2006a].</p>	23
5	<p>An example of unfavorable merging and matching of objects by MODE, with no convolution used in both the forecast (left column) and observation (right column) fields. The raw reflectivity field (top row), cluster and object field (middle row), and object number (bottom row) are shown. The Rubber-banded objects of the same color in the middle row represent merged objects, with similarly colored objects in both the forecast and observation columns indicating a match. Dark blue objects are objects that have not been merged or matched.</p>	25

6	Number of all forecasted (darker) and all observed (light with diagonal lines) cells greater than 5 dBZ for the entire POLCAST3 time period (June 27 - Aug. 6, 2012), binned according to cell size.	26
7	Number of cases and their corresponding differences between observed and forecasted convective areal coverage when convection existed in both domains across the entire POLCAST3 field campaign (June 22 - July 23, 2010)	27
8	The mean monthly precipitation (blue line) and standard deviation (black lines) of 33 years of precipitation data starting from 1981 for Grand Forks, ND airport (KGFK). The red diamonds indicate the monthly precipitation during the 2010 POLCAST3 campaign year and the green diamonds represent the monthly precipitation during the 2012 POLCAST4 campaign year.	30
9	The mean monthly average temperature (blue line) and associated standard deviation (black lines) of 33 years of temperature data starting from 1981 for Grand Forks, ND airport (KGFK). The red diamonds indicate the monthly average temperature during the 2010 POLCAST3 campaign year and the green diamonds represent the monthly average temperature during the 2012 POLCAST4 campaign year.	31
10	The number of forecasted (darker) and observed (lighter with diagonal lines) objects generated by MODE binned according to area across all P3 simulations at reflectivity thresholds of 5 (green), 30 (yellow), and 45 dBZ (red)	33
11	Similar to Figure 10 except for P4a simulations.	33
12	Similar to Figure 10 except for P4b simulations.	34
13	The percentage of cases across the P3, P4a, and P4b simulations that were either hits (dark blue), null cases (light blue), false alarms (yellow), or misses (red) at differing intensity thresholds of 5, 30, and 45 dBZ. . .	34
14	Differences between the forecasted and observed total area covered by convection relative to the domain size for P3 simulations at 5 (green), 30 (yellow), and 45 dBZ (red).	37
15	MODE comparison of the 3 km a) simulated reflectivity and b) observed reflectivity fields at 1 km AGL altitude at 12 UTC on July 6, 2012. . . .	38

16	The amount of area covered by simulated convection relative to the amount of area covered by observed convection at each matching time, at 5 dBZ for all P3 simulations. Percentages greater than 100% (green) indicate over-forecasting of areal coverage, while values below 100% (yellow) indicate under-forecasting of coverage.	39
17	Simulated reflectivity field at 1-km height for the a) 1 km resolution and b) 3 km resolution simulations valid at 2230 UTC on Aug. 4, 2012. . . .	43
18	Observed reflectivity field at 1 km AGL height showing noise in the domain during a matching time in the P3 campaign.	47
19	500 mb upper air observations, heights (solid lines), and isotherms (dashed red lines) valid at a) 00 and b) 12 UTC on Aug. 4, 2012 [Images courtesy of the Storm Prediction Center].	51
20	Surface observations and isobars (solid lines) valid at a) 00 and b) 12 UTC on Aug. 4, 2012 [Images courtesy of the Storm Prediction Center].	52
21	1 km composite radar reflectivity mosaic at a) 23 UTC on Aug. 3, 2012, b) 03 UTC on Aug. 3, 2012, and c) 18 UTC on Aug. 4, 2012 [Images courtesy from the College of DuPage NeXt Generation Weather Lab (NEXLAB)].	54
22	Time series of 10 minute averaged surface-based cloud condensation nuclei (CCN) concentrations starting 12 UTC Aug. 3, 2012.	56
23	Box plots of cloud condensation nuclei (CCN) concentrations at three different supersaturations of 0.2, 0.3, and 0.6% across the entire POLCAST4 field campaign. The line inside the box indicates the median, the upper and low bounds are the 75th and 25th percentiles respectively, and the extent of the whisker bars denotes the upper and lower fences. The mean is indicated by a 'x' and outliers are denoted by '+'.	59
24	The a) 9 km and b) 333 m, 1 km, and 3 km domains used for sensitivity testing. For b), the inner domain is the 333 m resolution grid, nested within the 1 km and 3 km grids.	61
25	The total number of objects present greater than 5 dBZ (top) and 30 dBZ (bottom) as identified by MODE, across the times of interest for the case day of Aug. 4, 2012.	63
26	333 m simulated reflectivity field valid on 20 UTC on Aug. 4, 2012. . . .	64

27	The total area covered by convection greater than (top) 5 dBZ and (bottom) 30 dBZ as identified by MODE, across the times of interest for the case day of Aug. 4, 2012.	66
28	Total hydrometeor mass of rain, snow, and graupel across the domain and times of interest for the case day of Aug. 4, 2012, at greater than 5 dBZ (top) and 30 dBZ (bottom).	72
29	Total mass of rain for convective areas greater than 5 dBZ across the domain and times of interest for the case day of Aug. 4, 2012.	73
30	The WSM6 autoconversion rates of cloud water to rain hydrometeors for varying cloud water mixing ratios. Different colored lines indicate different cloud droplet concentrations.	75
31	The total mass of cloud water (solid lines) and cloud ice (dashed lines) present across the domain of the 3-km simulations when convection was greater than 5 dBZ in intensity across the times of interest on the case day of Aug 4., 2012.	76
32	The total mass of snow present across domain of different model simulations when convection was greater than 5 dBZ in intensity across the times of interest on the case day of Aug. 4, 2012.	77
33	The WSM6 heterogeneous freezing rates of cloud water to cloud ice at a constant air density of 0.80 kg m^{-3} and temperature of -20°C for varying cloud water mixing ratios. Different colored lines indicate different cloud droplet concentrations.	78
34	The Thompson (solid lines) and WSM6 (dashed lines) autoconversion rates of cloud water to rain hydrometeors for varying cloud water mixing ratios. Different colored lines indicate different cloud droplet concentrations.	82
35	The area covered by convection (left) and the number of objects (right) greater than 5 dBZ, during the times of radar operation on the day of July 8, 2012.	85
36	Similar to Figure 35 except for the day of July 9, 2012.	85
37	Similar to Figure 35 except for the day of July 12, 2012.	87

38	MODE comparison of the 3 km grid spacing a) simulated reflectivity and b) observed reflectivity fields at 1 km AGL altitude at 22 UTC on Aug. 3, 2012.	92
----	--	----

LIST OF TABLES

Table	Page	
1	Physics used for the third and fourth Polarimetric Cloud Analysis and Seeding Test (POLCAST) campaign Weather Research and Forecasting (WRF) model runs.	18
2	Ratios of the number of forecasted objects across all POLCAST3 3 km (P3), POLCAST4 3 km (P4a), and POLCAST4 1 km (P4b) simulations to the number of observed objects by UND radar, binned by object sizes corresponding to Figures 10, 11, and 12. "-" denote cases where no observed objects matched the area and intensity thresholds, hence a ratio cannot be determined.	35
3	The number of times when either the forecasts or observations covered more area when convection existed in both domains. The FO ratio represents the number of times forecasted convective area was greater than the observed convective area.	40
4	The total number of objects, summation of all object areas (km ²), and the average area per object for all simulations and observations, when convection was present in both domains.	42
5	Same as Table 2, but for the P4b domain size and for analysis times that had both P4a and P4b model data.	45
6	The percentage of cases that were either successful in predicting convection (or lack thereof), produced false alarms, or missed observed convection, when both the 3-km (P4a) and 1- km (P4b) simulations were evaluated on matching domains and matching time periods.	46

7	Overview of the different WRF model simulations used for the case day of Aug. 8, 2012. '*' denotes the change in microphysics from WSM6 to Thompson, keeping the default Thompson 100 cm^{-3} cloud droplet concentration as opposed to the default 300 cm^{-3} in WSM6.	57
8	Average object sizes presented for the 14 to 00 UTC time period on Aug. 4, 2012.	67
9	The ratios of the number of forecasted objects to observed objects and the total forecasted area to observed area across the 14 to 00 UTC time period on Aug. 4, 2012. "-" denote cases where no observed objects matched the area and intensity thresholds, hence a ratio cannot be determined.	68

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor Dr. Gretchen Mullendore for all the guidance, support, patience, and opportunities she has presented me throughout both my undergraduate and graduate careers. I would also like to thank my committee members, Dr. David Delene and Dr. Paul Kucera for all their constructive feedback, and giving me opportunities to travel and work with them.

I also want to thank my family for always believing in me, and my friends who have made everyday fun and exciting.

I dedicate this thesis to my parents, Krystyna and Jozef Starzec. Because of the love, caring, and support they have always given me, I know I can achieve anything.

ABSTRACT

Forecast verification remains a crucial component of improving model forecasts, but still remains a challenge to perform. An objective method is developed to verify simulated reflectivity against radar reflectivity at a 1 km altitude utilizing the Method for Object-based Diagnostic Evaluation (MODE) Tool. Comparing the reflectivity field allows for an instantaneous view of what is occurring in simulations without any averaging that may occur when analyzing fields such as accumulated precipitation. The objective method is applied to high resolution 3 km and 1 km local convective WRF summertime forecasts in the Northern Plains region. The bulk verification statistics reveal that forecasts generate too many objects, over-forecast the areal coverage of convection, and over-intensify convection. No noteworthy increases in skill are found when increasing to 1 km resolution and instead lead to a significant over-forecasting of small cells.

A sensitivity study is performed to investigate the forecast biases found by varying the cloud droplet concentration, microphysical scheme, and horizontal resolution on a case day containing weakly forced convection mostly below the freezing level. Changing the cloud droplet concentration has a strong impact on the number of object and area biases. Increasing droplet counts to observed values generates a forecast that more closely resembles the observations in terms of area and object counts, but leads not enough rain generation. Changing the microphysical scheme produces the most pronounced effects on object counts and intensity, which is attributed to differences in autoconversion formulations. Coarsening the resolution from 3 km to 9 km leads to a decrease in skill, showing that 3 km simulations are more effective

at convective forecasts. Increasing the resolution to 1 km results in amplifying the object count bias, and is found to not be worth the additional computational expense.

CHAPTER 1

INTRODUCTION

Numerical weather prediction (NWP) models continue to rapidly evolve and expand as technological advances and new ideas come forward. Significant advances in computer technology allow models to run continuously at high resolutions of 4 km and higher at the regional and even continental United States (CONUS) scales. Advanced high-resolution convection-resolving models, such as the High-Resolution Rapid Refresh (HRRR; e.g. Alexander et al., 2010), have become instrumental in accurate forecasts of summertime convection with lead times of a few hours to a few days. These high-resolution forecasts are invaluable for predicting convective events, aiding forecasters in determining locations of surface features, and providing guidance on potential timing of initiation, convective morphology, and storm hazards. In addition to finer model grid resolution, increased computational power and resources allow for more complex parameterizations and finer scale processes, such as chemistry, to be included. However, continuous increases in model complexity also increase the chances for model inconsistencies to occur. Releases of the Weather Research and Forecasting (WRF; Skamarock et al., 2008) model now occur on an annual basis, with new code, new physics, and updates to older code being added consistently, but there is limited evaluation is conducted to see whether there is an actual increase in forecast skill or performance. It is important to step back, assess, and evaluate forecasts generated by WRF and other NWP models to see how much value these advances are adding. Model evaluations should determine whether the model changes

are even addressing the primary sources of forecast error and not adding more biases to the model.

Forecast verification is a crucial component in the development of more accurate models. Hence, forecasters have subjectively verified forecasts since the first model forecasts were generated, in order to evaluate performance and determine where potential errors exist. Subjective forecast verification is an extremely valuable way of evaluating a forecast since it allows for a person to select regions of interest and focus on an in-depth analysis of multiple fields that influence that region. However, it is impractical and inefficient when trying to find long term systematic model biases or evaluate simulations across extended periods of time. With the popularity of ensemble forecasts rising, subjective verification becomes unfeasible as there are too many forecasts to evaluate continuously. Instead, automated objective verification methods allow for processing of large amounts of data with no subjective biases, and if any biases do exist due to how the object method operates, they are systematic. Traditional objective verification (i.e. skill scores) that perform point-to-point or grid-to-point comparison, are appropriate for fields like temperature, dew point and winds. However, point-to-point and grid-to-point comparisons are not suitable for evaluating high-resolution convective forecasts (Mass et al., 2002) since the chaos within the boundary layer and atmosphere in general (which is unresolvable) can cause discrepancies in timing and location.

This study assesses the Advanced Research WRF (ARW) model forecasts of summertime convection using an object-based verification technique that accounts for spatial differences. The object-based verification method allows forecast skill in morphology, areal coverage and convective magnitude to be determined for two summers of convective forecasts. A statistical analysis of the forecast errors is performed with the ultimate goal of determining to what extent different aspects of the model

(e.g. microphysics, resolution) contribute to the overall error in convective forecasts. Understanding the cause of the largest sources of error details what parts of the WRF model require more focus for future improvements and enables developers to effectively improve the WRF model.

CHAPTER 2

BACKGROUND

2.1 Convective Verification

Convective forecast verification plays a crucial role in advancement of NWP models but meaningful evaluations remain challenging. A form of verification that accounts for spatial differences is essentially the only way to verify convection since traditional skill scores have limited usefulness when spatial anomalies exist between forecasts and observations. It is well known that the exact location of convective initiation, and therefore convection, is nearly impossible to predict due to the chaotic nature of the boundary layer. Consequently, traditional grid point-to-point verification cannot be applied and often, a subjectively good forecast would be deemed objectively bad. An example of subjectively good forecast is shown in Figure 1. Although the forecast over-strengthened and over-extended certain regions of the convective system and slightly shifted the location of the feature, the structure, morphology and vicinity of the system is well represented by a 24-hr forecast. Several point-to-point verification methods are applied to the forecast. The True Skill Statistic (Hanssen-Kuiper skill score; Hanssen and Kuipers, 1965) gave the forecast the best skill of just below 0.3 (where 1 is a perfect forecast), while the other methods gave the forecast effectively no skill (skill scores below 0.1). The verification instead should be more focused on the intensity, morphology and temporal distribution of convection and less on the precise location.

Several different objective verification methods have been developed and tested in an effort to account for the variability encountered with convective forecasts (i.e. timing, location, morphology). Gilleland et al. (2009) gives a review of 15 different modern methods, which include neighborhood evaluation (e.g. Roberts, 2005), field deformation (e.g. Keil and Craig, 2007), scale separation (e.g. Casati et al., 2004), and object-based methods (e.g. Ebert and McBride, 2000). Neighborhood methods focus on applying spatial filters by smoothing down the field of interest to different resolution scales to generate meaningful statistics. While the smoothing increases with decreasing resolution, the field stills contains the same general features. By utilizing single-band spatial filters, scale separation techniques attempt to coarsen the data fields and present verification statistics on the scales that had the most skillful forecasts in attempt to best match the forecast and observation fields. Field deformation methods can be thought of as displacement methods. One data field is deformed by rotating, re-sizing, scaling and orienting it to best match the other data field. For field deformation methods, once the field has been transformed the remaining differences (i.e. forecast errors) produce lower forecast skill. Likewise, the amount of transformation and altering of data fields can be used as a source of skill. Object-based methods identify features, or regions of interest, which may be individual targets or groups of targets in both forecast and observation fields. Different forecasted individual objects, or object clusters, are matched to observed objects with similar attributes.

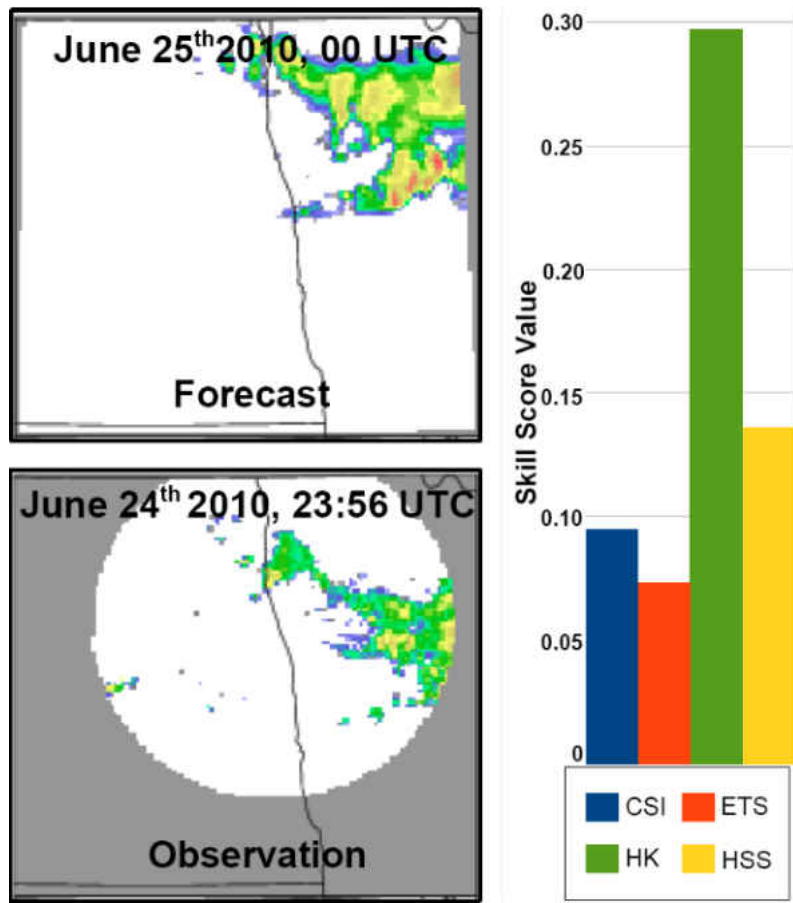


Figure 1: Simulated reflectivity (top), observed reflectivity (bottom), and corresponding traditional skill scores (right) where a perfect forecast is a 1. The traditional skill scores are the Critical Success Index (CSI), Equitable Threat Score (ETS), Hanssen Kuipers (HK) and Heidke Skill Score (HSS).

The verification conducted compares simulated reflectivity against radar reflectivity, with additional focus placed on individual cell information. The smoothing of neighborhood approaches would remove too much detail, especially with smaller and weaker convective cells. Current neighborhood and scale separation methods do not provide information on location errors or do not provide nearly as much information as object-based and field decomposition methods. Similarly, neighborhood and scale separation techniques also do not provide information on intensity errors and distributions, although certain scale decomposition methods (e.g. Marzban and

Sandgathe, 2006) can be thresholded to provide some information on intensity. While all verification techniques discussed give skill on the spatial structures and data distribution within the entire field, only object-based methods allow for spatial information on each individual object of interest, such as area, length, width and aspect ratio. In regards to timing, all techniques can be applied to different time periods or be manually designated to compare different time periods together. However, no technique automatically checks multiple forecasts against multiple observations in the same time frame (i.e. comparing multiple forecast fields across time to the same observation field) to look for time biases. Due to these reasons, focus is placed on object-based methods as they seem more versatile and satisfy the requirements of the study better.

According to Gilleland et al. (2009), the only methods that are fully object-based are the Contiguous Rain Area (CRA; Ebert and McBride, 2000), the Method for Object-based Diagnostic Evaluation (MODE; Davis et al., 2006a), Structure, Amplitude, and Location (SAL; Wernli et al., 2008), and Procrustes shape analysis (Micheas et al., 2007). However, only CRA, MODE and Procrustes give information on individual object structures and both CRA and MODE have been more widely used and evaluated (e.g. Gallus, 2010) than the other two methods. While both CRA and MODE can be utilized, CRA requires forecasted systems and observed systems to be contiguous while MODE clustering allows objects to be grouped together even if there is distance between objects, which is important when dealing with the no echo regions in the reflectivity fields. CRA also requires forecast and observed objects to have at least a little overlap in order to be matched. Although this matching method could be relaxed so no overlap is necessary, issues may arise when two or more forecast objects are near the observed object, leading to incorrect matches or no matches being chosen (Ebert and Gallus, 2009). MODE attempts to deal with this issue by apply

smoothing in the merging step. Tartaglione et al. (2005) recommended the CRA method be applied to domains that are much larger than the objects found within them. Having domains larger than forecasted objects ensures matching is correct and performed on complete systems. However, domain size is an issue that is present in all verification techniques, particularly object-based techniques, as objects may be caught on the edge of the domain leading to altered object attributes and incorrect matching. Following these reasons, MODE is chosen as the method for verifying convective forecasts.

Davis et al. (2006b) demonstrated the use of MODE by evaluating 4 km WRF forecasts of 1 hourly accumulated precipitation over the central portion of the United States. The Davis et al. (2006b) analysis found that compared to observations, forecasts generated too many large rain areas and rain systems lasted (1-2 hours) too long. Davis et al. (2006b) also found a large positive size and intensity bias in the Midwest. While looking at the impacts of grid spacing on precipitation forecasts, Johnson et al. (2013) found that both 1 km and 4 km simulations over-forecasted the amount of objects during the diurnal maximum with the 1 km forecasts performing worse at a 1-hr forecast lead time (i.e. 1-hr after initialization). However, Johnson et al.'s (2013) analysis showed that the 1 km forecasts predicted the average object area and average axis ratio more correctly than the 4 km forecasts. Other studies, such as one done by Johnson and Wang (2013), verified deterministic precipitation forecasts from an ensemble with differing model dynamics and physics. Fewer forecasted objects were found at a 1-hr lead time. However, after the 1-hr lead time, forecasted objects were more numerous, smaller in average area, more circular in average aspect ratio, and displaced to the east. Differing results between the Nonhydrostatic Mesoscale Model (NMM) and Advanced Research WRF (ARW) dynamical cores were also found. Generally, ARW object attributes were most similar to observations, while

the NMM had a higher rate of over-forecasting the number of objects than the ARW, had a higher precipitation bias, but performed better on average object area. Using a time-domain version of MODE (keeping track of objects across time), Clark et al. (2014) found that diurnal cycle shape was well represented by simulations. However, forecasts initiated too many objects during the first few hours and likely dissipated most of those objects after 3-5 hours. A slow bias was also found in simulations, although the authors hypothesized that this was due to bias introduced by the data assimilation used. Due to the extreme versatility of object-based methods, they have been used to identify and classify extreme precipitation events (Hitchens et al., 2012), to determine spread between ensemble members created by varying the physics and lateral boundary conditions (Gallus, 2010), and to assess the predictability of convective initiation (Duda and Gallus, 2013; Burghardt et al., 2014). The aforementioned studies show that MODE has a practical application to the simulated and observed reflectivity fields.

The majority of the convective verification studies have been performed on precipitation fields accumulated over an hour or more. While the accumulated precipitation field does provide information on convective track, coverage, morphology, magnitude and timing, time-averaged fields lose detail of what was occurring with the convective system at an instantaneous time (especially true of weak forcing events). Verifying the accumulated precipitation field diagnoses bias in the precipitation field, but the convective systems need to be evaluated to understand the cause of the precipitation bias.

2.2 Model Sensitivity

An added challenge to the process of forecast verification is that the model itself is systematically changing. Characterization of error becomes difficult as many

aspects of the model (i.e. cumulus parameterizations, planetary boundary layer, microphysics, etc.) can influence a single element. Due to the complicated interaction between model physics, several previous sensitivity studies investigated the effects of various model aspects, such as microphysical schemes, cumulus parameterizations and resolution. Weisman et al. (2008) performed a sensitivity study using 4 km WRF forecasts and 12 km Eta model (Black, 1994) forecasts by varying the microphysical schemes, planetary boundary layer (PBL) schemes and resolutions. Weisman et al.'s (2008) evaluation showed that none of the parameters tested accounted for the most significant forecast errors, although they discovered a systematic bias in the Yonsei University (YSU; Hong et al., 2006) PBL scheme. Weisman et al. (2008) stated that increasing the resolution did help the convective elements resemble more realistic features and an improvement in the diurnal cycle of convection was seen.

Bryan and Morrison (2012) compared simulations of a squall line with observations taken during the Second Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX2) campaign. The squall line simulations were created using both single and double-moment versions based off the Morrison et al. (2005a) microphysical scheme and using different horizontal grid spacing (4, 1, and 0.25 km). Bryan and Morrison (2012) found that simulated squall lines are sensitive to both the microphysics and the horizontal resolution; however, rainfall was more sensitive to resolution. Single-moment simulations generated less precipitation, cold pools that were too strong, and failed to generate stratiform precipitation near the surface due to too extreme evaporation rates. Further analysis revealed that smaller grid-spacing resulted in more evaporation due to more resolved turbulence, while the coarser resolutions caused the convective system to develop more slowly resulting in more larger convective cores and more precipitation. Bryan and Morrison (2012) noted that simulations that included microphysics incorporating graupel generated unrealistically

large convective areas. Comparing 4 km and 2 km simulated reflectivity fields, Kain et al. (2008) found that although 2 km forecasts provided more detailed depictions of convection, there was little to no added forecast skill as evaluation metrics were similar between resolutions. If there was any improvement in skill, it was found to not be worth the computational costs of increased resolution as model simulated reflectivity in both the 4 km and 2 km simulations tended to look more like each other than the observed reflectivity field.

Since the mid-1990s, there have been several studies performed to determine the resolution needed to correctly represent convective-scale features numerically. Grid spacing of 1 km has been considered in the past as a standard or target resolution for accurate representations of the atmosphere at convective-scale. However, whether any skill is added by the increased resolution remains questionable. 1 km grid spacing was shown to be able to replicate the formation, structure and evolution of convection better than coarser resolutions, thereby improving explicitly resolved convection (e.g. Weisman et al., 1997; Petch et al., 2002; Bryan et al., 2003; Petch, 2006). These results were especially true of organized linear features such as squall lines (e.g. Skamarock et al., 1994; Weisman et al., 1997; Bryan et al., 2003; Bryan and Morrison, 2012) and supercells (e.g. Droegemeier et al., 1994; Alderman and Droegemeier, 2002; Fiori et al., 2011). However, Petch (2002; 2006) argued that resolutions below 1 km are needed to accurately resolve convection due to differential surface heating. Boundary layer turbulence and eddies generated by differential heating need to be resolved to accurately illustrate convective initiation. Droegemeier (1994) also stated that to marginally resolve smaller cell-scale features that occur, you would need resolutions of 500 m or smaller. Using an idealized model to evaluate the effects of resolution, Verrelle et al. (2014) found that 4 km resolution clearly failed to resolve convective motions. Verrelle et al. (2014) stated that differences between 2 km and 1

km resolution were much greater than the difference between 1 km and 500 m spacing, causing them to suggest the possible beginning of converging solutions around 500 m. However, Bryan et al. (2003) concluded that simulations using 250 m and 125 m resolution still did not converge and 1 km grid spacing should not be used as a benchmark.

After a significant number of studies, complete understanding of the effects of resolution are still uncertain. Another question that arises is whether current parameterizations are still valid for such high-resolution simulations. Microphysical processes are tied to resolution changes as they are influenced and forced by resolved features. A resolution and microphysical sensitivity study performed on a simulated squall line by Bryan and Morrison (2012) showed that decreasing the grid spacing led to higher evaporation rates of cloud water, resulting in a 10-30% decrease in total surface precipitation when the grid spacing was reduced from 1 km to 250 m. In some cases, sub-grid scale features that were parameterized are now in the gray-zone of resolution, whereas certain processes are partially resolved and partially parameterized. Due to the nature of this problem, fields such as aggregated precipitation may be double counted as convective cores are partially explicitly resolved while convective parameterizations also generate precipitation. Deng and Stauffer (2006) found that for 4 km convective-resolving model runs, updrafts and downdrafts became the size of grid resolution (which are too large), resulting in over-predicting the amount of rainfall and increasing the evaporation rates. This increase in evaporation cools the air, causing stronger, more unrealistic downdrafts, disrupting lower level flows.

2.3 Microphysical Properties

Current microphysical schemes attempt to accurately represent several fine-scale atmospheric phenomena. Assumptions within microphysical schemes are re-

quired, but certain assumptions may only be representative of one location or season and can reduce forecast skill in other regions. One major assumption made throughout most single-moment schemes is a constant cloud droplet concentration (CDC). In the WRF Single-Moment 6-class (WSM6) scheme (Hong and Lim, 2006), for example, a constant cloud droplet concentration of 300 cm^{-3} is assumed, but in reality, the number of cloud droplets has a wide range depending on several environmental factors such as specific humidity and aerosol concentrations (e.g. Leaitch, 1992). The number of cloud droplets has a well-known relationship with the amount of cloud condensation nuclei (CCN; e.g. Twomey and Squires, 1959; Jiusto, 1966; Warner, 1969b; Hudson, 1980). Delene et al. (2011) found significant differences in the instantaneous number concentrations of CCN in eastern North Dakota (same region as used in this study) during the summer season, ranging from low values of around 300 cm^{-3} to concentrations over 3000 cm^{-3} , with significant fluctuations in daily averages. A constant cloud droplet assumption may adversely affect convective development, especially spatially and temporally, and may affect both the strength of the convection and the amount of precipitation generated.

The CCN concentration plays an important role in the development of convective clouds. Low amounts of CCN, especially larger CCN, may result in rapid droplet growth, as plentiful water vapor is readily available, quickly creating large cloud drops and enhancing the warm rain process preceding a precipitation event (Rosenfeld, 1999). A quick precipitation event may prevent the upscale growth and evolution of convection as rapid generation of precipitation may produce hydrometeor loading on the updraft that is too significant, effectively collapsing the updraft and washing-out the cloud. Fewer, larger drops would also allow for incoming shortwave radiation to be more easily transmitted between cloud drops, which reduces absorption and, therefore, evaporation of drops. On the other hand, higher CCN amounts

can result in delayed droplet growth as CCN compete for water vapor, delaying the precipitation process and changing cloud dynamics (Freud et al., 2011). However, using a triple moment bulk hail scheme, Loftus and Cotton (2014) found that by increasing the CCN concentrations from 100 to 3000 cm^{-3} led to largely unaffected storm dynamics and evolution, but they did find that increases in CCN increased hail size. The radiative properties of the cloud are also affected as higher CCN counts reduce droplet sizes but increase cloud droplet concentrations (Twomey, 1974). The increased cloud droplet concentrations allow for more absorption of shortwave radiation that leads to more evaporation in the cloud (Tao et al., 2007). Rosenfeld and Woodley (2000) found that polluted air with high observed CCN concentrations not only suppressed the warm rain process but also enhanced the cold rain process. The findings suggest that smaller cloud drops, corresponding to high aerosol concentrations, suppress collision and coalescence because there is limited differential growth. However, the updraft may more easily loft the smaller drops above the freezing level which results in more freezing of drops and the releasing of more latent heat. The latent heat release reinforces the updraft by making the air more buoyant. Rosenfeld and Woodley (2000) found that these processes helped enhance hail growth and the cold rain process, as did similar studies (e.g. Andreae et al., 2004; Lin et al., 2006), hence having a direct impact on convective strength. Rosenfeld's (1999) investigation found that extremely high CCN concentrations found in smoke from forest fires and biomass-burning actually stopped the warm rain process in tropical clouds, and enabled precipitation to occur only when cloud tops reached around -10°C .

The objectives of this study are to utilize MODE to spatially verify WRF forecasts of the simulated reflectivity field. Since the reflectivity field is not averaged, verifying the reflectivity field allows for an instantaneous look into what is occurring in the model. The results of the quantitative verification are evaluated to look for

potential biases in forecasts. The biases found are analyzed by performed a sensitivity study in order to determine what potentially generates or contributes to the biases.

CHAPTER 3

DATA

The data being used originates from the Polarimetric Cloud Analysis and Seeding Test (POLCAST; Kucera et al., 2008; Delene et al., 2011). POLCAST is an ongoing research project in eastern North Dakota, focused on evaluating the effectiveness of hygroscopic seeding flares in summertime convection. The POLCAST field campaigns were held during the years of 2006, 2008, 2010, and 2012. Starting in 2010 (POLCAST3), local 3 km resolution WRF (v3.1.1) model runs were used to predict the timing, intensity, and distribution of convection in the study region. The model grid included a 27 km parent domain nested with two-way 9 km and 3 km grids (Fig. 2a), and 45 vertical levels. Forecasts for POLCAST3 were generated from June 21 to July 23, 2010. In 2012 (POLCAST4), 1 km resolution model runs were added to forecast convection in the domain. Therefore, two independent model realizations were used: the original 3 km from POLCAST3 with an adjusted grid ((Fig. 2b) and the new 1 km resolutions runs (Fig. 2d). During POLCAST4, the 3 km simulations operated from June 27 to August 6, 2012. The 1 km simulations were operational starting July 8, but it was not until July 26 where the full 1 km domain was used. From July 8 to July 25, the 1 km domain was half its original size (Fig. 2c), with only the northern portion of the domain generating forecasts. The physics options used for the operational forecasts during all the POLCAST field campaigns were unchanged and are shown in Table 1. The Kain-Fritsch (Kain, 2004) cumulus parameterization scheme was utilized for the 27 km and 9 km grids only,

as convection was allowed to be resolved in the higher resolution domains. Model cold-start initialization occurred at 00 UTC everyday using three-hourly NAM 40-km data for lateral boundary conditions, with forecasts running for 24-hrs (00 UTC next day) and output being generated hourly.

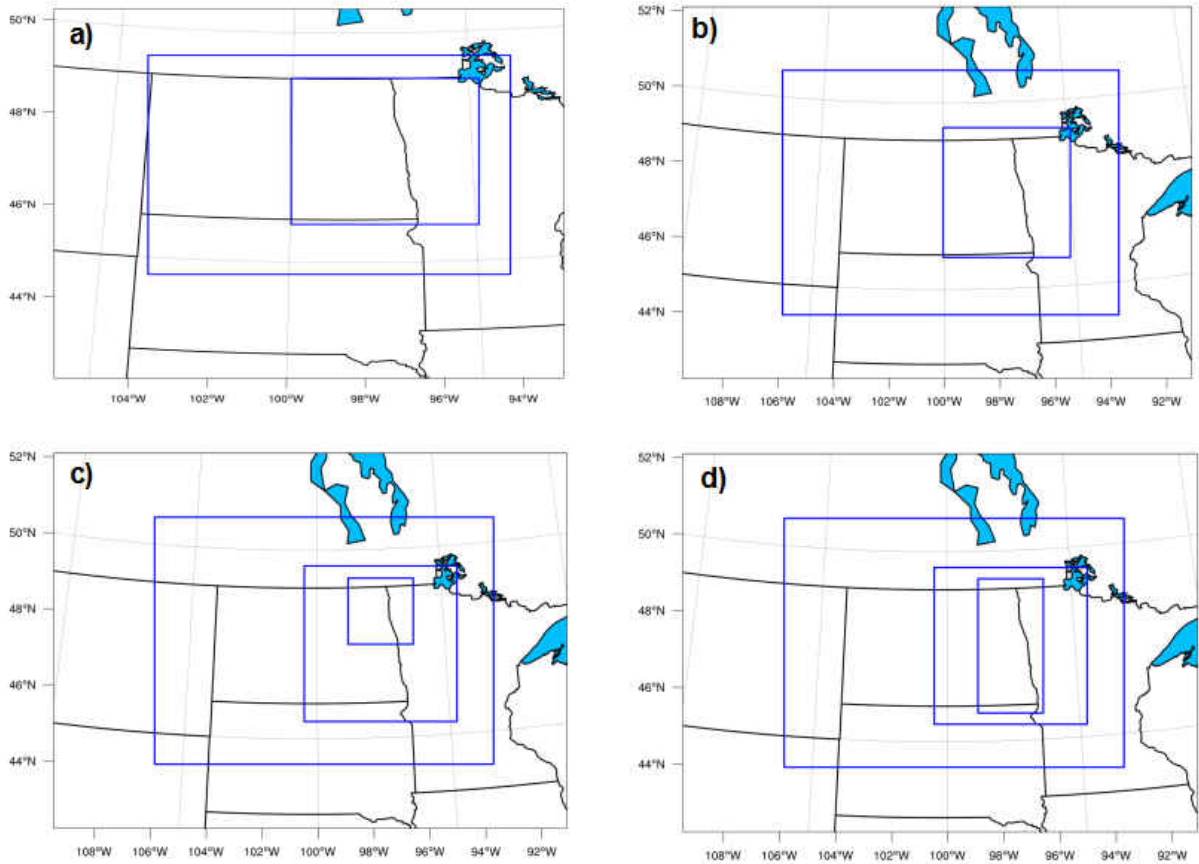


Figure 2: Domains used for a) POLCAST3 3 km, b) POLCAST4 3 km, and c) & d) POLCAST4 1 km simulations where the blue boxes indicate nested domains within the parent grid (outer black box).

Table 1: Physics used for the third and fourth Polarimetric Cloud Analysis and Seeding Test (POLCAST) campaign Weather Research and Forecasting (WRF) model runs.

Model Physics	
Microphysics	WSM6
PBL	YSU
Surface Layer	MM5 Similarity
Land Surface	Noah
Shortwave	Dudhia
Longwave	RRTM
CP	Kain-Fritsch

The University of North Dakotas (UND’s) polarimetric C-band Doppler radar1 actively scanned the study region over the operating range of 150 km during times of expected or ongoing convection. The reflectivity data were quality controlled by utilizing dual polarization data to eliminate noise and unwanted artifacts by performing several parameter tests. Data with low correlation coefficient (<0.6) below 4 km in height were removed. A range dependent metric using reflectivity, correlation coefficient, and velocity data removed instances of ground clutter and anomalous propagation. Neighborhood checks on the data made sure noise was not included in the domain. Reflectivity returns were checked to ensure they have vertical extent and echo tops. The reflectivity data also underwent bias correction in the quality control process.

During the POLCAST4 field campaign, four cloud condensation nuclei (CCN) counters, two from the University of Wyoming (UWyo) (SN 107 and 112) and two from Droplet Measurement Technologies (DMT) (SN 072 and 062), sampled the air at the UND radar site and at cloud base via the UND Citation II aircraft. CCN concentrations at three different supersaturations of 0.2%, 0.3%, and 0.6% were measured by the DMT counter and at a constant supersaturation of 0.6% by the UWyo counter.

To be able to measure the CCN concentrations across several supersaturations, the DMT counter cycled through each supersaturation, sampling for five minutes at 0.2%, three minutes at 0.3%, and four minutes at 0.6%. Since it takes a period of time for the conditions in the chamber to adjust and stabilize to the new supersaturation, the last 30 seconds of concentration data at each supersaturation were averaged to produce the ambient CCN concentration at that supersaturation and time.

CHAPTER 4

METHODOLOGY

4.1 Preprocessing for Forecast Verification

Forecasts are spatially verified by comparing the simulated reflectivity field derived from the WRF model at a 1 km height (above ground level, AGL) to the radar reflectivity field at 1 km height AGL constant altitude plan position indicator (CAPPI) derived using the native resolution UND radar volume scan data. Evaluation at 1 km is chosen due to the extensive use of the simulated reflectivity field by forecasters at that level. The reflectivity at a constant height also reveals more detail than fields such as composite reflectivity. Since the radar beam at the lowest elevation angle is above the 1 km height at further ranges, the reflectivity data are interpolated down to 1 km AGL from the lowest beam, using a vertical profile algorithm. The simulated reflectivity field is calculated using rain, snow, and graupel mixing ratios following Koch et al. (2005) and subsequently interpolating the results from model levels to 1 km AGL height grid. The verification is performed for every 00 UTC forecast, for each valid hour starting at 03 UTC, unless no radar data exists, which results in that hour being skipped. Verification is performed starting at 03 UTC, to allow three hours for the simulation to spin-up. The 2 km resolution radar data closest to the forecast valid time (within 30 minutes) is used, and is interpolated to the 3 km model grid. The radar data is interpolated to the model grid, since the 3 km model resolution is theoretically the smallest object size that can be present in the forecast field, and thus cannot generate objects 2 km in size. For the 1 km

simulations, simulated reflectivity is calculated on the 1 km grid but is interpolated to 3 km resolution in order to have a direct comparison against 3 km simulations.

Since the 3 km model domains are also larger and cover more area than the region covered by the radar scans, the model domains are masked to include only the area covered by the 150 km radar range in order to have a direct comparison between forecasts and observations (Fig. 3). During times where the radar performed sector scans due to targets of interest to the POLCAST campaign, the forecast is also masked to only include the sectors that contained data. In 1 km simulations, the forecast domain is smaller than the radar range in several locations but extends further than the radar range in others. Any region that does not have both model and radar data is masked out.

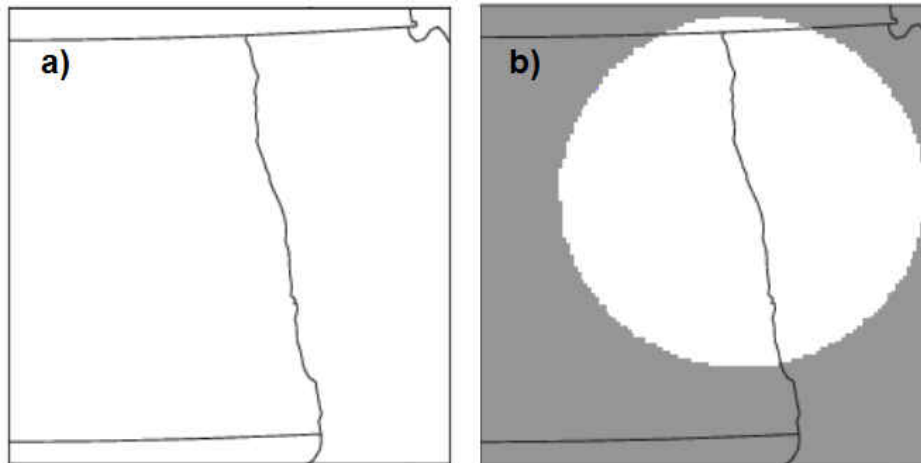


Figure 3: An example of the masking procedure used when comparing forecasts to observations, where the white denotes areas that contains data used and gray denotes a masked region, where no data is available. Panel a) denotes the original model domain and b) denotes the model domain after being masked to the radar range.

4.2 Convective Object Verification

Forecast performance is assessed by utilizing the MODE Tool (v4.1), a part of the Model Evaluation Toolkit (MET) package (<http://www.dtcenter.org/met/users/>).

MODE is an object-based verification approach designed to compare gridded forecasts with gridded observations. MODE operates in four steps, specifically the raw, convolved, mask, and objects steps (Fig. 4), for both the forecast and observation fields. First, the raw data field is processed and plotted on the analysis grid (Fig. 4a). A circular convolution is applied to the raw data field (Fig. 4b), which is essentially an application of a smoothing operator or gradient on the data. Any data that are adjacent to each other and share the same characteristics (e.g. orientation, axis angle), or that intersect due to the convolution application can be merged, or clustered together using a fuzzy logic method. In the masking step (Fig. 4c), the data is filtered by applying different thresholds, such as area or intensity thresholds, which masks any undesired values. Finally, the raw data is placed back into the convolved and masked regions (Fig. 4d), and objects are generated. Object characteristics such as centroid, area, median intensity, aspect ratio, and axis angle are determined. MODE then takes each object and cluster from the forecast field and attempts to match it with the objects and clusters in the observation field, calculating statistics on each pair. Statistics calculated by MODE include traditional skill scores and a total interest value. Total interest is generated on all object matches, which is based on the similarities between object characteristics and matched pair statistics such as intersection area, centroid distance, and convex hull distance. When total interest values are above a set threshold (default is set to 0.7, where 1 is a perfect match), objects and clusters with the best total interest become matched.

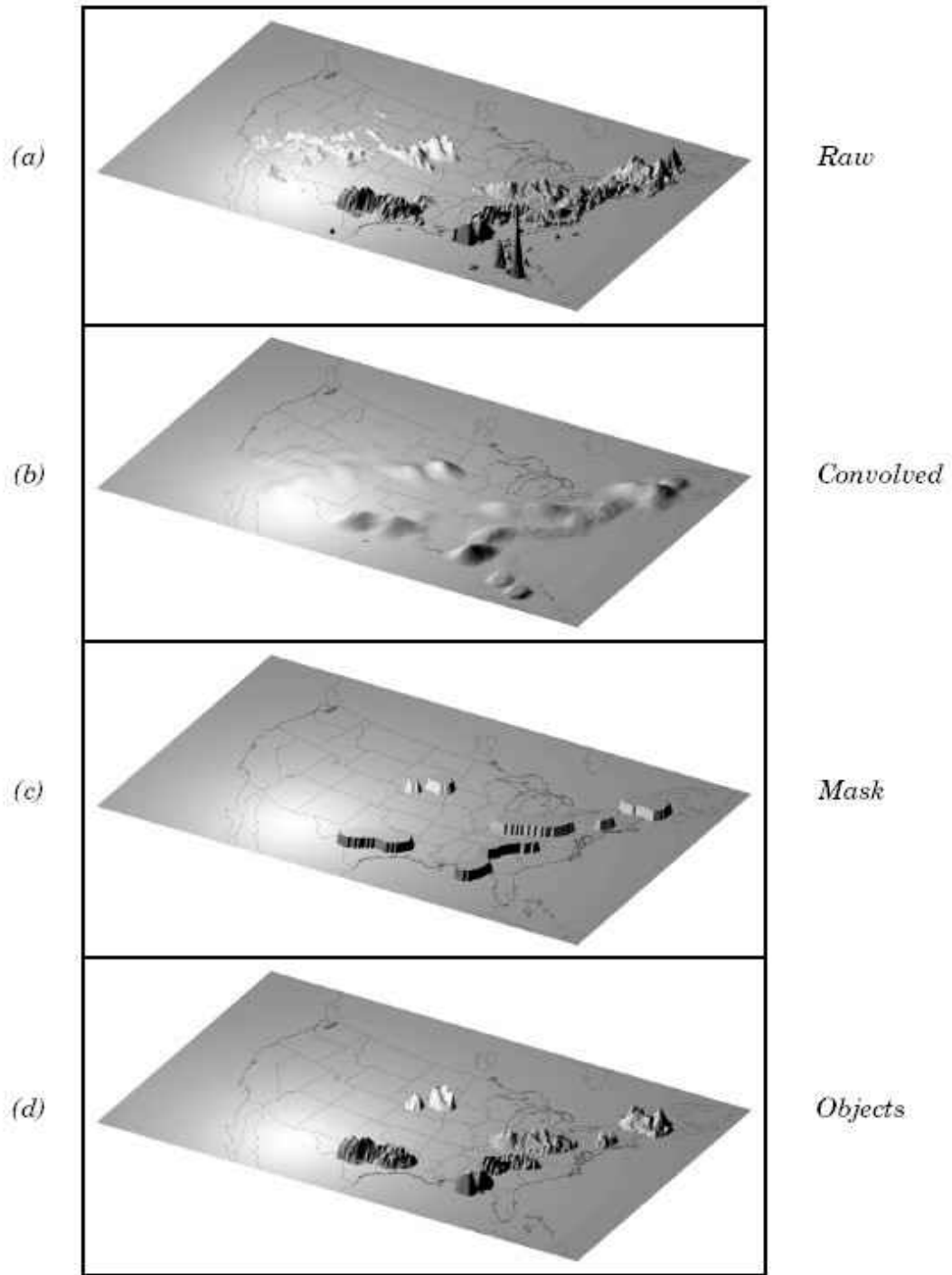


Figure 4: The following figure details how the Method for Object-based Diagnostic Evaluation (MODE) tool analyzes data. MODE begins with the (a) raw data field, (b) applies a circular convolution for object merging, (c) masks data and applies any thresholds, and (d) declares objects [Image from Davis et al., 2006a].

MODE was originally created to generate skill for quantitative precipitation forecasts. To be able to utilize MODE for reflectivity comparisons, usage of the precipitation-combine (PCP-Combine) tool, another tool part of the MET package, is initially required. The purpose of PCP-Combine is to sum accumulated precipitation data from multiple model output files in a single input file. The PCP-Combine tool is used to generate input files for MODE, but instead of ingesting precipitation, the MODE work-flow is altered to allow reflectivity (both observed and simulated) to be used instead. Additionally, because MODE was originally created to verify accumulated precipitation, the forecast and observed fields would normally contain larger regions of interest that are smoothed out and continuous. In contrast to accumulated precipitation, the reflectivity field often contains objects with more widely varied sizes and morphologies, and with sharper magnitude gradients. Due to these issues, MODE falsely clusters and matches objects (Fig. 5) frequently, rendering the statistical information generated by MODE less meaningful as the automated clustering and matching process does not necessarily replicate how a person would subjectively cluster objects and match them across fields. While it is possible to cluster all or most objects together in both the forecast and observation fields and match them together by changing the configuration script in MODE, only one object would be generated in both fields. Statistics generated by comparing these single objects would be heavily influenced by amount and distribution of cells within the object. Lastly, to best merge objects, a convolution radius should be used to group individual cells together. However, by applying a convolution to the raw data, objects created will lose the original attribute information on each individual cell that existed in the raw data field. Many small cells would also be removed during convolution, with the objects left being larger stronger cells. Therefore, for the purposes of this study, MODE's automated statistical methods performed on matched objects are not used

due to issues with how MODE handles high temporal and spatial resolution data such as reflectivity at a certain height. Instead, the raw object attributes generated by MODE, with no convolution, merging, or matching, such as number of objects and each objects' area, are used for forecast verification.

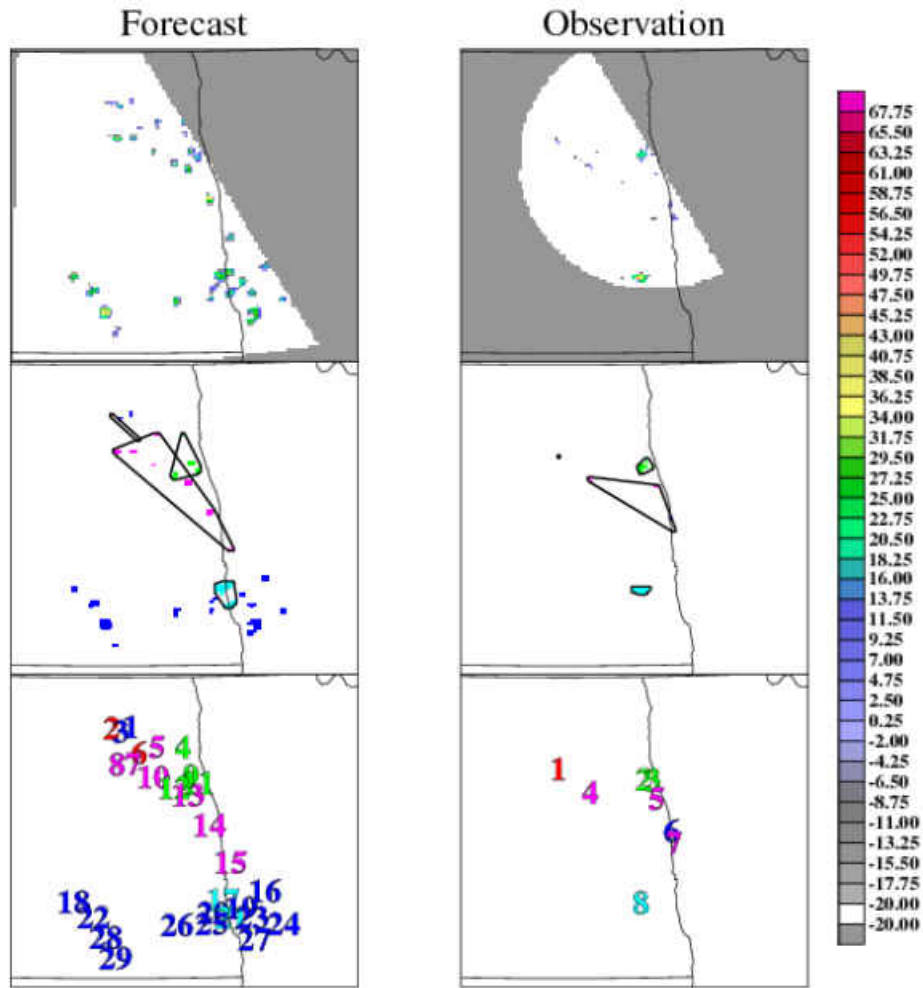


Figure 5: An example of unfavorable merging and matching of objects by MODE, with no convolution used in both the forecast (left column) and observation (right column) fields. The raw reflectivity field (top row), cluster and object field (middle row), and object number (bottom row) are shown. The Rubber-banded objects of the same color in the middle row represent merged objects, with similarly colored objects in both the forecast and observation columns indicating a match. Dark blue objects are objects that have not been merged or matched.

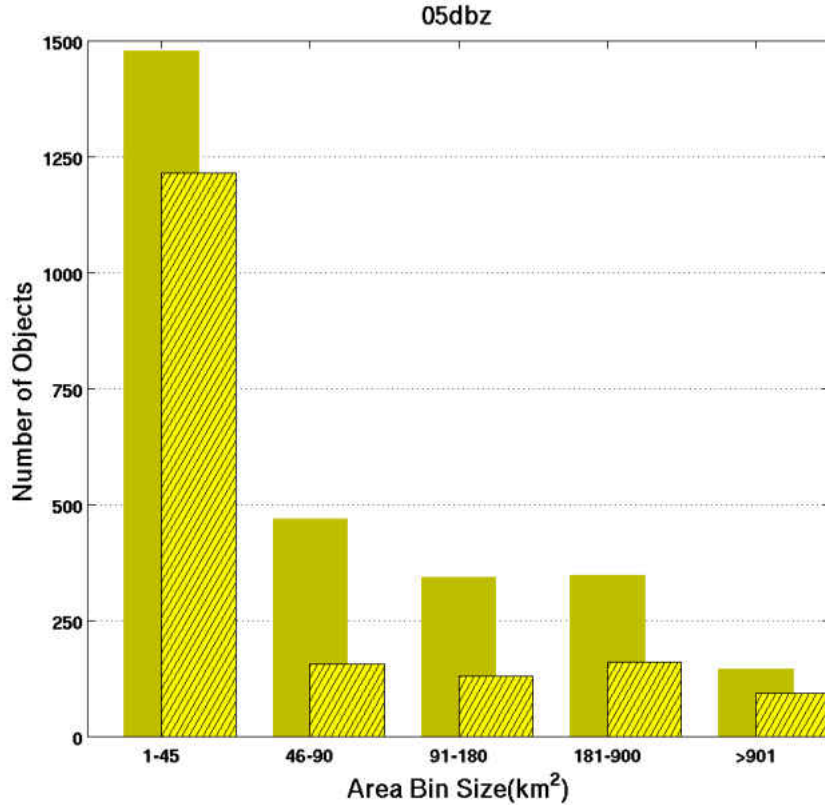


Figure 6: Number of all forecasted (darker) and all observed (light with diagonal lines) cells greater than 5 dBZ for the entire POLCAST3 time period (June 27 - Aug. 6, 2012), binned according to cell size.

Forecasts are assessed by looking at the hit, null, false alarm, and miss rates, dependent upon whether convection is present in both the forecast and observed domains. Three additional methods of verification used to assess skill of forecasts in this study are comparisons of convective morphology, areal coverage, and magnitude. Convective morphology is based upon whether or not the forecasts generate structures with the appropriate sizes. Each object in both domains is binned according to its area, which shows an overview of whether the correct morphology and size of convective systems is forecasted (Fig. 6). For example, if the forecast predicts a large linear feature, it would contain one large object, while if scattered multi-cells

were observed, the observations would contain several smaller objects. Areal coverage corresponds to how many grid squares of both the observed and forecasted fields are covered by convection. Coverage of convection is found for every hour that convection was present in both domains.

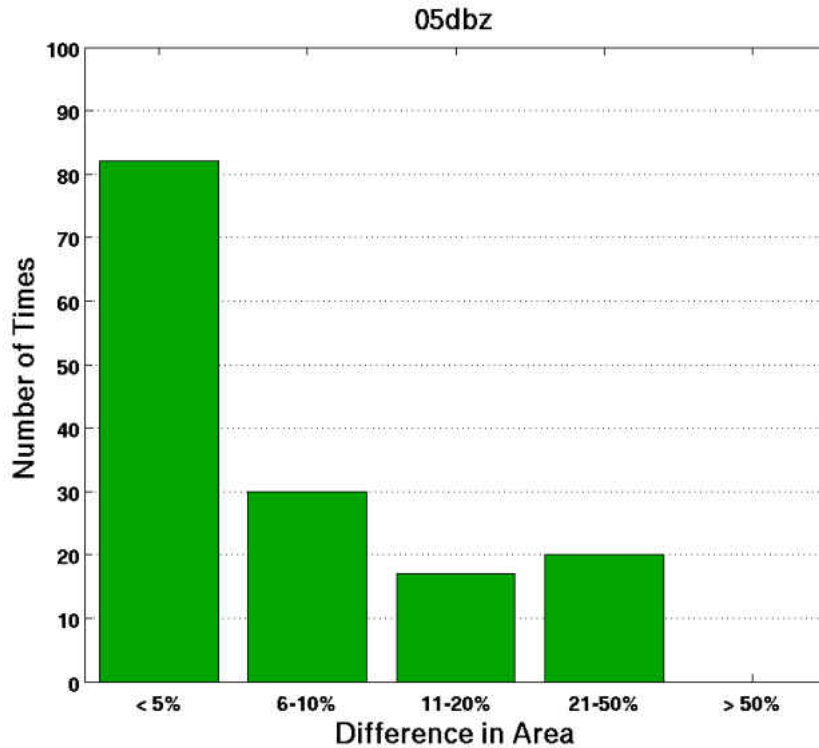


Figure 7: Number of cases and their corresponding differences between observed and forecasted convective areal coverage when convection existed in both domains across the entire POLCAST3 field campaign (June 22 - July 23, 2010)

By taking the summation of area covered by each object and dividing it by the total domain area, an areal coverage ratio is determined. Finding the difference between the observation and forecast ratios shows to what degree the forecast and observations differ in the areal extent of convection (Fig. 7). Aided by the convection morphology statistics, the areal coverage provides a good general estimate of forecast accuracy. To see how the forecasts perform at different magnitudes, various intensity

thresholds within MODE are used (5 dBZ, 30 dBZ, and 45 dBZ). Utilizing the morphology (object size) and areal coverage statistics at these differing thresholds, shows whether the WRF runs represent the correct magnitude of convection, and whether forecast skill improves or depreciates with increasing intensities.

CHAPTER 5

RESULTS

5.1 Campaign Overview

To get a sense of how representative both the POLCAST3 and POLCAST4 campaign years are relative to climatological average, both years are compared to 33 years of precipitation (Fig. 8) and temperature (Fig. 9) data. A 33-year span is selected as the representative period to give enough time to capture yearly variations yet still be considered recent climatology, without biasing to past climates. The 30-year period of 1981 to 2010 is also currently being utilized as the climate normal period by the National Weather Service.

The POLCAST4 (2012) year is drier (17.12 in) than the climate average of 20.31 in. Excluding July, where precipitation amounts are around the mean, precipitation is lacking from May until about September, affecting the POLCAST campaign. In contrast, 2010 is the wettest year on record (since 1890), with annual rainfall of 27.77 in. The mean monthly average temperatures (Fig. 9) show that January through August of 2012 is much warmer than the climate average. In terms of temperature, 2010 is close to the average yearly mean, while 2012 is the second warmest in the 33-year period, and over 1.5 standard deviations from the mean. Hence, the temperature in 2010 is near the climatological average, but the precipitation is much above average (outside two standard deviations from the mean), while 2012 is characterized as being drier and warmer than average. Comparing the operational months of June, July, and August, 2010 and 2012 received 10.31 and 7.74 in of rain, respectively,

which is relatively close to the average of 9.34 in. July is warmer for both years, especially in 2012. Since differing environmental conditions are prevalent across the POLCAST3 and POLCAST4 campaigns, the years are analyzed separately to ensure temporal robustness of the verification.

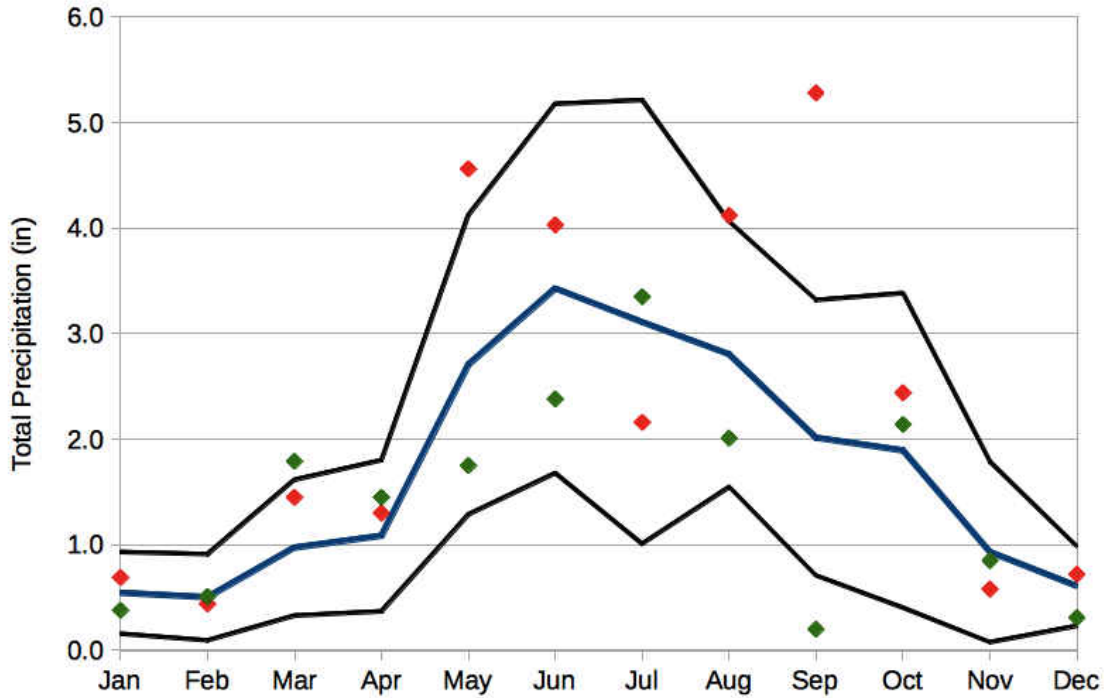


Figure 8: The mean monthly precipitation (blue line) and standard deviation (black lines) of 33 years of precipitation data starting from 1981 for Grand Forks, ND airport (KGFK). The red diamonds indicate the monthly precipitation during the 2010 POLCAST3 campaign year and the green diamonds represent the monthly precipitation during the 2012 POLCAST4 campaign year.

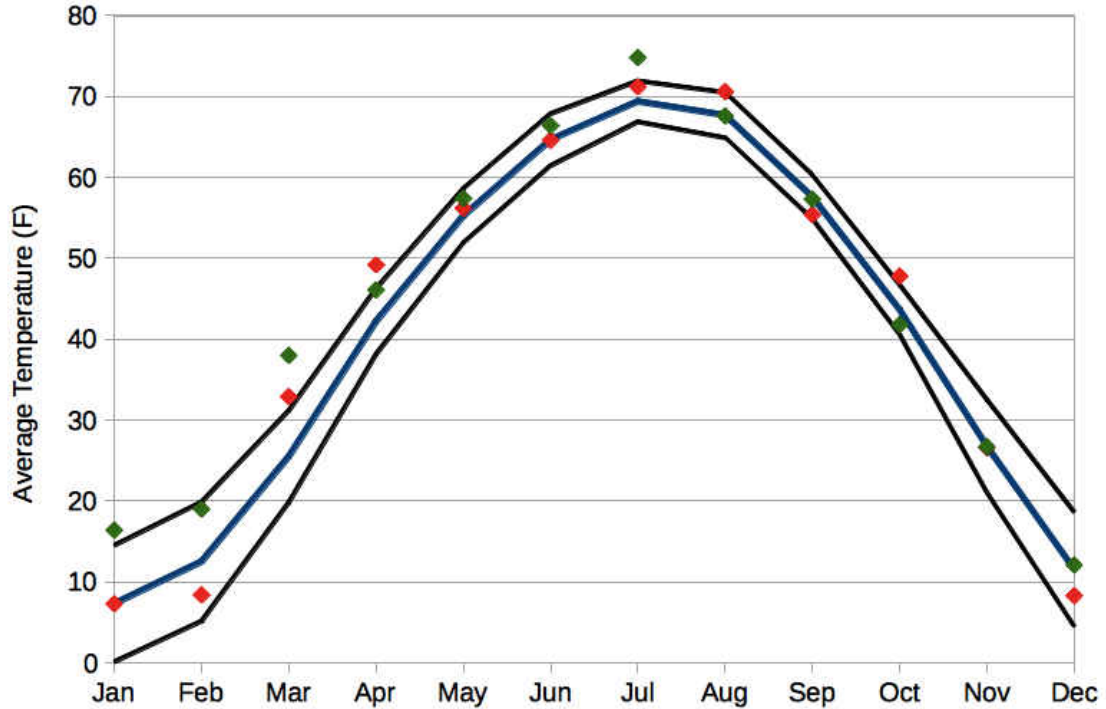


Figure 9: The mean monthly average temperature (blue line) and associated standard deviation (black lines) of 33 years of temperature data starting from 1981 for Grand Forks, ND airport (KGFK). The red diamonds indicate the monthly average temperature during the 2010 POLCAST3 campaign year and the green diamonds represent the monthly average temperature during the 2012 POLCAST4 campaign year.

5.2 Object Verification and Bias Summary of Forecasted Convection

Across the entire POLCAST3 3 km, POLCAST4 3 km, and POLCAST4 1 km (henceforth denoted as P3, P4a, and P4b) forecast periods, there are 276, 197, and 172 times across 28, 21, and 19 days respectively, where radar data matches with a forecast valid time. In all simulation periods, forecasts (solid bars) over-predict the number of convective cells as compared to observations (hashed bars) across all size bins and for all magnitude thresholds (Figs. 10, 11, 12). Table 2 shows that the best overall performance in terms of forecast to observation (FO ratio) is by the P4a realizations at the 30 dBZ threshold, but even then the over-forecast rate for object counts is

23%. All remaining simulations, regardless of intensity or campaign, generate more than 50% of the radar objects. The large FO ratios above the reflectivity threshold of 45 dBZ exist because only a few radar objects match that criteria while several such model objects exist.

A general improvement in forecasts is visible for the 30 dBZ reflectivity threshold compared to the 5 dBZ threshold (Table 2). The general improvement indicates that forecasts contain too many weak cells. However, all forecasts have significantly less skill at the 45 dBZ threshold, which indicates forecasts also over-intensify convection, as only a small amount of radar objects exist at the higher reflectivity threshold requirement. In all, there are 365, 350, and 244 model objects above the 45 dBZ threshold for P3, P4a, and P4b, but there are only 113, 64, and 28 radar objects, respectively. In addition, across both POLCAST campaigns there is only one radar object greater than the 45 dBZ threshold and greater than 181 km², while there are 90 such model (3 km) objects, which indicates again that the model over-intensifies convection.

A standard hit, miss, and false alarm rate histogram, based on whether any convection is present in either domain is shown in Figure 13. A hit case means convection exists in both the forecast and observation domains, while a null case signifies that neither domain contains convection. At a reflectivity threshold of 5 dBZ, success rates (hits plus null cases; blue colors in Fig. 13) of 61.0%, 67.3%, and 65.5% are visible for P3, P4a, and P4b, respectively. Figure 13 shows that P3 has a higher miss rate (red) of 32.2%, almost double the miss rate of P4a (18.4%) and P4b (17.0%), while the false alarm rate (yellow) for P3 (6.8%) is less than half the rate of 14.3% for P4a and 17.5% for P4b. The higher miss rate and lower false alarm rate in the P3 campaign is caused by radar artifacts (i.e. clutter, noise) being present

in the domain that failed to be removed during the radar quality control procedure (discussed further in Section 5.4).

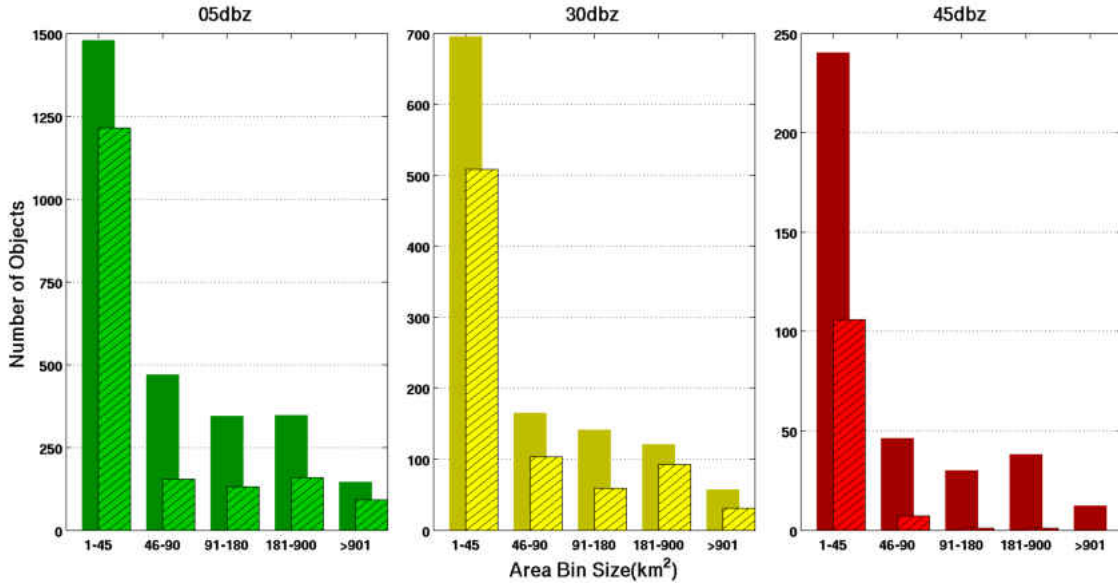


Figure 10: The number of forecasted (darker) and observed (lighter with diagonal lines) objects generated by MODE binned according to area across all P3 simulations at reflectivity thresholds of 5 (green), 30 (yellow), and 45 dBZ (red)

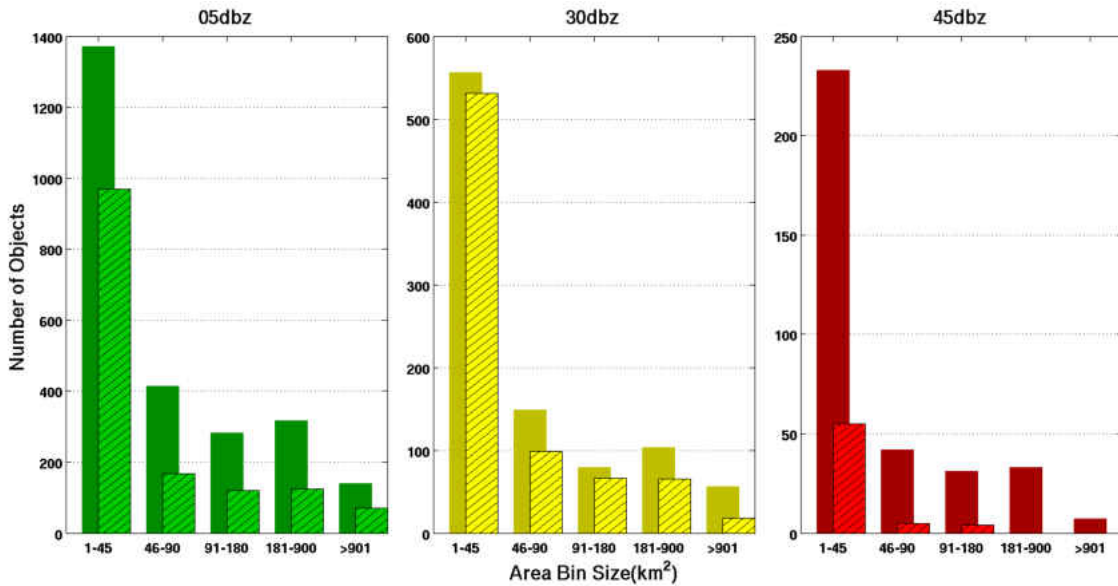


Figure 11: Similar to Figure 10 except for P4a simulations.

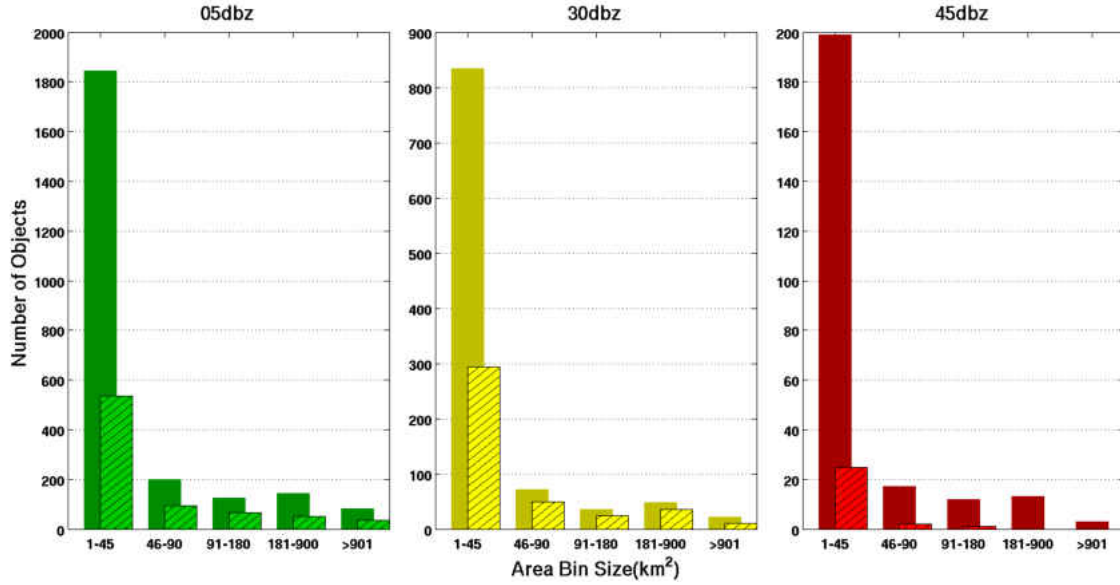


Figure 12: Similar to Figure 10 except for P4b simulations.

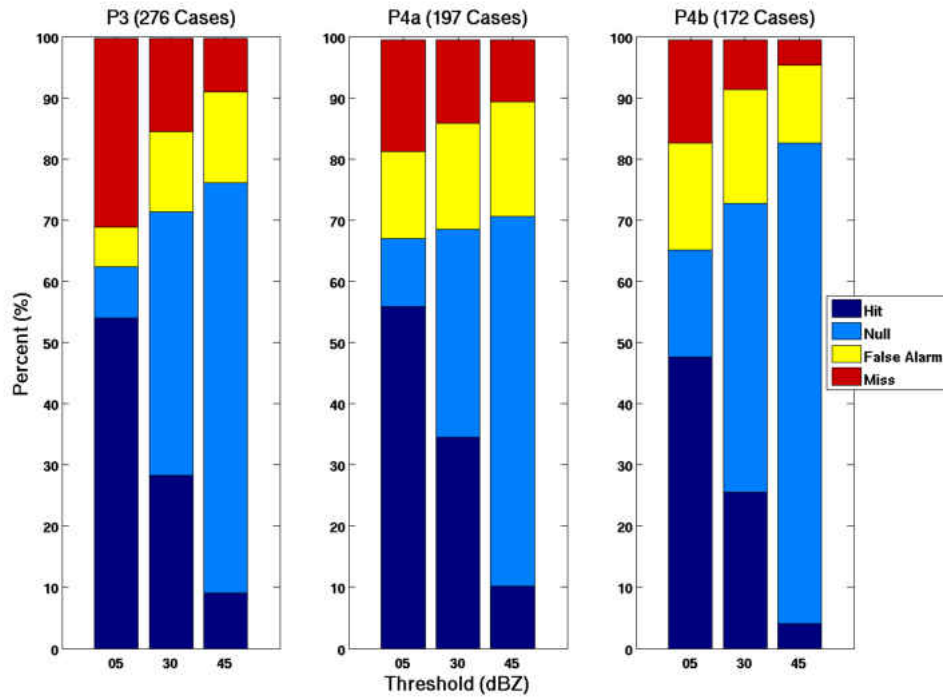


Figure 13: The percentage of cases across the P3, P4a, and P4b simulations that were either hits (dark blue), null cases (light blue), false alarms (yellow), or misses (red) at differing intensity thresholds of 5, 30, and 45 dBZ.

Table 2: Ratios of the number of forecasted objects across all POLCAST3 3 km (P3), POLCAST4 3 km (P4a), and POLCAST4 1 km (P4b) simulations to the number of observed objects by UND radar, binned by object sizes corresponding to Figures 10, 11, and 12. "-" denote cases where no observed objects matched the area and intensity thresholds, hence a ratio cannot be determined.

Threshold	Simulation	1-45 km ²	46-90 km ²	91-180 km ²	181-900 km ²	>900 km ²	Total	Total (Without 1-45 km ²)
5 dBZ	P3	1.19	2.98	2.59	2.24	1.64	1.57	2.44
	P4a	1.41	2.43	2.40	2.58	1.94	1.74	2.39
	P4b	3.44	2.12	1.94	2.84	2.28	3.06	2.24
30 dBZ	P3	1.41	1.66	2.44	1.42	1.84	1.53	1.73
	P4a	1.08	1.46	1.24	1.57	3.17	1.23	1.56
	P4b	2.84	1.44	1.40	1.37	2.20	2.44	1.48
45 dBZ	P3	2.30	6.57	30.00	38.00	-	3.23	14.00
	P4a	4.24	8.80	8.25	-	-	5.47	13.00
	P4b	7.96	8.50	12.00	-	-	8.71	15.00

As the reflectivity thresholds increase to 30 and 45 dBZ, the hit rate drops and the null rate increases as expected, since lower reflectivity values (cells) are now being masked, which leads to an increase in success rates (Fig. 13). The miss rate is lower at the 30 dBZ threshold than at the 5 dBZ threshold and is the lowest at the 45 dBZ threshold for all campaign simulations. However, all false alarm rates increase when increasing the threshold to 30 dBZ, and all but the P4b false alarm rates increase when the magnitude is greater than the 45 dBZ threshold. At the 5 dBZ threshold, misses contribute the most to the lack of skill in forecasts while for both the 30 and 45 dBZ thresholds, false alarms dominate the miss rate with about two false alarms for every miss at intensities of 45 dBZ and greater. The increase in false alarms signifies that simulated convective magnitudes are too strong, as model convective cores remain while convection in the radar domain is not present because most convective cores are weaker than 45 dBZ.

For the differences in areal coverage ratios between model convective area and radar convective area when convection exists in both domains (Fig. 14), P3 and P4b forecasts perform relatively well, while P4a has the least skill. Over all cases, 75% (P3), 66% (P4a), and 77% (P4b) of all forecasts are within 10% areal coverage of observations at the 5 dBZ (green) threshold, increasing to 91%, 90%, and 96% at the 30 dBZ (yellow) threshold. At the 45 dBZ (red) reflectivity threshold, all forecasts, except one case in P3, are within 10% area coverage of observations, which is expected since at such high magnitudes only smaller and stronger convective cores remain. Only P4a has any cases with areal differences greater than 50% (Fig. 14b). The large differences in area are due to objects being cut-off at the edge of the domain when the simulation is not correct temporally (Fig. 15). In Figure 15, the model is too slow in generating convection, and the convective region is just entering the forecast domain while the post-convective stratiform region is covering the observation domain. Even

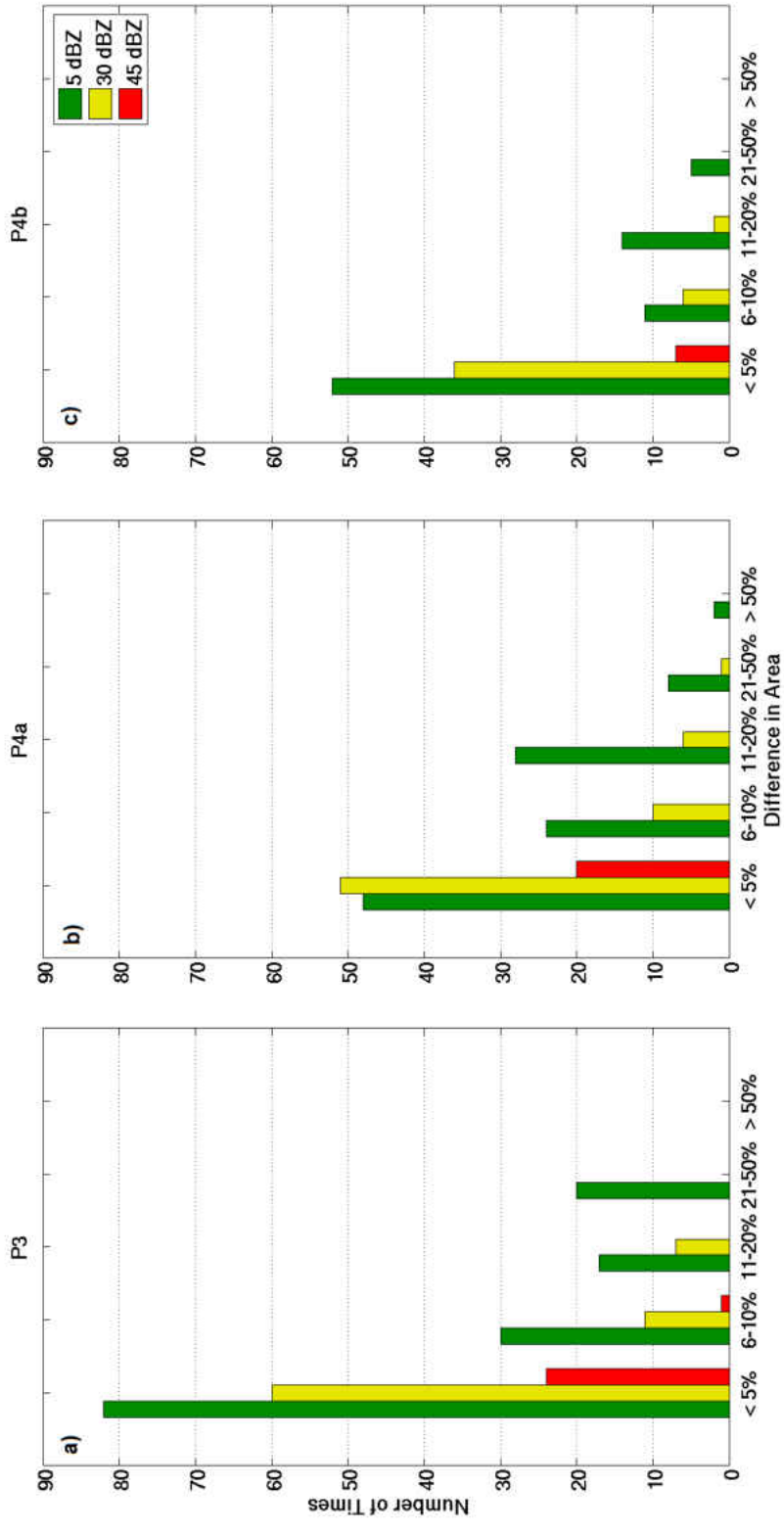


Figure 14: Differences between the forecasted and observed total area covered by convection relative to the domain size for P3 simulations at 5 (green), 30 (yellow), and 45 dBZ (red).

so, almost a quarter of all P4a simulations have area differences between 11-20%; or differences in coverage of around 7,700 to 14,000 km² during a full radar scan.

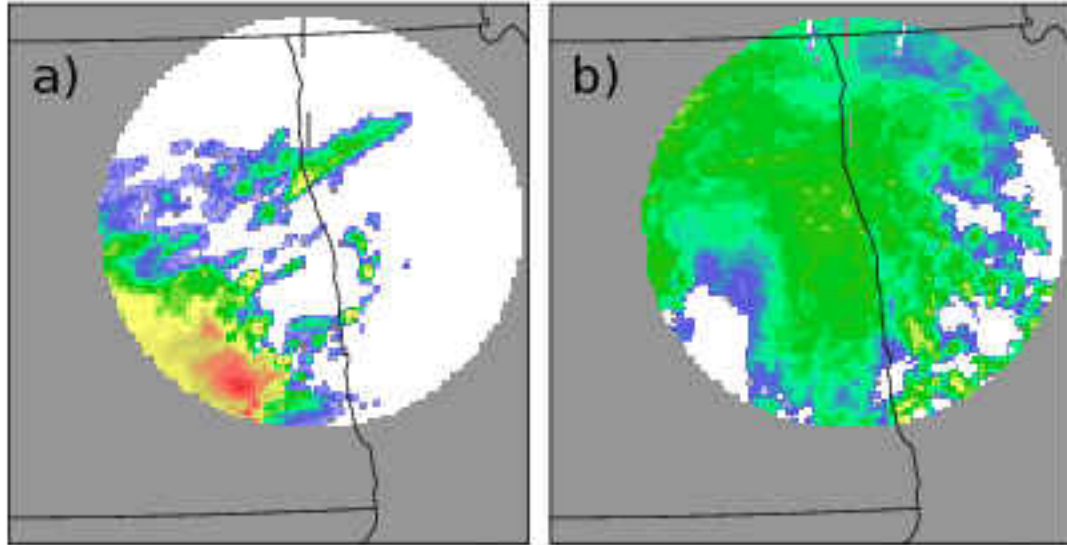


Figure 15: MODE comparison of the 3 km a) simulated reflectivity and b) observed reflectivity fields at 1 km AGL altitude at 12 UTC on July 6, 2012.

Table 3 shows that the over-prediction of coverage by forecasts is dominant across all simulations. Figure 16 shows the percent of areal coverage in the forecast domain, relative to the amount of areal coverage in the observation domain at each matching time across the P3 campaign, at the 5 dBZ reflectivity threshold. For example, a value of 110% indicates that the model contains 10% more area than the corresponding matching observation. The biggest differences in percent area are seen when forecasts over-predict areal extent of convection (Fig. 16; green bars). When forecasts under-predict convective area, the differences between forecasts and observations are less drastic, except in the few instances where a difference in timing of a strongly forced event occurs as mentioned previously (Fig. 15).

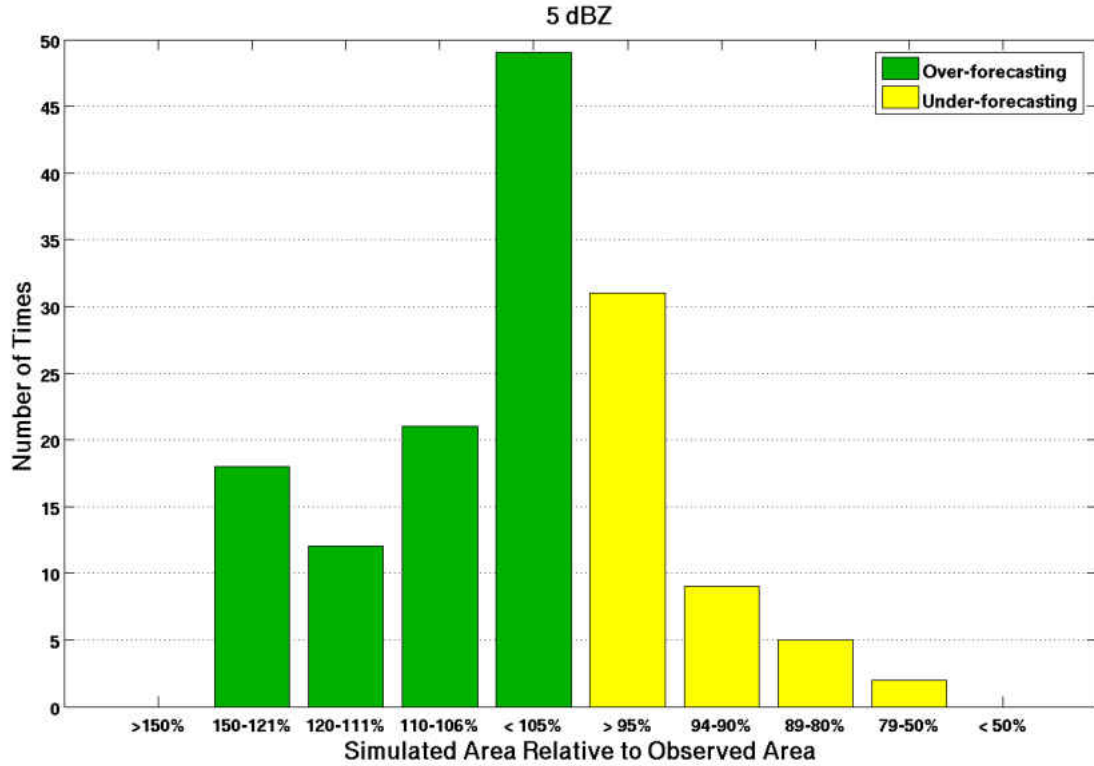


Figure 16: The amount of area covered by simulated convection relative to the amount of area covered by observed convection at each matching time, at 5 dBZ for all P3 simulations. Percentages greater than 100% (green) indicate over-forecasting of areal coverage, while values below 100% (yellow) indicate under-forecasting of coverage.

For all cases at the reflectivity threshold of 5 dBZ, the total area coverage by radar objects is about 530,00 km², 560,000 km², and 210,000 km², and the total area coverage by model objects is about 1,050,000 km², 970,000 km², and 410,000 km² for P3, P4a, and P4b, respectively. Although the total size of forecast convective areal coverage is approximately double the total size of the observation areal coverage, dividing the total object areas by the total object counts during matching times to get the average object size yields model objects that are very close to the size of radar objects (Table 4) at the 5 dBZ threshold for 3 km simulations (P3 and P4a). The average model and observed object areas are very similar at the 5 dBZ threshold; therefore, the over-forecasting convective of area is due to the over-forecasting of the

Table 3: The number of times when either the forecasts or observations covered more area when convection existed in both domains. The FO ratio represents the number of times forecasted convective area was greater than the observed convective area.

Threshold	Simulation	Forecasted	Observed	FO Ratio
5 dBZ	P3	96	42	2.29
	P4a	83	27	3.07
	P4b	64	16	4.00
30 dBZ	P3	50	25	2.00
	P4a	52	17	3.06
	P4b	32	12	2.67
45 dBZ	P3	21	3	7.00
	P4a	18	2	9.00
	P4b	6	1	6.00

number of model objects instead of over-extending the area of objects. However, the 30 dBZ and 45 dBZ thresholds show that individual object sizes are too large (Table 4). Another method to determine the average object sizes is to find the average object size in both the forecast and observation domains at each matching time. The differences between average object sizes at each matching time reveal that average model object sizes are greater than average radar object sizes 64% and 68% of the time for P3 and P4b, respectively. It obvious the model both generates too many objects and alternatively over-extends object areas. However, the over-extent of objects does not account for the majority of over-forecasting of convective area. Therefore, the over-forecasting of the number of objects leads to the greatest lack of skill at the 5 dBZ threshold, while at the 30 and 45 dBZ thresholds, model convection is too strong and generates higher reflectivity regions that are too large. On the other hand, the 1 km mean object size is significantly smaller than the observations at the 5 dBZ threshold due to the over-forecasting of the number of smaller cells (discussed in more detail in Section 5.3)). The over-forecasting of small objects also explains the misleading result of the 1 km model performing better at the 30 dBZ threshold in

terms of average object sizes (Table 4), since the large number of small cells skew the average size down.

In summary, the verification results show the following biases: simulations generate too many objects, simulations over-intensify convection, and simulations cover too much area by convection, which is due to the generation of too many objects at low intensities and over-intensification of convection at higher intensities.

5.3 Comparison of 3 km and 1 km Forecasts

Since both 1 km and 3 km simulations are conducted during the POLCAST4 campaign, it enables an assessment of differences in forecast skill at different model resolution. The 1 km simulations (P4b) significantly over-predict the number of objects (344% over-forecast rate) at the 5 dBZ threshold in the smallest bin (<45 km²; Table 2). In addition, even though P4b has 25 less matching times than P4a and the domain is smaller than in P4a, containing only 55% of P4a's radar cells, it generates 134% more cells than P4a (1843 compared to 1367) in the smallest bin. Subjectively analyzing a sample of these times indicates that the 1 km run is overly-convective and generates a vast number of small cells near boundaries or near other ongoing convective elements with some small cells having very high reflectivity values (e.g. Fig. 17). When the smallest size bin is discarded across all simulations, P4b simulations perform better as the total FO ratios result in better skill when compared to P3 and P4a at both 5 and 30 dBZ thresholds (Table 2; column 9). Similarly, the P4b simulations perform slightly better than both P3 and P4a in terms of areal coverage.

Table 4: The total number of objects, summation of all object areas (km^2), and the average area per object for all simulations and observations, when convection was present in both domains.

Threshold	Simulation	Total Area (km^2)	Total Objects	Avg. Obj. Area (km^2)	Obs. Total Area (km^2)	Obs. Total Objects	Obs. Avg. Obj. Area (km^2)
5 dBZ	P3	1046691	2335	448.26	530217	1273	416.51
	P4a	968832	2150	450.62	559953	1222	457.93
	P4b	414108	2030	203.99	209547	645	324.88
30 dBZ	P3	243657	928	262.56	129591	637	203.44
	P4a	216747	726	298.55	76887	683	112.57
	P4b	76716	743	103.25	37215	379	98.19
45 dBZ	P3	21924	170	128.96	1854	74	25.05
	P4a	22833	199	114.74	675	30	22.50
	P4b	10008	101	99.09	198	15	13.20

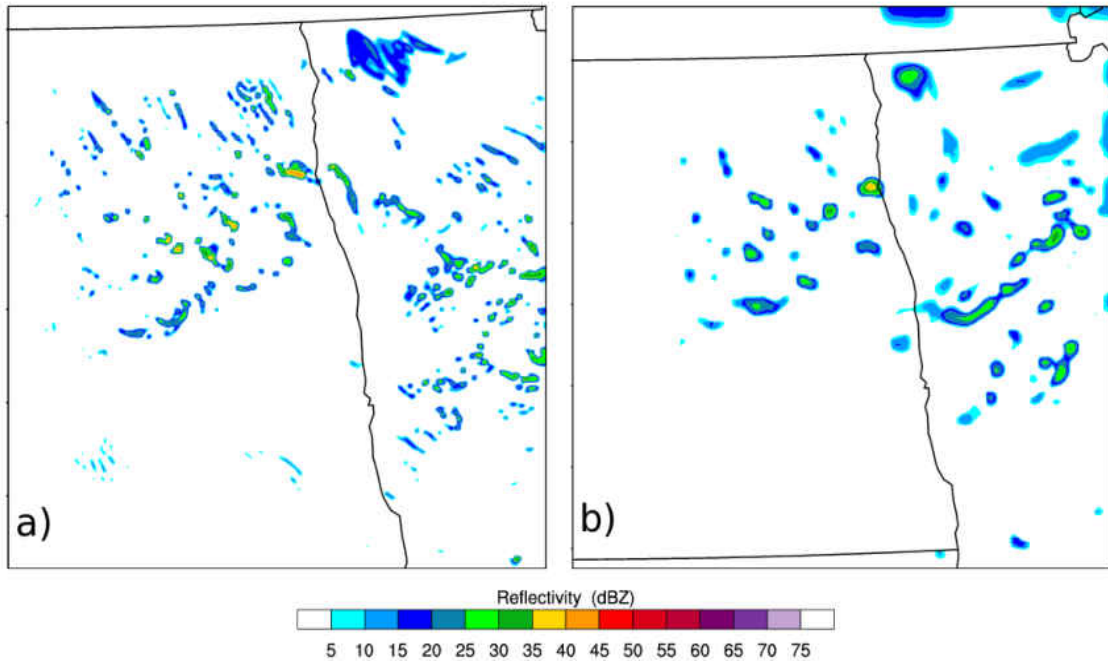


Figure 17: Simulated reflectivity field at 1-km height for the a) 1 km resolution and b) 3 km resolution simulations valid at 2230 UTC on Aug. 4, 2012.

To investigate whether the 1 km simulations actually did outperform the 3 km forecasts, the P4a 3 km domain is reduced in size to match the P4b 1 km domain at all times. Only the matching times that exist for both 3 km and 1 km runs are compared to ensure the domains and times are the same. The comparison resulted in 157 matches across 16 days. In terms of the number of objects, at the 5 dBZ reflectivity threshold, the 1 km run only performs better for objects sized 91-180 km², and performs worse than the 3 km for all other bins (Table 5), including the total FO ratios. At the 30 dBZ threshold, the 1 km performs better at the 46 to 900 km² sized objects; however, the overall improvement in those size bins is below 7%. At the 45 dBZ threshold, there is no improvement with increasing resolution. Without the smallest bin (Table 5; column 9), total FO ratios improve significantly for the 1 km; however, the 3 km actually loses skill in total ratios without the smallest

bin at all intensity thresholds, as the best performance occurs in those bins (except at 45 dBZ, where the 91-180 km² has slightly better skill).

In terms of area coverage, it is discussed previously that the 1 km runs (P4b) outperforms the 3 km runs (P3 and P4a; Fig. 13), with a larger percentage of cases falling within 10% difference in area coverage. However, when the P4a domain and times are restrained to that of the 1 km simulations, the 3 km runs slightly outperform the 1 km at all thresholds, showing that there is no skill added by 1 km simulations in coverage. For P4a, 79% and 96% of all times are within 10% areal coverage while P4b forecasts are within 10% coverage 78% and 95% of the time for the 5 and 30 dBZ thresholds, respectively. Likewise, the 3 km also has a slightly better success rate and less false alarms than the 1 km, although the 1 km did improve upon on misses (Table 6). Overall, the comparison results suggest no considerable improvement in skill (i.e. not worth the increased computational cost) when changing from 3 km to 1 km resolution based on comparing simulated reflectivity to observed reflectivity, matching the conclusions reached by Kain et al. (2008) and others.

5.4 Uncertainty Discussion of Objective Verification

A common problem with any verification is that objects become 'cut-off' by the edge of the analysis domain, as discussed in Section 2.1 and by Tartaglione et al. (2005). If a large mesoscale convective system (MCS) is present in the observation domain and the forecast is an hour behind, only a small piece of the simulated MCS is caught in the forecast domain during the comparison time, which leads to a big difference in area. The lag in time causing large differences in area is especially true of strong forcing events, where propagation speed is determined by the generation and advancement of frontal systems and outflow boundaries. A representative example of an object's area being misinterpreted due to being partially outside the domain is

Table 5: Same as Table 2, but for the P4b domain size and for analysis times that had both P4a and P4b model data.

Threshold	Simulation	1-45 km ²	46-90 km ²	91-180 km ²	181-900 km ²	>900 km ²	Total	Total (Without 1-45 km ²)
5 dBZ	P4a	1.04	1.92	2.21	2.85	2.22	1.43	2.25
	P4b	3.40	2.15	1.95	2.86	2.31	3.04	2.27
30 dBZ	P4a	0.85	1.55	1.25	1.27	1.80	1.02	1.42
	P4b	2.78	1.36	1.24	1.23	2.00	2.35	1.35
45 dBZ	P4a	4.85	8.50	4.50	-	-	5.67	9.75
	P4b	7.08	8.50	10.00	-	-	7.82	14.00

Table 6: The percentage of cases that were either successful in predicting convection (or lack thereof), produced false alarms, or missed observed convection, when both the 3-km (P4a) and 1- km (P4b) simulations were evaluated on matching domains and matching time periods.

Threshold	Simulation	Success (%)	False Alarm (%)	Miss (%)
5 dBZ	P4a	66.9	13.4	19.7
	P4b	66.7	15.4	17.9
30 dBZ	P4a	73.2	14.6	12.1
	P4b	72.4	18.6	9.0
45 dBZ	P4a	85.4	10.2	4.5
	P4b	82.1	13.5	4.5

seen during the P4a verification (Fig. 15), which leads to differences in area of over 50% (Fig. 14b; green).

Another potential source of error is in radar retrievals. While quality control is performed on the radar data, mitigation of non-meteorological artifacts (e.g., ground returns, insects, anomalous propagation) while ensuring that accurate data is left untouched is not a trivial problem. During certain scans, there is still a grid square or two of noise present in the domain that is not removed (Fig. 18). If the reflectivity value associated with the noise value is above the reflectivity thresholds, the noise is counted as an object (i.e. active convection). The implications are minor in terms of object counts and areas as the noise objects are small and not very numerous. However, the noise object will account for a forecast miss if noise is present in the observation domain when no convection is present in the forecast domain, which occurs during P3 at the 5 dBZ reflectivity threshold (Fig. 13). Similarly, if the forecast generates a false alarm when an artifact is present in the observed domain, it would be incorrectly counted as a hit and a difference in area would be computed. An additional potential source of error due to the radar retrievals is that the radar was not operating continuously during the POLCAST campaigns. The radar operated

almost exclusively during daytime hours, varying the hours of operation from day to day, when convection was possible or expected. The lack of radar coverage at certain times, or, coverage during the times of when convection was expected, may bias the results, especially the success rates.

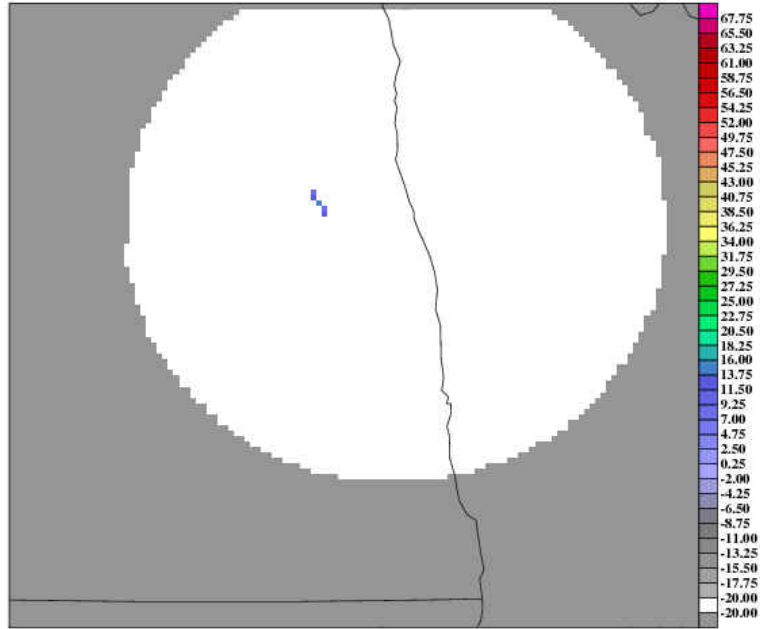


Figure 18: Observed reflectivity field at 1 km AGL height showing noise in the domain during a matching time in the P3 campaign.

Interpolating the radar data to the model resolution adds uncertainty; however, there is no other method to easily compare forecasts and observations. The 1 km resolution simulations also undergo interpolation to the 3 km model grid in order to have a direct comparison with 3 km forecasts. Johnson et al. (2013) 1 km runs against 4 km runs at their native resolutions, and also interpolated the 1 km grid to a 4 km grid in order to perform a comparison at the same grid spacing. Johnson et al. (2013) found that any improvements occur on the scales resolvable by the model, and small-scale improvements on the high-resolution grid do not increase skill when up-scaled to large-scale forecasts. Due to Johnson et al.'s (2013) findings, a

future verification comparing the 1 km against the 3 km can be performed without interpolation to allow for an additional comparison on native grids.

5.5 Case Study

5.5.1 Selection of Case Day

An investigation into what potentially yields the forecast biases in Section 5.2 is performed by sensitivity testing the horizontal resolution, microphysical scheme, and cloud droplet concentration. The goal of the sensitivity study is to understand the biases found and to find methods to improve forecasts. To accomplish the sensitivity study, a case day is chosen from the POLCAST campaign, representative of the biases found in the verification. The day has to contain at least two of forecast biases, as seen in the results section, prevalent across the chosen case day. The criteria for choosing the case study day is given below with some explanations also provided:

- over-forecasting the number of discrete cells by around 50% at any threshold,
 - Over-forecasting of objects of all sizes is present by both 3 km and 1 km simulations at every reflectivity threshold. At the 5 dBZ threshold, 3 km runs over-predict object counts by at least 57% whereas the P4b 1 km over-predicts by 306%, with significant over-forecasting of objects in the smallest bin).
- too many weak objects or over-intensifying of convection, and
 - There was an increase in skill when increasing the intensity threshold from 5 to 30 dBZ, showing model runs generate too many weak reflectivity regions. Simulations have the least skill at the 45 dBZ threshold, due to over-forecasting the number of convective objects.

- over-forecasting the area covered by convection (by either over-forecasting the number of objects or over-extending object areas).
 - During P3 and P4a, forecasts over-predict the areal coverage by convection 70% and 67% of the time, and overall, 75% and 66% of matching times were within 10% difference in areal coverage of the observed area.

The case day selected must also contain:

- simulated convection and observed convection occurring for at least six hours (capturing the entire event),
- radar coverage across the entire event and at least two hours prior to the convective event for a good temporal comparison,
- weakly-forced convection,
- no significant pre-existing boundaries and cold pools from possible convection earlier and lack of other strong forcing features in the domain that may initiate convection, and
- availability of cloud condensation nuclei (CCN) concentration data.

A weak forcing event is chosen to investigate the large over-prediction of smaller, weaker cells and their corresponding areas. Choosing a weak forcing event removes any location, and hence, additional timing issues potentially being caused by strong forcing events (e.g. strong cold front). Similarly, a component of the sensitivity study involves looking at the effects on forecast skill caused by changing the cloud droplet concentration (CDC) within microphysical schemes, influenced by the changes visible in in-situ CCN concentration data. A strong forcing feature will generate vigorous convection regardless of CCN concentration (although convective

features may differ), while weakly forced events should theoretically be more heavily influenced by changes in concentration.

Following the guidelines above, the case day of Aug. 4, 2012 is chosen for sensitivity testing. In terms of the biases found, the operational 3 km forecast generates peaks in model object counts that are similar (slight over-forecasting) to the radar object peaks; however, the simulation generates too many objects when all objects are summed up across the day. The total convective areal coverage is significantly greater than observations across the entire day, which indicates that individual object areas are too large. Likewise, the 1 km operational run over-predicts both the areal coverage and the number of objects in addition to generating cells over 45 dBZ in magnitude that are not present in observations.

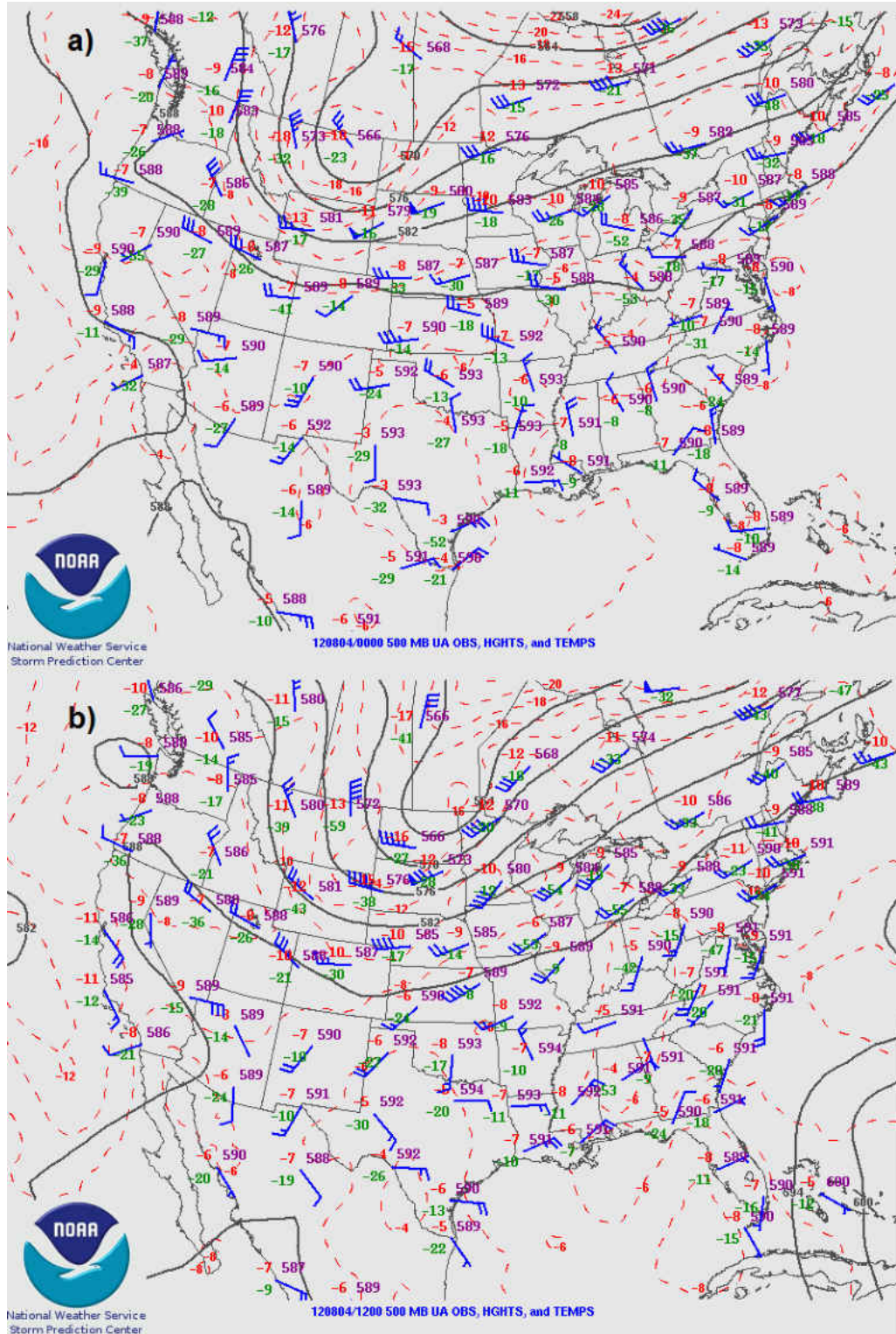


Figure 19: 500 mb upper air observations, heights (solid lines), and isotherms (dashed red lines) valid at a) 00 and b) 12 UTC on Aug. 4, 2012 [Images courtesy of the Storm Prediction Center].

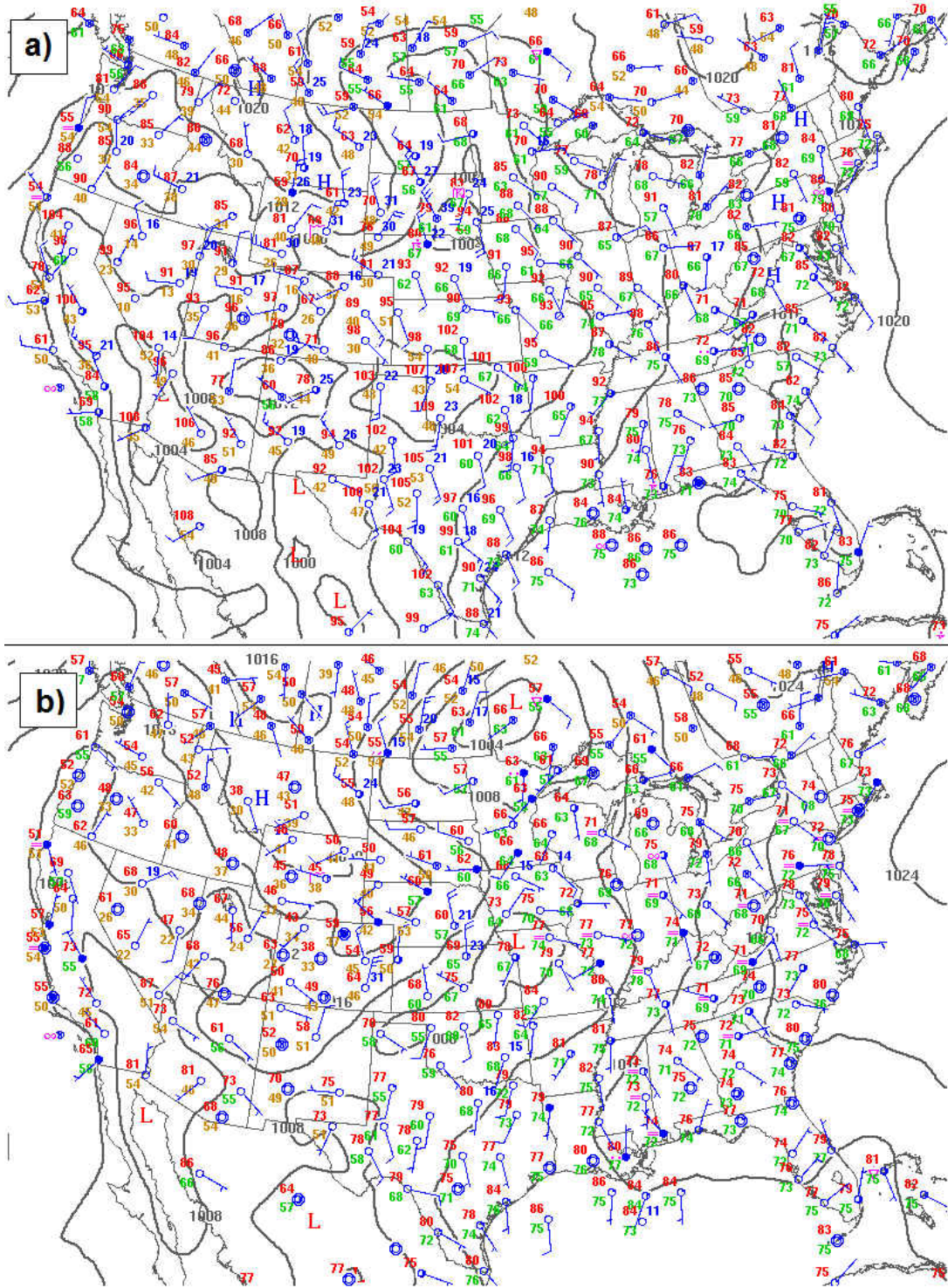


Figure 20: Surface observations and isobars (solid lines) valid at a) 00 and b) 12 UTC on Aug. 4, 2012 [Images courtesy of the Storm Prediction Center].

A positively-tilted shortwave trough was located in north-central Montana on Aug. 3, moving eastward towards the southern provinces of Saskatchewan and Manitoba, and North Dakota (Fig. 19). Plentiful diurnal heating mixed with low-level warm air and moisture advection, favorable shear profiles, and steep mid-level lapse rates generated an atmosphere favorable for deep convection. An associated surface low-pressure system developed and propagated toward south-central Manitoba with a strong cold front stretching down across the Dakotas (Fig. 20), acting as the primary area for convective development during the afternoon hours of Aug. 3. Around 19 UTC, a broken line of deep convective cells entered the domain and propagated eastward. By 23 UTC (Fig. 21a), the convection became more organized and began evolving into a continuous squall line along the cold front located in the center of the study domain. The squall line developed several bowing segments as it completely left the domain after 03 UTC (Fig. 21b). As the central low-pressure region moved northeastward, a trough developed on the backside of the system, generating weaker scattered (<45 dBZ) convective cells, which are the main focus of the sensitivity study. The scattered convection propagated into the domain between 14 and 15 UTC, with new convective initiation in the domain. At 18 UTC, the target convection was in the center of the domain (Fig. 21c). The last cells exited the domain around 00 UTC on Aug. 5. While there was a strong squall line progressing through the domain in the very start of the day, the scattered cells between 14 and 00 UTC were the main focus of the sensitivity study. There were no significant boundaries or cold pools visible in radar, in satellite, or surface data, so any boundaries present earlier were assumed to have been mixed away due to the strong frontal forcing.

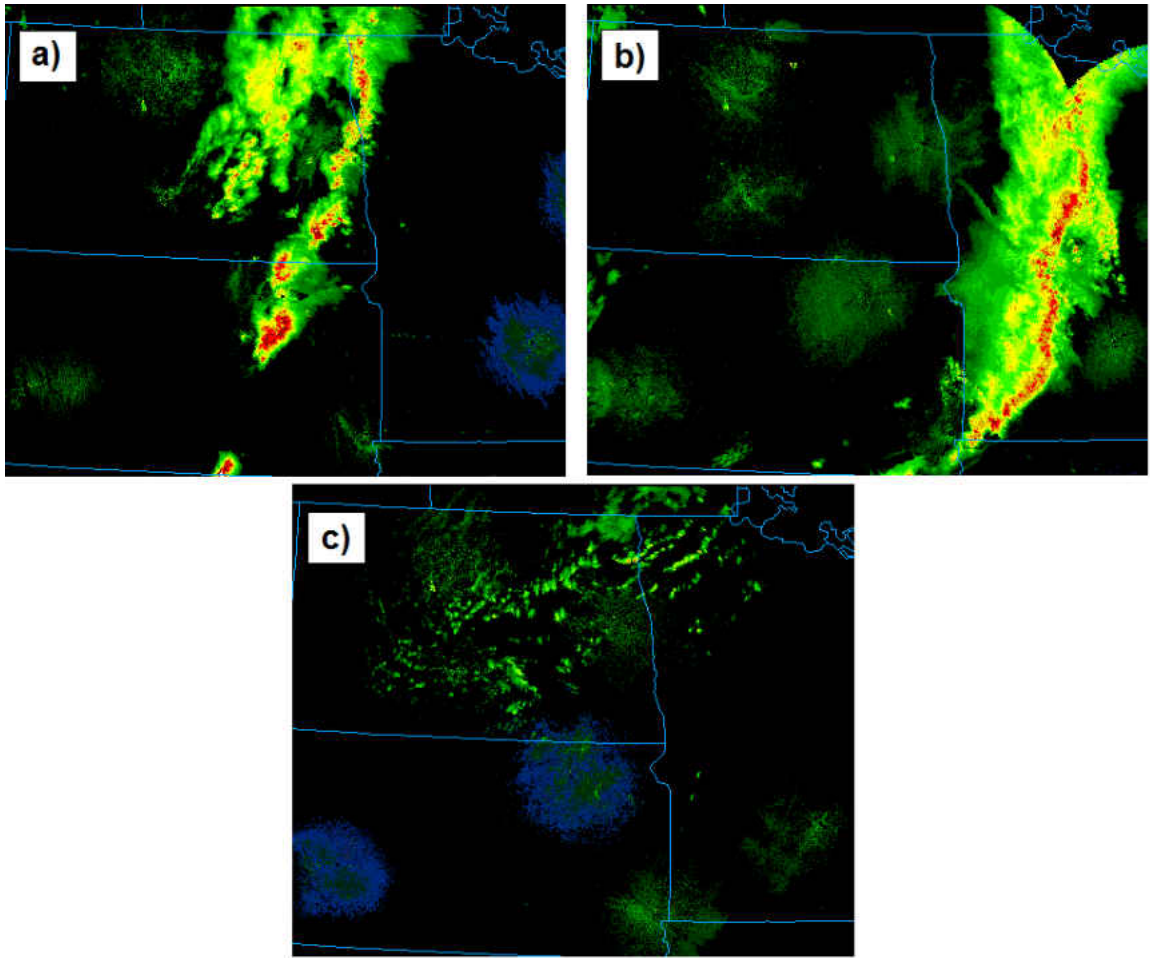


Figure 21: 1 km composite radar reflectivity mosaic at a) 23 UTC on Aug. 3, 2012, b) 03 UTC on Aug. 3, 2012, and c) 18 UTC on Aug. 4, 2012 [Images courtesy from the College of DuPage NeXt Generation Weather Lab (NEXLAB)].

5.5.2 Setup of Sensitivity Study

Since the forecast verification is performed using the WSM6 scheme, the WSM6 scheme is used as the microphysical scheme for the basis of the sensitivity studies. In addition to the control simulation, six simulations are generated for the case day sensitivity study and are summarized in Table 7. The control run has the same physics and domain set-up as the simulations across POLCAST4. However, due to the unavailability of 40 km NAM data used to initialize the operational models that

underwent verification, the 12 km NAM data is used instead. The WRF model version 3.2.1 is used in place of WRF version 3.1.1 due to availability on the supercomputing platform. Despite the differing WRF version and initialization data, results from the control run on the case day are relatively unchanged when compared to the original POLCAST run and show the same biases.

The Thompson simulation in Table 7 is similar to the control except the microphysics are changed from the WSM6 scheme to the new Thompson scheme (Thompson et al., 2008), with the appropriate fixes released for the scheme in the WRF v3.2.1 model (WRF User's Page, 2014). However, the Thompson default CDC of 100 cm^{-3} is kept. A comparison in microphysics from WSM6 to Thompson scheme is performed to determine how much uncertainty in convective forecasts is attributed to just changing microphysical schemes.

Similar to the majority of other single-moment microphysical schemes, the WSM6 scheme uses a constant cloud droplet concentration (CDC) of 300 cm^{-3} , which inherently implies a constant CCN concentration. The high CDC run increases the CDC within WSM6 from the default 300 to 600 cm^{-3} and likewise the low CDC run decreases the CDC down to 100 cm^{-3} . For the case day, the radar and the surface DMT CCN counter were operating across the entire 24-hour period. Since no flights of interest to the POLCAST campaign occurred during Aug. 4, there is no airborne CCN data. The UWyo CCN counter data were also only recorded when a flight of interest to POLCAST occurred. Therefore, only surface-based CCN data from the DMT counter is available. Pre-convective CCN concentrations at 0.3% supersaturation ($\text{CCN}_{0.3}$) of around 500 cm^{-3} are observed in the air mass directly preceding the scattered convective event (Fig. 22; 0813 UTC).

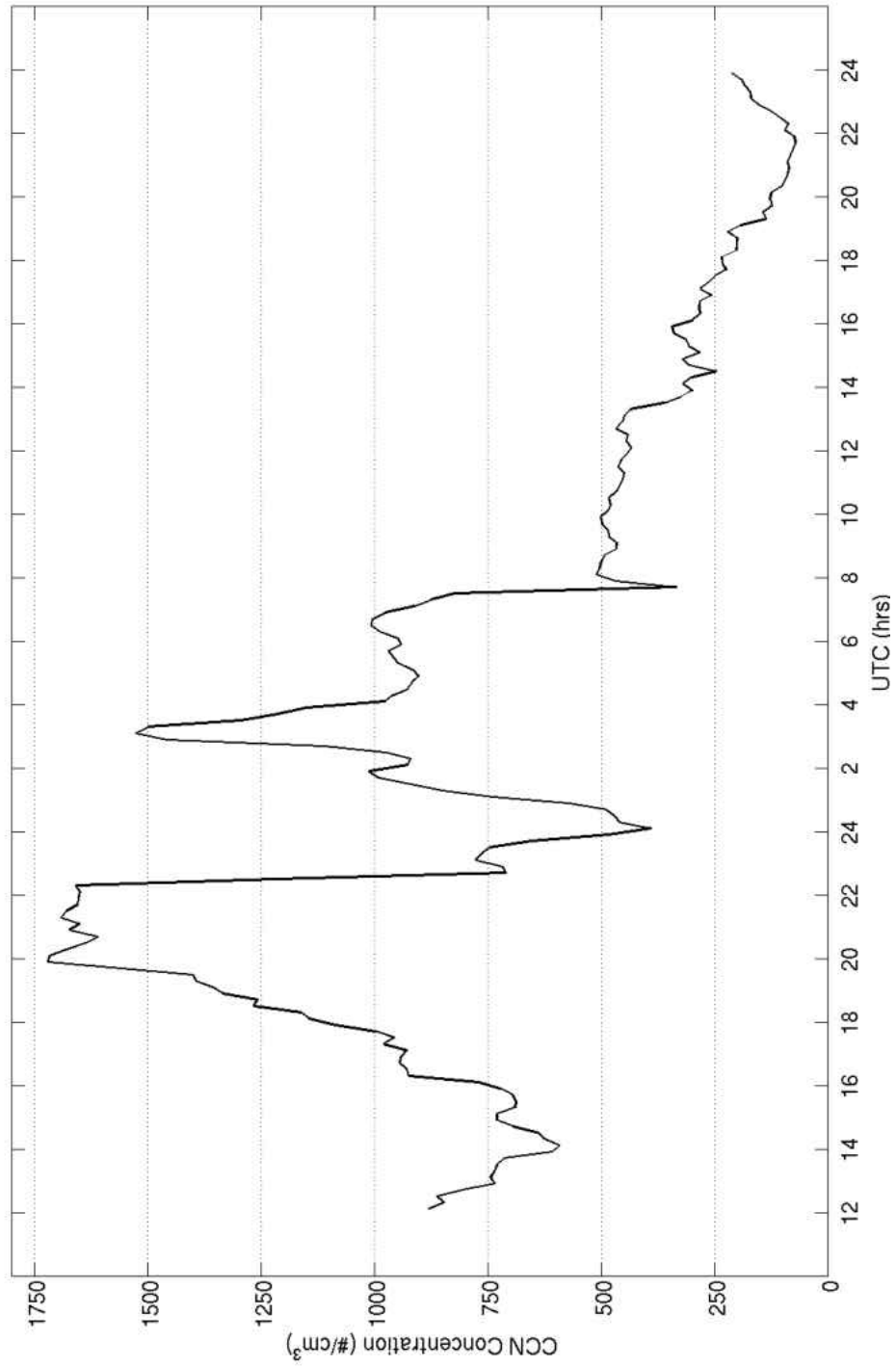


Figure 22: Time series of 10 minute averaged surface-based cloud condensation nuclei (CCN) concentrations starting 12 UTC Aug. 3, 2012.

Table 7: Overview of the different WRF model simulations used for the case day of Aug. 8, 2012. '*' denotes the change in microphysics from WSM6 to Thompson, keeping the default Thompson 100 cm^{-3} cloud droplet concentration as opposed to the default 300 cm^{-3} in WSM6.

Simulation Name	Horizontal Resolution	Microphysical Scheme	Cloud Droplet Concentration (cm^{-3})
Control	3-km	WSM6	300
Thompson	3-km	Thompson	100*
High CDC	3-km	WSM6	600
Low CDC	3-km	WSM6	100
9-km	9-km	WSM6	300
1-km	1-km	WSM6	300
333-m	333-m	WSM6	300

Hegg et al. (2012) generated a linear regression of the cloud droplet concentration onto CCN at 0.3% supersaturation (their Fig. 2) by using data from three different studies. Although their regression is low ($R^2 = 0.33$), similar studies have found the same regressions for different supersaturations (e.g. Delene et al., 2011). Using Hegg et al.'s (2012) linear regression, a pre-convective $\text{CCN}_{0.3}$ environment of 500 cm^{-3} corresponds to CDC of about 400 cm^{-3} . However, Hegg et al.'s (2012) observed $\text{CCN}_{0.3}$ concentrations of 500 cm^{-3} during their field campaigns gave CDC values that ranged from about 350 to 550 cm^{-3} . Additionally, the data was collected from airborne measurements of marine stratocumulus decks, which do not have strong vertical velocities as compared to mid-latitude convection. CDC is known to be affected by both CCN and vertical velocity. Vertical velocities tend to dominate changes in CDC at high CCN concentrations (Hudson and Noble, 2014), with increasing CDC associated with stronger vertical velocities. According to the 3 km operational run, the convection of interest has vertical velocities of up to 1.5 ms^{-1} at cloud base, while the 1 km generates vertical velocities above 2 ms^{-1} . Segal et al. (2007) developed a relation-

ship between vertical velocity and CDC for air that developed under thermodynamic conditions present in the Mediterranean and Texas. Using the vertical velocity of 1.5 ms^{-1} , the results from Segal et al. (2007) produce CDCs of 800 cm^{-3} when using the thermodynamic conditions for Texas. However, the thermodynamic conditions in the Northern Plains are not the same as in Texas. Therefore, a high value of 600 cm^{-3} is used to represent CDC in the modeled environment for the high CDC case. Similarly, a low value of 100 cm^{-3} (representative of post-convective concentrations) is chosen to represent the low CDC case. The chosen droplet concentrations of 600 cm^{-3} and 100 cm^{-3} also represent values that showcase the variability that is seen in eastern North Dakota. Surface CCN concentrations across the entire POLCAST4 campaign show a large variability at all supersaturations (Fig. 23). At 0.3% supersaturation, the 500 cm^{-3} pre-convective environment on the case day is below the 25th percentile, with the mean being around 950 cm^{-3} , further justifying the value of 600 cm^{-3} for the high CDC case.

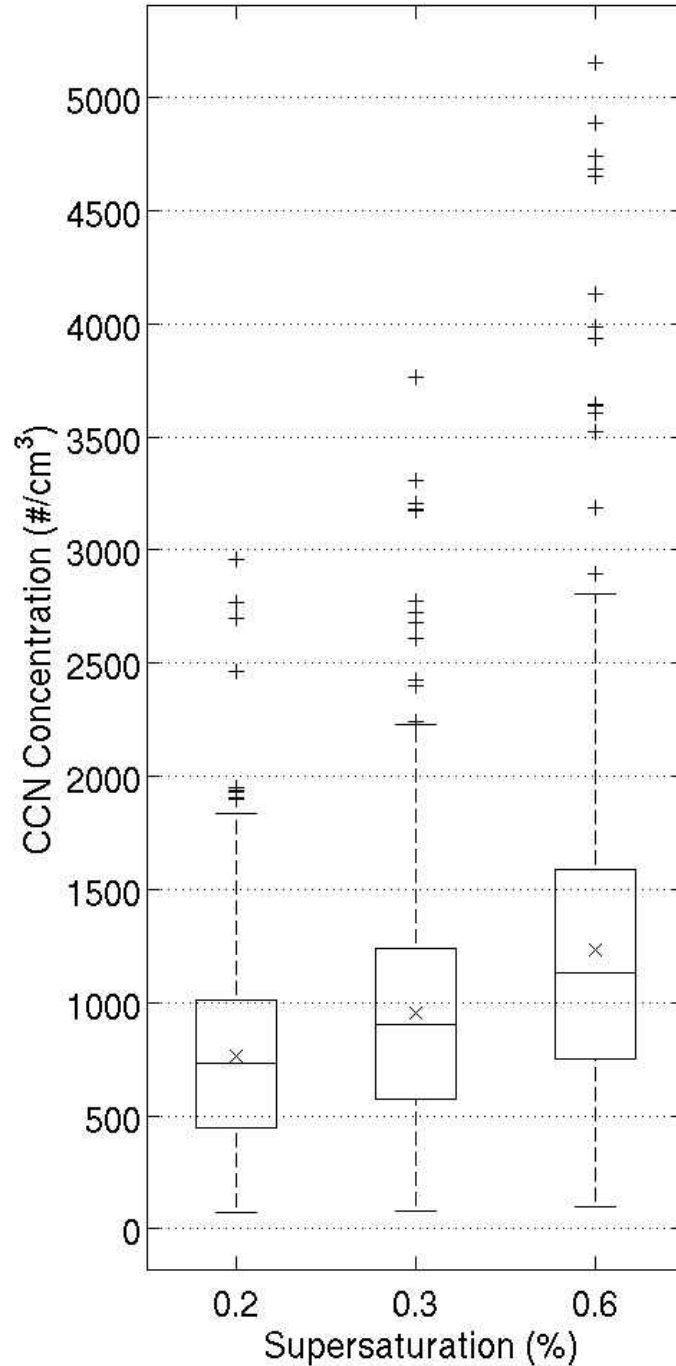


Figure 23: Box plots of cloud condensation nuclei (CCN) concentrations at three different supersaturations of 0.2, 0.3, and 0.6% across the entire POLCAST4 field campaign. The line inside the box indicates the median, the upper and low bounds are the 75th and 25th percentiles respectively, and the extent of the whisker bars denotes the upper and lower fences. The mean is indicated by a 'x' and outliers are denoted by '+'.

The remaining three simulations in Table 7 are associated with changes in resolution while retaining identical physics. Differing horizontal resolutions of 9 km, 1 km, and 333 m are used to simulate convection on the case day. The 9 km simulations shows whether there is any actual improvement in forecasts when convection is explicitly resolved at the 3 km spacing. Since the 1 km performs similarly to the 3 km runs in terms of areal coverage, but does poorly in forecasting the correct amounts of objects, a higher resolution of 333 m is also evaluated to see if there is any improvement with further increases in resolution.

Both higher-resolution 1 km and 333 m simulations are nested within the 3 km domain (Fig. 24b). The 1 km domain resides 10 grid-squares away from the 3 km domain boundaries at all sides. Similarly, the 333 m resolution has 15 grid-squares of transitioning space from the 1 km domain. For the 9 km run, the coarser domain is adjusted to be the same size as the 3 km control run (Fig. 24a). Model output for sensitivity testing is generated every 30 minutes to capture weakly forced convection, which may have very short lifetimes. To compare different resolution simulations to observations, the lowest resolution grid is used for interpolation. For example, the 1 km model run will be interpolated to 2 km resolution (the resolution of the radar data), since 2 km is the smallest object size that can exist in the radar data, and no 1 km objects would be present in the radar data. Likewise for the 3 km model simulations, the 2 km radar data is interpolated to 3 km to ensure that the data are comparable.

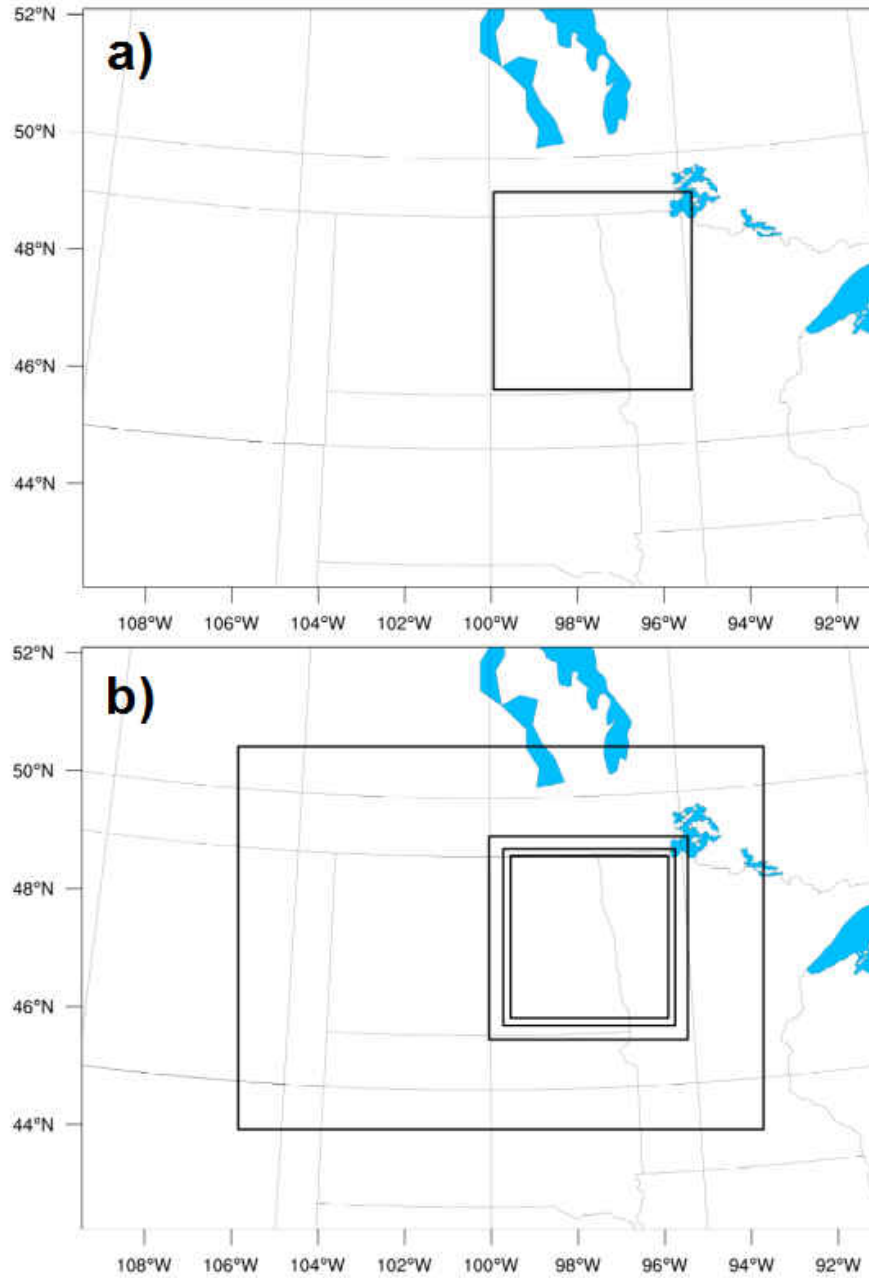


Figure 24: The a) 9 km and b) 333 m, 1 km, and 3 km domains used for sensitivity testing. For b), the inner domain is the 333 m resolution grid, nested within the 1 km and 3 km grids.

5.6 Case Study Sensitivity Analysis and Findings

A time-series of the number of objects present in each simulation for the times of interest across the case day at the 5 dBZ reflectivity threshold and 1 km AGL altitude (Fig. 25a) show that all simulations begin generating convection too early by upwards of 2.5 hours. Model convection also lasts too long, past 00 UTC when convection is no longer present in the observation domain. The simulations are instead continuing to initiate more convection in the domain than what is observed. Two peaks in the number of objects are visible in the observations. The control run (thin black line) performs relatively well compared to observations (thick black line), with both peaks in object numbers being captured; however, the timing is incorrect. The simulation is an hour too slow in generating the first maximum and two hours too slow for the second maximum. The Thompson (red line) and high CDC (blue line) simulations closely resemble each other and do not contain the maxima, but have a more gradually increasing slope. The low CDC (green line) run has a similar trend to the observations, but over-predicts the number of cells when matching up the peaks to observations. While the 1 km (brown line) and 333 m (orange line) simulations both capture the two maximums, they generate too many small cells, which is seen in the verification of the 1 km runs. The 333 m significantly over-forecasts object numbers (Fig. 26), with approximately six times more objects than the observations. The 9 km contains the least amount of objects.

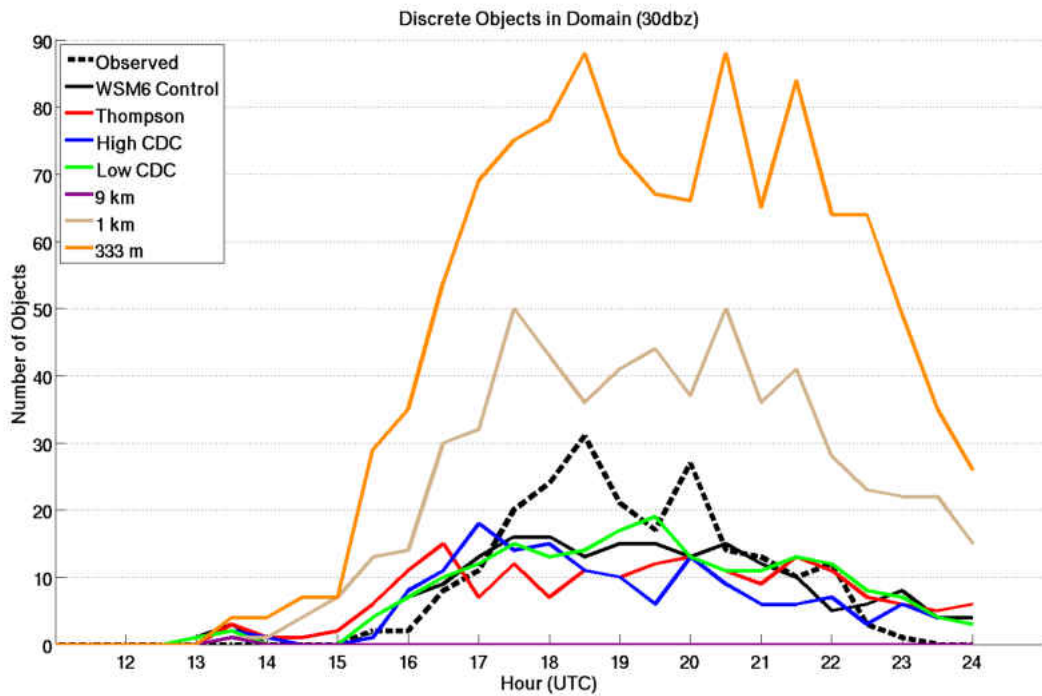
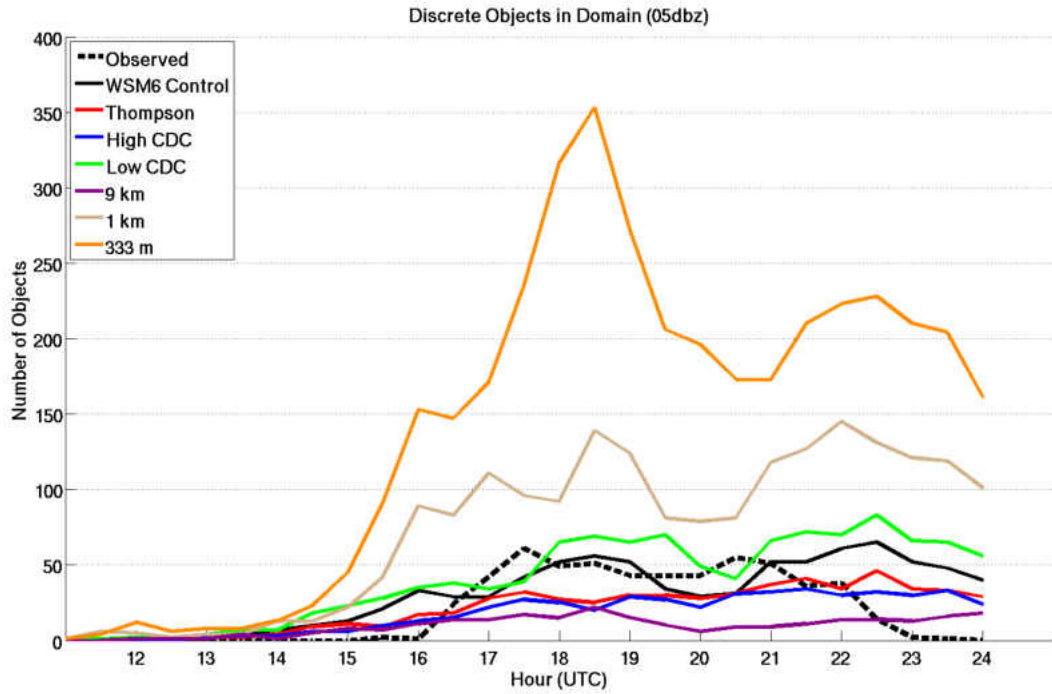


Figure 25: The total number of objects present greater than 5 dBZ (top) and 30 dBZ (bottom) as identified by MODE, across the times of interest for the case day of Aug. 4, 2012.

At 30 dBZ (Fig. 25b), all 3 km simulations under-predict the maximum number of objects. Although the 9 km simulation subjectively matches the observations well in terms of location at 5 dBZ reflectivity threshold, the resolution is too coarse to generate stronger updrafts and hence contains only one object above 30 dBZ early on. The higher resolution runs contain too many cells but also over-intensify the cells. Both the 1 km and 333 m are the only runs that generate cells above 45 dBZ, which are not in observations.

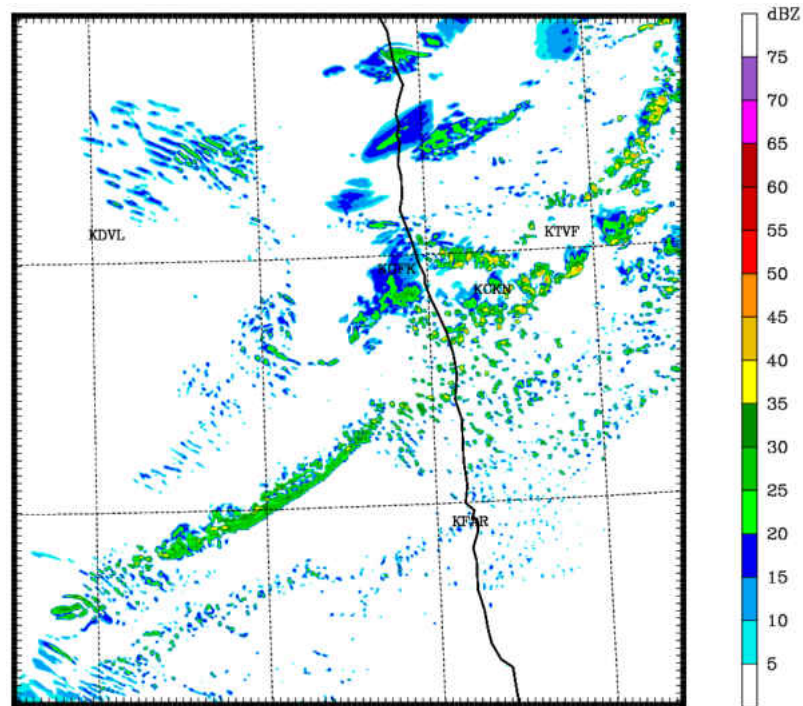


Figure 26: 333 m simulated reflectivity field valid on 20 UTC on Aug. 4, 2012.

Since the tendency for higher resolution runs is to generate too many objects, focus is shifted onto the area covered by convection to evaluate the forecasts instead. All simulations greatly over-predict the coverage of convection at 5 dBZ (Fig. 27a). The high CDC (blue line) case performed the best in terms of total coverage, while the higher resolution runs (orange and brown lines) and Thompson (red line) runs capture the general shape of the distribution better. The low CDC run (green line) has

the lowest performance in terms of greatest over-prediction at peak convective period, however the 9 km (purple line) performs the worst overall. Changing the CDC greatly influences convective areas and affects the area coverage bias. Thompson performs better than the control scheme in terms of area coverage and timing, however the differences weren't major. Even though the higher resolution realizations generate significantly more objects, they perform better in areal coverage than all other runs except the high CDC case. At 30 dBZ (Fig. 27b), all simulations contain a rapid increase in areal coverage that proceeds observations by two to three hours, with Thompson being the earliest and containing the most area. The rapid increase in area corresponds with a series of cells translating and initiating in the northwestern region of the domain in all simulations.

By analyzing the vertical velocity field, it is visible that the 333 m simulation starts to become unstable in the southeastern corner of the domain at around 15 to 16 UTC. By the end of the simulation, the instability is realized across the entire domain. Before the simulation becomes unstable, the tendency of the simulation is to generate even more objects than the 1 km simulation and similar areal coverage. However, the fact that the simulation goes unstable at all indicates that the instability may have affected the solution from the start, and the results should not be trusted. Therefore, the 333 m simulation is not analyzed further.

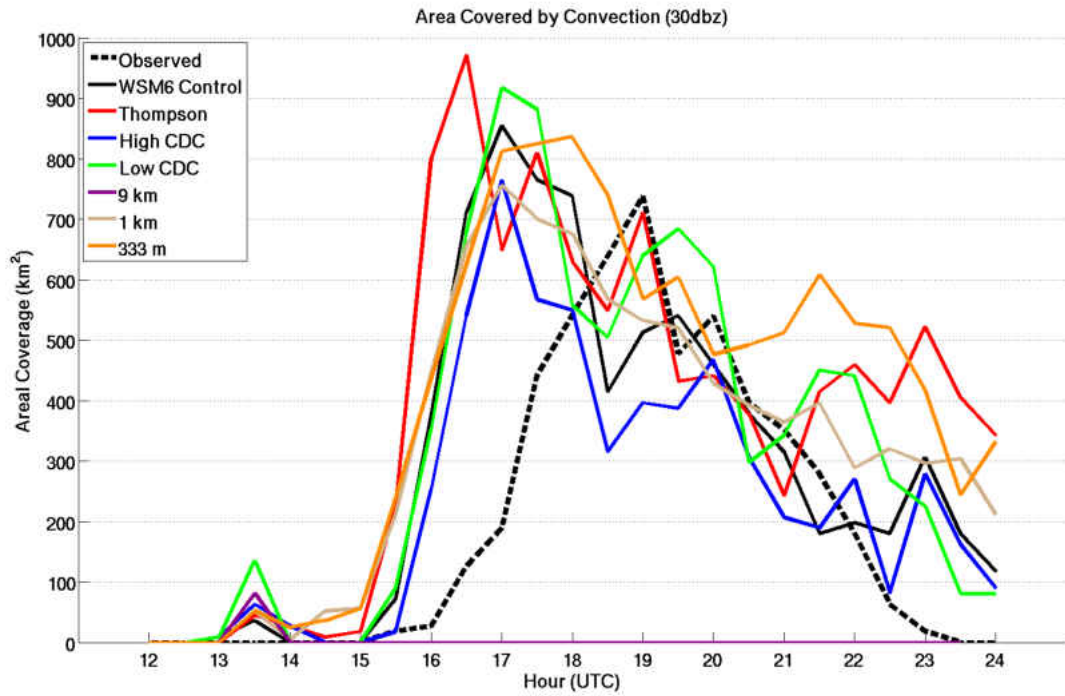
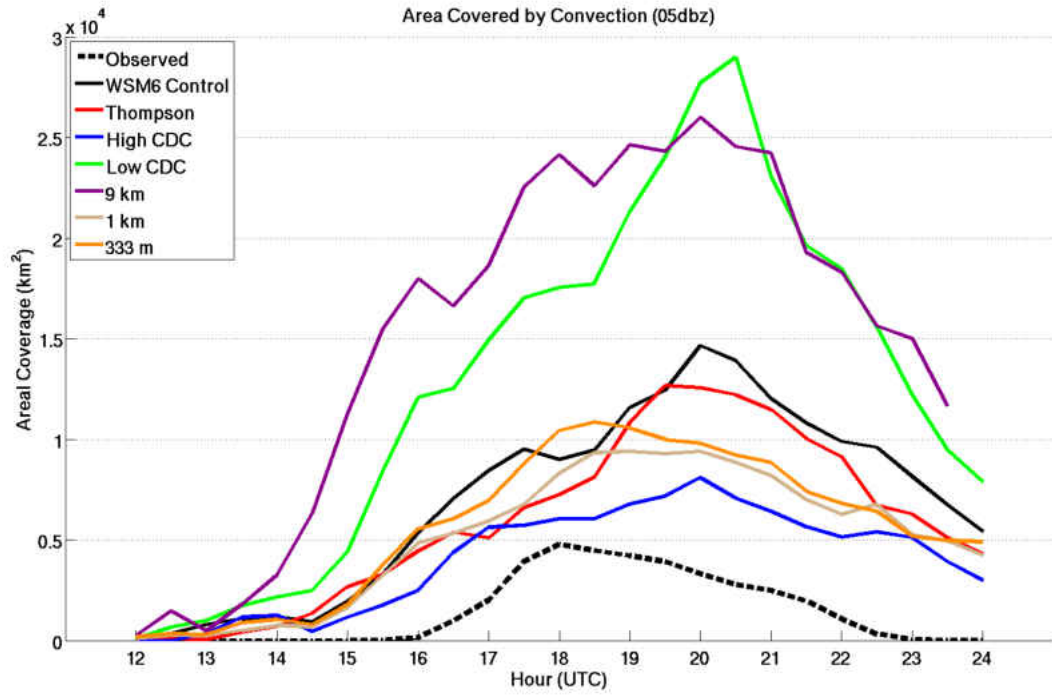


Figure 27: The total area covered by convection greater than (top) 5 dBZ and (bottom) 30 dBZ as identified by MODE, across the times of interest for the case day of Aug. 4, 2012.

Table 8: Average object sizes presented for the 14 to 00 UTC time period on Aug. 4, 2012.

Avg. Obj. Size (km ²)	5 dBZ		30 dBZ	
Simulation	Fcst	Obs	Fcst	Obs
Control	211.99	65.53	39.45	23.25
Thompson	264.19	65.53	53.59	23.25
High CDC	209.39	65.53	38.61	23.25
Low CDC	299.82	65.53	42.04	23.25

Overall, there is significantly too much area covered by convection across all simulations. However, objects counts are generally not as different when compared to observations (except by the high resolution runs). The over-forecasting of area with object counts being closer to observations indicates that for the case day, the areal extent of individual objects is too large leading to the large amounts of area covered by convection. A breakdown of average object sizes for the simulations and observation (Table 8) shows that simulations generate object sizes that are three or more times larger than the observed value of 66 km². The observed average object sizes change between different resolution runs because the radar data is scaled to the lower resolution model grid (see Section 5.5.2). The 9 km simulation is not shown, since it did not improve upon any of the biases found. The 9 km simulation instead amplified the area bias to about 10.5 times greater than observed area and created a weak bias in generated convection. The 9 km resolution did reduce the number of objects, which is essentially due to smaller objects observed in higher resolution runs being smoothed to match the lower grid spacing of the 9 km. The large grid squares create large regions of precipitation that are also too weak due to the grid spacing being unable to generate stronger updrafts, and hence the correct magnitude is not simulated. Therefore the 9 km resolution run is not analyzed further. The higher

resolution runs are also not presented due to the average object sizes being biased by the overwhelming amounts of small objects.

Table 9: The ratios of the number of forecasted objects to observed objects and the total forecasted area to observed area across the 14 to 00 UTC time period on Aug. 4, 2012. ”-” denote cases where no observed objects matched the area and intensity thresholds, hence a ratio cannot be determined.

FO Ratios	5 dBZ		30 dBZ	
Simulation	Objects	Area	Objects	Area
Control	1.45	4.70	0.86	1.45
Thompson	0.99	4.01	0.81	1.88
High CDC	0.85	2.71	0.70	1.17
Low CDC	1.90	8.72	0.89	1.61
9-km	1.36	10.54	-	-
1-km	3.52	4.56	3.07	2.48
333-m	6.91	4.89	5.90	2.88

Across the 3 km runs, changing the cloud droplet concentration and microphysical scheme influences both the number of objects and total area. FO ratios for the total number of objects and total areal coverage across the 14 to 00 UTC time span are presented in Table 9. Increasing the CDC reduces object counts and total area coverage by convection, while decreasing the CDC increases both the number of objects and the areal coverage. Lowering the CDC also increases the average object size, while there is no significant changes for the high CDC case (Table 8). Compared to the control, the high CDC case has a decrease in FO object ratio from 1.45 in the control to 0.85 and decrease in FO area ratio from 4.70 to 2.71. However, average object areas between the control and high CDC simulations are similar. The similar object sizes implies that raising the CDC removes objects and hence the area they provide instead of changing the areal extent of the objects themselves. In contrast,

the low CDC runs add more objects (1.90 FO ratio) and increase object sizes to an average of about 300 km², which adds significantly more area (8.72 FO ratio).

A change from WSM6 to Thompson decreases object counts (Table 9) total areal coverage of convection. However, average object sizes at both 5 and 30 dBZ reflectivity thresholds are greater for Thompson (Table 8), which indicates that any improvement in total areal coverage FO ratios is simply because the number of objects decreases. At 30 dBZ, the control run has better skill than the Thompson scheme in terms of the number of objects and area. The average object size for the Thompson scheme is significantly larger at 30 dBZ than at the other 3 km simulations, which shows that changes in microphysical scheme affects convective intensity.

5.6.1 Discussion of Observed Cloud Droplet Concentration Sensitivities

To assess what generates the differences in reflectivity area among the simulations, a stratification of the hydrometeor concentrations used in the simulated reflectivity calculation (rain, snow, and graupel) is performed. To ensure that the hydrometeor comparisons are equal between the different sized grids due to nesting (e.g. 1 km grid is slightly smaller than the 3 km grid), only the 333 m grid is used for comparison, meaning any data outside the extent of the 333 m grid for the other resolution grids is disregarded. Hydrometeors are summed across the entire domain for every model output period to generate a total hydrometeor mass present in the domain for grid cells that are above the 5 and 30 dBZ thresholds.

Hydrometeor concentrations in the observation domain are determined by using Gao and Stensrud's (2012) simple Z-q relationships of

$$q_r = \frac{10^{\frac{Z-98.1}{17.5}}}{\rho} \quad (5.1)$$

for rain and

$$q_s = \frac{10^{\frac{Z-89.9}{17.5}}}{\rho} \quad (5.2)$$

for dry snow, where Z is the radar reflectivity (in dBZ), ρ is the density of air, and q_r and q_s are the rain and snow mixing ratios. Total hydrometeor mass is obtained from radar reflectivity by summing mixing ratios horizontally across the entire domain starting at a 0.5 km height and increasing every half kilometer until 5 km. To distinguish between rain and snow, 3-hourly North American Regional Reanalysis (NARR; Mesinger et al., 2006) data is utilized to determine the freezing level altitude. NARR is a 32 km resolution combined model and assimilation dataset. The coarseness of the resolution allows for little interference that any potential convection may generate in finding the freezing level but is high-resolution enough to still have good spatial variation in the domain. Since the NARR data is present on 29 pressure (height) levels, the freezing level is found by averaging the first height that contains a temperature below freezing with the height of the previous warmer level, at each point in the grid. The freezing levels are then rounded to the closest radar level (every 0.5 km). Since the dataset is 3-hourly and hydrometeor concentrations are compared to the model output (every 30 minutes), the freezing levels determined are linearly interpolated between two 3-hour periods. Any reflectivity values above the freezing level are assumed to be snow hydrometeors and below that level to be rain. Since there is no graupel category included in observations, as only snow and rain are assumed, the graupel mass within simulations is added to the snow mass, resulting in a comparison between liquid (rain) and solid (snow and graupel) hydrometeors. This method of determining hydrometeor type is simplistic but should provide enough information about the observed hydrometeor types for verification purposes. It is not the goal to find a precise hydrometeor retrieval algorithm.

The maximums in hydrometeor mass for the 3 km simulations occur later than in observations (Fig. 28; top), and all simulations but the low CDC case contains less mass. Increasing the CDC does improve the timing in generation of precipitants, taking longer for the generation of hydrometeors by around 30 minutes as compared to the control run. On the other hand, the low CDC case begins generating precipitants at least 30 minutes before the control run. The high CDC case also begins to decrease the amount of hydrometeors sooner than the control and low CDC cases which more closely matching observations. While timing of hydrometeor generation does improve, there is less hydrometeor mass present in the domain for the high CDC simulation than in observations. After 1630 UTC, hydrometeor mass is not increasing as rapidly as the observations indicate. Therefore, increasing the CDC lowers the total mass of hydrometers while lowering the CDC raises the mass of hydrometeors.

To investigate the reason for the changes in hydrometeor mass due to changes in CDC, the liquid hydrometeors (rain) are analyzed. The liquid hydrometeors are focused on because the cloud droplets (and therefore CDC) directly affect the warm rain process, and a significant majority of the cells are above the freezing temperature (less than 5 km in altitude). Figure 29 shows that the same trends are visible in the total liquid hydrometeor mass as in the total hydrometeor mass; however, the magnitudes are different when compared to observations. The control best represents the magnitude of rain hydrometeors, while the low CDC simulation generates too much rain and the high CDC simulation generates too little rain. However, the high CDC simulation improves on the timing of generation and dissipation of rain hydrometeors. Within WSM6, the CDC directly affects two processes: the autoconversion of cloud water to rain and the heterogeneous freezing of cloud water to cloud ice.

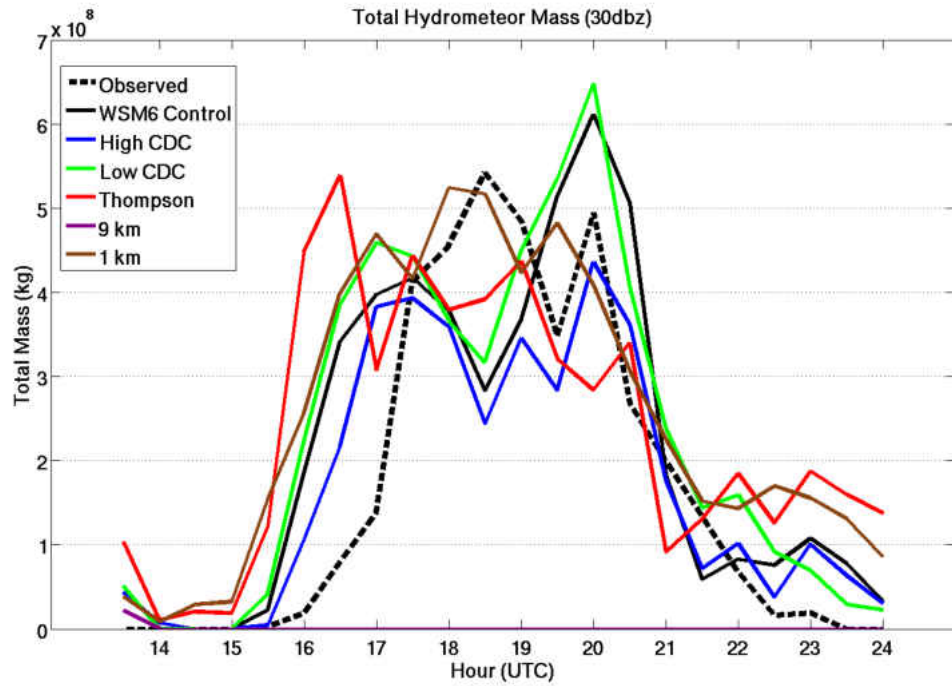
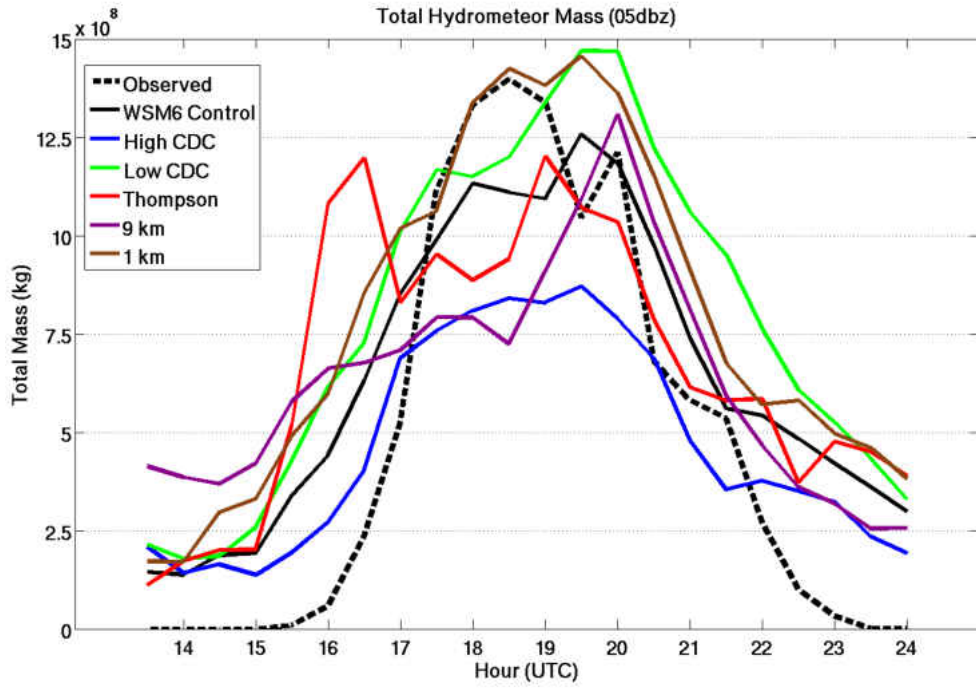


Figure 28: Total hydrometeor mass of rain, snow, and graupel across the domain and times of interest for the case day of Aug. 4, 2012, at greater than 5 dBZ (top) and 30 dBZ (bottom).

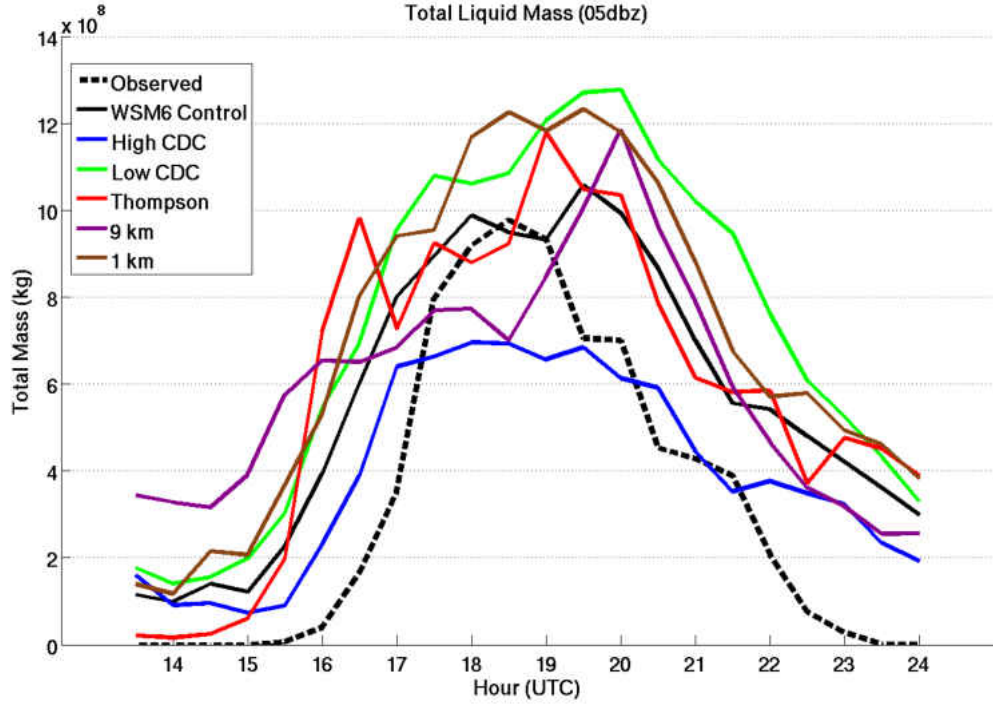


Figure 29: Total mass of rain for convective areas greater than 5 dBZ across the domain and times of interest for the case day of Aug. 4, 2012.

First, evaluating the WSM6 autoconversion rate (i.e. instantaneous coalescence rate) adapted from Tripoli and Cotton (1980), the conversion of cloud water to rain is given by:

$$\frac{dq_r}{dt} = \frac{0.104gE_c\rho_c^{4/3}}{\mu(N_c\rho_w)^{1/3}}q_cH(q_c - q_{co}), \quad (5.3)$$

where g is the gravitational acceleration, E_c is the mean collection efficiency, ρ_c and ρ_w are the densities of air and water, μ is the dynamic viscosity of air, N_c is the cloud droplet concentration, q_c is the cloud water mixing ratio, and q_{co} is the critical cloud water mixing ratio. H denotes the Heaviside step function, returning either a '0' or '1' depending on if the cloud water mixing ratio is greater than the critical cloud water

mixing ratio, which is determined by:

$$q_{co} = \frac{4\pi\rho_w r_{cr}^3 N_c}{3\rho}, \quad (5.4)$$

where r_{cr} is the mean critical cloud droplet radius ($8 \mu\text{m}$) and ρ is the density at the location of analysis. For Eq. (5.3), all variables are constant except for the cloud water mixing ratio, showing that the more cloud water that is available, the greater the conversion from cloud water to rain.

For the high CDC case, which has the N_c value increased from 300 cm^{-3} to 600 cm^{-3} , which results in a lower rate of conversion into rain since you are dividing by a larger number in Eq. (5.3) (i.e. autoconversion inversely correlated with CDC). Likewise, because the critical cloud water mixing ratio is dependent on the CDC (Eq. (5.4)), increasing the CDC results in a higher critical mixing ratio (i.e. autoconversion activation is directly proportional to CDC). The higher critical cloud water mixing ratio means that more cloud water is required to begin the autoconversion into rain hydrometeors. Lowering the CDC has the result of increasing the rate of conversion from cloud water to rain and lowers the critical cloud water mixing ratio, which allows the autoconversion to start at lower cloud water mixing ratios.

A simple model of the autoconversion function (Fig. 30) at varying q_c shows the variability in the cloud water coalescence rates. At a q_c of 2 g kg^{-1} , the autoconversion rate for CDC concentrations of 100, 300, and 600 cm^{-3} is 0.49, 0.34, and $0.27 \text{ kg kg}^{-1} \text{ s}^{-1}$ ($\times 10^{-5}$), respectively. However, at a q_c of 4 g kg^{-1} the rates increase and grow in spread to 2.48, 1.72, and $1.36 \text{ kg kg}^{-1} \text{ s}^{-1}$ ($\times 10^{-5}$). Assuming a q_c of 4 g kg^{-1} , if the autoconversion rate calculation takes place at one location (one grid square) for 20 minutes, representing the life cycle of cell that may contain such a mixing ratio of cloud water, the low CDC case contains 47% more rain hydrometeors compared

to the control while the high CDC case contains 17.6% less rain hydrometeors than the control. Summed across several locations, the low CDC case is therefore able to convert significantly more cloud water to rain.

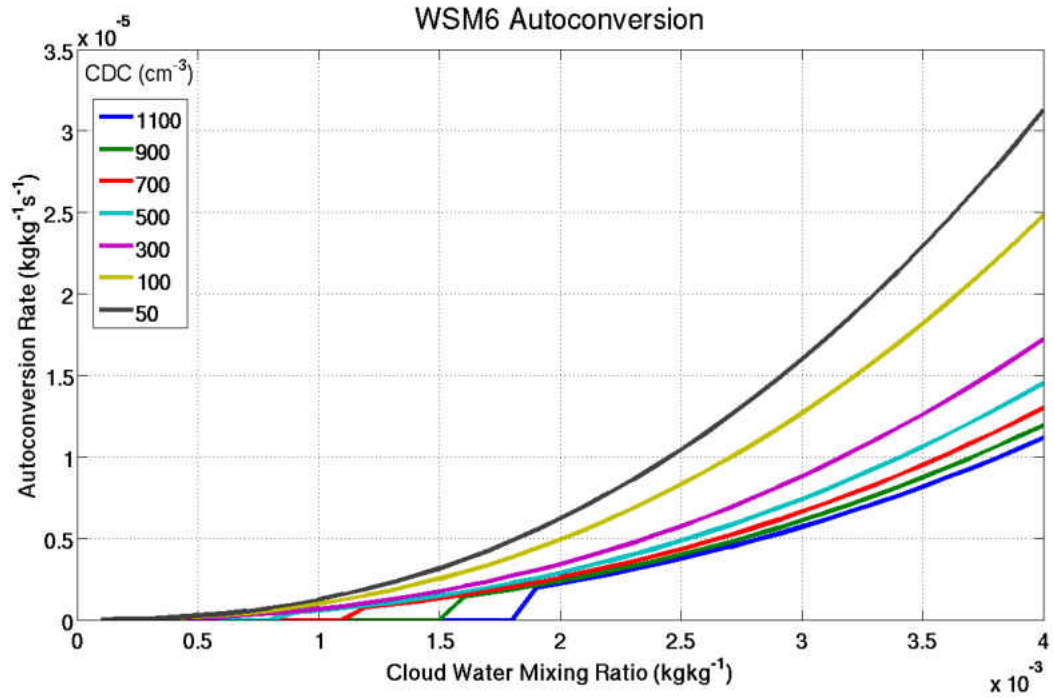


Figure 30: The WSM6 autoconversion rates of cloud water to rain hydrometeors for varying cloud water mixing ratios. Different colored lines indicate different cloud droplet concentrations.

Due to the presence of the step function, critical mixing ratios of 0.2, 0.8, and 1.6 g kg^{-1} , corresponding to the CDCs of 100, 300, and 600 cm^{-3} , must also be achieved in order for the autoconversion to take place. Therefore, for a q_c of 1 g kg^{-1} , the high CDC case would not generate any rain, and hence any objects, while the low CDC case may contain several objects if multiple regions containing such q_c exist. Once the critical mixing ratio is reached, the rate of conversion is always greater for the low CDC case than the control or high CDC cases (Fig. 30) resulting in the generation of more rain. Hence, the low CDC case may contain several objects converting cloud water to rain and at a greater rate than the high CDC case, which

may just contain a few. Therefore, lowering the CDC leads to the significant over-forecasting of objects and area (Table 9), including larger average object-sizes (Table 8) and more rain hydrometeors (Fig. 29). Since the control simulation is closer in magnitude to amount of rain mass present in the domain compared to the observations than both the low and high CDC cases (Fig. 29), instead of changing the cloud droplet concentration it may be useful to investigate increasing the mean critical radius in Eq. (5.4) in order to suppress activation of the autoconversion process. This suppression may result in better representation of the timing in generation of rain hydrometeors. WSM6 assumes a commonly used constant radius of $8 \mu\text{m}$, while the average particles in North Dakota may be larger due to combinations of stronger average wind speeds than most continental locations and large areas covered by agricultural fields, which may loft larger particles.

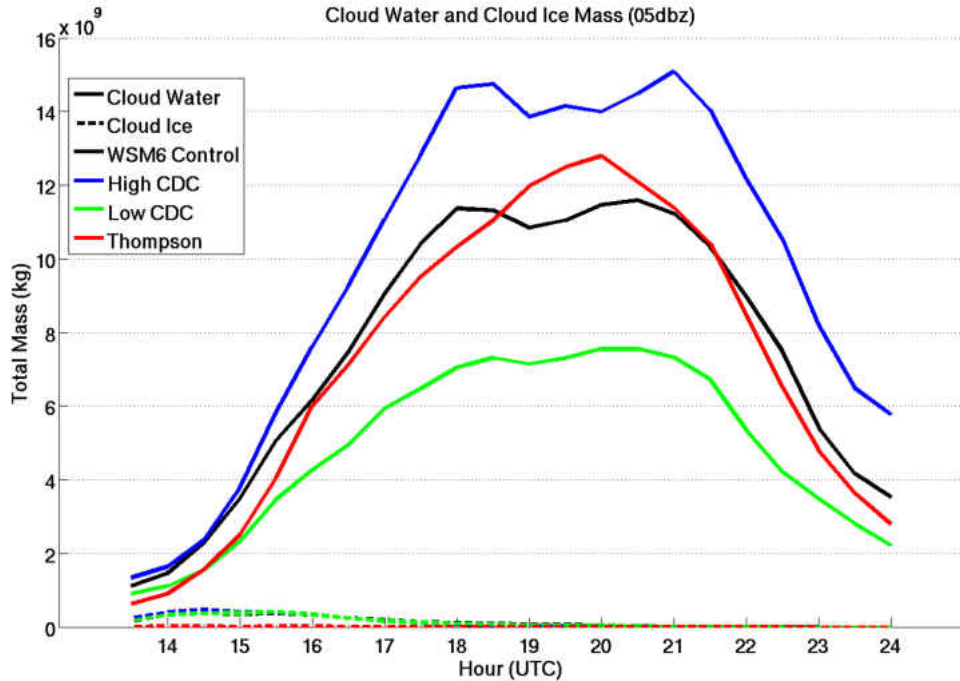


Figure 31: The total mass of cloud water (solid lines) and cloud ice (dashed lines) present across the domain of the 3-km simulations when convection was greater than 5 dBZ in intensity across the times of interest on the case day of Aug 4., 2012.

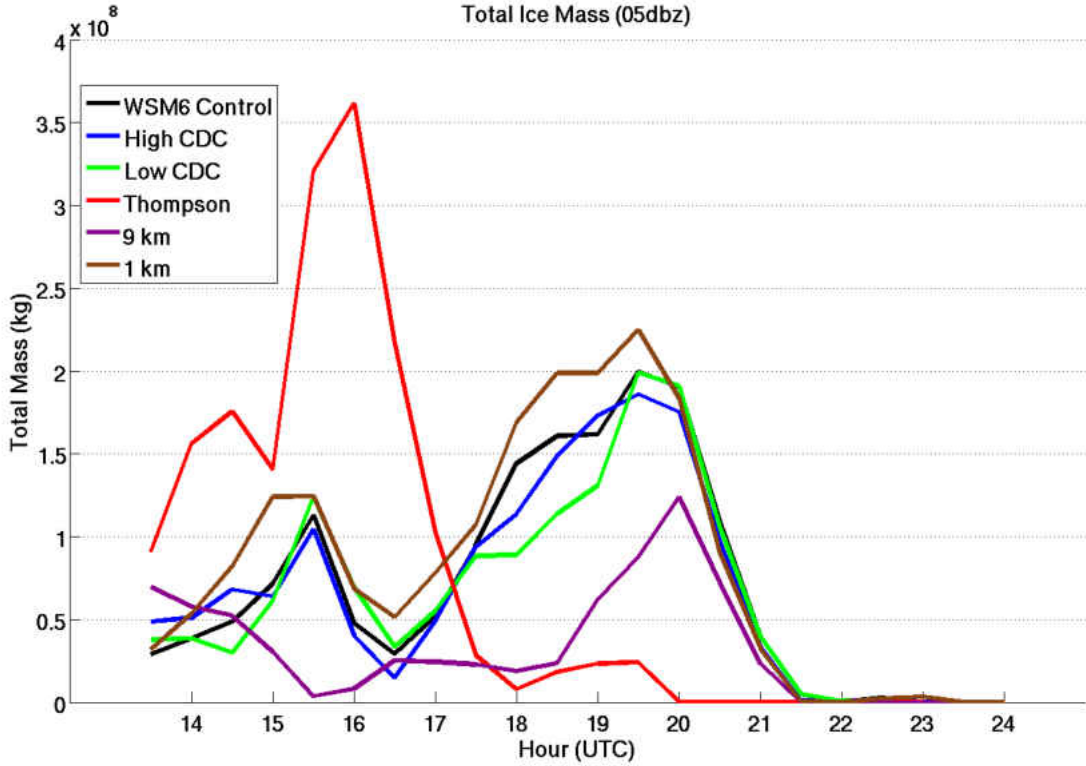


Figure 32: The total mass of snow present across domain of different model simulations when convection was greater than 5 dBZ in intensity across the times of interest on the case day of Aug. 4, 2012.

The other parameter that is directly affected by the cloud droplet number concentration is the heterogeneous freezing of cloud water to cloud ice. Following a similar procedure as above, the WSM6 freezing rate is adapted from Biggs (1953) and is given by

$$\frac{dq_i}{dt} = 100(e^{(0.66[T_o - T])} - 1) \frac{\rho q_c^2}{\rho_w N_c}, \quad (5.5)$$

where where q_i is the cloud ice mixing ratio, T_o is the reference temperature (273.15 K), and T is the temperature. Similar to the autoconversion rate, as the CDC increases the freezing rate decreases. Hence, decreasing the CDC increases the freezing rate, leading to more cloud ice mass. The increase in cloud ice allows for more cloud ice to aggregate to snow or be accreted by snow, which thereby melts and increases

the amount of rain hydrometeors present. However, while this process may be important for other days or convective events, the majority of the simulated convection is above the freezing temperature so there is relatively little cloud ice as compared to cloud water (Fig. 31) for the case day. Similarly there is little snow mass (Fig. 32) across the WSM6 runs as compared to rain (Fig. 29) which significantly minimizes this effect. Lastly, a simple model of the freezing rate shows that while differences in the rate exist (Fig. 33) especially when lowering the CDC below 300 cm^{-3} , the rate is not nearly as significant as the autoconversion rate at these low cloud water mixing ratios. Therefore, the freezing rate of cloud water to cloud ice is not considered a significant factor in the forecast bias.

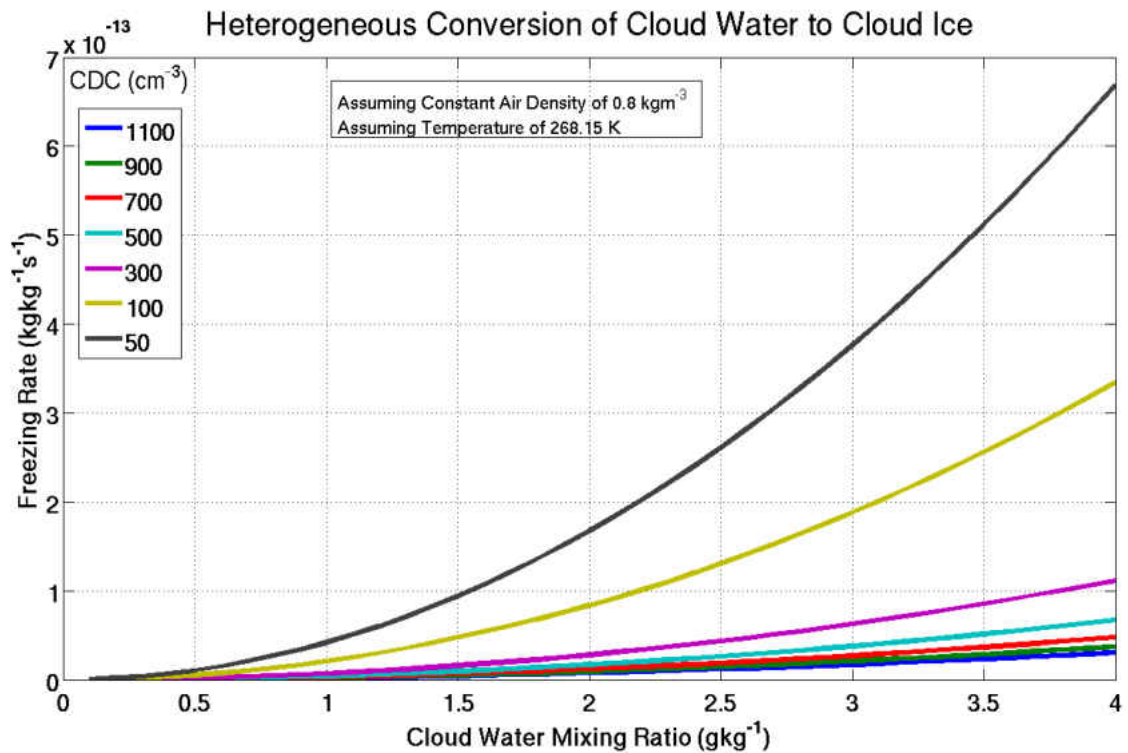


Figure 33: The WSM6 heterogeneous freezing rates of cloud water to cloud ice at a constant air density of 0.80 kg m^{-3} and temperature of -20°C for varying cloud water mixing ratios. Different colored lines indicate different cloud droplet concentrations.

5.6.2 Discussion of Microphysical Parameterization Sensitivities

Analyzing the differences between the Thompson and WSM6 control runs, the total hydrometeor masses appear to be relatively similar (Fig. 28); however, the Thompson scheme produces an initial peak at 16 UTC followed by a secondary peak hydrometeor mass at 19 UTC. The peak at 16 UTC is also present at the 30 dBZ threshold, while no other simulation contains the peak. Investigating the peak in hydrometeor mass further, it is evident that it originates in the snow hydrometeor field (Fig. 32). The Thompson scheme generates significant amounts of snow over the WSM6 scheme between 15 and 16 UTC, although a lower magnitude maximum is also observed across the other WSM6 scheme simulations. This maximum originates from a convective cell being translated into the inner analysis domain. As the cell enters the 3 km nest from the 9 km domain, it contains an increased amount of snow aloft aided by a stronger updraft. The cell rapidly weakens as the updraft quickly collapses, leaving behind an anvil. While all simulations contain snow hydrometeors during this event, the Thompson scheme generates significantly more snow and keeps in the domain for a longer period. Focusing then on the liquid hydrometeor mass present in the domain, the effects of the snow generated by the Thompson scheme are visible around the 16 and 1630 UTC time frame (Fig. 29). As the snow hydrometeors melt, a maximum in rain is visible. Excluding the effect of snow, the Thompson total liquid mass would be very similar to the WSM6 mass between the 1530 to 17 UTC period, leading to the Thompson and WSM6 schemes performing very similar to each other.

It is important to reiterate that the Thompson scheme is used with the standard preset CDC of 100 cm^{-3} , which is representative of clean or marine air and is the low value used for the low CDC runs. Thompson et al. (2008) strongly advised that the CDC be changed to ambient or known concentrations, because of the way

the Thompson scheme autoconversion rate is calculated. While WSM6 uses a fixed critical diameter and CDC (i.e. assuming monodispersed droplets) when calculating the critical q_c used in the autoconversion (Eq. (5.3)), the Thompson scheme uses the fixed CDC to generate a gamma distribution of droplet sizes. Therefore, the autoconversion rate, based off Berry and Reinhardt (1974),

$$\frac{dq_r}{dt} = \frac{0.027\rho q_c(\frac{1}{16} * 10^{20} D_b^3 D_f - 0.4)}{\frac{3.72}{\rho q_c}(\frac{1}{2} * 10^6 D_b - 7.5)^{-1}}, \quad (5.6)$$

requires the diameters of the corresponding gamma distribution, D_b and D_f , to be calculated. The diameters are expressed as

$$D_f = \left(\frac{6\rho q_c}{\pi\rho_w N_c}\right)^{1/3} \quad (5.7)$$

and

$$D_b = (D_f^3 D_g^3 - D_f^6)^{1/6} \quad (5.8)$$

where D_g is given by

$$D_g = \frac{[\frac{\Gamma(\mu_c+7)}{\Gamma(\mu_c+4)}]^{1/4}}{\lambda_c}, \quad (5.9)$$

where μ_c and λ_c are the shape parameter and slope intercept of the gamma distribution, respectively (Thompson et al., 2008). For cloud water, the shape parameter is defined as

$$\mu_c = \frac{10^9}{N_c} + 2 \quad (5.10)$$

and is minimized to be no greater than 15. At the default CDC value of 100 cm^{-3} used in the Thompson simulation, the value of μ_c is 12. There is no step function used by the Thompson scheme. Instead, a minimum q_c of $1 \times 10^{-5} \text{ kg kg}^{-1}$ is required for autoconversion.

A simple model of the Thompson autoconversion rates (solid lines) and the WSM6 autoconversion rates (dashed lines) is shown in Figure 34. Comparing the Thompson simulation at a CDC of 100 cm^{-3} against the WSM6 low CDC run (100 cm^{-3} ; blue lines), the autoconversion rate is similar at a q_c of 3.8 g kg^{-1} (1×10^{-3}). However, it is evident that the rate of autoconversion in the Thompson simulation is rapidly increasing for increasing q_c values after 3.8 g kg^{-1} , above that of the WSM6 scheme. The Thompson scheme at the default CDC thus contains much more rain hydrometeors for higher q_c , strengthening the convective core more than even the low CDC WSM6 simulation. The increase in rain within Thompson is the reason for the larger object sizes at 30 dBZ (Table 8). Below 3.8 g kg^{-1} , the WSM6 rate is greater, even for the high CDC (600 cm^{-3}) simulation. The WSM6 scheme converts more cloud water to rain at lower cloud water mixing ratios, even for air that contains more cloud droplets and is considered more polluted than the clean air in the Thompson scheme. If the CDC within the Thompson scheme was increased to 300 cm^{-3} (the value of the WSM6 control run), at least for q_c values less than 4 g kg^{-1} , no reasonable CDC within WSM6 would allow for such similar low autoconversion rates.

Another visible difference between the WSM6 and Thompson autoconversion rates is the differences in each schemes' rates when changing the CDC. For example, at a q_c of 4 g kg^{-1} , changing the CDC from 100 cm^{-3} to 300 cm^{-3} in WSM6 changes the autoconversion rate from about $2.5 \text{ kg kg}^{-1} \text{ s}^{-1}$ (10^{-5}) to $1.7 \text{ kg kg}^{-1} \text{ s}^{-1}$ (10^{-5}) which is a decrease of 32%. The Thompson rate decreases from about $2.7 \text{ kg kg}^{-1} \text{ s}^{-1}$ (10^{-5}) to $0.6 \text{ kg kg}^{-1} \text{ s}^{-1}$ (10^{-5}) which is a decrease of 77%. Therefore, not only are simulations sensitive to changes in CDC, but also to changes in the formulation used. The Thompson formulation derived from Berry and Reinhardt (1974) is more sensitive to the cloud droplet concentration, especially in the 100 to 300 cm^{-3} range.

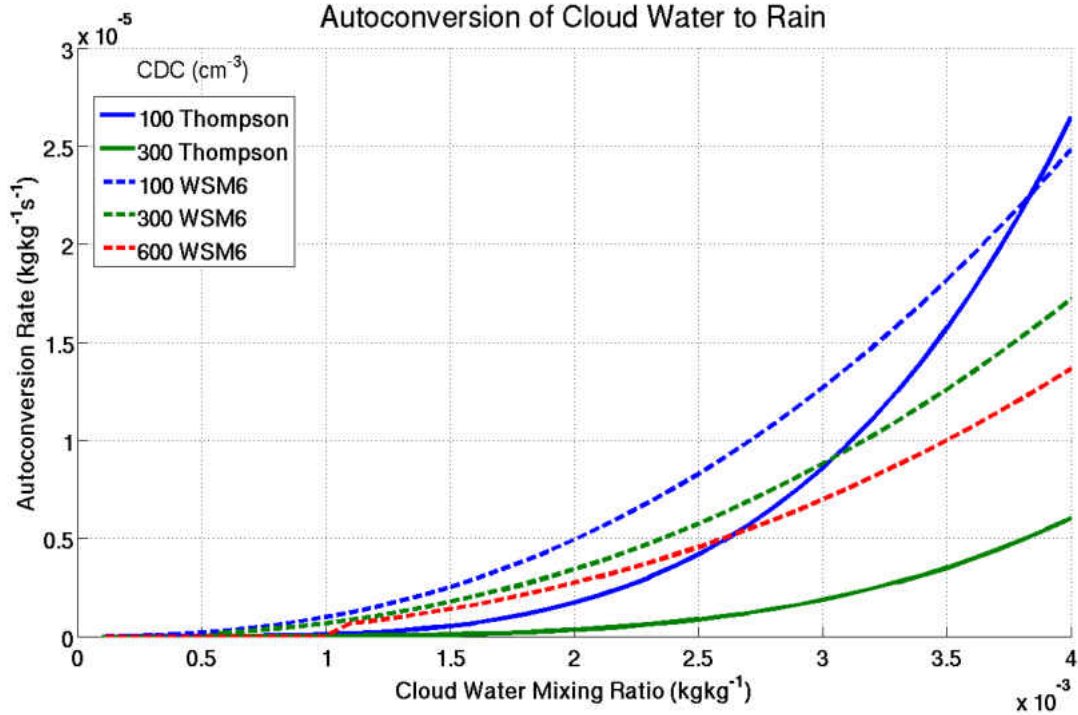


Figure 34: The Thompson (solid lines) and WSM6 (dashed lines) autoconversion rates of cloud water to rain hydrometeors for varying cloud water mixing ratios. Different colored lines indicate different cloud droplet concentrations.

Similar to the comparison of WSM6 simulations with differing CDCs, there is little snow mass present in the Thompson runs (beyond the initial maximum as described above), and very little cloud ice mass (Fig. 31). Therefore, the effects of heterogeneous freezing of drops should not be very pronounced. There should also be no direct differences between schemes as a result of the heterogeneous freezing rate as both the Thompson and WSM6 scheme employ the heterogeneous freezing function (Eq. (5.5)) from Biggs (1953).

5.6.3 Discussion of Horizontal Resolution Sensitivities

The comparison of the total mass of hydrometeors reveals that the peak in observed mass closely matched the initial peak in the 1 km simulation for both 5

(Fig. 28a) and 30 dBZ (Fig. 28b) reflectivity thresholds, respectively. However, for a comparison of liquid hydrometeors (Fig. 29), the high-resolution forecast generates too much rain. The higher-resolution 1 km simulations allow for higher vertical velocities as compared to 3 km simulations as the grid spacing is smaller, allowing for better resolved updrafts. Higher updraft velocities are able to generate more rain hydrometeors, and since the simulations generated too many objects at both 5 (Figs. 25a) and 30 dBZ (Figs. 25b) thresholds, this allows for generation of significantly more rain hydrometeors.

While the 1 km simulation contains too much rain mass, the high-resolution simulation still generates a maximum that qualitatively coincides better with the observation peak (Fig. 29). Similarly, the 1 km performs the best in terms of timing the peak in areal coverage of convection (Fig. 27). The improvement in timing and over-forecasting of objects are likely linked together, as the increased resolution allows for greater diabatic heating of the surface for individual grid squares. The increased heating allows the maximum temperature to rise locally, instead of being averaged across a wider area (i.e. 1 km² for 1 km grid spacing as opposed to 9 km² for 3 km grid spacing). The increased maximum temperatures more easily surpass the convective temperature, generating more numerous rising thermals that will generate more objects if the vertical motion is strong enough. Similarly, the improvement in timing comes from the fact that individual grid squares heat up more, allowing them to reach the convective temperature sooner.

Reducing the resolution out to 9 km does not provide any improvements in the biases found but instead amplifies the area bias and creates a weak bias in convection seen. The 9 km simulation did reduce the number of objects by instead creating less but significantly larger objects. Due to the coarseness of the grid spacing it is unable to generate strong updrafts and hence does not generate the higher intensity objects

seen in observations. In terms of location of objects, the 9 km simulation did place convective features in the correct locations similar to the other simulations.

5.7 Robustness of Sensitivity Analysis

To determine the robustness of the sensitivity analysis, three additional cases, important to the objectives of the POLCAST campaigns, are chosen: July 8, July 9, and July 12, 2012. Two of the cases, July 8 and 9, 2012, consist of weakly-forced isolated cells. On July 8 (15-22 UTC), the forecast results in a bad match to observations due to initiating convection too late and being outside the range of the radar. Since the model convection is outside the radar range, the analysis results in a general under-prediction of area and objects (Fig. 35), as only object edges and a few small cells are captured within the domain. However, even though the match between simulations and observations is bad, the same trends exist between model runs as is visible in the analysis case of Aug 4. The 1 km simulation and low CDC simulation generate the most objects, and likewise the 9 km and low CDC runs generate the most area. The high CDC and Thompson simulations contain the least area and objects.

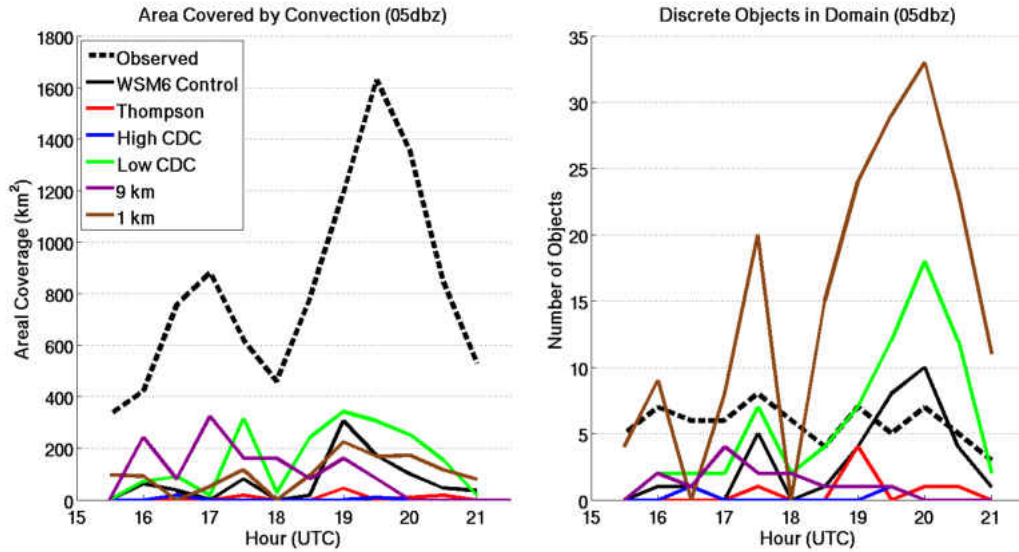


Figure 35: The area covered by convection (left) and the number of objects (right) greater than 5 dBZ, during the times of radar operation on the day of July 8, 2012.

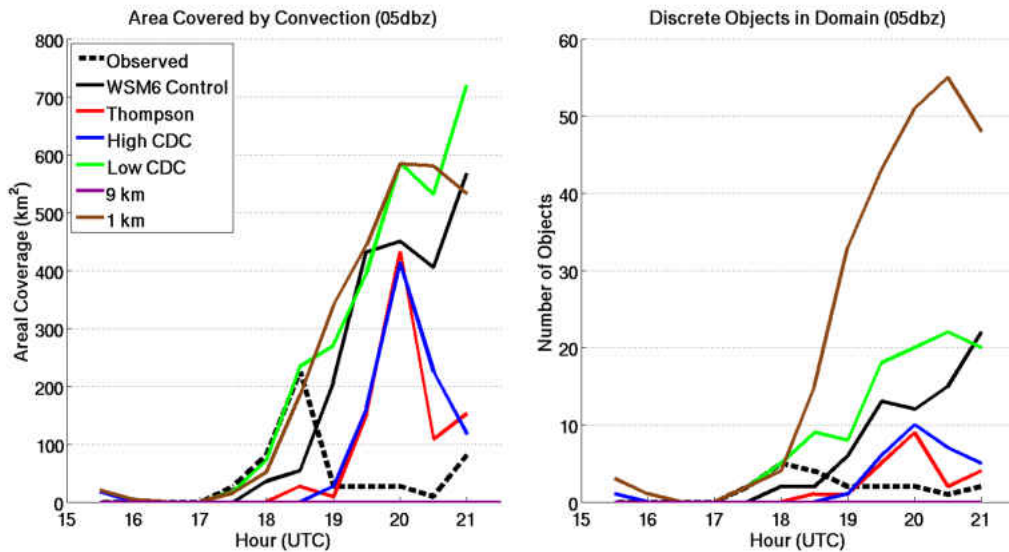


Figure 36: Similar to Figure 35 except for the day of July 9, 2012.

The case day of July 9 (15-22 UTC) is a more appropriate match, with a few small cells being located near the radar. While the simulations initiate the cells west of the cells visible in observations, the model cells are still within radar range. The 1 km and low CDC runs capture the initiation period very well, and initial area

and object count trends are very similar to observations (Fig. 36). However, both the 1 km and low CDC simulations continue to generate more objects and increase the areal coverage leading to an over-forecasting of both. The pre-convective surface DMT CCN concentrations at 0.3% supersaturation are around 1300 cm^{-3} . This CCN concentration would imply that the high CDC should perform better than the low CDC case. The high CDC run better represents the observed distribution in area and object counts qualitatively, but is too slow by two hours, as is the Thompson scheme. The analysis on both Aug. 4 and July 9 show that the Thompson scheme and high CDC simulations qualitatively perform better on weak-forcing cases, while increasing the resolution affects timing of events more. The 9 km simulation fails to generate the small cells.

The last case of July 12, 2012 (1530–24 UTC) is strongly-forced by an occluding frontal system. All simulations except the 9 km simulation generate a very large convective system, significantly over-predicting the area (Fig. 37a). The 9 km run fails to generate any significant convection in the region during the analysis time, which is attributed to the convective parameterization scheme. The Kain-Fritsch scheme compares the vertical velocity in a grid square to the mean vertical velocity across the domain, and associates a temperature perturbation at the lifted condensation level for that vertical velocity. The temperature perturbation is then utilized to determine whether convection should be represented in the domain. However, while the 9 km contains increased vertical velocities during the convective period visible in other simulations, it is broader and not as high in magnitude, which likely results in a lack of convection by the convective parameterization scheme.

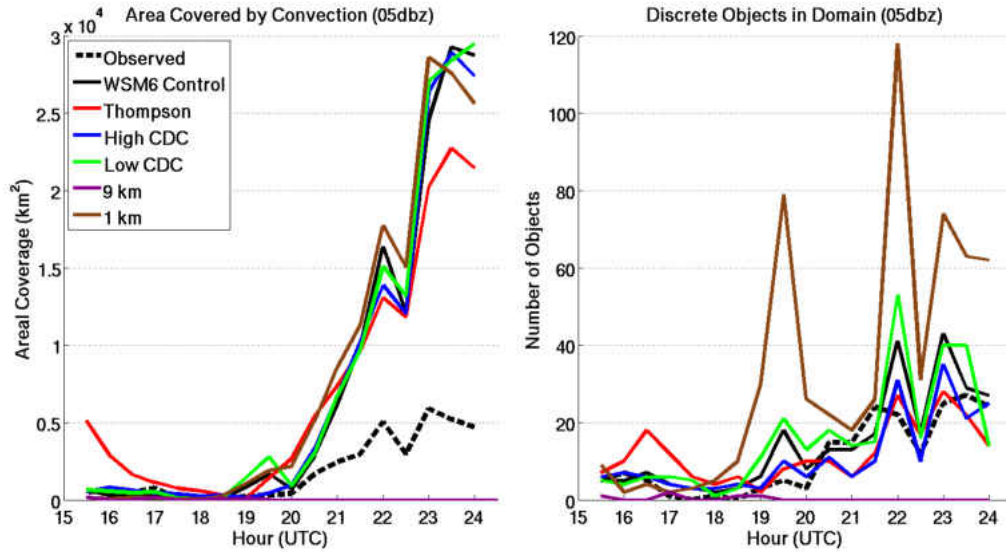


Figure 37: Similar to Figure 35 except for the day of July 12, 2012.

All simulations perform similarly in terms of area, other than the Thompson scheme. No major differences in area are objectively or subjectively noted between the different CDC runs, as they all look very similar. The lack of differences in the CDC are attributed to the forcing. Strong forcing generates large amounts of cloud water that is above any autoconversion activation thresholds regardless of CDC, and begins the autoconversion process to rain. Simulations are generally not far off in terms of objects counts early (Fig. 37b), but more significantly over-predict the number of objects during the most active period (22 - 00 UTC). A local minimum in the area and object counts is seen at 2230 UTC, because the radar scanning strategy switched to sector scans at that time. The sector scanning blocked a significant portion of the domain, lowering object area and counts. The same trends in the object count fields (i.e. 1 km over-predicting counts, low CDC run containing more objects than high CDC run, and the high CDC and Thompson schemes being similar) are noted as in the analysis case day and additional case day of July 9.

CHAPTER 6

DISCUSSION

6.1 Implications of the Objective Verification

The bulk statistics resulting from the objective forecast verification show that overall, too many objects are produced by both 3 km and 1 km simulations. Model simulations also over-forecast the areal coverage of convection, which the verification shows is due to both the sheer number of model objects and individual objects that are too large. These biases are seen across all reflectivity thresholds (5, 30, and 45 dBZ). Lastly, forecasts also generate convection that is too intense.

Kain et al. (2008) compared base reflectivity to simulated reflectivity at 1 km height at horizontal grid resolutions of 4 km and 2 km across 33 days, with matching physics, lateral boundary conditions, and simulated reflectivity calculation as this study. Kain et al. (2008) found that the number of objects significantly increased during the diurnal maximum with higher-resolution simulations and surpassed observations, similar to the 1 km simulations. However, when comparing the forecasted area to observed area at different intensity thresholds, Kain et al. (2008) found that forecasts under-predicted the area coverage at all analyzed lead times, except at the reflectivity threshold of 30 dBZ and lower intensities at the 18 hour lead time. Kain et al.'s (2008) area also dropped off at higher reflectivity values, indicating forecasts were too weak. The area and intensity results found in this study contradict Kain et al.'s (2008) results. Forecasts are found to cover more area than observations, and the over-forecasting of area becomes more significant as the intensity threshold increases.

The differences in area and intensity between Kain et al. (2008) and this study may be due to several factors. The domain of this study is smaller than the domain in Kain et al. (2008). However, the location of evaluation and the types of convection likely contributed the most to the differences. Strong and highly organized convection were not commonly observed in the Northern Great Plains area during POLCAST operations (Fig. 10-12). The model evaluation conducted by Kain et al. (2008), focused on a domain that was always the same size, but shifted daily to the location of the greatest severe threat. Simulations in Kain et al. (2008) were generally evaluated for vigorous convective cases. In addition, the location of evaluation was most commonly focused over the Southern Great Plains, which, especially during the evaluation period of April 18 to June 3, consisted of larger and/or stronger convection.

Another potential source of differences is the way simulated and observed reflectivity are compared. In this study, simulated reflectivity is compared to radar reflectivity at 1 km altitude. Kain et al. (2008) evaluated the simulated reflectivity field at 1 km AGL against the radar base scan reflectivity (elevation of 0.5°) field retrieved by the WSR-88D network. At the furthest range of 230 km (the max range of the short range base reflectivity product), the center of the beam is about 5.4 km high, with a beamwidth of 3.7 km. Therefore, during evaluation points at longer distances from the radar, the radar may have been detecting stronger convective cores (possibly hail aloft) than at 1 km altitude. Similarly there is a possibility of capturing the brightband. The higher beam height may also be expanding on rain areas if it detects stratiform regions aloft, which produce precipitation that does not reach the 1 km altitude. The differences stated showcase the difficulty in performing a quantitative verification when such variability may exist in differing locations.

Further examining the biases found in this study, forecasts are expected to over-predict the amount of precipitation, in terms of areal coverage and magnitude, which

has an important implication on quantitative precipitation forecasts. A majority of precipitation verification studies already performed found that simulations also tended to over-forecast the amounts of precipitation (e.g. Schwartz et al., 2010). For example, the previously referenced Davis et al. (2006b) study found that precipitation areas were too large and too intense. The results found in this study correspond directly with the results found by Davis et al. (2006b), in which the convective areas themselves were over-intensified and covered too much area. Likewise, their finding of precipitation areas lasting longer than observed was shown by the case study, in which simulated convection was still ongoing and being actively generated while no observed convection existed. Unfortunately due to varying times of radar operation, an evaluation into the timing of convective events would have not produced a robust analysis. However, a further avenue of research could be the utilization of the nearby WSR-88D radar, which has limited downtime. The number of objects may or may not be misrepresented in the accumulated precipitation field, as precipitation forecasts are typically verified over several hours, which can eliminate individual object signals if precipitation falls on locations that previously received precipitation. However, by looking at hourly accumulated precipitation totals, Clark et al. (2014) did find that too many precipitation objects were generated.

Another study by Davis et al. (2009) found that by comparing the total area forecasted by 4 km ARW and 4.5 km Non-hydrostatic Mesoscale Model (NMM) WRF simulations divided by the total area observed resulted in both dynamical cores over-forecasting the area covered by rain, with NMM having the poorest performance. Their results also showed that there were too many precipitation areas. By comparing precipitation totals across well matched forecasted and observed objects at the 90th percentile, both ARW and NMM simulations generated over 1.5 mm more rain than observations (over 15% more). Davis et al. (2009) also found that more

intense rainfall features are over-predicted by the models, and had no corresponding regions in observations. One can deduce that these precipitation biases show similar patterns to the biases found in the simulated reflectivity verification performed in this study. The verification shows that the model over-intensifies convection, with significantly more reflectivity areas above the 45 dBZ threshold, that would result in the over-generation of rain. Nevertheless, the next step would be to perform an analysis comparing precipitation amounts to determine how the biases in the simulated reflectivity translate down into the precipitation field.

In their concluding remarks, Davis et al. (2009) mentioned that model performance had significant daily variation, which is also true of this study. However, some general conclusions can be added by subjective analysis on several cases. When a MCS is present in both domains, simulations over-intensify the convective regions and generate convective cores that are too large (e.g. Fig. 38), which is especially true of linearly organized systems. Similarly, stratiform rain regions are generally poorly represented, which is a well-known problem in single-moment schemes such as WSM6 (e.g. Morrison et al., 2009; Luo et al, 2010; Wu et al., 2013). While forecasts of strongly-forced events are overall too large and overly intense, weaker-forced convective events show a larger daily variation in skill.

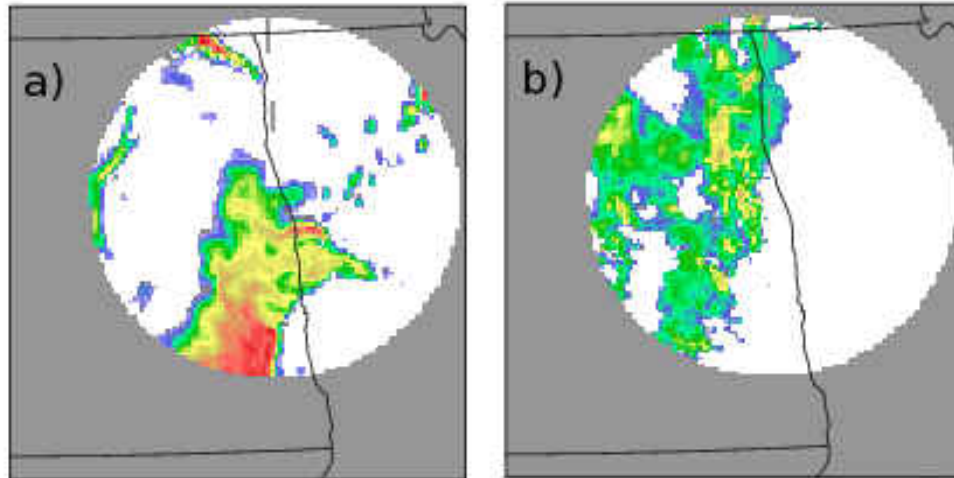


Figure 38: MODE comparison of the 3 km grid spacing a) simulated reflectivity and b) observed reflectivity fields at 1 km AGL altitude at 22 UTC on Aug. 3, 2012.

Since two simulations with spacing of 3 km and 1 km were operational during POLCAST4, an analysis was performed on different resolutions. Based upon the comparison of the matching domains and times of the 3 km and 1 km simulations, no noteworthy objective increases in skill are seen in 1 km simulations, and the area bias seen are almost identical. Burghardt et al. (2014) investigated convective initiation within high resolution simulations utilizing object-based methods. For all forecasts that had matches in the observational field, they found that there was too much area covered by objects at even the 429 m grid spacing. They raised an interesting point noting that these area biases are commonly seen in studies of convection-resolving models, even at an order of magnitude difference in grid resolutions, and likely suggest that they are resultant of the current handling of convection and not an issue of the resolution. While no improvement in the area bias is seen, the bias in the number of objects generated is amplified due to very small ($<45 \text{ km}^2$) cells. These small cells are also over-intensified, which leads to an even worse over-intensification for cells that size. Hence, in terms of the biases found, increasing the resolution to 1 km is found to

not be worth the computational resources. However, the 1 km does add significantly more detail to convection, which may ultimately aid forecasters in subjective analysis of potential storm types and hazards.

6.2 Implications of the Sensitivity Study

The sensitivity study shows that changing the CDC has a significant impact on simulated convection, supporting the results found in Cohen and McCaul Jr. (2006) and Thompson et al. (2004). Altering number concentrations strongly affects the areal coverage and number of object biases. The intensity bias is minimally influenced. The area bias is affected by both changing the number of objects and changing the areal extent of convection.

The high CDC simulation, which uses CDCs derived from CCN observations, results in improved skill in areal coverage and object counts. Likewise, there is an improvement in the timing of generation and dissipation of rain hydrometeors in the domain. Hence, it is important set the CDC to values observed, or at the very least climatologically relevant values, reaffirming Thompson et al.'s (2008) guidance. However, the high CDC simulation does not generate enough rain hydrometeors and is therefore unable to generate as many convective cores above 30 dBZ as are present in observations. Since the control simulation generated similar mass of rain hydrometeors as compared to observations, the rate of conversion of cloud water to rain may be adequate in the control simulation. However, the activation of the autoconversion function as defined by the critical cloud water mixing ratio in the Heaviside step function (Eq. (5.4)) may be set too low. The reason the high CDC case produces less objects is because the critical mixing ratio increases with the increase in CDC. However, increasing the CDC lowers the autoconversion rate, which generates less rain hydrometeor mass than lower CDCs with the same mass of cloud water present.

Setting the CDC to observed values is recommended, but while keeping to the current nature of WSM6 utilizing the Heaviside function, the effects of altering the mean droplet radius should also be investigated. By increasing the mean radius and holding the CDC constant, it is expected that the autoconversion rates remain the same but the required cloud water mass increases, theoretically leading to less objects and reduced areal extents of objects. The greater critical mixing ratio would also suppress rain activation, addressing potential issues with convection being generated too early as seen on the case day.

The strong forcing case (July 12, 2012) shows that changes in CDC have a lesser effect on intense convection. By increasing the CDC, the number objects still slightly decreases due to less smaller objects being generated, but the large convective system was relatively unchanged. Therefore, there are no major differences between the control, low, and high CDC cases and the areal coverage and rain hydrometeor mass are all relatively similar. This result is not surprising, as increasing the CDC influences when autoconversion begins due to the critical cloud water mixing ratio in WSM6. Under strong forcing, large amounts of cloud water are generated and the autoconversion process occurs regardless of CDC, covering about the same amount of area if strong uplift covers the same region.

Because of the pronounced effects of the CDC on the skill of high resolution forecasts in weak forcing regimes, it is ideal for microphysical schemes to explicitly predict the number of cloud drops and their corresponding radii. Two-moment schemes have been found to perform better than their one-moment counterparts in terms of generating simulations that are closer to corresponding observations (e.g. Bryan and Morrison, 2012; Morrison et al., 2009; Milbrandt and Yau, 2005). The improvements are attributed to the fact that the number concentration can be used to determine first order size information that is lacking in single moment microphysics.

Instead, single moment schemes need more assumptions (e.g. fixed slope intercept, set distribution of particles) to create a way to describe the growth rate of drops (Chen and Liu, 2004). Further advances have combined aerosol and CCN data in order to simulate explicit nucleation of cloud droplets within WRF (e.g. Ming et al, 2007; Morrison et al., 2005). More recently, Thompson and Eidhammer (2014) used seven years of Goddard Chemistry and Aerosol Radiation and Transport (GOCART; Ginoux et al., 2001) model runs to generate a monthly aerosol climatology. They used the aerosol climatology as sources of nucleation for cloud water and cloud ice within the Thompson microphysical scheme. They found the new scheme accurately simulated a large winter system across the continental United States. A future point of investigation would be to evaluate the same dataset used in the verification but by using simulations using a two-moment version of the same microphysical scheme, and using the Thompson scheme that utilizes the new aerosol data to determine the effects on the biases found.

Investigating the changes between microphysical schemes shows that the number of object and intensity biases are affected more than the areal coverage. The Thompson simulation predicts less objects than WSM6 (better matching observations), but generates similar areal coverage and similar mass of hydrometeors at the 5 dBZ reflectivity threshold. However, Thompson generates much larger objects at the 30 dBZ threshold than both the observations and WSM6, leading to an over-intensification of convection. The over-intensification is due to the autoconversion formulation used and due to the default CDC of 100 cm^{-3} used. At higher cloud water mixing ratios, the Thompson autoconversion rates were significantly higher than both the WSM6 control and low CDC cases, causing more rain generation. If the Thompson scheme is set to 300 cm^{-3} (i.e. representing more continental conditions), then it is expected that the Thompson simulation would actually better resemble the

observations in terms of intensity, since the autoconversion rate at a CDC of 300 cm^{-3} is much lower than WSM6 at a CDC of 300 cm^{-3} .

Changing the microphysical scheme shows the large differences that can arise just by utilizing different formulations of how to determine the rate of autoconversion. The Thompson scheme is found to be more sensitive to the CDC than WSM6, and at CDCs of 300 cm^{-3} , no reasonable CDC within WSM6 can reproduce the Thompson rate. While the WSM6 scheme incorporates the Heaviside step function to begin autoconversion while assuming a monodispersive droplet spectrum (constant diameter), the version of the Thompson scheme generates differing gamma distributions using the constant set CDC. The gamma distribution is used to better represent the sizes of varying concentrations of cloud droplets (i.e. higher concentrations of smaller sized particles for high CDC values to represent continental conditions). The shape parameter describing the gamma distribution is utilized to determine the diameters used in the autoconversion rate. Therefore, CDC can significantly influence forecast skill, but also the formulation chosen to represent the autoconversion process will significantly affect forecasts. The microphysical processes that occur need to be understood better in order to develop more accurate and more consistent schemes.

While both the WSM6 and Thompson simulations still generate too much area, are too intense, and have average object sizes that are too large, the decreasing trend in object counts and area coverage by the Thompson scheme is more representative of observations. The better representation of observations by Thompson is also seen on the additional cases (July 9 and July 12). However, to get a more conclusive argument, the seasonal simulations verified would have to be re-run using the Thompson scheme with a CDC of 300 cm^{-3} .

The results between different microphysical schemes was also seen by Jankov et al. (2009). By comparing different microphysical scheme simulations of brightband

and non-brightband precipitation events, Jankov et al. (2009) found that both simulated reflectivity and accumulated precipitation was over-predicted by all schemes, including WSM6 and Thompson. WSM6 was also found to generate and contain too much graupel, which over-intensified convection and led to more precipitation. Jankov et al. (2009) found that WSM6 generated significant amounts of cloud water and rain hydrometeors, which is most likely due to the constant hydrometeor intercept being set high, keeping large amounts of drops (Bryan and Morrison, 2012). Since the autoconversion process in WSM6 is strongly affected by amounts of cloud water (Eq. (5.3)), it would be beneficial to investigate how accurately the model depicts cloud water. Gallus and Pfiefer (2007) also found that both WSM6 and Thompson schemes over-predicted the maximum reflectivity and the average simulated reflectivity values when analyzing a simulated squall line. Thompson was found to be the worst with overestimation of reflectivity, which was found to be true in the case day (although attributed to the low CDC of 100 cm^{-3}). Gallus and Pfiefer (2007) concluded that all runs produced too many areas with high reflectivity values, and not enough areas with lower reflectivity values.

All forecast biases are affected by changes in resolution. Reducing the resolution to 9 km did remove the number of objects present, but this improvement is accomplished by generating objects that are too large and not intense enough. The 9 km simulation is too coarse and does not generate strong enough updrafts, since vertical velocities are effectively average across a wide area (81 km^2). Instead, larger, more broad regions of lift are generated, leading to larger and weaker convection with too much rain being generated. Due to the coarse resolution, the 9 km also misses a weakly forced event with a few small cells (July 9). Similarly for the additional strongly-forced case on July 12, the 9 km fails to produce a convective system and several smaller associated objects in the domain. The convective parameterization

scheme fails to trigger convection along a slowly advancing frontal system, showing that the 3 km resolution simulation has more skill than the 9 km simulation. Additionally, no reliance has to be placed on convective parameterizations for 3 km simulations. Even though convection at scales of 3 km is still not fully resolved, it has been showed to be more skillful than 9 km spacing with convective parameterizations.

It is difficult to determine what causes the over-generation of small, almost individual grid-square sized objects within 1 km simulations, since this resolution approaches a limit of current parameterizations. A major issue in simulations of increasingly high resolution is defining a clear-cut scale of when phenomena are resolvable, at what resolutions should parameterizations be applied, and at what scales do parameterizations not produce accurate representations of the processes they are tasked with representing. When processes become partially resolvable by the model but parameterizations are still required, the situation is complicated further, and the feedback between the two may lead to erroneous forecasts. For this reason, current parameterizations may not be suitable for higher-resolution simulations (e.g. Coniglio et al., 2010). An example is how the planetary boundary layer (PBL) parameterizations generate turbulence with increased resolution. The verification shows that the 1 km simulation generates too many objects, and similar initial trends are seen across the case day when attempting to increase the resolution to 333 m. With such high-resolution simulations, the model is able to resolve larger-scale turbulence located in the PBL. There is no longer a clear separation of scales between resolvable and sub-grid turbulence, which is implied in the use of PBL parameterizations. The effects of the resolved turbulence may combine with the circulations generated by the PBL scheme and could lead to a double-feedback generation of turbulence within the PBL. The additional feedback of turbulence may aid parcels in becoming deep moist convection and may have resulted in the generation of numerous small objects in the

domain. Hence it is important when choosing a fine resolution to evaluate physics and parameterization options and how they may influence each other. More studies need to be done investigating model physics and the intricate combined effects of various physics in these gray areas.

It is possible that the 1 km simulation is generating too many objects due to vertical resolution that is too coarse. While changes in horizontal resolution are investigated, changes in vertical resolution are not. Increases in vertical resolution may provide better depictions of convection, especially convection that is weakly-forced, due to better refinement in the location of capping features and better resolved fluxes. A poorly resolved (or unresolved) cap may allow a parcel to become deep moist convection when it should be inhibited instead. However, Aligo et al. (2009) showed the forecast skill in QPF actually decreased with increasing resolution. Increasing the vertical resolution throughout the model led to a greater over-prediction of rainfall, causing larger biases. QPF skill only increased in some instances in weakly forced cases when more levels were added near the surface. Overall, QPF skill increased by decreasing the vertical resolution. Aligo et al. (2009) speculated that lower skill with increasing resolution was due to more frequent grid cell saturation, since enhanced upward velocities and higher humidity values were present before convective initiation. Nevertheless, because there was some improvement seen in weakly forced convection and because the vertical resolution can significantly influence precipitation, which correlates back to convective strength, further investigation into vertical resolution may potentially address some of the biases found.

In terms of resolution, the conclusion reached in this study reflects the concluding remarks by Schwartz et al. (2009). There was no significant increase in skill with increasing the resolution below 3 km or 4 km resolutions. Instead, attention should now be placed on developing physics, parameterizations, and algorithms suit-

able for implementation in high-resolution forecasts that are now in operational use. Additionally, new data assimilation techniques have been show promising results, and future research can be placed on efforts to better ingest observational data into the model.

CHAPTER 7

CONCLUSIONS

An objective method to quantify forecast skill by verifying simulated reflectivity against radar reflectivity by utilizing an object-based method approach is developed. The method is applied to high resolution (3 km and 1 km) forecasts across two summers in the Northern Great Plains region. Results show that about two-thirds of all forecasts successfully predict convection, or lack thereof, at a reflectivity threshold of 5 dBZ, with the leading loss of skill attributed to forecast misses. At higher reflectivity thresholds, the number of misses decrease, but the number of false alarms increase, which indicates that forecasts are over-intensifying convection. All simulations show significant over-prediction of the number of objects at all sizes and intensities. The best model performance is at the 30 dBZ reflectivity threshold; however, the 3 km simulations still over-predict the total amount of objects by 37%. Simulations have the least skill at the 45 dBZ reflectivity threshold, with a 4:1 model to radar object ratio, which again shows simulations generate convection that is too intense.

Simulations consistently over-forecasted the area coverage of convection. The over-forecasting of area is mainly attributed to the generation of too many objects by simulations. Additionally, an analysis of individual objects shows that 3 km forecasts over-predict the average areal extent of objects 66% of the time, further amplifying the differences in area.

A direct comparison of 3 km simulations against 1 km simulations at equivalent times and domain sizes shows that there are no noteworthy increases in skill with further increases in resolution. The success rates and areal coverage statistics are relatively similar between 3 km and 1 km forecasts. Instead, the 1 km simulations significantly over-forecast the number of small objects (smaller than 45 km²) by over 300% and also over-intensify them.

A sensitivity study is performed on a case day containing weakly forced convection to determine which parameters are susceptible to generating the biases seen in the verification. The sensitivity study investigates the variation in the number of cloud droplets, changing the microphysical scheme, and changing resolutions.

Altering the CDC has the greatest effect on the area and object biases. Raising the CDC values to derived pre-convective values improves the forecast biases and improves the timing in generation and dissipation of rain hydrometeors. Lowering the CDC generates more objects and significantly more areal coverage by convection. The area bias is amplified by the generation of more objects and by extending individual object areas. The differences are attributed to the autoconversion process within WSM6. Lowering the CDC increases the rate of conversion of cloud water to rain, and similarly lowers the critical mass of cloud water required to begin the autoconversion process. These changes lead to more objects and expands on object areas since the required mass of cloud water to generate rain is now decreased. Raising the CDC has the opposite effect, improving the biases. However, the high CDC autoconversion rate did not generate enough rain hydrometeors. Changes in CDC within WSM6 are found to only significantly affect weak-forcing events, as there are no differences between the simulations during a strongly-forced case. The lack of differences is likely due to strong forcing generating large amounts of cloud water, above any critical activation thresholds, and instantly initiating the autoconversion process.

Changing microphysical schemes from WSM6 to Thompson has the most significant impact on object counts and intensity biases. The Thompson scheme reduces the number of objects across the evaluation time period to almost match observations, resulting in a reduction of the total area of convection. However, the average areal extent of individual object increases in the Thompson simulation. At the 30 dBZ reflectivity threshold, Thompson contains significantly more areal coverage than both observations and WSM6, indicating that the Thompson simulation is over-intensifying convection. The over-intensification is attributed to the Thompson simulation using the default CDC of 100 cm^{-3} , which enables the autoconversion process to generate significantly too much rain at higher masses of cloud water. There is a significant difference between the WSM6 and Thompson autoconversion schemes. WSM6 utilizes the autoconversion from Tripoli and Cotton (1980) and assumes a monodisperse population of drops. Thompson uses an altered version of the Berry and Reinhardt (1974) formulation, using the fixed CDC to generate different gamma distributions in an attempt to represent differing environments. The differences between autoconversion formulations between the two schemes have effects on the same order of magnitude as changing the CDC within each individual scheme. Likewise, for low cloud water mixing ratios, no realistic CDC within WSM6 can represent the lower autoconversion rates in the Thompson simulation.

The significant differences between the schemes shows the need for more understanding of microphysical processes that occur within clouds. Currently, because the microphysical processes are not fully understood, parameterizations are essentially tasked with addressing the issues of what typical cloud environments should look like instead of parameterizing the processes themselves. As horizontal resolutions are becoming fine enough to begin resolving individual convective clouds, the amount of error within forecasts that can be contributed to microphysical inconsistencies be-

comes significant, which is especially true of weakly-forced convection or possible convective initiation. The autoconversion process affects the initiation and evolution of convection, yet significant uncertainty in the warm-rain processes remains. More in-depth field campaigns are needed to collect data utilizing several platforms simultaneously to get a fundamental understanding these microphysical processes. Droplet spectrum analysis within the warm clouds combined with ample radar coverage and radiative retrievals over time will help constrain our understanding of the processes occurring within the clouds.

The analysis of differing resolutions of 9 km, 3 km, 1 km, and 333 m, shows that all three biases are affected. Using 9 km grid spacing instead of 3 km grid spacing reduces object numbers, but significantly expands the areal coverage of objects. The 9 km simulation is unable to resolve smaller updrafts and hence, smaller objects. Instead, wide areas of vertical motion create large objects in the region. Since the 9 km simulation is so coarse, it is unable to simulate proper updraft strength, which results in simulated convection that is significantly weaker than convection in observations. Likewise, because the 9 km simulations depend on cumulus parametrizations to produce convection, it fails to produce both a weakly-forced convective event and a strongly-forced event visible in two additional case days. The 1 km simulation generates numerous amounts of small objects (by up to six times more than observations). These small convective objects are also too intense, matching the results from the verification. The 1 km simulation generates too much rain in the domain because of the small intense objects. The area coverage of convection did improve over the control run across the main case day, which is not seen in the bulk statistical verification. However, application to the two additional case days shows that the 1 km did not improve in areal coverage. Unfortunately, the 333 m simulation becomes unstable half-way through the simulation. However, initial trends suggest that the

results are similar to the 1 km (i.e. too many small objects that are too intense), but even further amplified. Evaluation of the different resolutions shows that there is no significant increase in skill that would warrant increasing resolution further from 3 km spacing. However, the 3 km spacing did show significant improvement over the 9 km simulations with convective parameterizations.

Therefore, for the ongoing cloud studies focusing on weakly-forced convection in North Dakota and surrounding regions, this study recommends that simulations:

- retain 3 km grid spacing,
- change to the Thompson microphysical scheme,
- and set the cloud droplet concentration value to 300 cm^{-3} within the Thompson scheme.

Since setting the CDC to daily values or pre-convective values is not possible, the CDC of 300 cm^{-3} is chosen as the best representation of the region, appropriate for the Thompson scheme's autoconversion function.

REFERENCES

- Alderman, E. J., and K. K. Droegemeier, 2002: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters. *Mon. Wea. Rev.*, **130**, 2671-2691.
- Alexander, C. R., S. S. Weygandt, T. G. Smirnova, S. Benjamin, P. Hofmann, E. P. James, and D. A. Koch, 2010: High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2010 convective season. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 9.2.
- Aligo, E. A., W. A. Gallus Jr., and M. Segal, 2009: On the Impact of WRF Model Vertical Grid Resolution on Midwest Summer Rainfall Forecasts. *Wea. Forecasting*, **24**, 575-594.
- Andreae, M. O., D. Rosenfeld, P. Artaxo, A. A. Costa, G. P. Frank, K. M. Longo, and M. A. F. Silva-Dias, 2004: Smoking rain clouds over the Amazon, *Science*, **303**, 1337-1342.
- Asai, Tomio, 1965: A Numerical Study of the Air-Mass Transformation over the Japan Sea in Winter, *J. of the Meteor. Soc. Japan*, **43**, 1-15.
- Berry, E. X., and R. L. Reinhardt, 1974: An analysis of cloud droplet growth by collection. Part II: Single initial distributions. *J. Atmos. Sci.*, **31**, 2127-2135.
- Bigg, E. K., 1953: The supercooling of water. *Proc. Phys. Soc. London*, **B66**, 688-694.

Black, T. L., 1994: The new NMC mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting*, **9**, 265-278.

Bryan, G. H., and H. Morrison, 2012: Sensitivity of a Simulated Squall Line to Horizontal Resolution and Parameterization of Microphysics. *Mon. Wea. Rev.*, **140**, 202-225.

Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394-2416.

Burghardt, B. J., C. Evans, and P. J. Roebber, 2014: Assessing the Predictability of Convection Initiation in the High Plains Using an Object-Based Approach. *Wea. Forecasting*, **29**, 403-418.

Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141-154.

Chen, J.-P., and S.-T. Liu, 2004: Physically based two-moment bulkwater parametrization for warm-cloud microphysics. *Quart. J. Roy. Meteor. Soc.*, **130**, 51-78.

Clark, A. J., R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong, 2014: Application of Object-Based Time-Domain Diagnostics for Tracking Precipitation Systems in Convection-Allowing Models. *Wea. Forecasting*, **29**, 517-542.

Climate Prediction Center/National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, and Joint Office for Science Support/University Corporation for Atmospheric Research, 2000: NCEP/CPC 4km Precipitation Set, Gage and Radar. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. Accessed 25 Nov 2013.

Cohen, C., and E. W. McCaul Jr., 2006: The Sensitivity of Simulated Convective Storms to Variations in Prescribed Single-Moment Microphysics Parameters that Describe Particle Distributions, Sizes, and Numbers. *Mon. Wea. Rev.*, **134**, 2547-2565.

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

Davis, C. A., B. G. Brown, R. Bullock, and J. H.-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.

Davis, C., B. Brown, and R. Bullock, 2006a: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Mon. Wea. Rev.*, **134**, 1772-1784.

Davis, C., B. Brown, and R. Bullock, 2006b: Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Mon. Wea. Rev.*, **134**, 1785-1795.

Delene, D. J., C. Grainger, P. Kucera, D. Langerud, M. Ham, R. Mitchell, and C. Kruse, 2011: The Second Polarimetric Cloud Analysis and Seeding Test. *J. of Wea. Mod.*, **43**, 14-28.

Deng, Aijun, and D. R. Stauffer, 2006: On Improving 4-km Mesoscale Model Simulations. *J. Appl. Meteor. Climatol.*, **45**, 361-381.

Droegemeier, K. K., G. Bassett, and M. Xue, 1994: Very high-resolution, uniform-grid simulations of deep convection on a massively parallel processor: Implications for small-scale predictability. *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 376-379.

Duda, J. D., and W. A. Gallus, 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*, **28**, 994-1018.

Ebert, E. E., and W. A. Gallus Jr., 2009: Toward Better Understanding of the Contiguous Rain Area (CRA) Method for Spatial Forecast Verification. *Wea. Forecasting*, **24**, 1401-1415.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrology*, **239**, 179-202.

Fiori, E., A. Parodi, and F. Siccardi, 2011: Uncertainty in prediction of deep moist convective processes: Turbulence parameterizations, microphysics and grid-scale effects. *Atmos. Research*, **100**, 447-456.

Freud, E., D. Rosenfeld, D. Axisa, and J. R. Kulkarni, 2011: Resolving both entrainment, mixing and number of activated CCN in deep convective clouds, *Atmos. Chem. Phys.*, **11**, 12887-12900.

Gallus, W. A., 2010: Application of Object-Based Verification Techniques to Ensemble Precipitation Forecasts. *Wea. Forecasting*, **25**, 144-158.

Gallus Jr., W. A. and Pfeifer, M., 2008: Intercomparison of simulations using 5 WRF microphysical schemes with dual-Polarization data for a German squall line, *Adv. Geosci.*, **16**, 109-116.

Gao, J., and D. J. Stensrud, 2012: Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hydrometeor classification. *J. Atmos. Sci.*, **69**, 1054-1065.

Gilleland, E, D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of Spatial Forecast Verification Methods. *Wea. Forecasting*, **24**, 1416-1430.

Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. Holben, O. Dubovik, and S.-J. Lin, 2001: Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geo-phys. Res.*, **106**, 20255-20273.

Hanssen, A. W., and W. J. A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Koninklijk Ned. Meteor. Instit., Meded. Verhand.*, **81**, 2-15.

Hegg, D. A., D. S. Covert, H. H. Jonsson, and R. K. Woods, 2012: A simple relationship between cloud drop number concentration and precursor aerosol concentration for the regions of Earth's large marine stratocumulus decks. *Atmos. Chem. Phys.*, **12**, 1229-1238.

Hitchens, N. M., M. E. Baldwin, and R. J. Trapp, 2012: An Object-Oriented Characterization of Extreme Precipitation-Producing Convective Systems in the Mid-western United States. *Mon. Wea. Rev.*, **140**, 1356-1366.

Hong, Song-You, and J.-O. J. Lim, 2006: The WRF single-moment microphysics scheme (WSM6), *J. Korean Meteor. Soc.*, **42**, 129-151.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318-2341.

Hudson, J. G., and S. Noble, 2014: CCN and vertical velocity influences on droplet concentrations and supersaturations in clean and polluted stratus clouds. *J. Atmos. Sci.*, **71**, 312-331.

Hudson, J. G., 1983: Effects of CCN on stratus deck. *J. Atmos. Sci.*, **40**, 480-486.

Jankov, I., J.-W. Bao, P. J. Neiman, P. J. Schultz, H. Yuan, and A. B. White, 2009: Evaluation and Comparison of Microphysical Algorithms in ARW-WRF Model Simulations of Atmospheric River Events Affecting the California Coast. *J. Hydrometeorol.*, **10**, 847-870.

Jiusto, J. E., 1966: Maritime concentration of condensation nuclei. *J. Rech. Atmos.*, **2**, 245-250.

Johnson, A., X. Wang, F. Kong, and M. Xue, 2013: Object-Based Evaluation of the Impact of Horizontal Grid Spacing on Convection-Allowing Forecasts. *Mon. Wea. Rev.*, **141**, 3413-3425.

Johnson, A., and X. Wang, 2013: Object-Based Evaluation of a Storm-Scale Ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Mon. Wea. Rev.*, **141**, 1079-1098.

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

Kain, J. S., 2004: The Kain-Fritsch Convective Parameterization: An Update. *J. Appl. Meteor.*, **43**, 170181.

Keil, C., and G. C. Craig, 2007: A displacement-based error measure applied in a regional ensemble forecasting system. *Mon. Wea. Rev.*, **135**, 3248-3259.

Koch, S. E., B. S. Ferrier, J. S. Kain, M. T. Stoelinga, E. J. Szoke, and S. J. Weiss, 2005: The use of simulated radar reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts. *Preprints, 12th Conference on Mesoscale Processes*. October 24-28, 2005, Albuquerque, NM.

Kucera, P. A., A. Theisen, and D. Langerud, 2008: Polarimetric Cloud Analysis and Seeding Test (POLCAST), *J. Wea. Mod.*, **40**, 64-76.

Leitch, W.R., G. A. Isaac, J. W. Strapp, C. M. Banic, and H. A. Wiebe, 1992: The relationship between cloud droplet number concentrations and anthropogenic pollution: Observations and climatic implications. *J. Geo-phys. Res.*, **97**, 0148-0227.

Lin, J. C., T. Matsui, R. A. Pielke Sr., and C. Kummerow, 2006: Effects of biomass-burning derived aerosols on precipitation and clouds in the Amazon Basin: a satellite-based empirical study. *J. Geo-phys. Res.*, **111**, D19204.

Loftus, A. M., and W. R. Cotton, 2014: Examination of the CCN impacts on hail in a simulated supercell storm with triple-moment hail bulk microphysics. *Atmos. Research*, **147**, 183-204.

Luo, Y., Y. Wang, H. Wang, Y. Zheng, and H. Morrison, 2010: Modeling convective-stratiform precipitation processes on a Mei-Yu front with the weather research and forecasting model: Comparison with observations and sensitivity to cloud microphysics parameterizations, *J. Geophys. Res.*, **115**, D18117.

Marzban, C., and S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, **21**, 824-838.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407-430.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovi?, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343-360.

Micheas, A. C., N. I. Fox, S. A. Lack, and C. K. Wickle, 2007: Cell identification and verification of QPF ensembles using shape analysis techniques. *J. Hydrol.*, **343**, 105-116.

Milbrandt, J. A., and M. K. Yau, 2005: A multimoment bulk microphysics parameterization. Part II: A proposed three-moment closure and scheme description. *J. Atmos. Sci.*, **62**, 3065-3081.

Ming, Y., V. Ramaswamy, L. J. Donner, V. T. Phillips, S. A. Klein, P. A. Ginoux, and L. W. Horowitz, 2007: Modeling the interactions between aerosols and liquid water clouds with a self-consistent cloud scheme in a general circulation model. *J. Atmos. Sci.*, **64**, 1189-1209.

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

Morrison, H., and J. O. Pinto, 2005: Mesoscale modeling of springtime arctic mixed-phase stratiform clouds using a new two-moment bulk microphysics scheme. *J. Atmos. Sci.*, **62**, 3683-3704.

Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991-1007.

NOAA National Climatic Data Center, State of the Climate: Wildfires for August 2012. Accessed on September 9, 2014. [Available Online at <http://www.ncdc.noaa.gov/sotc/fire>]

Petch, J. C., 2006: Sensitivity studies of developing convection in a cloud-resolving model. *Quart. J. Roy. Meteor. Soc.*, **132**, 345-358.

Petch, J. C., A. R. Brown, and M. E. B. Gray, 2002: The impact of horizontal resolution on the simulations of convective development over land. *Quart. J. Roy. Meteor. Soc.*, **43**, 2619-2635.

Roberts, N. M., 2005: An investigation of the ability of a storm-scale configuration of the Met Office NWP model to predict flood-producing rainfall. Forecasting Research Tech. Rep. 455, Met Office, 80 pp.

Rosenfeld, D. and W. L. Woodley, 2000: Convective clouds with sustained highly supercooled liquid water down to 37C. *Nature*, **405**, 440-442.

Rosenfeld, D., 1999: TRMM observed first direct evidence of smoke from forest fires inhibiting rainfall. *Geophys. Res. Lett.*, **26**, 3105-3108.

Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M.C. Coniglio, and M. S. Wandishin, 2010: Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance with Small Ensemble Membership. *Wea. Forecasting*, **25**, 263-280.

Schwartz, C. S., J. S. Kain, S.J. Weiss, M. Xue, D.R. Bright, F. Kong, K. W. Thomas, J. J. Levit, and M.C. Coniglio, 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.

Segal, Y., M. Pinsky, and A. Khain, 2007: The role of competition effect in the raindrop formation. *Atmos. Res.*, **83**, 106118.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the advanced research WRF version 3. NCAR Technical Note TN-475. 113 pp.

Skamarock, W. C., M. L. Weisman, and J. B. Klemp, 1994: Three-dimensional evolution of simulated long-lived squall lines. *J. Atmos. Sci.*, **51**, 2563-2584.

Tartaglione, N., S.Mariani, C.Accadia,A. Speranza, and M. Casaioli, 2005: Comparison of rain gauge observations with modeled precipitation over Cyprus using contiguous rain area analysis. *Atmos. Chem. Phys.*, **5**, 2147-2154.

Tao, W.-K., X. Li, A. Khain, T. Matsui, S. Lang, and J. Simpson, 2007: Role of atmospheric aerosol concentration on deep convective precipitation: Cloud-resolving model simulations, *J. Geophys. Res.*, **112**, D24S18.

Thompson, G., and T. Eidhammer, 2014: A Study of Aerosol Impacts on Clouds and Precipitation Development in a Large Winter Cyclone. *J. Atmos. Sci.*, **71**, 3636-3658.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. 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.

Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.*, **132**, 519-542.

Tripoli, G. J., and W.R. Cotton, 1980: A Numerical Investigation of Several Factors Contributing to the Observed Variable Intensity of Deep Convection over South Florida. *J. Appl. Meteor.*, **19**, 1037-1063.

Twomey, S., 1974: Pollution and planetary albedo, *Atmos. Environ.*, **8**, 1251-1256.

Twomey, S., and P. Squires, 1959: The influence of cloud nucleus population on the microstructure and stability of convective clouds. *Tellus*, **11**, 408-411.

Verrelle, A., D. Ricard, and C. Lac, 2014: Sensitivity of high-resolution idealized simulations of thunderstorms to horizontal resolution and turbulence parametrization. *Q. J. R. Meteorol. Soc.*

Warner, J., 1969b: The microstructure of cumulus cloud. Part II: The effect on droplet size distribution of the cloud nucleus spectrum and updraft velocity. *J. Atmos. Sci.*, **26**, 1272-1282.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 036-h Explicit Convective Forecasts with the WRF-ARW Model. *Wea. Forecasting*, **23**, 407-437.

Weisman, M. L., W. C. Shamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527-548.

Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SALA novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Wea. Rev.*, **136**, 4470-4487.

WRF Users Page, 2014: WRF Model Version 3.2 : Known Problems and Fixes. [Available Online at <http://www2.mmm.ucar.edu/wrf/users/wrfv3.2/known-prob-3.2.html>.]

Wu, D., X. Dong, B. Xi, Z. Feng, A. Kennedy, G. Mullendore, M. Gilmore, and W.-K. Tao, 2013: Impacts of microphysical scheme on convective and stratiform characteristics in two high precipitation squall line events, *J. Geophys. Res.*, **118**, 11,119-11,135.