UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IDENTIFICATION AND VERIFICATION OF MESOSCALE CONVECTIVE SYSTEMS IN THE MODELS OF THE HIGH RESOLUTION ENSEMBLE FORECAST SYSTEM

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

JEFFREY MILNE
Norman, Oklahoma
2024

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/340628

IDENTIFICATION AND VERIFICATION OF MESOSCALE CONVECTIVE SYSTEMS IN THE MODELS OF THE HIGH RESOLUTION ENSEMBLE FORECAST SYSTEM

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Harold Brooks, Chair

Dr. Steven Cavallo

Dr. Stacey Hitchcock

Dr. Cameron Homeyer

Dr. Patrick Marsh

Dr. Renee McPherson

© Copyright by JEFFREY MILNE 2024
All Rights Reserved.

Acknowledgments

I would like to thank my advisor, Dr. Harold Brooks, and Dr. Patrick Marsh for their guidance and support. I would also like to thank Dr. Israel Jirak. I would like to thank Dr. David Jahn for providing the survey data used here. I would also like to thank Dr. Andrew Wade for sharing the code of the original observational MCS tracking algorithm. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce.

Table of Contents

Acknowledgments **iv**

List Of Tables **vii**

List Of Figures **viii**

Abstract **xii**

1 Introduction **1**
1.1 Historical Context of MCSs 1
1.2 MCS Impacts 6
1.3 MCS Forecasting 9
1.4 Research Questions 13

2 What do modeled storms look like? **16**
2.1 Introduction 16
2.2 Data & Methods 20
2.2.1 Data 20
2.2.2 Methods 21
2.3 Results 23
2.4 Conclusion 31

3 How can MCSs be identified in the models? **34**
3.1 Introduction 34
3.2 Data and Methods 38
3.2.1 Model Data 38
3.2.2 Expert Labeling 39
3.2.3 MCS identification and tracking algorithm 40
3.2.3.1 MCS identification 40
3.2.3.2 MCS tracking 42
3.2.4 Optimizing the identification algorithm 43
3.3 Results 48
3.3.1 Parameter optimization 48
3.3.2 Full year analysis 52
3.4 Conclusion 55

4 How good are the models at forecasting MCSs? 58
4.1 Introduction . . . 58
4.2 Data and Methods . . . 62
4.2.1 Model Data . . . 62
4.2.2 MCS identification and tracking algorithms . . . 62
4.2.2.1 MCS identification . . . 63
4.2.2.2 Observed MCS tracking . . . 63
4.2.2.3 Model MCS tracking . . . 63
4.2.3 Evaluation and Verification . . . 64
4.3 Results . . . 65
4.3.1 Time Series Comparison . . . 65
4.3.2 Verification . . . 67
4.3.3 Error evaluation . . . 71
4.3.4 Cases . . . 79
4.3.4.1 14 June 2022 . . . 79
4.3.4.2 21 August 2022 . . . 84
4.4 Discussion and Conclusion . . . 88

5 Summary and Conclusions 93

Bibliography 105

List Of Tables

1.1 Configuration of HREFv3 models used in this study 12

3.1 Summary of the number of responses for each case. 40
3.2 Summary of the default and optimized parameters for the modified Wade algorithm. 50

List Of Figures

1.1 Conceptual model of a trailing stratiform MCS as viewed from a vertical cross section perpendicular to the convective line. The orange shading is the front to year inflow, while the green shading is the rear inflow. The medium and dark gray shading is the radar reflectivity. Figure 17-12 from Houze (2018), which is a color version of Figure 1 from Houze et al. (1989). 4

1.2 Schematic showing how an updraft can be influenced by shear and a cold pool. (a) with neither wind shear nor a cold pool, the storm is vertical (b) with a cold pool but no shear, the structure is biased by the negative vorticity from the cold pool and the storm leans upshear (c) with shear but no cold pool, the positive vorticity from the shear causes the storm to lean (d) with both a cold pool and wind shear, the vorticities may balance and cause the storm to remain vertical. Figure 18 from Rotunno et al. (1988). 5

2.1 (a) A highly non-round ellipse with an orientation near 30° and an r' near 0.8, (b) a rounder ellipse with an orientation near 60° and an r' near 0.4, (c) the ellipses from (a) and (b) plotted on a polar plot. . . . 22

2.2 Density plot of object area against r', with a minimum area of 900 km^2 for the (a) HRW NSSL, (b) NAM CONUS Nest, (c) HRW FV3, (d) HRRR, and (e) HRW ARW models and (f) observations. Rounder objects have an r' closer 0, and less round objects have an r' closer to 1. The color scale uses a power law and is relative for each subfigure. The power law scale was chosen to emphasize differences at lower object counts. 24

2.3 As in Figure 2.2 for area and maximum reflectivity in the object with a linear color scale. 26

2.4 Aspect ratio and orientation plotted on polar axes, as described in Figure 2.1, subject to a 900 km^2 minimum area for the (a) HRW NSSL, (b) NAM CONUS Nest, (c) HRW FV3, (d) HRRR, and (e) HRW ARW models and (f) observations. The labeled red line in each plot represents the mean orientation angle for the given plot. Rounder objects are closer to the origin, and less round objects are closer to the edge. The color scale is relative for each subfigure. 27

2.5 As in Figure 2.4 with a minimum area of $4500km^2$. 28

2.6 Ellipses of various sizes, with aspect ratio and orientation for each determined by the mean aspect ratio and orientation for ellipses of those sizes for the (a) HRW NSSL, (b) NAM CONUS Nest, (c) HRW FV3, (d) HRRR, and (e) HRW ARW models and (f) observations. The locations of the ellipses are not representative and are meant to help provide a sense of scale. The number of objects at a given size is shown in the legend. 30
2.7 As in Figure 2.6, but for smaller ellipses. 30

3.1 The counts of expert-labeled storm mode for each model in the HREF. SC are supercells, QL are linear, DI are disorganized. 41
3.2 A schematic of the MCS tracking algorithm for time T. MCSs were identified at times $T-3$, $T-2$, $T-1$, and T. The intense reflectivity contour (black dashed line) is 50 dBZ. 44
3.3 Performance diagram showing how AUC-PR varies with several different forecasts. The x-axis is success ratio (SR), which is equivalent to 1 minus the false alarm ratio (FAR). They y-axis is probability of detection (POD). The black contours are lines of constant critical success index (CSI). 47
3.4 Performance diagrams showing the ability of the modified Wade algorithm to identify MCSs as labeled by the expert classification for each model. Each colored line is the performance of a single parameter set. The solid black line is the parameter set with the highest AUC-PR for each model and the dashed black line is the default parameters. 49
3.5 The number of modeled MCSs in 2022 for the (a) HRW NSSL, (b) HRRR, (c) HRW ARW, (d) NAM CONUS Nest, and (e) HRW FV3. The black X marks the location of the maximum count over land. The count at the maximum is denoted in the legend of each plot. 51
3.6 (A) Figure 2 from Wade et al. (2023) showing the average number of MCSs for 1996-2021. (B) Figure 1 from Haberlie and Ashley (2019) (a) shows the annual number of MCSs for 1996-2017, (b) and (c) show seasonal counts from May-August and September-April, respectively. . 53
3.7 The average area of modeled MCSs in 2022 for the (a) HRW NSSL, (b) HRRR, (c) HRW ARW, (d) NAM CONUS Nest, and (e) HRW FV3. . 53
3.8 KDE of modeled MCS centroids in 2022 for the (a) HRW NSSL, (b) HRRR, (c) HRW ARW, (d) NAM CONUS Nest, and (e) HRW FV3. Only centroids over land were considered for the analysis. 55

4.1 The total number of MCSs that occur during 2022 by time of day. . . . 66
4.2 Average number of MCSs per day, using a 31-day smoothing window. . 67

4.3 Performance diagram using area of overlap to count as hits. The y-axis is probability of detection and the x-axis is success ratio, which is equivalent to 1 minus the false alarm ratio. Note that the y-axis only goes from 0 to 0.4. The curved black lines are contours of critical success index (CSI). 68
4.4 Performance diagram using centroid distance. The threshold distance to count as a hit varied from 0 km to 1000 km. Results are plotted every 1km to produce a curve. The dots are plotted every 100 km. The hollow dots are associated with the dashed lines and the optimized parameters while the solid dots are associated with the solid lines and the default parameters. The stars represent the limit where every forecast object on a day with observed objects is counted as a hit (in other words, if the distance threshold went towards infinity). Again note that the y-axis only goes from 0 to 0.4. 69
4.5 As in Figure 4.4 but for minimum edge distance between MCSs. 71
4.6 The storm-relative error vector from the centroid of each modeled MCS to the nearest observed MCS centroid using the default MCS identification parameters. The radius is in kilometers. The modeled storm direction is along the positive x-axis. The hollow square markers indicate when the modeled MCS overlapped with the observed MCS, and the filled dots indicate when there was no overlap. The models are (a) the HRW NSSL, (b) the HRW ARW, (c) the HRW FV3, (d) the NAM CONUS Nest, and (e) the HRRR. The numbers in the legend indicate the number of forecast MCSs in each category. 72
4.7 As in Figure 4.6 but for the optimized MCS identification parameters. . 73
4.8 The difference in storm direction between the modeled MCS and the nearest observed MCS. The mean angle difference is plotted. The radius is the distance between the modeled MCS centroid and the nearest observed MCS centroid. The modeled storm direction is along the positive x-axis. The hollow square markers indicate when the modeled MCS overlapped with the observed MCS, and the filled dots indicate when there was no overlap. The models are (a) the HRW NSSL, (b) the HRW ARW, (c) the HRW FV3, (d) the NAM CONUS Nest, and (e) the HRRR. The numbers in the legend indicate the number of forecast MCSs in each category. 77
4.9 As in Figure 4.8 but for the optimized MCS identification parameters. . 78
4.10 SPC Day 1 outlook and storm reports for 14 June 2022. 80
4.11 Modeled and observed MCSs for 14 June 2022. The solid colored lines are the individual MCSs and the black dashed lines are the convex hulls of each full MCS. The modeled MCSs were identified using the default parameters. (a) HRW NSSL, (b) HRW ARW, (c) HRW FV3, (d) NAM CONUS Nest, (e) HRRR, (f) Observations. 81

4.12 As in Figure 4.11, but modeled MCSs were identified using the MCS identification parameters optimized to match expert labeling. 82
4.13 SPC Day 1 outlook and storm reports for 21 August 2022. 85
4.14 As in Figure 4.11 for 21 August 2022. 86
4.15 As in Figure 4.12 for 21 August 2022. 87

Abstract

To better understand the characteristics of storms produced by forecast models and those seen in observations, contiguous and nearly contiguous radar objects with reflectivity above a convective threshold were identified and their shapes analyzed. The area, aspect ratio, orientation angle and maximum reflectivity within the object were calculated. For all members of the High Resolution Ensemble Forecast system, the modeled and observed storm objects had similar distributions in area-aspect ratio space. Modeled storm objects had a preferred band of maximum reflectivity that was not seen in the observed storm objects. The modeled storm objects also had a more north-south orientation than observed. Despite these differences, the shape characteristics of the modeled storm objects were still close enough to modeled storm objects to proceed with the creation of an MCS identification and tracking algorithm. The identification algorithm was based on one developed for MCS identification and tracking of observed storms at the NOAA Storm Prediction Center. A novel tracking algorithm was developed based on continuous swaths of maximum reflectivity and tuned to match expert labeling. Together, the identification and tracking algorithm produced a one year spatial distribution of MCSs that was somewhat consistent with an observed climatology. Additional verification of the identification and tracking algorithm revealed an underforecast from all of the models. At both the diurnal and annual scale, the models generally did not match the shape of the observed distribution. Matching the forecast MCSs with observed MCSs further showed low bias, though the forecasts did have very low FAR. The tuned parameters did not necessarily improve the forecast compared to the default parameters. The positional error between the modeled and observed MCSs showed that when the model produced an MCS, it was usually close to an observed MCS and oriented the same direction as the observed MCS.

Chapter 1

Introduction

Mesoscale convective systems (MCSs, Maddox, 1980; Houze, 2004) are large organized storms that are often associated with weather impacts such as intense rainfall (Da Silva and Haerter, 2023) or severe wind, hail, and tornadoes (Smith et al., 2012). MCSs are also projected to increase in frequency in a warming climate (Haberlie et al., 2023). While numerical forecast models can forecast for MCSs (Done et al., 2004), this research evaluates the ability of the models of the High Resolution Ensemble Forecast systems (Roberts et al., 2019) to forecast for MCSs using an automated detection algorithm based a method used operationally to identify observed MCSs.

1.1 Historical Context of MCSs

MCSs were first described by Maddox (1980). MCCs were defined based on the size, shape, and duration of their infrared satellite signature: they were required to have length scales of at least 125 km and eccentricities greater than 0.7 (that is, the minor axis length must be 0.7 times the major axis length or longer) and last for at least 6 hours. Maddox (1980) classifies MCCs as a circular type of mesoscale convective system (MCS). Linear type MCSs are classified by Maddox (1980) as squall lines. Maddox (1983) then explored the environments associated with MCCs and showed that MCCs were associated with the progression of shortwave troughs.

Bluestein and Jain (1985) categorize linear MCSs into four types based on the structure of the convective line: broken line, back building, broken areal, and embedded areal. They considered storms that were at least 50 km long, less than 50 km wide, with a length-to-width ratio of at least five-to-one. Broken line and back building storms tended to occur earlier in the convective season, while broken areal and embedded areal were more frequent later in the season. These storms tended to form in conditionally unstable environments with strong low level shear and weak shear aloft. The storms are oriented along the low level shear, and the storm motion's steering level is between 6-7 km.

Parker and Johnson (2000) investigated linear MCSs and classified them based on the location of the stratiform precipitation: leading, trailing, or parallel stratiform. Trailing stratiform, where the stratiform region of an MCS trails the convective line, accounted for 60% of the cases studied by Parker and Johnson (2000). Leading stratiform, where the stratiform region is ahead of the convective line, and parallel stratiform, where the stratiform region is parallel to (that is, at one end or the other) of the convective line, each accounted for 20% of the cases. Regardless of the type, a majority of MCSs evolved towards a trailing stratiform structure.

Some linear MCSs are also known as quasi-linear convective systems (QLCSs, Trapp et al., 2005). QLCSs are linearly oriented and may contain bow echoes or cause derechos (Johns and Hirt, 1987). The impacts of MCSs will be discussed further in the next section, but Trapp et al. (2005) found that in Indiana, half of all tornado days were associated with QLCSs.

A conceptual model of the structure of a trailing stratiform MCS was developed by Houze et al. (1989, Figure 1.1). Low level inflow from ahead of the convective line is ingested into the storm and ascends into the updraft associated with the strongest radar reflectivity. New convective cells form as this inflow is lifted. The air then continues

to lift through the stratiform region, though not as quickly as in the convective region. Beneath the ascending front-to-rear inflow is the descending rear inflow. This air is drawn from the mid levels behind the convective line. It is driven downward by the melting, sublimation, and evaporation of precipitation in the stratiform region. If this inflow merges with the convective downdrafts, it is known as the rear inflow jet (Smull and Houze, 1987), and the rear inflow jet can lead to the formation of a bow echo and severe winds at the surface (Squitieri et al., 2023).

A conceptual theory of MCS maintenance is described by Rotunno et al. (1988, RKW theory). According to RKW theory, one way for an MCS to be maintained at the leading edge of the cold pool is for the horizontal vorticity generated by the gust front and cold pool to be balanced by the ambient line-normal wind shear (Figure 1.2). When the cold pool and wind shear are balanced (Figure 1.2(d)), the storm can be maintained. When the wind shear or cold pool generate stronger vorticity than the other (Figures 1.2(b,c)), the updraft will be tilted either upshear or over the cold pool. RKW theory assumes that vertical updrafts last longer than tilted updrafts since vertical updrafts are usually stronger than tilted updrafts.

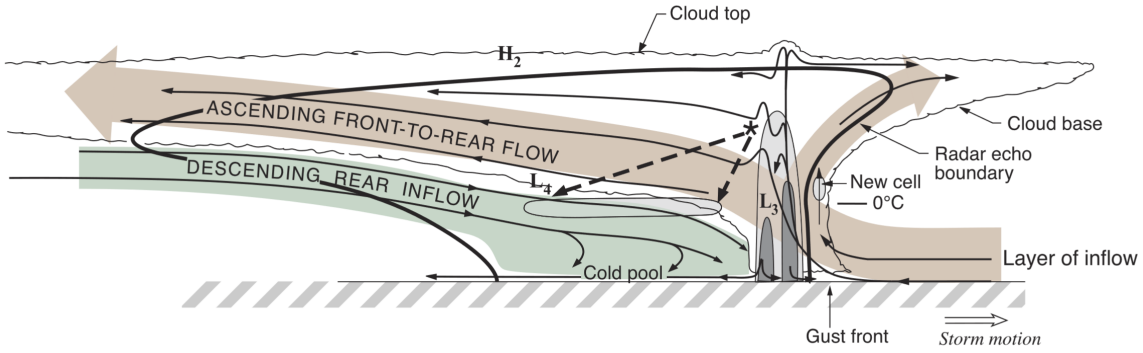


Figure 1.1: Conceptual model of a trailing stratiform MCS as viewed from a vertical cross section perpendicular to the convective line. The orange shading is the front to year inflow, while the green shading is the rear inflow. The medium and dark gray shading is the radar reflectivity. Figure 17-12 from Houze (2018), which is a color version of Figure 1 from Houze et al. (1989).

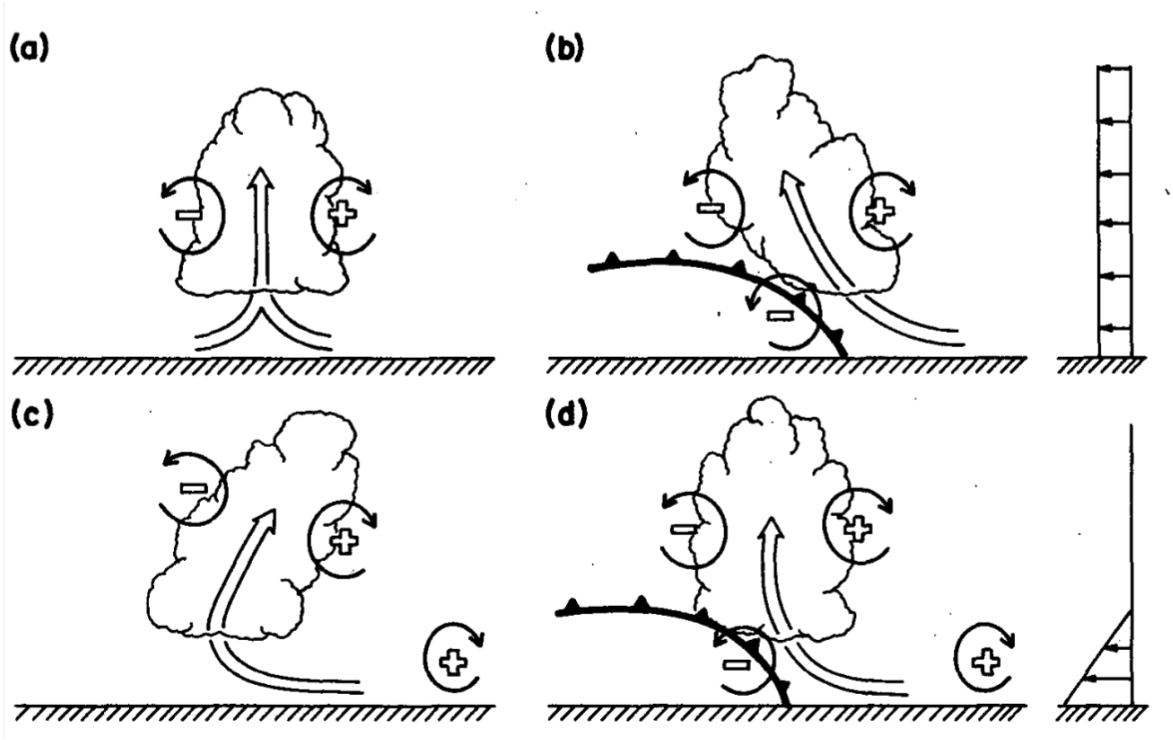


Figure 1.2: Schematic showing how an updraft can be influenced by shear and a cold pool. (a) with neither wind shear nor a cold pool, the storm is vertical (b) with a cold pool but no shear, the structure is biased by the negative vorticity from the cold pool and the storm leans upshear (c) with shear but no cold pool, the positive vorticity from the shear causes the storm to lean (d) with both a cold pool and wind shear, the vorticities may balance and cause the storm to remain vertical. Figure 18 from Rotunno et al. (1988).

1.2 MCS Impacts

MCSs are often associated with severe weather (as defined by the NWS: tornadoes, wind gusts greater than 50 kt, hail greater than 1"). Smith et al. (2012) found that 25.3% of tornadoes, 6.4% of significant hail reports (diameter $\geq$ 2"), and 44.3% of significant wind reports (gusts $\geq 65kt$) were associated with linear MCSs. Tornadoes associated with linear systems tend to occur further east and later in the year than those associated with right moving supercells (Smith et al., 2012). Significant wind events associated with linear systems occur more often in the northern plains and upper Midwest during the summer (Smith et al., 2012).

The tornado threat is also different in MCSs than in supercells given an otherwise similar environment. Jahn et al. (2022) found that tornadoes were less likely in MCSs and used that knowledge to improve model-based calibrated tornado guidance for each grid point within the model domain. Supercells and MCSs were distinguished by the distribution of updraft helicity (UH) in the vicinity of each point. The distribution of UH in modeled MCSs was found to have a lower skewness than the distribution within modeled supercells. The method used by Jahn et al. (2022) classifies each gridpoint within a storm as being part of an MCS or a supercell, so it does not necessarily provide large scale storm mode classification. In other words, the Jahn et al. (2022) method does not necessarily guarantee that the storm classification will be spatially consistent. A single grid square may be identified as part of a supercell if the UH distribution matches that of a supercell even if that grid square is surrounded by other grid squares identified as being part of an MCS.

While there is much focus on MCSs impact on severe weather (as defined by the NWS), MCSs also produce intense rainfall. Da Silva and Haerter (2023) found that MCSs accounted for up to 70% of extreme rainfall events in parts of Europe. As the

Earth's climate continues to warm, Haberlie et al. (2023) found that the fraction of precipitation coming from MCSs are projected to increase across the US. Schumacher and Rasmussen (2020) found that MCSs contribute up to half of the annual rainfall in some parts of the US. Despite their importance, Schumacher and Rasmussen (2020) note that there are difficulties in forecasting for MCSs because many processes that drive MCSs occur at scales that are unresolved in numerical models. Fairman et al. (2016) explored "banded" precipitation features (length greater than 100 km and a major to minor axis ratio of at least three to one) and showed that precipitation features are banded up to 30% of the time east of the Rockies. For the entire US, 15% of precipitation features are banded, but these banded features account for 22% of the annual precipitation. Though Fairman et al. (2016) does not explicitly call the "banded" features MCSs, the definition used is consistent with the definition of MCSs from other studies such as Bluestein and Jain (1985).

In addition to the immediate weather threats associated with MCSs, the existence and placement of MCSs can have impacts on future weather. A cold pool left behind by an MCS can stabilize the atmosphere (Houze, 2004), and if a model does not accurately represent the MCS or its cold pool, forecasts after the MCS will be degraded. Similarly, outflow boundaries associated with MCSs can then provide foci for initiation of later convection (Parker and Johnson, 2000). Thus a model failing to produce an MCS early in the model run can result in the model failing to produce an MCS later, especially if subsequent MCSs were initiated by the outflow of the first MCS.

At a larger scale, Parsons et al. (2019) and Rodwell et al. (2013) show a link between poorly modeled MCSs in the central US and inaccurate model forecasts over Europe. Parsons et al. (2019) attribute the connection to downstream propagation of Rossby wave packets that result in errors in the 500 hPa field over Europe. If a model inaccurately represents the impact of the MCS on the upper level flow, that

error propagates downstream. Rodwell et al. (2013) attribute the error in the 500 hPa field to MCSs over the US causing Rossby waves to slow down. Thus if a model misses an MCS, there will be downstream timing errors of the large-scale flow.

The presence or absence of an MCS in a numerical model can also influence forecast evolution and ensemble forecast spread. Clarke et al. (2019a) show that MCSs induce potential vorticity anomalies that can be captured with a regional convection allowing model, but the anomalies were weaker in a global convection-parameterizing model. When these anomalies were artificially introduced to a convection-parameterizing model, they were found to influence the strength of an upper level trough within the first 24 hours after the MCS. The potential vorticity anomalies were still impacting geopotential height fields five days after the MCS. The perturbations induced by the MCS also impacted the position of a surface cyclone. Clarke et al. (2019b) did a similar analysis with ensembles and found that the potential vorticity anomalies induced by an MCS resulted in a deepening of the developing cyclone, but the perturbations were damped. The perturbations induced by the MCS do persist through five days, but there was no discernible change in skill between an ensemble with and without MCS-induced potential vorticity perturbations. Clarke et al. (2019b) found that the perturbations from the MCS were less impactful than the initial perturbation conditions used to initialize the ensemble.

For example, a poor MCS forecast over Iowa would have many impacts, both immediately and downstream. For a missed event (that is, an MCS occurs but was not forecast by the forecast model), rain will occur that was not forecast, and there may be severe weather (as defined by the NWS) that was not forecast. The rainfall may be enough to induce flooding and severe weather may cause damage to property or injuries or even deaths. Beyond the immediate impacts of a missed event, there are also downstream impacts. The outflow of the MCS could result in the creation of an outflow

boundary, which may serve as a focus for subsequent initiation of thunderstorms later. If a forecast model missed an MCS, it would not have the outflow boundary and so may not initiate convection later in the forecast. Even further downstream, a missed forecast of an MCS in the US could cause degraded 500 hPa forecasts over Europe days later. For a false alarm (that is, an MCS is forecast but does not occur), the opposite impacts might happen: a lack of rainfall or severe weather, no outflow boundary for later initiation, and degraded 500 hPa forecast over Europe.

1.3 MCS Forecasting

Forecasting MCSs is a challenge (Corfidi, 2003; Coniglio et al., 2007). Corfidi (2003) and Coniglio et al. (2007) both show the challenges associated with forecasting MCS motion and developement. Herman et al. (2018) found that the Storm Prediction Center's forecast skill had a relative minimum in the warm season, which coincides with an observed maximum in MCSs seen in Schumacher and Rasmussen (2020).

Part of the challenge in MCS forecasting is that the conditions conducive to convective initiation are not necessarily those conducive to MCS growth and development (Jirak and Cotton, 2007). Jirak and Cotton (2007) also find that MCSs initiate mostly along frontal boundaries rather than along the dryline.

Once an MCS has formed, MCS motion is related to the cold pool-relative flow and the distribution of instability along the gust front (Corfidi, 2003). Corfidi (2003) develops a forecasting technique to forecast MCS storm motion in the 3-6 hour time frame based on the cold pool and instability.

Corfidi et al. (1996) explored the motion of MCCs and showed that their motion was driven by mean flow in the cloud layer and the low level jet. Together, these components represent advection and propagation, respectively. Corfidi et al. (1996)

was able to accurately predict the motion of MCCs using the mean flow in the cloud layer and the low level jet.

Forecasting for MCS maintenance is also challenging. Coniglio et al. (2007) explore differences between mature and dissipating quasi-linear convective systems (QLCSs) and finds that higher wind shear is associated with mature QLCSs and is the best discriminator between mature and dissipating QLCSs. High lapse rates, high instability, and high mean wind associated with a QLCS indicates that the QLCS is mature rather than dissipating.

Letkewicz and Parker (2010) look at the differences between MCSs that cross the Appalachian Mountains and those that do not. MCSs that crossed the mountain range tended to have more unstable environments, weaker wind shear, and a slower mean wind in the lee of the mountain range. Some of these findings, especially weaker winds in the lee of the mountains, seem to run counter to the findings of Coniglio et al. (2007). Letkewicz and Parker (2010) attribute the difference to several potential reasons relating to the orographic downslope flow or improved cold pool dynamics after storms initially weaken in the mountains.

Though convection-allowing models (CAMs) are commonplace now, convection-parameterizing models also forecast MCSs. Koch (1985) found that their model produced fields that were useful to diagnose mesoscale convection. Their model produced too little convection more frequently than it produced too much convection, and errors were largely driven by synoptic-scale errors. There were also forecast errors related to the physics of deep convection. More recent studies use CAMs, but Koch (1985)was the first to evaluate model performance in forecasting for MCSs. Given the importance of MCSs to severe weather and rainfall and their impact on subsequent weather, it is important to understand and quantify a model's ability to forecast MCSs. CAMs and CAM ensembles can forecast MCSs (Done et al., 2004; Bauer et al., 2015; Pinto et al.,

2015), but further study is needed to assess how accurately models forecast them. This information will allow forecasters to know the model biases and model developers to know areas where models can improve.

Some CAMs have been shown to be able to forecast MCSs well. Pinto et al. (2015) identifies MCSs in both the High Resolution Rapid Refresh model (HRRR) and observations by using vertically integrated liquid (VIL). Objects are identified where VIL exceeds a threshold. From those objects, MCSs are identified where objects have a largest dimension longer than 100km, allowing for gaps below the VIL threshold within an object of up to 20 km. The VIL threshold used to identify modeled MCSs was optimized such that the distribution of modeled MCS areas matched the distribution of observed MCS areas, and the final threshold was 50% less than that used to identify observed MCSs. There was a high bias in the number of MCSs produced early in the forecast period, but a low bias by the end of the period. Spatially, the HRRR had a high bias over the Great Plains, but there was a low bias in the southeast and Mississippi River valley.

CAM ensembles can also forecast MCSs well, though the structure of modeled MCSs did not always match observed MCSs. Britt et al. (2024) found that their ensemble Warn On Forecast System (WoFS) forecast QLCSs skillfully. Modeled QLCSs were less straight, oriented more counterclockwise, larger, more intense, and more round than observed storms. When comparing individual forecasts, Britt et al. (2024) also found that forests of non-QLCSs had higher false alarm ratio (FAR) and lower probability of detection (POD) and critical success index (CSI) than forecasts of QLCSs, which suggests that while the model is overforecasting other modes, it does better at predicting QLCSs.

Member	ICs	LBCs	Microphysics	PBL	dx (km)	Vertical levels
HRRR	RAP - 1h	RAP - 1h	Thompson	MYNN	3.0	50
HRW ARW	RAP	GFS - 6h	WSM6	YSU	3.2	50
HRW FV3	GFS - 6h	GFS - 6h	GFDL	GFS EDMF	3.0	60
HRW NSSL	NAM	NAM - 6h	WSM6	MYJ	3.2	40
NAM CONUS Nest	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60

Table 1.1: Configuration of HREFv3 models used in this study

There are many CAMs routinely available to operational forecasters. In this study, the non-time lagged members of the High Resolution Ensemble Forecasting System version 3 (HREF) will be used (summarized in Table 1.1). The HREF began as the Storm Scale Ensemble of Opportunity (Jirak et al., 2012), and it has since become operational as the HREF (Roberts et al., 2019). The HREF is a multi-model, multi-physics ensemble that has been shown to have more spread than other single-physics ensembles, which allows the HREF to better capture forecast uncertainty (Roberts et al., 2020). The HREF is used operationally by forecasters, and the output from the HREF is used to generate calibrated guidance (Jahn et al., 2022; Harrison et al., 2022).

The individual members of the HREF have been evaluated in various studies. The HRRR has been found to roughly match observed MCS area while underforecasting the number of storms, particularly early in the forecast period (Grim et al., 2024). Though it was not identical to the HRW FV3, the Rapid Refresh Forecast System (RRFS) evaluated by Grim et al. (2024) was similar to the HRW FV3. Grim et al. (2024) found that the RRFS overforecasts the area of the MCSs while also underforecasting

the number of them. Gallo et al. (2018) found that the NSSL WRF could be used to generate probabilities of tornadoes. The NAM CONUS Nest has been found to struggle with convection initiation due to inaccurate representation of the vertical structure of the atmosphere (Colbert et al., 2019).

When examining the HREF as an ensemble, it has been shown to forecast convection well. Roberts et al. (2020) found that the HREF outperformed other CAM ensembles when forecasting for 40 dBZ simulated reflectivity. The other ensembles had less spread and thus produced higher probability of occurrence for events, even when the models were incorrect. The HREF had more spread and was thus able to better sample the range of possible forecasts. Clark et al. (2023) also found that the HREF had higher subjective scores of forecasts of convective fields than an RRFS-based ensemble.

1.4 Research Questions

While previous research has shown that CAMs can forecast for MCSs and that the HREF outperforms other CAM ensembles, there has been little research into the ability of the models of the HREF to forecast for MCSs. In particular, the HREF has been shown to outperform other ensembles in forecasting convection in general, but the ability of the HREF members to forecast for organized convection has not been thoroughly explored. To address these research gaps, this research asks three questions: What do modeled storms look like? How can MCSs be identified in the models? and How good are the models at forecasting MCSs?

To determine what modeled storms look like, the shape characteristics of modeled and observed storms are explored to determine if the shapes of storms produced by the models in the HREF are similar, in bulk, to observed storms. The HREF has been shown to perform well compared to other CAM ensembles, especially with regard to

simulated reflectivity, so it was expected that each individual member will also produce storms that have shapes similar to observations. This is done to determine how similar the radar objects produced by the models were to observations. The similarities and differences are analyzed to assess whether, in bulk, each model produces storms that look like observed storms. This analysis informs the development of an automated MCS identification method.

Once determined that modeled and observed radar object shapes are largely similar to each other, an automated method of identifying MCSs is developed. This method is based on a method for automated detection of observed MCSs using radar reflectivity (Wade et al., 2023). The method is modified to work on modeled radar reflectivity and parameters are tuned to match expert forecaster labeling. The areas and distribution of MCSs identified using the tuned parameters are also explored in an attempt to determine how the tuning impacted MCS identification. It is hypothesized that the automated detection algorithm can be modified and tuned to identify MCSs in a way that matches expert labels and that the tuned algorithm, when applied for a full year of data, identifies MCSs in a way that is consistent with multi-year climatologies derived from the detection algorithm applied to observations.

Finally, to evaluate the models' ability forecast MCSs, MCSs are identified over a full year using the automated detection algorithm. These model MCSs are verified directly against MCSs identified by the MCS identification algorithm. The hypothesis is that the tuning done to identify MCSs in a way that matches expert labeling will also improve the models' ability to identify MCSs. The verification is done on a daily basis over an entire year. In addition to matching forecast and observed MCSs to determine hits, misses, and false alarms, the displacement and orientation errors are also evaluated. The time series of the various models are also explored and compared to the observed time series, both annually and diurnally. Evaluating both the time

series and day-by-day performance of each model give an improved picture of each model's ability to forecast MCSs.

Chapter 2

What do modeled storms look like?

2.1 Introduction

CAMs and CAM ensembles can give forecasters insight into the character of storms expected during the forecast period (Jirak et al., 2012). The Storm Scale Ensemble of Opportunity described in Jirak et al. (2012) has since operationalized as the High Resolution Ensemble Forecast System (HREF, Roberts et al., 2019). The HREF is a multi-model, multi-physics ensemble and has been shown to have more spread than other single-physics ensembles, which allows the HREF to better capture model uncertainty (Roberts et al., 2020). The HREF is used by the Storm Prediction Center as the basis for short-term calibrated probabilistic thunderstorm guidance (Harrison et al., 2022).

Mesoscale convective systems (MCSs) can cause impactful weather (e.g., severe weather and heavy rainfall; Smith et al., 2012; Haberlie et al., 2023) but CAMs can be useful in forecasting them (Done et al., 2004). Done et al. (2004) found that their model under-forecast the number of MCSs, though the model did have some correlation with the number of observed MCSs. The model used in Done et al. (2004) also had skill in forecasting the dominant convective mode of the day. There was a significant northward bias in the initiation of MCSs, but because of a rightward bias in propagation speed and a long bias in MCS duration, modeled MCSs dissipated southward and eastward of observed MCSs (Done et al., 2004).

Previous studies have used object-based verification to verify CAM forecasts. Duda and Turner (2021) found that the HRRR was unable to reproduce the spatial radar reflectivity structure of the observed storms. This was more pronounced with smaller objects because the model was unable to resolve finer-scale features of the environment. Duda and Turner (2021) also found that the HRRR generally over-predicted the number of storms, especially for smaller objects. Forecasts early in the forecast period (from hours 1 - 3) had the greatest sensitivities to the impacts of data assimilation (Duda and Turner, 2021).

Spatial forecast verification, especially for high resolution models, is a challenge (Gilleland et al., 2009). One issue with spatial verification techniques is the potential for a double penalty (Ahijevych et al., 2009). A slight displacement error in a forecast can lead to a single incorrect forecast being classified as both a missed event and a false alarm (as demonstrated by Figure 1 in Ahijevych et al., 2009). Gilleland et al. (2009) divides spatial verification techniques into four broad categories: neighborhood, scale separation, object-based, and field deformation. Neighborhood and scale separation methods can be even more broadly categorized as filtering methods, while object-based and field deformation methods can be categorized as displacement methods (Gilleland et al., 2009). The two filtering methods attempt to find the scale at which the forecast is useful by applying a spatial filter (for neighborhood methods) or a spectral filter (for scale separation methods). The two displacement methods attempt to quantify how much a forecast field would need to change to match an observation field (Gilleland et al., 2009).

The Method for Object-Based Detection and Evaluation (MODE, Davis et al., 2006) is an object-based verification method that attempts to solve the double penalty by identifying objects in both the observation and forecast fields, then using a fuzzy logic algorithm to determine whether the forecast and observed objects match. This

method gives forecasts credit for being "close enough" to the observations without unduly penalizing for small displacement errors.

Another object-based method is the Contiguous Rainfall Area (CRA, Ebert and McBride, 2000). As the name suggests, the CRA identifies areas of contiguous rainfall in a field that is the merge of the forecast and observations. The forecast is then translated over the observations until the "best" forecast is found ("best" can be determined from any number of measures) and the displacement vector is recorded. The shifted fields can then be directly compared without having to consider any potential double penalty.

Gallus (2010) explored the differences between MODE and CRA and found that both methods were useful. Using both methods, Gallus (2010) showed that a CAM ensemble with perturbed initial and lateral boundary conditions had more spread than an ensemble with mixed physics.

Another technique for spatial verification is the Fractions Skill Score (FSS, Roberts, 2008; Roberts and Lean, 2008). FSS is a neighborhood method that compares areal coverage of where both forecast and observation fields exceed a given threshold across a range of spatial scales. Increasing the spatial scale gives credit to "close enough" forecasts while still keeping track of displacement errors. Comparing the FSS at various spatial scales for a single forecast can provide information about the forecast's skill. The threshold used to identify areas can also be varied to further evaluate a forecast's performance.

Lack et al. (2010) describe the Procrustes verification scheme, which is an object-based scheme that identifies objects based on reflectivity, but only stores certain object characteristics to speed up the verification process so it can be done in near-real time. Forecast and observed objects are matched using a penalty function based on shape error (size, rotation, and displacement) and intensity error (Lack et al., 2010). The

penalty functions for each object on a given day can be summed to yield a total penalty function, with better forecasts having lower penalty functions (Lack et al., 2010). Lack et al. (2010) found that the Procrustes scheme was useful for comparing forecasts, especially members of an ensemble, but the single number from the penalty function was unable to characterize the forecast as good or bad.

Pinto et al. (2015) used object-based verification to verify HRRR forecasts of MCSs using vertically integrated liquid to identify MCSs. The HRRR produced too many MCSs early in the forecast period and too few late in the period, though there was spatial variability in the frequency errors (Pinto et al., 2015). The HRRR predicted too many daytime MCSs over the Great Plains, but too few over southeastern US (Pinto et al., 2015). There was also a low bias in the number of MCSs predicted overnight in the Mississippi River Valley, which was attributed to the model dissipating MCSs too quickly (Pinto et al., 2015).

Johnson and Wang (2012) used MODE to identify objects from CAM ensembles. Johnson and Wang (2012) found that, for lead times longer than one hour, the ensemble overforecast the number and aspect ratio and underforecast the size of one-hour accumulated precipitation objects. This was explained by the model producing numerous, spurious, small small circular objects. Johnson and Wang (2012) also found that the ensemble's accuracy had a diurnal cycle, with a maximum in forecast accuracy at either 12-hour or 24-hour lead time, depending on if the accuracy measure favored small objects or weighted small and large objects the same, respectively. ARW-based members had less round objects than NMM members (Johnson and Wang, 2012). The orientation angle of storms in the model were found to subjectively match those seen in observations (Johnson and Wang, 2012). The objects were identified in Johnson and Wang (2012) using the 1 hour accumulated rainfall field, which means the orientation angle is highly dependent on the storm motion, especially for small objects.

Many verification methods have an underlying assumption that modeled storms "look like" observed storms. That is, they assume that, while the model may not produce storms in the same places or times as observations, the storms produced by the model are realistic looknig. It is hypothesized that all members of the HREF will produce storm objects with similar shape characteristics to observations. This chapter explores the differences in shape between modeled and observed radar objects to determine how similar modeled storms are to those that are observed.

2.2 Data & Methods

2.2.1 Data

To identify and characterize observed storms using observed radar reflectivity, the Iowa Environmental Mesonet radar archive was used (Iowa State University, 2001). The Iowa Environmental Mesonet archive was also used to develop the algorithm in Wade et al. (2023) which will be used as the basis for an MCS identification algorithm in Chapter 3. The data from the Iowa Environmental Mesonet archive has a horizontal resolution of 500 m and a data resolution of 0.5 dbZ. Radar mosaics from the top of every hour from 12 UTC on 1 January 2022 through 11 UTC on 1 January 2023 were used. Modeled storms were identified from the non-time lagged members of the High Resolution Ensemble Forecast system version 3 (summarized in Table 1.1). Hourly reflectivity for forecast hours 12-35 from the 00 UTC run were used for the entirety of 2022. Forecast hours 12-35 were used to avoid any potential impacts of model spin-up (Schwartz et al., 2021; Short and Petch, 2022; Duda and Turner, 2021) and to capture a full day without counting any hour twice. In other words, forecasts valid from 12 UTC on 1 January 2022 through 11UTC on 1 January 2023 will be used.

2.2.2 Methods

Observed storms were identified based on contiguous areas of reflectivity above 35 dBZ. A threshold of 35 dBZ is often used to identify convective precipitation (Henderson et al., 2021; Cecchini et al., 2020; May and Ballinger, 2007). All identified storm objects were considered. Modeled storms were identified based on contiguous areas of reflectivity above the 99.662th percentile of reflectivity for the whole year. The same percentile was used for each model because it is equivalent to the percentile of 35 dBZ in observations. Nearby objects were merged with each other if there was 3 km or less separation between objects.

Once objects were identified and merged, the area, maximum reflectivity within the object, aspect ratio and orientation angle were calculated. Area was calculated by counting the number of grid squares within the object and multiplying by the area of each grid square. Maximum reflectivity is simply the maximum value of reflectivity within the object. The aspect ratio, r', is defined for this study as one minus the ratio of the minor axis to the major axis. Aspect ratio is thus bounded by $0 \leq r' < 1$, with r' near one representing a highly elliptical object and r' of zero representing a perfect circle. Aspect ratio was chosen and defined this way to be bounded and to provide a metric that better matches the visual differentiation between round and highly elliptical objects. Orientation angle, θ, is the direction in which the major axis points and is bounded by $-180° < \theta \leq 180°$. An orientation of $0°$ corresponds to a major axis pointing directly east, with orientation increasing counterclockwise. Since the objects are not perfectly elliptical, the major and minor axis are determined based on the ellipse that has the same normalized second central moments of the object.

Once the objects and their characteristics have been identified and their characteristics recorded, area was plotted against maximum reflectivity and aspect ratio for each

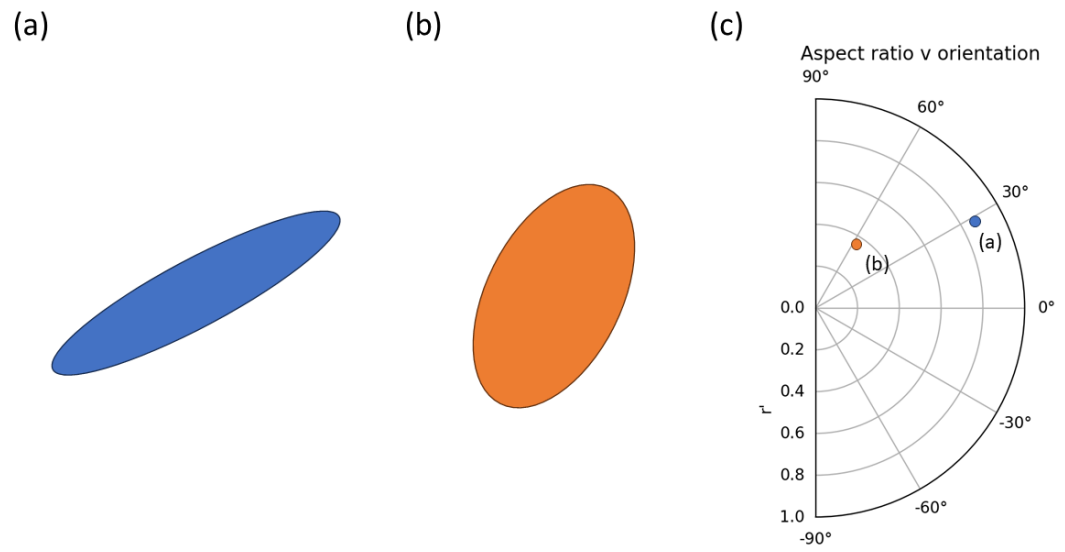


Figure 2.1: (a) A highly non-round ellipse with an orientation near 30° and an r' near 0.8, (b) a rounder ellipse with an orientation near 60° and an r' near 0.4, (c) the ellipses from (a) and (b) plotted on a polar plot.

model and the observations. Aspect ratio and orientation were also plotted against each other to assess any relationship between the object's shape and orientation. This plot was made in a polar coordinate system, with the orientation angle determining theta and the aspect ratio determining the radius. The aspect ratio was defined so that rounder objects would be closer to the origin and highly elliptical objects would be further away. This highlights elliptical objects, since for rounder objects, orientation is less important. For example, the orientation of a near-perfect circle could vary greatly without greatly changing the appearance of the object. On the other hand, for a highly elliptical object, varying the orientation angle would cause a larger change in the appearance of the object. Figure 2.1 demonstrates how objects with different shapes are plotted. Highly elliptical objects are plotted at a longer radius and rounder objects are plotted at a shorter radius. The orientation of the ellipse determines the

angle at which the object is plotted on the polar plot. This will show the differences in the shape and orientation of objects produced by the models and the observed objects.

To more fully assess and visualize the shape characteristics of modeled and observed storm objects, ellipses of various sizes are drawn on a map using the aspect ratios and orientations associated with objects of that size. For an object of a given size, the mean aspect ratio and orientation of all objects with areas within $450km^2$ of the given size were calculated. An ellipse was then drawn with the given area and mean aspect ratio and orientation. Since orientation is bounded by (-180°, 180°], the mean orientation was calculated by doubling the value, computing the mean using sines and cosines, and halving the result. For example, to calculate an average ellipse with an area of $4500km^2$, all objects with areas between $4050km^2$ and $4950km^2$ would be used. The mean aspect ratio and orientation angle would be calculated for all objects within that area limit, and a single ellipse would be drawn with an area of $4500km^2$ and the calculated average aspect ratio and orientation angle for objects with areas between $4050km^2$ and $4950km^2$. The ellipses are drawn on a map, but the location is not representative and is only to provide a reference for object size.

2.3 Results

Plotting object area against aspect ratio (Figure 2.2) shows that, in general, larger objects are less round. The models also fail to produce large enough object. This pattern holds for all five models and for observations. This suggests that the largest reflectivity objects tend to be more linear in shape. Such large, linear objects could be due to quasi-linear convective systems (QLCSs). For the largest objects, however, the models seem to favor less round objects compared to the observations. This indicates that the models produce more very large, very linear reflectivity objects, while the

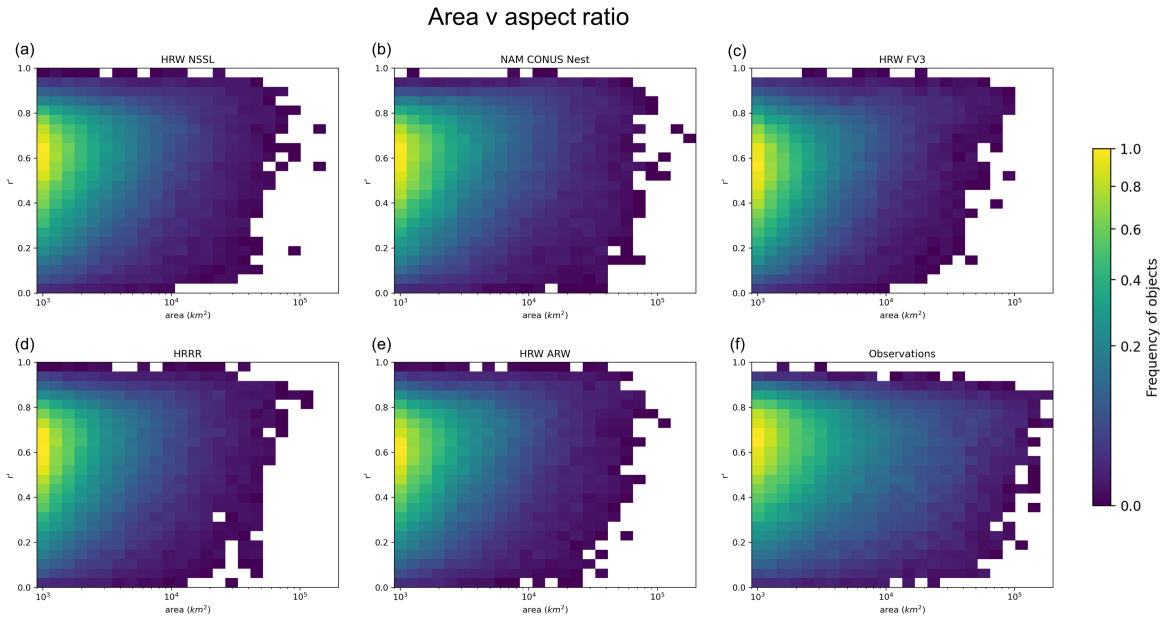


Figure 2.2: Density plot of object area against r', with a minimum area of 900 km^2 for the (a) HRW NSSL, (b) NAM CONUS Nest, (c) HRW FV3, (d) HRRR, and (e) HRW ARW models and (f) observations. Rounder objects have an r' closer 0, and less round objects have an r' closer to 1. The color scale uses a power law and is relative for each subfigure. The power law scale was chosen to emphasize differences at lower object counts.

observed reflectivity objects of a similar size are not as linear, though the observed objects do get more linear with increasing size.

Figure 2.3 shows the relationship between the area of reflectivity objects and the maximum reflectivity within those objects. The HRW NSSL, NAM CONUS Nest, HRW FV3, and HRW ARW all have maximum reflectivities below about 60dBZ. The HRRR, however, has a band of increased counts of maximum reflectivities below 60 dBZ, but there are also objects at all areas with reflectivities above that limit. The observations do not appear to have any firm upper bound like what is seen in the models. This potentially indicates that the reflectivity algorithms in the non-HRRR models have a cap on reflectivity. The HRRR has a favored band of maximum reflectivity which indicates a similar cap that can be exceeded in certain circumstances. The models and observations all tend to have higher maximum reflectivity with increasing object area. This suggests that the biggest objects tend to be the most intense.

Figures 2.4 and 2.5 show the relationship between the orientation and aspect ratio of radar objects. It can clearly be seen that the models all have average orientation angles that are similar to each other but different from the observed average orientation angle. For the 900 km^2 object threshold, the average orientation angle for the models is between 51° and 56°, while the observed average orientation angle is 20.41°. For the 4500 km^2 threshold, the average orientation angles increase for all models and for the observations, to between 53° and 57° for the models and 25.15° for observed radar objects. The difference between the models and the observations indicates that all of the models are incorrectly orienting reflectivity objects. The high orientation angle of the modeled reflectivity objects suggests that the models favor storms associated with cold fronts. The increased orientation angle with the higher minimum area threshold

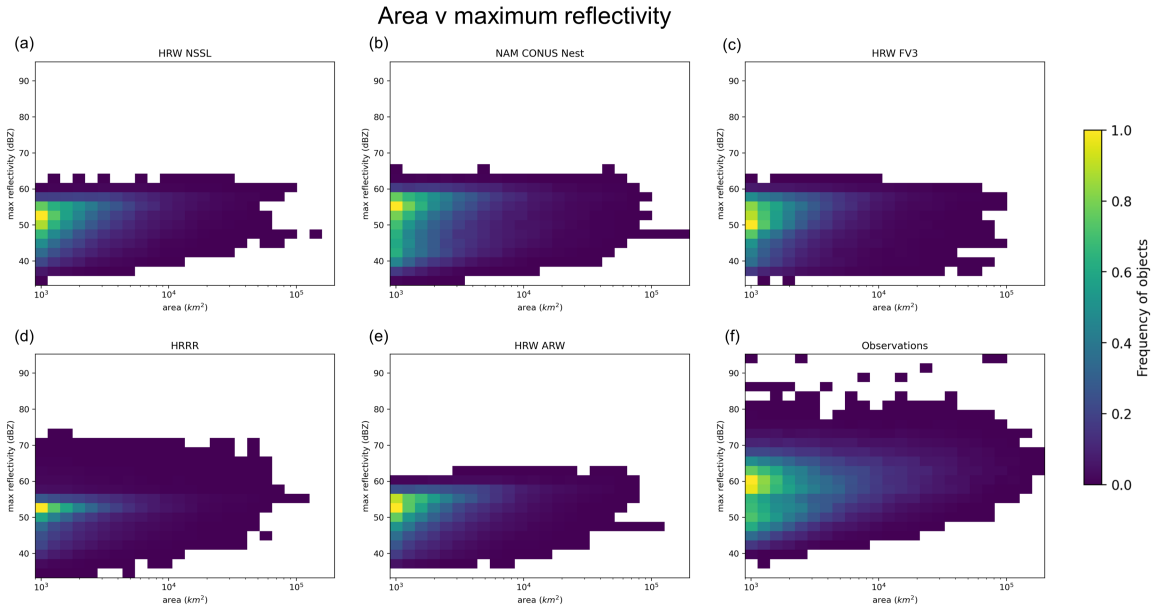


Figure 2.3: As in Figure 2.2 for area and maximum reflectivity in the object with a linear color scale.

indicates that bigger storms are oriented more north-south for both the observed and modeled storms.

The distribution of objects in aspect ratio-orientation space also varies. There is a wider spread of orientation angles for the models than there is in the observations, more noticeably in Figure 2.4. This spread is also visible in Figure 2.5 but it is not as apparent. With a 900 km^2 minimum area, the models and observations have similar aspect ratios, but with a 4500 km^2 minimum area threshold, the models have slightly higher r' (i.e. less round objects). This matches what was seen in Figure 2.2, where the largest modeled objects had less round aspect ratios than the largest observed objects. The modeled reflectivity objects have more variable orientations and the largest modeled objects are less round than the observed objects. The increased spread in the orientation angle in the modeled reflectivity objects suggests that the observed

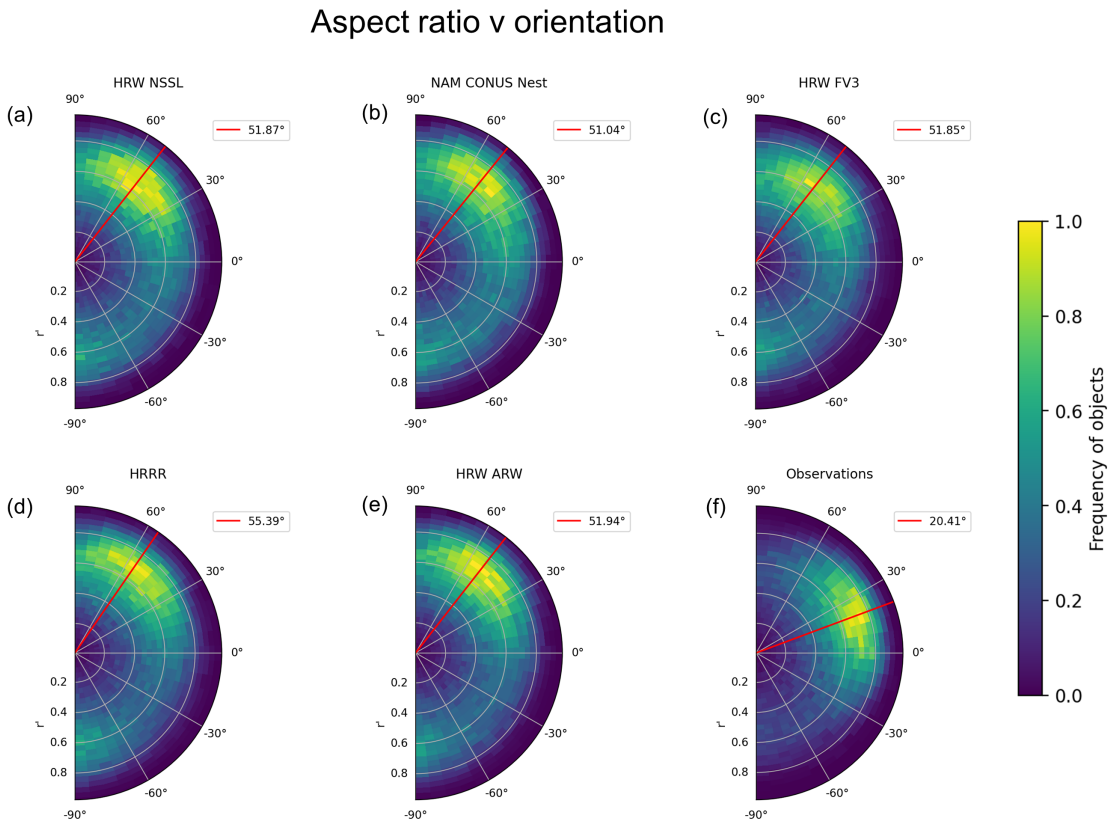


Figure 2.4: Aspect ratio and orientation plotted on polar axes, as described in Figure 2.1, subject to a 900 km^2 minimum area for the (a) HRW NSSL, (b) NAM CONUS Nest, (c) HRW FV3, (d) HRRR, and (e) HRW ARW models and (f) observations. The labeled red line in each plot represents the mean orientation angle for the given plot. Rounder objects are closer to the origin, and less round objects are closer to the edge. The color scale is relative for each subfigure.

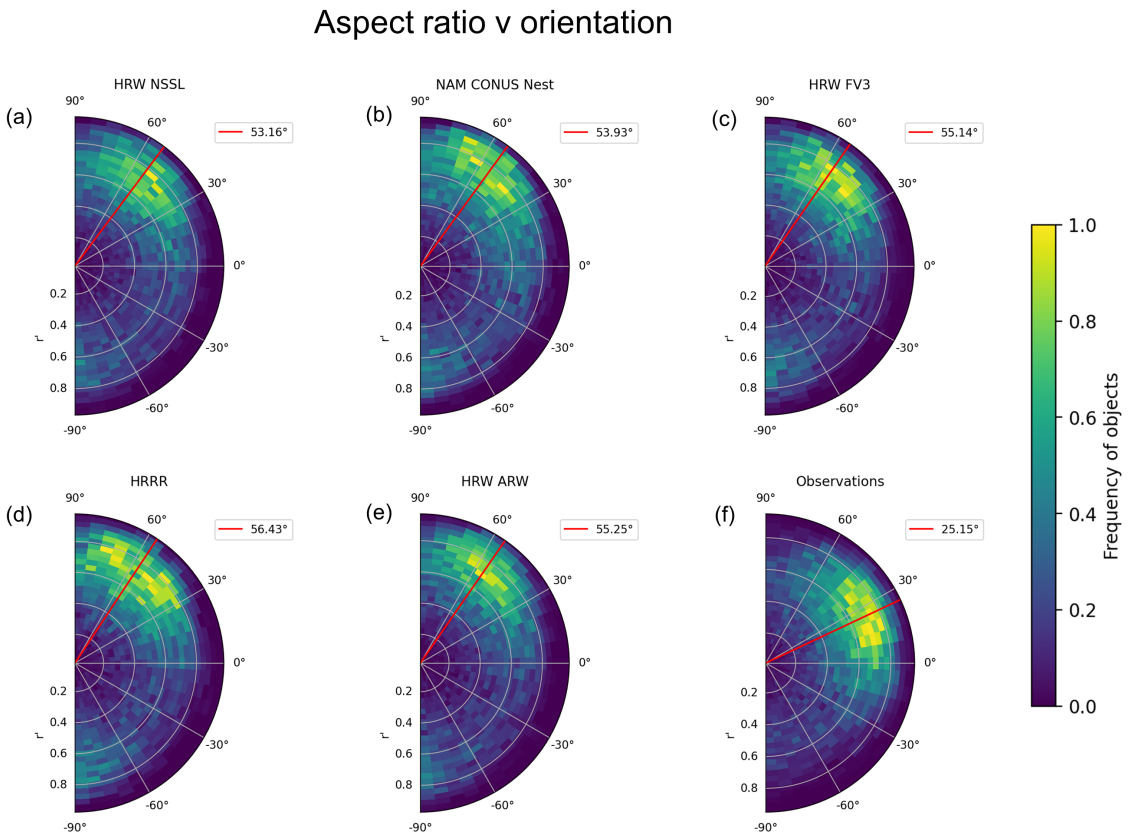


Figure 2.5: As in Figure 2.4 with a minimum area of $4500km^2$.

reflectivity objects have a more consistent preferred orientation angle than the modeled reflectivity objects.

Using the average aspect ratio and orientation angle for objects of a given size can help visualize the differences in object characteristics between models and the observations. The most obvious difference between the models and observations in Figure 2.6 is that the observed ellipses have a much lower orientation angle than the modeled objects (i.e. the observed objects are oriented in a more east-west direction), though the difference is less pronounced for the largest objects. For all models, but especially for the HRW ARW and the HRW FV3, the ellipses are much less round with increasing size, whereas the observed ellipses are only slightly less round. The models differ most with each other in the change in orientation angle with increasing object size. The HRW NSSL, HRW FV3, and HRRR all have greater orientation angles with increasing object size except for the largest object (36000 km^2), but there are very few objects in the largest size bin. The orientation angles of the HRW ARW and NAM CONUS Nest ellipses do not change very much with increasing size. The observed reflectivity object ellipses have an increasing orientation with increasing size, even for the largest ellipse.

For smaller reflectivity object ellipses, the same overall orientation difference between the models and observations is apparent as with the larger ellipses (Figure 2.7). For all models, the ellipses have decreasing orientation angle with decreasing object sizes (i.e. smaller ellipses are oriented more east-west than larger ellipses). The observed ellipses have roughly the same orientation angle for all of the small sizes. The aspect ratios of the ellipses for the models and observations remain roughly constant with decreasing size.

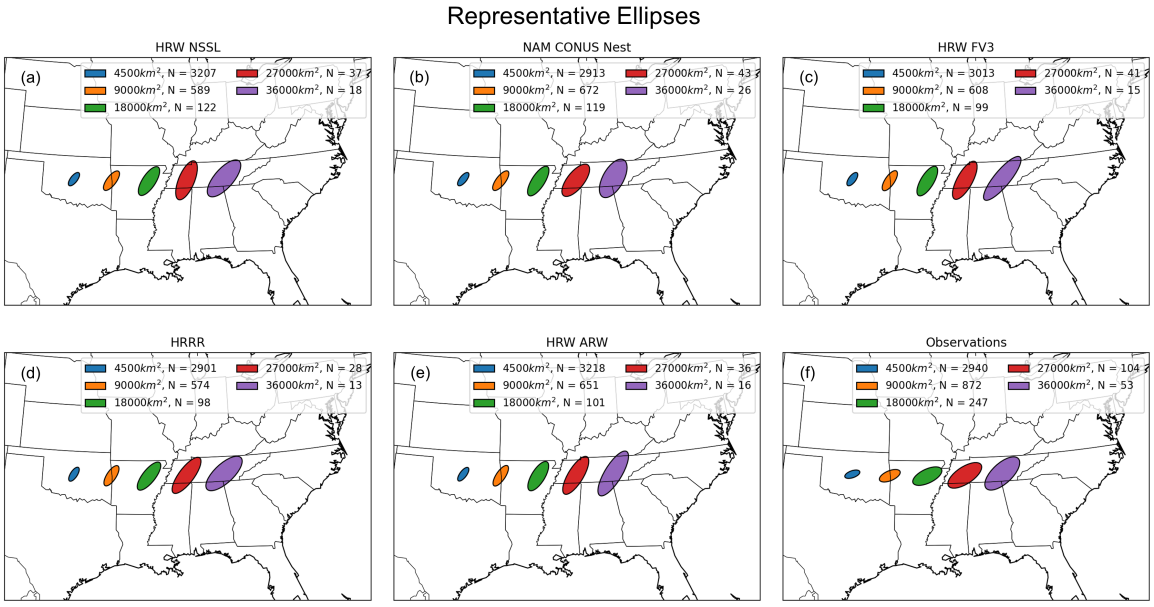


Figure 2.6: Ellipses of various sizes, with aspect ratio and orientation for each determined by the mean aspect ratio and orientation for ellipses of those sizes for the (a) HRW NSSL, (b) NAM CONUS Nest, (c) HRW FV3, (d) HRRR, and (e) HRW ARW models and (f) observations. The locations of the ellipses are not representative and are meant to help provide a sense of scale. The number of objects at a given size is shown in the legend.

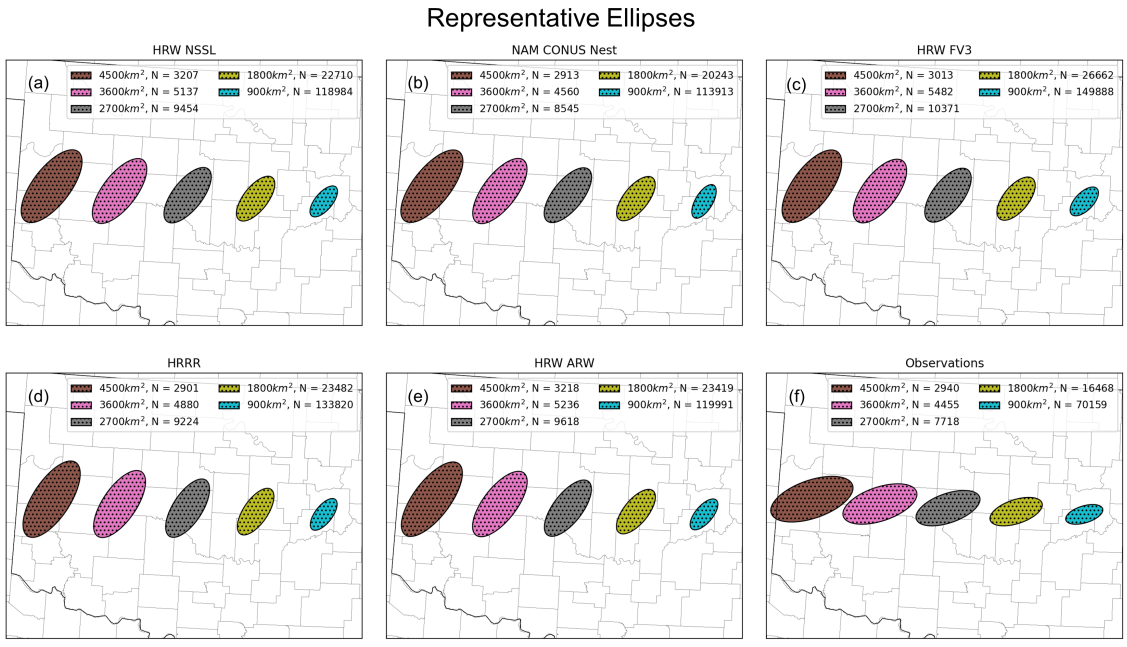


Figure 2.7: As in Figure 2.6, but for smaller ellipses.

2.4 Conclusion

Forecast and observed reflectivity fields were analyzed by identifying convective reflectivity objects for the period from 12 UTC 1 January 2022 through 11 UTC 1 January 2023. Forecasts from the non-time-lagged members of the HREF were used. The area, aspect ratio, orientation, and maximum reflectivity of each object was recorded. The distributions of each of the variables was analyzed for the models and observations by using two dimensional plots of area against aspect ratio, area against maximum reflectivity, and aspect ratio against orientation. To assist in visualizing the differences between models and observations, representative ellipses were generated at various sizes using the mean aspect ratio and orientation associated with objects of those sizes.

The modeled and observed distributions of area against aspect ratio are all broadly similar. Larger objects tend to be less round. This trend is less strong in the observations than in the models, especially for the absolute largest objects. For both the models and observations, the maximum reflectivity within an object increases with increasing object size. For the models, there was a preferred band of maximum reflectivity that was not seen in the observations. For the HRW NSSL, NAM CONUS Nest, HRW FV3, and HRW ARW, this preferred band appears to be a strict cap on the reflectivity, while in the HRRR, there were some objects with higher reflectivities than the preferred band.

The largest difference between the modeled and observed objects was in the orientation angle. Observed objects have a mean orientation angle roughly 20° counterclockwise above the horizontal, while the modeled objects have an orientation angle roughly 50° counterclockwise above the horizontal. In other words, modeled objects were oriented closer to north-south while observed objects were oriented closer to east-west. While both the forecast and observed objects had higher orientation angles with

increasing area, the large difference in orientation is quite notable. The spread of the object placement in aspect ratio-orientation space is also different between the modeled and observed objects. Modeled objects have much more spread in orientation, while observed objects have much less variation.

Johnson and Wang (2012) explored the shape of storms produced by models and compared them to observations, but they identified storms based on the one hour accumulated precipitation field. This approach used top of the hour reflectivity (simulated or observed). The difference between using the accumulated precipitation and top of the hour reflectivity is that accumulated precipitation is also dependent on storm motion, whereas top of the hour reflectivity does not. By identifying storms using accumulated precipitation, modeled storm motion is also being evaluated. Johnson and Wang (2012) acknowledge that the orientation angle in their study, particularly for small objects, is at least partially driven by the storm motion.

The differences in orientation angle can potentially be explained by the models failing to accurately reproduce the storm structure seen in observations. The models may be failing to represent back building cells at the southern end of a convective line (Hitchcock and Schumacher, 2020). These storms extend further to the west and would cause the orientation angle of the observed storms to become more east-west. When the models fail to represent these, the orientation angle would be oriented closer to north-south. Even if the model does represent convection at the southern end of a convective line, if the storms are discrete, they may not be merged into the larger line for this analysis if the storms are too far apart. Models may also struggle to accurately represent stratiform regions of convective lines. These regions would cause the object shape to have a smaller aspect ratio and, if the stratiform region was large enough, an orientation angle closer to east-west. This would also help to explain the lower aspect ratios seen for the largest observed objects. Models may also be struggling to accurately

represent bow and arrow structures (Keene and Schumacher, 2013), which would also have a similar impact on object shapes as failing to accurately represent stratiform regions. Line orientation is also driven by the direction of the mean wind (Johns and Hirt, 1987), so any model error in the mean wind may influence the orientation of the modeled storm.

The ellipses further demonstrate the differences between the modeled and observed reflectivity objects. The difference in orientation angle is obvious at all object sizes. The aspect ratio difference is also apparent. The largest modeled objects are more linear than the largest observed objects. For small objects, there are fewer differences in shape, though the difference in orientation angle remains apparent.

In general, the models look more like each other than the observations. The differences between the models and observations are most apparent in the orientation angle, but the maximum object reflectivity in the models is also quite different from the observed reflectivity. The differences in aspect ratio are less apparent. The hypothesis that modeled storm objects would have similar shape characteristics to observed storm objects was supported for the area and aspect ratio, but the modeled maximum reflectivity within the object and modeled orientation angle were not similar to observations.

These differences in object shape illustrate some potential challenges in identifying modeled storm features from reflectivity. Despite these challenges, modeled reflectivity may still provide useful information about storm structure and evolution. The similarities in object shapes between the models and observations, even with the different orientation angles, do suggest that it is possible to develop an MCS identification method. A method for detecting MCSs in models derived from a method for detecting observed MCSs will be developed.

Chapter 3

How can MCSs be identified in the models?

3.1 Introduction

Convection-allowing models (CAMs) and CAM ensembles can forecast MCSs (Done et al., 2004). The High Resolution Ensemble Forecast System (HREF, Jirak et al., 2012; Roberts et al., 2019) is a CAM ensemble that is used by forecasters at the Storm Prediction Center as the basis for calibrated severe weather guidance (Harrison et al., 2022). The HREF is a multi-physics ensemble, and it has been shown to have more spread than other single-physics ensembles, allowing it to better capture model uncertainty (Roberts et al., 2020). MCSs have been identified in the HREF by using the distribution of updraft helicity (Jahn et al., 2022).

While MCSs can be identified in models by inspection of the output, doing so can be time consuming, especially with ensembles. Automated MCS identification can highlight to forecasters where the model predicts an MCS will occur. Lakshmanan and Smith (2009) developed a method of extracting large amounts of shape characteristics from gridded data and applied that method to identify storm mode from observed reflectivity (though they note that the method could be applied to CAM output, too). Storms were identified following the method in Lakshmanan et al. (2009), which finds clusters of pixels above a certain threshold in the vicinity of each other. Lakshmanan

and Smith (2009) then developed an algorithm to categorize storms as pulse, supercellular, QLCS, or disorganized that utilizes the data mined from the gridded data of the storms.

Storms can also be identified and classified based on reflectivity and mid-level rotation. Potvin et al. (2022) developed such a method that can be applied to either observations or model output. This method was able to identify embedded rotation within larger linear storms. Potvin et al. (2022) first identified storms by using a reflectivity and an area threshold, then merging nearby storms, before applying a maximum reflectivity threshold within each storm. This removed small storms and large storms without an intense core. Objects were then classified as cells or QLCS based on their shape characteristics, and object rotation within each object is determined based on the presence of rotation exceeding a threshold in the model updraft helicity or radar-derived azimuthal shear. Cells were identified within objects by progressively increasing the reflectivity threshold. This identified the most intense areas within each object. The cells were then classified using the same procedure as was done for the larger objects. Potvin et al. (2022) evaluated their method using a survey of 400 storms with 29 respondents. The storm classification algorithm and survey generally agreed with each other.

Machine learning can also be used to classify storms in models. Sobash et al. (2023) compared several different machine learning methods and their ability to identify storm mode (MCS, supercell, and disorganized). Sobash et al. (2023) used a large set of hand-labeled storms, where each storm was labeled by one expert. Each of the machine learning methods in Sobash et al. (2023) produced an annual climatology that roughly matched the observed climatology for each storm mode. Jergensen et al. (2020) also used machine learning to classify storms using observed radar reflectivity and Rapid Update Cycle modeled proximity soundings. The database from Smith et al. (2012)

was used to label each storm. Jergensen et al. (2020) found that gradient-boosted forests yielded the best classifications. Shape characteristics were found to be the most important in classifying storms, followed by wind shear and storm relative wind.

Storms identified using machine learning can be tracked using non-machine learning methods. Haberlie and Ashley (2018a,b) identified and tracked MCSs using observed reflectivity and machine learning. Storms were initially identified in Haberlie and Ashley (2018a) using a reflectivity threshold, then nearby storms were merged. Small storms were discarded, and, where applicable, storms were associated with any nearby stratiform regions. A total of 3000 storms were hand labeled and used to develop a machine learning algorithm. Features related to area were found to be the most important, namely area, solidity (ratio of area to convex hull) and area of reflectivity exceeding an intense threshold. The tracking algorithm developed in Haberlie and Ashley (2018b) used radar reflectivity every 15 minutes to identify MCSs using the procedure from Haberlie and Ashley (2018a). If MCSs from consecutive time steps overlap, they were considered to be part of a swath. If an MCS at one time step overlapped with more than one MCS at another time step, the most similar MCS was kept and added to the swath. Temporal discontinuities were allowed if the length of the discontinuity was less than one hour and 100 km, as measured from the last MCS identified in the first swath to the first MCS identified in the later swath. The two swaths to be connected also needed to be at least 30 minutes long.

The studies mentioned so far use base or composite (two dimensional) radar reflectivity to classify storms, but three dimensional modeled or observed reflectivity can also be used. Short et al. (2023) developed a method that allowed for separate cells at low levels to be merged together into an MCS if they are connected by reflectivity at higher levels. Tracking was accomplished by estimating the storm motion of each separate low level cell and matching cells at a subsequent time step with the forecast

position. Cells were then merged back into lines using the reflectivity at higher levels. The identification and tracking algorithm was initially developed for observed radar reflectivity, but it was modified for use with model data by Short and Lane (2023).

Observational storm tracking algorithms often rely on estimating storm motion to forecast a storm's position at a subsequent time step and and then identifying storms near the forecast position at the subsequent time step (Johnson et al., 1998; Han et al., 2009). Other tracking algorithms use satellite observations (Feng et al., 2021) or model-derived outgoing longwave radiation as a proxy for satellite brightness temperature (Hsu et al., 2023) to track storms.

Though Houze (2004) defines an MCS as having a largest dimension of 100 km or more, preliminary work found that maximum length was not a good discriminator between MCSs and non-MCSs. When using the expert labels described ahead in Section 3.2.2, there was overlap in the maximum length of MCSs and non-MCSs. In other words, the smallest MCSs had a shorter maximum length than the longest non-MCSs. This, together with the development of an observational MCS detection algorithm at the SPC (Wade et al., 2023), were the main reasons for the decision to develop an MCS identification algorithm.

This chapter builds on an algorithm developed by Wade et al. (2023), which is based on the one developed by Haberlie and Ashley (2018a). The algorithm developed here will be optimized by leveraging a survey of Storm Prediction Center forecasters. The algorithm from Wade et al. (2023) is described more fully in the following sections. Because the tracking algorithm developed by Wade et al. (2023) is based on fifteen minute reflectivity that is not available in model output, a new tracking algorithm is developed that takes advantage of hourly maximum reflectivity output from the forecast models.

It has been shown that models can produce MCSs (Done et al., 2004) and they can be identified using automated techniques (Potvin et al., 2022; Sobash et al., 2023; Jergensen et al., 2020; Short et al., 2023). There are also methods of identifying MCSs in observations (Haberlie and Ashley, 2018a; Wade et al., 2023; Johnson et al., 1998; Han et al., 2009). However, there has yet to be a study that applies an algorithm to identify observed MCSs to model output.

The identification algorithm in Wade et al. (2023, hereafter the Wade algorithm) is based on a large dataset of observed MCSs. The Wade algorithm is also used at the Storm Prediction Center to identify MCSs in near real-time as part of an effort to classify severe wind-producing MCSs. By leveraging that algorithm, the work presented here will identify model-produced MCSs the same way as they are identified operationally in observations. It is hypothesized that the Wade algorithm can be optimized to identify MCSs in a way that matches expert labeling, and that the optimizations will identify MCSs in a way that better matches expert labeling than when using the default parameters used in Wade et al. (2023).

3.2 Data and Methods

3.2.1 Model Data

Model output from the non-time lagged members of HREF version 3 were used (configurations summarized in Table 1.1). MCSs were identified using model reflectivity valid at every forecast hour and tracked using hourly maximum reflectivity valid for the preceding hour. To analyze the full year, model MCSs were identified and tracked beginning at forecast hour 9 and continuing through forecast hour 39. The only MCSs considered for analysis were those that existed at any point between forecast hour 12 and forecast hour 35. Even if an MCS existed from forecast hour 9 through forecast

hour 12, it would still be tracked and counted. This was done for all 00 UTC model runs from 1 January 2022 through 31 December 2022. In other words, MCSs were tracked and identified at the top of every hour from 12 UTC on 1 January 2022 through 11 UTC on 1 January 2023.

3.2.2 Expert Labeling

To make sure that the Wade algorithm is identifying MCSs in a way that matches what an expert human would identify, it was necessary to have a dataset of modeled storms labeled by experts. Jahn et al. (2022) initially conducted a survey of Storm Prediction Center (SPC) forecasters that has since been expanded. Forecasters responded to 715 individual cases distributed across all members of the HREF. Cases had between four and eight responses (summarized in Table 3.1). In total, there were fourteen different respondents who answered an average of 280 cases each. The average case had 5.9 responses.

Respondents were given plots of 1 km above ground level model reflectivity at the forecast hour of maximum updraft helicity, the 1 km above ground level model reflectivity one hour prior, and the hourly maximum 2-5 km updraft helicity. They had the option to classify storms as quasi-linear convective system (QLCS), QLCS hybrid, supercell in line, supercell in cluster, discrete supercell, and disorganized. For this study, QLCS in line, QLCS hybrid, and supercell in line were all counted as linear storms. Supercell in cluster and discrete supercell were both counted as supercells. If the plurality of responses for a given case were disorganized (i.e. the number of disorganized responses was greater than any of the other five classifications), the case was counted as disorganized and was discarded from further analysis. Otherwise, the case was classified as the greater of linear or supercell classification. Ties were recorded

Table 3.1: Summary of the number of responses for each case.

Number of responses	Number of cases
4	1
5	241
6	348
7	100
8	24

but were discarded from further analysis. The results of the expert labeling for each model are shown in Figure 3.1. For each case, the expert categorization, the location of the storm, the model, the forecast time, and the date and run of the model were recorded.

3.2.3 MCS identification and tracking algorithm

3.2.3.1 MCS identification

The MCS identification algorithm is based on Wade et al. (2023). Wade et al. (2023) incorporated characteristic MCS scales and the most useful predictors in Haberlie and Ashley's (2018a) machine learning model into a non-machine learning algorithm tunable to maximize detection of derecho-producing MCSs. The algorithm described in Wade et al. (2023) masks areas outside of a square neighborhood of size 81 km around intense radar reflectivity (above 50 dBZ) to only capture the most intense storms. It then masks areas where the neighborhood mean radar reflectivity within a 21 km square is too low (below 25 dBZ) to remove isolated areas of intense reflectivity. From that masked field, a convective reflectivity (40 dBZ) contour is drawn and objects are identified. Nearby objects within 5 km are merged together using a buffer. Merged objects

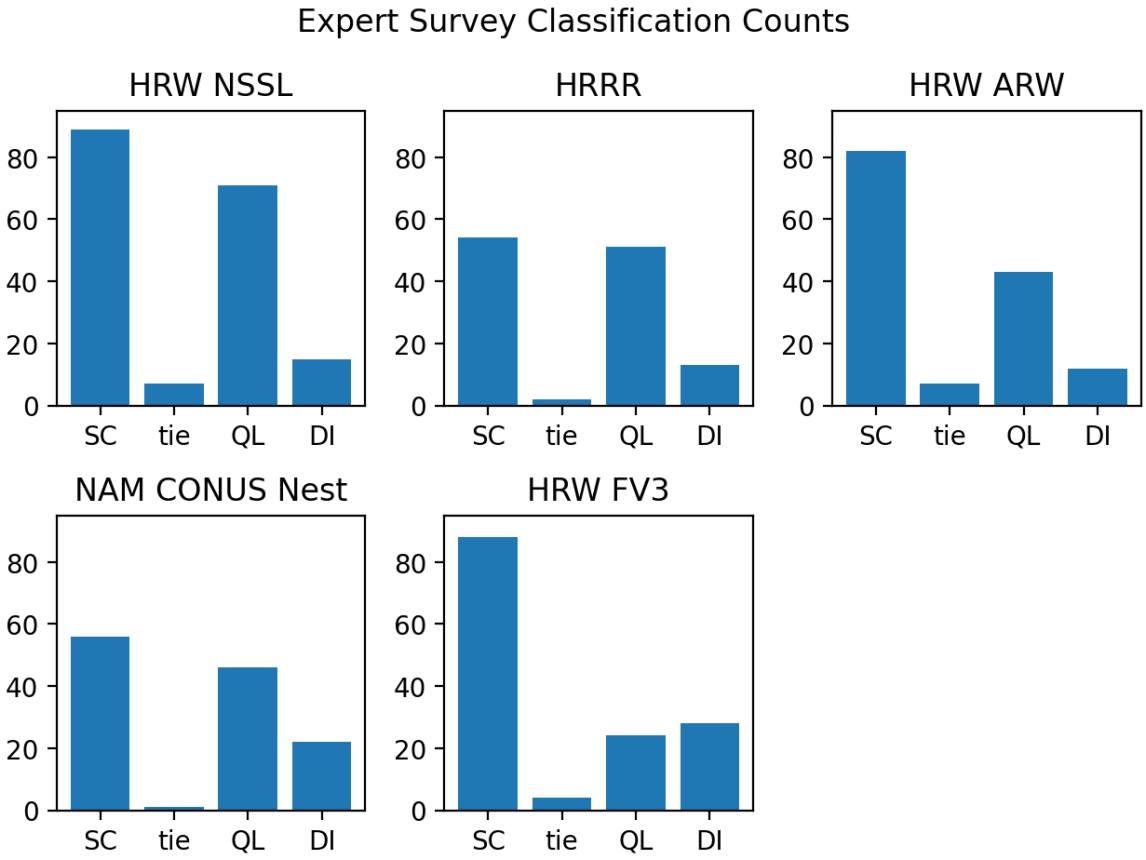


Figure 3.1: The counts of expert-labeled storm mode for each model in the HREF. SC are supercells, QL are linear, DI are disorganized.

are then checked for sufficient length (120 km), area of intense reflectivity (100 km^2), and areal density of intense reflectivity (0.01 pixels per square kilometer). Another buffer of 20 km is applied to any remaining objects to merge nearby objects.

The Wade algorithm was written for observed radar reflectivity on a 1 km grid. To modify the Wade algorithm for use with models on a 3 km grid, all parameters based on grid size were changed by a factor of 3 or 9 for length and area, respectively. These parameters include the size of the neighborhood used to identify intense radar reflectivity, the size of the neighborhood used to calculate neighborhood means, the area of intense reflectivity within identified objects, and the areal density of intense reflectivity. In addition, the reflectivity values used were changed to percentile values for each model because each model does not reproduce observed reflectivity in a distribution that matches observations. The remaining parameters were based on actual sizes of objects so did not need to be changed.

3.2.3.2 MCS tracking

Wade et al. (2023) uses a tracking algorithm based on Haberlie and Ashley (2018b) that relies on radar mosaics every fifteen minutes. Other studies, such as Short and Lane (2023), base tracking on estimated storm motion from model reflectivity every ten minutes. The tracking methods developed in Britt et al. (2024) and Skinner et al. (2018) also rely on model reflectivity available every five minutes. Because model reflectivity from the members of the HREF is only available hourly and to avoid estimating storm motion, a new tracking algorithm was developed.

The tracking algorithm is based on the hourly maximum radar reflectivity field from the model (Figure 3.2). For every forecast hour, the hourly maximum radar reflectivity fields from the three previous hours are combined to produce a three hourly maximum radar reflectivity field. MCSs are identified using the modified Wade algorithm for

the forecast hour and each of the previous four forecast hours (i.e., for the entire duration of the three hourly maximum radar reflectivity field). The intense reflectivity threshold (equivalent percentile of 50 dBZ) from the modified Wade algorithm was used to identify contours in the three hourly maximum reflectivity field. Since intense reflectivity is a necessary condition of an MCS, a break in the intense reflectivity contour means that the MCS must not have existed at some point in the duration of the hourly maximum field. For each intense reflectivity contour, the number of intersections between the contour and the field of identified MCSs was counted. If there were at least four such intersections, the intense reflectivity contour was checked again to determine if there was an intersection with an MCS at each hour. If there was one at each hour, all of the MCSs identified that intersected with the intense reflectivity contour were retained. This method keeps all sides of line splits or mergers as long as there is a contiguous region of intense reflectivity in the hourly maximum reflectivity field.

3.2.4 Optimizing the identification algorithm

All parameters in the modified Wade algorithm were checked for their sensitivity to object detection. The minimum length requirement (cline_len), the buffer size for the initial object mergers (cline_buffer), and the size of the neighborhood used to compute the mean reflectivity (nsize_mean) were found to impact the object detection. To determine the set of parameters that best identify MCSs, a range of values for each of the three parameters was used. The minimum length requirement varied from 100-155 km in 5 km increments, the buffer size varied from 1-15 km in 1 km increments, and the size of the neighborhood used to compute the mean reflectivity varied from 1-15 grid squares (3-45 km per side) in 1 grid square increments. This yielded a total

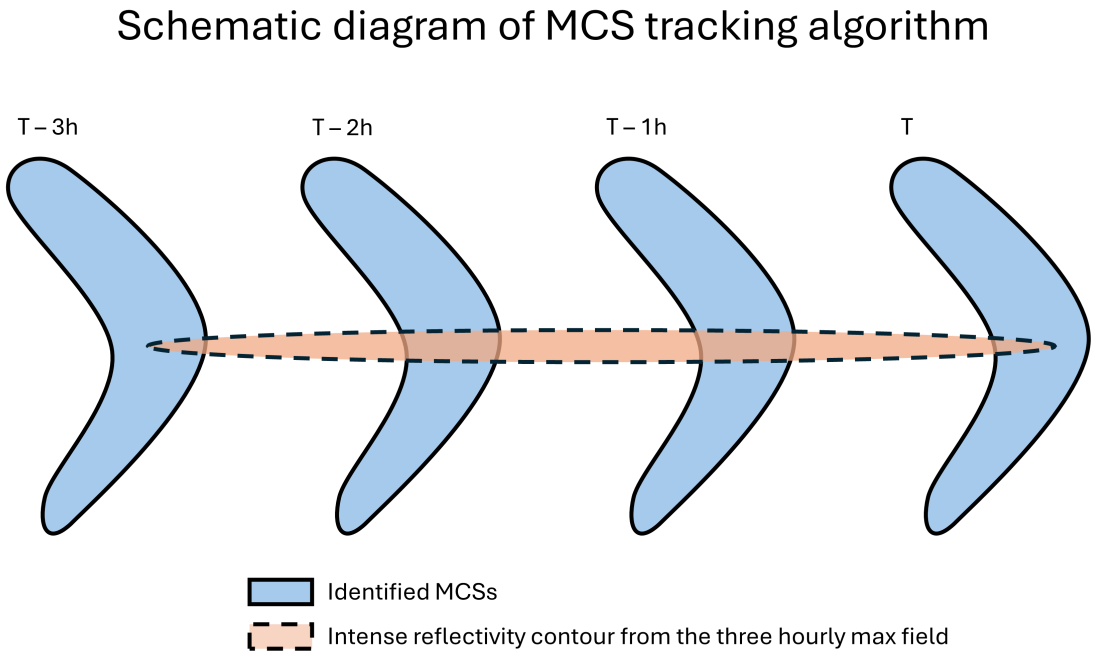


Figure 3.2: A schematic of the MCS tracking algorithm for time T. MCSs were identified at times $T-3$, $T-2$, $T-1$, and T. The intense reflectivity contour (black dashed line) is 50 dBZ.

of 2700 different combinations of parameters. For each case from the expert survey and for each combination of parameters, the modified Wade algorithm was applied. The distance from the edge of the nearest object identified by the Wade algorithm to the storm in the expert survey was determined (if the storm was inside of an object identified by the Wade algorithm, the distance was recorded as 0).

To optimize the set of parameters such that algorithm best matches the expert labeling, the distance threshold used to determine a hit, miss, or false alarm was varied. Objects within the threshold distance counted as a hit if the expert survey classified the storm as linear. If there was no object within the threshold distance and the expert survey classified the storm as linear, it was counted as a miss. If there was an object within the threshold distance but the expert survey classified the storm as non-linear, it was counted as a false alarm. If there was no object within the threshold distance and the expert survey classified the storm as non-linear, it was counted as a correct null.

By varying the distance threshold, a curve on a performance diagram (Roebber, 2009) could be generated. To determine which parameter combination best identified MCSs, the area to the left of the performance diagram curve was calculated. This is equivalent to the area under a precision-recall curve, which is used to assess the performance of rare event binary classification (AUC-PR, Sofaer et al., 2019). AUC-PR is distinct from the area under a receiver operating characteristic curve (AUC-ROC, Marzban, 2004), which has statistical meaning in determining whether the forecast can discriminate between "yes" and "no" events (Hanley and McNeil, 1982). AUC-PR does not have a statistical meaning, but it can still assist in qualitatively evaluating the performance of a system. A schematic showing AUC-PR for several different forecasts is presented in Figure 3.3. The blue forecast has the highest AUC-PR, followed by the green, red, and orange forecasts. Even though the green forecast is "above" the blue

for some areas, the much lower SR of the green forecast means it has a lower AUC-PR than the blue. The red forecast has a lower AUC-PR than the green even though at points the red forecast is "above" the green. The orange forecast is "below" all of the other forecasts at all times and thus has a lower AUC-PR than all of them. Each model will be optimized separately, since Jahn et al. (2022) showed that each model of the HREF did not organize storms in the same way and since the distribution of the number of each storm type is not the same across all of the models (Figure 3.1)

The set of parameters with the highest AUC-PR for each model was used to identify modeled MCSs, and combined with the tracking algorithm, produced a full year analysis of MCSs produced by each model in the HREFv3.

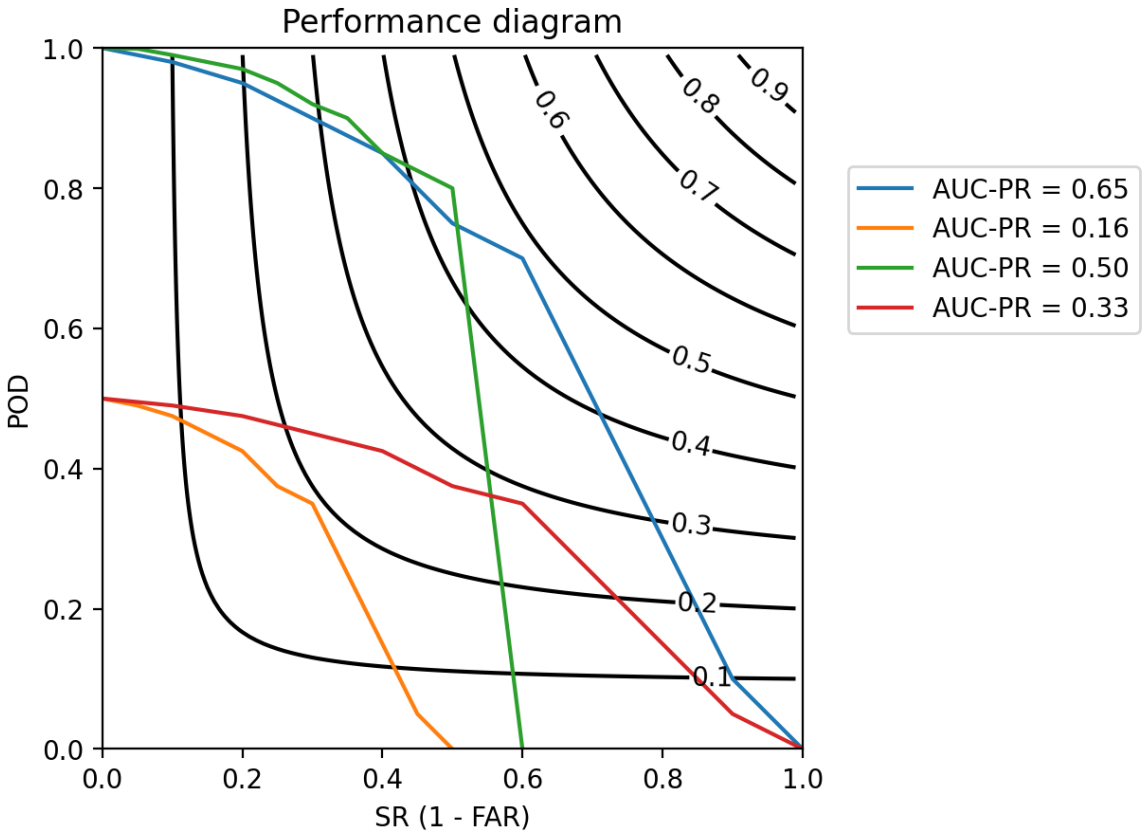


Figure 3.3: Performance diagram showing how AUC-PR varies with several different forecasts. The x-axis is success ratio (SR), which is equivalent to 1 minus the false alarm ratio (FAR). They y-axis is probability of detection (POD). The black contours are lines of constant critical success index (CSI).

3.3 Results

3.3.1 Parameter optimization

For each model in the HREFv3, the optimized parameter set that best matches expert labeling has a higher AUC-PR than the default parameter set (Figure 3.4). The optimized parameters for both HRW NSSL and HRW ARW have an AUC-PR that is very close to the default parameters. The other three models all have a much greater difference between the optimized and default parameters. Though it was not used as an optimization score, the maximum CSI for the optimized parameters is higher than the maximum CSI for the default parameters.

The NAM CONUS Nest and HRW ARW optimized parameters have the highest AUC-PR, which indicates that the modified Wade algorithm performs best (that is, identifies MCSs in a way that best matches expert labeling) for identifying MCSs in those models. The HRW NSSL and HRRR optimized parameters have the next highest AUC-PR. The HRW FV3 optimized parameters have the lowest AUC-PR, which suggests that the modified Wade algorithm struggles to identify MCSs in that model more than in any of the other four models. Even though the HRW FV3 optimized parameters have the lowest AUC-PR out of all of the optimized parameters, it is still higher than the AUC-PR of the default parameters for the HRW FV3. The survey results for the HRW FV3 also indicate that it produces more supercells and fewer MCSs than the other four models.

The optimized parameters for each model and the default parameters are shown in Table 3.2. Since nsize_mean controls the size of the window over which a 25dBZ (or percentile equivalent) reflectivity minimum threshold is applied, a larger value means more smoothing. This means that more areas will be masked out. Conversely, a smaller

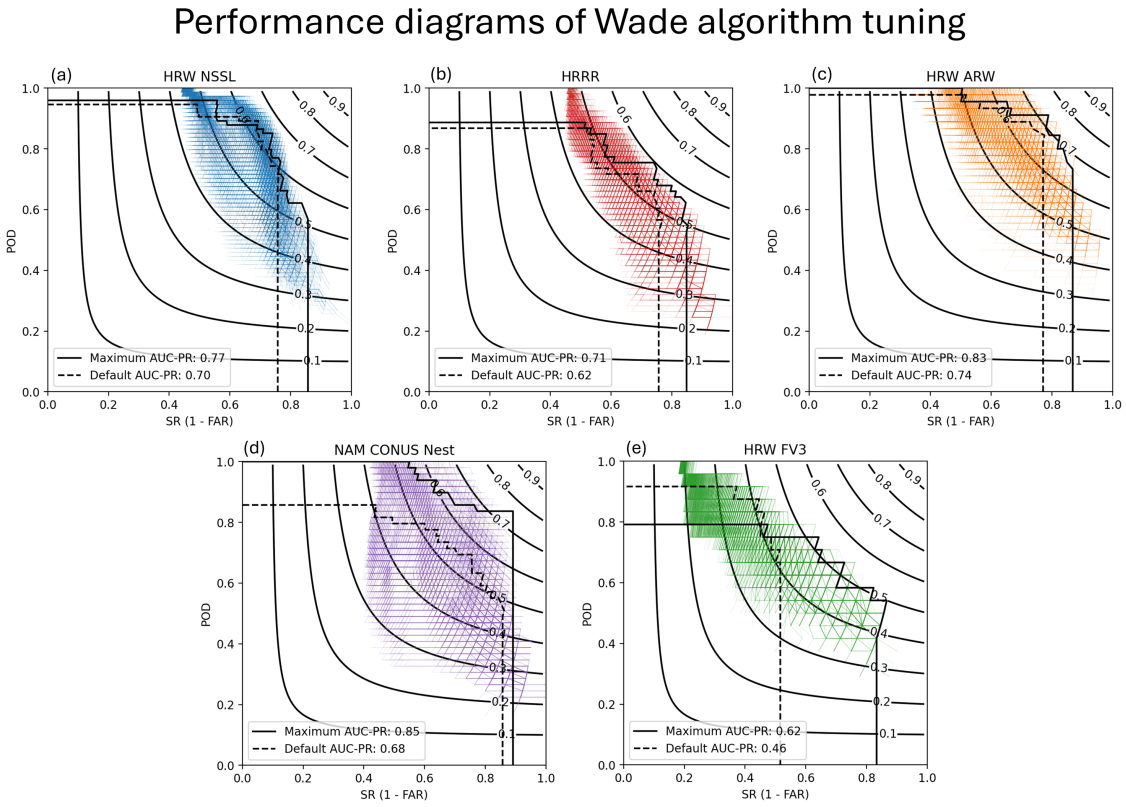


Figure 3.4: Performance diagrams showing the ability of the modified Wade algorithm to identify MCSs as labeled by the expert classification for each model. Each colored line is the performance of a single parameter set. The solid black line is the parameter set with the highest AUC-PR for each model and the dashed black line is the default parameters.

Table 3.2: Summary of the default and optimized parameters for the modified Wade algorithm.

Model	nsize_mean (grid squares)	cline_buffer (km)	cline_len (km)
Default	7	5	120
HRW NSSL	6	1	110
HRRR	11	9	150
HRW ARW	5	3	130
NAM CONUS Nest	5	10	150
HRW FV3	9	2	145

threshold means less smoothing and less masked out areas. The HRRR and the HRW FV3 have a higher optimized nsize_mean than the default so will identify fewer MCSs. The other three models all have an optimized nsize_mean that is lower than the default, so will identify more MCSs. The cline_buffer parameter merges nearby objects before applying the cline_len filter, so a larger cline_buffer means more objects will be merged so more objects will be kept by the cline_len filter. The HRRR and NAM CONUS Nest have a larger optimized cline_buffer than the default so more objects will be identified. The other models have a lower optimized cline_buffer so fewer objects will be identified. All models except the HRW NSSL have a larger optimized cline_len than the default. This suggests that the MCSs identified by those models will be larger than the default, since objects longer than 120km but shorter than the optimized cline_len will be discarded. There will also be fewer objects identified since the filter is more strict. The HRW NSSL has a smaller cline_len, which means smaller objects and more objects will be identified.

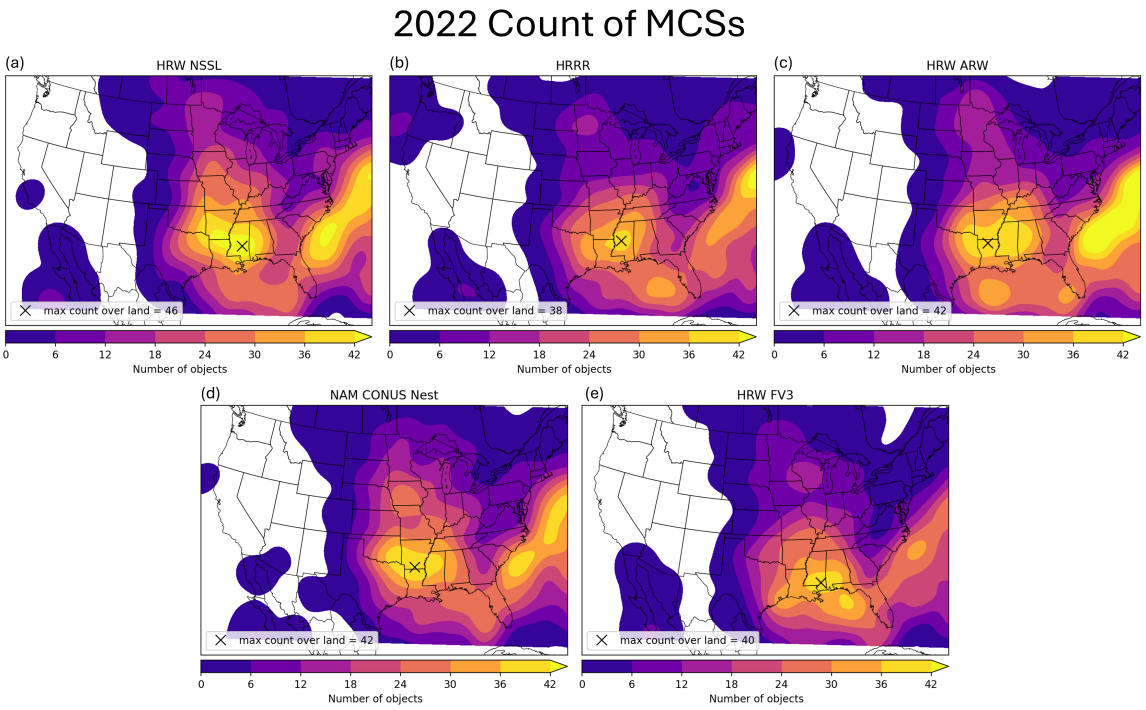


Figure 3.5: The number of modeled MCSs in 2022 for the (a) HRW NSSL, (b) HRRR, (c) HRW ARW, (d) NAM CONUS Nest, and (e) HRW FV3. The black X marks the location of the maximum count over land. The count at the maximum is denoted in the legend of each plot.

3.3.2 Full year analysis

A smoothed full year count of MCSs for each of the models is shown in Figure 3.5. The smoothing is based on Hitchens et al. (2013), with a 40 km maximum filter applied, followed by Gaussian smoother with a σ of 120 km. All five models have a continental maximum in the southern Mississippi River valley. The maximum number of MCSs per year for each model range from 38 for the HRRR to 46 for the HRW NSSL. The distribution of the HRW NSSL and the NAM CONUS Nest both extend further west and north than any of the other three models. The distributions of the HRW FV3 and HRW ARW extend further east than the remaining models. The differences in distribution may be due to certain optimized parameters favoring larger MCSs or it may be due to the model producing more MCSs.

The smoothed one year counts compare favorably to the unsmoothed multiyear climatologies in Haberlie and Ashley (2019, Figure 1) and Wade et al. (2023, Figure 2) (reproduced in Figure 3.6). The climatology in Haberlie and Ashley (2019) has a maximum above 45 MCSs per year in southern Alabama, while the climatology in Wade et al. (2023) has two maxima above 36 MCSs per year in the Ohio River valley and near the Oklahoma, Kansas, Missouri, and Arkansas border.

To attempt to assess whether the differences in the model counts of MCSs are due to parameter differences or inherent differences in the models, the average area of MCSs that occurred at each grid point was found (Figure 3.7). The HRW NSSL has the lowest average areas of all five models, which is because the HRW NSSL has the lowest cline_len. Smaller objects that would have been filtered out by the optimized cline_len parameter from the other models are kept. The HRRR, HRW FV3, and NAM CONUS Nest have higher areas because they have the highest cline_len parameter. The

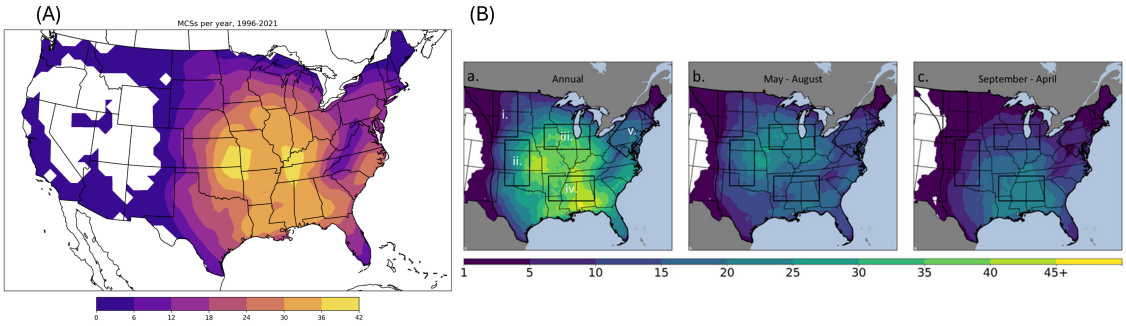


Figure 3.6: (A) Figure 2 from Wade et al. (2023) showing the average number of MCSs for 1996-2021. (B) Figure 1 from Haberlie and Ashley (2019) (a) shows the annual number of MCSs for 1996-2017, (b) and (c) show seasonal counts from May-August and September-April, respectively.

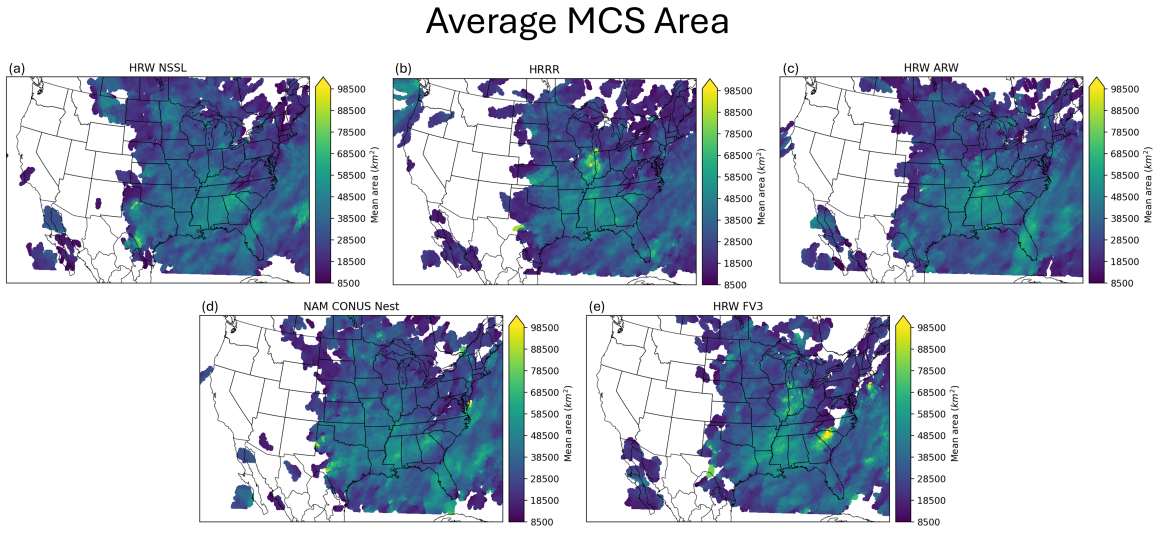


Figure 3.7: The average area of modeled MCSs in 2022 for the (a) HRW NSSL, (b) HRRR, (c) HRW ARW, (d) NAM CONUS Nest, and (e) HRW FV3.

differences in area suggests that at least some of the differences in count are driven by different MCS average area between models.

A kernel density estimate of the centroid of each individul MCS identified at the top of each hour was made to better assess the differences in spatial distribution between the models (Figure 3.8). Only centroids over land were considered for this analysis. Because it only considers centroids, this analysis removes any influence of MCS size. Overall, the distributions are largely similar, however there are some differences. The distribution of NAM CONUS Nest and HRW ARW centroids extends farther into eastern Oklahoma than any of the other three models. While all five models have an extension of the distribution into the upper Mississippi River, the NAM CONUS Nest and HRW NSSL both have a bigger extension than the other three models. The distribution of the HRW FV3 extends further east than any of the other models. The differences in the distributions here demonstrate that there are some inherent differences between the models' MCS identification that are not driven by differences in the size of the MCSs.

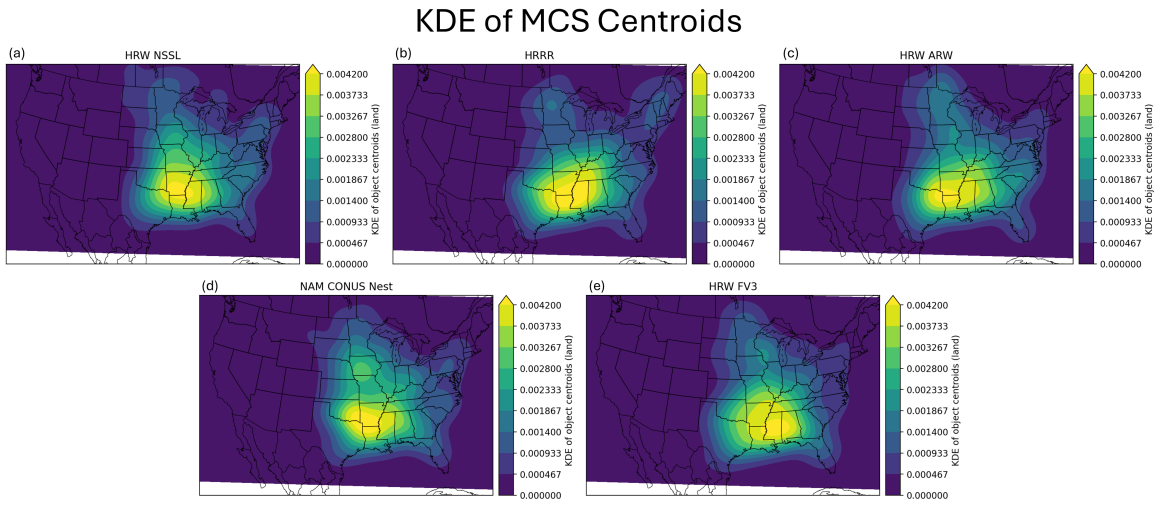


Figure 3.8: KDE of modeled MCS centroids in 2022 for the (a) HRW NSSL, (b) HRRR, (c) HRW ARW, (d) NAM CONUS Nest, and (e) HRW FV3. Only centroids over land were considered for the analysis.

3.4 Conclusion

A method of identifying observed MCSs was modified to identify model MCSs. The parameters used to identify modeled MCSs were optimized using results from a survey of SPC forecasters. MCS identification was sensitive to the following parameters: the minimum length of the MCS (cline_len), the size of the buffer used to merge nearby potential MCSs (cline_buffer), and the size of the averaging window used to identify areas of convection (nsize_mean). The forecasters labeled modeled storms as supercells or MCSs, and the parameters of the MCS identification algorithm were chosen to maximize the AUC-PR. The optimized algorithm identified modeled MCSs in a way that matched expert labeling better than the default algorithm. This was done for each model in the HREFv3. A novel tracking algorithm was developed that tracked MCSs based on the hourly maximum reflectivity.

The cline_len parameter filters out objects smaller than the value of the parameter, so a small cline_len tends to keep more, smaller objects than a larger value.

The cline_buffer parameter merges nearby potential MCS objects before applying the cline_len filter, so a larger cline_buffer means more objects are merged together and will be kept by the cline_len filter. A larger nsize_mean means more smoothing, which will lead to more areas being masked out and fewer MCSs being identified.

The resulting optimized parameters and the tracking algorithm were used to analyze a full year of model data. The single year spatial distribution of modeled MCSs was consistent with multi-year observed climatologies, thought the models all had fewer MCSs than were observed. The HRW NSSL's and NAM CONUS Nest's distributions both extended further west and north than the other models, while the HRW FV3 and HRW ARW both extended further east. All models had their continental maximum in the lower Mississippi River valley.

The HRW NSSL had the smallest MCSs because it had the smallest optimized cline_len. The smaller cline_len means that smaller objects not filtered out. The other models have larger optimized cline_len, and so discard smaller objects.

A KDE analysis of MCS centroids was done to isolate the MCS distribution from the area. This analysis revealed differences in the spatial distribution of MCSs between the models that were less apparent when comparing the MCS counts. The NAM CONUS Nest and HRW ARW centroid distribution extends further into eastern Oklahoma than the other three models. The NAM CONUS Nest and HRW NSSL both have bigger extensions of the centroid distribution into the upper Mississippi River than the other three. The distribution of the HRW FV3 also extends further east than any of the other three models.

The hypothesis, that the Wade algorithm could be optimized and that optimization would improve the algorithm's ability to identify MCSs in a way that matches expert labeling was supported. The Wade algorithm was optimized to match expert labeling and, when using the optimized parameters, the algorithm identifies MCSs in a way

that matches expert labeling better than the default parameters used in Wade et al. (2023).

The MCS identification and tracking algorithm developed here will be used to compare modeled MCSs to observed MCSs. Model performance in producing MCSs will be evaluated by comparing MCSs identified in the models using the modified Wade algorithm with MCSs identified in observations using the original Wade algorithm. In addition to simply classifying MCSs as hits, misses, or false alarms, the placement and direction errors of the modeled MCSs will also be evaluated.

Chapter 4

How good are the models at forecasting MCSs?

4.1 Introduction

MCSs can cause significant severe weather. Smith et al. (2012) found that 25.3% of tornadoes, 6.4% of significant hail reports (diameter $\geq$ 2"), and 44.3% of significant wind reports (gusts $\geq 65kt$) were associated with linear systems, compared to 71.8% of tornadoes, 95.8% of significant hail reports, and 31% of significant wind reports being associated with supercells. Smith et al. (2012) also found differences in the spatial and temporal climatologies between supercells and linear storms. In addition to the differing frequencies of tornadoes between supercells and linear storms, Jahn et al. (2022) found that the probability of a tornado was less in an MCS than for an isolated supercell given otherwise similar environmental conditions. MCSs are also associated with heavy rainfall (Da Silva and Haerter, 2023), and the fraction of precipitation associated with MCSs is projected to increase across the US as the Earth's climate warms (Haberlie et al., 2023). MCSs also have downstream impacts on weather. MCSs can leave behind cold pools and outflow boundaries that impact weather later on in the day, acting as forcing for convective initiation. Poor MCS forecasts over the US are also associated with poor forecasts of the 500 hPa field over Europe (Parsons et al., 2019).

Spatial verification of high resolution model output can be challenging (Gilleland et al., 2009). Small displacement errors result in a double penalty, where the forecast system is penalized both for missing an observed event and for generating a false alarm

(as demonstrated by Figure 1 in Ahijevych et al., 2009). Object-based verification is one way to verify spatial forecasts that gives credit for "near misses" (Gilleland et al., 2009). Objects are identified in both the forecast and observation fields, then forecast and observed objects are matched with each other. The matching can be done using using simple measures such as distance or more complex fuzzy logic algorithms that take into account more object characteristics (as in Brown et al., 2007).

MCSs can be identified using a three dimensional reflectivity field, either forecast or observed. Short and Lane (2023) developed a method of MCS identification identifies cells at lower levels and merges them based on reflectivity at higher levels. Short and Lane (2023) used a convection-allowing model (ACCESS-C) centered over Darwin, Australia, and found that their model produced too many MCSs oriented parallel to the ambient wind and ambient wind shear. These storms are often associated with extreme rainfall, so the model also produced too much rainfall. Modeled storm orientation was found to be different from observed. Observed storms most often were oriented meridionally, reflecting cold pool-driven storms. The model, however, suggested that storms were oriented along the ambient flow. There were also differences in the time of day of the storms, with the model having a peak in storms one to two hours after observed, and the model also had a broad secondary peak that was not seen in observations. Their classification scheme divided MCSs by the location of the stratiform region, and their model produced an MCS distribution across the classification scheme that was largely similar to what was observed. Overall though, Short and Lane (2023) found that the model was fundamentally different from the observations in terms of convective organization.

MCSs can also be identified by using vertically integrated liquid (VIL). Using VIL, Pinto et al. (2015) found that the HRRR was able to reproduce both the count and size distribution of MCSs over the eastern US. The aspect ratios, orientation, and diurnal

cycle were also reproduced by the model. There was a high bias in the number of MCSs produced early in the forecast period, but a low bias by the end of the period. Spatially, the model had a high bias over the Great Plains, but there was a low bias in the southeast and Mississippi River valley.

Skinner et al. (2018) developed an object-based verification procedure based on the Method for Object-Based Diagnostics and Evaluation (MODE, Brown et al., 2007). The method in Skinner et al. (2018) identifies storms by using percentile thresholds of reflectivity in both models and observations, while rotation was identified using updraft helicity in the models and radar-derived azimuthal shear in observations. Objects are matched between the model and observed field using an interest score that takes into account the objects' positions and shapes. The interest scores attempts to quantify objects that are "close enough" to be matched. The interest score used in Skinner et al. (2018) is the fuzzy logic matching method used in MODE (Brown et al., 2007). The model had good performance (high probability of detection, low false alarm ratio) early in the forecast period, though performance degraded with increasing lead time. There was also a high bias early in the period that decreased towards one throughout the forecast time. Skinner et al. (2018) did note that their object-based verification, in general, does not consider the intensity of objects beyond the intensity required to be identified.

Potvin et al. (2022) developed another method of identifying and classifying storms in modeled or observed radar reflectivity and mid-level rotation. Storms were first analyzed using reflectivity and area thresholds, before merging nearby storms together. A maximum reflectivity threshold was then applied to remove large storms without a strong reflectivity core. Storms were classified based on their shape and rotation characteristics into one of seven classes. The reflectivity threshold to identify storms

was successively increased to identify and classify cells within the overall storm. This method was found to agree reasonably well with expert labels.

Flora et al. (2019) identified mesocyclones using a watershed algorithm, which requires a minimum rotation threshold and a higher threshold that must be met within the object. Mesocyclone objects were identified in the forecasts using the ensemble probability of exceeding a certain updraft helicity threshold. Observed mesocyclones were identified using radar-derived rotation. Verification was done using the procedure set out in Skinner et al. (2018). Flora et al. (2019) found that some of the largest errors occurred in MCSs because the modeled mesocyclones associated with MCSs were much larger than the observed rotation tracks. Flora et al. (2019) note that while some of the modeled mesocyclone objects associated with MCSs overlapped with the observed rotation tracks, the modeled objects were so much larger that the centroid of the modeled objects were often much further away.

Britt et al. (2024) used the algorithm developed in Potvin et al. (2022) and the verification procedure described in Skinner et al. (2018) to compare quasi-linear convective systems (QLCSs) between the Warn-on-Forecast System (WoFS) and observations. WoFS is a ensemble forecast system designed to provide forecasts at lead times less than 6 hours. Britt et al. (2024) found that WoFS has skill in forecasting QLCSs, with CSI of 0.44 for a 3-hour lead time and 0.42 for a 6-hour lead time. Non-QLCSs had higher FAR and lower POD and CSI than QLCSs, which suggests that while WoFS is overforecasting other modes, it does better at predicting QLCSs. QLCSs in the WoFS were also larger, more intense, and more round than the observed QLCSs. Britt et al. (2024) also developed a novel method to evaluate a QLCS's curviness and found that WoFS QLCSs were more curved and were oriented more counterclockwise with increasing forecast time. These orientation and shape differences were attributed to

WoFS's overforecast of the intensity of the northern bookend vorticies and rear inflow jets.

Chapter 3 modified the Wade algorithm to be able to identify MCSs in forecast models. Expert labels were used to optimize the parameters used in the Wade algorithm. The models produced a one-year climatology that, when smoothed, roughly matched the multi-year climatology from Wade et al. (2023).

This work compares MCSs identified in observations using the Wade algorithm with those identified in models using the modified Wade algorithm. This will allow an assessment of the models' ability to forecast MCSs. It is hypothesized that applying the optimized parameters developed in Chapter 3 to the Wade algorithm will result in improved forecasts compared to the default parameters.

4.2 Data and Methods

4.2.1 Model Data

MCSs were identified and tracked using model output from the non-timelagged members of the HREF (configuration summarized in Table 1.1) as described in Section 3.2.1.

4.2.2 MCS identification and tracking algorithms

The MCS identification algorithm for both models and observations and the MCS tracking algorithm for observations are based on the algorithm in Wade et al. (2023). The modeled MCS identification algorithm was modified and tuned and a different MCS tracking algorithm for models was developed in Chapter 3. The modification and tuning are described briefly in the next section.

4.2.2.1 MCS identification

The Wade algorithm identifies large areas of convective reflectivity, with minimum area and density requirements of convective reflectivity. Nearby objects are also merged together. The full algorithm is described more fully in Section 3.2.3.1.

Chapter 3 found that MCS identification was sensitive to the minimum length requirement (cline_len), the buffer size for the initial object mergers (cline_buffer), and the size of the neighborhood used to compute the mean reflectivity (nsize_mean). The optimized and default values for each of those parameters is shown in Table 3.2. MCSs were identified using both the optimized and default parameters. The optimization was done to identify MCSs in a way that matched expert-labeled storms.

4.2.2.2 Observed MCS tracking

Wade et al. (2023) uses a tracking algorithm based on 15 minute radar imagery. Every 15 minutes, MCSs are identified using the algorithm described in Section 4.2.2.1. A 100 km radius buffer is added to each MCS polygon. At the next time step, the MCS polygon with the most overlap to the one at the prior time step is counted as the tracked MCS. Splits and mergers are not kept, so the smaller pre-merger object is discarded, and the post-split object with the smaller overlap is discarded. A one time step gap was allowed to account for MCSs that briefly fail to meet the identification criteria due to radar blockage or artifacts.

4.2.2.3 Model MCS tracking

Modeled MCSs are tracked using the algorithm described in Chapter 3. For every forecast hour, hourly maximum reflectivity from the last three hours and top of the hour instantaneous reflectivity from the last 4 hours are used (i.e., from the start to the end of the three hourly maximum period). MCSs are identified in the instantaneous

reflectivity fields using the modified Wade algorithm described in Section 4.2.2.1. The three hourly maximum reflectivity fields are combined into a three hourly maximum reflectivity field. A continuous swath of 50 dBZ (or percentile equivalent) reflectivity in the three hourly maximum that connects identified MCSs in each of the instantaneous reflectivity fields is needed for an MCS to count. This method keeps all sides of splits and mergers.

4.2.3 Evaluation and Verification

The model MCS forecasts were evaluated in 24 hour pieces, starting at 12 UTC and continuing through 11 UTC the next day. For each MCS, both modeled and observed, the convex hull of the MCS was taken. This yielded an MCS swath for each MCS. Only MCSs that occurred over the contiguous United States were considered. For each modeled MCS, the distance and direction from the centroid of the swath to the centroid of the nearest observed MCS swath was recorded. The minimum distance from the edge of each modeled MCS to the edge of the nearest observed MCS was also calculated. For a modeled MCS that overlapped with an observed MCS, the minimum edge distance was counted as 0km. For both the centroid distance and minimum edge distance, hits will be counted by varying the distance threshold. Misses are determined by counting the total number of observed MCSs on a given day and subtracting the number of hits. False alarms are similarly countedas the total number of modeled MCSs on a given day minus the number of hits.

For each day, the total area of modeled MCSs and observed MCSs and the area of overlap between the two was recorded. The overlapping area is counted as a hit, while the difference between the modeled area and the hit area is counted as a false alarm. Miss area is calculated as the difference between the observed area and the hit area. The storm motion of each MCS was estimated by taking the bearing of a line

between the centroid of the MCS at the first time step it existed and the centroid of the MCS at the last time step it existed. For MCSs that persist both before and after 12 UTC, the storm direction was estimated using the full time period of the MCS. For each hour and for each day, the number of observed and modeled MCSs was counted. The differences in position, orientation, and frequency will be analyzed.

4.3 Results

4.3.1 Time Series Comparison

The number of observed MCSs is higher than in any of the models for all hours of the day (Figure 4.1). There is also a strong diurnal cycle in the observations that peaks at near 400 MCSs at around 01 UTC. Some models have a slight suggestion of a diurnal cycle, but none are as pronounced as in the observations. The NAM CONUS Nest using the optimized parameters and HRW ARW using the default parameters have the highest numbers of MCSs, but both have much fewer than observations. Most of the models also appear to have a downward trend with time, which suggests that the models produce more MCSs at the start of the run and then fewer towards the end. This is likely driven by the model initiating fewer MCSs later in the period. It could also be driven by early MCSs growing upscale and merging; however, that seems less likely since there is no peak in modeled MCSs corresponding to the peak in observed MCSs.

When using the default MCS identification parameters, the model counts are clustered together, when using the optimized parameters, there is much more difference between the models. This suggests that much of the difference between the models

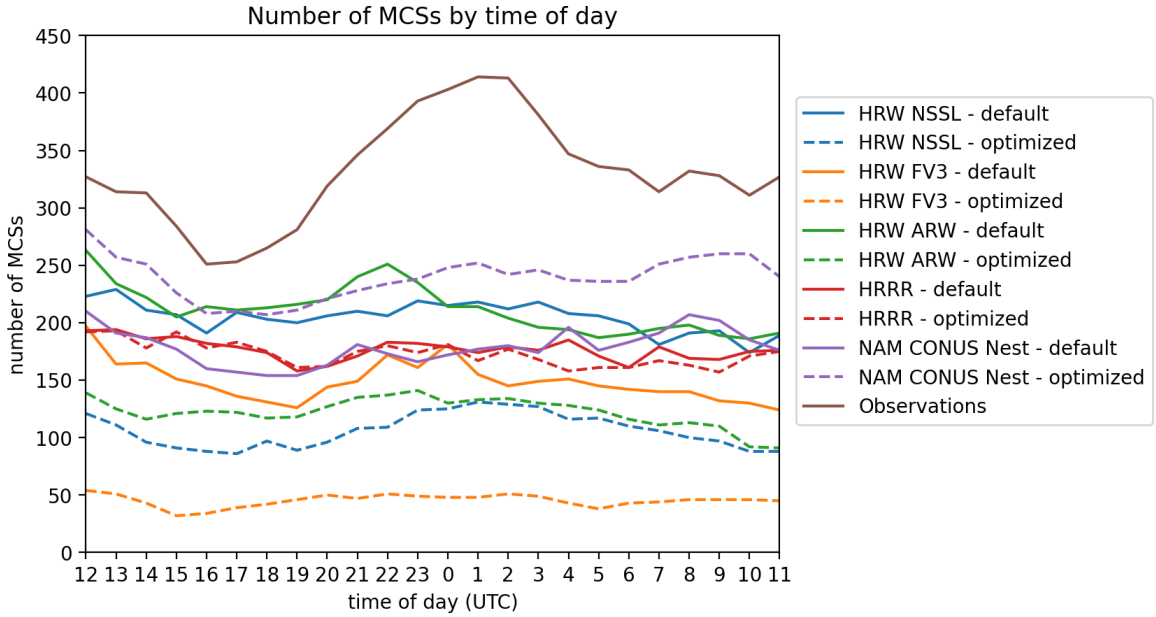


Figure 4.1: The total number of MCSs that occur during 2022 by time of day.

when using the optimized parameters is driven by the optimization, rather than by differences between the models themselves.

Comparing the annual cycle of MCSs identified in observations and the models also shows that the models struggle to match both the frequency and shape of the observed distribution (Figure 4.2). Between October and May, the modeled MCS frequency roughly matched the observed. Between May and October, the number of observed MCSs greatly increases, while the models mostly have more modest increases in the warm season. While all models have an increase in the warm season, the models all have a decrease in the number of MCSs in late July and August when the number of observed MCSs is near the peak. When using the optimized parameters, the HRW FV3 has an absolute minimum in August, corresponding with a peak in the observed MCSs.

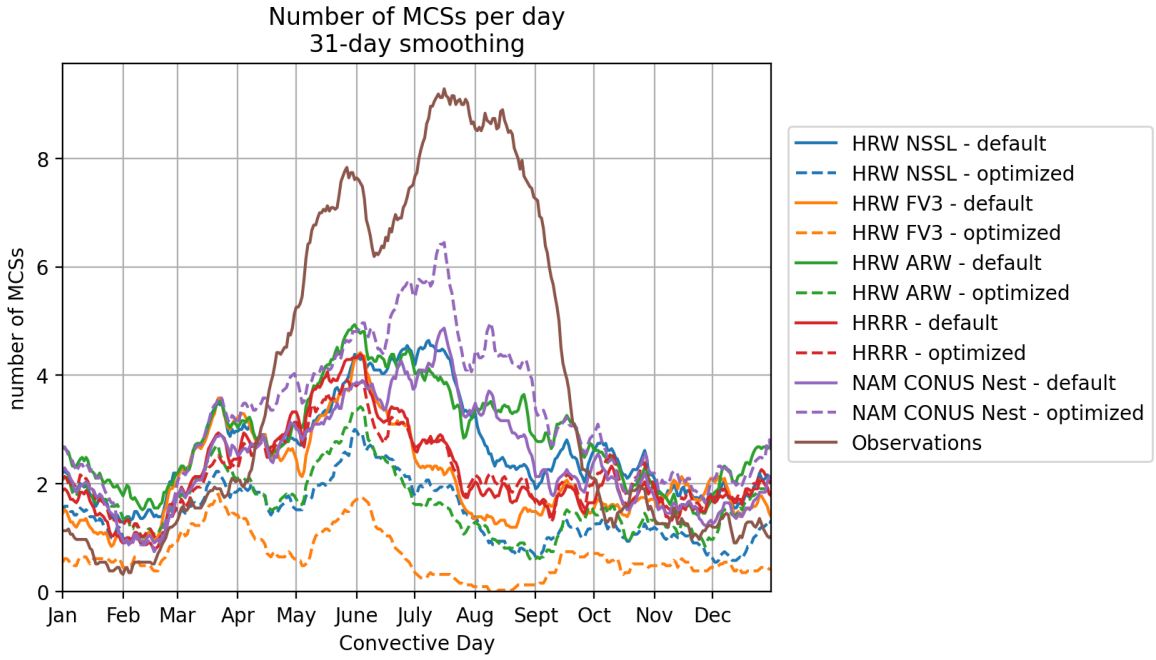


Figure 4.2: Average number of MCSs per day, using a 31-day smoothing window.

As in the time of day time series, the models cluster together when using the default MCS identification parameters and are more spread out when using the optimized parameters.

4.3.2 Verification

When verifying modeled MCSs by area, the underforecast seen in the time series is evident (Figure 4.3). The SR of all of the forecasts is at or above 0.8, but the highest probability of detection is near 0.3. This suggests that, while the models do not produce enough MCSs, the MCSs produced by the model are rarely false alarms. When using the default MCS identification parameters, the models tend to cluster together, matching what was seen in the time series. There is much more spread in the verification when using the optimized parameters. The HRW FV3, using the optimized parameters to identify MCSs, had the lowest POD and CSI, but had the highest SR.

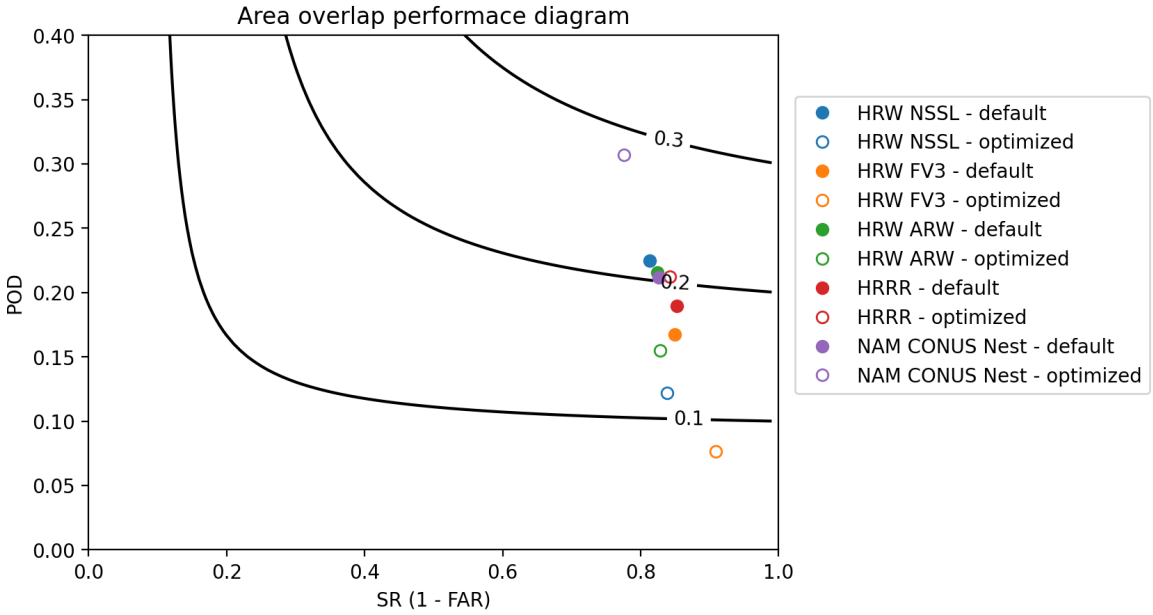


Figure 4.3: Performance diagram using area of overlap to count as hits. The y-axis is probability of detection and the x-axis is success ratio, which is equivalent to 1 minus the false alarm ratio. Note that the y-axis only goes from 0 to 0.4. The curved black lines are contours of critical success index (CSI).

When using the optimized parameters, the HRW FV3, HRW ARW, and HRW NSSL had lower CSIs than when using the default parameters. By far the highest CSI was achieved by the NAM CONUS Nest using the optimized parameters, but it had the lowest SR.

When using area overlap to verify, there is the potential for a double penalty. That is, a slight displacement error resulting in no overlap penalizes the forecast doubly: the forecast area with no overlap is counted as false alarm area and the observed area with no overlap is counted as miss area. This also means that area overlap is a very strict verification method. To help combat this, verification can be done using object-based principles to give credit to "close" forecasts.

One way to combat the double penalty is to use proximity between forecast and observed objects as the way to determine hits (i.e., accurate forecasts). Figure 4.4 shows

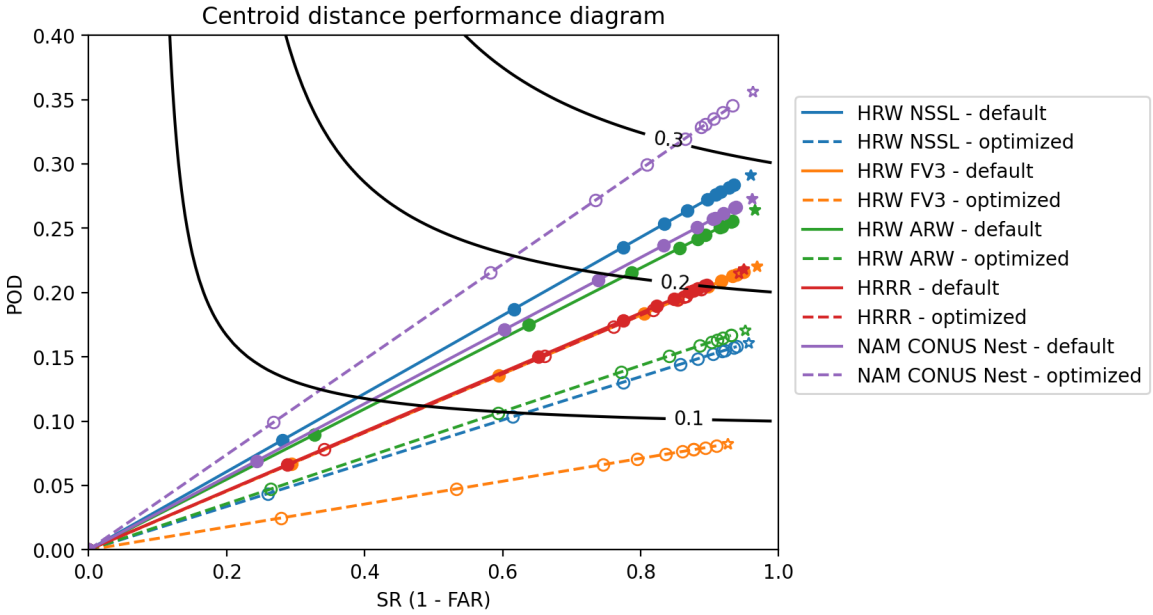


Figure 4.4: Performance diagram using centroid distance. The threshold distance to count as a hit varied from 0 km to 1000 km. Results are plotted every 1km to produce a curve. The dots are plotted every 100 km. The hollow dots are associated with the dashed lines and the optimized parameters while the solid dots are associated with the solid lines and the default parameters. The stars represent the limit where every forecast object on a day with observed objects is counted as a hit (in other words, if the distance threshold went towards infinity). Again note that the y-axis only goes from 0 to 0.4.

the results of varying the distance between forecast and observed centroids required to count as a hit. When a distance of 0 km is required (i.e. the forecast centroid must be in exactly the same place as the observed centroid), all of the models are in the lower left of the performance diagram, indicating that all forecasts were false alarms and all observations were missed. As the threshold increases up to 1000 km, POD and SR both increase. In general, the scores rapidly increase for the first 300 km before slowing. This suggests that most of the forecast MCSs are within 300 km of the nearest observed object. An underforecast bias is again evident, as seen the area overlap performance diagram and the time series. Each forecast also falls on a line of constant bias because varying the threshold required to count as a hit does not

change the overall number of forecast objects. Once again, MCSs identified using the default parameters are clustered together, while there is more spread when using the optimized parameters.

The HRW FV3 using the optimized parameters performed the worst, with the maximum POD never rising above 0.1, though the SR was similar to the other models. When using the default parameters, the HRW FV3 performed similarly to the HRRR using either set of parameters. The NAM CONUS Nest was the only model in which the optimized parameters outperformed the default parameters. All of the models are in the high SR, but low POD space, which again indicates that while the models do not produce enough MCSs, those that are produced usually correspond to an observed MCS. The clustering of the models when using the default MCS identification parameters indicates that a lot of the differences when using the optimized parameters are driven by the parameters, rather than differences inherent in the models.

The minimum distance from the edge of each modeled MCS to the nearest observed MCS is an even less strict method that can be used to evaluate the forecast performance (Figure 4.5). For this analysis, a distance of 0 km indicates that the forecast MCS overlaps with the nearest observed MCS, so the left end of each curve represents when hits are counted only if the modeled MCS overlaps with the observed. Increasing the distance threshold increases the SR and POD, with the curve following the same line of constant bias as in Figure 4.4. The small changes with increasing threshold suggest that many of the modeled MCSs overlap with an observed MCS. Once again, the HRW FV3 using the optimized parameters has the lowest CSI and POD, while the NAM CONUS Nest using the optimized parameters has the highest. When using the default parameters, the results for all of the models cluster together, whereas there is more spread when using the optimized MCS parameters.

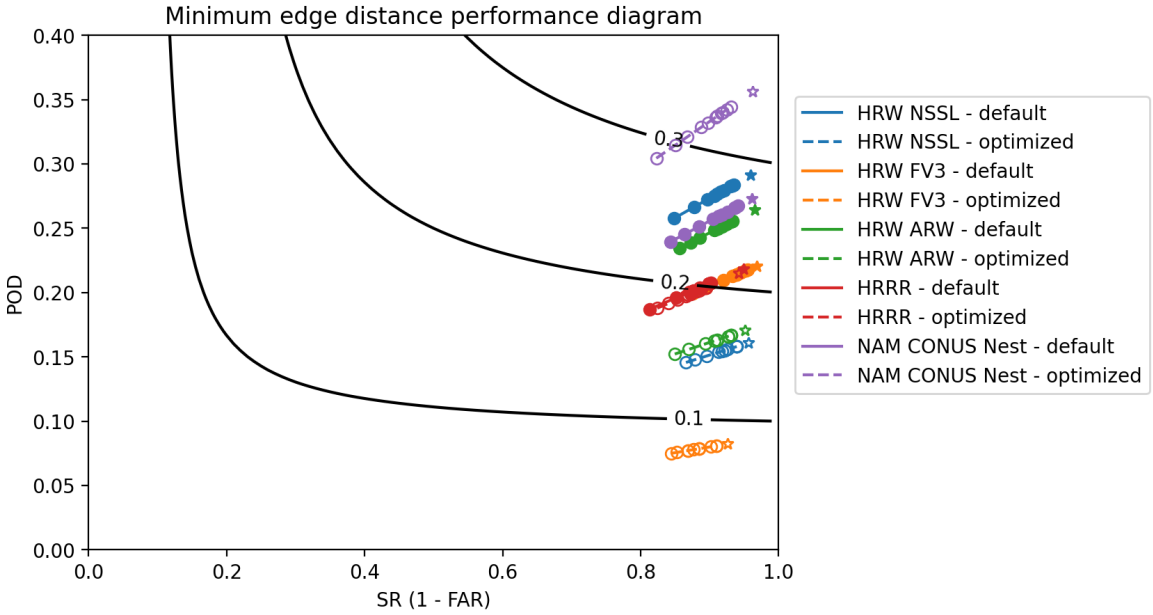


Figure 4.5: As in Figure 4.4 but for minimum edge distance between MCSs.

4.3.3 Error evaluation

Comparing the storm-relative locations of the modeled MCSs to the nearest observed MCS can give information about the errors in the model (Figures 4.6 and 4.7). In this analysis, a marker on the positive x-axis indicates that the nearest observed MCS was ahead of the modeled MCS. Similarly a marker on the negative x-axis would indicate that the nearest observed MCS was directly behind the modeled MCS, and a marker on the positive or negative y-axis would indicate that the nearest observed MCS was directly to the left or right, respectively, of the modeled MCS. For modeled and observed MCSs that overlap, there do not appear to be any trends in the errors. In other words, if the modeled and observed MCSs overlapped, the centroids were close together and the displacement does not have a preferred direction. This holds true when using both the optimized and default parameters to identify modeled MCSs. This indicates that

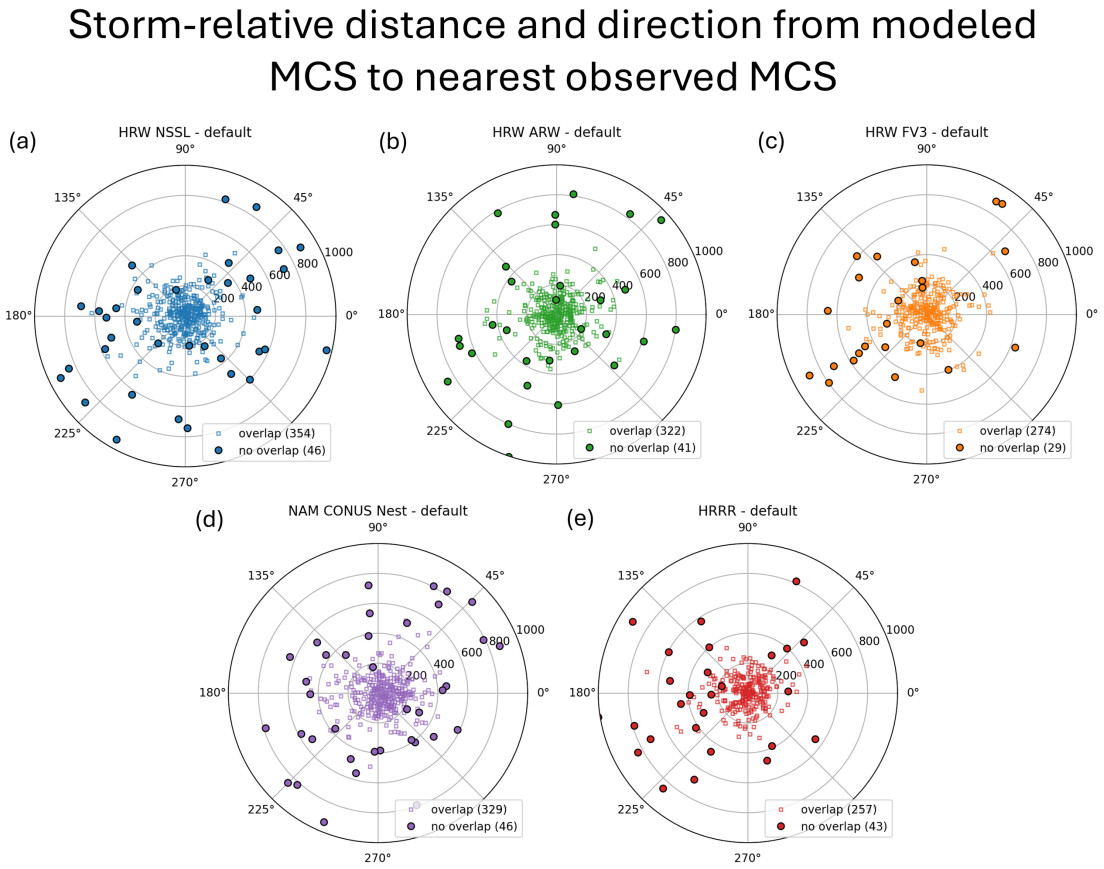


Figure 4.6: The storm-relative error vector from the centroid of each modeled MCS to the nearest observed MCS centroid using the default MCS identification parameters. The radius is in kilometers. The modeled storm direction is along the positive x-axis. The hollow square markers indicate when the modeled MCS overlapped with the observed MCS, and the filled dots indicate when there was no overlap. The models are (a) the HRW NSSL, (b) the HRW ARW, (c) the HRW FV3, (d) the NAM CONUS Nest, and (e) the HRRR. The numbers in the legend indicate the number of forecast MCSs in each category.

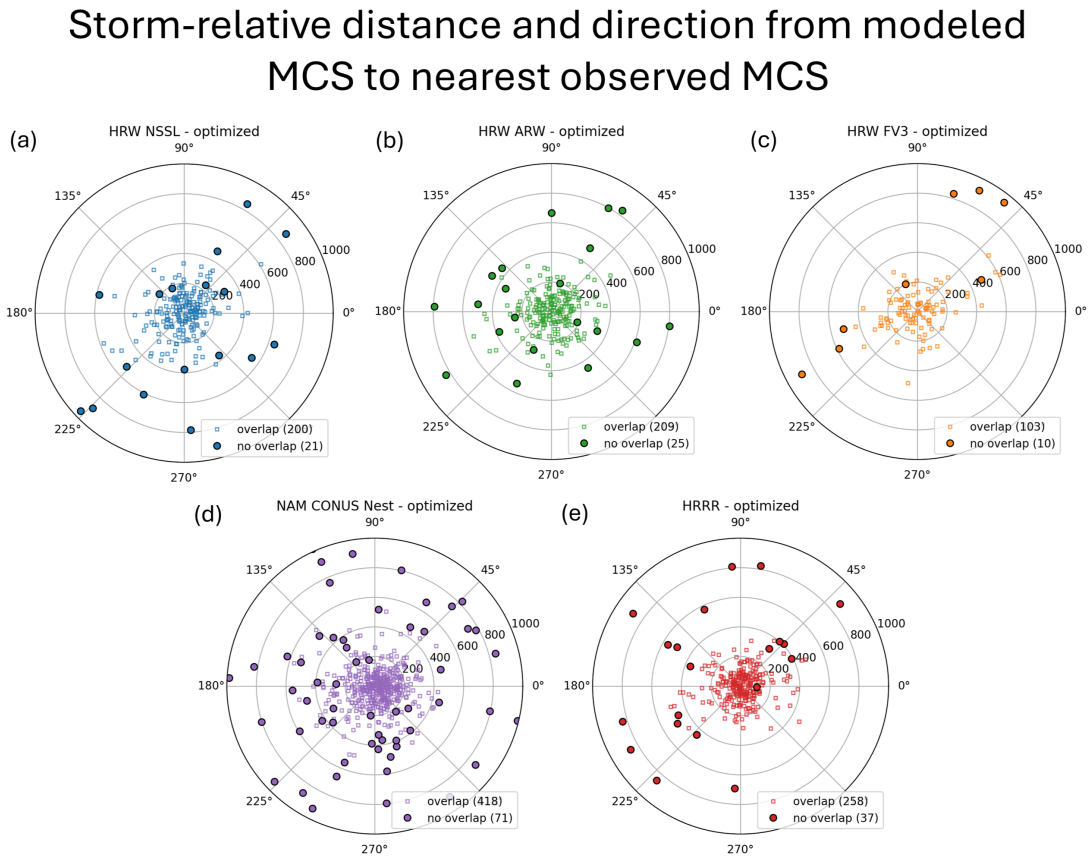


Figure 4.7: As in Figure 4.6 but for the optimized MCS identification parameters.

when the forecast and observed MCSs overlap, the forecast MCSs are very close to those observed.

In general, when using the default parameters to identify MCSs, the models produce similar numbers of MCSs (Figure 4.6). The HRW FV3 and HRRR both produce around 300 MCSs in total, while the remaining three models produce nearly 400 MCSs. The HRW FV3 produces the fewest amount of non-overlapping MCSs, though it also produces fewer objects overall than the other models, except for the HRRR. The HRW NSSL produces the most, most overlapping and most non-overlapping MCSs overall. Going back to the performance diagrams (Figures 4.4 and 4.5), the HRW NSSL had the highest POD when using the default parameters. The other models all follow the trend of fewer MCSs producing lower POD, even though the SR is similar for all models. This suggests that producing more MCSs could improve the models' POD and CSI.

When using the optimized parameters to identify MCSs, the models produce greatly varying numbers of MCSs (Figure 4.7). The HRW FV3 produces by far the fewest number of MCSs: 113, of which 91% overlap with an observed MCS. The NAM CONUS Nest produces 489 MCSs in total, by far the most out of all of the models. Of those, 85% overlap with an observed MCS. Both the number of overlapping and non-overlapping MCSs is also the highest out of all of the models. The HRRR produces the next highest number of MCSs, with 87% of model MCSs overlapping with an observed MCS out of a total of 295 model MCSs. The HRW NSSL and HRW ARW produce similar numbers of MCSs. The HRW NSSL produces 221 with 90% overlapping and the HRW ARW produces 234 with 89% overlapping. As with the default parameters, when looking back at the performance diagrams (Figures 4.4 and 4.5), producing more objects is associated with higher POD and CSI without much change in SR.

For instances where the modeled and observed MCSs do not overlap, there appear to be some trends in the errors. The HRW FV3 when using the default MCS identification parameters has most of the non-overlapping observed objects behind and slightly to the right of the modeled object (Figure 4.6). The observed objects being behind the modeled indicates potential timing issues, either with the timing of initiation or the speed of the initiation. The HRRR also has many of the non-overlapping observed objects behind the modeled MCSs. This again indicates a potential timing error. The HRW NSSL has very few non-overlapping observed MCSs behind and to the left of the modeled MCSs. This indicates that when the model produces an MCS that does not overlap with an observed MCS, the observed MCS is rarely to the left and behind the modeled MCS. When the observed MCS is ahead of the modeled MCS, it can be either to the left or the right. The HRW ARW has slightly more errors to the right of the modeled storm direction than the left, indicating that missed observed MCSs are more likely to be to the right of the modeled MCSs. The errors to the left or right are likely positional errors, where the model places the storm in the wrong place, due to errors in the environmental fields.

There are also some trends in the non-overlapping errors when using the optimized parameters (Figure 4.7). The HRW FV3 has very few non-overlapping MCSs because it has very few MCSs in general. Of the non-overlapping MCSs, the errors tend to be behind and slightly to the right of the modeled MCSs or ahead and to the left of the modeled MCSs. The non-overlapping errors are also mostly far away (400 km+), which indicates that these are potentially false alarms rather than close by misses. The forward and backward errors indicate potential timing issues, and the left and right errors indicate potential displacement errors. The non-overlapping MCSs produced by the HRRR are mostly behind and to the left of the 45° and 135° line, which potentially indicates both timing and displacement issues. Many of the non-overlapping HRRR

MCSs are also quite far from the nearest observed MCS. The NAM CONUS Nest has many more non-overlapping MCSs than the other models, but it produces more MCSs than any other model. There appear to be slightly more non-overlapping MCSs to the right of the modeled MCSs. In the HRW ARW, the modeled MCSs that do not overlap with the nearest observed MCS tend to be closer to the observed MCS. This potentially indicates more close by misses and fewer outright false alarms. The HRW NSSL appears to once again have more errors in non-overlapping observed MCSs being ahead and to the right of the modeled MCS. This matches what was seen when using the default MCS identification parameters.

All models have a small average difference in storm direction between the models and observations, using either the default (Figure 4.8) or optimized (Figure 4.9) MCS identification parameters. When using the default MCS identification parameters (Figure 4.8), the NAM CONUS Nest has the greatest mean difference between the modeled and observed MCS direction. While some of this difference is driven by the non-overlapping MCSs, even in the overlapping MCSs, there is a slight bias towards a negative difference. The mean error indicates that observed MCSs have a slightly more clockwise orientation compared to the modeled MCSs. The HRW FV3 has the next smallest average angle difference. The orientation error in the non-overlapping MCSs appears to be more positive than the overall mean error. In other words, the observed non-overlapping MCSs have a more counterclockwise orientation compared to the model MCS than the overlapping observed MCSs. The HRW ARW and HRW NSSL have the next smallest mean difference angle. The HRW ARW has no apparent trend in the errors from the non-overlapping MCSs; there are about as many non-overlapping MCSs above and below the mean. The HRW NSSL follows a similar pattern as the HRW FV3 in that observed non-overlapping MCSs are oriented more counterclockwise

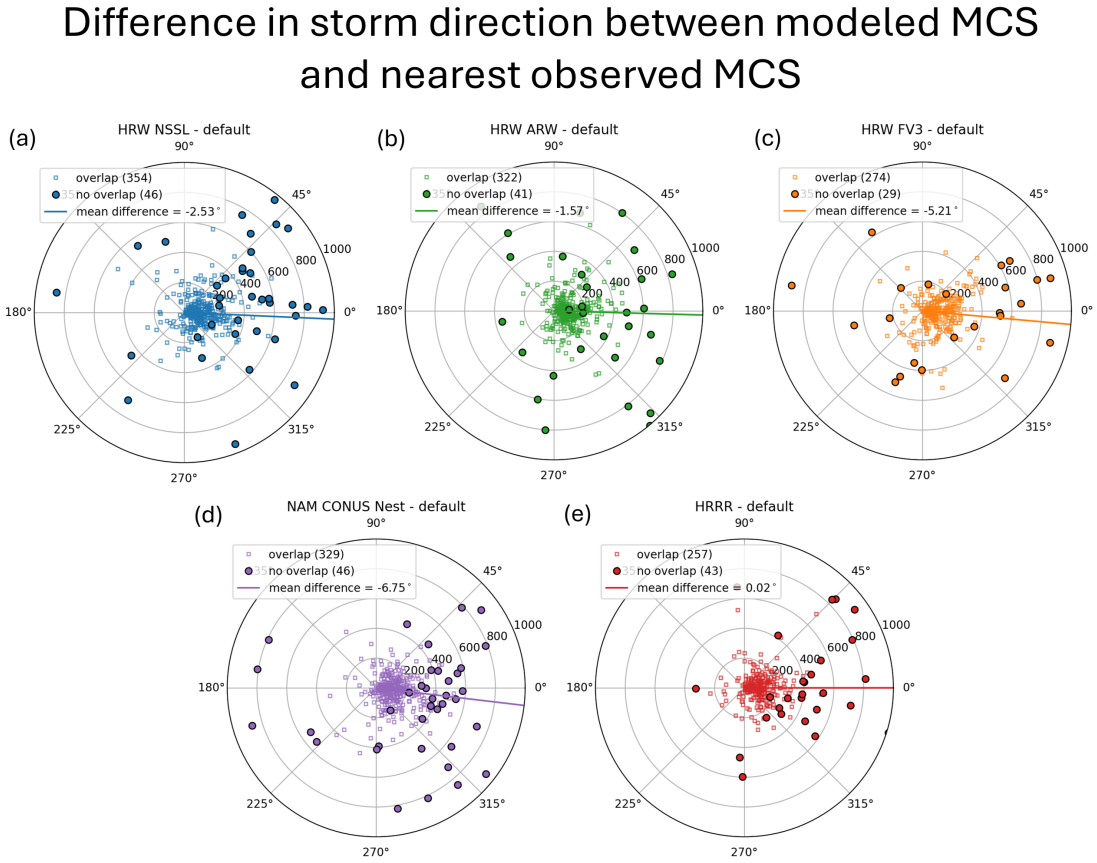


Figure 4.8: The difference in storm direction between the modeled MCS and the nearest observed MCS. The mean angle difference is plotted. The radius is the distance between the modeled MCS centroid and the nearest observed MCS centroid. The modeled storm direction is along the positive x-axis. The hollow square markers indicate when the modeled MCS overlapped with the observed MCS, and the filled dots indicate when there was no overlap. The models are (a) the HRW NSSL, (b) the HRW ARW, (c) the HRW FV3, (d) the NAM CONUS Nest, and (e) the HRRR. The numbers in the legend indicate the number of forecast MCSs in each category.

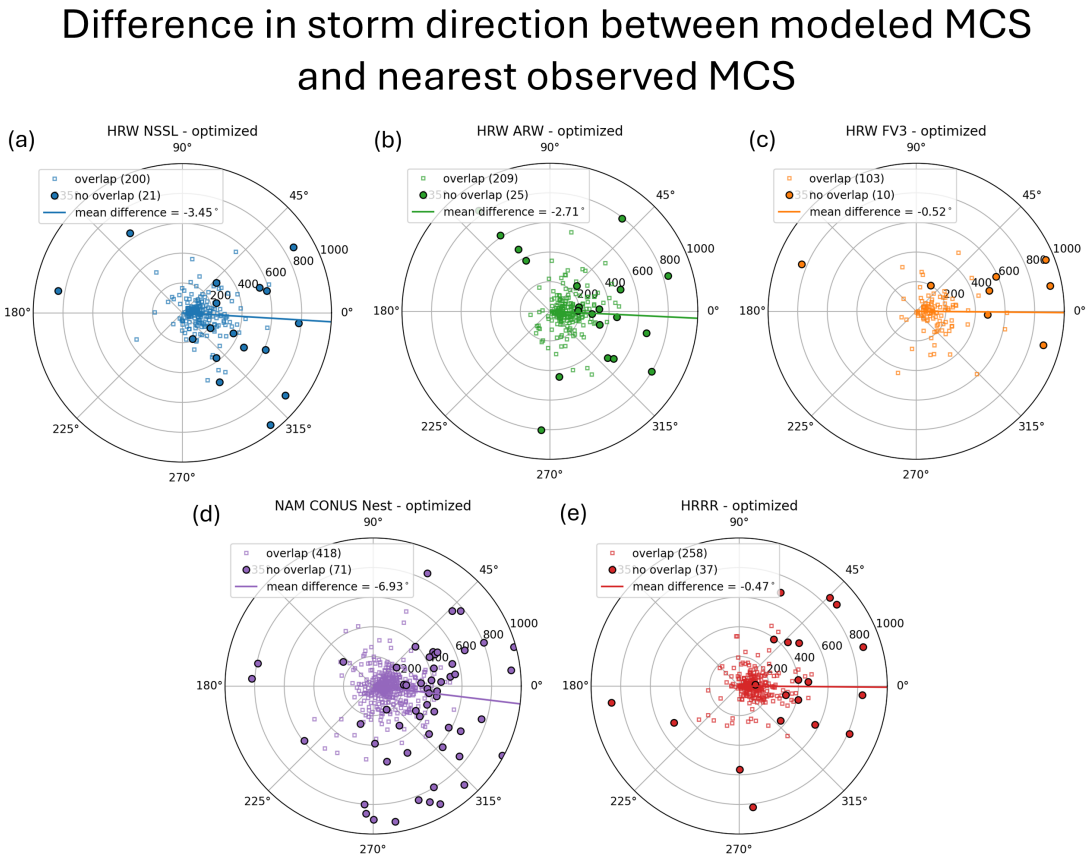


Figure 4.9: As in Figure 4.8 but for the optimized MCS identification parameters.

than the corresponding modeled MCS. The HRRR has the smallest mean orientation angle difference. The errors in the non-overlapping observed MCSs appear to be more clockwise than the mean error, and the errors in the overlapping observed MCSs appear to be slightly more counterclockwise.

While the broad trend of small mean orientation angle difference continues when using the optimized MCS identification parameters, there are some important differences in the details (Figure 4.9). The NAM CONUS Nest still has the biggest mean orientation angle difference between the modeled and observed MCSs. The non-overlapping observed MCSs appear to be oriented in a much more counterclockwise direction than when using the default MCS identification parameters. The HRW NSSL and HRW ARW again have similar mean orientation angle differences, and both appear to have a similar distribution of non-overlapping observed MCSs with orientations both above and below the mean orientation angle difference. The HRW FV3 has the next smallest mean orientation angle difference. As seen previously, there are very few modeled MCSs overall. Of the MCSs that are identified using the optimized parameters, the average difference in orientation angle with the nearest observed MCS is very small. This suggests that the optimized parameters identify MCSs that more closely match the observed.

4.3.4 Cases

4.3.4.1 14 June 2022

On 14 June 2022, there were a number of observed MCSs, most of which were associated with severe weather reports (reports shown in Figure 4.10, observed MCSs shown in Figure 4.11(f)). The largest number of severe reports were primarily severe wind

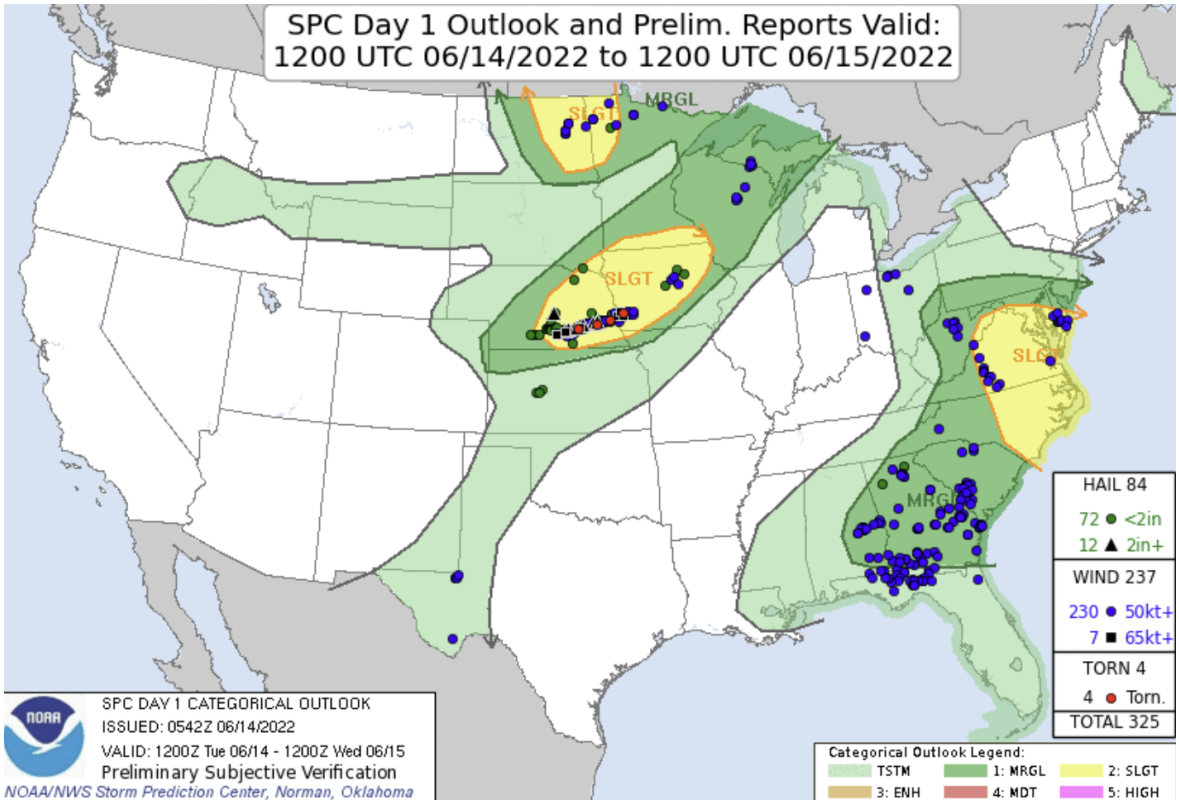


Figure 4.10: SPC Day 1 outlook and storm reports for 14 June 2022.

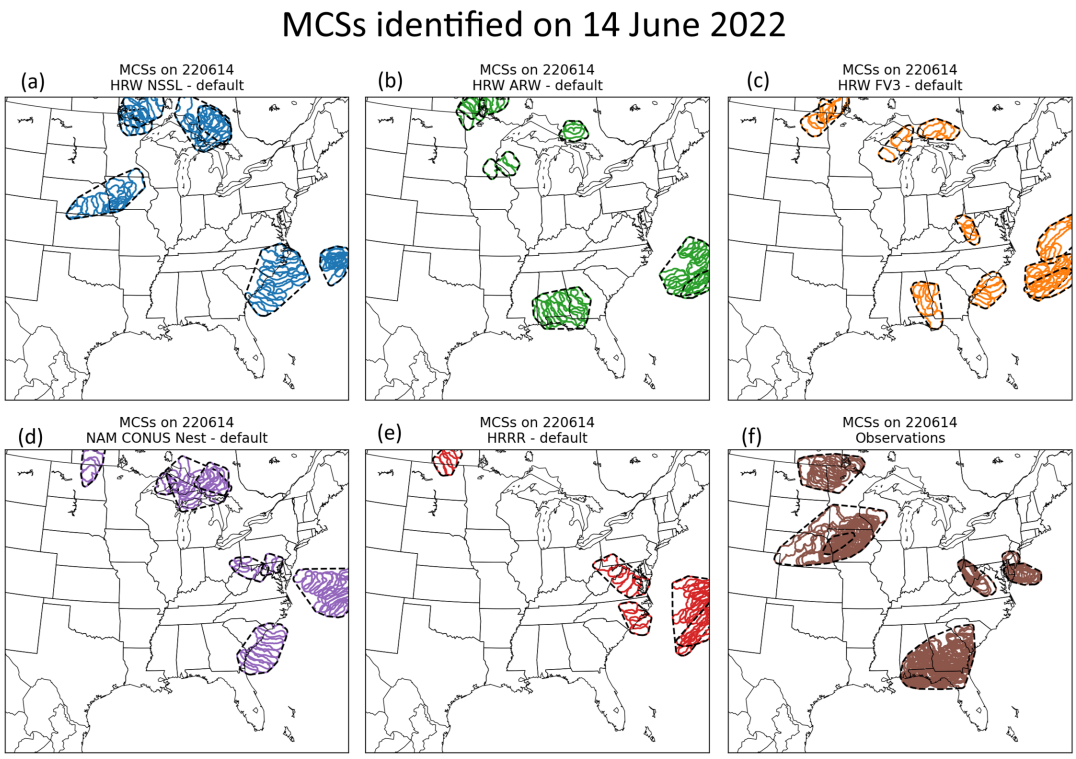


Figure 4.11: Modeled and observed MCSs for 14 June 2022. The solid colored lines are the individual MCSs and the black dashed lines are the convex hulls of each full MCS. The modeled MCSs were identified using the default parameters. (a) HRW NSSL, (b) HRW ARW, (c) HRW FV3, (d) NAM CONUS Nest, (e) HRRR, (f) Observations.

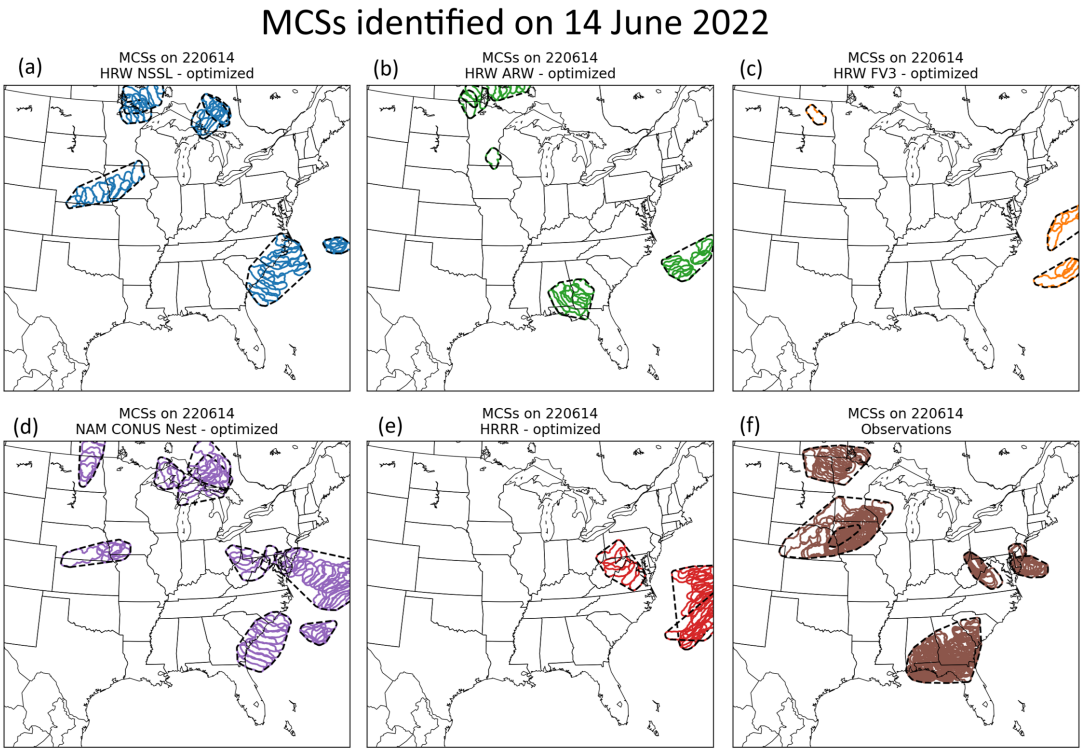


Figure 4.12: As in Figure 4.11, but modeled MCSs were identified using the MCS identification parameters optimized to match expert labeling.

reports associated with the observed MCS in South Carolina, Georgia, and Alabama. The observed MCS in eastern Nebraska and Iowa was also associated with a number of severe weather reports, include four tornado reports. The other MCSs identified in observations were associated with fewer severe weather reports. While in general the models fail to produce enough MCSs throughout the entire warm season, June is roughly when the models produce the most number of storms. For 14 June in particular, several MCSs were identified in each of the models (Figures 4.11 and 4.12).

When using the default MCS identification parameters (Figure 4.11), the HRRR was the only model to not identify any MCS corresponding to the observed one over Alabama, Georgia, and South Carolina. The NAM CONUS Nest and HRW NSSL both suggested that the MCS in the southeastern US would be more offshore and further north than observed. This is potentially a timing issue, if those two models did not initiate the storms until later in the forecast period, the forcing for that initiation could have moved further east, impacting the placement of the modeled storms. The HRW FV3 and HRW ARW both had MCSs identified in Georgia and Alabama, but those MCSs did not have the same areal extent as the observed MCS. The HRW NSSL was the only model to identify an MCS in eastern Nebraska and Iowa. The HRW ARW did identify an MCS in southeastern Minnesota and western Wisconsin, which did overlap with the observed MCS. The displacement error in this case is also likely a timing error, with the model not initiating or organizing the storms until later than what was observed. The other models had storms in that area, but they failed to organize enough to count as an MCS. All of the models identified MCSs near the Canadian border in far northern Minnesota and North Dakota that corresponded to an observed MCS in the same area. The HRW FV3, HRRR, and NAM CONUS Nest were the only models to identify MCSs in the mid-Atlantic that corresponded with two observed MCSs in the region. There were additional MCSs identified in the models in northern Ontario and

in or near the Gulf Stream. Those should not be treated as false alarms, however, due to the lack of radar coverage in those areas, which prevents MCSs from being observed there.

When using the MCS identification parameters optimized to match expert labels (Figure 4.12), all models except the HRW NSSL and NAM CONUS Nest identified fewer MCSs than when using the default MCS identification parameters. The HRW FV3 has the greatest decrease in number of MCSs when using the optimized MCS identification parameters, which matches what has been seen throughout this chapter. The NAM CONUS Nest now identifies an MCS in eastern Nebraska and western Iowa, which shows that there were storms in that region, they were just not organized enough to be identified as an MCS when using the default identification parameters.

4.3.4.2 21 August 2022

On 21 August 2022, there were many MCSs (Figure 4.14(f)) but few severe weather reports (Figure 4.13). The only MCS associated with multiple severe weather reports was one observed in central and southern Ohio. There were also a number of MCSs extending from north Texas through Alabama and the Florida panhandle. Many MCSs were also identified along the east coast from Georgia through upstate New York. Besides the MCS in Ohio, there were only a few severe weather reports associated with the other MCSs. August is when the number of observed MCSs reaches a maximum, but most of the models have decreasing numbers of MCSs through the month.

When identifying MCSs using the default identification parameters (Figure 4.14), MCSs are identified in north Texas in all of the models. For all models except the NAM CONUS Nest, no other MCSs are identified. The NAM identifies an MCS in eastern South Carolina and another in central Pennsylvania, which both overlap with

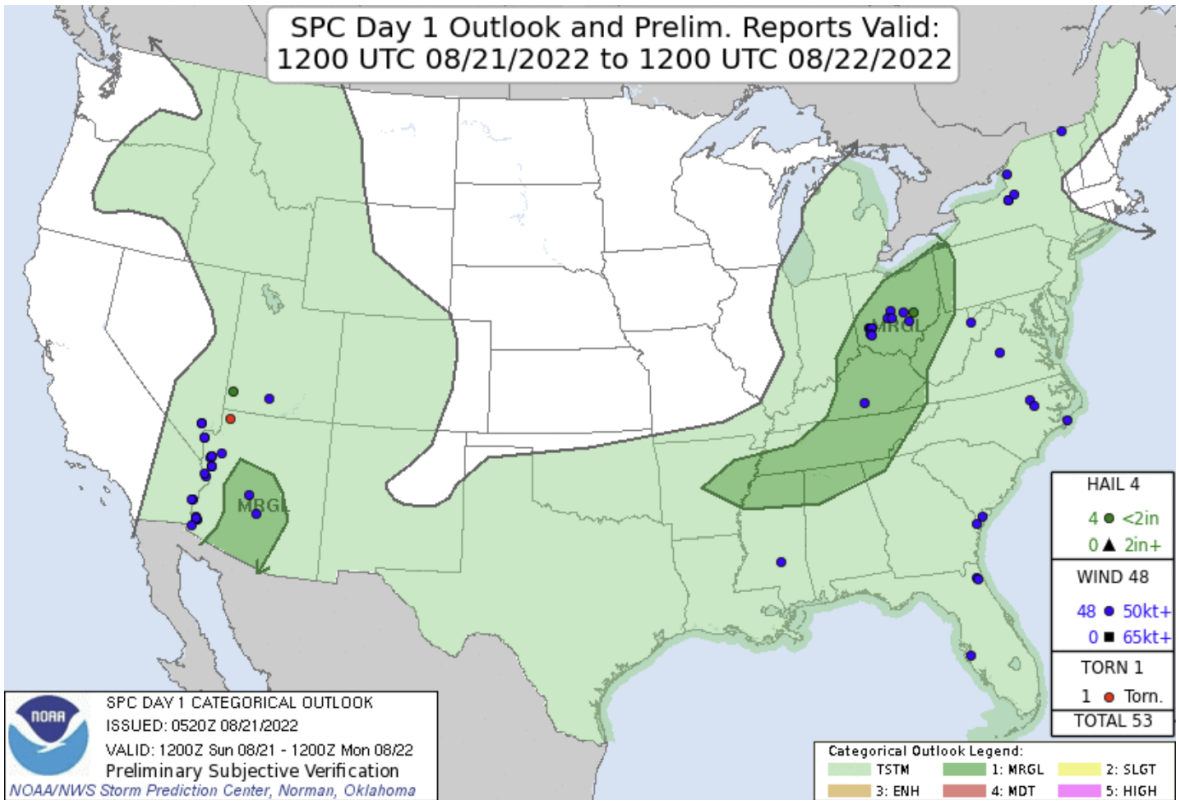


Figure 4.13: SPC Day 1 outlook and storm reports for 21 August 2022.

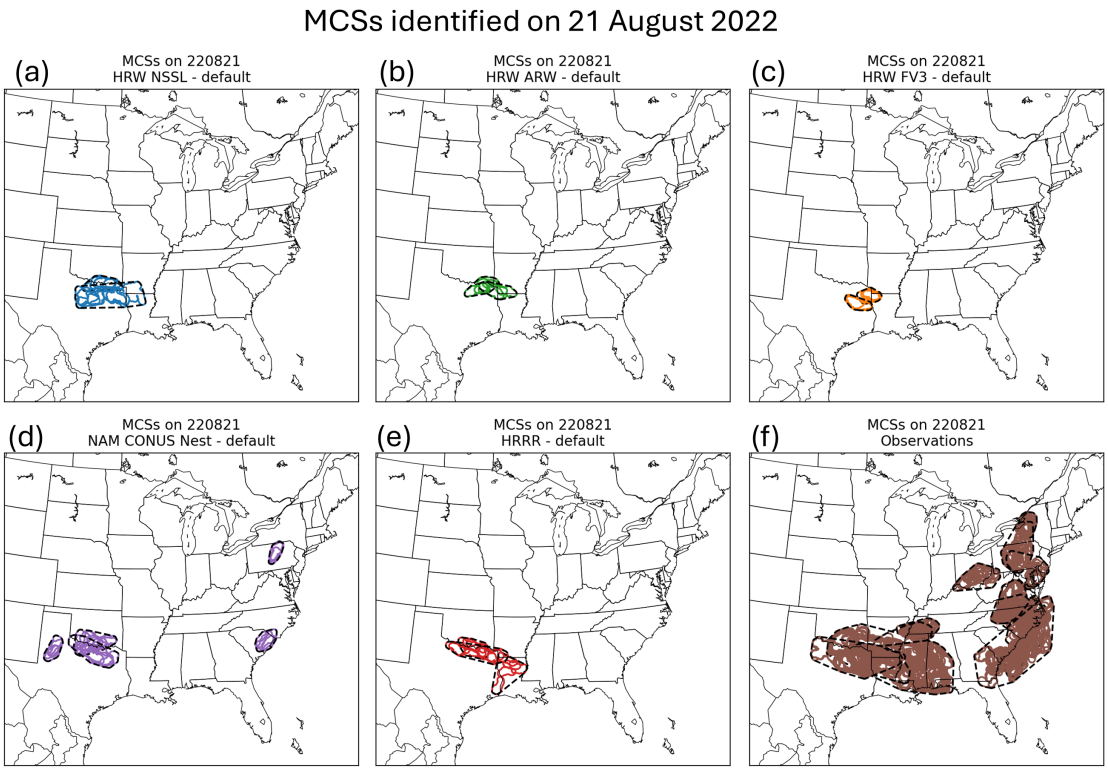


Figure 4.14: As in Figure 4.11 for 21 August 2022.

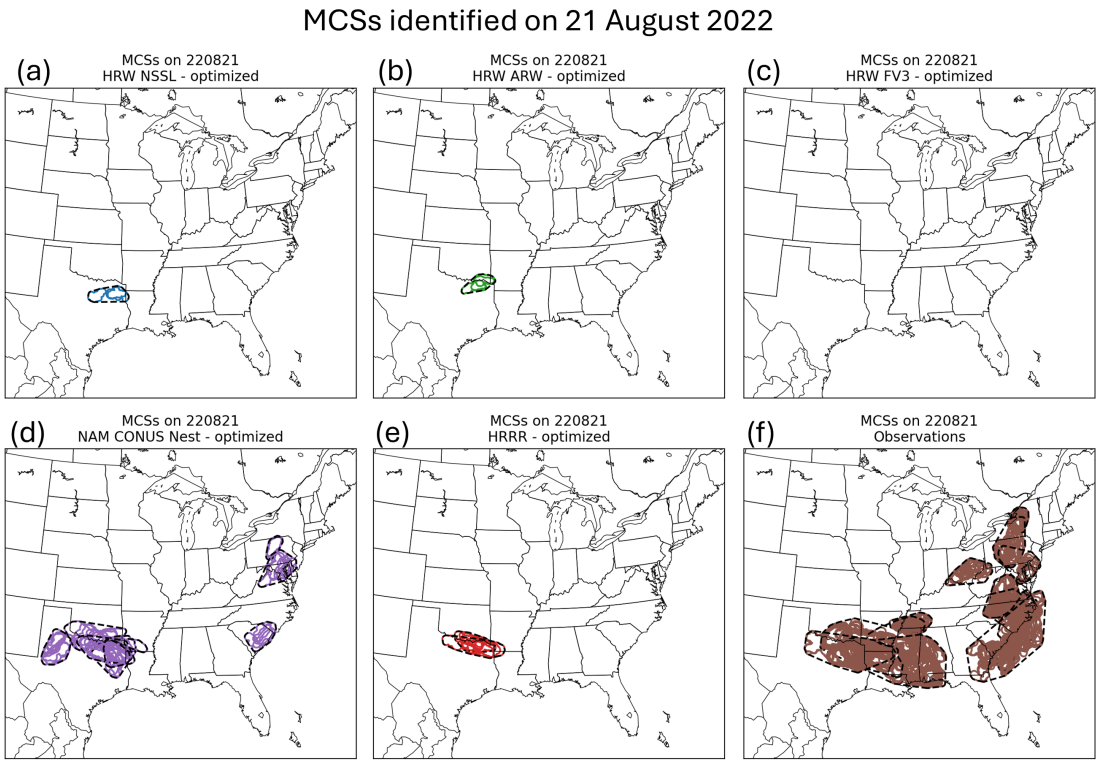


Figure 4.15: As in Figure 4.12 for 21 August 2022.

observed MCSs. For all of the models, all of the MCSs identified are far smaller than observed.

This pattern continues when identifying MCSs using the parameters optimized to match expert labeling (Figure 4.15). Once again, the NAM CONUS Nest is the only model to identify MCSs outside of north Texas. The HRW FV3 identifies no MCSs. The MCSs identified in the HRW NSSL, HRW ARW, and HRRR are smaller than when using the default parameters. The NAM CONUS Nest identifies a larger area of MCSs in north Texas which extends further west than the observed MCSs.

4.4 Discussion and Conclusion

The models generally produced fewer MCSs than were observed. The models also failed to produce MCSs with a diurnal or annual cycle that matched the observations. On the diurnal scale, there is a peak in the number of observed MCSs at 01 UTC, but the number of modeled MCSs has no corresponding peak. The HRW FV3 using the optimized parameters has the fewest number of MCSs. The NAM CONUS Nest using the optimized parameters has the greatest number, but it still has fewer MCSs than observed. On the annual time scale, the observed number of MCSs has an initial peak in May, with a larger sustained peak in July and August. The models do have more MCSs in the warm season, but they all fail to replicate the second late summer peak seen in the observations. Some models even have a local minimum when the observations have a maximum. When using the default MCS identification parameters, the models cluster together much more than when using the optimized parameters. This suggests that the parameters used to identify MCSs have an impact on the number of MCSs identified. Even with the parameters optimized to match expert labeling, the fact that the models struggled to match the observed diurnal and annual cycle suggests that

these model climatologies are fundamentally flawed. This fundamental error places obvious limits on model performance, both for forecasting the immediate weather and for longer time scale climate simulations.

Verifying the modeled MCSs also showed a strong underforecast bias. Using area-based verification showed that HRW FV3 using the optimized parameters had the lowest POD and CSI, but the highest SR. The NAM CONUS nest using the optimized parameters had the highest POD and CSI, but the lowest SR. The area-based method is vulnerable to a double penalty for small displacement errors, so an object-based method was used instead. Using centroid distance or minimum edge distance between the modeled MCSs and the nearest observed MCS, the threshold distance used to determine a hit was varied to produce a verification curve for each forecast. Using either verification method produced similar results. The HRW FV3 using the optimized parameters performed the worst (worst bias, lowest POD), while the NAM CONUS Nest using the optimized parameters performed the best. Using the default parameters produced a better forecast than using the optimized parameters for all models except the NAM CONUS Nest. Once again, using the default MCS identification parameters produces forecasts that are much more clustered than using the optimized parameters.

The underforecast seen in all models of the HREF is potentially caused by several different factors. The summertime underforecast is potentially driven by the models failure to reproduce the weakly forced environments seen in the summertime that often cause convective initiation. CAMs and CAM ensembles have been shown to struggle to forecast rainfall in weakly forced environments (Surcel et al., 2016; Nielsen and Schumacher, 2016; Herman and Schumacher, 2016). In addition, Surcel et al. (2016) notes that weakly forced environments are often driven by the diurnal cycle. The models of the HREF shown here struggle to reproduce the diurnal cycle of MCSs seen in the observations. A potential cause of the failure of the models to reproduce the diurnal

cycle is poor initialization. All of the models examined here were the 00 UTC model run with the forecasts analyzed from hour 12-35. The frequency of observed MCSs reaches a maximum between 01-02 UTC. If the models fail to reproduce the MCSs early in the forecast period, they may struggle to accurately represent the environment later in the day. Models have been shown to struggle with nocturnal MCSs (Peters et al., 2017), so those errors may be influencing forecasts later in the day.

Britt et al. (2024) and Flora et al. (2019) found that the WoFS was able to forecast MCSs over a short time period ($\leq$ 6 hours). This study, however, examined 12-36 hour model forecasts of MCSs. The difference in performance seen in those studies and this one may be driven at least partially by the models' declining performance with increasing lead time.

The tracking algorithm developed for this study is also much stricter than the observational tracking algorithm developed by Wade et al. (2023). The observational tracking algorithm allows the MCS to continue to be tracked even if the identification criteria are not met at an individual (15 minute) time step, as long as it exists both before and after. This more generous criteria is to limit the impacts of observed radar artifacts and ground clutter. The novel tracking algorithm based on hourly maximum reflectivity and does not allow an MCS to be tracked if the tracking criteria (reflectivity greater than 50 dBZ) is not met for the entire time period. Because there is no ground clutter or radar artifacts in the modeled reflectivity, there were no discontinuities considered. This more strict criteria, however, potentially means that some MCSs are not being counted because the hourly maximum reflectivity field has a discontinuity.

Exploring the storm-relative error characteristics showed that when the model forecast MCS overlapped with an observed MCS, the distance between the two objects' centroids was small and there was no systematic error direction. This holds true when

using either the default or optimized MCS identification parameters. When the modeled MCSs did not overlap with an observed MCS, there were some trends in the errors. Displacement errors ahead of or behind the modeled MCS are indicative of timing errors, while errors to the left or right are likely positional errors. The NSSL WRF had more errors to the right, while the HRW FV3 had more errors ahead or behind. The HRRR appeared to have more errors behind the modeled MCSs than in front, and when using the optimized parameters, errors to the left were more common.

The storm direction error also showed that when the modeled MCSs overlapped with an observed MCS, there was very little difference between the modeled and observed storm direction. The NAM CONUS Nest had the largest direction error when using either the default or optimized MCS identification parameters, but the mean error was less than 7°. All of the mean storm direction errors were negative, indicating that the observed MCS was oriented in a more clockwise direction than observed. There appeared to be a greater difference when the modeled MCS did not overlap with an observed MCS.

The error characteristics also further reinforced the differences in the number of MCSs forecast by each model. While the models identified roughly similar number of MCSs using the default parameters, there were much larger differences when using the optimized parameters. However, the vast majority of the model MCSs did overlap with an observed MCS. This indicates that while models do underforecast the number of MCSs, if a model produces an MCS, it usually overlaps with an observed MCS.

The two cases shown here demonstrate the low POD, high SR forecasts produced from models using this MCS identification method. The first case, 14 June 2022, showed that MCSs can be identified in the models and that many of those MCSs overlap with observed MCSs. Some of the models identified some of the observed MCSs however none of the models identified all of the observed MCSs. The MCSs

observed on this day were associated with a large number of severe weather reports. In the second case, 21 August 2022, there were far fewer severe weather reports associated with observed MCSs. The models also struggled to identify any MCSs outside of north Texas on this day. Together, these two days suggest that there may be a relationship between the strength or intensity of an observed MCS and the ability of the models to produce one, though more research would be required to draw any meaningful conclusion.

While a high POD, high SR forecast would be best, there is still useful information in a high SR, low POD forecast, especially if the forecast MCSs are similar to the observed MCSs. If the models produce an MCS using the method in Chapter 3, it is likely that there will be an observed MCS nearby or overlapping the modeled MCS. The optimized parameters, while they do identify MCSs in a way that best matches expert identification, do not seem to improve the forecast for most of the models. The NAM CONUS Nest was the only model to have a higher POD and higher CSI when using the optimized parameters. It was also the only model that produced more MCSs when using the optimized identification parameters. The other models had lower POD, lower CSI, and lower MCS counts when using the optimized parameters. The hypothesis, that the optimized parameters would improve the forecasts, was only supported for the NAM CONUS Nest. For the other models, the hypothesis was not supported. This suggests that the identification parameters could be optimized to best match the observed MCSs from the Wade algorithm instead of matching expert labeling. Alternatively, using the default parameters based on the Wade algorithm still produced reasonable forecasts, so perhaps optimization would not need to be done as models are updated and new models are introduced.

Chapter 5

Summary and Conclusions

MCSs are responsible for many types of impactful weather, from direct impacts such as severe weather (as defined by the NWS: hail larger than one inch, wind gusts greater than 50 kt, or tornadoes) and intense rainfall to downstream impacts on due to outflow boundaries and modifications of the 500 hPa pattern. The prevalence of MCSs is also projected to increase in a warming climate. Given the weather impacts of MCSs and their potential increasing prevalence, it is important to understand the forecast models' ability to forecast for MCSs.

To determine numerical models' ability to forecast MCSs, an automated method of identifying MCSs in model output was needed. As part of developing that method of identification, the shape characteristics of modeled and observed storms were explored. Storms were identified in each model of the HREF and in observations using a quantile-based convective radar reflectivity threshold for the period between 12 UTC on 1 January 2022 and 11 UTC on 1 January 2023.

When comparing modeled and observed storm objects, the distribution of objects in aspect ratio-area space was broadly similar. Objects tended to get less round with increasing size, though the trend is slightly less strong in observations than in the models. When comparing the maximum reflectivity within each object, the models all seemingly have an upper limit on reflectivity. The HRW NSSL, HRW ARW, HRW FV3, and NAM CONUS Nest have very few objects with reflectivity above about 60 dBZ, which suggests an upper limit in the models' reflectivity algorithm. The

HRRR apparently has a soft upper limit, where most HRRR objects have a maximum reflectivity below about 55 dBZ with few objects having a maximum reflectivity above that level. In observations the maximum reflectivity within each object was distributed without an apparent upper bound.

There were more differences between modeled and observed objects when comparing orientation angles. Modeled objects have a mean orientation angle roughly 50° counterclockwise above an east-west line, while observed objects have a mean orientation angle roughly 20° counterclockwise above an east-west line. For larger objects, both modeled and observed objects had higher (i.e. more counterclockwise) orientation angles for larger objects, though the difference between the forecast and observations persists even with larger objects. Modeled objects have more spread in orientation angle than observed objects.

The differences demonstrate some potential challenges to identifying MCSs in numerical models, though the similarities in the aspect ratio-area space indicate that modeled storms are at least shaped somewhat similarly to observed objects, even if their orientations are different. The differences in orientation angle have several potential explanations. The forecast models fail to accurately recreate observed storm structure. This could be due to the models' failure to represent back building storm cells at the southern end of a convective line (Hitchcock and Schumacher, 2020). For an observed storm with these back building cells, this would tilt the orientation closer to east-west as the cells extend to the west and southwest. A modeled storm that lacks these cells would be oriented closer to north-south. The models may also fail to accurately represent the stratiform region of a convective line or fail to reproduce a bow and arrow structure (Keene and Schumacher, 2013). This region would cause an observed storm to have a lower, rounder, aspect ratio and potentially an orientation angle closer to east-west. A modeled storm lacking a similar stratiform region would

have a less round aspect ratio and an orientation angle closer to north-south. A model failing to reproduce the stratiform region could also explain the differences in aspect ratio seen with the biggest objects.

The Wade algorithm, initially developed to identify observed MCSs, was modified to identify modeled MCSs. Initial modifications were done to account for model grid spacing and differences in reflectivity distributions. MCS identification was found to be sensitive to certain parameters, and those parameters were tuned using a set of storms labeled by experts as the truth. A new tracking algorithm was also developed that uses modeled hourly maximum reflectivity to track MCSs identified using the modified Wade algorithm across time.

For each model in the HREF, a set of parameters was found that yielded the best matches to the expert labeling. The best match was determined by generating a curve on a performance diagram and computing the area to the left of that curve. The curve was generated by varying the distance between the labeled storm and the nearest identified MCS required to count as a hit. This curve represented the overall performance of the algorithm using each set of parameters. The area to the left of this curve, which is equivalent to the area under a precision-recall curve, is used to assess the performance of rare event forecasts. The parameters that maximized the area to the left of the performance diagram were used as the optimized parameters.

The tracking algorithm and optimized parameters were used to analyze a full year of model output, from 12 UTC on 1 January 2022 through 11 UTC on 1 January 2023. With some aggressive smoothing, the spatial distribution of the modeled MCSs matched the shape of the observed MCSs from Wade et al. (2023), though the model counts were lower than observed. The HRW NSSL had the smallest MCSs; it also had the smallest cline_len, which is the final size filter applied in the Wade algorithm. It excludes MCSs with a longest dimension shorter than the cline_len parameter. This would

favor keeping smaller MCSs that the other models, with larger optimized cline_len parameters would discard. A KDE analysis of modeled MCS centroids showed that at least some of the variation in the spatial map was driven by the differences in MCS area between the modeled MCSs. Though the KDE analysis of each model was largely similar, there were some differences. The highest densities of the KDE plot of the NAM CONUS Nest and HRW ARW extended further west into eastern Oklahoma than any of the other models; the NAM CONUS Nest and the HRW NSSL extended further north up the Mississippi River valley than any of the other models; and the HRW FV3's distribution extended further east than the other four models.

The Wade algorithm using the optimized parameters and the novel tracking algorithm were subjectively able to identify MCSs and reproduce a spatial distribution that can reasonable be described as being drawn from the observed multi-year climatology. Since the spatial distribution appeared reasonable, the forecast MCSs were also verified and evaluated.

Verification and evaluation was done by creating convex hulls of both the forecast and observed MCSs for each 24-hour period, from 12UTC on one day through 11UTC on the next (forecast hours 12-35 of the 00 UTC model runs), from 12 UTC on 1 January 2022 through 11 UTC on 1 January 2023. Observed MCSs that existed both before and after 12 UTC were counted on both days, while modeled MCSs that existed before forecast hour 12 or after forecast hour 35 were only counted for the portions that existed between forecast hours 12 and 35.

Comparing the time series of modeled and observed MCSs showed an underforecast at both the diurnal and annual time scale. The model forecasts struggled to reproduce the diurnal cycle seen in observations. When using the default parameters from the Wade algorithm, all of the models have a similar number of MCSs, but when using the optimized parameters, there is considerably more variation. This also holds true

when looking at the annual distribution of modeled MCSs. Using the default MCS identification parameters results in the models clustering together, while the optimized parameters produces more variation. The models also struggle to reproduce the annual cycle seen in the observations. There are two peaks in the observed number of MCSs, one in May and a longer, higher peak in late summer. The models do generally have more MCSs in the warm season than the cool season, but many have a decrease in the late summer when the observed MCSs are at their highest. When using the optimized MCS identification parameters, the HRW FV3 reaches an absolute minimum at the same time that the observed MCS count reaches an absolute maximum. During the cool season, all of the models more closely match the observed number of MCSs.

Verification by area, in which the overlapping area counts as hits and the non-overlapped forecast and observed areas count as false alarms and misses, respectively, further reinforced the underforecast seen in the time series analysis. All models had high SR but low POD. The HRW FV3 using the optimized parameters had the lowest CSI and lowest POD but the highest SR. The NAM CONUS Nest using the optimized parameters had the highest CSI and POD but the lowest SR. To avoid potential double penalty in the area verification, modeled MCSs were also matched to the nearest observed MCS. The distance from the centroid of a modeled MCS to the centroid of the nearest observed MCS and the minimum edge distance from a modeled MCS to the nearest observed MCS were recorded. By varying the distance threshold to count as a hit, the forecasts could be verified. These results further show the low bias, with all forecasts in high SR, low POD space. Once again, the NAM CONUS Nest with the optimized parameters had the highest CSI and POD, while the HRW FV3 with the optimized parameters had the lowest. The NAM CONUS Nest was the only model for which the optimized MCS identification parameters improved the verification scores.

The minimum edge distance verification also demonstrated that many of the modeled MCSs overlapped with an observed MCS.

Exploring the error characteristics showed that, for all models, when a model produced an MCS, it was usually quite close to an observed MCS. When a forecast MCS overlapped with an observed MCS, the distance between the objects' centroids was low, and there was no systematic error direction. This was true for both the default and the optimized MCS identification parameters. For modeled MCSs that did not overlap with observed MCSs, there appeared to be some trends in the errors. Some models had more errors in front or behind the modeled storm, which indicates timing errors, while other models had errors to the left and the right, indicating displacement errors. The storm direction error again showed that, when a modeled MCS overlaps with an observed MCS, there is very little difference in storm direction.

The high SR, low POD forecast indicates that, for all models, when a model does forecast an MCS, it is usually close to or overlapping with an observed MCS. This kind of model information is extremely useful to forecasters because, while the absence of an MCS does not indicate that one will not happen, the presence of an MCS in the model indicates that one is likely to happen.

The optimized MCS identification parameters did not necessarily improve the model forecast of MCSs even though the parameters were optimized to match expert labeling. The performance of the MCS identification algorithm when using the default parameters, while having a low CSI, potentially indicates that optimization may not be necessary because of the high SR, low POD nature of the forecast. If the default MCS identification parameters are good enough, then there would be no need to re-optimize them for use as models are upgraded or as new models are introduced. Not

needing to re-optimize would save time when models are upgraded or changed. Forecasters would not need to label new cases or the new models would not need to run old cases.

There were several potential limitations of this work. When identifying storm objects in Chapter 2, the choice of reflectivity threshold could potentially have influenced the shape of objects. While 35 dBZ is often used to identify convective precipitation (Henderson et al., 2021), 40 dBZ is also often used (Roberts et al., 2020) to identify storms. The difference would have the most impact in the shape of the stratiform region of the storm, where a small overforecast of reflectivity by the model may have pushed reflectivity above 35 dBZ. Choosing 40 dBZ as a threshold would also introduce a similar effect around 40 dBZ. While the tuning of the MCS identification algorithm in Chapter 3 was intended to be as objective as possible, it still relied on the expert forecasters to label storms. The choice of score to maximize in the tuning was also a potential limitation. While AUC-PR was chosen to best represent the ability of the particular set of parameters to identify MCSs, critical success index (CSI) could also have been used. Some of the parameter sets maximize both AUC-PR and CSI, but some of the models would have had different tuned parameter sets. Finally, the two object-based verification methods shown (centroid and minimum edge distance) were perhaps overly strict and overly lenient, respectively. The centroid distance could heavily penalize a forecast if the MCS shapes were slightly different or if two nearby MCSs were merged in the forecast but not the observations (or vice versa). Conversely, minimum edge distance could not penalize a forecast enough if the forecast and observed MCSs overlap even if the two MCSs were otherwise extremely dissimilar. An interest score would potentially better represent when the forecast and observed MCSs truly match, but since both the centroid and minimum edge distance show the same high SR, low POD regime, using a more advance interest score may not change the overall

result very much. While using forecast hours 12 through 35 removed any potential impact of model spin up, potential MCS forecast inaccuracies in the first 12 hours could have driven later inaccuracies.

Though the severe weather impacts of MCSs were the driving motivation for this study, MCSs also have non-severe impacts and not all MCSs produce severe weather. The high SR, low POD forecasts shown here can still be useful for rainfall events associated with MCSs. The presence of an MCS in the models indicates that one will likely occur, but the absence of one does not provide information about whether one will happen. Similarly, the presence of an MCS early in the forecast period indicates that one is likely to happen. Any impacts on weather later in the period, such as outflow boundaries and convection initiated off of them, may also still occur. The models' failure to regularly predict sufficient numbers of MCSs has impacts further downstream. For any global CAMs, a poor forecast of MCSs in the central United States can negatively impact the 500 hPa forecast over Europe days later. Additionally, in a warming climate, a greater proportion of rainfall is projected to come from MCSs, so accurate MCS forecasts will become more important for forecasting rainfall.

Though the error evaluation graphics would also be useful in evaluating any future MCS detection algorithm, they could also be useful for evaluating any object-based forecast. This kind of error evaluation graphics visually depict motion-related errors. It can identify potential timing errors by showing when identified model objects are ahead of or behind the nearest observed object. Spatial errors can also be identified by showing when a modeled object is to the left or right of the nearest observed object. The object direction error can also demonstrate whether modeled objects are moving in the same direction as the nearest observed objects. Highlighting the non-overlapping objects can also separate which errors are coming from larger misses and which errors are coming from smaller displacement errors.

Besides simply showing where the models forecast an MCS, the automated MCS identification can also be used in post-processed guidance. Jahn et al. (2022) demonstrated a technique to use automated MCS identification to improve tornado forecasts. The MCS identification method used here could be applied using the technique from Jahn et al. (2022) to see if the radar-based MCS identification method can improved forecasts over the UH-based MCS identification method from Jahn et al. (2022). A similar technique could be developed to identify the wind or hail threat associated with MCSs. Since 44.9% of significant severe wind reports are associated with MCSs (Smith et al., 2012), a wind signal from the model may be more indicative of a wind threat if it is associated with an MCS. Conversely, since only 6.4% of significant hail reports are associated with MCSs, an otherwise favorable hail signal may be less indicative of a hail threat if it is associated with an MCS.

The model forecasts of MCSs can be used directly by forecasters as simply an outline of where the model predicts an MCS will occur. Since the forecasts can be made for all members of the HREF, they can also be combined into an ensemble probability. An ensemble probability would also simplify the output and prevent a forecaster from having to check each individual member. The ensemble probabilities would also take advantage of the time lagged members of the HREF, which were not evaluated here. The ensemble probabilities would help show uncertainty in the forecast. Especially given the high SR, low POD nature of the models' MCS forecasts, multiple models forecasting an MCS in the same area and producing high ensemble probabilities could be a very important signal for forecasters. The ensemble probabilities would need to be evaluated both objectively by using verification metrics and subjectively by getting feedback from forecasters and other users.

The modified Wade algorithm can be used to identify MCSs in calibrated guidance, such as in Jahn et al. (2022). The Jahn method uses UH to identify model MCSs and

applies a modified tornado threat to grid squares that are identified as part of an MCS. In other words, given otherwise similar environments, a grid square within an MCS has a lower probability of a tornado than a grid square not in an MCS. Instead of only using UH to identify MCSs, the modified Wade algorithm could be applied. While the method in Jahn et al. (2022) was developed for tornado forecasting, a similar method could be applied to improve calibrated severe hail or severe wind guidance. Smith et al. (2012) found that significant severe wind was more common within MCSs while significant severe hail was less common. MCSs could be identified using the modified Wade algorithm, then the probability of severe wind within an MCS could be increased while the probability of severe hail within an MCS could be decreased, given otherwise similar environment that are not in MCSs. As with the ensemble probabilities, any new calibrated guidance product would also need to be evaluated. For the calibrated guidance in particular, forecaster feedback would be useful to

The development of the original Wade algorithm is part of an ongoing effort by the Storm Prediction Center to classify severe wind producing MCSs, and the two cases presented in Section 4.3.4 potentially suggest that the models may forecast more intense MCSs more accurately than weaker MCSs. More research is needed to confirm that initial suggestion, but if it holds true, it would be extremely useful information for forecasters since model MCSs could be associated with more intense observed MCSs.

The research presented here demonstrated that while numerical forecast models produce storm objects that are similar in shape to observed objects, the model objects are not oriented in the same direction as observed, and maximum reflectivity within each model object does not match observations. Despite the differences in orientation, the similarities in shape between forecast and observed objects helped inform the development of the modified Wade MCS detection algorithm, which was based on an algorithm used to identify MCSs in observations. Due to the limited temporal

resolution of model data compared to observed, a novel tracking algorithm based on hourly maximum reflectivity was developed. Finally, the modified Wade algorithm was applied to a full year of model output and model MCSs were verified against observed MCSs. Though the models can and do produce MCSs, they do not produce enough of them. However, the MCSs that the models do produce often correspond to an observed MCS. Two cases suggested that there may be a relationship between MCS strength and the models' forecasting ability. In addition, this research could potentially drive further research into the forecasting applications of the modified Wade algorithm. MCSs could be identified and used as part of an ensemble probability of MCS occurrence and MCSs identified in the models could also be used to improved calibrated severe weather guidance. As part of the evaluation of the model MCSs, new error evaluation graphics were developed that can help to visualize model errors. These graphics can be used not only for MCS forecasts, but for any object-based forecast to visualize the direction of model errors.

Bibliography

Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of Spatial Verification Methods to Idealized and NWP-Gridded Precipitation Forecasts. *Weather and Forecasting*, **24 (6)**, 1485–1497, https://doi.org/10.1175/2009WAF2222298.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2009WAF2222298.1.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature 2015 525:7567*, **525 (7567)**, 47–55, https://doi.org/10.1038/nature14956, URL https://www.nature.com/articles/nature14956.

Bluestein, H. B., and M. H. Jain, 1985: Formation of Mesoscale Lines of Pirecipitation: Severe Squall Lines in Oklahoma during the Spring. *Journal of the Atmospheric Sciences*, **42 (16)**, 1711–1732, https://doi.org/10.1175/1520-0469(1985)042, URL https://journals.ametsoc.org/view/journals/atsc/42/16/1520-0469_1985_042_1711_fomlop_2_0_co_2.xml.

Britt, K. C., P. S. Skinner, P. L. Heinselman, C. K. Potvin, M. L. Flora, B. Matilla, K. H. Knopfmeier, and A. E. Reinhart, 2024: Verification of Quasi-Linear Convective Systems Predicted by the Warn-on-Forecast System (WoFS). *Weather and Forecasting*, **39 (1)**, 155–176, https://doi.org/10.1175/WAF-D-23-0106.1, URL https://journals.ametsoc.org/view/journals/wefo/39/1/WAF-D-23-0106.1.xml.

Brown, B. G., L. Holland, J. E. H. Gotway, R. Bullock, D. A. Ahijevych, E. Gilleland, and C. A. Davis, 2007: Application of the MODE object-based verification tool for the evaluation of model precipitation fields. *22nd Conference on Weather Analysis and Forecasting / 18th Conference on Numerical Weather Prediction*, 8, URL https://ams.confex.com/ams/22WAF18NWP.

Cecchini, M. A., M. A. Silva Dias, L. A. Machado, C. A. Morales Rodriguez, and T. Biscaro, 2020: Macrophysical and Microphysical Characteristics of Convective Rain Cells Observed During SOS-CHUVA. *Journal of Geophysical Research: Atmospheres*, **125 (13)**, e2019JD031 187, https://doi.org/10.1029/2019JD031187, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2019JD031187https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031187https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019JD031187.

Clark, A. J., and Coauthors, 2023: The First Hybrid NOAA Hazardous Weather Testbed Spring Forecasting Experiment for Advancing Severe Weather Prediction. *Bulletin of the American Meteorological Society*, **104 (12)**, E2305–E2307, https://doi.org/10.1175/BAMS-D-23-0275.1.

Clarke, S. J., S. L. Gray, and N. M. Roberts, 2019a: Downstream influence of mesoscale convective systems. Part 1: influence on forecast evolution. *Quarterly Journal of the Royal Meteorological Society*, **145 (724)**, 2933–2952, https://doi.org/10.1002/QJ.3593, URL https://onlinelibrary.wiley.com/doi/full/10.1002/qj.3593https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3593https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3593.

Clarke, S. J., S. L. Gray, and N. M. Roberts, 2019b: Downstream influence of mesoscale convective systems. Part 2: Influence on ensemble forecast skill and spread. *Quarterly Journal of the Royal Meteorological Society*, **145 (724)**, 2953–2972, https://doi.org/10.1002/QJ.3613, URL https://onlinelibrary.wiley.com/doi/full/10.1002/qj.3613https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3613https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3613.

Colbert, M., D. J. Stensrud, P. M. Markowski, and Y. P. Richardson, 2019: Processes Associated with Convection Initiation in the North American Mesoscale Forecast System, Version 3 (NAMv3). *Weather and Forecasting*, **34 (3)**, 683–700, https://doi.org/10.1175/WAF-D-18-0175.1, URL https://journals.ametsoc.org/view/journals/wefo/34/3/waf-d-18-0175_1.xml.

Coniglio, M. C., H. E. Brooks, S. J. Weiss, and S. F. Corfidi, 2007: Forecasting the Maintenance of Quasi-Linear Mesoscale Convective Systems. *Weather and Forecasting*, **22 (3)**, 556–570, https://doi.org/10.1175/WAF1006.1.

Corfidi, S. F., 2003: Cold Pools and MCS Propagation: Forecasting the Motion of Downwind-Developing MCSs. *Weather and Forecasting*, **18 (6)**, 997–1017, https://doi.org/10.1175/1520-0434(2003)018⟨0997:CPAMPF⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0434(2003)018%3C0997:CPAMPF%3E2.0.CO;2.

Corfidi, S. F., J. H. Meritt, and J. M. Fritsch, 1996: Predicting the Movement of Mesoscale Convective Complexes. *Weather and Forecasting*, **11 (1)**, 41–46, https://doi.org/10.1175/1520-0434(1996)011, URL https://journals.ametsoc.org/view/journals/wefo/11/1/1520-0434_1996_011_0041_ptmomc_2_0_co_2.xml.

Da Silva, N. A., and J. O. Haerter, 2023: The Precipitation Characteristics of Mesoscale Convective Systems Over Europe. *Journal of Geophysical Research: Atmospheres*, **128 (23)**, e2023JD039 045, https://doi.org/10.1029/2023JD039045, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2023JD039045https://onlinelibrary.wiley.com/doi/abs/10.1029/2023JD039045https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023JD039045.

Davis, C., B. Brown, and R. Bullock, 2006: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Monthly Weather Review*, **134 (7)**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmospheric Science Letters*, **5 (6)**, 110–117, https://doi.org/10.1002/asl.72, URL http://doi.wiley.com/10.1002/asl.72.

Duda, J. D., and D. D. Turner, 2021: Large-sample application of radar reflectivity object-based verification to evaluate HRRR warm-season forecasts. *Weather and Forecasting*, **36 (3)**, https://doi.org/10.1175/WAF-D-20-0203.1.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. *Journal of Hydrology*, **239 (1-4)**, 179–202, https://doi.org/10.1016/S0022-1694(00)00343-7.

Fairman, J. G., D. M. Schultz, D. J. Kirshbaum, S. L. Gray, and A. I. Barrett, 2016: Climatology of Banded Precipitation over the Contiguous United States. *Monthly Weather Review*, **144 (12)**, 4553–4568, https://doi.org/10.1175/MWR-D-16-0015.1, URL https://journals.ametsoc.org/view/journals/mwre/144/12/mwr-d-16-0015.1.xml.

Feng, Z., and Coauthors, 2021: A Global High-Resolution Mesoscale Convective System Database Using Satellite-Derived Cloud Tops, Surface Precipitation, and Tracking. *Journal of Geophysical Research: Atmospheres*, **126 (8)**, e2020JD034 202, https://doi.org/10.1029/2020JD034202, URL https://onlinelibrary.wiley.com/doi/full/10.1029/2020JD034202https://onlinelibrary.wiley.com/doi/abs/10.1029/2020JD034202https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020JD034202.

Flora, M. L., P. S. Skinner, C. K. Potvin, A. E. Reinhart, T. A. Jones, N. Yussouf, and K. H. Knopfmeier, 2019: Object-Based Verification of Short-Term, Storm-Scale Probabilistic Mesocyclone Guidance from an Experimental Warn-on-Forecast System. *Weather and Forecasting*, **34 (6)**, 1721–1739, https://doi.org/10.1175/WAF-D-19-0094.1, URL https://journals.ametsoc.org/view/journals/wefo/34/6/waf-d-19-0094_1.xml.

Gallo, B. T., A. J. Clark, B. T. Smith, R. L. Thompson, I. Jirak, and S. R. Dembek, 2018: Blended Probabilistic Tornado Forecasts: Combining Climatological Frequencies with NSSL–WRF Ensemble Forecasts. *Weather and Forecasting*, **33 (2)**, 443–460, https://doi.org/10.1175/WAF-D-17-0132.1, URL https://journals.ametsoc.org/view/journals/wefo/33/2/waf-d-17-0132_1.xml.

Gallus, W. A., 2010: Application of object-based verification techniques to ensemble precipitation forecasts. *Weather and Forecasting*, **25 (1)**, https://doi.org/10.1175/2009WAF2222274.1.

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, **24 (5)**, 1416–1430, https://doi.org/10.1175/2009WAF2222269.1, URL https://journals.ametsoc.org/view/journals/wefo/24/5/2009waf2222269_1.xml.

Grim, J. A., J. O. Pinto, and D. C. Dowell, 2024: Assessing RRFS versus HRRR in Predicting Widespread Convective Systems over the Eastern CONUS. *Weather and Forecasting*, **39 (1)**, 121–140, https://doi.org/10.1175/WAF-D-23-0112.1, URL https://journals.ametsoc.org/view/journals/wefo/39/1/WAF-D-23-0112.1.xml.

Haberlie, A. M., and W. S. Ashley, 2018a: A method for identifying midlatitude mesoscale convective systems in radar mosaics. Part I: Segmentation and classification. *Journal of Applied Meteorology and Climatology*, **57 (7)**, https://doi.org/10.1175/JAMC-D-17-0293.1.

Haberlie, A. M., and W. S. Ashley, 2018b: A method for identifying midlatitude mesoscale convective systems in radar mosaics. Part II: Tracking. *Journal of Applied Meteorology and Climatology*, **57 (7)**, https://doi.org/10.1175/JAMC-D-17-0294.1.

Haberlie, A. M., and W. S. Ashley, 2019: A radar-based climatology of mesoscale convective systems in the United States. *Journal of Climate*, **32 (5)**, https://doi.org/10.1175/JCLI-D-18-0559.1.

Haberlie, A. M., W. S. Ashley, V. A. Gensini, and A. C. Michaelis, 2023: The ratio of mesoscale convective system precipitation to total precipitation increases in future climate change scenarios. *npj Climate and Atmospheric Science 2023 6:1*, **6 (1)**, 1–5, https://doi.org/10.1038/s41612-023-00481-5, URL https://www.nature.com/articles/s41612-023-00481-5.

Han, L., S. Fu, L. Zhao, Y. Zheng, H. Wang, and Y. Lin, 2009: 3D Convective Storm Identification, Tracking, and Forecasting—An Enhanced TITAN Algorithm. *Journal of Atmospheric and Oceanic Technology*, **26 (4)**, 719–732, https://doi.org/10.1175/2008JTECHA1084.1, URL https://journals.ametsoc.org/view/journals/atot/26/4/2008jtecha1084_1.xml.

Hanley, J. A., and B. J. McNeil, 1982: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143 (1)**, https://doi.org/10.1148/radiology.143.1.7063747.

Harrison, D. R., M. S. Elliott, I. L. Jirak, and P. T. Marsh, 2022: Utilizing the High-Resolution Ensemble Forecast System to Produce Calibrated Probabilistic Thunderstorm Guidance. *Weather and Forecasting*, **37 (7)**, https://doi.org/10.1175/WAF-D-22-0001.1.

Henderson, D. S., J. A. Otkin, and J. R. Mecikalski, 2021: Evaluating Convective Initiation in High-Resolution Numerical Weather Prediction Models Using GOES-16 Infrared Brightness Temperatures. *Monthly Weather Review*, **149 (4)**, 1153–1172, https://doi.org/10.1175/MWR-D-20-0272.1, URL https://journals.ametsoc.org/view/journals/mwre/149/4/MWR-D-20-0272.1.xml.

Herman, G. R., E. R. Nielsen, and R. S. Schumacher, 2018: Probabilistic Verification of Storm Prediction Center Convective Outlooks. *Weather and Forecasting*, **33 (1)**, 161–184, https://doi.org/10.1175/WAF-D-17-0104.1, URL https://journals.ametsoc.org/view/journals/wefo/33/1/waf-d-17-0104_1.xml.

Herman, G. R., and R. S. Schumacher, 2016: Extreme Precipitation in Models: An Evaluation. *Weather and Forecasting*, **31 (6)**, 1853–1879, https://doi.org/10.1175/WAF-D-16-0093.1, URL http://journals.ametsoc.org/doi/10.1175/WAF-D-16-0093.1.

Hitchcock, S. M., and R. S. Schumacher, 2020: Analysis of Back-Building Convection in Simulations with a Strong Low-Level Stable Layer. *Monthly Weather Review*, **148 (9)**, 3773–3797, https://doi.org/10.1175/MWR-D-19-0246.1, URL https://journals.ametsoc.org/view/journals/mwre/148/9/mwrD190246.xml.

Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective Limits on Forecasting Skill of Rare Events. *Weather and Forecasting*, **28 (2)**, 525–534, https://doi.org/10.1175/WAF-D-12-00113.1, URL http://journals.ametsoc.org/doi/full/10.1175/WAF-D-12-00113.1.

Houze, R. A., 2004: Mesoscale convective systems. *Reviews of Geophysics*, **42 (4)**, RG4003, https://doi.org/10.1029/2004RG000150, URL http://doi.wiley.com/10.1029/2004RG000150.

Houze, R. A., 2018: 100 Years of Research on Mesoscale Convective Systems. *Meteorological Monographs*, **59 (1)**, 1–17, https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0001.1, URL https://journals.ametsoc.org/view/journals/amsm/59/1/amsmonographs-d-18-0001.1.xml.

Houze, R. A., S. A. Rutledge, M. I. Biggerstaff, and B. F. Smull, 1989: Interpretation of Doppler weather radar displays of midlatitude mesoscale convective systems. *Bulletin - American Meteorological Society*, **70 (6)**, https://doi.org/10.1175/1520-0477(1989)070⟨0608:IODWRD⟩2.0.CO;2.

Hsu, W. C., G. J. Kooperman, W. M. Hannah, K. A. Reed, A. A. Akinsanola, and A. G. Pendergrass, 2023: Evaluating Mesoscale Convective Systems Over the US in Conventional and Multiscale Modeling Framework Configurations of E3SMv1. *Journal of Geophysical Research: Atmospheres*, **128 (23)**, e2023JD038 740, https://doi.org/10.1029/2023JD038740,

URL https://onlinelibrary.wiley.com/doi/full/10.1029/2023JD038740https://onlinelibrary.wiley.com/doi/abs/10.1029/2023JD038740https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023JD038740.

Iowa State University, 2001: Iowa Environmental Mesonet. URL https://mesonet.agron.iastate.edu/docs/nexrad_mosaic/.

Jahn, D. E., I. L. Jirak, A. Wade, and J. Milne, 2022: Storm Mode and Tornado Potential Determination Using Statistical Moments of Updraft Helicity Distribution. *31st Conf. Weather Analysis and Forecasting/27th Conf. Numerical Weather Prediction*, AMS, Virtual.

Jergensen, G. E., A. McGovern, R. Lagerquist, and T. Smith, 2020: Classifying Convective Storms Using Machine Learning. *Weather and Forecasting*, **35 (2)**, 537–559, https://doi.org/10.1175/WAF-D-19-0170.1, URL https://journals.ametsoc.org/view/journals/wefo/35/2/waf-d-19-0170.1.xml.

Jirak, I. L., and W. R. Cotton, 2007: Observational Analysis of the Predictability of Mesoscale Convective Systems. *Weather and Forecasting*, **22 (4)**, 813–838, https://doi.org/10.1175/WAF1012.1, URL https://journals.ametsoc.org/view/journals/wefo/22/4/waf1012_1.xml.

Jirak, I. L., S. J. Weiss, and C. J. Melick, 2012: The SPC Storm-Scale Ensemble of Opportunity: Overview and Results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *26th Conference on Severe and Local Storms*, American Meteorological Society, Nashville, TN, P9.137, URL http://www.spc.noaa.gov/publications/jirak/sseo_hwt.pdf.

Johns, R. H., and W. D. Hirt, 1987: Derechos: Widespread Convectively Induced Windstorms. *Weather and Forecasting*, **2 (1)**, 32–49, https://doi.org/10.1175/1520-0434(1987)002⟨0032:DWCIW⟩2.0.CO;2.

Johnson, A., and X. Wang, 2012: Object-Based Evaluation of a Storm-Scale Ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Monthly Weather Review*, **141 (3)**, 1079–1098, https://doi.org/10.1175/MWR-D-12-00140.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-12-00140.1.

Johnson, J. T., P. L. Mackeen, A. Witt, E. D. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, **13 (2)**, https://doi.org/10.1175/1520-0434(1998)013⟨0263:TSCIAT⟩2.0.CO;2.

Keene, K. M., and R. S. Schumacher, 2013: The Bow and Arrow Mesoscale Convective Structure. *Monthly Weather Review*, **141 (5)**, 1648–1672, https://doi.org/10.1175/MWR-D-12-00172.1, URL https://journals.ametsoc.org/view/journals/mwre/141/5/mwr-d-12-00172.1.xml.

Koch, S. E., 1985: Ability of a Regional-Scale Model to Predict the Genesis of Intense Mesoscale Convective Systems. *Monthly Weather Review*, **113 (10)**, 1693–1713, https://doi.org/10.1175/1520-0493(1985)113, URL https://journals.ametsoc.org/view/journals/mwre/113/10/1520-0493_1985_113_1693_aoarsm_2_0_co_2.xml.

Lack, S. A., G. L. Limpert, and N. I. Fox, 2010: An object-oriented multiscale verification scheme. *Weather and Forecasting*, **25 (1)**, https://doi.org/10.1175/2009WAF2222245.1.

Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An Efficient, General-Purpose Technique for Identifying Storm Cells in Geospatial Images. *Journal of Atmospheric and Oceanic Technology*, **26 (3)**, 523–537, https://doi.org/10.1175/2008JTECHA1153.1, URL https://journals.ametsoc.org/view/journals/atot/26/3/2008jtecha1153_1.xml.

Lakshmanan, V., and T. Smith, 2009: Data Mining Storm Attributes from Spatial Grids. *Journal of Atmospheric and Oceanic Technology*, **26 (11)**, 2353–2365, https://doi.org/10.1175/2009JTECHA1257.1, URL https://journals.ametsoc.org/view/journals/atot/26/11/2009jtecha1257_1.xml.

Letkewicz, C. E., and M. D. Parker, 2010: Forecasting the Maintenance of Mesoscale Convective Systems Crossing the Appalachian Mountains. *Weather and Forecasting*, **25 (4)**, 1179–1195, https://doi.org/10.1175/2010WAF2222379.1, URL https://journals.ametsoc.org/view/journals/wefo/25/4/2010waf2222379_1.xml.

Maddox, R. A., 1980: Mesoscale convective complexes. *Bulletin, American Meteorological Society*, **61 (11)**.

Maddox, R. A., 1983: Large-Scale Meteorological Conditions Associated with Midlatitude, Mesoscale Convective Complexes. *Monthly Weather Review*, **111 (7)**, 1475–1493, https://doi.org/10.1175/1520-0493(1983)111, URL https://journals.ametsoc.org/view/journals/mwre/111/7/1520-0493_1983_111_1475_lsmcaw_2_0_co_2.xml.

Marzban, C., 2004: The ROC Curve and the Area under It as Performance Measures. *Weather and Forecasting*, **19 (6)**, 1106–1114, https://doi.org/10.1175/825.1, URL https://journals.ametsoc.org/view/journals/wefo/19/6/825_1.xml.

May, P. T., and A. Ballinger, 2007: The Statistical Characteristics of Convective Cells in a Monsoon Regime (Darwin, Northern Australia). *Monthly Weather Review*, **135 (1)**, 82–92, https://doi.org/10.1175/MWR3273.1, URL https://journals.ametsoc.org/view/journals/mwre/135/1/mwr3273.1.xml.

Nielsen, E. R., and R. S. Schumacher, 2016: Using Convection-Allowing Ensembles to Understand the Predictability of an Extreme Rainfall Event. *Monthly Weather Review*, **144 (10)**, 3651–3676, https://doi.org/10.1175/MWR-D-16-0083.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-16-0083.1.

Parker, M. D., and R. H. Johnson, 2000: Organizational Modes of Midlatitude Mesoscale Convective Systems. *Monthly Weather Review*, **128 (10)**, 3413–3436, https://doi.org/10.1175/1520-0493(2001)129⟨3413:OMOMMC⟩2.0.CO;2.

Parsons, D. B., S. P. Lillo, C. P. Rattray, P. Bechtold, M. J. Rodwell, and C. M. Bruce, 2019: The Role of Continental Mesoscale Convective Systems in Forecast Busts within Global Weather Prediction Systems. *Atmosphere 2019, Vol. 10, Page 681*, **10 (11)**, 681, https://doi.org/10.3390/ATMOS10110681, URL https://www.mdpi.com/2073-4433/10/11/681/htmhttps://www.mdpi.com/2073-4433/10/11/681.

Peters, J. M., E. R. Nielsen, M. D. Parker, S. M. Hitchcock, and R. S. Schumacher, 2017: The Impact of Low-Level Moisture Errors on Model Forecasts of an MCS Observed during PECAN. *Monthly Weather Review*, **145 (9)**, 3599–3624, https://doi.org/10.1175/MWR-D-16-0296.1, URL https://journals.ametsoc.org/view/journals/mwre/145/9/mwr-d-16-0296.1.xml.

Pinto, J. O., J. A. Grim, and M. Steiner, 2015: Assessment of the High-Resolution Rapid Refresh Model's Ability to Predict Mesoscale Convective Systems Using Object-Based Evaluation. *Weather and Forecasting*, **30 (4)**, 892–913, https://doi.org/10.1175/WAF-D-14-00118.1, URL https://journals.ametsoc.org/view/journals/wefo/30/4/waf-d-14-00118_1.xml.

Potvin, C. K., and Coauthors, 2022: An Iterative Storm Segmentation and Classification Algorithm for Convection-Allowing Models and Gridded Radar Analyses. *Journal of Atmospheric and Oceanic Technology*, **39 (7)**, 999–1013, https://doi.org/10.1175/JTECH-D-21-0141.1, URL https://journals.ametsoc.org/view/journals/atot/39/7/JTECH-D-21-0141.1.xml.

Roberts, B., B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What Does a Convection-Allowing Ensemble of Opportunity Buy Us in Forecasting Thunderstorms? *Weather and Forecasting*, **35 (6)**, 2293–2316, https://doi.org/10.1175/WAF-D-20-0069.1, URL https://journals.ametsoc.org/view/journals/wefo/35/6/WAF-D-20-0069.1.xml.

Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bulletin of the American Meteorological Society*, **100 (7)**, https://doi.org/10.1175/BAMS-D-18-0041.1.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteorological Applications*, **15 (1)**, 163–169, https://doi.org/10.1002/MET.57, URL https://onlinelibrary.wiley.com/doi/full/10.1002/met.57https://onlinelibrary.wiley.com/doi/abs/10.1002/met.57https://rmets.onlinelibrary.wiley.com/doi/10.1002/met.57.

Roberts, N. M., and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather*

Review, **136 (1)**, 78–97, https://doi.org/10.1175/2007MWR2123.1, URL https://journals.ametsoc.org/view/journals/mwre/136/1/2007mwr2123.1.xml.

Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bulletin of the American Meteorological Society*, **94 (9)**, https://doi.org/10.1175/BAMS-D-12-00099.1.

Roebber, P. J., 2009: Visualizing Multiple Measures of Forecast Quality. *Weather and Forecasting*, **24 (2)**, 601–608, https://doi.org/10.1175/2008WAF2222159.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2008WAF2222159.1.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A theory for strong, long-lived squall lines. *Journal of the Atmospheric Sciences*, **45 (3)**, https://doi.org/10.1175/1520-0469(1988)045⟨0463:ATFSLL⟩2.0.CO;2.

Schumacher, R. S., and K. L. Rasmussen, 2020: The formation, character and changing nature of mesoscale convective systems. *Nature Reviews Earth & Environment 2020 1:6*, **1 (6)**, 300–314, https://doi.org/10.1038/s43017-020-0057-7, URL https://www.nature.com/articles/s43017-020-0057-7.

Schwartz, C. S., G. S. Romine, and D. C. Dowell, 2021: Toward Unifying Short-Term and Next-Day Convection-Allowing Ensemble Forecast Systems with a Continuously Cycling 3-km Ensemble Kalman Filter over the Entire Conterminous United States. *Weather and Forecasting*, **36 (2)**, 379–405, https://doi.org/10.1175/WAF-D-20-0110.1, URL https://journals.ametsoc.org/view/journals/wefo/36/2/WAF-D-20-0110.1.xml.

Short, C. J., and J. Petch, 2022: Reducing the spin-up of a regional NWP system without data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **148 (745)**, 1623–1643, https://doi.org/10.1002/QJ.4268, URL https://onlinelibrary.wiley.com/doi/full/10.1002/qj.4268https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.4268https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4268.

Short, E., and T. P. Lane, 2023: Objectively Assessing Characteristics of Mesoscale Convective Organization in an Operational Convection-Permitting Model. *Monthly Weather Review*, **151 (10)**, 2819–2841, https://doi.org/10.1175/MWR-D-23-0033.1, URL https://journals.ametsoc.org/view/journals/mwre/151/10/MWR-D-23-0033.1.xml.

Short, E., T. P. Lane, and C. L. Vincent, 2023: Objectively Diagnosing Characteristics of Mesoscale Organization from Radar Reflectivity and Ambient Winds. *Monthly Weather Review*, **151 (3)**, 643–662, https://doi.org/10.1175/MWR-D-22-0146.1, URL https://journals.ametsoc.org/view/journals/mwre/151/3/MWR-D-22-0146.1.xml.

Skinner, P. S., and Coauthors, 2018: Object-Based Verification of a Prototype Warn-on-Forecast System. *Weather and Forecasting*, **33 (5)**, 1225–1250, https://doi.org/10.1175/WAF-D-18-0020.1, URL https://journals.ametsoc.org/view/journals/wefo/33/5/waf-d-18-0020_1.xml.

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective Modes for Significant Severe Thunderstorms in the Contiguous United States. Part I: Storm Classification and Climatology. *Weather and Forecasting*, **27 (5)**, 1114–1135, https://doi.org/10.1175/WAF-D-11-00115.1, URL https://journals.ametsoc.org/view/journals/wefo/27/5/waf-d-11-00115_1.xml.

Smull, B. F., and R. A. Houze, 1987: Rear inflow in squall lines with trailing stratiform precipitation. *Monthly Weather Review*, **115 (12)**, https://doi.org/10.1175/1520-0493(1987)115⟨2869:RIISLW⟩2.0.CO;2.

Sobash, R. A., D. J. Gagne, C. L. Becker, D. Ahijevych, G. N. Gantos, and C. S. Schwartz, 2023: Diagnosing Storm Mode with Deep Learning in Convection-Allowing Models. *Monthly Weather Review*, https://doi.org/10.1175/MWR-D-22-0342.1.

Sofaer, H. R., J. A. Hoeting, C. S. Jarnevich, and C. R. Helen Sofaer, 2019: The area under the precision-recall curve as a performance metric for rare binary events. *Methods Ecol Evol*, **10**, https://doi.org/10.1111/2041-210X.13140.

Squitieri, B. J., A. R. Wade, I. L. Jirak, and B. S. Gov, 2023: A Historical Overview on the Science of Derechos: Part II: Parent Storm Structure, Environmental Conditions, and History of Numerical Forecasts. *Bulletin of the American Meteorological Society*, **104 (10)**, E1734–E1763, https://doi.org/10.1175/BAMS-D-22-0278.1, URL https://journals.ametsoc.org/view/journals/bams/104/10/BAMS-D-22-0278.1.xml.

Surcel, M., I. Zawadzki, and M. K. Yau, 2016: The Case-to-Case Variability of the Predictability of Precipitation by a Storm-Scale Ensemble Forecasting System. *Monthly Weather Review*, **144 (1)**, 193–212, https://doi.org/10.1175/MWR-D-15-0232.1, URL https://journals.ametsoc.org/view/journals/mwre/144/1/mwr-d-15-0232.1.xml.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from Squall Lines and Bow Echoes. Part I: Climatological Distribution. *Weather and Forecasting*, **20 (1)**, 23–34, https://doi.org/10.1175/WAF-835.1, URL https://journals.ametsoc.org/view/journals/wefo/20/1/waf-835_1.xml.

Wade, A., B. Squitieri, and I. L. Jiral, 2023: 27-Year Climatology of Severe Wind-Producing Mesoscale Convective Systems in the United States. *32nd Conf. Weather Analysis and Forecasting (WAF)/28th Conf. Numerical Weather Prediction/20th Conf. Mesoscale Processes*, AMS, Madison, WI.