January 2016

Improving Forecasts Of Winter Storm Tracks Using A Local Ensemble

Brittany A. Peterson

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

IMPROVING FORECASTS OF WINTER STORM TRACKS
USING A LOCAL ENSEMBLE

by

Brittany Ann Peterson
Bachelor of Science, Iowa State University, 2008

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

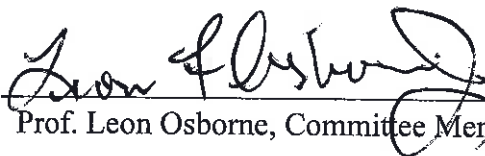Grand Forks, North Dakota

December
2016

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

_____

Dr. Gretchen Mullendore, Committee Chair
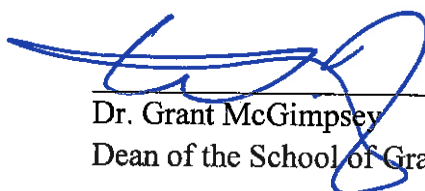
_____

Dr. Mark Askelson, Committee Member

_____

Prof. Leon Osborne, Committee Member

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

_____

Dr. Grant McGimpsey
Dean of the School of Graduate Studies

December 12, 2016
_____
Date

PERMISSION

Title   Improving Forecasts of Winter Storm Tracks Using a
       Local Ensemble

Department Atmospheric Sciences

Degree   Master of Science

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

<div align="right">

Brittany A. Peterson
October 27, 2016

</div>

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

ABSTRACT

The strong winds, extreme snowfall, and low visibilities that often accompany winter storms can affect millions of people, with impacts such as disruptions to transportation, hazards to human health, reduction in retail sales, and structural damage. Blizzard forecasts for Alberta Clippers can be a particular challenge in the Northern Plains, as these systems typically depart from the Canadian Rockies, intensify, and impact the Northern Plains all within 24 hours. The purpose of this study is to determine whether probabilistic forecasts derived from a local physics-based ensemble can improve specific aspects of winter storm forecasts for three Alberta Clipper cases.

The ensemble includes three different initialization times to capture temporal uncertainty in initial system intensification and combinations of different initializations and various radiation, land surface, and microphysics schemes. Verification is performed on the individual ensemble members and the ensemble mean with a focus on quantifying uncertainty in the overall storm track, two-meter winds, and precipitation using the MERRA and NOHRSC SNODAS datasets. This study finds that addition improvements are needed in order to proceed with operational use of the ensemble blizzard products, but the use of a proxy for blizzard conditions yields promising results.

CHAPTER I

INTRODUCTION

Hazardous winter weather plays an important role in central North America. The strong winds, extreme snowfall, and low visibilities that often accompany winter storms and blizzards can affect millions of people (Schwartz and Schmidlin 2002), with impacts such as disruptions to transportation, hazards to human health, reduction in retail sales, and structural damage. Alberta clippers, for example, may result in relatively low precipitation amounts across a broad area. However, clippers often include strong winds and narrow but significant bands of snowfall (Thomas and Martin 2007), which can lead to blizzard conditions. Forecasting these conditions can be a particular challenge in central North America, as clippers typically depart from the Canadian Rockies, intensify, and impact the northern plains all within 24 hours.

An example of one of these events occurred 11 March 2011. By the morning of Monday March 7, NWS forecasters were aware of a clipper system forecast to pass through the central and eastern Dakotas on Friday March 11 (National Weather Service Grand Forks 2011). The primary hazards associated with the system were expected to be mixed precipitation and strong surface wind. By early morning on March 10, hazardous travel was anticipated for the following day, but forecasters noted a lack of model consistency in the clipper track and high forecast uncertainty in surface wind speed and precipitation type and amounts. A winter storm watch was issued 21 hours prior to the

event, and by early morning on March 11, the forecast for eastern North Dakota and northwest Minnesota included up to five inches of snow and surface wind gusts up to 40 mph. Strong surface wind speeds became the primary concern, prompting the issuance of a blizzard warning, in effect from 18:00 UTC Friday March 11 through 06:00 UTC Saturday March 12.

Due to uncertainty in the pressure gradient and precipitation amounts and locations, it was difficult for forecasters to determine whether blizzard conditions would be present until the day of the event. A local ensemble model may have added valuable details regarding wind speed and precipitation probabilities, which were needed to improve the forecast and provide more advanced warning of hazardous winter weather conditions. This study consists of establishment of a local ensemble (using both variation in physics and initialization) and analysis of both the ensemble and its forecasts for three wintertime clipper events. Ensemble forecasts like these would aid operational forecasting for hazardous winter weather scenarios (e.g. blizzard conditions) by supplying information regarding storm track and cyclone intensity from multiple simulations, and uncertainties and probabilistic forecasts such as the likelihood of measurable precipitation and surface wind speed greater than 11 m s^{-1} (25 mph). Additional analysis using differing initialization times provides further information about the spatial and temporal uncertainty of the storm track and intensity. Recommendations for local ensemble creation and the ensemble's value for operational hazardous winter weather scenarios are also provided.

CHAPTER II

BACKGROUND

Winter storm forecasts for Alberta clipper systems that affect the Northern Great Plains can be challenging since these systems propagate rapidly, producing strong surface winds and narrow but significant bands of snowfall (Thomas and Martin 2007). Past studies of cyclone storm tracks and intensities have focused upon improving operational forecasting (Hurley 1954) and theories of cyclogenesis, cyclolysis, and life cycles (Bjerknes and Solberg 1922, Carlson 1980, Shapiro and Keyser 1990). Careful examination of the uncertainty in the clipper's storm track and intensity forecasts is crucial for predicting mesoscale features with high impacts such as strong surface winds and the location of heavy snow bands. Recent ensemble studies focused on other high-impact events indicate that local multi-physics ensemble forecasts could provide beneficial information regarding forecast uncertainties associated with such storms (e.g., Novak and Colle 2012, Deppe et al. 2013).

2.1 The Alberta Clipper

According to the American Meteorological Society (AMS 2013), an Alberta clipper is "a low pressure system that is often fast-moving, has low moisture content, and originates in western Canada in or near Alberta province. In wintertime, it may be associated with a narrow but significant band of snowfall and typically affects portions of the Plains, Midwest, and East Coast." This definition is consistent with that used in

several studies (Thomas and Martin 2007, Schultz and Doswell 2000, Hutchinson 1995) that recognize a clipper based on its Canadian Rockies origination, rapid movement, and defined sea level pressure (SLP) minimum associated with a 500 hPa vorticity maximum.

While the narrow but significant band of snowfall is a noteworthy hazard, the most significant hazard often associated with a clipper is strong wind (Thomas and Martin 2007). This is caused by a tight pressure gradient created by the exiting clipper and typically a strong trailing anticyclone. These strong winds disturb recently fallen or currently falling snow from the clipper, causing blowing snow, reduced visibility and possibly blizzard conditions (Schwartz and Schmidlin 2002, Thomas and Martin 2007). Alberta clippers are the most frequent blizzard-producing storm in the Canadian Prairie (immediately northwest of the area of interest in this study), producing approximately five blizzards each winter (Stewart et al. 1995).
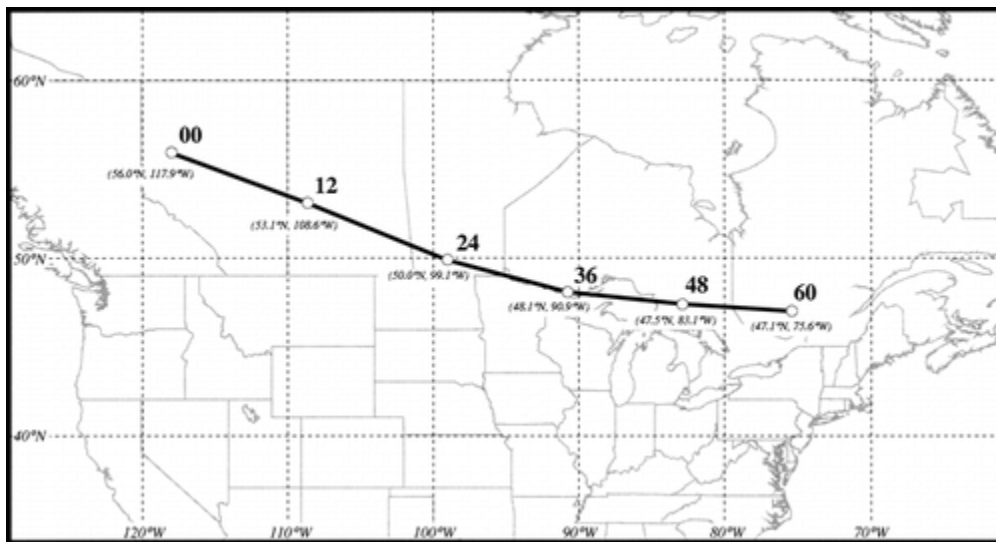
Figure 1. Average position of all Alberta clippers in the climatology at the given time (hrs) after departure is given by the gray dots (Thomas and Martin 2007, Fig. 6). The black line connecting the dots represents the average track of the clippers in the climatology. Latitude and longitude of each average position are also indicated.

4

Thomas and Martin (2007) suggested the dissipation of a Pacific cyclone as a precursor to the development of a new clipper within a lee trough in the lee of the Canadian Rockies. They found that, on average, clippers track across the Northern Great Plains for the first 24 hours after departure from the Canadian Rockies as shown in Fig. 1. Post-departure clippers typically have a central SLP from 1000 to 1009 hPa, which is not generally considered to be a strong cyclone (e.g., Angel and Isard 1997, Sanders and Gyakum 1980). After an average of 60 hours from departure from the Canadian Rockies, clippers have tracked across the Northern Great Plains and Great Lakes and are positioned in the Northeast U.S. and/or Quebec.

2.2 Storm Tracks

Storm tracks have been analyzed in previous studies using either an Eulerian filtered variance method (e.g., Blackmon 1976, Hoskins and Valdes 1990, Hoskins and Hodges 2002, Pinto et al. 2007) or a feature tracking method (e.g., Hurley 1954, Sanders and Gyakum 1980, Angel and Isard 1997, Thomas and Martin 2007, Ancell 2013). The Eulerian method uses band-pass filtering (low-pass and high-pass) to explore the variability of a field (e.g., 500 hPa geopotential height). This method works well for large data sets (i.e., GCMs), as influences due to elevation and the background pressure field are typically removed.

In contrast, the feature tracking method allows for the analysis of an individual synoptic system and its life cycle. Feature tracking generally uses SLP minima relative to surrounding points to create low pressure-center tracks. For example, Fig. 2 from Michaelis and Lackmann (2013) shows two manual track analyses of SLP minima based upon surface observations from two previous studies (Kocin and Uccellini 2004, AMS

1888a and AMS 1888b) and a WRF-simulated SLP minima track initialized with the

20CRv2 ensemble mean.
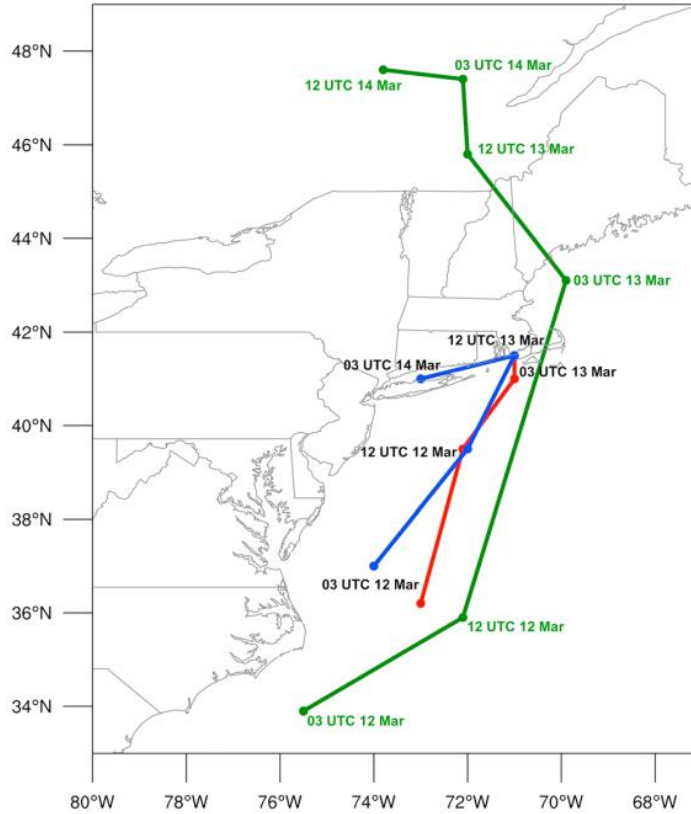
Figure 2. Cyclone center tracks for each analysis and model simulation from Michaelis and Lackmann (2013; Fig. 5). The red, blue, and green lines represent tracks of Kocin and Uccellini 2004 (hereafter KU04), AMS 1888a and 1888b (combined hereafter MWR88), and the Michaelis and Lackmann (2013) simulation, respectively. The black time labels correspond to KU04 and MWR88 and the green time labels correspond to the simulation. Locations for KU04 were only available through 12:00 UTC 13 March.

The feature tracking method using SLP has potential drawbacks that must be

avoided (Pauley 1998, Sinclair 1997, Mesinger and Treadon 1995), as determining actual

tracks can be complicated by the presence of multi-vortex cyclone centers (i.e., multiple

SLP minima), complex terrain, or the occurrence of center jumps. Sinclair (1994)

recommended that the use of cyclonic vorticity would be more consistent with other

observations (e.g., comma-shaped cloud signatures) in their study focused on Southern

6

Hemisphere climatology of cyclone tracks. Use of cyclonic vorticity could confirm weaker, secondary circulations that the SLP may correctly (or incorrectly) attribute to the background pressure field. However, systems having a similar central vorticity value may exhibit different sizes, structures, and intensity, making it more difficult to analyze individual systems (Sinclair 1997). Sinclair (1997) avoided some of the drawbacks listed above by analyzing cyclones using a combination of SLP, vorticity, and cyclonic circulation, but this approach excludes weak cyclones, which may include clippers. Wang et al. (2006) implemented feature tracking but used the local Laplacian of the pressure field to measure cyclone intensity, focused mainly on strong cyclones, and excluded areas with an elevation at or above 1000 m.

Several studies have addressed orographic features associated with storm track analysis. Sinclair (1997) excluded quasi-stationary orographic features (e.g., thermal lows) from analyses by requiring that a cyclone translate at least 1200 km if it spent its entire lifespan over or within 500 km of land. Pinto et al. (2007) addressed errors due to underground extrapolation of SLP by excluding systems in areas with terrain-heights greater than above 1,500 meters above sea-level. They also attempted to remove spurious weak lows by requiring at least a 24-hr cyclone lifetime and that the pressure gradient surrounding the low was above a certain threshold value.

Latitude can also introduce bias when using pressure to identify the most prominent cyclones. Thomas and Martin (2007) recognized that, despite having a similar pressure gradient, cyclones at different latitudes produce different geostrophic winds. Therefore, they adjusted SLP tendency with respect to a reference latitude (50 ºN) to compare intensities. Eichler et al. (2013) also accounted for latitudinal bias in a Northern

Hemisphere climate simulation study by computing storm track intensity relative to mean storm track intensity climatology.

2.3 Ensemble Forecasts

Probabilistic forecasting has become an integral component in the forecast process and has a wide range of applications. While many ensemble studies can be considered, this section only discusses those that are pertinent to this study. Toth et al. (2001) and Gneiting and Raftery (2005), along with many other studies discussed in this section, noted the importance of ensemble forecasts in improving forecast value, including their ability to express forecast uncertainty. Stensrud et al. (1999) stated that an ensemble approach can even compensate for the accuracy lost from using a coarser model resolution.

Much of ensemble research has a primary focus of improving tropical cyclone forecasts (e.g., Goldenberg et al. 2015), regional or global climate modeling (e.g., Pinto et al. 2007, Flaounas et al. 2011), or general forecasting (e.g., Du et al. 1997), but several recent ensemble studies are focused on improving forecast skill of high impact events for various applications. Jankov et al. (2007) quantified the impact of physics schemes and initial conditions on simulated orographically induced cold season rainfall and reduced bias in their WRF ensemble quantitative precipitation forecasts (QPF) to improve flood forecasting. Michaelis and Lackmann (2013) produced ensemble forecasts using initial and boundary conditions from 20th Century Reanalysis version 2 (20CRv2) data in order to explore the ensemble members' ability to depict the U.S. Northeast Blizzard of 1888. None of the ensemble member forecasts contained the solution (i.e., the correct storm track and intensity through the analysis time period) in their study, but the correct

cyclone intensity was forecast in approximately the right locations in the earlier stages of the event. Novak and Colle (2012) focused on the skill of an ensemble in predicting heavy snowbands in the Northeast. They found that while ensemble forecasts helped differentiate between cases with high and low predictability and hinted at the threat of band development in favored time periods and locations, significant uncertainty often remained in the specific details of band development. Deppe et al. (2013) looked to reduce mean absolute error in their WRF ensemble forecasts of significant wind speed increases or decreases (i.e., a ramp event) with applications to the wind industry. While the model forecasts of ramp events were poor, their results reinforced the need for further improvement in the PBL schemes' representation of turbulent and mixing processes. Hacker et al. (2011) agreed on the importance of the PBL schemes and noted that the multi-physics and perturbed observation ensembles performed similarly and quite poorly for rarer higher wind events. Although the above studies focus on different high-impact events, they agree that local ensemble forecasts add valuable details regarding high-impact event predictability and propose that further research is needed in this area. Thus, they imply that local or regional ensemble forecasts, such as those analyzed in this study, could elucidate intensity and track variability of a clipper system and its surface wind and precipitation fields as the clipper progresses through the Northern Great Plains.

Ensembles have traditionally been classified based on method of ensemble member production by altering one of three elements: initial and/or boundary conditions, physical parameters (physics ensemble), and physics parameterizations (multi-model ensemble). Several recent studies have adopted a hybrid approach of combining the physics and multi-model ensemble into a "multi-physics" ensemble. These studies have

led to better understanding of local ensemble production and the benefits of using a multi-physics ensemble. Stensrud et al. (2000) concluded that a multi-physics ensemble is more skilled than an initial condition ensemble when the large-scale forcing for upward motion is strong and vice versa when the large scale signal is weak. Evans et al. (2012) determined that their multi-physics ensemble's spread was much smaller for relatively weak systems than for strong or extreme systems. In Yuan et al. (2012), the WRF ensemble downscaling forecasts with differing physics schemes had larger spread than the Climate Forecast System (CFS) ensemble forecasts with differing initial conditions. Hacker et al. (2011) performed a comprehensive analysis of nine techniques for local Global Ensemble Forecast System (GEFS) ensemble production and concluded that the multi-physics ensemble showed significant improvement and greater spread than the control, a downscaled global ensemble. While all of the production techniques (besides the control) in Hacker et al. (2011) showed improvement in at least one metric over the multi-physics ensemble, physics diversity was critical for probabilistic prediction in the PBL and therefore the multi-physics approach showed the most skill in the PBL. Finally, Novak and Colle (2012) cautioned that predictability differences amongst cases are related to a complex combination of the initial conditions, complexity of the initial flow, and sensitivities to the model core, physics, and lateral boundary conditions. Thus, further work is needed to produce a more comprehensive examination of the relationships amongst all of these factors.

Several studies (e.g., Jankov et al. 2005 and Evans et al. 2012) found that no single ensemble member (i.e., physics configuration) performed best overall. Despite this, studies often suggested that the impacts of model physics schemes and

complementary physics scheme combinations could be deduced to reduce biases and improve forecasts through optimization for geographic regions or certain types of events (e.g., Yuan et al. 2012, Jankov et al. 2005). In Jankov et al. (2005), warm season rainfall ensemble forecast variability was introduced through changes in the microphysics, PBL, and most importantly convective schemes. Evans et al. (2012) deduced that Australian East Coast Lows are more sensitive to PBL and cumulus parameterization schemes than other types of schemes for both MSLP and wind ensemble forecasts. However, Yuan et al. (2012) found that radiation and land surface schemes had a more significant impact on forecasts of winter precipitation extremes than did microphysics and cumulus parameterization schemes. The RUC land surface and RRTMG radiation schemes performed best in general for this study, but performance differed based on location.

CHAPTER III

DATA AND METHODOLOGY

3.1 Weather Research and Forecasting (WRF) Ensemble Design

The WRF model is a numerical weather prediction system designed for a broad range of research and operational applications (Skamarock et al. 2008). This community model is suitable for real-time or idealized data, and it has two dynamics solvers: the Advanced Research WRF (ARW) and the Nonhydrostatic Mesoscale Model (NMM). This study uses ARW version 3.5.1 to generate a WRF physics ensemble. The WRF ensemble is used to simulate three recent cases of Northern Great Plains winter storms forced by Alberta clippers: 2011 March 11, 2009 January 12, and 2013 December 28. These cases were chosen for their differing tracks and precipitation amounts.
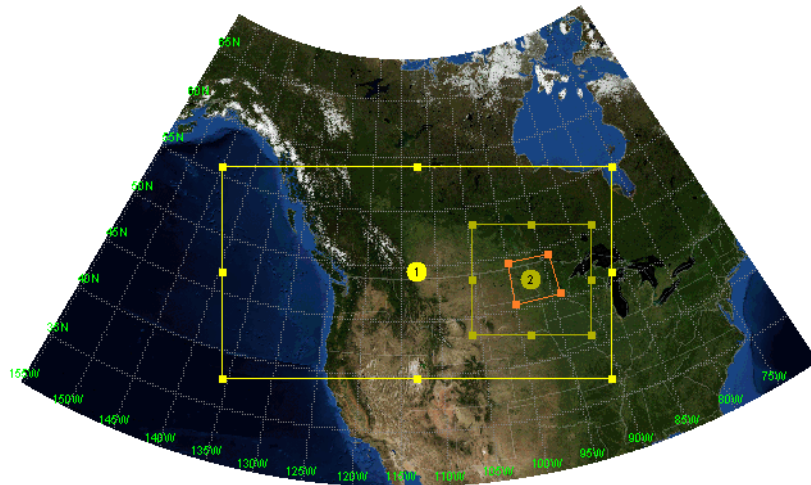
Figure 3. The 32 and 12 km domains, Domains 1 and 2, are outlined in yellow. The forecast area is outlined in orange (see Fig. 5 for details).

Figure 3 depicts the outer and inner domains of the WRF forecasts surrounding a third innermost forecast area (hereby referred to as Domain 1, Domain 2, and the forecast area, respectively). Because of this study's focus on the Northern Great Plains, Domain 1 (i.e. the outer domain) covers southwest and south central Canada and the northwest and north central contiguous United States. Domain 1 must capture incoming and upper level flow, but this large, coarse domain is also limited by the bounds of the initialization model domains (discussed later in detail). Domain 2 is nested (one-way nesting) within Domain 1, allowing for higher detail for forecasts over eastern North Dakota and western Minnesota. Domain 2 covers southeast Saskatchewan, southern Manitoba, southwest Ontario, North Dakota, South Dakota, Minnesota, northern Iowa, and northwest Wisconsin. The spatial grid spacing is 36 km for Domain 1 and 12 km for Domain 2. The majority of the results in this study are focused on the forecast area. Figure 4 provides details regarding the specific bounds of the forecast area (Fig. 5), which consists of eastern North Dakota and northwest Minnesota.

Boundary and initial conditions for the WRF ensemble member forecasts are derived every three hours from one of two operational model datasets: the Global Forecast System (GFS) and North American Model (NAM). These model datasets are provided by the National Center for Environmental Prediction (NCEP) and are available from the National Oceanic and Atmospheric Administration (NOAA) through the National Operational Model Archive and Distribution System (NOMADS). The NCEP GFS and NAM runs occur at 00, 06, 12, and 18 UTC, and are initialized using a broad set of satellite, rawinsonde, and surface station observation data. The GFS, also known as the

GFS Aviation (AVN) model, is a global spectral model that replaced the NCEP's
Medium Range Forecast (MRF) model in 2002 (EMC 2003). GFS-AVN 003 data were
used in this study, and they have 27 vertical levels with a grid spacing of 1° by 1°
(latitude by longitude). The NAM, or WRF-NMM, replaced the Eta model in 2006 as
NCEP's regional short-term model (Rogers et al. 2009). NAM 218 data were used in this
study and have 40 vertical levels with a grid spacing of 12 km by 12 km. Both models
have horizontal grid spacing appropriate for exploring Alberta clipper cases, given the
computing power available for this study.
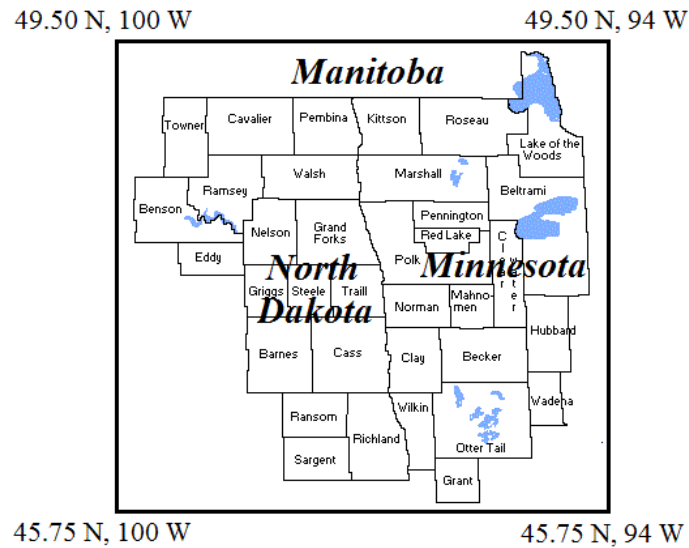
Figure 4. The forecast area is comprised of eastern North Dakota and northwestern
Minnesota (i.e., the National Weather Service Grand Forks office forecast area). This
area will be the primary focus for verification and analysis.
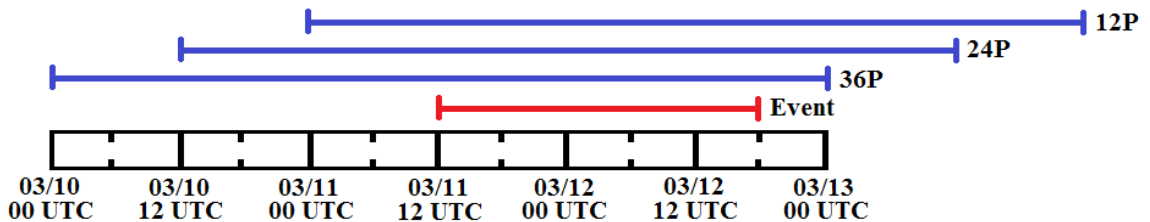
Figure 5. Timeline of ensemble forecasts for the March 2011 case relative to the event
(red) for the earliest (bottom blue; 36P), middle (middle blue; 24P), and latest (top blue;
12P) ensemble forecasts.

One goal of this study is to explore the use of different initialization times to capture temporal uncertainty in the initial intensification of a cyclone. Thomas and Martin (2007) found that clippers' quick intensifications within the first 60 hours post-departure from the Canadian Rockies is a source of uncertainty that affects the accuracy of simulated clippers' spatial and temporal characteristics. Additionally, Deppe et al. (2013) commented that differing time initializations produced a compromise in increased ensemble spread without degrading the mean absolute error. For this study, 72-hour cold-start model forecasts are initialized 12, 24, and 36 hours prior to the initial time that the cyclone impacts the forecast area (Fig. 4). The three model forecasts for each case are hereafter referred to as the 12P, 24P, and 36P forecasts, respectively. The start time for each event is defined as the 00 UTC or 12 UTC time prior to when the cyclone center enters the western or northern bounds of the forecast area. The end time for each event is defined as the 00 UTC or 12 UTC time after the cyclone center exits the eastern bounds of the forecast area.

Another goal of this study is to investigate the ensemble performance using the NAM or GFS forcing with certain combinations of microphysics, land surface, and longwave and shortwave radiation parameterizations. These combinations result in ten ensemble members for each initialization time, which will hereafter be referred to by their abbreviated names from Table 1. The member naming convention was derived from the NAM (Nam; green or cyan) or GFS (Gfs; purple or pink) forcing, NOAH (Ln; green or purple) or RAP/RUC (Lr; cyan or pink) land surface, RRTMG (Rr; solid line) or GFDL (Rg; dashed line) radiation, and WSM6 (no letters; no markers) or Thompson (Mt; black markers) microphysics schemes.

15

Table 1. Descriptions of the forcing and parameterizations used for each ensemble member. The ensemble member name is listed in the first column with its forcing and parameterization schemes along each row. The corresponding forcing, microphysics, land surface, and radiation schemes are listed by name in the subsequent columns.

Member	Forcing	Microphysics	Land Surface	Longwave and Shortwave Radiation	Line Color
NamLnRr	NAM	WSM6	NOAH	RRTMG	Green solid
NamLnRg	NAM	WSM6	NOAH	GFDL	Green dashed
NamLrRr	NAM	WSM6	RAP/RUC	RRTMG	Cyan solid
NamLrRg	NAM	WSM6	RAP/RUC	GFDL	Cyan dashed
GfsLnRr	GFS	WSM6	NOAH	RRTMG	Purple solid
GfsLnRg	GFS	WSM6	NOAH	GFDL	Purple dashed
GfsLrRr	GFS	WSM6	RAP/RUC	RRTMG	Pink solid
GfsLrRg	GFS	WSM6	RAP/RUC	GFDL	Pink dashed
NamLnRrMt	NAM	Thompson	NOAH	RRTMG	Green with markers
GfsLnRrMt	GFS	Thompson	NOAH	RRTMG	Purple with markers

Similarly to Michaelis and Lackmann (2013), all ensemble members constructed in this study employ the Monin-Obukhov (old MM5) surface layer scheme, Yonsei University (YSU) planetary boundary layer (PBL) scheme, and the Kain-Fritsch convective scheme. The old MM5 surface layer scheme applies Monin-Obukhov similarity theory and four stability functions to calculate its friction velocities and exchange coefficients (Zhang and Anthes 1982, Skamarock et al. 2008). This surface layer scheme was designed to be paired with either the YSU or Medium Range Forecast (MRF) PBL scheme. The YSU PBL scheme produces the PBL structure while using counter-gradient mixing and a parabolic K profile for the unstable mixed layer, an explicit entrainment layer, and a PBL top dependent on the buoyancy profile (Hong et al. 2006, Skamarock et al. 2008).The Kain-Fritsch cumulus parameterization (Kain 2004, Kain and Fritsch 1993) scheme is a mass flux scheme allows for deep and shallow

convection, includes cloud, rain, ice, and snow detrainment, and has a CAPE removal closure assumption.

The microphysics scheme is also important in parameterizing water, cloud, and precipitation processes. This study employs the WRF Single-Moment 6-class (WSM6; Hong and Lim 2006) and Thompson microphysics (Thompson et al. 2008). Both schemes incorporate ice-phase and mixed-phase processes, but Thompson microphysics allows for double-moment cloud ice. Additionally, Thompson microphysics has snow size distribution dependent on both ice water content and temperature, instead of just ice water content (WSM6 scheme).

While microphysics and cumulus schemes are essential to convective and large-scale precipitation parameterization, Yuan et al. (2012) found that LSM and radiation schemes were the dominant uncertainties in predicting characteristics of simulated winter precipitation forecasts. They also noted that the most accurate physics combinations were location-specific. While the members with a combination of the Rapid Update Cycle LSM (RUC; Smirnova et al. 1997, 2000) and updated Rapid Radiation Transfer Model (RRTMG; Iacono et al. 2008, Mlawer et al. 1997) radiation physics performed best in general during their study, the best member utilized the RRTMG radiation, Grell cumulus (Grell 1993, Grell and Dévényi 2002), Noah LSM (Ek et al. 2003), and Thompson microphysics schemes. These findings support further study on the role of LSMs and radiation schemes in winter precipitation forecasts, which is explored in this study.

LSMs approximate the canopy, ground surface, and soil layers and, in conjunction with the parameterized forcings from other physics schemes, derive heat and moisture fluxes. The LSMs selected in this study are the Noah LSM and the RUC LSM. The Noah

17

LSM is similar to that used in the NAM and accounts for four soil layers and one fractional snow cover layer. It solves energy and moisture equations explicitly for the boundary-layer scheme, while the RUC LSM solves them implicitly with a layer approach. The RUC LSM accounts for six soil layers and multi-layer snow and allows ice and water to more deeply penetrate the soil than the Noah LSM.

This study uses the RRTMG and the Eta Geophysical Fluid Dynamics Laboratory (GFDL; Schwarzkopf and Fels 1991) longwave and shortwave radiation schemes to calculate heating of the surface and the atmosphere. Both schemes account for effects of atmospheric water vapor, ozone, and carbon dioxide and have random cloud overlap.

3.2 Ensemble Forecast Verification

This study's analysis consists of a brief synoptic overview of each case and verification for individual ensemble members and the ensemble mean. Forecast verification focuses on the simulated cyclone track, 2 m winds, and precipitation, and compares their simulated values to observed values across the forecast area using designated threshold values. The ensemble output fields investigated in this study describe the surface cyclone's intensification, moisture, and winds that are crucial to operational winter storm forecasts.

3.2.1 Synoptic Overview

Background synoptic information regarding each case is necessary for a thorough analysis of the observed and ensemble forecast fields. Therefore, a best estimate of the state of the atmosphere during each case is provided by the NOAA Storm Prediction Center (SPC) Mesoscale Analysis Archive (Storm Prediction Center 2005). These diagnostic fields are a combination of a two-pass Barnes surface objective analysis and

18

RUC model forecast data upper-air data, resulting in a detailed atmospheric analyses of events.

3.2.2 Sea Level Pressure (SLP) Track

For cyclone track verification, the locations of simulated (WRF) and observed surface cyclone centers are compared within the forecast area. The hourly verification SLP data are provided by the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis dataset. MERRA data focus on analyses of the hydrological cycle for a broad range of weather and climate time scales (Lucchesi et al. 2012). It assimilates roughly 2 x 10^6 surface and remote sensing observations for each six-hourly analysis cycle. This allows for a multitude of available products including temperature, moisture, wind, and chemistry fields. Available MERRA output fields range from hourly to monthly time scales from 1979 to present. For this study, the MERRA dataset incorporates 72 vertical levels with a grid spacing of 0.50˚ by 0.66˚ (latitude by longitude; Rienecker et al. 2011). Use of MERRA's gridded hourly data is beneficial, given the short temporal analysis period of this study's cases, but it should be noted that other reanalyses with hourly temporal resolution are also available (e.g., North American Regional Reanalysis).

SLP is a diagnostic variable that is calculated in the WRF model. However, to obtain SLP data from the MERRA dataset, it is derived from surface pressure, elevation, and virtual temperature using

$$p_0 = p_s \times e^{\frac{g_0 z}{R_d T_V}} \tag{1}$$

Equation (1) is a form of the hypsometric equation where p_0 is sea level pressure, p_s is surface pressure, g_0 is a gravitational constant (9.81 ms^{-2}), z is the terrain height, R_d is the dry air gas constant (287 Jkg^{-1}K^{-1}), and T_v is virtual temperature.

A simple automated tracking method calculates a surface cyclone track forecast for each member track forecast and the MERRA data by finding the latitude-longitude location of the simulated hourly SLP minima within Domain 2 for each forecast hour. This tracking method was effective for the cases in this study because it was important to capture the locations of "true values" of the hourly SLP minima forecast with the ensemble members. Additionally, the time period and area of the track forecast verification analysis were selected to only focus on the targeted cyclone within forecast area. During this process, another check excludes points on the west and east edges of Domain 2, which may not be truly forecast as a location on the edges and should not be included as part of a physical track. One should use caution if attempting to reapply this exact method to other cases or for operational purposes because an additional quality control check should be applied to ensure that the automated tracking method continuously follows the same surface cyclone center. (For further recommendations on this, see discussion in Chapter 5.)

Track verification consists of comparing the locations of the WRF ensemble member track forecasts to the MERRA track at each forecast hour (Fig. 6a). The absolute distance (in kilometers) between the simulated and MERRA cyclone tracks is used to quantify track error, as shown in Fig. 6b. Distance is calculated by finding the absolute value of the difference in latitude and longitude values between the simulated member (and mean) track and the MERRA track. Because points on the west and east edges of

Domain 2 are not included in the automated tracking method, they are excluded from the

absolute distance calculations as well.
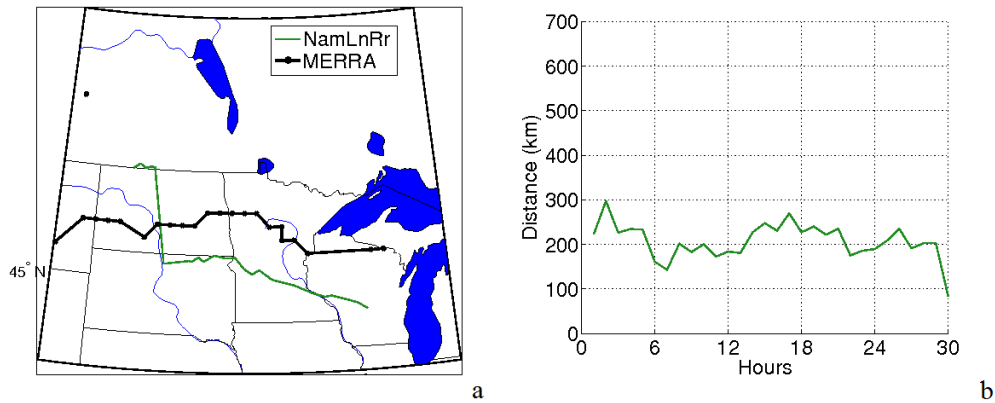
Figure 6. (a) The 24P NamLnRr cyclone track forecast (green) and the MERRA cyclone
track (black) across Domain 2 for the March 2011 case and (b) its track error (from the
MERRA track) in kilometers.

3.2.3 Precipitation

For verification analysis of the ensemble precipitation forecasts, this study utilizes

the NOAA National Operational Hydrologic Remote Sensing Center (NOHRSC) Snow

Data Assimilation System (SNODAS) solid precipitation data. SNODAS is a modeling

and data assimilation system that integrates observational data from satellite, airborne

platforms, and ground stations with model estimates of snow cover (Carroll et al. 2001).

The snow model has 1 km horizontal grid spacing and daily temporal grid spacing and is

forced by the NCEP RUC2 model and a multi-layer snow model (Barrett 2003). Model

output products include snow water equivalent (SWE), snow depth, and solid and liquid

precipitation. SNODAS products are only available for public download with a domain

of the conterminous United States, but all data products including a domain of the

conterminous United States and Canada can be viewed on the NOHRSC website's

interactive snow information viewer (NOHRSC 2015).

Snowfall and snow depth measurements and verification come with a significant

number of challenges, especially in a windy environment (e.g., Rasmussen et al. 2012,

Goodison et al. 1998, Yang et al. 1998). Hourly precipitation verification is typically

addressed using liquid equivalent precipitation observations because the current snowfall

observational network is spatially and temporally sparse compared to the standard

network of hourly precipitation observations from AWOS and ASOS data. Because of

these challenges associated with snowfall and snow depth observations, this study

primarily uses SNODAS daily solid precipitation data as a proxy for observations. Solid

precipitation data is defined as the total liquid water accumulation of falling snow

hydrometeors for a specified time period (e.g., hourly or daily). Hourly SNODAS solid

precipitation data is only available for the entirety of one of the three cases chosen for

this study. Thus, hourly precipitation verification will only be performed on the March

2011 case.

The primary method for precipitation verification for all three cases will focus on

event total solid precipitation amounts. Event total solid precipitation verification is

performed by comparing the sum of the daily SNODAS solid precipitation data values

during the event to the accumulated total grid scale snow and ice from the WRF

ensemble forecasts. It is important to note that the time period of the daily SNODAS data

is representative of 05 UTC Day 1 to 05 UTC Day 2. This time period of the daily

SNODAS data must be carefully compared to the ensemble forecasts so that all of the

snowfall for each event is captured in the SNODAS data. For the March 2011 case, the

event total period is from 05 UTC 11 March to 05 UTC 13 March 2011 for the 24P and

12P ensemble forecasts and the SNODAS data. However, the 36P ensemble forecasts

ending at 00 UTC 13 March 2011 because the ensemble forecasts are only 72 hours. For the December 2013 case, the event total period is from 05 UTC 28 December to 05 UTC 29 December 2013 for all three ensemble forecasts and the SNODAS data. For the January 2009 case, event total period is from 05 UTC 11 January to 05 UTC 13 January 2009 for the 36P and 24P ensemble forecasts and the SNODAS data. However, the 12P ensemble forecasts use an event total period starting seven hours later (12 UTC 11 January 2009).

Because of the discrepancy during for two of the forecasts listed above, additional analysis has been performed using the online Interactive Snow Information for hourly SNODAS snow precipitation data (NOHRSC 2015) and NOAA Storm Prediction Center (SPC) Mesoscale Analysis Archive base reflectivity mosaic (SPC 2005). For the 36P March 2011 ensemble precipitation forecasts, the SNODAS Interactive Snow Information shows only trace amounts of precipitation in isolated areas of the forecast area during this period, and the SPC Archive base reflectivity mosaic depicts weak isolated reflectivity (likely light snowfall) during this period. For the 12P January 2009 ensemble precipitation forecasts, the SNODAS Interactive Snow Information shows precipitation amounts ranging from a trace (much of the forecast area) to a maximum of 0.06 inches (in the far southeast corner of the forecast area). The SPC Archive base reflectivity mosaic depicts weak isolated reflectivity (likely light snowfall) during this period. Because precipitation accumulated during this period for both cases is not a product of either cyclone being studied and is light in comparison to the remainder of the event total precipitation, the effect on the results is minimal.

This study's precipitation verification consists of a three-fold approach, as follows: 1) probability of exceedance over a zoomed-in version of Domain 2 (hereafter referred to as Domain 2*), 2) coverage area exceeding a threshold value, and 3) grid point forecast verification. Figures 7a and 7b show examples of the probability of exceedance for threshold values of 0.01 inches of hourly precipitation and 0.10 inches of event total precipitation, respectively. Probability of exceedance analysis is valuable for exploring variability in the ensemble precipitation forecast locations and amounts and is the ratio of the number of ensemble members exceeding the threshold value to the total number of ensemble members (ten). It is noted that the area considered for Fig. 7 is Domain 2*, not Domain 2 or the forecast area. While the forecast area was too small to sufficiently analyze precipitation forecasts, the sharp gradient features depicted by the precipitation forecasts were difficult to observe given the larger area of Domain 2. Thus, Domain 2* is a compromise between those two areas.
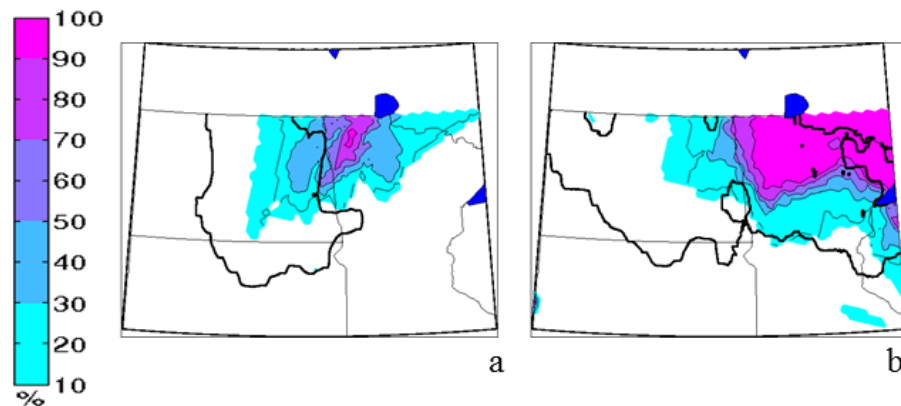
Figure 7. Percentage of 36P ensemble members (color fill) exceeding (a) 0.01 inches of hourly precipitation and (b) 0.10 inches of event total precipitation across Domain 2*. Area of SNODAS liquid equivalent snowfall is shown with the black outline.

Probability of exceedance analysis is also used to determine a representative threshold value for comparing the forecast and observed coverage areas (Figs. 8a and 8b).

24

In order to compare the coverage area of these differing datasets, the percentage of the total forecast area (in km^2; see area shown Figs. 3 and 4) exceeding the threshold value is calculated for each dataset during the event. Then, a difference in these percentages is determined for hourly (Fig. 8a) or event total precipitation (Fig. 8b) during the event. In this study, a positive (negative) value indicates that the forecast overestimates (underestimates) coverage.
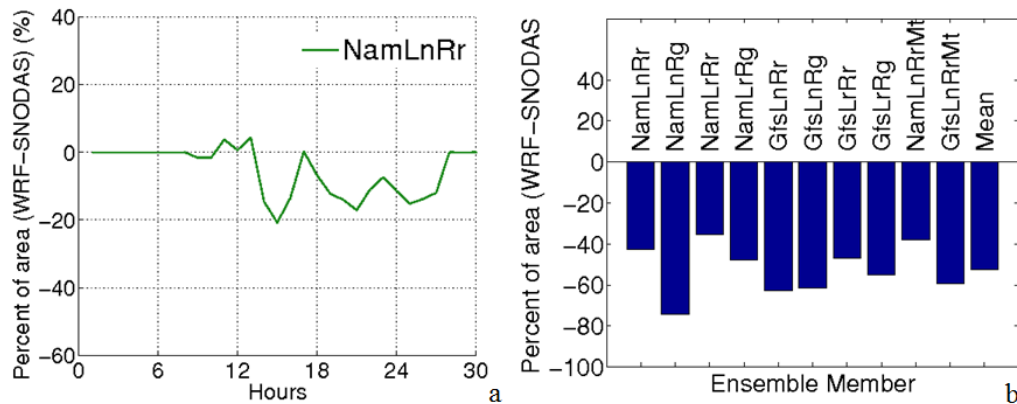
Figure 8. Difference in coverage area of (a) hourly precipitation exceeding 0.01 inches and (b) event total precipitation exceeding 0.10 inches for the 36P ensemble mean and member forecasts. Hourly and event total precipitation are for the forecast area (eastern North Dakota and northwestern Minnesota), not Domain 2.

Much of the ensemble verification is this study relies on threshold values to assess the quality, in terms of location, timing, and coverage area, of the ensemble precipitation. In choosing threshold values, it is important to remember that Alberta clippers tend to have the lowest moisture content compared to the two other common cyclone types impacting the Northern Great Plains (Mercer and Richman 2007). The threshold values of hourly liquid precipitation were determined by choosing values representative of this study's three Alberta clipper events' datasets: 0.01 inches (i.e., measurable) and 0.05 inches. The threshold values of event total precipitation chosen for verification were determined for this study by choosing four values of interest to operational

meteorologists: 0.10 inches, 0.20 inches, 0.25 inches, and 0.40 inches. The middle 50%

of snow liquid equivalent ratios fall between 0.09 and 0.17 inches of liquid precipitation

per one inch of snow (Baxter et al. 2005) and make for an easy comparison to the lowest

two threshold values. The greatest two threshold values are comparable to two to four

inches of snowfall and would be also useful in non-winter operations. The results from

these four event total precipitation threshold values are compared in Chapter 4 for the

March 2011 case results, and two primary threshold values are chosen to compare event

total precipitation for all three cases.

  While forecast quality over the forecast area is addressed with the probability of

exceedance and coverage area analyses, it is also important to assess the quality of a grid

point forecast. Grid point forecast verification is performed for the ensemble mean and

member precipitation forecasts against both the SNODAS solid precipitation and local

ASOS liquid precipitation observations, as shown in Fig. 9. Both the hourly and event

total accumulated precipitation are considered. NOAA and Department of Transportation

(DOT) Automated Surface Observing System (ASOS) and Automated Weather

Observing System (AWOS) jointly serve as the primary surface weather observing

network for the United States (NOAA 1998). ASOS and AWOS stations are managed by

the NOAA National Weather Service (NWS) and the Federal Aviation Administration

(FAA). These automated stations report hourly weather conditions for a variety of needs

(e.g., aviation and climatology). Weather conditions are reported at two meters above the

surface and include wind speed, direction, and gusts, precipitation, air temperature, and

dew point temperature. For this study, the Grand Forks, North Dakota ASOS station

26

(KGFK; 47.969 ˚N, -97.176 ˚W) provides a secondary source of verification for liquid precipitation.
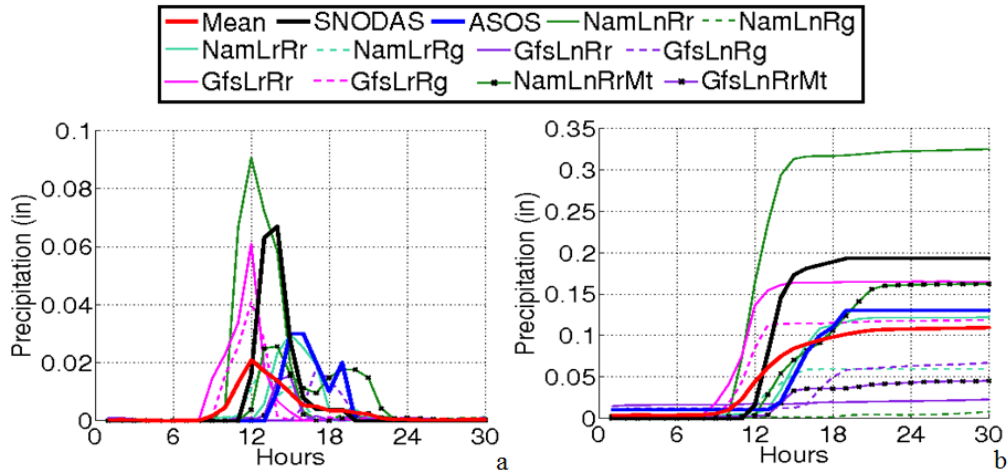
Figure 9. (a) Hourly and (b) accumulated precipitation for SNODAS (black), ASOS (blue), ensemble mean (red) and members' (remaining colors) forecasts at the Grand Forks International Airport (KGFK). Valid from 12 UTC 11 March 2011 to 00 UTC 13 March 2011 for the 36P ensemble forecasts. Note this time range is different from the other figures.

3.2.4 Wind Speed

For wind speed verification, the magnitudes of the WRF simulated 10 m winds are compared to the MERRA reanalysis 10 m wind and KGFK ASOS wind speed observations. The MERRA data products are described earlier in this chapter, but it is also important to note that while MERRA 10 m winds are time averaged, they are intended for comparison with surface meteorological stations (Lucchesi et al. 2012). Wind speed verification is performed similarly to the precipitation verification and consists of probability of exceedance over Domain 2*, coverage area exceeding a threshold value, and grid point forecast verification. Four operationally relevant threshold values are chosen for the initial probability of exceedance verification: 20 mph (17.4 kts), 25 mph (21.7 kts), 30 mph (26.1 kts), and 35 mph (30.4 kts). The results from these four

27

wind speed threshold values are compared in Chapter 4 for the March 2011 case results, and one threshold value is chosen to compare wind speeds for all three cases.

Grid point forecast verification is performed against both the MERRA winds and ASOS observations at the Grand Forks, ND site, as shown in Fig. 10. Because of this relatively coarse MERRA grid (0.5˚ by 0.67˚), it is unlikely that the MERRA data can resolve the magnitude of peak wind; instead these maximum values will likely be smoothed out, affecting the accuracy of the MERRA verification for any one grid point location. This smoothing effect is apparent when comparing the MERRA and ASOS data in Fig. 10, particularly at local minima and maxima.
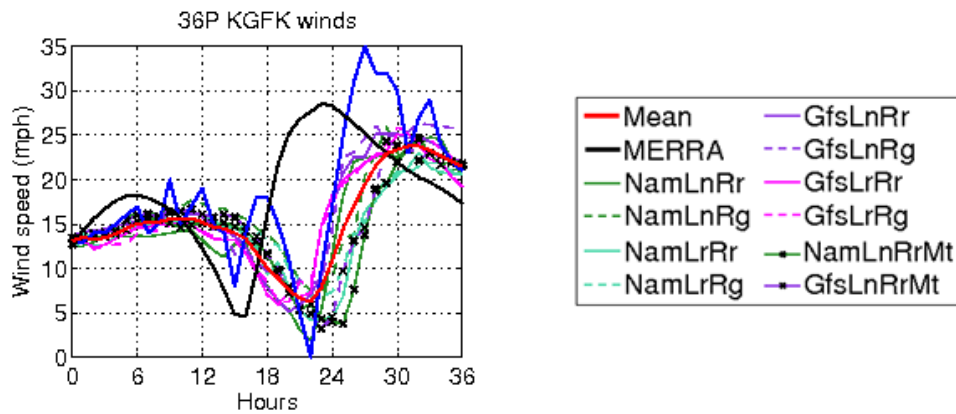
Figure 10. Hourly wind speeds for MERRA (black), ASOS (blue), ensemble mean (red) and members' (remaining colors) forecasts at the Grand Forks International Airport (KGFK). Valid from 12 UTC 11 March 2011 to 00 UTC 13 March 2011 for the 36P ensemble forecasts. Note this time range is different from the other figures.

3.3 Ensemble Evaluation

An assessment of the ensemble is also performed and consists of comparing the performance of the ensemble mean fields to the individual members and MERRA and SNODAS datasets, determining a best-performing member (i.e., preferable combination of parameterizations), investigating whether the truth is represented by the ensemble, and

identifying primary areas of uncertainty (e.g., initialization). The probabilistic skill of the ensemble as a whole cannot be evaluated due to the limited number of cases in this study, but case-dependent qualitative value and spread can be assessed, as in Novak and Colle (2012). At minimum, the member with the minimal error will lead to an optimal physics scheme combination for each case or for all of the cases. Ideally, the best ensemble member would have minimal error and would represent the observed state of the track, winds, and precipitation. However, a best-performing member may not be apparent (as discussed in Chapter 2).

While mean field calculations are not a physical representation of the respective fields, they do provide insight into the quality of the ensemble forecasts. The ensemble mean calculation varies depending upon the model field. The ensemble mean track forecasts are calculated as an average of the values of the members' latitude and longitude locations at each time. However, the mean values of the absolute error in the ensemble track forecasts are calculated as an average of the absolute error values each hour. A similar method is applied to obtain the mean hourly and event total differences in coverage area of snow and wind. Ensemble mean surface station wind speed and precipitation forecasts are calculated as an average of the members' forecast values.

CHAPTER IV

RESULTS AND DISCUSSION: March 2011 Case

In this chapter, the best ensemble forecast and members are determined for the March 2011 case. First, a synoptic overview that describes the general atmospheric pattern, features of interest, and pertinent impacts to the forecast area using figures from NOAA SPC mesoscale analysis (SPC 2005) is provided. Next, results of the ensemble forecast verification for the cyclone track, precipitation, and surface wind speed are presented. Note that due the large volume of verification plots, several appendices are referred to in this chapter. This was done to maintain readability while also providing the full set of verification results. The verification results are then summarized before concluding with discussion of the overall ensemble performance and its limitations for the March 2011 case. Assessment of this case and additional cases presented in Chapter 5 in the context of previous studies, along with recommendations for future work, is provided in Chapter 6.

4.1 Synoptic Overview

During the evening of 10 March 2011, an Alberta clipper departed southeast from the Canadian Rockies. Figures 11-13 depict the development of the clipper for the 300 hPa, 500 hPa, 850 hPa, and the surface levels at 0000 UTC 11, 12, and 13 March 2011, respectively. For a timeline of the ensemble forecasts compared to the event duration for the March 2011 case, refer to Fig. 5. The initial synoptic pattern over North America

consisted of a developing 500 hPa shortwave trough over the Pacific Northwest, a

building positively tilted ridge over the Plains, and a deep negatively tilted trough with an

axis stretching from western Ontario to South Carolina (Figs. 11b and 12b). The

dominant downstream trough matured into an upper level closed-off cyclone over the

eastern Great Lakes by 0000 UTC 12 March 2011 (Fig. 12b).
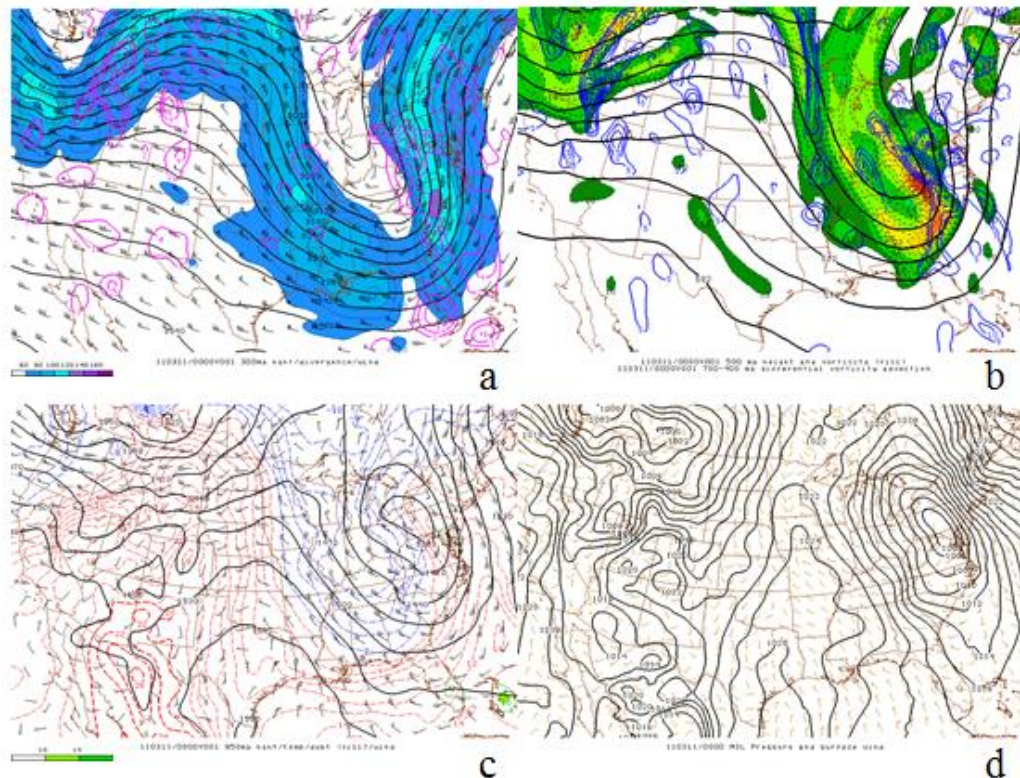
Figure 11. NOAA Storm Prediction Center (SPC) mesoscale analysis at 0000 UTC 11 March 2011, just after the cyclone's departure from the Rockies. This analysis depicts: (a) 300 hPa heights (black contours), ageostrophic wind (barbs), and 700-500 hPa layer-averaged omega (magenta-up/red-down), (b) 500 hPa heights (black contours) and 700-400 hPa differential vorticity advection (blue contours), (c) 850 hPa heights (black contours), temperature (red and blue contours), and wind (barbs), and (d) mean sea level (MSL) pressure (black contours) and surface wind (barbs). The color fills show (a) 300 hPa isotachs (i.e., the jet stream), (b) 500 hPa relative vorticity, and (c) 850 hPa dew point temperature.

During 11 March 2011, a 300 hPa jet streak developed over Wyoming, Montana,

and the western Dakotas along the aforementioned shortwave trough (Figs. 11a and 12a).

At 500 hPa, the dominant downstream trough matured into an upper level closed-off

cyclone over the eastern Great Lakes, and the 500 hPa shortwave trough deepened over

North Dakota as its associated relative vorticity maximum intensified (Figs. 11b and

12b). The focused intensification of the 300 hPa jet streak and 500 hPa shortwave trough

prompted a shift in location of the primary center of the surface cyclone from southeast

Saskatchewan to western North Dakota around 09 UTC 11 March 2011 (Figs. 11d and

12d). The clipper's surface cyclone deepened as it propagated southeastward into west

central North Dakota. At the surface, preexisting snow cover was present and wind

speeds increased throughout the day as the pressure gradient tightened. A forecast

discussion from NWS Grand Forks suggests that a combination of these factors and

anticipated additional precipitation prompted forecasters to issue a blizzard warning for

eastern North Dakota and northwestern Minnesota (NWS Grand Forks 2011).

By 0000 UTC 12 March 2011, the strengthening clipper had propagated eastward

into northern Minnesota (Fig. 12d). A strong jet streak peaked at over 115 mph (100 kts)

in South Dakota (Fig. 12a) with a strong relative vorticity maximum and shortwave

trough at 500 hPa (Fig. 12b) and the surface cold front was draped along the North

Dakota/Minnesota border (Fig. 12c). Upstream of the clipper, a surface high pressure

system began to form, and the pressure gradient strengthened, resulting in a north-

northwest surface wind at 35-40 mph (30-35 kts) and gusts up to 63 mph (55 kts) (Fig.

12d). Increased surface winds and narrow post-frontal north-south snow bands

propagated into eastern North Dakota. Prior to the cold frontal passage that afternoon,

warm air advection allowed for high temperatures nearing 10 ˚C, so many people were

caught off guard by the rapidly deteriorating conditions (NWS Grand Forks 2011).

Blizzard conditions occurred across the warned area during the overnight hours as the clipper continued with an east-southeast track through northern Minnesota and came into phase with the stacked downstream cyclone (Fig. 13).
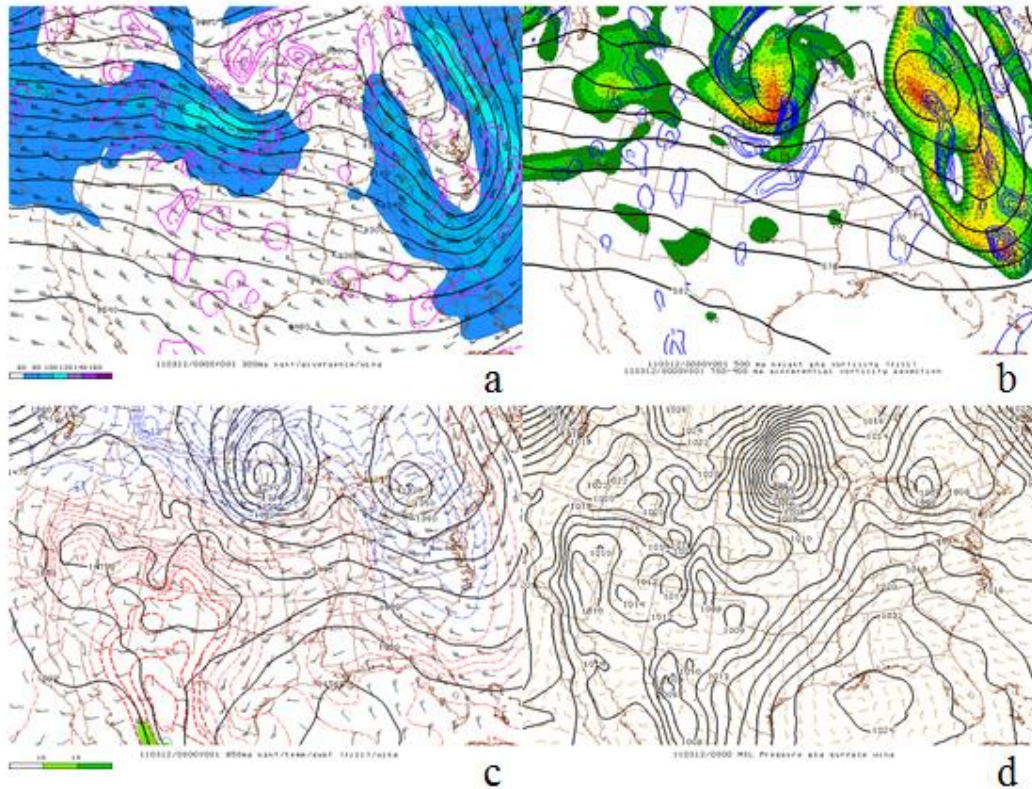
Figure 12. As in Fig. 11 but at 0000 UTC 12 March 2011.

By 0000 UTC 13 March 2011, the clipper had weakened and had come into phase with the downstream mature cyclone (Fig. 13). With the exception of the closed off upper level cyclone in eastern Canada, upper and mid-level flow over the United States was primarily zonal (Figs. 13a-c). A strong surface high pressure system trailed the clipper of interest and entered the Northern Great Plains, bringing continued cold air advection (Fig. 13c) and northerly winds (Fig. 13d). Through the morning of 13 March, blizzard conditions began to subside from west to east as the surface pressure gradient weakened.

Event total snowfall across the forecast area ranged from one to four inches (NWS Grand
Forks 2011).
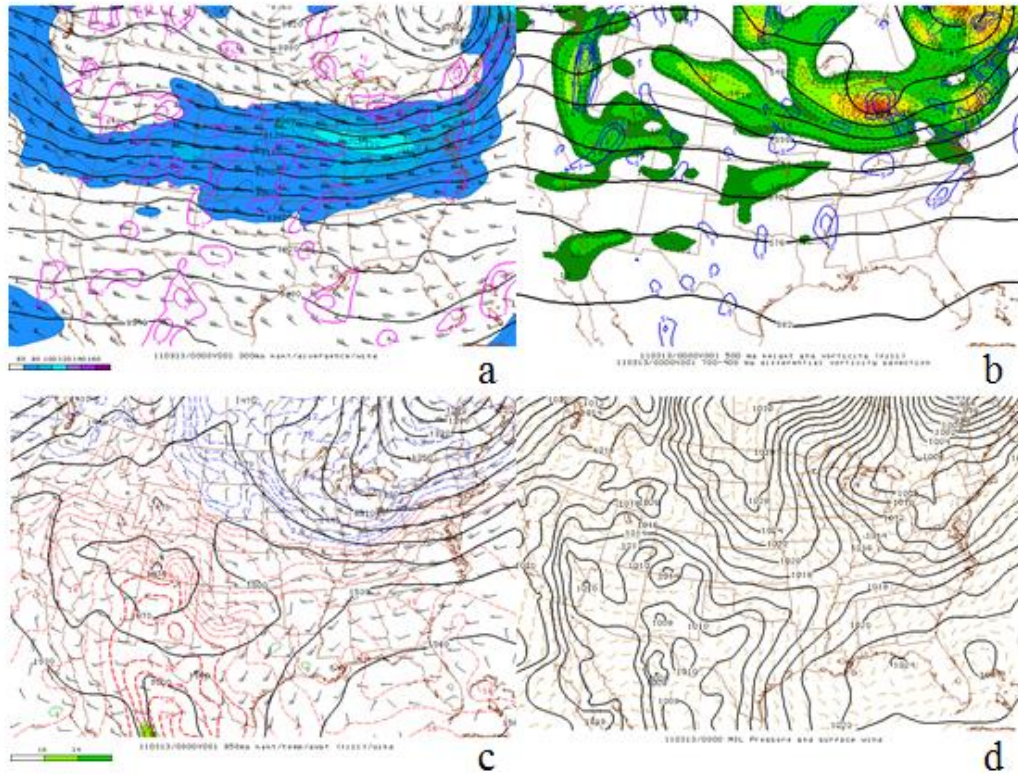
Figure 13. As in Fig. 11 but at 0000 UTC 13 March 2011.

4.2 Ensemble Forecast Verification

Forecast verification of the March 2011 cyclone track, two-meter wind speed, and
precipitation is presented in this section. First, the ensemble mean and member cyclone
tracks are spatially and temporally compared to the MERRA observed track. Following
the cyclone track verification, precipitation and wind speed ensemble forecasts are
verified using multiple techniques: comparison of the area covered given a threshold
value, probability of exceedance, and time series at a surface station location.

4.2.1 Ensemble Mean Track Verification

Figure 14a shows the MERRA and ensemble mean cyclone center track forecasts in Domain 2 throughout the March 2011 event. All three ensemble mean track forecasts depict a surface cyclone starting in Saskatchewan and progressing eastward into Manitoba. Then the cyclone propagates southeastward from Manitoba into the United States. In contrast, the SPC mesoanalysis and MERRA tracks portray an initial surface cyclone center in western North Dakota with propagation eastward across central North Dakota and northern Minnesota (Figs. 11d, 12d, and 14a). By 12 UTC 12 March 2011, all three forecasts agree with the observed track's eastward propagation.

Figure 14b shows the absolute error in distance from the MERRA track for each of the ensemble mean forecasts. The greatest absolute difference in distance is nearly 500 km and occurs at 00 UTC 11 March 2011 (i.e., the beginning of the event) for all three ensemble mean track forecasts. All three ensemble mean track forecasts are to the north of the observed track from 00 UTC 11 March 2011 until around 21 UTC 11 March 2011 (Fig. 14a). Until 21 UTC 11 March 2011, all three ensemble mean track forecasts trend toward the observed track. Figure 14b indicates this trend with decreasing distances. After 21 UTC 11 March 2011, the 36P track forecast depicts the most southern cyclone track bias (Fig. 14a), while the 24P track forecast depicts the most northern track bias.

The 36P and 12P ensemble mean track forecasts consistently perform best after 21 UTC 11 March 2011 (Fig. 14b). In fact, both forecasts show an absolute difference in distance from the observed track $\leq \sim 100$ km. Out of these two forecasts, the 12P ensemble mean track forecast generally performs best from 00 UTC to 19 UTC 11 March 2011, and the 36P ensemble mean track forecast generally performs best from 19 UTC 11

35

March 2011 to 12 UTC 12 March 2011. The 24P ensemble mean track forecast remains

far from the observed track, barely within 100 km of the observed track during its best

performance.
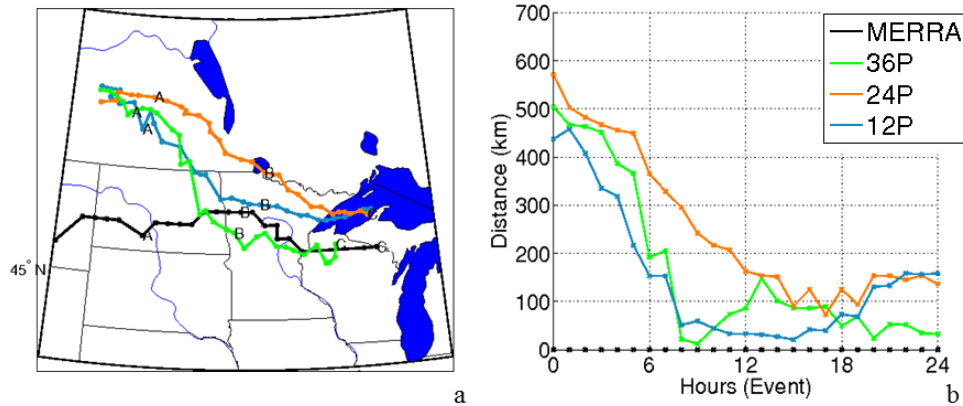
Figure 14. (a) Surface cyclone center track forecasts across Domain 2 and (b) the absolute error (distance in km) in the track forecasts for the 12P (green), 24P (orange), and 36P (blue) ensemble means. Hours 12 UTC 11 March 2011, 00 UTC 12 March 2011, and 12 UTC 12 March 2011 are indicated by A, B, and C, respectively. (a) represents 06 UTC 11 March to 12 UTC 11 March 2011 and (b) represents 12 UTC 11 March to 12 UTC 12 March 2011.

4.2.2 Ensemble Member Track Verification

Figures 15a, 15c, and 15e show ensemble means' and members' track forecasts

throughout the March 2011 event at each of the three initialization times . The majority of

the individual ensemble member cyclone track forecasts show cyclone propagation from

southern Manitoba to northern or central Wisconsin. From 00 UTC to 12 UTC 11 March

2011, most of the track forecasts are in close proximity to one another (Figs. 15a, 15c,

and 15e).

The track forecasts differ substantially after 12 UTC 11 March 2011 (Figs. 15a,

15c, and 15e). The majority of the 36P track forecasts begin to indicate a bias to the south

between 12 UTC 11 March 2011 and 00 UTC 12 March 2011. The 36P track forecasts

indicate a cyclone propagating south from southern Manitoba to far northeast South

36

Dakota, continuing southeastward through central or southern Minnesota and ending in southern Wisconsin (Fig. 15a). The 24P track forecasts are the most northern of the three forecasts and show a southeastward track through southern Manitoba, continuing along the Minnesota-Canadian border, and ending over Lake Superior (Fig. 15b). Most of the 12P track forecasts are far north of the observed track, depicting a southeastward track from southern Manitoba into east central North Dakota, followed by a generally eastward track through northern Minnesota and ending over Lake Superior (Fig. 15c).

Figures 15b, 15d, and 15e show the absolute error in distance from the MERRA track relative to the event for each of the ensemble mean forecasts. With respect to the individual ensemble member track forecasts, the best performing forecast overall is 12P. Most of the 12P individual ensemble member track forecasts perform almost equal to or better than the 12P ensemble mean track forecast throughout the event (Fig. 15g). Additionally, the 12P absolute errors in the track forecast are the least for the longest portion of the event, compared to the 24P and 36P forecasts (Figs. 15b, 15d, and 15g). While the 36P ensemble mean track forecast has the least absolute error under 100 km for the longest period during the event (Fig. 15b), the 12P track forecast is under 50 km for the longest period during the event and has many members with absolute error less than 200 km for much of the event (Fig. 15g).

In relation to the ensemble members' physics schemes, Figs. 15a, 15c, and 15e show most members with a NOAH land surface scheme and RRTMG radiation scheme producing the most southern cyclone tracks for each forecast. In fact, some of these members (e.g., 36P and 24P NamLnRr, 12P GfsLnRr) had some of the noticeably better performing individual ensemble member track forecasts. Members with the Thompson

microphysics, RRTMG radiation, or RUC LSM schemes tended to perform better than the WSM6 microphysics, GFDL radiation, or NOAH LSM schemes from 00 UTC to 12 UTC 12 March 2011 and vice versa during the remainder of the event (Figs. 15b, 15d, and 15g). The model forcing (NAM or GFS) did not appear to show a significant bias in track forecast performance throughout the event.
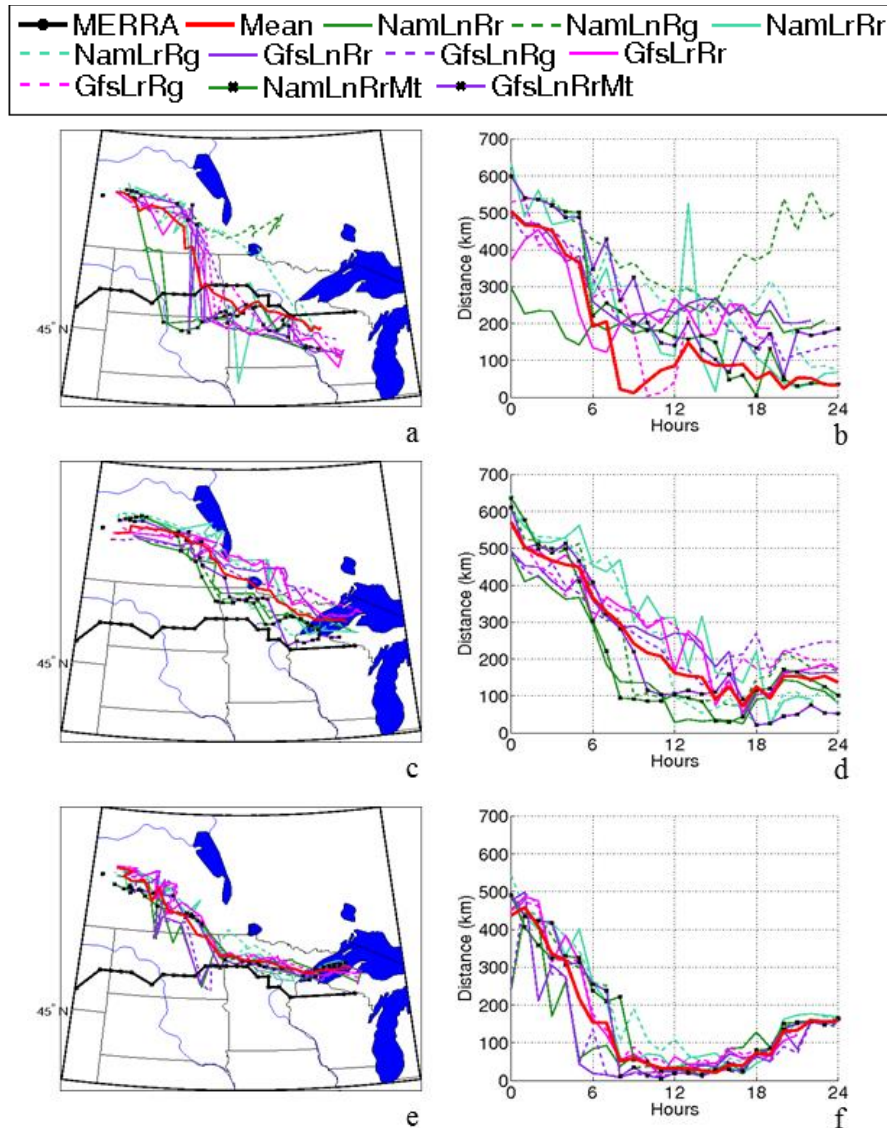
Figure 15. Surface cyclone center tracks across Domain 2 for the 36P (a), 24P (c), and 12P (e) ensemble members and the absolute error in the 36P (b), 24P (d), and 12P (f) ensemble track forecasts. Tracks from MERRA (black bold), ensemble mean (red), and ensemble members' forecasts (remaining colors). (a), (c) and (e) represent 06 UTC 11 March to 12 UTC 11 March 2011. (b), (d), and (f) represent 12 UTC 11 March to 12

UTC 12 March 2011.For more information on the member naming and color conventions, refer to Table 1.

4.2.3 Precipitation Threshold Verification

Appendices A and B depict the variability in the probability of exceeding 0.01 and 0.05 inches (respectively) of hourly precipitation for the 36P, 24P, and 12P ensemble member forecasts and SNODAS data. Typical snow-to-liquid ratios would equate these values to snowfall amounts of 0.1 to 0.5 inches. Because hourly precipitation values exceeding 0.05 inches are very limited in the ensemble forecasts during much of the event (Appendix B), an hourly precipitation threshold value of 0.01 inches will be the primary focus for the hourly precipitation analysis. While this threshold value is quite small, it is representative of whether snow was observed or forecast in a measurable quantity. For ensemble forecast analysis of greater threshold values, event total precipitation verification will be available later in this section.

The SNODAS data shows precipitation moving into eastern North Dakota around 22 UTC 11 March 2011, developing eastward as a light, broad swath, and exiting northwestern Minnesota by 13 UTC 12 March 2011 (Appendices A and B). The SNODAS dataset kept relatively light precipitation rates throughout the duration of the March 2011 case and has a consistent area of hourly precipitation totals exceeding 0.05 inches during the cyclone's progression through central North Dakota. The local automated surface observations of hourly precipitation across the area concur with the SNODAS timing and consistently indicated hourly precipitation totals of near or less than 0.05 inches with no hourly totals above 0.10 inches recorded (NCDC 2005).

Figures 16-18 represent three valid times throughout the 36P, 24P, and 12P forecasts and show the hourly probability of exceedance for 0.01 inches of hourly

precipitation for the three forecasts and SNODAS. All three ensemble forecasts start off

with poor band placement relative to the SNODAS data and continue to displace the

hourly precipitation coverage too far eastward compared to the SNODAS hourly

precipitaiton (Figs. 16-18 and Appendices A and B). When disregarding the displacement

errors in all three ensemble forecasts, the 12P ensemble forecast performs best for this

event considering the number of members depicting the location and intensity of

developing hourly precipitation (Appendix A).
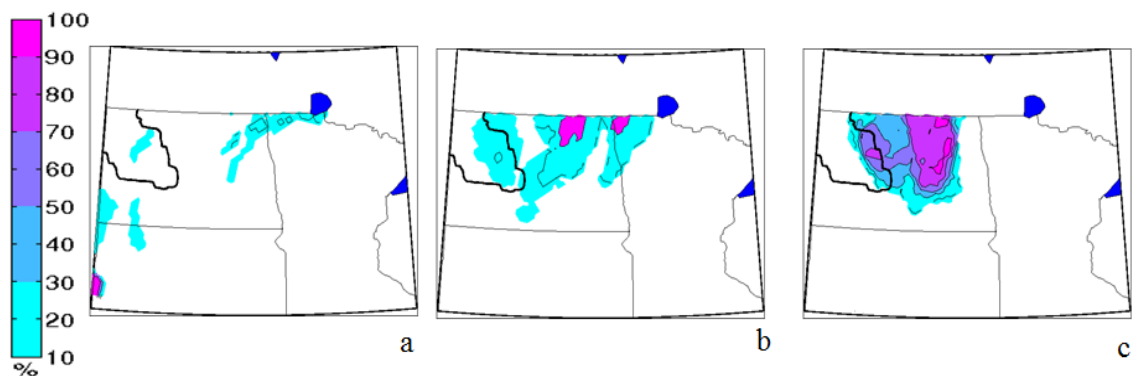
Figure 16. Probability of exceedance (color fill) at 0.01 inches of hourly precipitation
from the 36P (a), 24P (b), and 12P (c) ensemble forecasts for Domain 2* from 17 UTC to
18 UTC 11 March 2011. Coverage area of SNODAS hourly precipitation is outlined in
black.

Figure 16c shows 70-90% of the 12P ensemble member forecasts predict a

precipitation swath over central North Dakota at 18 UTC 11 March 2011, which is

similar in size to a large precipitation swath over northwest North Dakota from the

SNODAS data. Only 10-30% of the 36P and 24P member forecasts predict this swath's

existance and both greatly underpredict its size (Figs. 16a and 16b).

By 00 UTC 12 March 2011, this widespread swath of precipitation has

propagated eastward and into portions of eastern North Dakota (Fig. 17). Only 10-30% of

all three ensemble member forecasts predict a widespread swath of precipitation, albeit

predicting the swath's location too far eastward. The 12P ensemble member forecasts

perform best at this time with at least 50% of members predicting a precipitation band

close to the observed widespread swath (Fig. 17c).
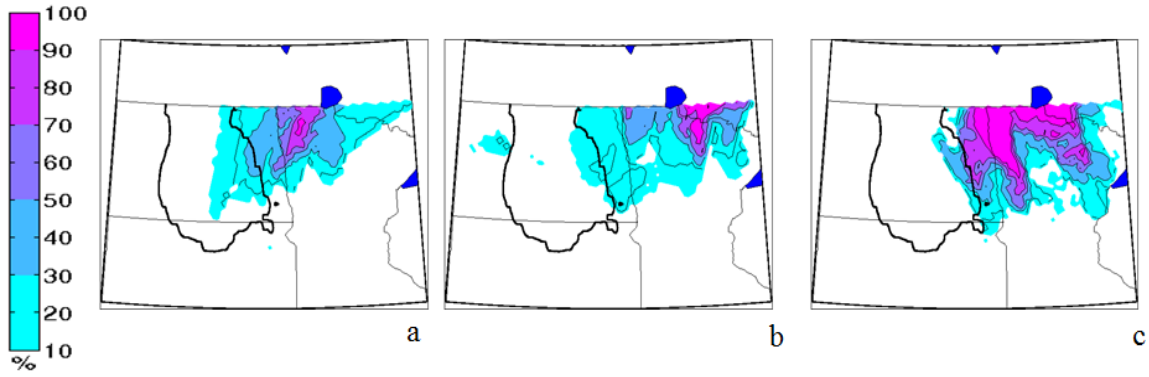
Figure 17. Same as Fig. 16 but from 23 UTC 11 March to 00 UTC 12 March 2011.

While the 12P hourly precipitation forecasts perform best during the beginning of

the event, the 36P hourly precipitation forecasts improve in performance by the second

half of the event. By 06 UTC 12 March, the 36P hourly precipitation forecasts perform

best with 50-70% of ensemble members predicting a broad precipitation swath over

northern Minnesota with a very similar location and size to the SNODAS data (Fig. 18).

The 12P and 24P forecasts have 50-70% and 10-30%, respectively, of ensemble members

predicting a similarly-sized swath to the SNODAS data, but these swaths are displaced

too far east. By 12 UTC 12 March, all three forecasts predict a swath of hourly

precipitation displaced too far eastward of the SNODAS data (Appendix A).

The remainder of the precipitation verification results in this study will focus on

event total precipitation values. Fig. 19 depicts the variability in the probability of

exceeding event total precipitation threshold values of 0.10 inches, 0.20 inches, 0.25

inches, and 0.40 inches for the 12P ensemble forecasts and the SNODAS data. Using a

typical March snow to liquid ratio of 0.13 inches for the forecast area (Baxter et al.

2005), these thresholds correspond to approximately 0.77 inches, 1.54 inches, 1.92

inches, and 3.08 inches, respectively.
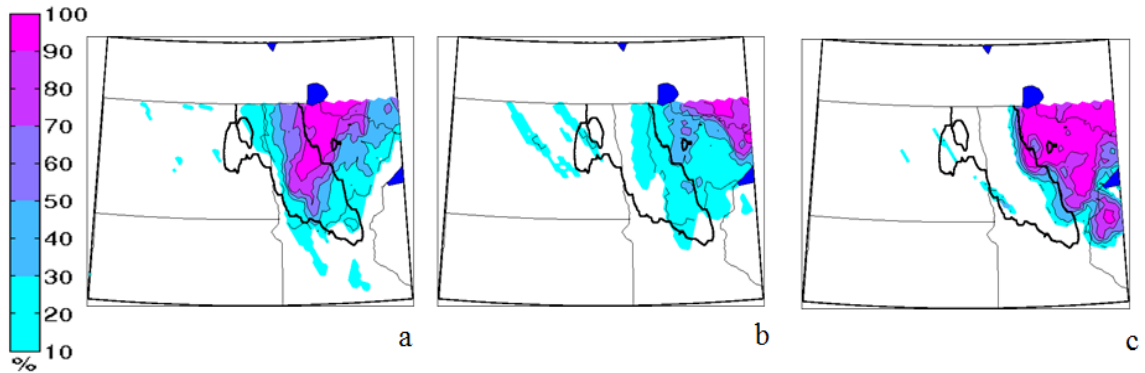
Figure 18. Same as Fig. 16 but from 05 UTC to 06 UTC 12 March 2011.

The SNODAS data shows a broad area of at least 0.25 inches of event total

precipitation across central North Dakota (Figs. 19b and 19c) and a widespread area of at

least 0.10 inches of event total precipitation across nearly all of North Dakota (excluding

southwest North Dakota) and northern Minnesota (Fig. 19a). Additionally, a small area in

central North Dakota received more than 0.40 inches of event total precipitation (Fig.

19d). The 12P ensemble member forecasts differ significantly in location and magnitude

from the SNODAS data. The majority of ensemble members predict event total liquid

precipitation exceeding 0.10 inches for much of northern Minnesota and northeast North

Dakota. However, the 12P ensemble member forecasts failed to predict higher snowfall

amounts in the correct locations. They predicted between 0.20 and 0.25 inches for

isolated locations along the Canadian border (Fig. 19), and most members (70% or

greater) did not forecast any event total precipitation greater than 0.40 inches (Fig. 19d).

Figure 20 shows the probability of exceeding event total precipitation of 0.10

inches for all three ensemble forecasts and the SNODAS data. For the variability amongst

the three forecasts with other three threshold values (0.20 inches, 0.25 inches, and 0.40

42

inches), refer to Appendix C. The 12P ensemble member forecasts perform best for the

0.10 inch threshold value (Fig. 20) but lack precipitation coverage in much of North

Dakota and central Minnesota. The 36P and 24P event total precipitation forecasts

perform even worse than the 12P forecasts with only 10-30% of their ensemble members

predicting event total precipitation exceeding 0.10 inches over North Dakota (Figs. 20b

and 20c). Overall, all three forecasts erroneously displace the heaviest precipitation bands

northeast or east of where they were observed and are unable to depict the area of

heaviest precipitation over central North Dakota (Appendix C).
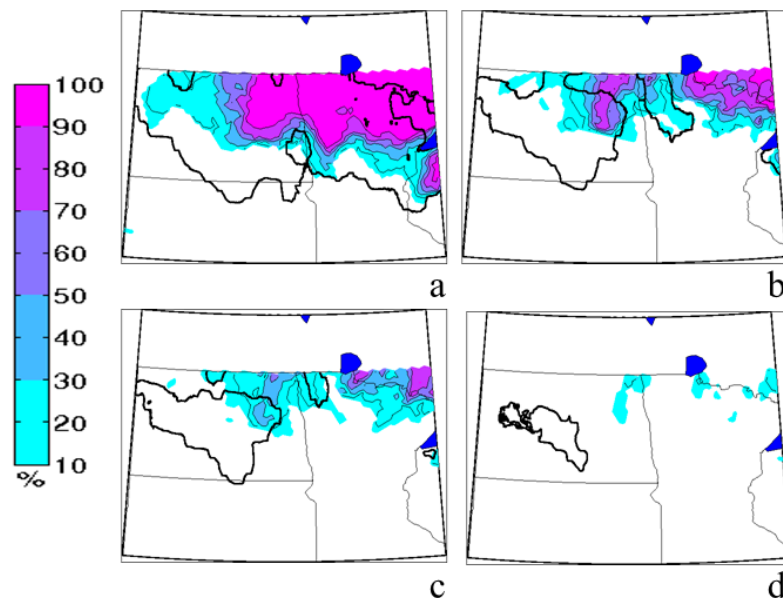
Figure 19. Percentage of 12P ensemble members (color fill) exceeding 0.10 (a), 0.20 (b), 0.25 (c), and 0.40 inches of (d) event total liquid equivalent snowfall across Domain 2*. Area of SNODAS liquid equivalent snowfall shown in black outline.

From examining Figs. 16-20 and Appendices A, B, and C, it is concluded that the

0.10 inch hourly precipitation threshold will best capture the variability of the ensemble

forecast precipitation throughout the event. The March 2011 case showed a signficant

difference between the SNODAS and ensemble forecast precipitation at this threshold.

As mentioned in previous chapters, Alberta clippers do not have a high moisture capacity

in comparison to cyclones originating from other locations. Thus, 0.10 inches of liquid

precipitation is the primary threshold used for comparison between the three forecasts for

the remainder of this section and study.
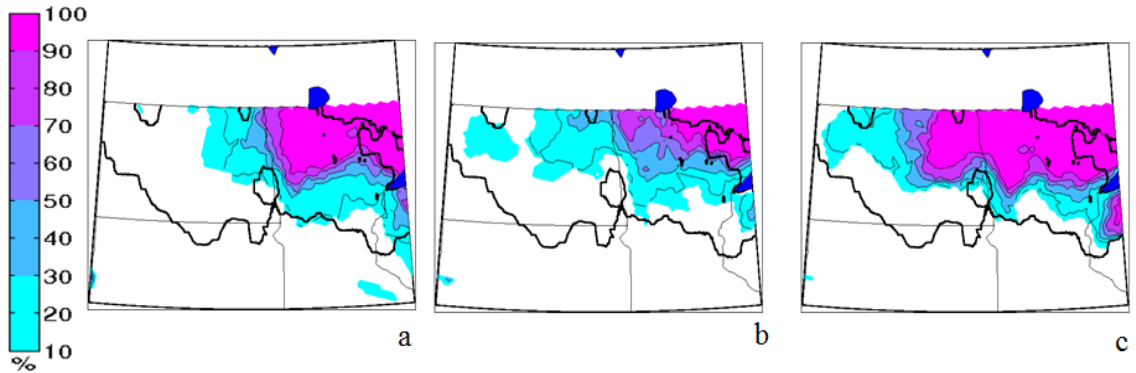
Figure 20. Percentage of members (color fill) exceeding 0.10 inches of event total liquid
equivalent snowfall across Domain 2* for the 36P (a), 24P (b), and 12P (c) ensemble
forecasts. Area of SNODAS liquid equivalent snowfall shown in black outline.

Fig. 21 provides the quantified difference (from the SNODAS coverage) in

coverage of hourly precipitation exceeding 0.01 inches for the 36P, 24P, and 12P

forecasts. The ensemble members with NAM forcing or RUC LSM tended to perform

better than the GFS forcing or NOAH LSM in all three precipitation coverage forecasts.

The NAM forcing predicted coverage is up to 20% closer to the observed (SNODAS)

coverage than the members with GFS forcing. The NamLnRrMt member was a top

performing member in all three forecasts. Signficant differences in performance based on

land surface model or cumulus parameterization schemes were not present.

While all three forecasts underestimate precipitation coverages (as negative

values in Fig. 21) throughout most of the event, the general trends, durations, and

magnitudes differ amongst the forecasts (Fig. 21). The 36P precipitation coverage

forecast performs best overall. Despite its poor performance around 04 UTC 12 March

(event hour 16), the 36P ensemble mean difference in coverage remains below 20% (Fig.

21a). The 24P precipitation coverage forecast performs worst, and its ensemble mean

difference in coverage is 20-40% below the SNODAS coverage for a significant portion

of the event (Fig. 21b). The 36P and 24P precipitation coverage forecasts perform better

than the 12P forecast from 18 UTC 11 March to 00 UTC 12 March (event hours 06 to

12), as most members remain within 20% of the observed coverage (Fig. 21). Although

the 12P forecast does not perform best during the beginning or end of the event, its best

performance is 10-20% better than the 36P and 24P forecasts from 00 UTC to 06 UTC 12

March (event hours 12 to 18), during the period of greatest uncertainty for all three

forecasts.
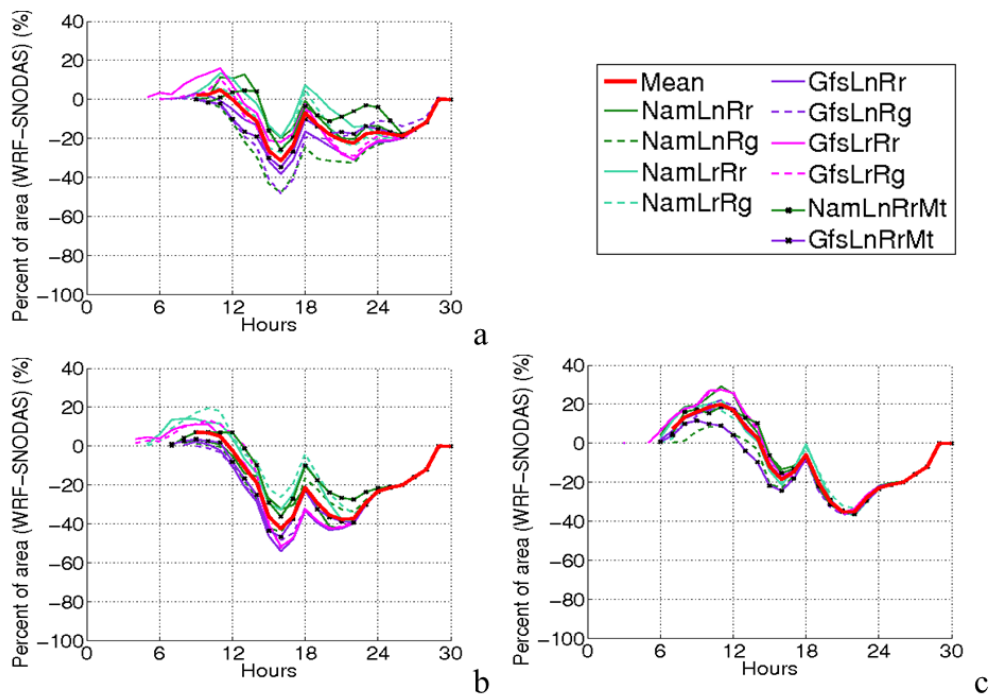
Figure 21. Difference in coverage area of hourly precipitation exceeding 0.01 inches for the (a) 36P, (b) 24P, and (c) 12P ensemble mean (red) and member (remaining colors) forecasts. Hourly precipitation is considered for the forecast area during the 30 hour event period of 12 UTC 11 March to 18 UTC 12 March. Note that the area considered is the forecast area (Fig. 5), not Domain 2.

Fig. 22 shows the difference in coverage area of event total precipitation exceeding 0.10 inches between each of the three ensemble forecasts and the SNODAS data. In looking at the event as a whole, all of the ensemble means and members underpredict the coverage area of event total precipitation exceeding 0.10 inches. Because Figs. 19 and 20 and Appendix C indicate that all of the ensemble means and members underpredict coverage area at higher threshold values of event total precipitation values (see outline of MERRA coverage area compared to member coverage), further analysis of the coverage area for event total precipitation is not necessary at higher thresholds.
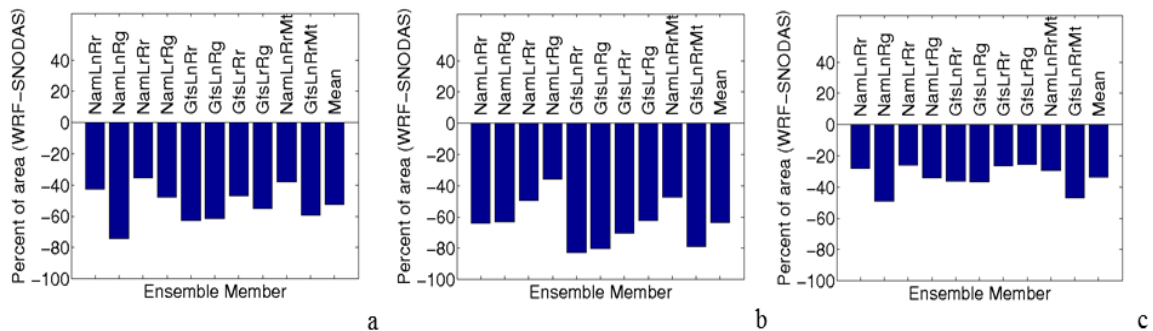
Figure 22. Difference in coverage area of event total precipitation exceeding 0.10 inches for the 36P (a), 24P (b), and 12P (c) ensemble mean and member forecasts. Event total precipitation is for the forecast area during the 30 hour period of 12 UTC 11 March to 18 UTC 12 March. Note that the area considered is the forecast area (eastern North Dakota and northwestern Minnesota), not Domain 2.

By examining both the ensemble means and the values of the individual members in Fig. 22, it is clear that the 12P forecasts performed best with an underprediction in coverage ranging from 25.7% to 49.0% and a mean of 33.9% (Fig. 22c). The 24P ensemble forecasts perform worst with ensemble members underpredicting coverage by an average of 63.5% (Fig. 22b), and the 36P ensemble members underpredict coverage by an average of 52.4% (Fig. 22a). The best members are typically the NamLrRr,

46

NamLrRg, and NamLnRrMt members (Fig. 22), but this is not always the case (e.g., the 12P ensemble member forecasts). Ensemble members with RUC LSM or Thompson microphysics schemes perform better than similar members with an opposing (Noah) LSM or (WSM6) microphysics scheme, but the radiation schemes and model forcing do not show a clear trend for all three forecasts.

4.2.4 Wind Speed Threshold Verification

Appendix D shows the variability in the probability of exceeding wind speed thresholds of 20 mph (17.4 kts), 25 mph (21.7 kts), 30 mph (26.1 kts), and 35 mph (30.4 kts) at 18 UTC March 11, 00 UTC March 12, 06 UTC March 12, and 12 UTC March 12 for the 12P ensemble member forecasts and MERRA data in Domain 2*. Around 18 UTC 11 March 2011, MERRA wind speeds exceeding 25 mph develop in southeastern North Dakota. By 00 UTC 12 March 2011, MERRA wind speeds exceeding 25 mph become widespread over eastern North Dakota and begin to develop in northwestern Minnesota (Fig. 23). These areas of winds exceeding 25 mph diminish in eastern North Dakota but remain in northwestern Minnesota until 12 UTC 12 March 2011 (Appendix D). For wind speeds exceeding 30 mph, MERRA data shows isolated locations in the Domain 2*, which occur between 18 UTC March 11 and 06 UTC March 12. Wind speeds exceeding 35 mph only develop in central North Dakota from 22 UTC 11 March to 02 UTC 12 March.

Automated surface observations over Domain 2* show wind speeds exceeding 25 mph starting at 22 UTC 11 March 2011 in eastern North Dakota, increasing to most locations exceeding 35 mph at 00 UTC 12 March 2011 (NCDC 2005). These widespread

47

wind speeds exceeding 35 mph continued until 09 UTC 12 March 2011, when wind

speeds begin to quickly diminish to below 25 mph.
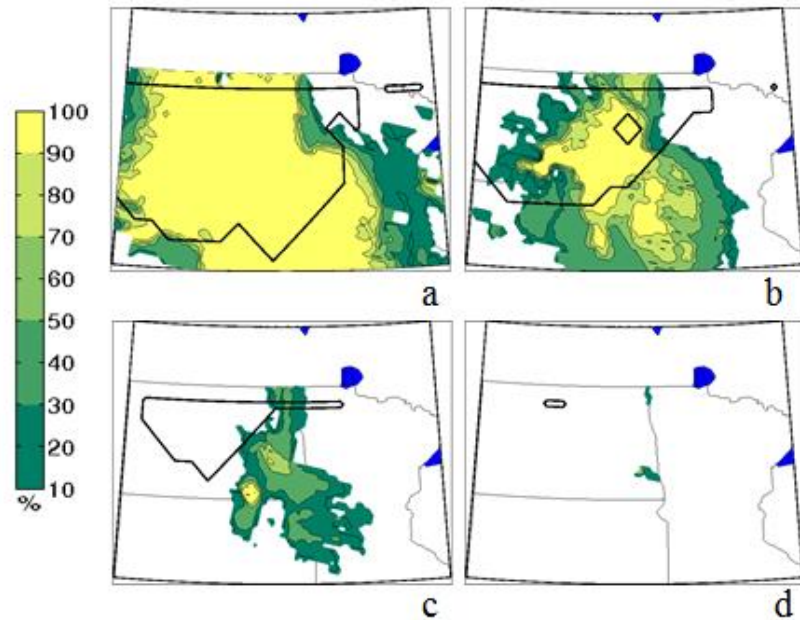
Figure 23. Probability of exceedance (color fill) at 20 mph (a), 25 mph (b), 30 mph (c), and 35 mph (d) for the 12P ensemble wind speed forecasts at 00 UTC 12 March 2011 in Domain 2*. Coverage area of MERRA wind speeds is outlined in black.

Figure 23 is a representative valid time of the probability of exceedance during

the event for the 12P ensemble forecasts and MERRA wind speeds exceeding four

different threshold values at 00 UTC 12 March 2011. This is a time during which the

clipper is greatly impacting the forecast area. Nearly all of the ensemble members

forecast wind speeds exceeding 20 mph during much of the event, but only a few

members forecast wind speeds exceeding 30 mph (Fig. 23 and Appendix D). Wind

speeds exceeding 35 mph are rarely forecast by the 12P ensemble members. It is clear

from examining Fig. 23 and Appendix D that the 25 mph threshold best captures the

variability in the ensemble member forecast wind speeds during the beginning, middle,

and end of the event. Thus 25 mph will be the primary threshold used for comparing the

48

three forecasts, even though the operational threshold for blizzard criteria stands at 35 mph.
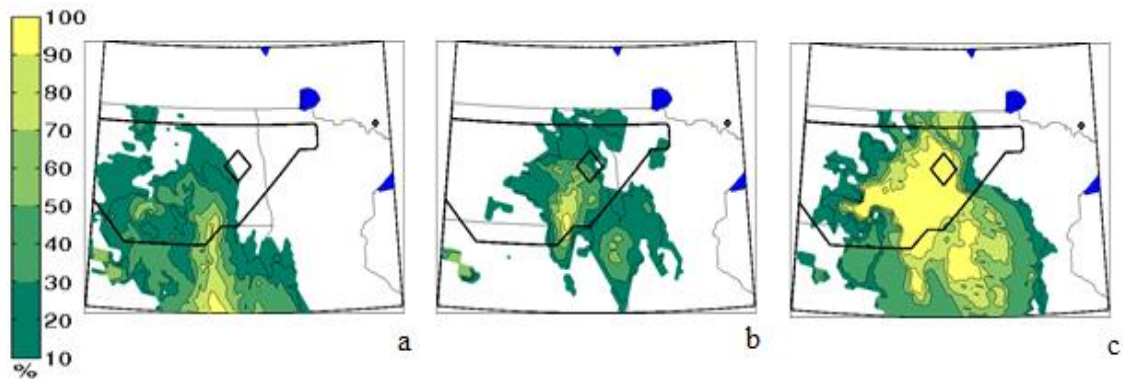
Figure 24. Probability of exceedance (color fill) at 25mph for the 36P (a), 24P (b), and 12P (c) ensemble forecasts wind speeds at 00 UTC 12 March 2011 in Domain 2*. Coverage area of MERRA wind speeds is outlined in black.

Figure 24 depicts the probability of exceeding 25 mph for the 36P, 24P, and 12P ensemble member wind speeds compared to the MERRA winds at 00 UTC 12 March 2011. For the variability in the probability of exceedance amongst the three forecasts at other times in Domain 2*, refer to Appendix E. Figure 24c shows all of the 12P member forecasts are predicting wind speeds exceeding 25 mph and is a representative depiction of the 12P member forecasts at and prior to 00 UTC 12 March 2011. While the 36P and 24P forecasts show smaller areas with all of the members predicting wind speeds exceeding 25 mph at and before 00 UTC 12 March 2011, they have only 10-30% of member forecasts predicting these higher magnitudes in the correct location (Figs. 24a and 24b, Appendix E). After 00 UTC 12 March 2011, all three forecasts have the majority of members forecasting wind speed coverage exceeding 25 mph too far south and east of where it was observed (Appendix E). It can be concluded from Appendix E that the 12P ensemble member wind speed forecasts performed better in terms of location and intensity than the 36P and 24P forecasts.

Figure 25 quantifies the difference (from the MERRA coverage) in areal coverage of wind speeds exceeding 25 mph for the 36P, 24P, and 12P forecasts. Regarding ensemble members' physics schemes, ensemble members with the NOAH LSM or GFDL radiation scheme tended to perform better for higher wind speeds than the ensemble mean and other members. In particular, the GfsLnRg member was a top performing member for all three wind speed forecasts and even overestimated the coverage of wind speeds over 25 mph at times. On the other hand, ensemble members with RRTMG radiation scheme and RUC LSM performed poorly and often underestimated coverage of wind speeds over 25 mph by 50% or more. The Thompson microphysics scheme did not perform well, except for the 24P NamLnRrMt member.
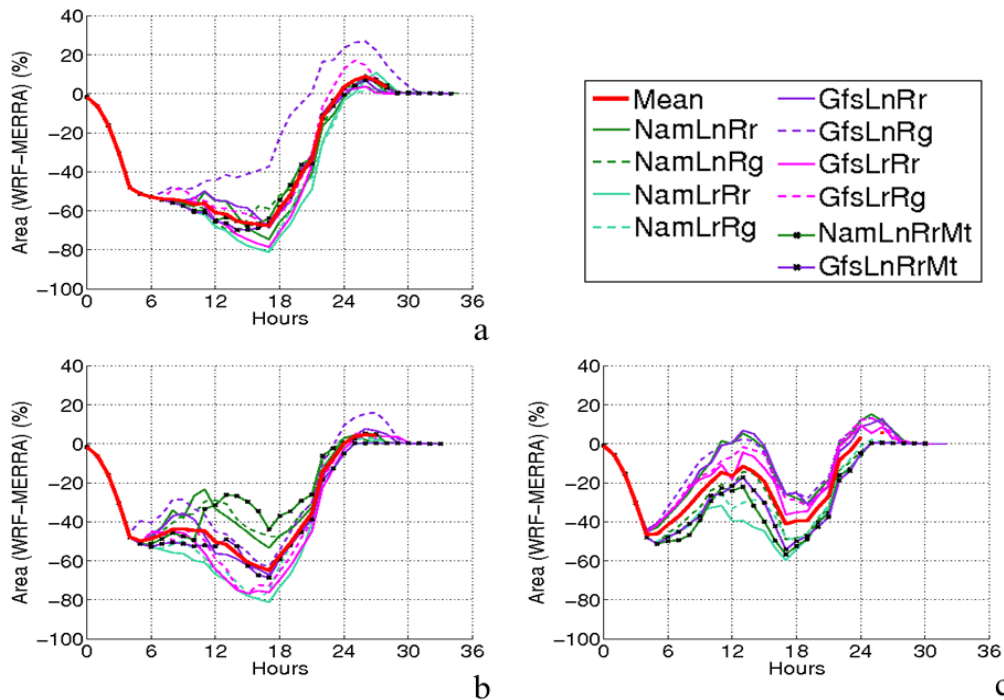
Figure 25. Difference in coverage area of wind speeds exceeding 25 mph for the 36P (a), 24P (b), and 12P (c) ensemble mean (red) and members' (remaining colors) forecasts. Wind speeds are considered for the forecast area during the 36 hour period of 12 UTC 11 March to 00 UTC 13 March. Note that the area considered is the forecast area (eastern North Dakota and northwestern Minnesota), not Domain 2.

All three forecasts underestimate wind speeds (as negative values) throughout most of the event, and similarly to the forecast precipitation coverages, the forecasts' general trends, durations, and magnitudes differ greatly. The 12P forecast performed best overall when forecasting wind speeds exceeding 25 mph (Fig. 25c). The 12P ensemble mean forecast and most members remained within 40% of the observed coverage for most of the event, and nearly all of the 12P ensemble members remained within 50% of the MERRA coverage. While the 12P ensemble members perform particularly poorly at 18 UTC 11 March and 06 UTC 12 March 2011, this underprediction is still an improvement on the 36P and 24P forecasts. The 24P forecast performs slightly better than the 36P forecast and has a few members with coverage within 30% (Figs. 25a and 25b). However, the 36P and 24P forecasts perform particularly poorly around 06 UTC 12 March 2011, when they both have members that underpredict coverage by nearly 80%.

4.2.5 Grid Point Forecast Verification

Figures 26 and 27 provide the observations and grid point forecasts for precipitation and wind speeds, respectively, at the Grand Forks International Airport (KGFK) in Grand Forks, ND. Hourly precipitation was reported at KGFK from 01 UTC to 08 UTC 12 March 2011 (Fig. 26), averaging 0.02 to 0.03 inches of precipitation per hour. The SNODAS data shows slightly higher precipitation rates and depicts a shorter event than the ASOS. Therefore, this study assumes that a relatively accurate grid point forecast should fall between or very near the values from the local ASOS and SNODAS data. The hourly precipitation values for all of the ensemble members are within 0.05 inches per hour of the ASOS and SNODAS values.
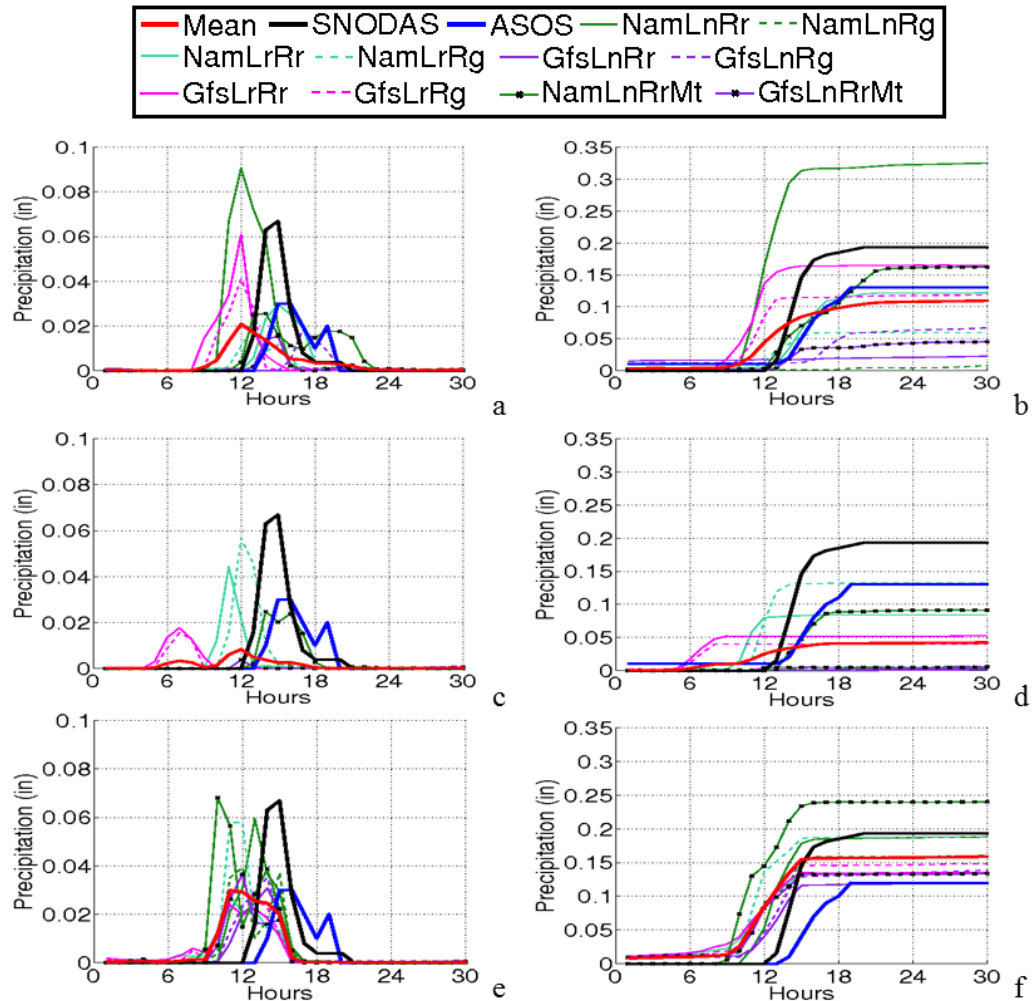
Figure 26. Hourly (left) and accumulated (right) liquid equivalent snowfall for SNODAS (black), ASOS (blue), ensemble mean (red) and members' (remaining colors) forecasts at the Grand Forks International Airport (KGFK). Valid from 12 UTC 11 March 2011 to 18 UTC 12 March 2011 for the 36P (a) hourly and (b) accumulated, 24P (c) hourly and (d) accumulated, and 12P (e) hourly and (f) accumulated ensemble forecasts.

The best hourly and accumulated precipitation forecasts at KGFK are the 12P ensemble members' and mean forecasts (Fig. 26). The 12P ensemble mean forecast parallels the ASOS very closely in amounts, trend, and length of time, despite 12P ensemble mean forecast precipitation starting and ending prior to the ASOS values. Several ensemble members mirror the hourly maximum precipitation values from SNODAS data, near 0.07 inches, and most members fall between the ASOS and

52

SNODAS hourly maximum precipitation values. Additionally, the 12P ensemble mean

and nearly all of the 12P ensemble members have total accumulated precipitation

between the ASOS and SNODAS values, 0.12 and 0.19 inches respectively.

Regarding timing, most ensemble members and even SNODAS show

precipitation starting and ending earlier than was observed (Fig. 26). While most

members started precipitation between 22 UTC 11 March 2011 and 03 UTC 12 March

2011, a few members (36P and 24P GfsLrRr, 24P GfsLrRg, and 12P NamLnRrMt)

started and ended precipitation noticeably differently from the ASOS and SNODAS data

and/or other members.
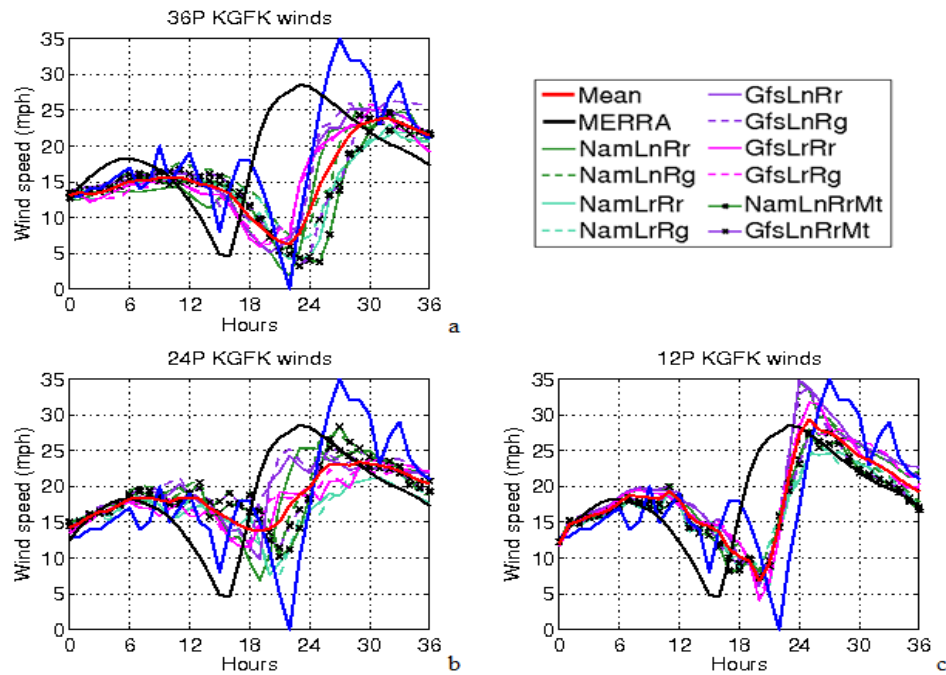
Figure 27. Hourly wind speeds for MERRA (black), ASOS (blue), ensemble mean (red) and members' (remaining colors) forecasts at the Grand Forks International Airport (KGFK). Valid from 12 UTC 11 March 2011 to 00 UTC 13 March 2011 for the 36P (a), 24P (b), and 12P (c) ensemble forecasts. Note this time range is different from the other figures.

The KGFK ASOS reported sustained wind speeds exceeding 25 mph from 01

UTC 12 March until 0700 UTC 12 March, with gusts exceeding 35 mph from 0136 UTC

53

12 March to 0810 UTC 12 March (Fig. 27). The MERRA data had sustained winds exceeding 25 mph from 21 UTC 11 March to 03 UTC 12 March and peaking at about 28 mph. The MERRA data shows a lower maximum wind speed, but its general trend and the duration of its maximum wind speeds are similar to the ASOS data.

The best wind speed forecasts at KGFK are the 12P ensemble members' and mean forecasts. The 12P ensemble mean forecast is nearly identical to the MERRA values in magnitude but is even closer to the trend and timing of the ASOS' peak wind speeds than the MERRA data. The three ensemble members with the Noah LSM and RRTMG radiation schemes have peak wind speeds of the same magnitude as the ASOS, 35 mph, but produce them three hours later. Additionally, most 12P member forecasts follow the trend and peak of the MERRA wind speeds well and are within a few miles per hour of the MERRA data throughout the event. Unfortunately, none of the ensemble forecasts excel at capturing the hour-to-hour variations in the ASOS wind speed data 18 hours prior to or 6 hours after the peak sustained wind speed.

The 36P and 24P wind speed forecasts at KGFK perform relatively poorly and often underpredict wind speed magnitudes. Two of the 24P ensemble member forecasts have closer wind speed maxima to the observed values than any of the 36P ensemble member forecasts. However, the 36P ensemble mean forecast minimum and maximum wind speeds have magnitudes and timing closer to the observed values, and the abrupt changes in wind magnitude from 20 UTC 11 March to 00 UTC 12 March 2011 are predicted better with the 36P ensemble member forecasts than the 24P forecasts.

4.3 March 2011 Case: Summary of Results

The results of this study for the March 2011 case show that nearly all of the three forecasts' ensemble members underpredict precipitation and wind speed intensity and coverage throughout the majority of the event. All of the three forecasts' ensemble members are also unable to capture the correct location of the cyclone at initialization. They depict a cyclone center hundreds of kilometers north of the observed cyclone center, even within a few hours of the cyclone center entering the forecast area (Figs. 14a, 14c, and 14e). Despite these initial errors, all of the track forecasts improved greatly by 22 UTC 11 March and performed best during the second half of the event.

It was expected that the 12P forecasts would verify best compared to the observed values because of the decreased error associated with an initialization time closer to the start of the event. Following this assumption, it would be also expected that the 36P forecasts would tend to perform worse than the 24P forecasts overall. 12P ensemble mean and member track, wind speed, and precipitation forecasts outperform the 36P and 24P forecasts for the majority of the March 2011 event (Sections 4.2.2, 4.2.3, and 4.2.4). The 36P wind speed forecasts perform worse than the 24P wind speeds forecasts, but the 36P track and precipitation forecasts perform better than the 24P track and precipitation forecasts.

All three ensemble precipitation forecasts predict snowfall too far east (Section 4.2.3). They predict the heaviest snowfall to occur along the Canadian border of Minnesota, while the SNODAS and local ASOS data display the heaviest snowfall in central North Dakota. Measurable precipitation (above 0.01 inches) coverage is underpredicted by an average of 20-30% for the 36P and 24P ensembles and 10-20% by

the 12P ensemble. For the coverage of event total precipitation exceeding 0.10 inches, the 12P ensemble members underpredict coverage by an average of 33.9%, the 24P ensemble members underpredict coverage by an average of 63.5%, and the 36P ensemble members underpredict coverage by an average of 52.4%. All of the ensemble wind speed forecasts underpredict wind speed intensity and coverage (Section 4.2.4). Wind speed coverage of 25 mph is underpredicted by an average of 40-50% for the 36P and 24P ensembles and 20-30% for the 12P ensemble.

The 12P precipitation and wind speed forecasts at KGFK performed quite well compared to the observed values (Figs. 26 and 27). The 12P ensemble mean precipitation forecast at KGFK was between the ASOS and SNODAS values for the hourly and total accumulated precipitation, even though most of the 12P ensemble member forecasts started precipitation too early (Figs. 26e and 26f). Additionally, the 12P ensemble mean and a few 12P ensemble member forecasts were able to predict similar peak sustained wind speeds values (near 35 mph) compared to those from the verification datasets (Fig. 27c). The 36P and 24P ensemble mean forecasts indicated peak sustained wind speeds of only 23-24 mph at KGFK with only a few members exceeding 30 mph (Figs. 27a and 27b).

4.4 Overall Ensemble Evaluation

True ensemble performance statistics are not possible with only one case study, but case-dependent qualitative value can be assessed. This section contains an evaluation of the overall ensemble spread, mean performance, and outliers by forecast, as well as the performance of the physics schemes and forcings used for the March 2011 case. This

section concludes with an investigation of the impacts of the "double-barrel low" feature on ensemble performance.

4.4.1 Ensemble Spread and Mean Performance

Figs. 15a, 15c, and 15e show moderate predictability (relative to the MERRA track) in the 36P, 24P, and 12P ensemble member track forecasts, but most of the track forecasts are consistently at least 100 to 300 km from the MERRA track. Nearly all of the ensemble members' forecasts of the surface low positions show relatively small spread, except for a few 36P and 12P member forecasts. These 36P and 12P members providing a large spread are either members with a NOAH LSM and RRTMG radiation scheme that are performing well early in the event or error-laden outlier members (e.g., 36P NamLnRg and 36P NamLrRg track forecasts and 36P GFSLnRg wind speed coverage forecast). While the quantitative evaluation in Figs. 15b, 15d, and 15f indicates a larger spread hour-by-hour, the lack of accuracy in the ensemble members' track forecast (relative to the MERRA track) is concerning during the first six hours of the event.

Given the offset in the ensemble track forecasts compared to the MERRA track, it is not surprising that the limited accuracy and lack of large spread in the track forecasts carries into the coverage forecasts (Figs. 21 and 25). Figs. 21 and 22 indicate moderate spread (relative to other time periods) in all three forecasts' precipitation coverage, and all of the ensemble members and means underpredict the precipitation coverage for most of the event. Fig. 25 indicates moderate spread in the 24P and 12P wind speed coverage and small spread in the 36P wind speed coverage, but the majority of the ensemble members and means underpredict the wind speed coverage at the 25 mph threshold for the majority of the event.

Regarding performance of the ensemble mean compared to the individual members, the mean track forecast only performs consistently better than the individual member forecasts for the 36P track forecast (Fig. 15). Several 24P and 12P individual member track forecasts, particularly those with the RUC LSM and RRTMG radiation scheme, tend to perform better than their respective mean track forecasts for much of the event. The underprediction of precipitation and wind speed coverage in nearly all of the ensemble member forecasts greatly limits the accuracy of the mean forecasts for these fields (Figs. 21 and 25). Thus, the ensemble means do not perform better than the best-performing individual member forecasts for precipitation and wind speed coverage.

4.4.2 Performance of Ensemble Physics Schemes and Forcings

No model forcing, microphysics scheme, or radiation scheme performed best for all three verification fields, but the RUC LSM tended to be the best performing LSM, particularly when paired with the RRTMG radiation scheme. For the cyclone track forecasts, no individual ensemble member consistently performed better than the ensemble mean in all three forecasts (Fig. 15). Ensemble members with the NOAH LSM performed well for the track forecasts overall, and ensemble members with the Thompson microphysics performed well for the track forecasts after 00 UTC 12 March.

For wind speed forecasts, most ensemble members with the RUC LSM and RRTMG radiation scheme performed poorly (Figs. 25 and 27). Generally, the ensemble member wind forecasts with the GFS forcing or Noah LSM performed best, and the GfsLnRg member generally performed better than the ensemble mean. For precipitation forecasts overall, most ensemble members with GFS forcing performed poorly, but ensemble

members with the RUC LSM tended to perform best, and the NamLnRrMt member performed better than the ensemble mean (Figs. 21, 22, and 26).

These results generally agree with those of Yuan et al. (2012) in regards to best performing physics schemes. In both studies, the best member (or one of the best members) tended to have a Noah LSM and Thompson microphysics, but members with the RUC LSM generally performed better than those with the Noah LSM for the precipitation results.

4.4.3 Impacts of Possible Observation Errors on Ensemble Verification

Other than the potential errors associated with the ensemble setup and threshold values selected specifically for this study, it is important to consider the possible observation-related errors that may impact the accuracy of the March 2011 ensemble forecast. One possible source of observation-related error is the initial conditions. The ensemble mean and member track forecasts from Figs. 15a, 15c, and 15e begin in south central Saskatchewan, instead of east central Montana as depicted by MERRA data. Those areas in and immediately upstream of Domain 2 are known to have sparse surface and upper air observation coverage (e.g., surface stations in Fig. 28). Both the lack of observations and their proximity to the inflow boundary of Domain 2 could introduce errors into the ensemble's initialization and initial forecast hours.

Another possible source of observation-related error in this study is error in the observational datasets used for verification. Figure 27 gives an example of this, showing that the MERRA peak sustained wind speed is nearly 12 mph slower than the ASOS observed value. Overall, the MERRA data for the March 2010 case perform poorly at KGFK; MERRA is unable to depict peak wind speeds that are vital to operational

59

decision making for this case. These discrepancies are likely a result of several factors, including the forcing models, data assimilation methods, and coarse horizontal grid spacing and time averaging of the MERRA wind data.

Significant limitations also arise in the verification between the gridded SNODAS data and the ASOS observations. Figure 26 shows that the SNODAS data overestimate hourly precipitation for several hours at KGFK. As a result, the SNODAS total accumulated precipitation is greater than the ASOS total accumulated precipitation at KGFK by 0.06 inches for the March 2011 case. Factors such as the forcing models, data assimilation methods, and inherent errors related to assimilated observational datasets impact the SNODAS data. Furthermore, many studies (e.g. Rasmussen et al. 2012, Goodison et al. 1998, Yang et al. 1998) have found that obtaining reliable and relatively accurate automated solid precipitation measurements continues to be a significant challenge. These errors are magnified in particularly windy environments (e.g., blizzard conditions), resulting in significant concerns related to undercatch and snow-to-liquid ratios. Current methods of collecting automated snow precipitation measurements tend to have more acute observational errors and are less reliable than any other variable field. These limitations make it difficult to accurately assess the ensemble forecast performance in this study.

4.4.4 Ensemble Sensitivity to a Double Barrel Low Feature

While initial conditions and observational datasets are noteworthy error sources, the automated identification of the cyclone is particularly important to examine with the March 2011 case. The cyclone track forecasts in this study are created through an automated identification process, as discussed in Sections 2.2 and 3.2.2. Certain features

60

existing in an extratropical cyclone, such as a double-barrel low, can be challenging for

an automated identification process. Double-barrel low is a common term in operational

meteorology and refers to a situation in which two local SLP minima (i.e., cyclone

centers) are present within an extratropical cyclone.

The NOAA WPC surface analysis at 12 UTC 11 March 2011 in Fig. 28 shows the

presence of a double barrel low with SLP minima in both western North Dakota and

southern Saskatchewan. Prior to this time, the northern-most minimum entering

Saskatchewan is the primary cyclone center (SPC 2005). As the parent shortwave trough

deepens and jet maxima increase over southeast Montana, the primary cyclone center

shifts to a secondary SLP minimum (south of the previous cyclone center) a few hours

after this time. While the MERRA cyclone track agrees with the WPC surface analyses'

intensification of the cyclone and the cyclone center shift, many of the ensemble member

and mean track forecasts do not (Figs. 15a, 15c, and 15e).
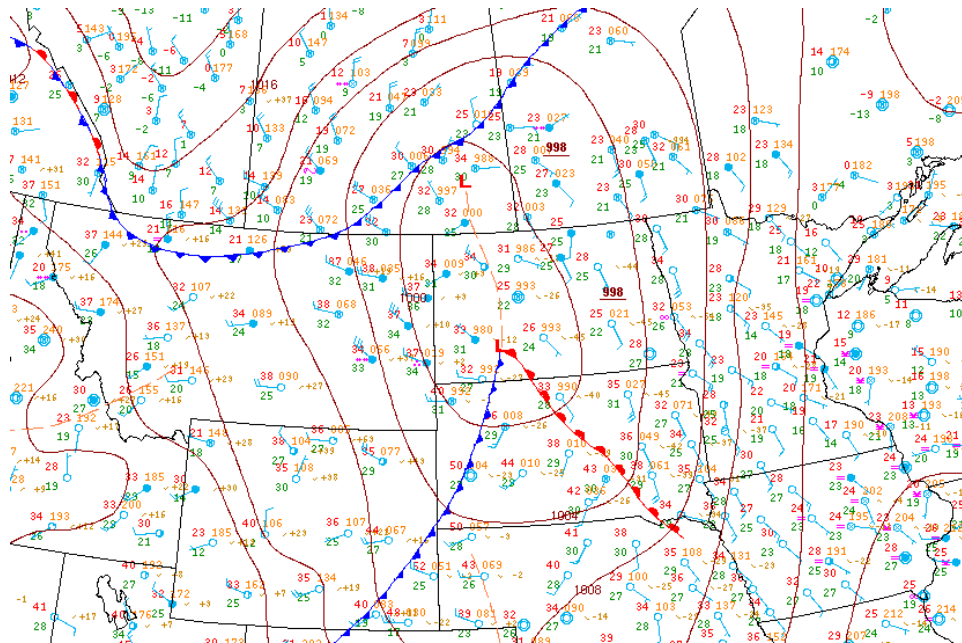
Figure 28. NOAA Weather Prediction Center (WPC) surface analysis at 1200 UTC 11
March 2011. This analysis depicts analyzed fronts, surface station locations and winds

(blue), temperatures (red), dew point temperatures (green), MSL pressure and pressure change (orange), and isobars (red contours) over the north central United States and south central Canada.

Table 2 summarizes the results of subjective analysis performed on the SLP forecasts for the 30 ensemble members. This subjective analysis determined whether a double barrel low was present and additionally whether the secondary SLP minima intensified (as discussed in Section 4.5.2). All of the 36P and 12P members were able to predict a double barrel low for at least one hour while the cyclone was inside of Domain 2. With the exception of the 24P GFS members, he 24P NAM members also predicted a double barrel low for at least one hour.

Table 2. A summary of surface SLP characteristics for each individual ensemble member. The ensemble member name is listed in the first column. The corresponding characteristics—double barrel low and secondary center intensification—are denoted in the subsequent columns for each of the three forecast times. Each ensemble member predicting the existence of a secondary SLP minimum in North Dakota during at least one hour of the analysis period is identified with an 'x' under the double barrel column. Each member predicting a secondary SLP minimum in North Dakota which intensifies enough to become the dominant cyclone center and alters the track forecast by at least 50 km (i.e., showing further intensification of the secondary center) is identified with an 'x' under the intensification column.

Member	Double Barrel Low			Intensification of Secondary SLP Minima		
	36P	24P	12P	36P	24P	12P
NamLnRr	x	x	x	x	-	x
NamLnRg	x	x	x	-	x	x
NamLrRr	x	x	x	-	-	x
NamLrRg	x	x	x	-	x	x
GfsLnRr	x	-	x	x	-	x
GfsLnRg	x	-	x	x	-	x
GfsLrRr	x	-	x	x	-	x
GfsLrRg	x	-	x	x	-	x
NamLnRrMt	x	x	x	x	x	x
GfsLnRrMt	x	-	x	x	-	x

Figure 29 shows a physical example of the double barrel low (Fig. 29a) and the intensification of the secondary SLP minima (Fig. 29b) for one of the ensemble members.

The 36P NamLnRr surface analysis forecast exhibits a double barrel low (Fig. 29a) at 18

UTC 11 March 2011 and no longer shows this feature by 00 UTC 12 March 2011 (Fig.

29b). This appears to create an abrupt north-to-south track shift from the primary

(northern) to the secondary (southern) SLP minimum during the period from 18 UTC 11

March to 00 UTC 12 March (Figs. 15, 29a, and 29b). This abrupt track shift is physically

realized as an intensification of the secondary SLP minimum as it becomes the primary

SLP minimum and is present in many of the ensemble member track forecasts (Figs. 15a,

15c, and 15e).
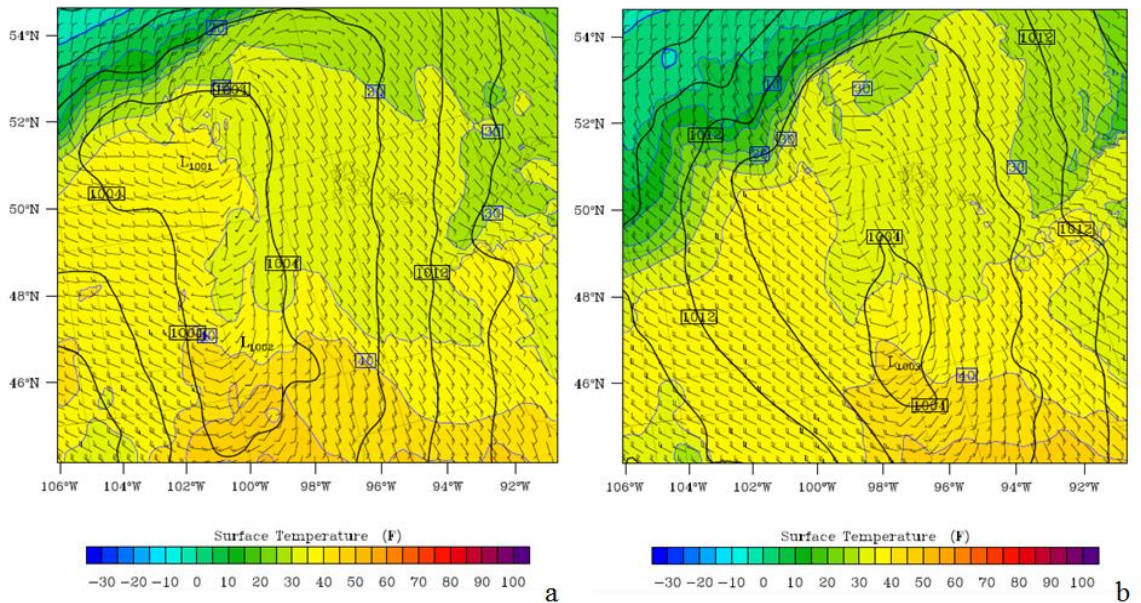
Figure 29. Surface analysis from the 36P NamLnRr member forecasts (a) with a double barrel low 18 UTC 11 March 2011 and (b) without a double barrel low 00 UTC 12 March 2011. This analysis depicts the forecast SLP in millibars (black contours), surface winds (black wind barbs; knots), and surface temperature in degrees Fahrenheit (color fill and blue contours) over Domain 2*.

All of the 12P members, many of the 36P members, and several of the 24P members

were able to predict intensification of the secondary SLP minimum, and this allowed

them to perform particularly well in comparison to the other forecasts (Table 2, Fig. 15).

Unfortunately, all of the ensemble member and mean track forecasts resolved the

intensification of the secondary low too slowly relative to the MERRA data and WPC analyses (Fig. 15).

With a difference of only a few millibars between the two SLP minima (Figs. 29a and 29b), it is important to explore whether the track forecasts could have improved if the ensemble members had intensified the primary and secondary SLP minima differently during the event. The 36P NamLnRr track forecast was an individual track forecast that performed well, as it predicted a double barrel low and intensification of the secondary minimum sooner than many other members in all three forecasts (Table 2 and Fig. 15). Thus, it is a good example for further analysis.

Fig. 30 depicts the MERRA and 36P NamLnRr automated track forecasts, along with a subjective track forecast (36P NamLnRr_2) created only with the locations of the secondary SLP minima. While the 36P NamLnRr automated track forecast exhibits two abrupt track shifts, the 36P NamLnRr subjective track forecast does not. Throughout the event, the subjective track forecast is much closer to the MERRA track than the automated (36 NamLnRr) track forecast. While it is not realistic to perform a subjective analysis on every ensemble member, this analysis proves that the error in the 36P NamLnRr track forecast is not as large as it previously appeared.

In summary, the double barrel low feature from the March 2011 case presented a major challenge to the ensemble, beyond the errors in the initial conditions and observational datasets. Even though many ensemble member forecasts presented a double barrel low feature and some individual members performed well, the overall results showed only a small to moderate spread, underpredicted coverage of near-blizzard criteria wind speeds, and underpredicted precipitation coverage. An automated

identification method was used to identify the ensemble forecast tracks, which often

struggled to depict intensification of the secondary cyclone center as quickly as observed.

This intensification occurring sooner in the simulations could have significantly

improved ensemble mean and member forecasts.
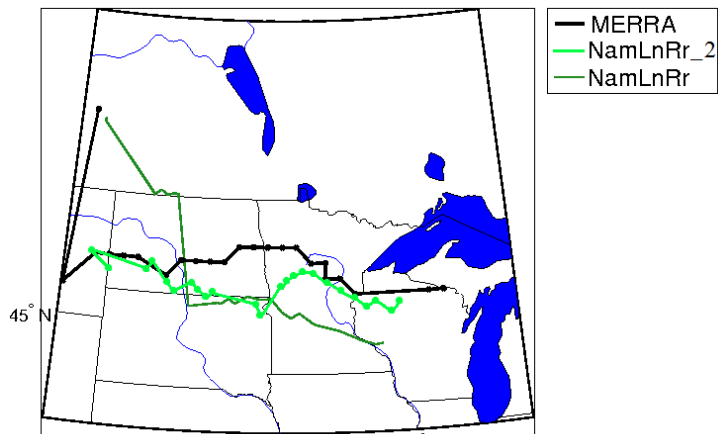
Figure 30. Surface cyclone center for the MERRA (black) track and 36P NamLnRr automated (light green) and 36P NamLnRr_2 subjective (dark green) ensemble member track forecasts across Domain 2 from 06 UTC 11 March 2011 to 12 UTC 12 March 2011.

CHAPTER V

RESULTS AND DISCUSSION: ADDITIONAL CASES

This chapter contains the results and discussion for the January 2009 and December 2013 cases. The first two sections consist of the results of the additional two cases. Each of these two sections begins with a synoptic overview, followed by forecast verification of the case's cyclone track, precipitation, and two-meter wind speed. The final section presents a discussion comparing the results from the January 2009 and December 2013 cases to the results from the March 2011 case.

5.1 January 2009 Case Results

5.1.1 Synoptic Overview

An Alberta clipper was expected to bring a band of two to six inches of snowfall and wind gusts up to 45 mph (20 m s^{-1}) to the forecast area during the night of 11 January 2009 to the morning of 12 January 2009 (NWS Grand Forks 2009a). A blizzard watch and winter storm watch were issued by NWS Grand Forks for the forecast area during the afternoon of 10 January. By the afternoon of 11 January, a blizzard warning was issued for southern and western portions of the forecast area, which were predicted to receive the heaviest snowfall and greatest wind gusts. For a timeline comparing the ensemble forecasts and the event for the January 2009 case, refer to Fig. 31.
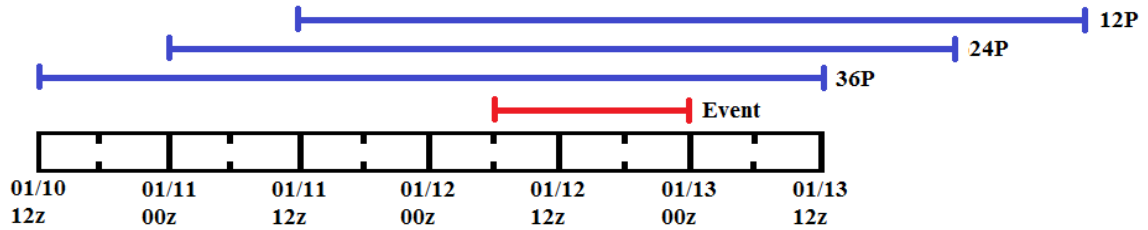
Figure 31. As in Fig. 3 but for the January 2009 case.
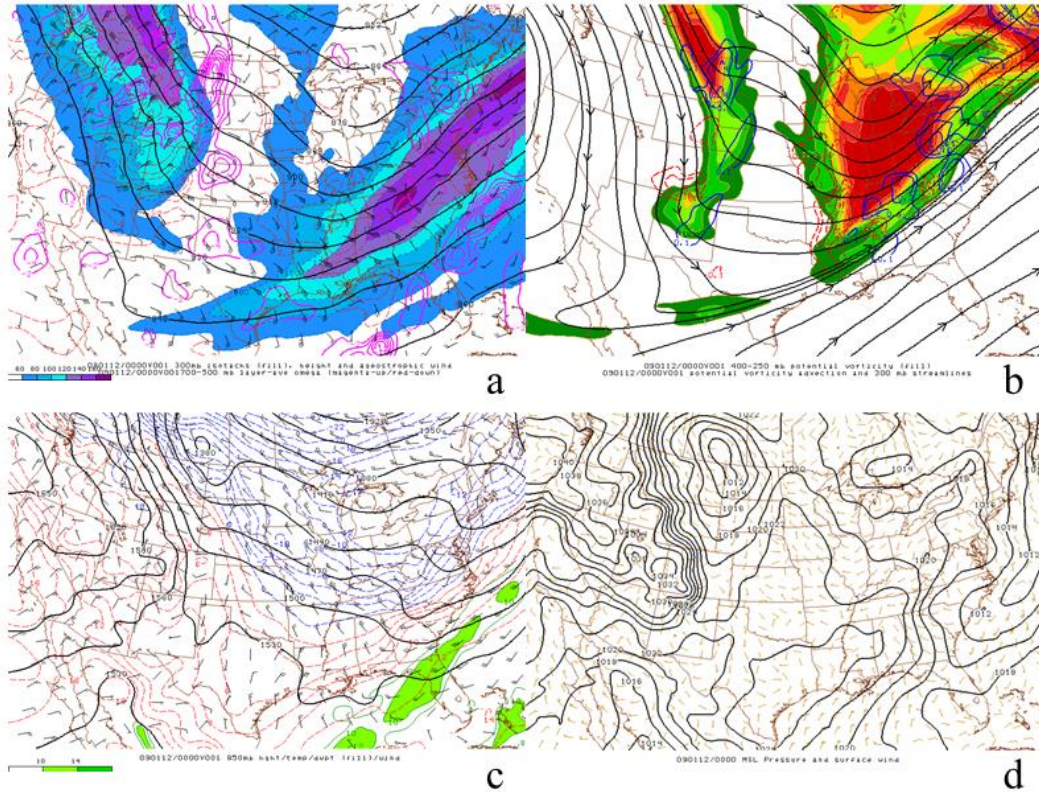
Figure 32. As in Fig. 11 but at 0000 UTC 12 January 2009.

Figures 32-34 depict the clipper's development for the 300 hPa, 500 hPa, 850 hPa, and surface levels at 0000 UTC 12 January, 1200 UTC 12 January, and 0000 UTC 13 January. A fast-moving Alberta clipper departed southeast from the Canadian Rockies a few hours before 0000 UTC 12 January 2009 (Fig. 32). The general synoptic pattern across CONUS remained the same throughout the event, with a 500 hPa longwave ridge anchored just off the West Coast and a deep, positively tilted 500 hPa longwave trough

67

over the central United States. However, there were developing features of interest in relation to the clipper that allowed for its rapid evolution: a corresponding 300 hPa jet streak and a potent 500 hPa shortwave trough associated with a relative vorticity maximum.

By 1200 UTC 12 January 2009, the clipper propagated into northeast South Dakota and strong cold air advection developed over the forecast area (Fig. 33). Through the morning of 12 January, the approaching 300 hPa jet streak and 500 hPa shortwave trough continued to intensify the aforementioned clipper. A high pressure system developed northwest of the clipper in Saskatchewan, causing a strong surface pressure gradient upstream of the clipper.
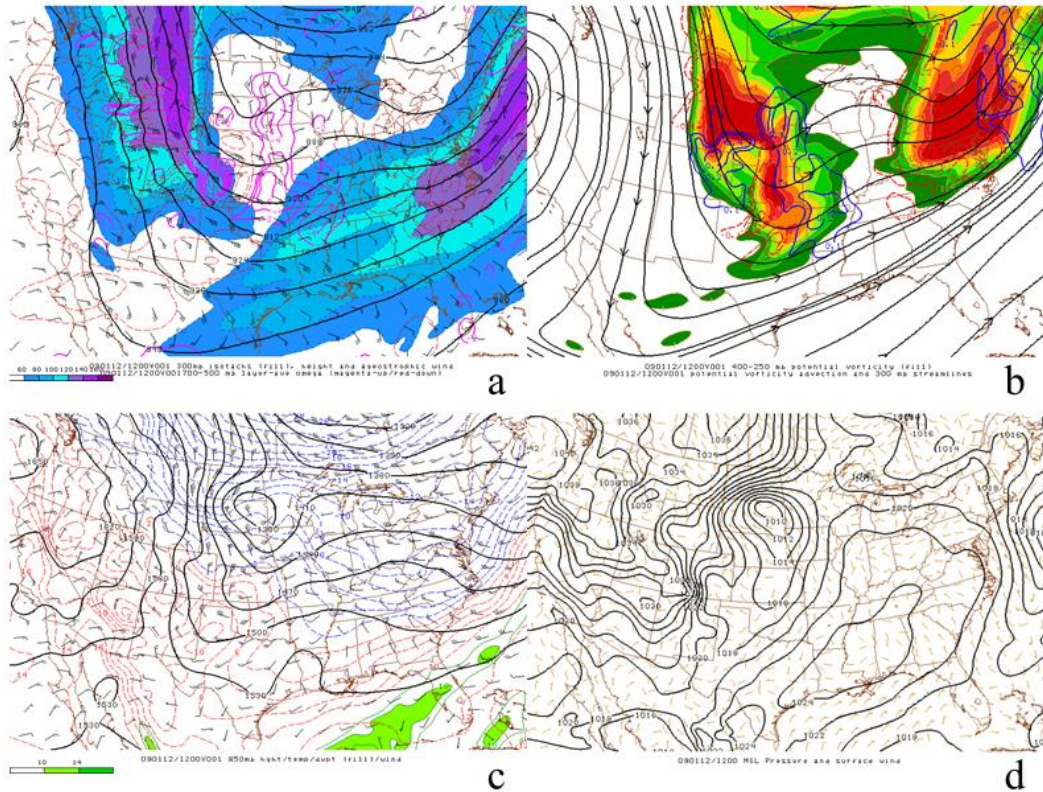
Figure 33. As in Fig. 11 but at 1200 UTC 12 January 2009.

During the afternoon hours of 12 January, the clipper tracked through northern

Iowa, and blizzard conditions began to subside from north to south as the surface

pressure gradient weakened. The approaching 300 hPa jet streak strengthened and

deepened the shortwave trough associated with the clipper, making it the primary

longwave trough with a positively tilted axis stretching from Quebec to West Texas (Fig.

34). By 0000 UTC 13 January, the clipper was centered over eastern Iowa, and blizzard

conditions had ceased. Event total snowfall across the forecast area ranged from two to

five inches with the greatest amounts falling in southeast North Dakota (NWS Grand

Forks 2009b).
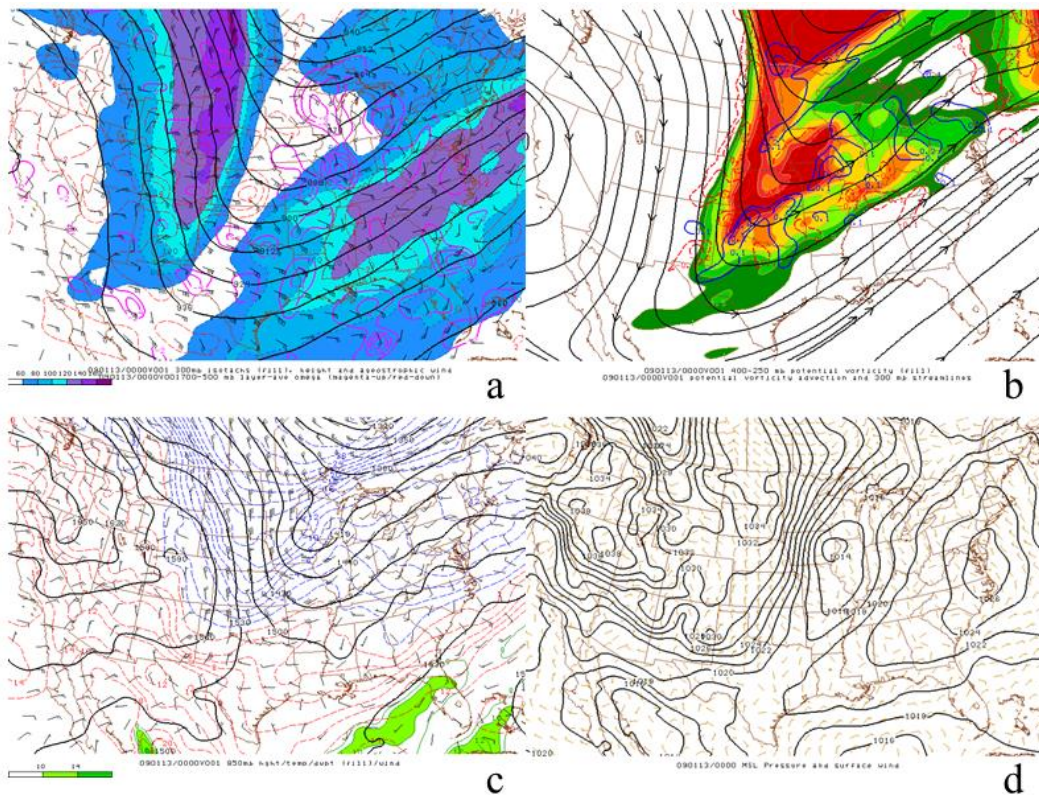
Figure 34. As in Fig. 11 but at 0000 UTC 13 January 2009.

5.1.2 Ensemble Mean and Member Track Verification

Figure 35a shows the MERRA and ensemble mean track forecasts in Domain 2 throughout the January 2009 event. All three ensemble mean cyclone track forecasts for the January 2009 event agree with the general progression of the SPC mesoanalysis and MERRA reanalysis cyclone tracks throughout the event (Figs. 32, 33, and 35a). Both observed and MERRA tracks show an initial surface cyclone center starting in Saskatchewan with rapid cyclone propagation southeastward into northwestern and central North Dakota (Fig. 35a). All three ensemble mean track forecasts agree with this initial surface pattern, but after the cyclone enters North Dakota, they predict tracks located northeast of the MERRA track for the remainder of the event. As of 12 UTC 12 January 2009, all three ensemble mean track forecasts depict a cyclone center in northeastern South Dakota (Fig. 35a), while the MERRA track and the SPC mesoanalysis depict the cyclone center further south in eastern South Dakota (Fig. 33d). During the last few hours of the event, the 36P and 24P ensemble mean track forecasts remain northeast of the MERRA track, but the 12P ensemble mean track forecast shows a cyclone center quite close to the MERRA track in north central Iowa (Fig. 35a).

Figure 35b shows the absolute error in distance from the MERRA track relative to the event for all three ensemble mean forecasts for the January 2009 event. Despite some minor variations, the 36P and 24P ensemble mean track forecasts perform best from 04 UTC to 13 UTC 12 January. These two ensemble mean track forecasts have an absolute distance of around or less than 100 km during the majority of the January 2009 event. The 12P ensemble mean track forecast performs worst of the three ensemble mean track forecasts and trends away from the MERRA track from 04 UTC to 12 UTC 12 January.

70

However, its peak absolute distance from the MERRA track is only about 180 km, which

is a great improvement in overall ensemble mean track performance compared to the

March 2011 event.
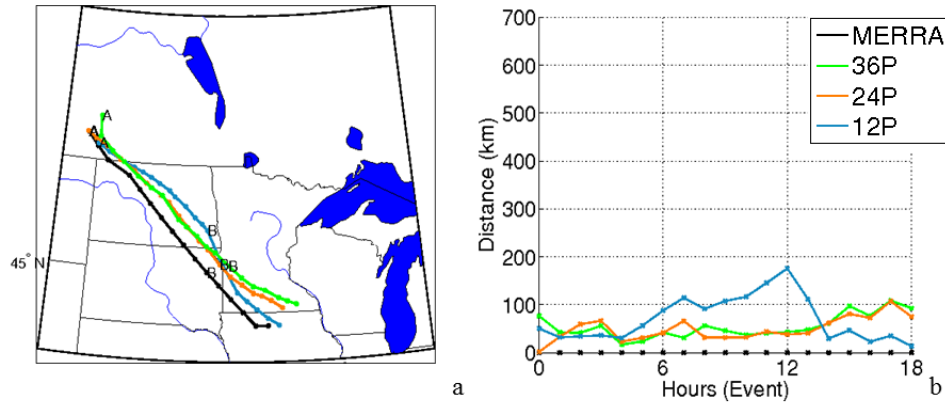
Figure 35. As in Fig. 14 but both figures represent the entirety of an 18 hour event from 00 UTC 12 January 2009 to 18 UTC 12 January 2009. Hours 00 UTC 12 January 2009 and 12 UTC 12 January 2009 are indicated by A and B, respectively.

Figure 36 shows all of the ensemble mean and member track forecasts and their

corresponding absolute distances from the MERRA track throughout the January 2009

event. The majority of the 24P and 12P ensemble member track forecasts and half of the

36P ensemble member track forecasts perform well and remain less than 200 km from the

MERRA track throughout the event. Most ensemble member track forecasts correctly

predict the cyclone propagation from northwestern north Dakota into far northeast South

Dakota and ending in north central Iowa.

The 24P ensemble member track forecasts perform best overall for the January

2009 event (Figs. 36c and 36d). Despite a slight decrease in performance throughout the

event, all of the 24P ensemble member track forecasts remain within 100 km of the

MERRA track during the majority of the event. While the 12P ensemble mean track

forecasts perform worst (of the three ensemble mean track forecasts), further examination

71

of all of the individual ensemble member track forecasts indicates that some of the individual 12P ensemble member track forecasts perform better than the 36P ensemble member track forecasts (Figs. 35 and 36). The 12P ensemble member track forecasts performed well during the first six hours of the event with all members predicting a track less than 100 km from the MERRA track (Figs. 36e and 36f). During the middle of the event, the 12P ensemble track forecasts predict that the track will veer further north and east than the MERRA track. This causes the absolute error in the 12P ensemble track forecasts to climb to between 100 km and 200 km, peaking near 240 km for the Nam LnRg, NamLrRr, and NamLrRg members.

Even though 36P ensemble mean track forecast remains less than 100 km from the MERRA track for most of the January 2009 event, Fig. 36a shows that the ensemble member track forecasts are spread evenly on either side of the MERRA track and are not as close to the MERRA track as the ensemble mean track may make it seem. Poor initialization of the 36P forecasts led to five 36P ensemble members (GfsLnRr, GfsLnRg, GfsLrRr, GfsLrRg, and NamLnRr) predicting a track that worsens as the event progresses, often remaining 150 km to 200 km from the MERRA track during the last 12 hours of the event. The remaining five 36P ensemble member forecasts initialized the cyclone slightly too far south and west but still within 100 km of the MERRA track for most of the event.
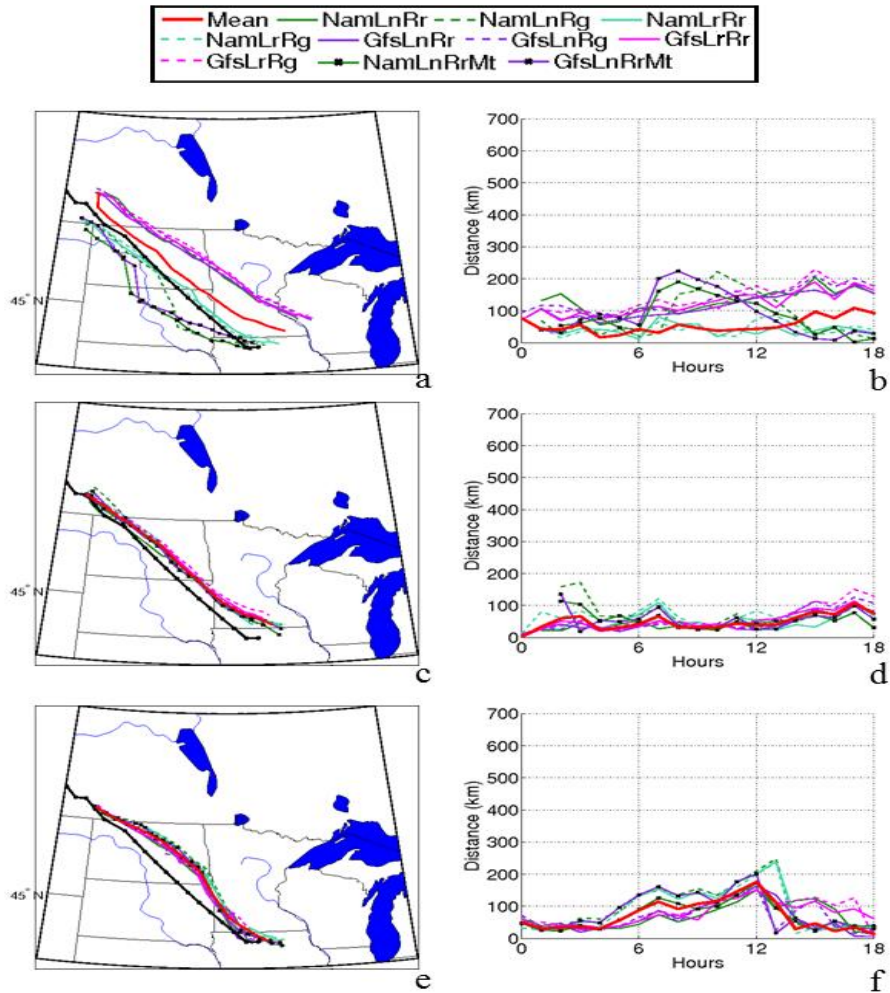
Figure 36. As in Fig. 15 but representing the entirety of the January 2009 18 hour event, starting at 00 UTC 12 January 2009 and ending at 18 UTC 12 January 2009.

Due to the low variability in the 24P and 12P ensemble member track forecasts, it is difficult to determine a best performing member (Fig. 36). Better performing physics schemes and model forcings can be recognized during certain periods in each forecast, but no physics scheme or model forcing consistently performs better throughout all three forecasts of the January 2009 event. For example, members with the NAM forcing performs better on average than members with the GFS forcing during hours 6 to 14 of the event for the 12P and 24P ensemble mean track forecasts and vice versa during hours

73

14 to 18 of the event. For the 36P track forecast, members with the NAM forcing perform better on average than members with the GFS forcing throughout the entire event.

5.1.3 Precipitation Threshold Verification

Figures 37 and 38 depict the variability in the probability of exceeding event total precipitation threshold values of 0.10 inches and 0.25 inches, respectively, for the all three ensemble forecasts and the SNODAS data. The SNODAS data shows a broad area of exceeding 0.25 inches of event total precipitation across southeast North Dakota and far northeast South Dakota (Fig. 28) and a widespread area exceeding 0.10 inches of event total precipitation spanning across an area from northwest and north central North Dakota to eastern Minnesota, Wisconsin, Iowa, and eastern South Dakota (excluding southwest, south central, and far northeast North Dakota and northern Minnesota; Fig. 37).
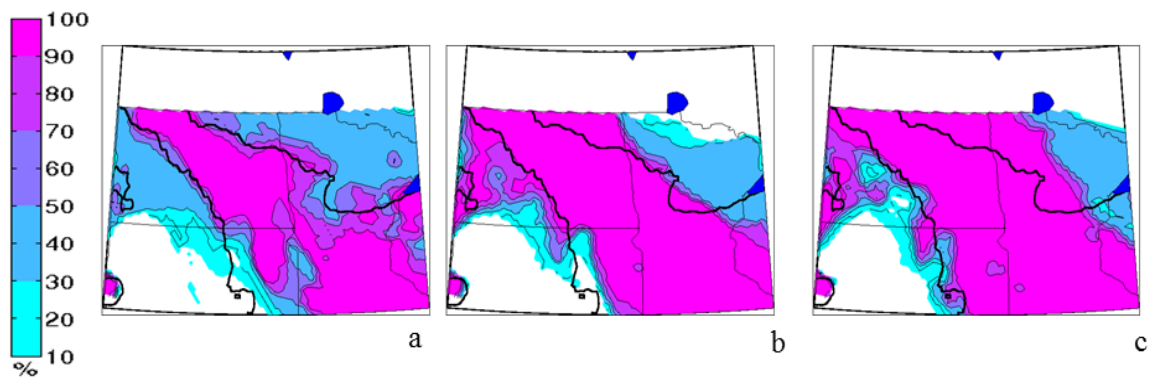
Figure 37. Percentage of January 2009 event members (color fill) exceeding 0.10 inches of event total liquid equivalent snowfall across Domain 2* for the 36P (a), 24P (b), and 12P (c) ensemble forecasts. Area of SNODAS liquid equivalent snowfall shown in black outline. The event total period is described in Chapter 3.

The majority of all of the ensemble member forecasts for the January 2009 event perform well and predict event total precipitation exceeding 0.10 inches over much of the area in which it occurred (Fig. 37). Half of the 36P ensemble members predict the

74

locations of event total precipitation exceeding 0.10 inches almost entirely correctly in

the forecast area (Fig. 37a). However, half of the 36P ensemble members also overpredict

the coverage area by placing additional precipitation over western North Dakota, where

this threshold was not met. On the other hand, all of the 24P and 12P ensemble members

predict the locations of event total precipitation exceeding 0.10 inches correctly, but they

also overpredict the coverage area and place additional precipitation in far northeast

North Dakota, far northwest Minnesota, and parts of western North Dakota (Figs. 37b

and 37c).
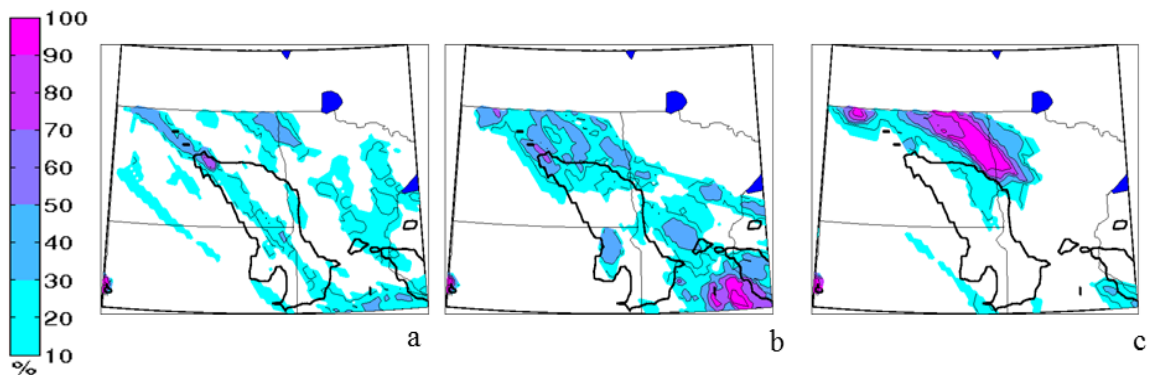
Figure 38. As in Fig. 37 but for percentage of members (color fill) exceeding 0.25 inches
of event total liquid equivalent snowfall.

When predicting the locations of event total precipitation exceeding 0.25 inches,

the three forecasts for the January 2009 event perform poorly (Fig. 38). While 70% of the

36P ensemble members predict some area with event total precipitation greater than 0.25

inches, none of the 36P ensemble members are able to capture the large area in which the

MERRA data depict that it occurred (Fig. 38a). Thirty percent of 36P ensemble members

do predict a swath of event total precipitation greater than 0.25 inches in the correct area,

but this swath is too narrow to accurately represent the broad area over which it occurred.

Both the 36P and 24P forecasts have up to 50% of members predicting isolated to

scattered event total precipitation values exceeding 0.25 inches across far northeast South

Dakota and central and southern Minnesota, but those greater values did not occur at

many of those locations (Figs. 38a and 38b). The 24P forecast performs better than the

36P forecast. Thirty percent of the 24P ensemble members predict a broad swath of event

total precipitation greater than 0.25 inches in north central and northeast North Dakota,

but the swath is depicted too far north (Fig. 28b). The 12P forecast also depicts a swath of

event total precipitation exceeding 0.25 inches too far north but is more confident in this

(Fig. 38c). However the 12P ensemble members do not depict isolated to scattered higher

amounts like the other two forecasts.

Figure 39 shows the difference in coverage area of event total precipitation

exceeding 0.10 inches for each of the three ensemble forecasts and the SNODAS data. In

looking at the event as a whole, all of the 24P and 12P and half of the 36P ensemble

members overpredict the coverage area of event total precipitation exceeding 0.10 inches.

After analyzing both the ensemble means and the values of the individual members in

Fig. 39, the 36P forecasts have less overprediction in coverage on average and perform

slightly better (mean of 14.1%; Fig. 39a) than the 24P forecasts (mean of 22.6%; Fig.

39b).
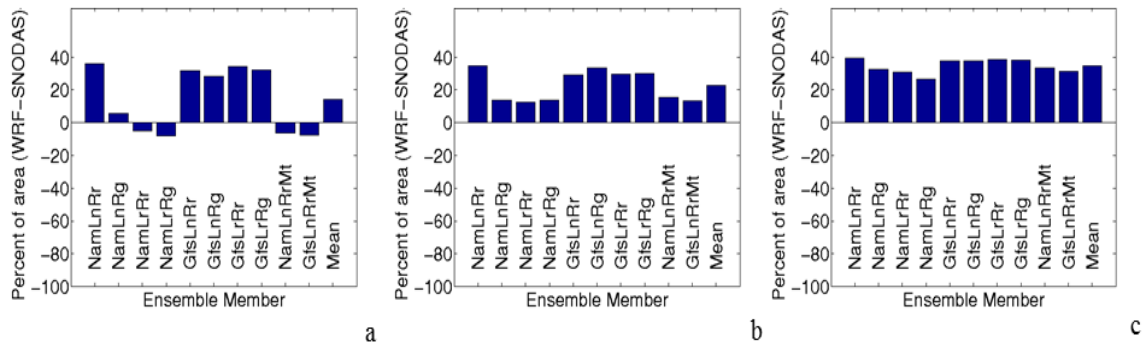
Figure 39. Difference in coverage area of event total precipitation exceeding 0.10 inches
for the 36P (a), 24P (b), and 12P (c) ensemble mean and member forecasts. Event total

precipitation is for the forecast area during the period described in Chapter 3. Note that the area considered is the forecast area (eastern North Dakota and northwestern Minnesota), not Domain 2.

The 12P ensemble forecasts perform worst with ensemble members overpredicting coverage by an average of 34.7% (Fig. 39c). All of the 12P and 24P ensemble member forecasts overpredicted coverage, ranging from 12.4% to 34.7% for the 24P ensemble members (Fig. 39b) and 26.6% to 39.6% for the 12P ensemble members (Fig. 39c). While the 36P ensemble mean difference in coverage was less than the respective 12P ensemble mean, the 36P ensemble member forecasts both overpredicted and underpredicted coverage with ranges of 5.8% to 36.1% and -5.1% to -8.0%, respectively (Fig. 39a). It is interesting to note that the trend remains the same for the ensemble members' physics and forcings. The NamLnRr, GfsLnRr, GfsLnRg, GfsLrRr, and GfsLrRg ensemble members overpredicted by the greatest magnitude for all three forecasts, and the remaining five ensemble members overpredicted by the least magnitude (or underpredicted) for all three forecasts (Fig. 39). Additionally NamLnRr member performed worst for all three forecasts.

5.1.4 Wind Speed Threshold Verification

Appendix F shows the variability in the probability of exceeding wind speed thresholds of 25 mph (21.7 kts) at 09 UTC, 12 UTC, 15 UTC, and 18 UTC 12 January 2009 for the 36P 24P, and 12P ensemble member forecasts and MERRA data in Domain 2*. At 09 UTC 12 January 2009, MERRA data depict wind speeds exceeding 25 mph across a widespread area from southwestern North Dakota to south central South Dakota and isolated areas of east central North Dakota. The MERRA data only show wind speeds exceeding 25 mph west and south of the forecast area for the rest of the event.

77

All three forecasts predict at least a 30-50% chance of wind speeds exceeding 25 mph in southeast North Dakota and west central Minnesota but indicate more uncertainty regarding wind speeds exceeding 25 mph in northeast North Dakota and especially northwest Minnesota (Appendix F). Figure 40 depicts the probability of exceeding 25 mph for the 36P, 24P, and 12P ensemble member wind speed forecasts compared to the MERRA winds at 12 UTC 12 January 2009. All of the forecasts predict wind speeds exceeding 25 mph, both in the locations that the MERRA data show that they occurred and east of those areas (in eastern North Dakota).
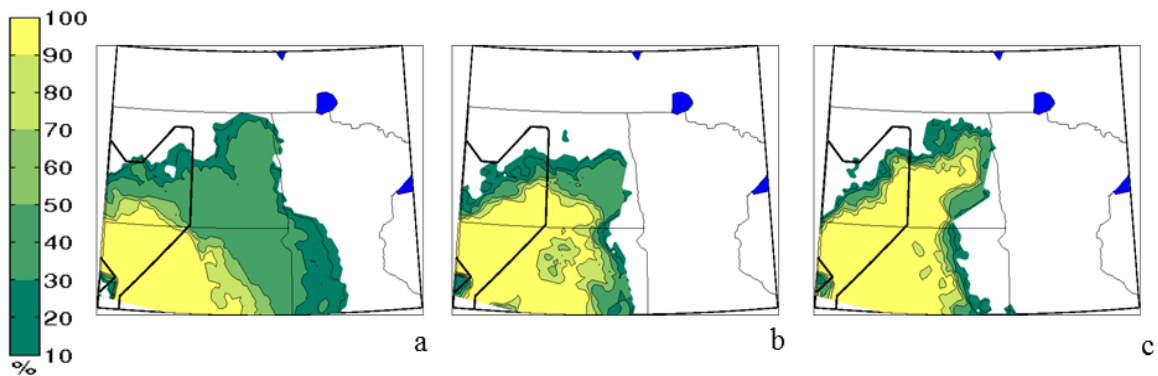
Figure 40. As in Fig. 24 but at 12 UTC 12 January 2009 in Domain 2*.

The 36P ensemble member wind speed forecasts are the weakest of the three forecasts and show 30-50% of members predicting wind speeds exceeding 25 mph in southeast North Dakota and west central Minnesota from 09 UTC to 18 UTC 12 January (Appendix F). All of the 24P and 12P ensemble member forecasts predict an increased chance of wind speeds exceeding 25 mph in parts of eastern and especially southeast North Dakota during the event. The 24P ensemble member forecasts predict up to a 50% chance of wind speeds exceeding 25 mph in northeast and east central North Dakota at 09 UTC and 12 UTC 12 January, but the 12P ensemble members predict a 70-100% chance of these greater wind speeds at those locations and times. The 12P ensemble member

forecasts also predict a 100% chance of wind speeds exceeding 25 mph in southeast

North Dakota at 15 UTC and 18 UTC 12 January.

When compared to the MERRA data, the ensemble member wind speeds

forecasts perform quite poorly and overestimate the wind speed coverage and intensity

during much of the event (Appendix F and Fig. 40). Figure 41 depicts surface station

observations of wind speeds in knots at two times during the January 2009 event and

shows that several observation stations in southeast North Dakota and west central

Minnesota observed wind speeds exceeding 25 mph (21.7 kts). Surface observations over

Domain 2* show wind speeds exceeding 25 mph starting at 13 UTC 12 January 2009 at

several locations in central/east central North Dakota (NCEI 2005 and Fig. 41). These

higher wind speeds develop in portions of eastern North Dakota and northwest Minnesota

until 20 UTC 12 January 2009, when the wind speeds begin to quickly diminish to below

25 mph. Based on Fig. 41 and other automated surface observing station data from NCEI

(2005), it can be concluded that the ensemble wind speed forecasts performed much

better than originally depicted in comparison to the MERRA data. This is likely an

example in which the MERRA data are unable to resolve the magnitude of the peak

sustained wind speeds (Chp. 3). Thus, the magnitude of the peak sustained wind speeds

have likely been smoothed out, as shown in Fig. 27 and discussed in Chapter 4.
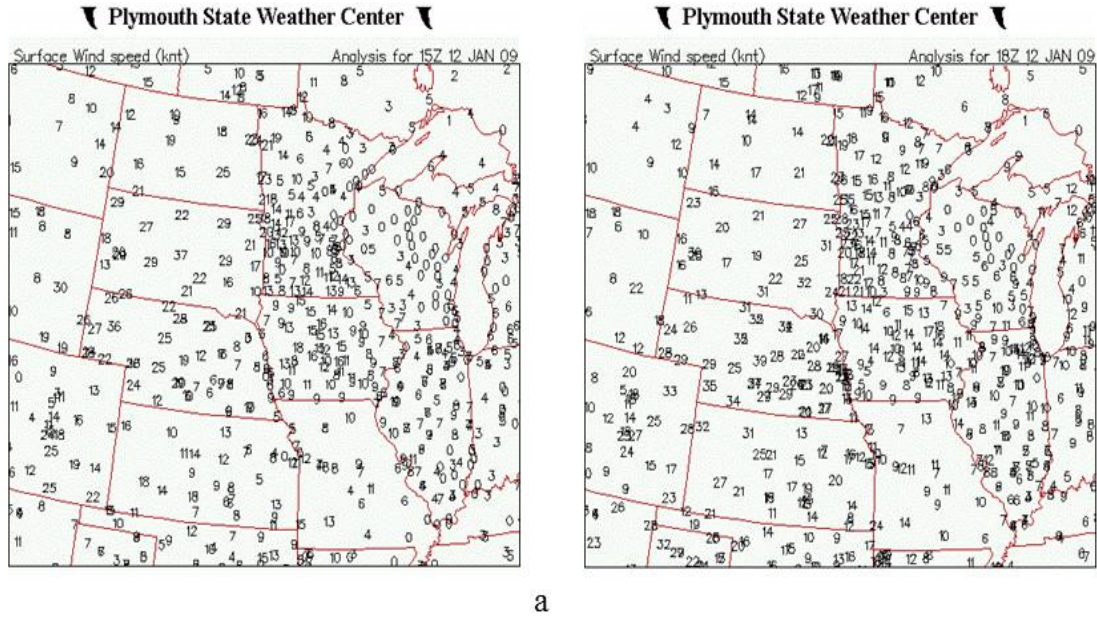
Figure 41. Plymouth State University Weather Center plotted surface data map of observed surface wind speed in knots at (a) 15 UTC and (b) 18 UTC 12 January 2009.

Although MERRA is underpredicting magnitude for this case, plotting differences in areal coverage is still valuable for assessing variability in the forecasts. Figure 42 quantifies the difference in area (from the MERRA coverage) of wind speeds exceeding 25 mph for the 36P, 24P, and 12P forecasts of the January 2009 event. While all of the ensemble member wind speed forecasts predicted coverage within 40% of the MERRA data, it is difficult to draw conclusions regarding performance of specific physics schemes when the performance differs so greatly from forecast to forecast. Some members do not predict wind speeds exceeding 25 mph in the forecast area, which agrees with the MERRA data but does not agree with actual surface observations (Figs. 40-42). The 36P and 24P ensemble member forecasts indicate uncertainty in predicting the wind speed throughout the event, and both show underestimation and overestimation of coverage for wind speeds exceeding 25 mph.
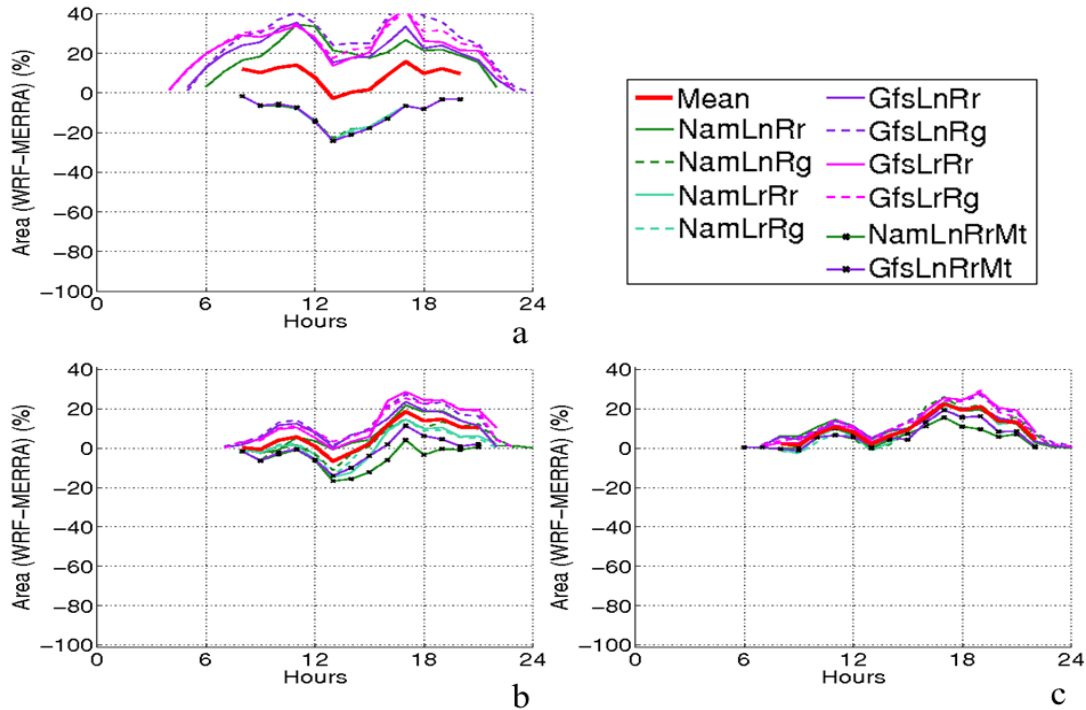
Figure 42. As in Fig. 25 but during the 24 hour period of 00 UTC 12 January to 00 UTC 13 January 2009. Note that the area considered is the forecast area (eastern North Dakota and northwestern Minnesota), not Domain 2.

Appendix F shows that the 36P and 24P ensemble member wind speeds forecasts only show 50% of members or less predicting wind speeds exceeding 25 mph at 9 and 12 hours into the event, and these are the members that overestimate coverage compared to the MERRA data in Figs. 42a and 42b. However several members underestimate coverage because they do not predict an area of higher wind speeds as far north as the MERRA data indicate (i.e., small area in north central North Dakota in Fig. 42). A slight drop in overestimation (increase in underestimation) occurs at 13 to 15 hours into the event, when ensemble members from all three forecasts predict a drop in the coverage area of wind speeds exceeding 25 mph. The 12P ensemble member wind speed forecasts primarily overestimate coverage, especially in central North Dakota around 09 UTC 12

81

January 2009 and in southeast North Dakota at 15 UTC to 18 UTC 12 January 2009 (Appendix F and Fig. 42).

5.1.5 Ensemble Spread and Mean Performance

Figures 36a, 36c, and 36e show very high predictability (relative to the MERRA track) in the 36P, 24P, and 12P ensemble member track forecasts. Most of the 36P and 12P track forecasts are consistently within 100 to 200 km of the MERRA track, and most of the 24P track forecasts are consistently less than 100 km from the MERRA track. The 36P member track forecasts show a moderate amount of spread between the 36P members with a GFS forcing and WSM6 microphysics (consistently north of the MERRA track), 36P members with the NAM forcing and RUC LSM (very close to the MERRA track), and 36P members with Thompson microphysics (consistently south of the MERRA track; Fig. 36a). The 36P members with the Thompson microphysics lose accuracy quickly from 06 UTC to 07 UTC 12 January 2009 when their track forecasts dive south, but the track forecasts improve through the rest of the event (Figs. 36a and 36b). The best performing 36P member track forecasts are from the NamLrRr and NamLrRg members, which consistently perform nearly as well as or more accurately than the 36P ensemble mean track. The remaining 36P members lose accuracy (relative to the MERRA track) throughout the event. Conversely, the 24P and 12P member track forecasts show such a small amount of spread and it is very difficult to distinguish differences in accuracy amongst the members (Fig. 36). While it can be concluded that the 24P members perform best overall, there is no clear best performing 24P member track forecast.

Although the track forecasts performed quite well for the January 2009 case, the coverage forecasts are as accurate as would be expected (Figs. 39 and 41). Figures 39 and 41 indicate moderate to large spread in the 36P and 24P precipitation and wind speed coverage forecasts and a small amount of spread in the 12P precipitation and wind speed coverage forecasts. A majority of the ensemble members and means overpredict precipitation coverage for the 0.10 inches threshold, except for some of the best performing and worst performing 36P members, which slightly underpredict precipitation coverage (Fig. 39). A majority of the ensemble members and means overpredict wind speed coverage at the 25 mph threshold, except for the 36P and 24P members with Thompson microphysics (Fig. 41). However, as discussed, the true performance of the wind forecasts is difficult to assess in this case due to underprediction in the MERRA data.

Regarding the performance of the ensemble mean compared to the individual members, the mean track forecast only performs consistently better than most of the individual member forecasts for the 36P track forecast (Fig. 36). The small spread in the 24P and 12P individual member track forecasts makes it particularly difficult to identify any best performing member or to determine if the ensemble mean performed best for both forecasts. The overprediction of precipitation and wind speed coverage in many of the ensemble member forecasts greatly limits the accuracy of the mean forecasts (Figs. 39 and 41). The NamLnRg, NamLrRr, NamLrRg, NamLnRrMt, and GfsRrMt members tend to perform similar to or better than the ensemble means for these fields.

5.2 December 2013 Case Results

5.2.1 Synoptic Overview

During the evening of 27 December 2013, an Alberta clipper propagated southeastward from the Canadian Rockies. For a timeline comparing the ensemble forecasts and the event for the December 2013 case, refer to Fig. 43. Figures 44-46 depict the development of the clipper for the 300 hPa, 500 hPa, 850 hPa, and surface levels at 00 UTC 28 December, 12 UTC 28 December, and 00 UTC 29 December, respectively. The initial synoptic pattern over North America consisted of a developing 500 hPa shortwave trough over the Pacific Northwest, a weak ridge extending from Wyoming to southern California, a weakening longwave trough departing the Northeast, and an upper level closed-off cyclone over the West Texas and Northern Mexico (Fig. 44).
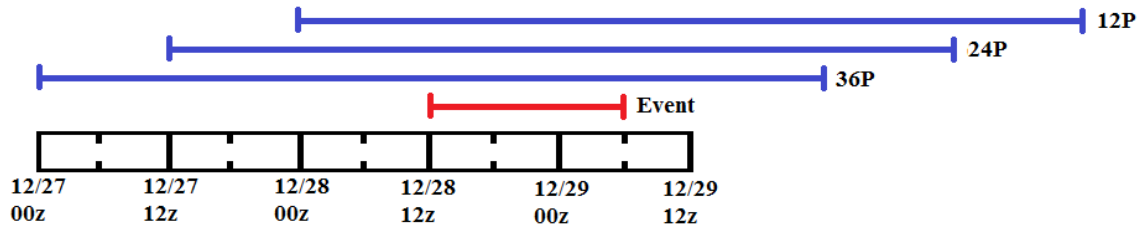
Figure 43. As in Fig. 3 but for the December 2013 case.

Overnight, the clipper progressed rapidly across the Northern Great Plains. In the early morning hours, NWS Grand Forks issued a winter weather advisory stating that snow accumulations of one to two inches and strong winds with gusts up to 40 mph were expected in the forecast area with near blizzard conditions possible at times (NWS Grand Forks 2013a). By 12 UTC 28 December, the clipper was centered over southeast North Dakota (Fig. 45). As a surface high pressure system built over Saskatchewan, a strong pressure gradient developed upstream of the clipper through the morning hours. At the same time, a jet streak developed at 300 hPa with concomitant digging of the 500 hPa

shortwave trough over the Pacific Northwest. At 1621 UTC 28 December, NWS Grand

Forks issued a blizzard warning, as gusts up to 50 mph were now expected in at least the

northern half of the forecast area throughout the afternoon (NWS Grand Forks 2013a).
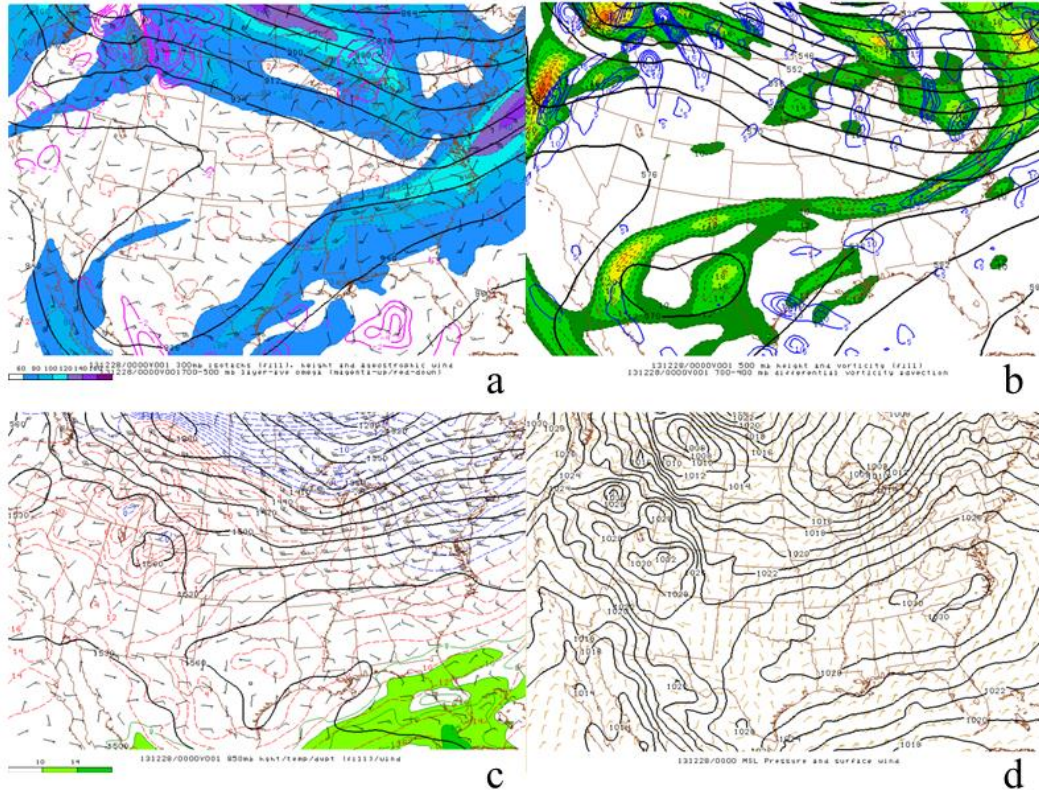
Figure 44. As in Fig. 11 but at 0000 UTC 28 December 2013.
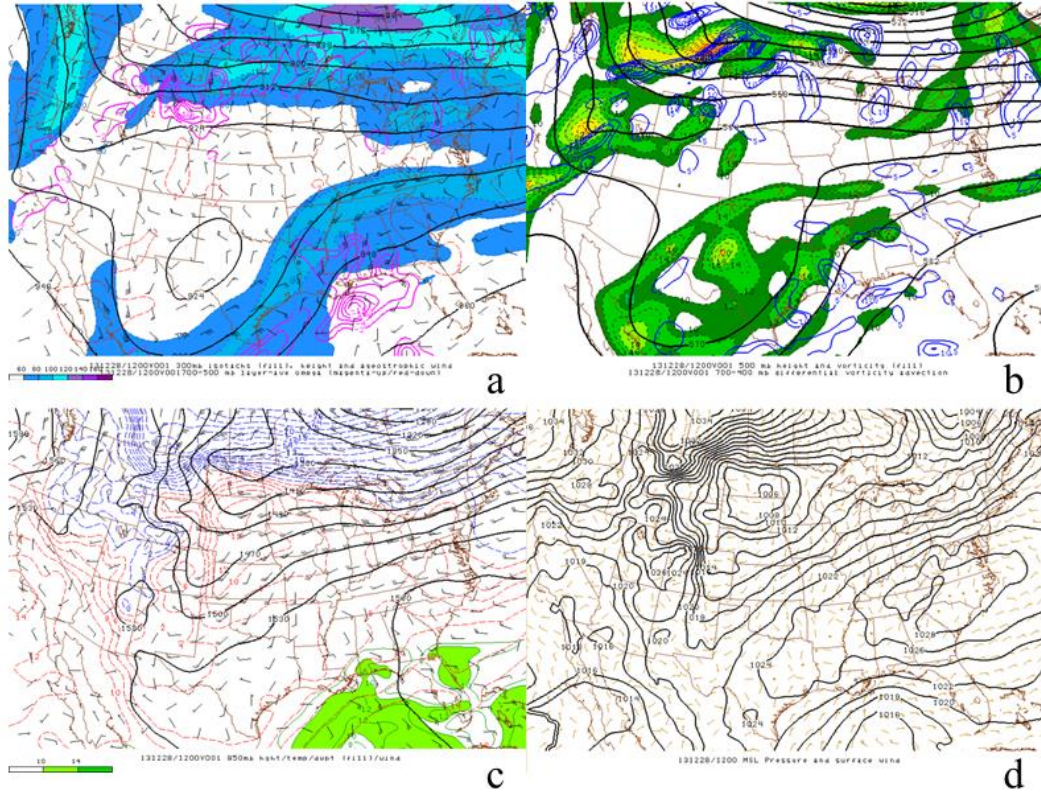
Figure 45. As in Fig. 11 but at 1200 UTC 28 December 2013.

By 00 UTC 29 December, the clipper was centered over southeast Minnesota, and strong wind gusts and light snowfall were beginning to weaken over the forecast area (Fig. 46). The 500 hPa shortwave trough became a positively tilted longwave trough stretching from the forecast area to southern California. By 12 UTC 29 December, blizzard conditions had ceased as a high pressure system moved into the forecast area. Event total snowfall across the forecast area ranged from a trace to two inches.
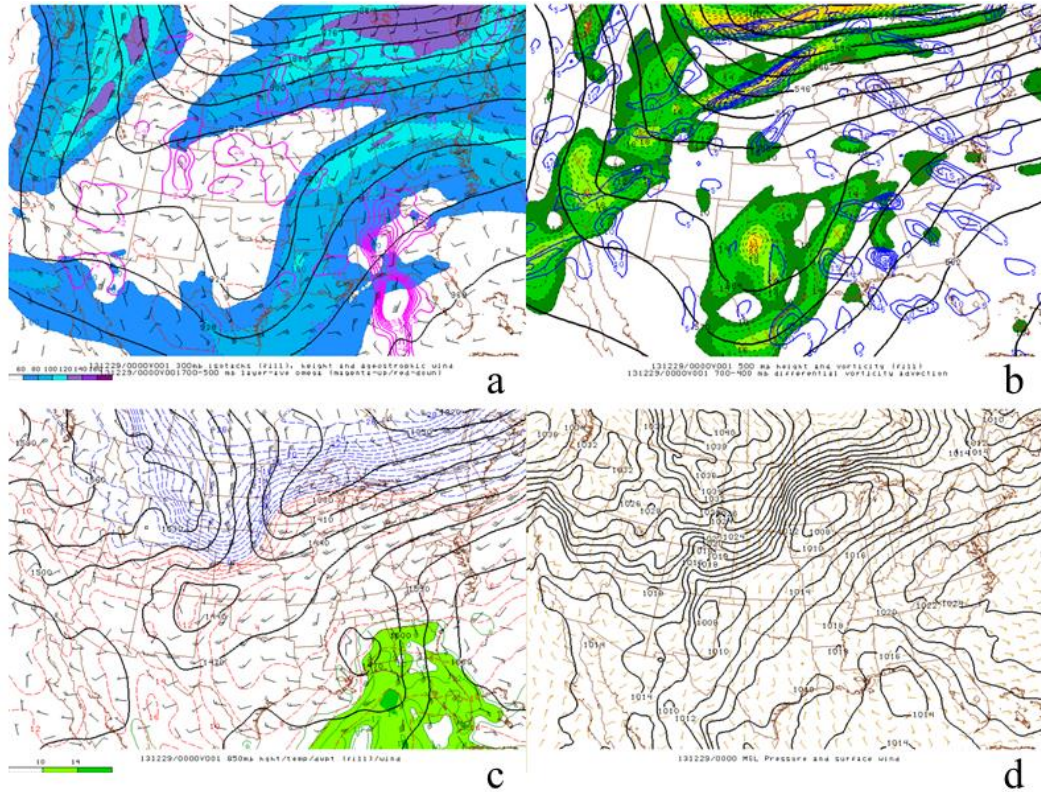
Figure 46. As in Fig. 11 but at 0000 UTC 29 December 2013.

5.2.2 Ensemble Mean and Member Track Verification

Figure 47a shows the MERRA and ensemble mean cyclone center track forecasts in Domain 2 throughout the December 2013 event. All three ensemble mean track forecasts depict a surface cyclone starting in Saskatchewan and progressing southeastward into northwest, north central, and southeast North Dakota. As the ensemble mean track forecasts pass through far northeast South Dakota, the forecasts diverge. The 12P ensemble mean track forecast progresses slightly northeast and then directly eastward through central Minnesota, ending in northeast Wisconsin. The 36P ensemble mean track forecast propagates directly eastward through central Minnesota and ends in northeast Wisconsin. The 24P ensemble mean track forecast continues southeast into south central Minnesota and then turns eastward, ending in central

87

Wisconsin. The MERRA track portrays an initial surface cyclone center in Saskatchewan, but the track quickly "jumps" to west central South Dakota. This indicates a brief shift from the primary cyclone center to a secondary center. (For further analysis and discussion of a similar shift that occurred in the March 2011 event, refer to Chapter 4.) By 12 UTC 28 December 2013, the MERRA track had jumped northeast and was located along the North Dakota/South Dakota border, in close proximity to the ensemble mean track forecasts again. The MERRA track continued southeast into far northeast South Dakota and southwest Minnesota, where it turned eastward and continued through southern Minnesota and into central Wisconsin.
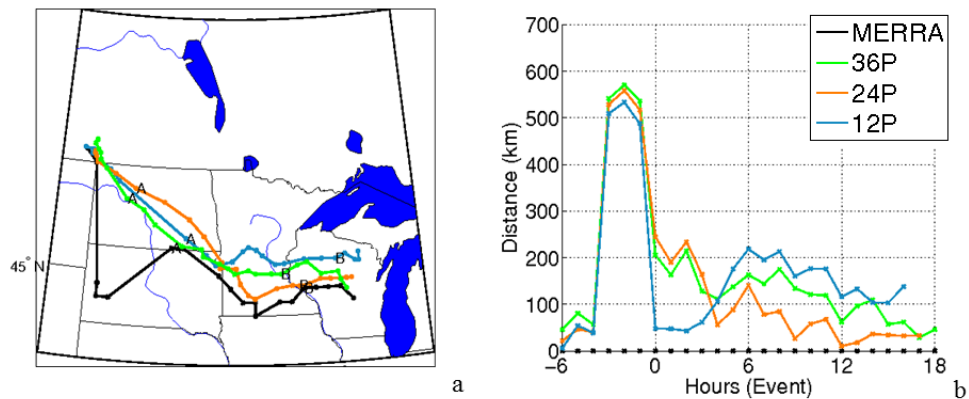
Figure 47. As in Fig. 14 but representing the period from 06 UTC 28 December 28 2013 to 06 UTC 29 December 29 2013. Hours 12 UTC 28 December 2013 and 00 UTC 29 December 2013 are indicated by A and B, respectively.

Figure 47b shows the absolute error in distance from the MERRA track, relative to the event time, for the ensemble mean forecasts. The greatest absolute difference in distance is nearly 600 km and occurs at 10 UTC 28 December 2009, just prior to the start of the event. This peak in track error is clearly due to the MERRA cyclone center shift (mentioned previously). During the event, the ensemble mean track forecasts are within 250 km of the MERRA track. All three ensemble mean track forecasts exhibit a northern

88

track bias throughout the December 2013 event (Fig. 47a). Initially, the 12P ensemble mean track forecast performs well, but around 16 UTC 28 December it shifts northward for a few hours and continues to depict the most northern track bias for the remainder of the event (Fig. 47b). After 16 UTC 28 December, the 24P ensemble mean track forecast continues to propagate southeast, in a manner similar to the MERRA track. It remains within 100 km of the MERRA track for the last 12 hours of the event and, thus, it performs better than the other ensemble mean track forecasts.

Figure 48 shows all of the ensemble mean and member track forecasts and their corresponding absolute distances from the MERRA track during the December 2013 event. Although all of the ensemble member track forecasts correctly depict the initialization in Domain 2, there is great variability amongst the 36P and 24P individual ensemble member track forecasts throughout the event. Both forecasts depict ensemble member tracks north and south of the MERRA track. Although the 36P ensemble mean track forecast is within 200 km of the MERRA track during the event, only a few of the 36P member track forecasts (with GFS forcing and WSM6 microphysics) are consistently close to the MERRA track (Figs. 48a and 48b). The 36P ensemble member track forecasts with Thompson microphysics divert quickly southeastward and then eastward away from the MERRA track. They propagate from northwest North Dakota to eastern North Dakota, through northern Minnesota, and into far northern Wisconsin and have a northern bias of at least 200 km during most of the event. Additionally, the 36P members with a NAM forcing and the RUC LSM shift in direction from hour to hour, as if they are shifting between cyclone centers, and remain 100 to 300 km from the MERRA track throughout most of the event.

89
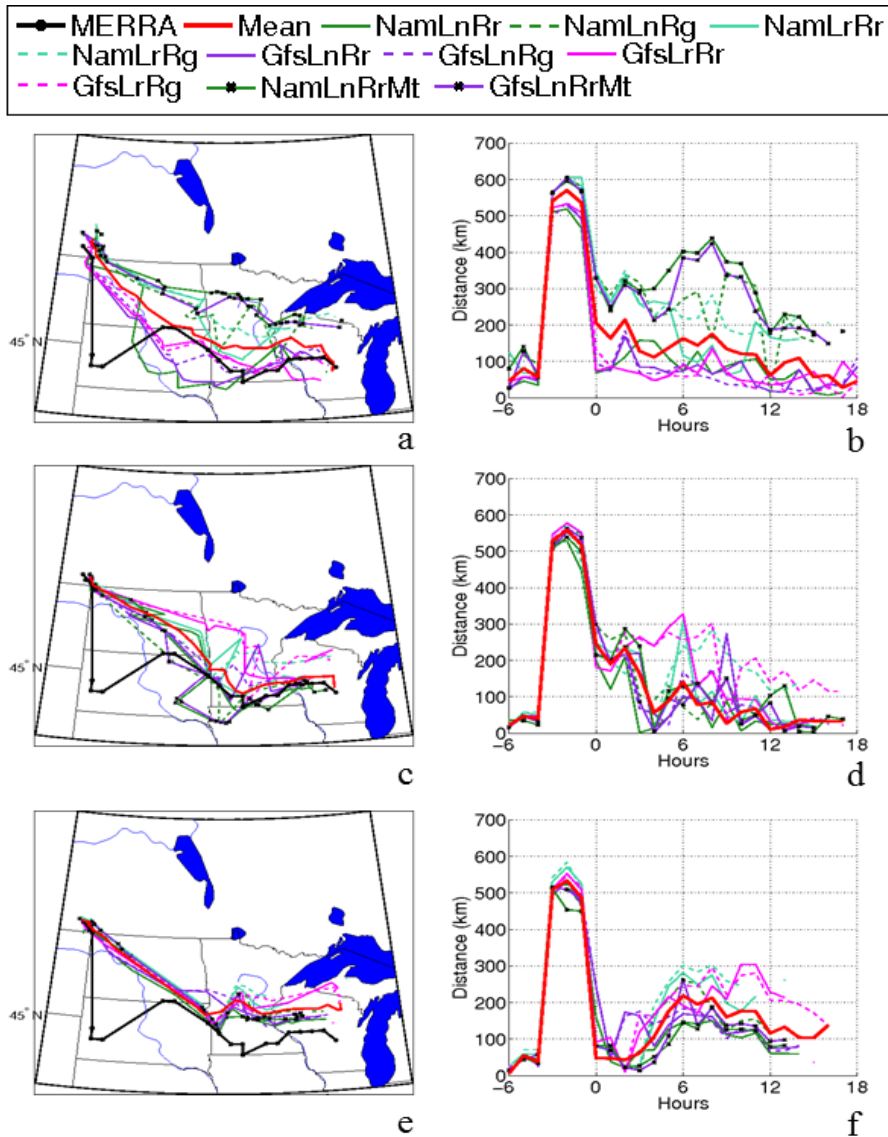
Figure 48. As in Fig. 15 but all figures represent the period from 06 UTC 28 December 28 2013 to 06 UTC 29 December 29 2013.

The 24P members perform better than the 36P and 12P members on average (Fig. 48). Unlike in the 36P ensemble member track forecasts, the 24P members with the Thompson microphysics for the December 2013 event remain within 150 km of the MERRA track during most of the event. However members with Thompson microphysics are still outliers in the 24P forecasts and have a bias south of the MERRA track. The 24P and 12P ensemble member track forecasts with the NOAH LSM tend to

perform better than those with the RUC LSM and stay within 200 km of the MERRA track during the last 12 hours of the event. While the 12P ensemble mean track forecasts perform worst (on average) of the three ensemble mean track forecasts, the 12P individual ensemble member track forecasts perform best until they begin to shift erroneously northward away from the MERRA track a few hours into the December 2013 event.

5.2.3 Precipitation Threshold Verification

Figures 49 and 50 depict the variability in the probability of exceeding event total precipitation threshold values of 0.10 inches and 0.25 inches, respectively, for the all three ensemble forecasts and the SNODAS data. The SNODAS data show a swath precipitation exceeding 0.10 inches of event total precipitation across most of northern Minnesota and scattered across portions of northeast North Dakota (Fig. 49) and no locations in the forecast area with event total precipitation exceeding 0.25 inches (Fig. 50).
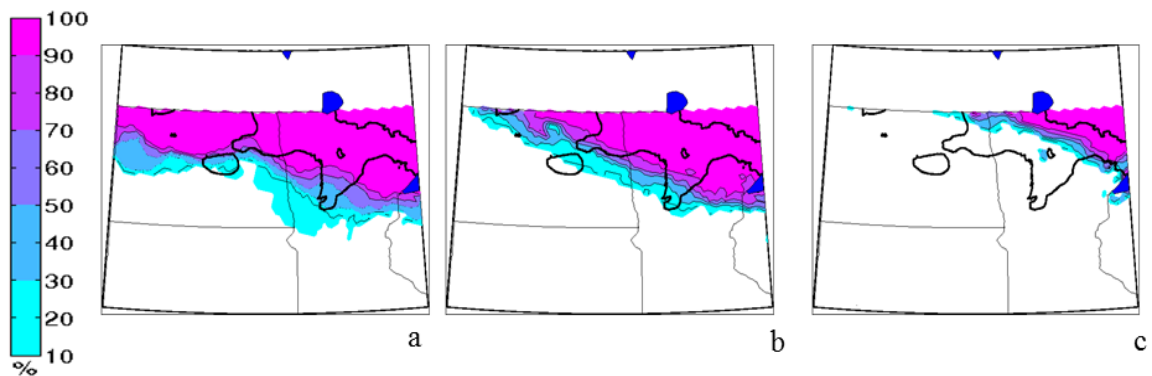
Figure 49. As in Fig. 16 but during the event total period for the December 2013 case as described in Chapter 3.

The 36P and 24P ensemble member precipitation forecasts for the December 2013 event perform very well and predict event total precipitation exceeding 0.10 inches

over much of the area in which it occurred (Fig. 49). All of the 36P and 24P ensemble

members correctly predict precipitation over far northeast North Dakota and northern

Minnesota, although many members struggle to capture the precise edges of the snow

band. The 12P ensemble member precipitation forecasts performed poorly and depicted

the swath of precipitation too far east (and likely north of the United States/Canada

border where SNODAS data were unavailable, as discussed in Chapter 3).

Figure 50. As in Fig. 17 but during the event total period for the December 2013 case as described in Chapter 3.

Assessing higher precipitation amounts, the 24P and 12P ensemble member

precipitation forecasts performed best for the December 2013 event and depicted little to

no event total precipitation exceeding 0.25 inches in the forecast area (Fig. 50). A few of

the 36P ensemble member precipitation forecasts incorrectly overestimated the

southward reach of a swath of event total precipitation exceeding 0.25 inches extending

into the forecast area. The majority of the 36P ensemble member precipitation forecasts

predicted over 0.25 inches in far northeast Minnesota and into Ontario (i.e., outside of the

forecast area), but this is a southward bias based on data from NOHRSC (2015).

SNODAS modeled snow precipitation from the daily driving data (NOHRSC 2015)

indicates precipitation ranging from 0.20 to 0.39 inches did occur just north of the

Minnesota/Ontario border. Even though all three forecasts predicted event total

precipitation exceeding 0.25 inches just north of the Minnesota/Ontario border, the

SNODAS data used in this study would be unable to reflect those values (as described in

Chapter 3).

Figure 51 shows the difference in coverage area of event total precipitation

exceeding 0.10 inches for each of the three ensemble forecasts and the SNODAS data. In

looking at the event as a whole, all of the 36P members overestimate the coverage area of

event total precipitation exceeding 0.10 inches, with a range of 2.1% to 33.4% and an

average of 13.6%. Most of the 24P ensemble members overestimate the coverage area of

event total precipitation exceeding 0.10 inches, with a range of 2.2% to 22.6%, but two

members underestimate the coverage by 1.1% and 5.4%. The 24P ensemble members

overestimate coverage by 7.6% on average . The 12P ensemble members perform worst

and underestimate the coverage area by an average of 26%. After analyzing both the

ensemble means and the values of the individual members coverage differences, the 24P

forecasts predict coverage closest to the SNODAS data for event total precipitation
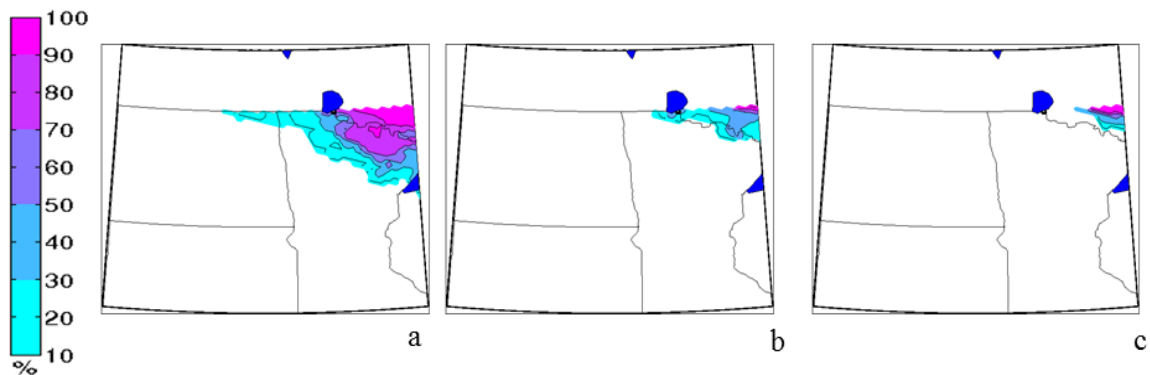
exceeding 0.10 inches.
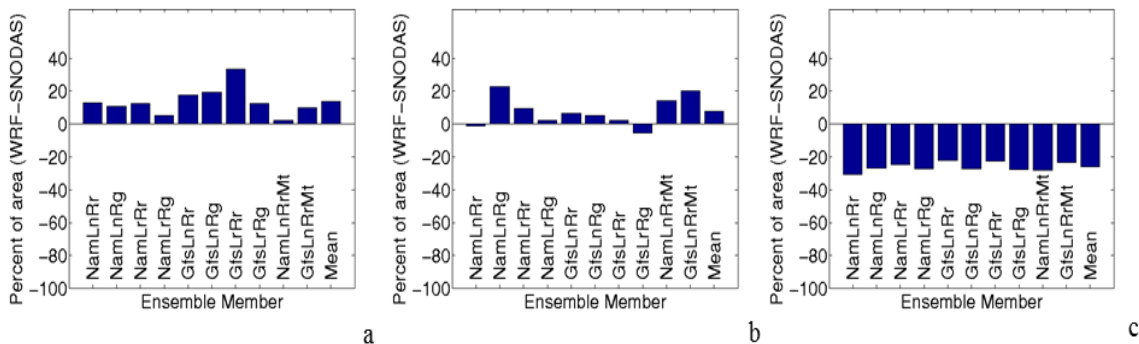
Figure 51. As in Fig. 18 but during the event total period for the December 2013 case as described in Chapter 3.

With very little spread in the ensemble member event total precipitation from forecast to forecast, it is difficult to determine best performing members or physics schemes (Figs. 49-51). The 36P GfsLrRr member, 24P NamLnRgmember , and 12P event total precipitation forecasts performed worst, and the 24P NamLnRr member and 24P members with GFS forcing and WSM6 LSM performed best (Fig. 51).

5.2.4 Wind Speed Threshold Verification

Appendix G shows the variability in the probability of exceeding wind speed thresholds of 25 mph (21.7 kts) at 21 UTC 28 January and 00 UTC, 03 UTC, and 06 UTC 29 December 2013 for the 36P, 24P, and 12P ensemble member forecasts and MERRA data in Domain 2*. The MERRA data depict wind speeds exceeding 25 mph developing across western North Dakota from 21 UTC 28 December to 00 UTC 29 December and then diminishing below 25 mph for the remainder of the event.

Figure 52 depicts the probability of exceeding 25 mph for the 36P, 24P, and 12P ensemble member wind speed forecasts compared to the MERRA winds at a valid time representative of the overall ensemble's wind speed forecast performance. Although MERRA data only depict wind speeds exceeding 25 mph in central North Dakota, eastern North Dakota is where the ensemble members predict those higher wind speeds are more likely to occur (Fig. 52 and Appendix G). The 36P ensemble member forecasts have 70-100% of members predicting wind speeds exceeding 25 mph in northeast and east central North Dakota at 21 UTC 28 December 2013 (Appendix G). The forecasts show this area of highest probability of higher winds gradually shifting southward along the North Dakota/Minnesota border throughout the event, exiting the forecast area into eastern South Dakota and west central Minnesota between 03 UTC and 06 UTC 29 December.

94

The 24P ensemble member wind speed forecasts are the weakest of the three forecasts. They only show 30-50% of members predicting wind speeds exceeding 25 mph in eastern North Dakota until around 00 UTC 29 December, when 70-100% of members predict wind speeds exceeding 25 mph until around 03 UTC 29 December in far southeast North Dakota. The 12P ensemble member wind speed forecasts are similar to the 36P ensemble member wind speed forecasts, but predict a slightly larger area of wind speeds exceeding 25 mph from 21 UTC 28 December to 03 UTC 29 December 2013.
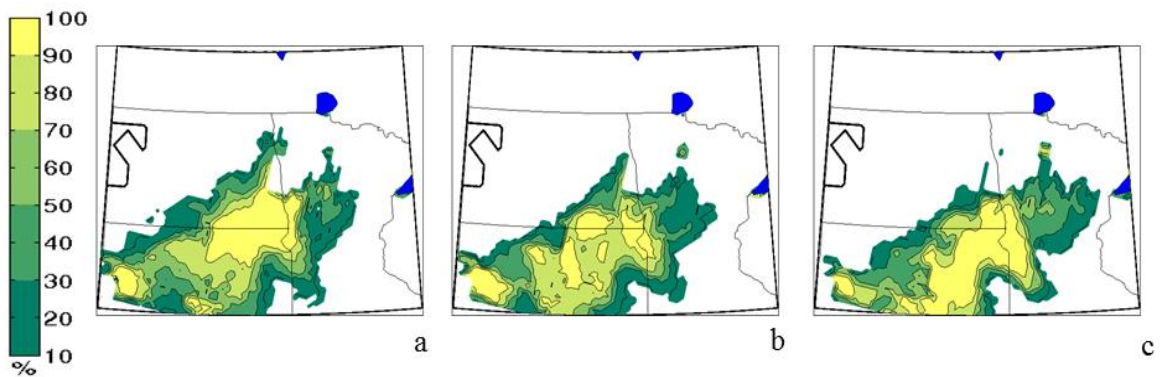
Figure 52. As in Fig. 19 but at 00 UTC 29 December 2013.

When compared to the MERRA data, the ensemble member wind speed forecasts perform quite poorly and overestimate the wind speed coverage and intensity during much of the event (Appendix G). Despite the MERRA data not showing development of wind speeds exceeding 25 mph across the forecast area for much of the December 2013 event, Fig. 53 shows automated surface observations over Domain 2* with wind speeds exceeding 25 mph (21.7 kts) at 18 UTC 28 December 2013 and 00 UTC 29 December 2013. These higher wind speeds values started at 16 UTC 28 December 2013 at several locations in northeast North Dakota (Fig. 53 and NCEI 2005). These higher wind speeds develop throughout eastern North Dakota and northwest Minnesota until 06 UTC 29

December 2013, when the wind speeds begin decreasing to below 25 mph. Peak gusts

ranging from 40 to 55 mph were recorded across the forecast area (NWS 2013b).
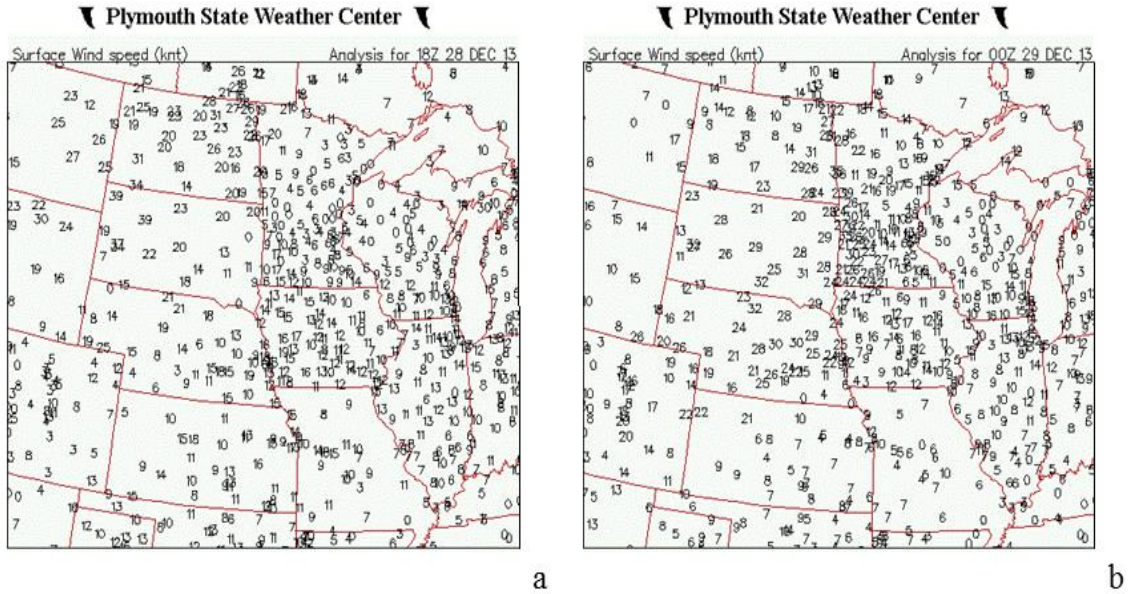
Figure 53. As in Fig. 41 but at (a) 18 UTC 28 December and (b) 00 UTC 29 December 2013.

Figure 54 illustrates the difference in area (from the MERRA coverage) of wind

speeds exceeding 25 mph for the 36P, 24P, and 12P forecasts of the December 2013

event. All of the ensemble member wind speed forecasts predicted coverage within 40%

of the MERRA data. The 36P ensemble member forecasts tended to underestimate

coverage of wind speeds exceeding 25 mph from 20 UTC 28 December until two to four

hours later and then overestimated coverage for the final six hours of the event (Fig. 54a).

Most of the 24P members underestimate coverage from 19 UTC 28 December until four

to eight hours later and then overestimate coverage for the remainder of the event (Fig.

54b). The 24P GfsLnRg and GfsLrRg members overestimate coverage through the entire

event, and the two 24P members with the Thompson microphysics underestimate

coverage through almost the entire event. The 12P ensemble members are split based on

the radiation scheme of the members (Fig. 54c); the four members with the GFDL

radiation scheme overestimate coverage of wind speeds exceeding 25 mph through most

of the event, and the six members with the RRTMG radiation scheme underestimate

coverage of wind speeds exceeding 25 mph through most of the event. In fact, when

looking at all three forecasts, members with the GFDL radiation scheme tend to predict a

larger area of coverage of wind speeds exceeding 25 mph than members with the

RRTMG radiation scheme throughout most of the December 2013 event. In the 24P and

12P forecasts, members with the Thompson microphysics also tend to predict a smaller

area of coverage of wind speeds exceeding 25 mph than members with WSM6

microphysics throughout most of the event.
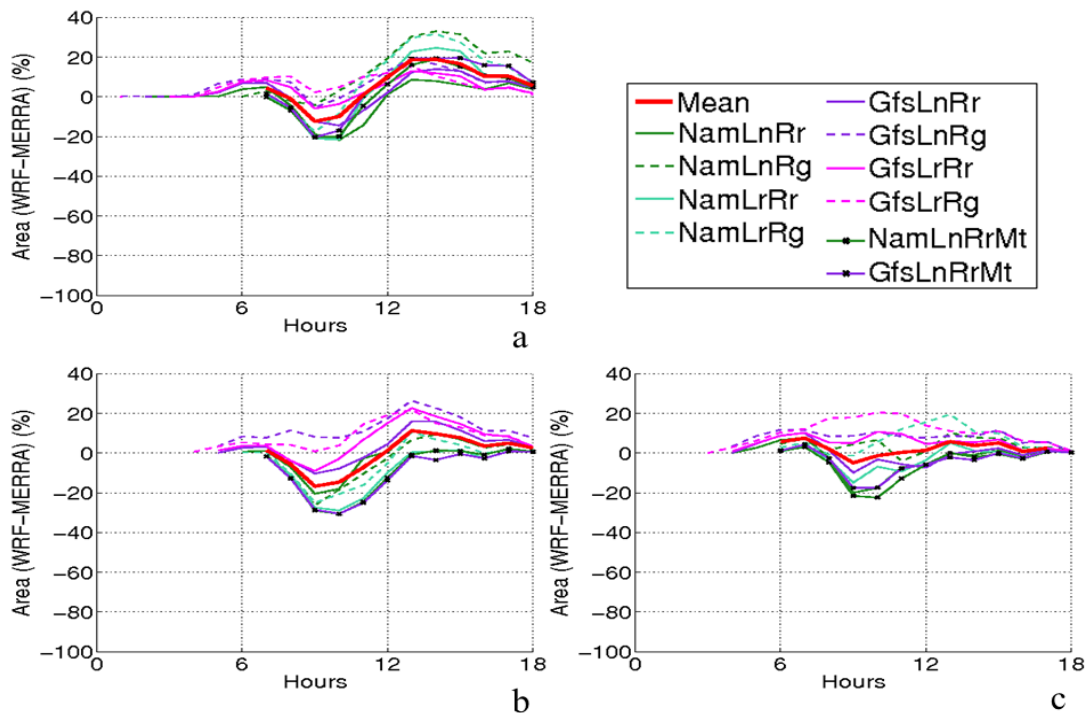
Figure 54. As in Fig. 20 but during the 18 hour period of 12 UTC 28 December to 06 UTC 29 December 2013.

97

5.2.5 Ensemble Spread and Mean Performance

Figures 48a and 48c show low predictability (relative to the MERRA track) in the 36P and 24P ensemble member track forecasts. The 36P track forecasts are within about 25 km to 450 km of the MERRA track, and the 24P track forecasts are generally within 350 km of the MERRA track (Figs. 38b and 38d). The 36P and 24P track forecasts show large spread. Figures 38e and 38f show moderate predictability (relative to the MERRA track) in the 12P ensemble member track forecasts, and the 12P track forecasts remain within about 25 to 300 km of the MERRA track. The 12P track forecasts show relatively small spread. All three forecasts show spread increasing during the middle of the event and then decreasing again toward the end.

Despite a relatively large spread in the ensemble track forecasts, the coverage forecasts have small spread and limited accuracy (Figs. 50 and 53). Figures 50 and 53 indicate small to moderate spread in all three forecasts' precipitation coverage. The 36P precipitation forecasts overpredict the precipitation coverage (Fig. 50a). Most of the 24P precipitation forecasts overpredict the precipitation coverage, but two 24P members underpredict precipitation coverage (Fig. 50b). Figure 52 indicates that the wind speed forecasts generally have small spread near the start and end of the event and moderate spread in the middle of the event (coinciding with the greater spread in track forecast locations). Each forecast has members with wind speed forecasts that underpredict and members with wind speed forecasts that overpredict coverage.

Regarding the performance of the ensemble mean compared to the individual members, the mean track, precipitation, and wind speed forecasts do not perform consistently better than individual members for any of the three forecasts (Figs. 48, 50,

and 53). While certain members and physics schemes tend to perform better than the ensemble mean at times, there is no clear trend for any of the forecast fields.

5.3 Overall Ensemble Performance for Three Cases

The primary goal of this study was to assess whether this ensemble could correctly predict blizzard conditions. The NWS traditionally defines a blizzard warning as sustained wind or wind gusts of 35 mph or greater with considerable falling and/or blowing snow that frequently reduces visibility to a quarter of a mile or less for a minimum three hours. This traditional NWS definition complicates verification methods due to the need to meet criteria for multiple variable fields over a certain period. While wind gust and visibility data are available in AWOS and ASOS data, they are not readily accessible in either of the verification datasets (MERRA or SNODAS) or the ensemble output. It is also difficult to quantify the frequency of falling and/or blowing snow and its frequency using these datasets. Thus, for this study's purposes, blizzard verification is focused on sustained winds and precipitation.

In addition to the requirement of multiple variable fields, it has been observed by the author that the occurrence of 35 mph sustained winds is uncommon in the Northern Plains. In fact, AWOS and ASOS observations indicate that sustained winds exceeding 35 mph rarely occurred in the forecast area for the January 2009 and December 2013 events (e.g., Figs. 41 and 53; additional times not shown). Since the March 2011 case was the only analyzed case with scattered to widespread sustained winds exceeding 35 mph, the ensemble significantly underpredicts coverage of traditional NWS blizzard conditions for this specific event.

99

During all three events, AWOS and ASOS observations indicate that frequent wind gusts exceeding 35 mph often coincided with sustained winds of 25 to 30 mph. For this reason and others discussed in Chapter 4, it is valuable to assess the ensemble performance regarding blizzard conditions using 25 mph sustained winds as a proxy for gusts of at least 35 mph. The use of this proxy for wind gusts is combined with event total precipitation exceeding 0.10 inches as a proxy for falling snow occurring frequently for at least three hours, and these criteria are referred to as North Dakota proxy blizzard criteria (hereafter NDP blizzard criteria).

There is significant improvement in ensemble performance when using NDP blizzard criteria compared to the traditional NWS blizzard criteria, but the ensemble wind forecasts still do not predict blizzard conditions in all of the locations that they were observed. For the March 2011 case, half of the 12P members meet NDP blizzard criteria for precipitation (i.e., event total precipitation exceeding 0.10 inches), and a few of the 36P and 24P members and half of the 12P members meet NDP blizzard criteria for wind (i.e., sustained wind exceeding 25 mph). For the January 2009 case, nearly all members meet NDP blizzard criteria for precipitation, and a few of the 36P, half of the 24P, and most of the 12P members meet NDP blizzard criteria for wind. For the December 2013 case, many of the 36P and 24P members meet NDP blizzard criteria for precipitation and a few of the members meet NDP blizzard criteria for wind.

Further work is needed in order to improve the ensemble and its verification methodology before operational use can be recommended. Nonetheless, the ensemble performed fairly well in predicting the three blizzard events using NDP blizzard criteria compared to ground-truth observations (i.e., ASOS and AWOS data). Verification

analysis of the ensemble performance was greatly impacted by the accuracy of the MERRA data, track bias, and the presence of multiple SLP minima. These challenges, discussion of the overall member performance and spread, and additional recommendations for future work are discussed in detail for the remainder of this section.

5.3.1 Impacts of MERRA Data on Wind Speed Verification

When utilizing the MERRA data to perform ensemble surface wind forecast verification without the context of surface station observations, it appears that the ensemble member wind forecasts perform poorly in predicting the locations and intensities of the sustained surface winds for all three cases. Conversely, Chapter 4 and Sections 5.1 and 5.2 conclude that the primary error is not in the ensemble wind forecasts. Instead, comparisons of the MERRA data and local AWOS and ASOS observations during the events show that the MERRA data are unable to resolve the peak sustained winds for all three events, likely as a result of MERRA's coarse grid spacing and time averaging.

There are several other reanalyses that could be used as an alternative to MERRA data for surface pressure (i.e., SLP) and 2 m wind fields for verification purposes in future studies. The second Modern-Era Retrospective analysis for Research and Applications (MERRA-2; Bosilovich et al. 2016) replaces the original MERRA analysis (Rienecker et al. 2011). MERRA-2 has the same horizontal and vertical grid spacing and variable fields as MERRA, but uses an upgraded data assimilation system (Goddard Earth Observing System Model, Version 5; GEOS-5), which is capable of ingesting several new important data types (Bosilovich et al 2016). As an alternative to MERRA-2, Phase 2 of the North American Land Data Assimilation System (NLDAS-2; Xia et al.

2012b) can provide hourly precipitation, surface pressure (i.e., SLP) and 2 m wind fields. The NLDAS-2 non-precipitation land surface forcing fields are derived from the analysis fields of the NCEI North American Regional Reanalysis (NARR), and the NLDAS-2 precipitation fields are derived from gauge-only Climate Prediction Center analyses of daily precipitation. Available NDLAS-2 output fields range from hourly to monthly time scales from 1979 to present with 0.125˚ by 0.125˚ horizontal grid spacing (latitude by longitude).

5.3.2 Impacts of Track Bias on Precipitation Verification

Track bias has a strong effect on the precipitation forecast verification in this study, and the quality of the initial conditions is likely a contributing factor to track bias. The results of this study are consistent with those of Novak and Colle (2012), who found that the quality of the initialization is likely a key factor in affecting case-to-case predictability. All of the March 2011 case ensemble mean and member track forecasts were unable to capture the correct initialization of the cyclone in Domain 2 (Chapter 4), and the January 2009 36P ensemble member track forecasts tended to initialize either too far north or south (Chapter 5.1.2). Consequently, the March 2011 and 36P January 2009 precipitation forecasts performed worse and were less predictable than the other forecasts. Adjustments to the boundaries of Domain 2, such as increasing the area of the Domain 2 along its westward extent, could improve this bias.

All three cases have a similar ensemble track forecast performance for their 36P, 24P, and 12P forecasts, which influences the precipitation forecasts in the same manner. The 36P track forecasts tend to have the greatest spread (i.e., lowest confidence), as there are typically 36P ensemble member track forecasts located north and south of the MERRA

track (e.g., Fig. 35a). However the 24P and 12P track forecasts have increasingly higher confidence in the track location, respectively, but show a persistent track bias to the north and east of the MERRA track (e.g., Figs. 35c and 35e). Consequently, many of the event total precipitation forecasts for all three cases have similar biases in magnitude and location based on the quality of the forecast initializations (Figs. 16, 36, and 47). This is particularly evident in the January 2009 event. While confidence is relatively high for all three initialization times in the January 2009 event, the 36P event total precipitation forecasts (Figs. 36a and 37a) show the lowest confidence, and the 24P and 12P event total precipitation forecasts show increasing confidence but not necessarily increasing skill (Figs. 36b, 36c, 37b, and 37c).

The track biases have the most influence on the skill of the 0.25 inches threshold value of event total precipitation for all three cases. For all three cases, the ensemble precipitation forecasts tend to underpredict coverage for this threshold value and to incorrectly predict locations for the heaviest precipitation bands compared to the SNODAS data. For example, the confidence of the December 2013 event total precipitation forecasts is greater on average than the other two events and relatively the same as that of the track forecasts (Fig. 37). However, as previously mentioned, increasing confidence does not necessarily lead to increasing skill, and in this case, the precipitation forecast skill is more dependent on the track forecast skill.

5.3.3 Impacts of Multiple SLP Minima

This study concludes that while an automatic feature tracking process for SLP minima can be used to identify a cyclone track, the process still requires human oversight in order to accurately assess whether a physical or nonphysical track shift has occurred in

ensemble track forecasts or MERRA SLP data. As discussed in Chapter 4, the presence

of multiple SLP minima (e.g., double barrel low feature) complicates the automatic

feature tracking process of determining a cyclone track. Figures 14 and 47 show multiple

SLP minima with an abrupt shift in the cyclone tracks for the March 2011 and December

2013 cases. Section 4.4.4 details the abrupt shift for the March 2011 event and offers a

subjective analysis technique that indicates that some of the track forecasts would

perform better than the absolute error in distance suggests if they had intensified the

correct SLP minimum. Conversely, a situation can also occur in which the ensemble

track forecasts follow the correct intensifying SLP minimum and the MERRA cyclone

track does not, such as in the December 2013 case (Figs. 47a and 47b). (Note that the

location of the correct cyclone track is based on subjective analyses of local ASOS and

AWOS observations and the WPC surface analyses—not shown.)

Before assessing the quality of the forecast, it is recommended that several steps

be added to the automatic feature tracking process to avoid the some of the most common

drawbacks associated with the occurrence of multiple SLP minima. Hoskins and Hodges

(2002) discusses previous studies that have indicated that feature tracking fields are

usually restricted to MSLP combined with geopotential at 500 hPa, geostrophic vorticity

using MSLP, 500 hPa relative vorticity, or 850 hPa relative vorticity. Vorticity data may

work well in a climatological or post-event analysis, but for operational use, vorticity

data are not ideal. Additionally, there are times in operations during which a primary

cyclonic surface circulation is difficult to ascertain. Several other logical checks could

include setting a maximum displacement distance from the cyclone center at the previous

time step or only allowing the cyclone track to propagate in the direction of the mean

upper level flow (e.g., with an eastward component in the Northern Hemisphere). Unfortunately, these logical checks can fail in certain scenarios. Other methods could include a neighborhood approach in searching for the SLP minimum at the next time step (e.g., Deppe et al. 2013) or a smoothing approach in which the SLP field values would be smoothed until a single "true" local minimum remains. Overall, further work is needed in this area as a subjective analysis of the plan-view hourly SLP observations is still typically necessary to identify the most accurate primary cyclone center.

5.3.4 Overall Member Performance and Spread

In this study, one physics scheme may perform best for one forecast field of one case, but no physics parameterization scheme or model forcing performs best for all of the verification fields and cases. The lack of a best performing member is not unusual, as other studies (e.g., Evans et al. 2012 and Jankov et al. 2007) have obtained similar results.

Although one or two members may have performed better than the ensemble mean for one field and one forecast initialization, the ensemble mean tends to perform better than the individual members on average. Individual members tended to perform better than the mean when there was low accuracy and high confidence in the all of the member forecasts in one initialization (e.g., poor initialization). For example, all of the March 2011 24P member track forecasts had a bias north of the MERRA track. Figures 15c and 15d show high confidence in the March 2011 24P track forecasts but low accuracy, which results in the member track forecasts with Thompson microphysics performing better than the mean track forecast.
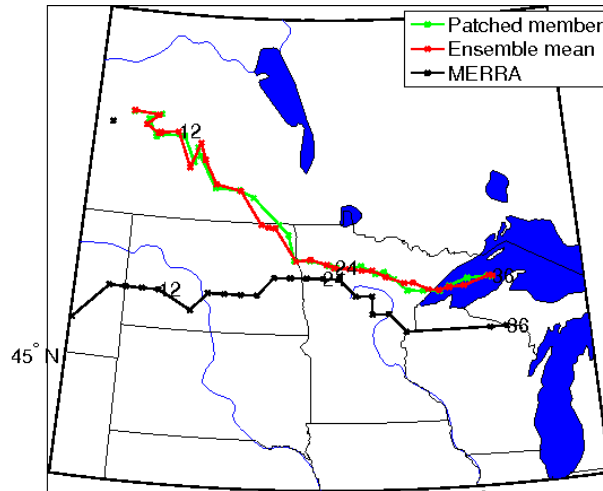
Figure 55. The MERRA track (black), "patched together" member track (green), and ensemble mean track (red) for the 12P forecast of the March 2011 case.

While the ensemble mean provides the statistically best performing forecast on average, Ancell (2013) indicates that the mean can provide poor forecast guidance for specific high-impact events. On the other hand, the best continuous forecast member differs in performance at each time interval and is likely to diverge from the mean at some point, due to its higher average error. Thus, Ancell (2013) takes another approach to finding the "best" member by creating a new "patched-together" member from the forecast member closest to the ensemble mean at each analysis time. This "patched-together" member is both realistic (i.e., physical) and contains the least error in comparison to the ensemble mean and "best" forecast members. The "patched together" member track and 12P ensemble mean track for the March 2011 event are shown in Figure 55.

5.3.5 Additional Ideas for Ensemble Improvement

Ensemble verification analysis in this study is dependent on threshold values in order to depict whether the ensemble predicts blizzard conditions for each event. Threshold values were carefully chosen based on a number of factors (see Chapters 3 and

106

4 for details), but additional threshold values could be added in order to provide more detail in the variability and uncertainty in the ensemble forecasts and to better address operational forecasting needs

For example, NOAA's Weather Prediction Center (WPC) creates probabilistic winter weather forecasts from SREF data. WPC uses one threshold value (0.25 inches of liquid equivalent snow/sleet) for their public experimental day 4-7 winter weather outlook (Weather Prediction Center 2015a) and multiple threshold values on their internal impact graphics website (Weather Prediction Center 2015b), based on intensity (e.g., visibility and precipitation rate), event duration, event timing, and other impacts (e.g., ensemble mean event total accumulation and blizzard criteria). Furthermore, the WPC SREF probability of exceeding blizzard criteria considers 10 m wind speeds exceeding 30 kts and surface visibility less than one-quarter mile in falling snow or ice pellets for at least three hours.

In future work, calculation of the probability of simultaneously exceeding an operational threshold value for more than one field (e.g., simultaneous peak wind and measurable snowfall) would be beneficial. Other future ensemble verification analyses could include comparing the ensemble forecasts to the NCEI's Short Range Ensemble Forecast (SREF) model forecasts, creating 48P ensemble forecasts for each case, or developing additional ensemble member forecasts with alternative surface layer or PBL schemes.

CHAPTER VI

BROADER IMPLICATIONS FOR OPERATIONS

6.1 Current Operational Use of Ensembles

Probabilistic information has become a crucial part of the forecast process (as discussed in Chapter 2). Operational forecasting is trending toward using ensemble output more and deterministic less in forecast processes and in products (Novak et al. 2008). In fact, recent developments in the weather and forecasting display and analysis package used by the NWS, the Advanced Weather Interactive Processing System II (AWIPS II; University Corporation for Atmospheric Research 2016), allow forecasters to view, manipulate, and create statistics for several ensemble datasets (e.g. operational Global Ensemble Forecasting System (GEFS) and Short Range Ensemble Forecasting (SREF) ensembles). For example, a NWS forecaster could use the AWIPS II Display 2-Dimensions (D2D) graphical user interface (GUI) to display the GEFS mean 700 hPa temperature field and the 700 hPa temperature fields for all of the ensemble members or to calculate statistical values from the 700 hPa temperature fields for all of the ensemble members (e.g. maximum, minimum, range).

NWS forecasters also have new probabilistic tools to assist them in making decisions as part of the severe weather warning decision-making process. For example, the NWS is operationally testing the NOAA/CIMSS ProbSevere product (Cintineo et al. 2014), which statistically forecasts the probability that a developing storm will produce

severe weather based on a combination of real-time satellite, radar, and RAP model data. NOAA's Weather Prediction Center (WPC) and Storm Prediction Center (SPC) also create operational products that give the likelihood of hazardous winter weather occurring in days one through seven of the operational forecast (WPC 2015a, 2015b). These products are available internally to NOAA NWS operational forecasters and externally to private sector forecasters, decision makers, and the public.

Recent social science research has found that users of operational forecasts prefer to have access to forecast uncertainty information. Morss et al. (2008) concludes that the U.S. public tends to prefer weather forecasts that express uncertainty to single-valued forecasts and that addressing forecast uncertainty increases users' trust in the forecasts. Additionally, decision-makers need detailed uncertainty information regarding timing, magnitude, and/or impacts of significant weather and require uncertainty information based on their threshold values in order to make the best decisions possible (Joslyn et al. 2012, 2013). Because of the aforementioned findings, the NWS is focused on improving its communication of forecast uncertainty and risk in the future (NOAA 2016). Products such as those mentioned above and those created in this study (similar to WPC 2015a, 2015b) will aid forecasters in achieving this goal.

6.2 Operational Use of UND Alberta Clipper Ensemble

As discussed in Section 5.3, a local ensemble would be further calibrated to focus on operational performance in the Northern Plains region and could provide additional information regarding forecast confidence (or lack thereof) for blizzard conditions. Ideally, the local ensemble would aid decision-making and increase forecaster confidence to enable timely issuance of blizzard watches and warnings (i.e., 24 to 72 hours preceding

the onset of blizzard conditions for a watch and 12 to 18 hours preceding the onset of blizzard conditions for a warning).

Figures 56 and 57 provide decision flow diagrams detailing the decision-making processes that operational forecasters would use in order to issue blizzard watches using currently available deterministic model guidance and the additional blizzard ensemble guidance, respectively. In both situations, forecasters must first use pattern recognition to ascertain that the global operational ensemble and deterministic models are indicating a signal of blizzard conditions. However, the available fields from current operational model guidance limit operational forecasters' skill in determining whether blizzard criteria are likely to occur in the next 24 to 72 hours (Fig. 56).

If the local blizzard ensemble described in this study were available for operational use, forecasters would first need to subjectively determine whether the previously mentioned signal in the operational model guidance was persistent enough to warrant spinning up the local blizzard ensemble (Fig. 57). Additionally, a decision regarding precisely when to spin up the ensemble could be impacted by available computational power. While computational cost assessment of the ensemble is beyond the scope of this study, a careful assessment must be performed before switching ensemble use from research to operations. Operational use will require access to computational power that would allow for ingest of initial and boundary conditions, completion of the ensemble members and ensemble products, and analysis of ensemble products by forecasters in a timely manner.

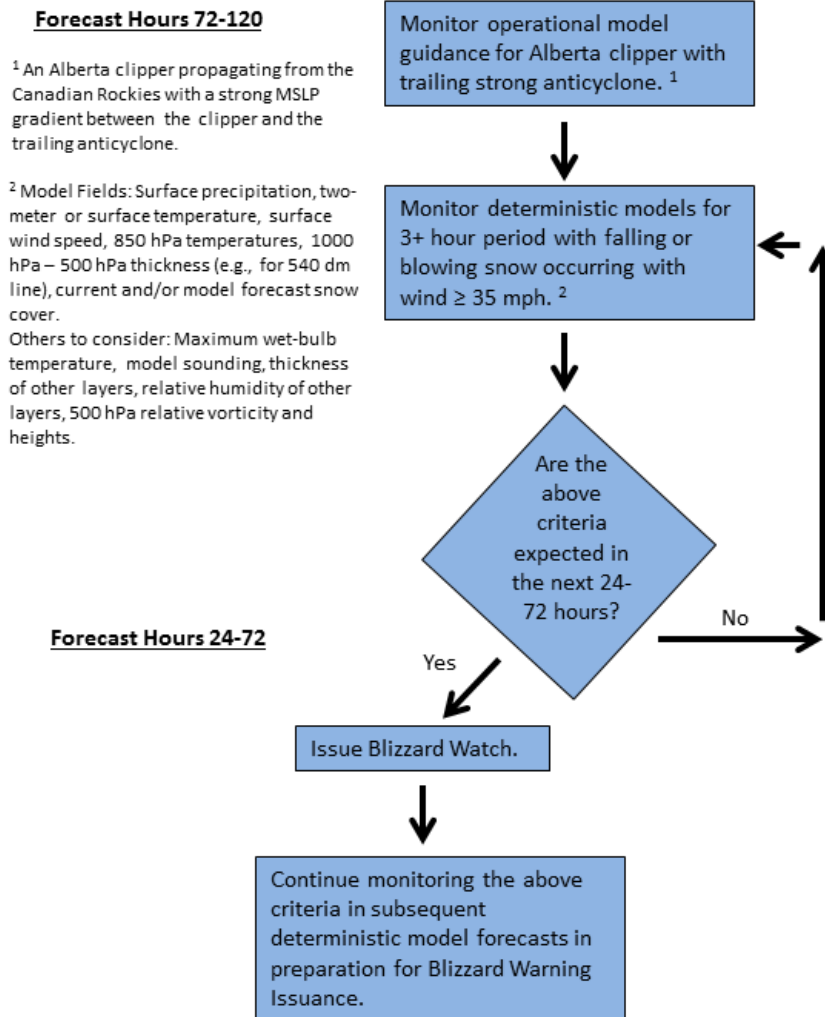## Deterministic Model Decision Diagram

Forecast Hours 72-120

[1] An Alberta clipper propagating from the Canadian Rockies with a strong MSLP gradient between the clipper and the trailing anticyclone.

[2] Model Fields: Surface precipitation, two-meter or surface temperature, surface wind speed, 850 hPa temperatures, 1000 hPa – 500 hPa thickness (e.g., for 540 dm line), current and/or model forecast snow cover.
Others to consider: Maximum wet-bulb temperature, model sounding, thickness of other layers, relative humidity of other layers, 500 hPa relative vorticity and heights.

Forecast Hours 24-72

Monitor operational model guidance for Alberta clipper with trailing strong anticyclone. [1]

Monitor deterministic models for 3+ hour period with falling or blowing snow occurring with wind ≥ 35 mph. [2]

Are the above criteria expected in the next 24-72 hours?

No

Yes

Issue Blizzard Watch.

Continue monitoring the above criteria in subsequent deterministic model forecasts in preparation for Blizzard Warning Issuance.

Figure 56. Decision diagram flowchart for blizzard watch issuance 24 to 120 hours prior to blizzard onset using deterministic model(s). Rectangular boxes indicate processing steps, or actions, and diamonds indicate a decision.

Once the local blizzard ensemble was started, it is recommended that an ensemble forecast be run every 12 hours until the event begins (Fig. 57). Based on the size and orientation of ensemble Domains 1 and 2 relative to typical clipper propagation (Thomas and Martin 2007), the Domain 1 (36 km) ensemble forecasts could be run as early as 72 to 120 hours preceding the predicted onset of blizzard conditions. For each ensemble forecast, Domain 1 forecasts should be run until 36 forecast hours prior to the blizzard

onset time. For the period from 36 forecast hours prior to the blizzard onset through the event duration, Domain 2 (12 km) ensemble forecasts should be run in order to create ensemble products.
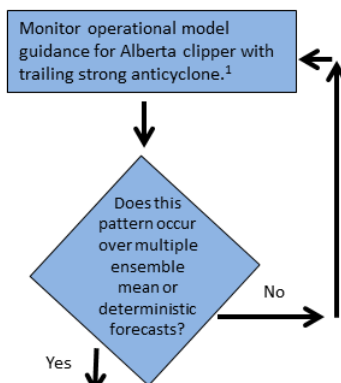
Operational model guidance and ensemble products should be monitored throughout the days leading up to the event. Ideally, the local ensemble would provide additional value for blizzard forecasting with hourly temporal resolution during the anticipated event. The ensemble products would ideally incorporate user-input threshold values with probability of exceedance. For example, ensemble probability of hourly sustained winds exceeding 25 mph (i.e., proxy for gusts exceeding 35 mph) and 35 mph could be overlaid with ensemble precipitation forecasts to provide an indication of the likelihood of blizzard conditions.

Starting one to three days before the onset of blizzard conditions, forecasters must assess whether blizzard criteria are expected. Ideally, the combination of local blizzard ensemble products and other available operational model guidance would provide increased confidence of the likelihood of blizzard conditions occurring. Thus, forecaster confidence would increase in determining whether issuance of a blizzard watch was necessary. Starting one to two days before the onset of blizzard conditions, ensemble forecasts could be run for both domains throughout the anticipated event. Forecasters should then continue monitoring local blizzard ensemble products and operational model guidance in preparation for blizzard warning issuance.
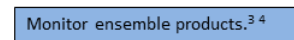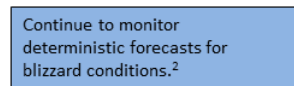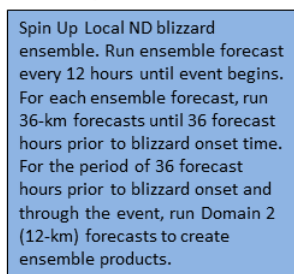
Ensemble Decision Diagram

Forecast Hours 72-120

[1] An Alberta clipper propagating from the Canadian Rockies with a strong MSLP gradient between the clipper and the trailing anticyclone.

Monitor operational model guidance for Alberta clipper with trailing strong anticyclone.[1]

Does this pattern occur over multiple ensemble mean or deterministic forecasts?

No

Yes

Spin Up Local ND blizzard ensemble. Run ensemble forecast every 12 hours until event begins. For each ensemble forecast, run 36-km forecasts until 36 forecast hours prior to blizzard onset time. For the period of 36 forecast hours prior to blizzard onset and through the event, run Domain 2 (12-km) forecasts to create ensemble products.

[2] Model Fields: Surface precipitation, two-meter or surface temperature, surface wind speed, 850 hPa temperatures, 1000 hPa – 500 hPa thickness (e.g., for 540 dm line), current and/or model forecast snow cover.
Others to consider: Maximum wet-bulb temperature, model sounding, thickness of other layers, relative humidity of other layers, 500 hPa relative vorticity and heights.

[3] Use ensemble mean and member track forecasts to determine likelihood of the clipper impacting the forecast area.

[4] Use probability of exceedance at user-defined precipitation and wind speed threshold values to determine the likelihood of blizzard conditions occurring in the forecast area.

Continue to monitor deterministic forecasts for blizzard conditions.[2]

Monitor ensemble products.[3] [4]

Forecast hours 24-72

Are NDP blizzard criteria expected in the next 24-72 hours?

No

Yes

Issue Blizzard Watch.

Continue monitoring the above criteria in subsequent ensemble products and deterministic model forecasts in preparation for blizzard warning issuance.
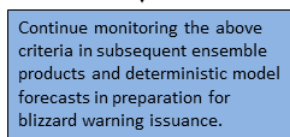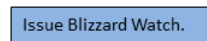
Figure 57. Decision diagram flowchart for blizzard watch issuance 24 to 120 hours prior to blizzard onset using ensemble and deterministic models. Rectangular boxes indicate processing steps, or actions, and diamonds indicate a decision.

CHAPTER VII

CONCLUSIONS

Verification analyses are performed for ensemble cyclone track, wind, and precipitation forecasts from three initialization times using MERRA SLP and wind data and SNODAS precipitation data for blizzard events from March 2011, January 2009, and December 2013.

1. The ensemble significantly underpredicts coverage of blizzard conditions for the March 2011 and December 2013 cases but not for the January 2009 case, primarily because of the wind forecasts. However, NWS criteria for blizzard warnings is unique in that it considers sustained winds and wind gusts. Thus, a threshold value of 25 mph can be used as a proxy to analyze wind forecasts for blizzard conditions.

2. No ensemble physics schemes perform best for all three cases, as expected from previous studies.

3. Accuracy of forecast initialization in Domain 2, with respect to the track, greatly influences the accuracy of the precipitation forecasts. Thus, precipitation forecasts show increasing confidence but not necessarily increasing skill as initialization approaches the event start time.

4. Improvements are needed in order to proceed with operational use of the ensemble:

a. The presence of multiple SLP minima presents a challenge for automatic tracking methods.

b. MERRA is unable to resolve the peak sustained winds for all three events, which has a significant impact on wind forecast verification.

c. There is still a great need for improving the reliability of automated solid precipitation measurements. The acute observational errors in the ASOS and SNODAS solid precipitation datasets used in this study make it difficult to accurately assess ensemble forecast performance.

5. Ensemble forecasts providing probability of exceedance for operational (and other user) threshold values would likely be beneficial to NWS (and other) users.

APPENDICES

Probability of Exceedance for Hourly Precipitation > 0.01 inches:
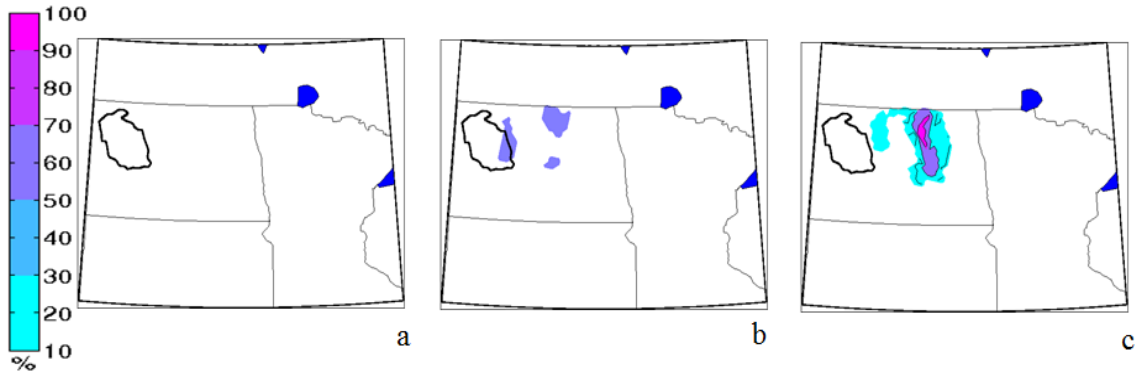March 2011 Case Variability Amongst Forecasts
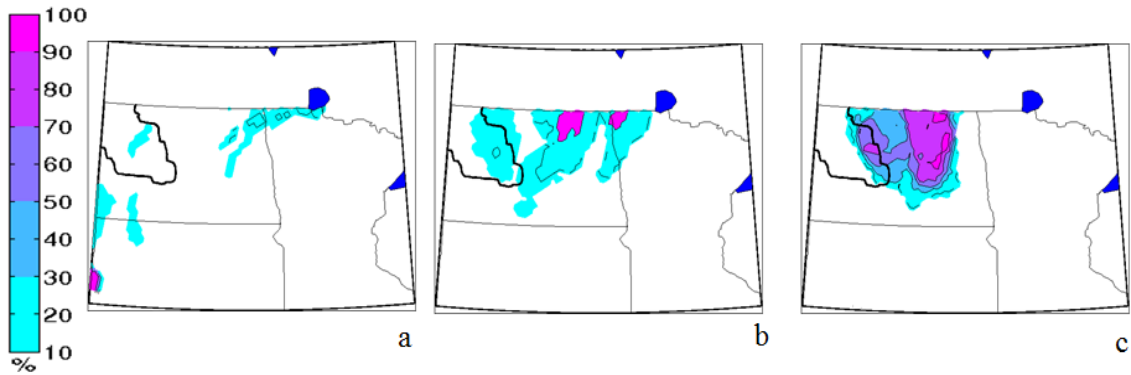
Figure A1. Percentage of members (color fill) forecasting hourly precipitation exceeding 0.01 inches from 17 UTC to 18 UTC 11 March 2011. Forecast wind speed values are given for Domain 2* from the 36P (a), 24P (b), and 12P (c) ensemble forecasts. Coverage area of MERRA wind speeds is outlined in black.
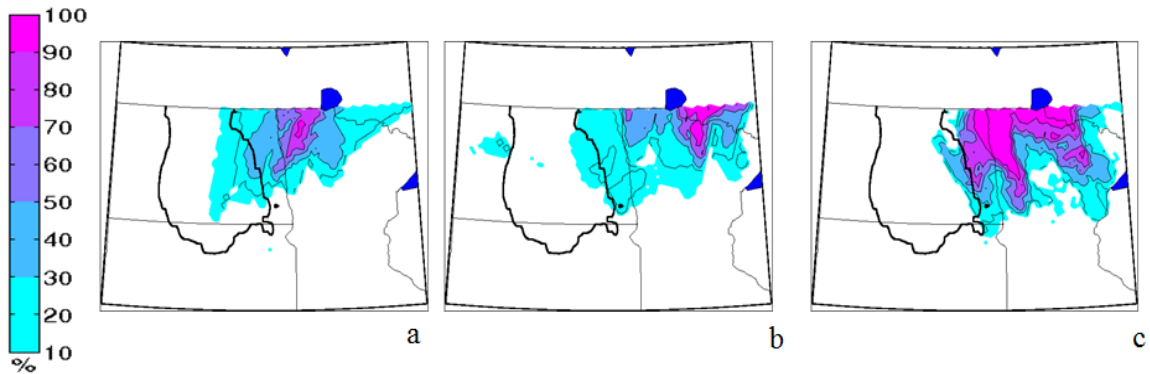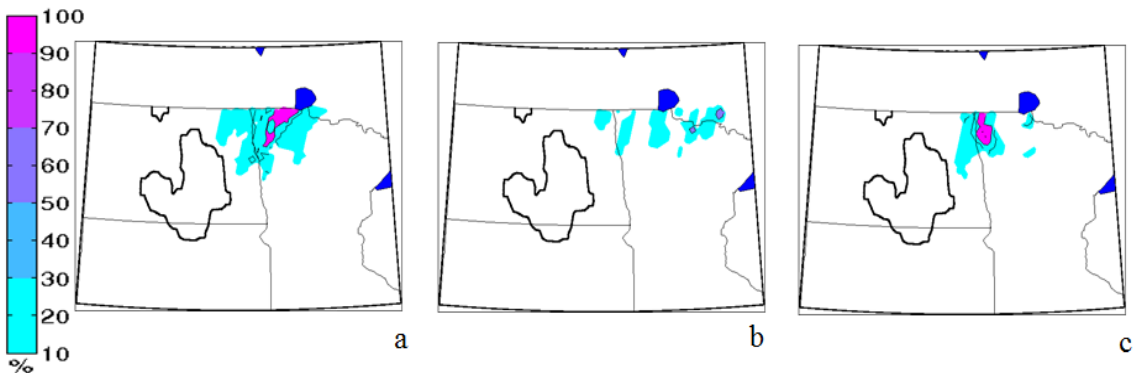

Figure A2. Same as Figure A1 but from 23 UTC 11 March to 00 UTC 12 March 2011.
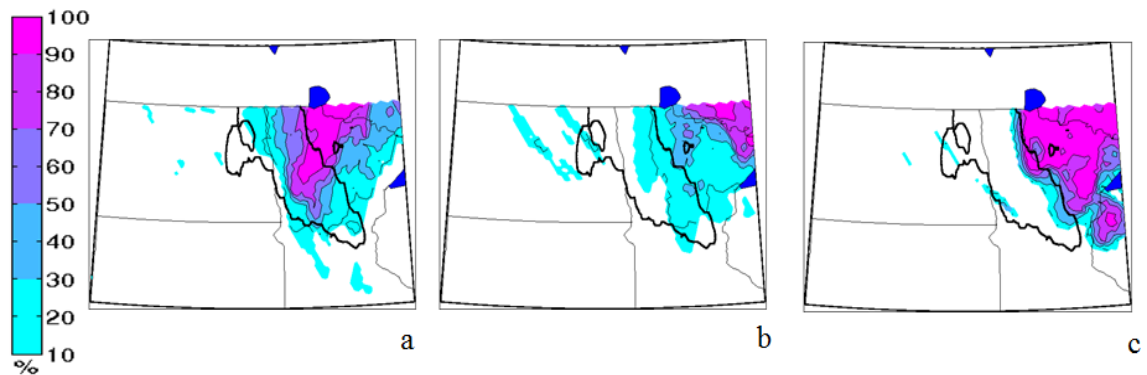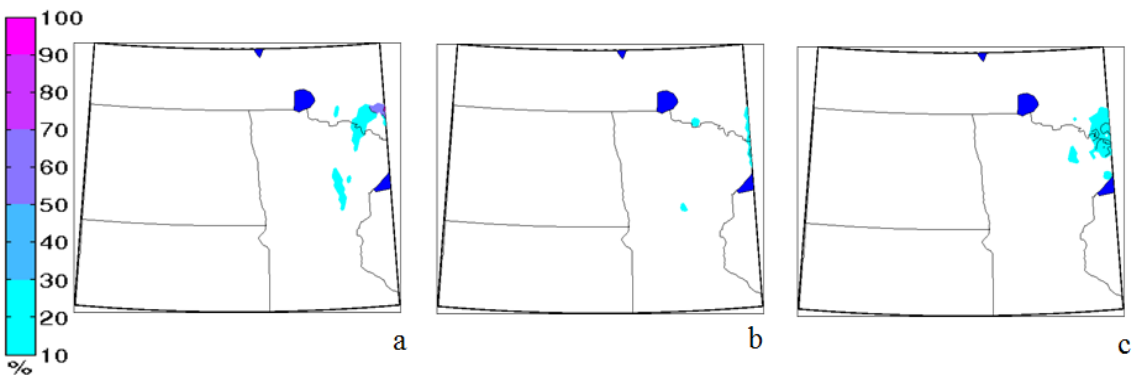

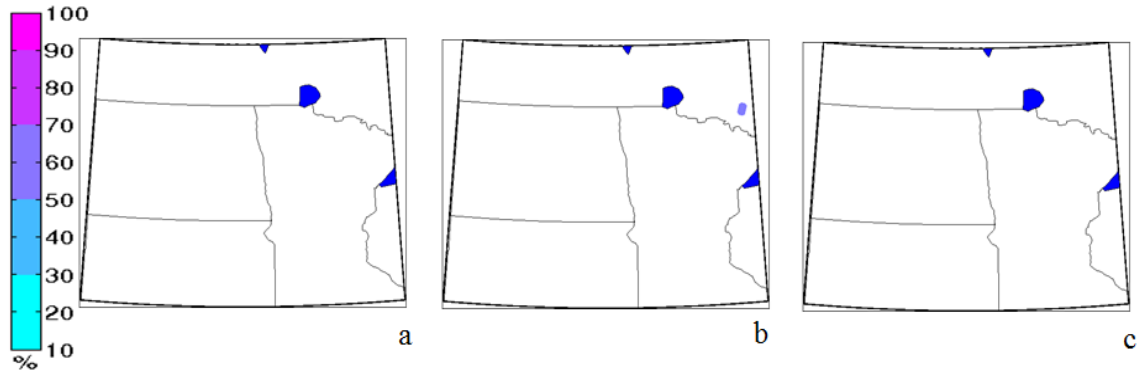Figure A3. Same as Figure A1 but from 05 UTC to 06 UTC 12 March 2011.
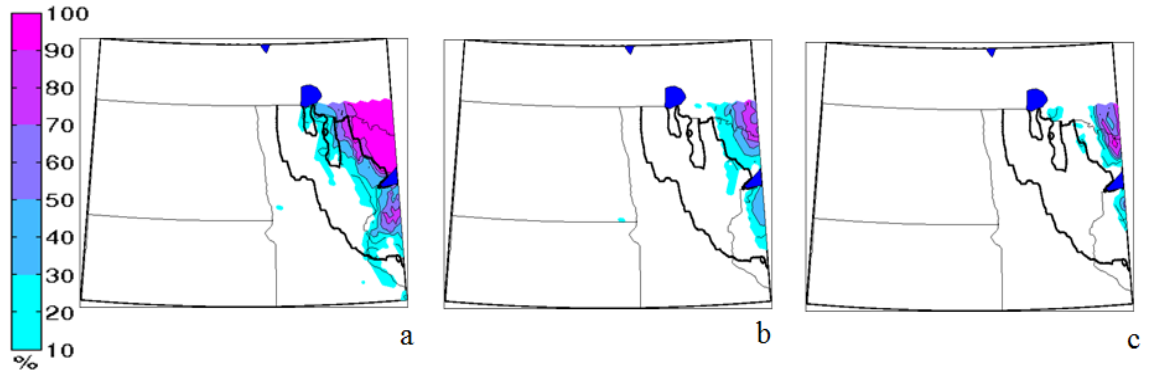
Figure A4. Same as Figure A1 but from 11 UTC to 12UTC 12 March 2011.

Probability of Exceedance for Hourly Precipitation > 0.05 inches:
March 2011 Case Variability Amongst Forecasts

Figure B1. Percentage of members (color fill) forecasting hourly precipitation exceeding 0.05 inches from 17 UTC to 18 UTC 11 March 2011. Forecast wind speed values are given for Domain 2* from the 36P (a), 24P (b), and 12P (c) ensemble forecasts. Coverage area of MERRA wind speeds is outlined in black.

Figure B2. Same as Figure B1 but from 23 UTC 11 March to 00 UTC 12 March 2011.

Figure B3. Same as Figure B1 but from 05 UTC to 06 UTC 12 March 2011.

Figure B4. Same as Figure B1 but from 11 UTC to 12 UTC 12 March 2011. No hourly precipitation greater than 0.05 inches was predicted by any of the ensemble forecasts or observed by the SNODAS dataset in the area of interest at this time.

Probability of Exceedance for Event Total Snowfall > 0.10 inches:
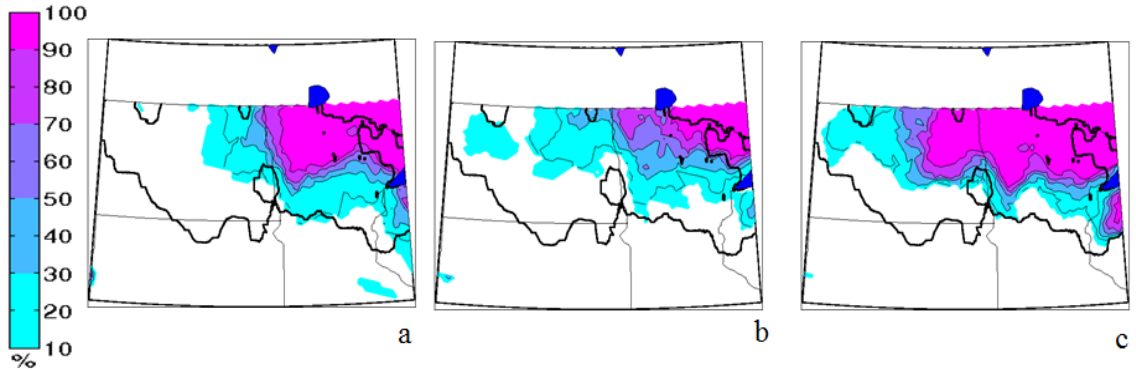March 2011 Case Variability Amongst Thresholds

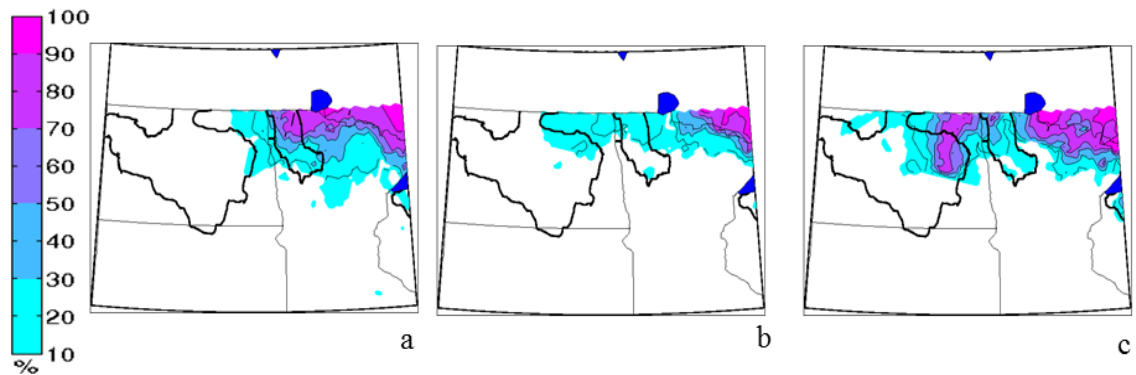Figure C1. Percentage of members (color fill) exceeding 0.10 inches of event total precipitation across Domain 2* for all three forecasts.

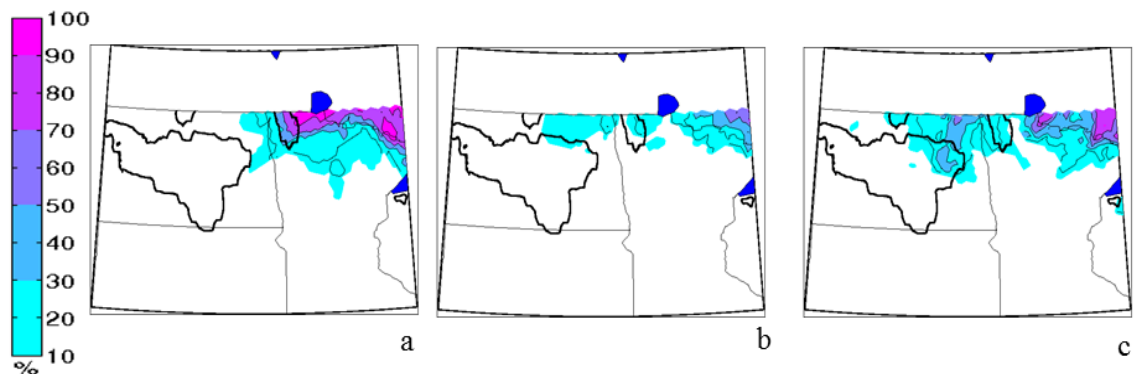Figure C2. Same as Figure C1 but exceeding 0.20 inches of event total precipitation.

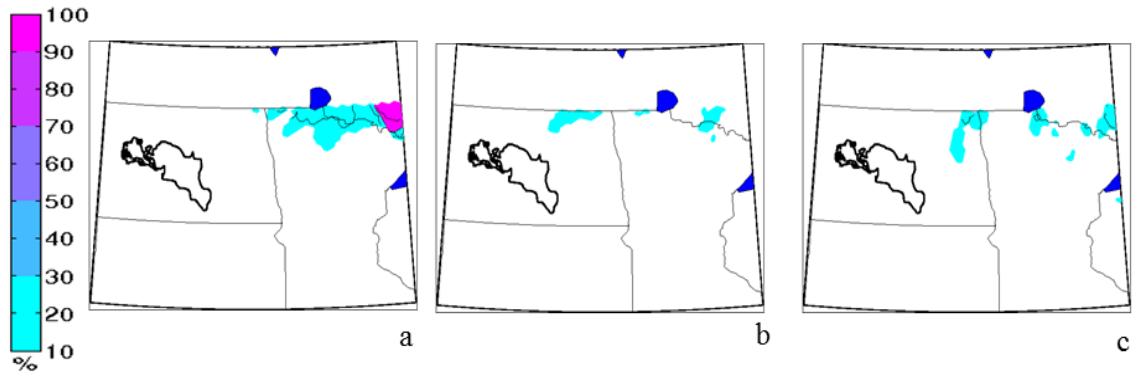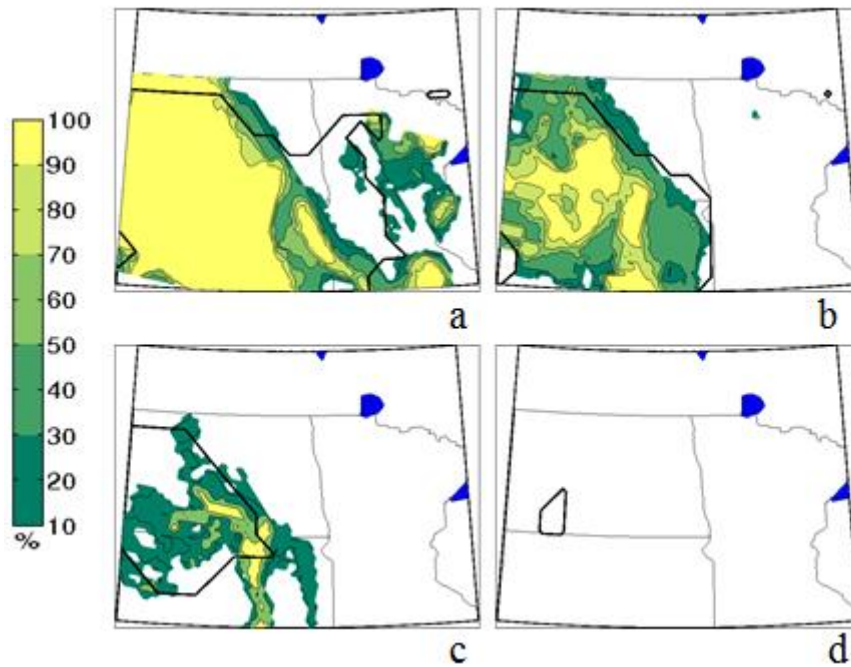Figure C3. Same as Figure C1 but exceeding 0.25 inches of event total precipitation.

Figure C4. Same as Figure C1 but exceeding 0.40 inches of event total precipitation.
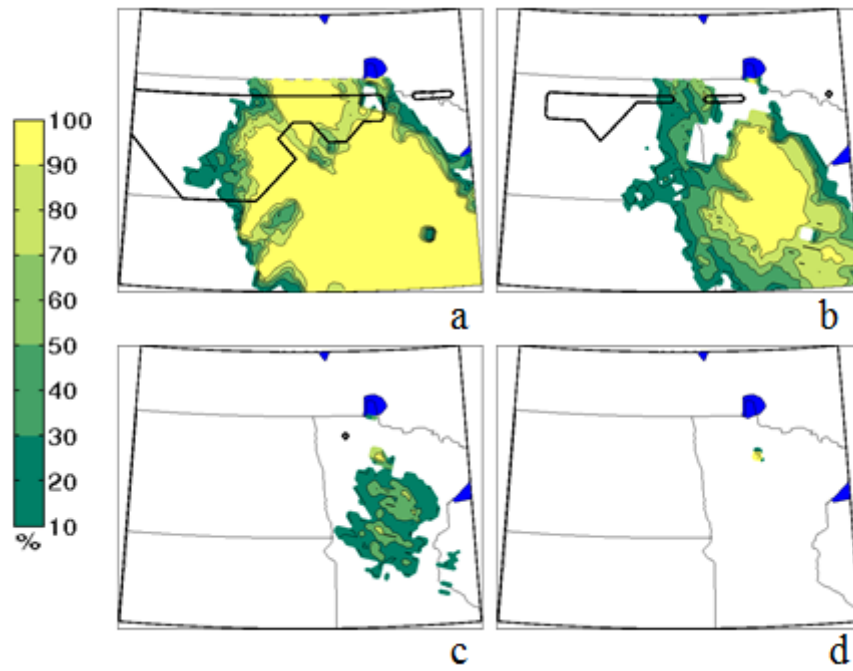
APPENDIX D
Probability of Exceedance for Wind Speed > 25 mph:
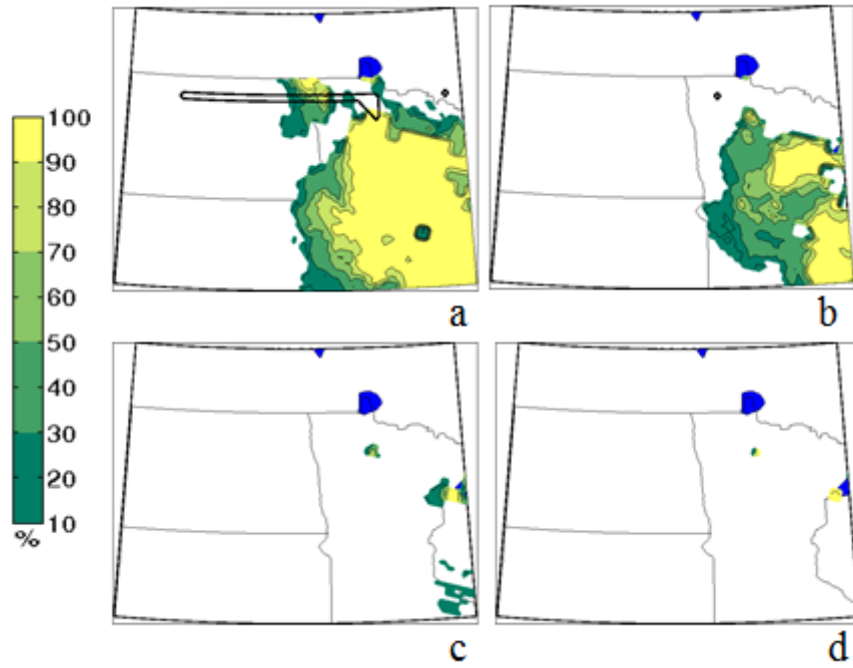March 2011 Case Variability Amongst Thresholds

Appendix D1: Percentage of members (color fill) exceeding 20 mph (a), 25 mph (b), 30 mph (c), and 0.25 (d) at 18 UTC 11 March 2011. 12P ensemble forecast wind speed values are given for Domain 2*. Coverage area of MERRA wind speeds is outlined in black.
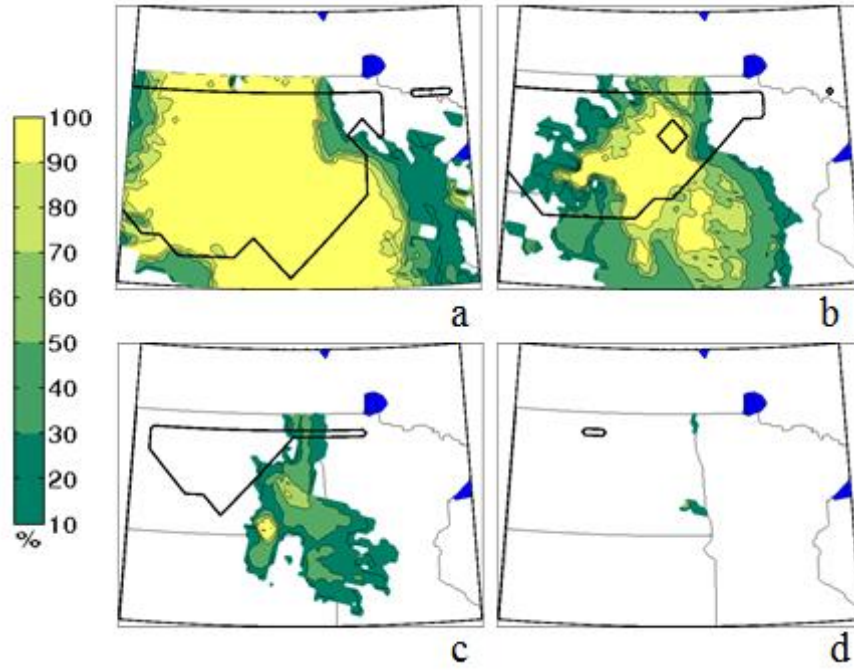
Appendix D2: Same as Figure D1 but at 03 UTC 12 March 2011.

Appendix D3: Same as Figure D1 but at 06 UTC 12 March 2011.

Appendix D4: Same as Figure D1 but at 12 UTC 12 March 2011.

Probability of Exceedance for Wind Speed > 25 mph:
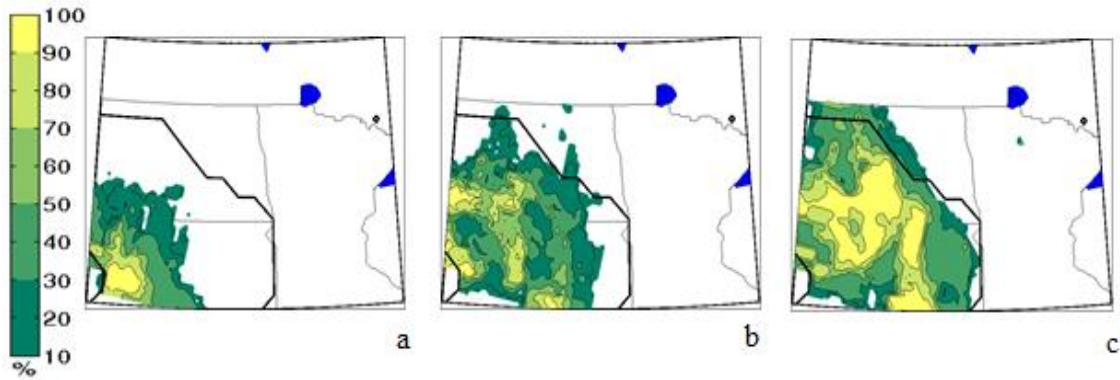March 2011 Case Variability Amongst Forecasts

Figure E1. Percentage of members (color fill) exceeding 25mph wind speeds at 18 UTC 11 March 2011. Forecast wind speed values are given for Domain 2* from the 36P (a), 24P (b), and 12P (c) ensemble forecasts. Coverage area of MERRA wind speeds is outlined in black.
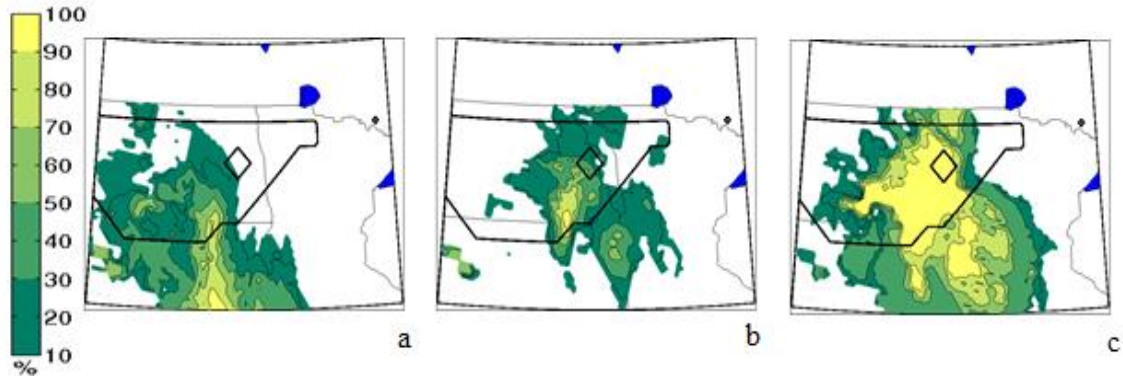

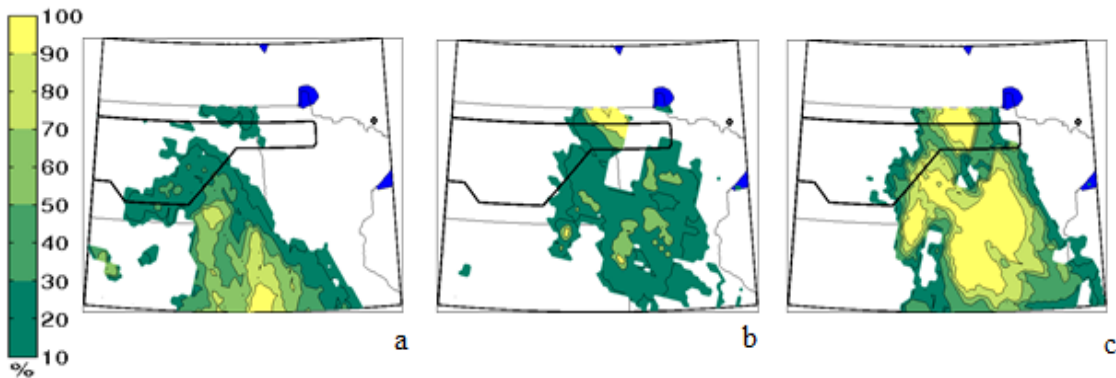Figure E2. Same as Figure E1 but at 00 UTC 12 March 2011.

Figure E3. Same as Figure E1 but at 03 UTC 12 March 2011.
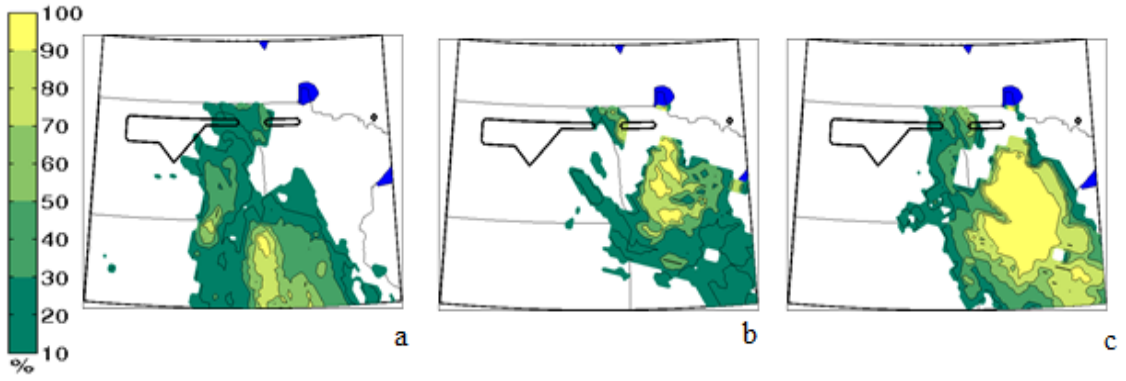
Figure E4. Same as Figure E1 but at 06 UTC 12 March 2011.
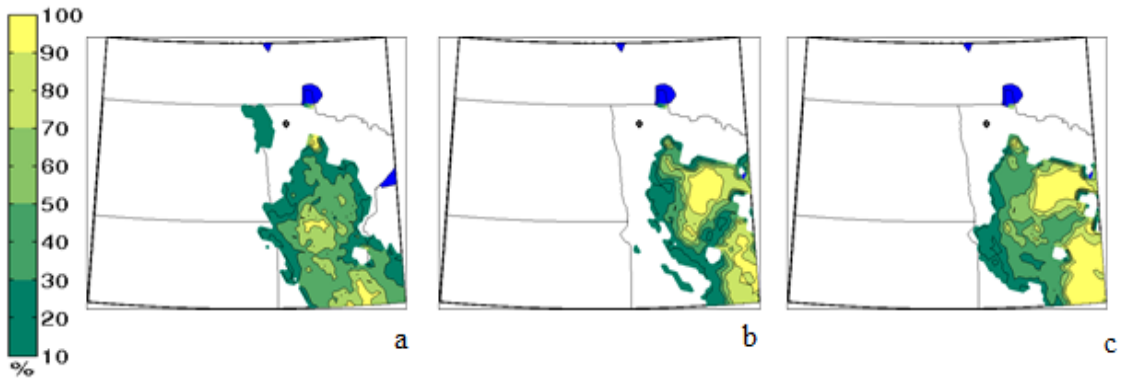
Figure E5. Same as Figure E1 but at 12 UTC 12 March 2011.
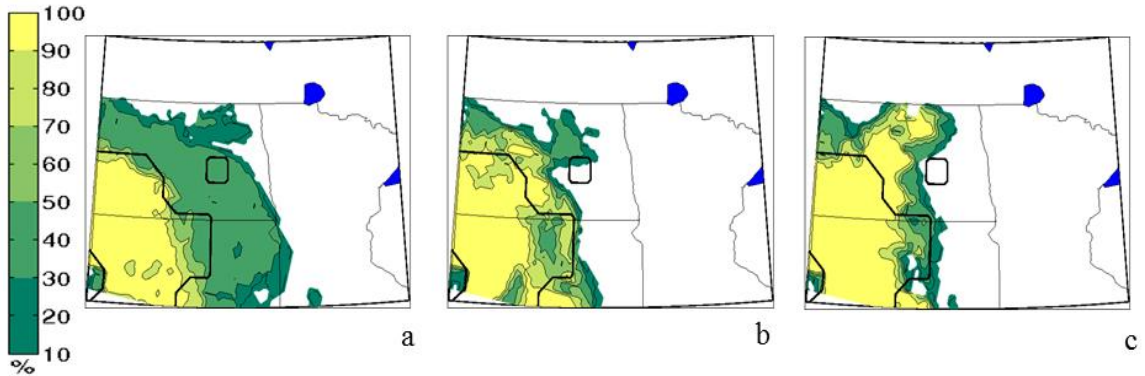
Figure F1. Percentage of members (color fill) exceeding 25mph wind speeds at 09 UTC
12 January 2009. Forecast wind speed values are given for Domain 2* from the 36P (a),
24P (b), and 12P (c) ensemble forecasts. Coverage area of MERRA wind speeds is
outlined in black.
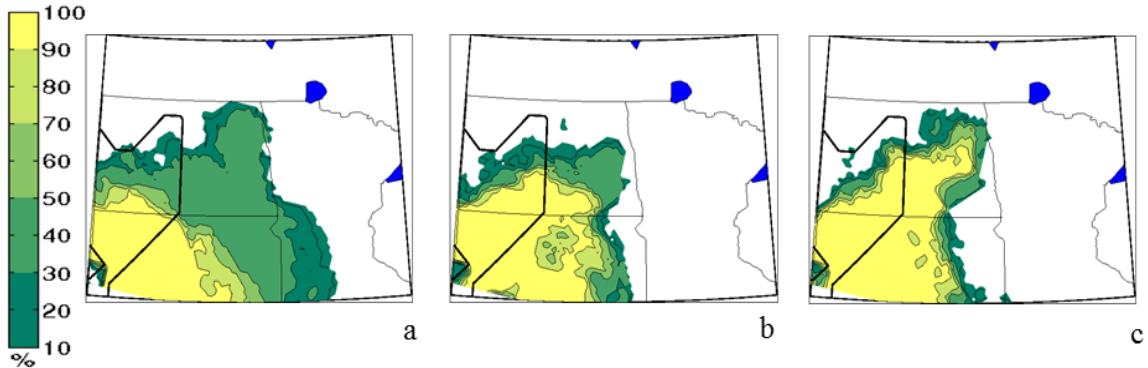
Figure F2. Same as Figure F1 but at 12 UTC 12 January 2009.
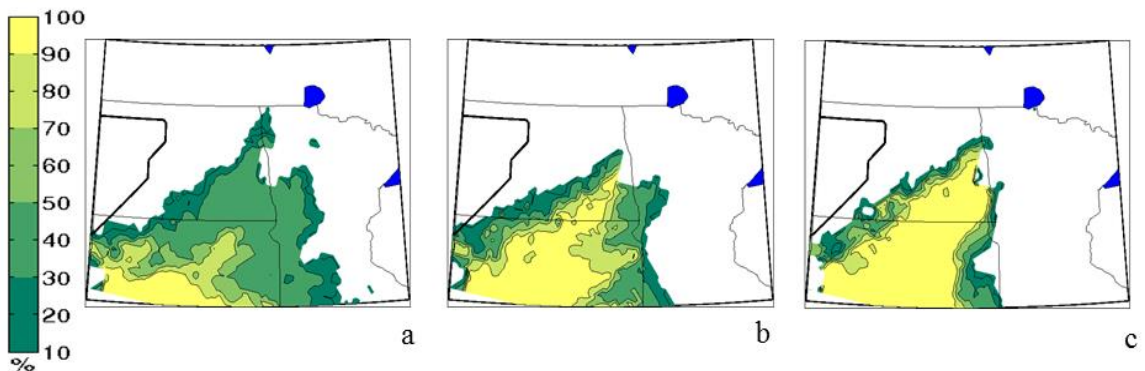
Figure F3. Same as Figure F1 but at 15 UTC 12 January 2009.
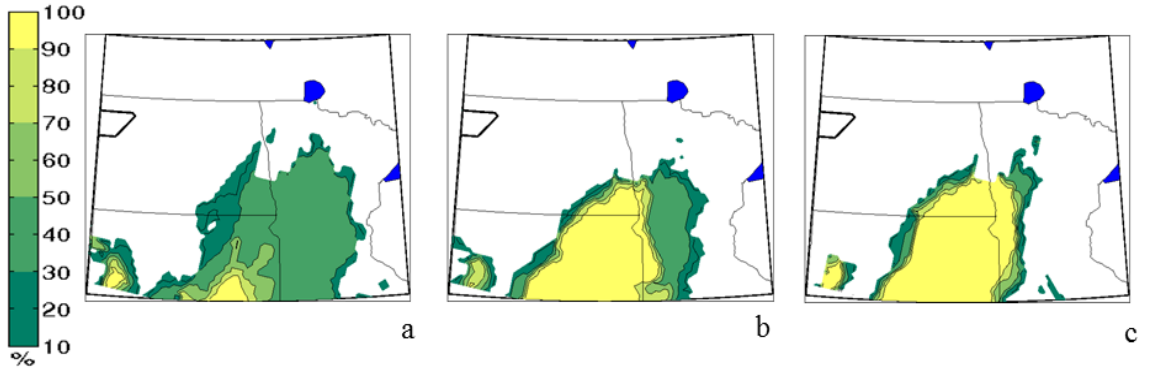
Figure F4. Same as Figure F1 but at 18 UTC 12 January 2009.
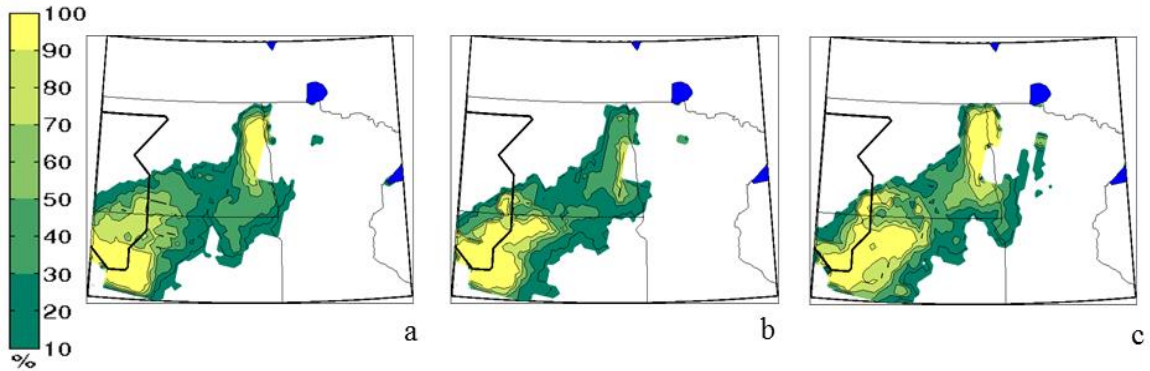
Figure G1. Percentage of members (color fill) exceeding 25mph wind speeds at 21 UTC 28 December 2013. Forecast wind speed values are given for Domain 2* from the 36P (a), 24P (b), and 12P (c) ensemble forecasts. Coverage area of MERRA wind speeds is outlined in black.

Figure G2. Same as Figure G1 but at 00 UTC 29 December 2013.
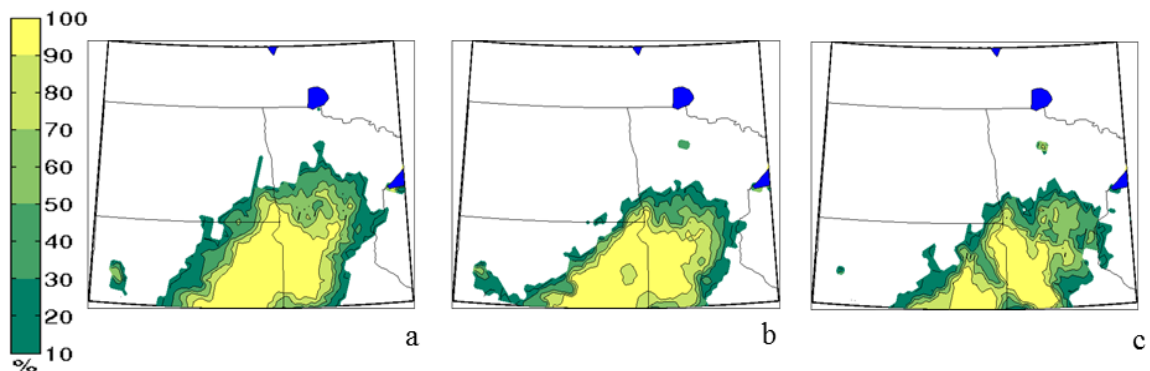

Figure G3. Same as Figure G1 but at 03 UTC 29 December 2013.
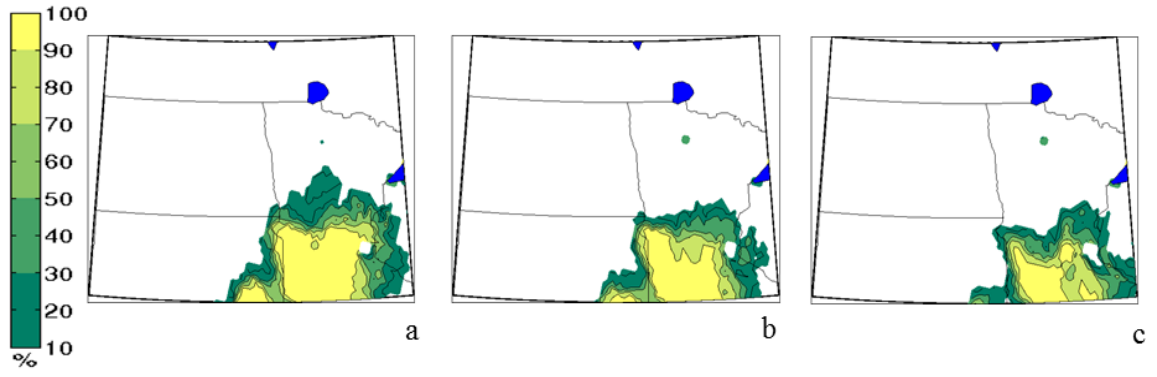
Figure G4. Same as Figure G1 but at 06 UTC 29 December 2013.

REFERENCES

American Meteorological Society, 1888a: Atmospheric pressure (expressed in inches and hundredths). *Mon. Wea. Rev.*, **16(3)**, 57-62.
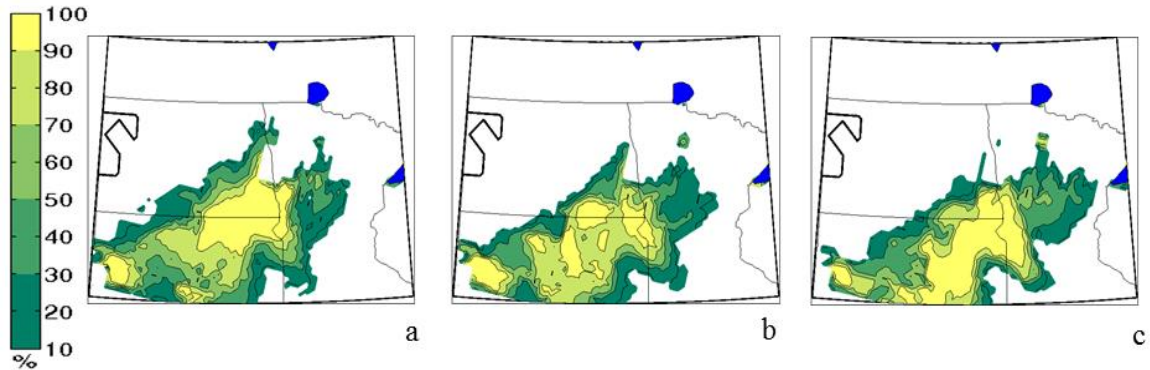
American Meteorological Society, 1888b: Chart V. *Mon. Wea. Rev.*, **16(3)**, c5-c5.

American Meteorological Society, 2002: AMS Statement: Enhancing Weather Information with Probabilistic Forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 450-452.

American Meteorological Society, 2013: "Alberta clipper." Glossary of Meteorology.

Ancell, B. C., 2013: Nonlinear Characteristics of Ensemble Perturbation Evolution and Their Application to Forecasting High-Impact Events. *Wea. Forecasting*, **28**, 1353–1365.

Angel, J. R. and S. A. Isard, 1997: An Observational Study of the Influence of the Great Lakes on the Speed and Intensity of Passing Cyclones. *Mon. Wea. Rev.*, **125**, 2228–2237.

Barrett, A., 2003: National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (SNODAS) Products at NSIDC. NSIDC Special Report 11. Boulder, CO USA: National Snow and Ice Data Center.

Baxter, M. A., C. E. Graves, and J. T. Moore, 2005: A Climatology of Snow-to-Liquid Ratio for the Contiguous United States. *Wea. Forecasting*, 20, 729-744.

Blackmon, M. L., 1976: A Climatological Spectral Study of the 500 mb Geopotential Height of the Northern Hemisphere. *J. Atmos. Sci.*, **33**, 1607–1623

Bjerknes, J., and H. Solberg, 1922: Life cycle of cyclones and the polar front theory of atmospheric circulation. *Geophys. Publ.,* **3**, No. 1, 1-18.

Bosilovich, M. G., R. Lucchesi, and M. Suarez, 2016: MERRA-2: File Specification. GMAO Office Note No. 9 (Version 1.1), 73. [Available from: http://gmao.gsfc.nasa.gov/pubs/office_notes.]

Carroll, T. D., D. Cline, G. Fall, A. Nilsson, L. Li, and A. Rost, 2001: NOHRSC Operations and the Simulation of Snow Cover Properties for the Conterminous U.S. *Proceedings of the 69th Annual Meeting of the Western Snow Conference*, 1-14.

Carlson, T. N., 1980: Airflow Through Midlatitude Cyclones and the Comma Cloud Pattern. *Mon. Wea. Rev.*, **108**, 1498–1509.

Cintineo, J., M. Pavolonis, J. Sieglaff, and D. Lindsey, 2014: An Empirical Model of Assessing the Severe Weather Potential of Developing Convection. *Wea. Forecasting*, **29**, 639-653.

Deppe, Adam J., William A. Gallus Jr., and Eugene S. Takle, 2013: A WRF Ensemble for Improved Wind Speed Forecasts at Turbine Height. *Wea. Forecasting*, **28**, 212–228.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-Range Ensemble Forecasting of Quantitative Precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459.

Eichler, T. P., N. Gaggini, and Z. Pan, 2013: Impacts of global warming on Northern Hemisphere winter storm tracks in the CMIP5 model suite, *J. Geophys. Res. Atmos.*, **118**, 3919–3932.

Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rodgers, G. Gayno, and V. Koren, 2003: Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108,** 8851.

Environmental Modeling Center (EMC), 2003: The GFS Atmospheric Model. *NCEP Office Note 442*.

Evans, J. P., M. Ekstrom, and F. Ji, 2012: Evaluating the performance of a WRF physics ensemble over South-East Australia. *J. Clim. Dyn.*, **39**, 1241-1258.

Flaounas E., S. Bastin, and S. Janicot, 2011: Regional climate modelling of the 2006 West African monsoon: Sensitivity to convection and planetary boundary layer parameterization using WRF. *Clim Dynam,* **36**, 1083–1105.

Gneiting, T. and A. E. Raftery, 2005: Weather Forecasting with Ensemble Methods. *Science*, 310, 248-249.

Goldenberg, S. B., S. G. Gopalakrishnan, V. Tallapragada, T. Quirino, F. Marks Jr., S. Trahan, X. Zhang, and R. Atlas, 2015: The 2012 Triply Nested, High-Resolution Operational Version of the Hurricane Weather Research and Forecasting Model (HWRF): Track and Intensity Forecast Verifications. *Wea. Forecasting*, **30**, 710–729.

Goodison, B. E., P. Y. T. Louie, and D. Yang, 1998: WMO solid precipitation measurement intercomparison. WMO Instruments and Observing Methods Rep. 67, WMO/TD-872, 212 pp.

Grell, G.A., 1993: Prognostic evaluation of assumptions used by cumulus parameterizations. *Mon. Wea. Rev.*, 121, 764-787.

Grell, G.A. and D. Dévényi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett*., 29, 1693-1697.

Hacker, Josh P., S.-Y. Ha, C. Snyder, J. Berner, F. A. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: scientific description and performance results. *Tellus*, **63A**, 625-641. Available at: http://www.tellusa.net/index.php/tellusa/article/view/15814.

Hong, S-Y. and J-O. J. Lim, 2006: The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). *J. Korean Met. Soc*, 42, 2, 129-151.

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

Hoskins, B. J. and K. I. Hodges, 2002: New Perspectives on the Northern Hemisphere Winter Storm Tracks. *J. Atmos. Sci.*, **59**, 1041–1061.

Hoskins, B. J. and P. J. Valdes, 1990: On the Existence of Storm-Tracks. *J. Atmos. Sci.*, **47**, 1854–1864.

Hurley, J. C., 1954: Statistics on the Movement and Deepening of Cyclones in the Middle West. *Mon. Wea. Rev.*, **82**, 116–122.

Hutchinson, T. A., 1995: An Analysis of NMC's Nested Grid Model Forecasts of Alberta Clippers. *Wea. Forecasting*, **10**, 632–641.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, J. Geophys. Res., 113, D13103, doi:10.1029/2008JD009944.

Jankov I., W. Gallus Jr., M. Segal, B. Shaw, and S. Koch, 2005: The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall. *Wea. Forecasting,* 20, 1048–1060.

Jankov I., P. Schultz, C. Anderson, and S. Koch, 2007: The impact of different physical parameterizations and their interactions on cold season QPF in the American River basin. *J Hydromet.*, 8, 1141–1151.

Joslyn, S. and J. LeClerc, 2012: Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error. *J. of Experimental Psychology*, **18**, 126-140.

Joslyn, S., L. Nemec, and S. Savelli, 2013: The benefits and challenges of predictive interval forecasts and verification graphics for end users. *Wea., Cli., and Soc.*, **5**, 133-147.

Kain, J. S., 2004: The Kain-Fritsch Convective parameterization: An update. *J. Appl. Meteor.*, 43, 170-181.

Kain, J. S. and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.*, 24, 165-170.

Kocin, P. J. and L. W. Uccellini (2004): Historical Overview. *Northeast Snowstorms: Volume II: The Cases, Meteorological Monographs Series,* **32**, 299-358.

Lucchesi, R. et al., 2012: File Specification for MERRA Products. GMAO Office Note No. 1 (Version 2.3). Available at: http://gmao.gsfc.nasa.gov/pubs/office_notes.

Mercer, A. E. and M. B. Richman, 2007: Statistical Differences of Quasigeostrophic Variables, Stability, and Moisture Profiles in North American Storm Tracks. *Mon. Wea. Rev.*, **135**, 2312–2338.

Mesinger, F. and R. E. Treadon, 1995: "Horizontal" Reduction of Pressure to Sea Level: Comparison against the NMC's Shuell Method. *Mon. Wea. Rev.*, **123**, 59–68.

Michaelis, A. C. and G. M. Lackmann, 2013: Numerical Modeling of a Historic Storm: Simulating the Blizzard of 1888. *Geo. Res. Letters*, **40**, 4092-4097.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* 102, 16663-16682, doi: 10.1029/97JD00237.

Morss, R. E., J. L. Demuth, and J. K. Lazo, 2008: Communicating Uncertainty in Weather Forecasts: A Survey of the U.S. Public. *Wea. Forecasting*, **23**, 974-991.

National Centers for Environmental Information (NCEI), 2005: *Quality Controlled Local Climatological Data (QCLCD)*. [Available at: http://www.ncdc.noaa.gov/qclcd/QCLCD/. Last accessed 10 May 2016.]

National Oceanic and Atmospheric Administration, 1998: *Automated Surface Observing System (ASOS) User's Guide*. [Available at: http://www.nws.noaa.gov/asos/pdfs/aum-toc.pdf/. Last accessed 6 September 2016.)

National Oceanic and Atmospheric Administration, 2013: *National Weather Service Weather-Ready Nation Roadmap Version 2*. [Available at: http://www.nws.noaa.gov/com/weatherreadynation/files/nws_wrn_roadmap_final_april17.pdf]

National Operational Hydrologic Remote Sensing Center (NOHRSC), 2004: *Snow Data Assimilation System (SNODAS) Data Products at NSIDC*, 11-12 March 2011. Boulder, Colorado USA: National Snow and Ice Data Center.

National Operational Hydrologic Remote Sensing Center (NOHRSC), 2004: *Snow Data Assimilation System (SNODAS) Data Products at NSIDC*, 11 January 2009. Boulder, Colorado USA: National Snow and Ice Data Center.

National Operational Hydrologic Remote Sensing Center (NOHRSC), 2004: *Snow Data Assimilation System (SNODAS) Data Products at NSIDC*, 28–29 December 2013. Boulder, Colorado USA: National Snow and Ice Data Center.

National Operational Hydrologic Remote Sensing Center (NOHRSC), 2015: *Interactive Snow Information*. [Available at: http://www.nohrsc.noaa.gov/interactive/. Last accessed 01 August 2016.]

National Weather Service Grand Forks, 2009a: Area Forecast Discussion. *NWS Text Products*. [Available at: https://mesonet.agron.iastate.edu/wx/afos/. Last accessed 05/10/2016.]

National Weather Service Grand Forks, 2009b: Public Information Statement.. *NWS Text Products*. [Available at: https://mesonet.agron.iastate.edu/wx/afos/. Last accessed 05/10/2016.]

National Weather Service Grand Forks, 2011a: Area Forecast Discussion. *NWS Text Products*. [Available at: https://mesonet.agron.iastate.edu/wx/afos/. Last accessed 05/10/2016.]

National Weather Service Grand Forks, 2011b: Public Information Statement. *NWS Text Products*. [Available at: https://mesonet.agron.iastate.edu/wx/afos/. Last accessed 05/10/2016.]

National Weather Service Grand Forks, 2013a: Area Forecast Discussion. *NWS Text Products*. [Available at: https://mesonet.agron.iastate.edu/wx/afos/. Last accessed 05/10/2016.]

National Weather Service Grand Forks, 2013b: Public Information Statement. *NWS Text Products*. [Available at: https://mesonet.agron.iastate.edu/wx/afos/. Last accessed 05/10/2016.]

Novak, D. R. and B. A. Colle, 2012: Diagnosing Snowband Predictability Using a Multimodel Ensemble System. *Wea. Forecasting*, **27**, 565-585.

Pauley, P. M., 1998: An Example of Uncertainty in Sea Level Pressure Reduction. *Wea. Forecasting*, **13**, 833–850.

Pinto, J. G., et al. 2007: Changes in storm track and cyclone activity in three SRES ensemble experiments with the ECHAM5/MPI-OM1 GCM, *Climate Dynamics*, **29**, 195-210.

Rasmussen, R. et al., 2012: How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bull. Amer. Meteor. Soc.*, 93, 811-829.

Rienecker, M.M. et al., 2011: MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. J. Clim. 24, 3624-3648, doi:10.1175/JCLI-D-11-00015.1.

Rogers, E., G. J. DiMego, T. L. Black, M. B. Ek, B. S. Ferrier, G. A. Gayno, Z. Janjic, Y. Lin, M. E. Pyle, V. C. Wong, W.-S. Wu, and J. Carley, 2009: The NCEP North American mesoscale modeling system: Recent changes and future plans. Preprints, *23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, Amer. Meteor. Soc., Omaha, NE. paper 2A.4.

Sanders, F. and J. R. Gyakum, 1980: Synoptic-Dynamic Climatology of the "Bomb. *Mon. Wea. Rev.*, **108**, 1589–1606.

Schultz, D. M. and C. A. Doswell, 2000: Analyzing and Forecasting Rocky Mountain Lee Cyclogenesis Often Associated with Strong Winds.*Wea. Forecasting*, **15**, 152–173.

Schwartz, R. M. and T. W. Schmidlin, 2002: Climatology of Blizzards in the Conterminous United States, 1959–2000. *J. Climate*, **15**, 1765–1772.

Schwarzkopf, M. D. and S. B. Fels, 1991: The simplified exchange method revisited – an accurate, rapid method for computation of infrared cooling rates and fluxes. J. Geophys. Res., 96, 9075-9096.

Shapiro, M. A. and D. Keyser, 1990: Fronts, jet streams and the tropopause. *Extratropical Cyclones, The Erik Palmen Memorial Volume.* C. W. Newton and E. Holopainen, Eds., Amer. Meteor. Soc., 167-191.

Sinclair, M. R., 1994: An Objective Cyclone Climatology for the Southern Hemisphere. *Mon. Wea. Rev.*, **122**, 2239–2256.

Sinclair, M. R., 1997: Objective Identification of Cyclones and Their Circulation Intensity, and Climatology. *Wea. Forecasting*, **12**, 595–612.

Skamarock, W.C., J. B. Klemp, J. Dudhia, D. Gill, D. Barker, M. Duda, X. Huang, W. Wang, and J. Powers, 2008: A description of the advanced research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 125 pp.

Smirnova, T. G., J. M. Brown, and S. G. Benjamin, 1997: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes, *Monthly Wea. Rev.*, 125, 8, 1870-1884.

Smirnova, T. G., J. M. Brown, S. G. Benjamin, and D. Kim, 2000: Parameterization of cold-season processes in the MAPS land-surface scheme, *J. Geophys. Res.*, 105, 3, 4077-4086.

Stewart, R.E., D. Bachand, R. R. Dunkley, A. C. Giles, B. Lawson, L. Legal, S. T. Miller, B. P. Murphy, M. N. Parker, B. J. Paruk, and M. K Yau, 1995: Winter storms over Canada, *Atmos.-Ocean*, **33**, 223-247.

Stensrud, David J., Harold E. Brooks, Jun Du, M. Steven Tracton, and Eric Rogers, 1999: Using Ensembles for Short-Range Forecasting. *Mon. Wea. Rev.*, **127**, 433–446.

Stensrud, D. J., J-W. Bao, and T. T. Warner, 2000: Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems. *Mon. Wea. Rev.*, **128**, 2077–2107.

Storm Prediction Center, 2005: *Introduction to the SPC's Hourly Mesoscale Analysis Page*. Available at: http://www.spc.noaa.gov/exper/mesoanalysis/help/begin.html.

Thomas, B. C. and J. E. Martin, 2007: A Synoptic Climatology and Composite Analysis of the Alberta Clipper. *Wea. Forecasting*, **22**, 315–333.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Mon. Wea. Rev.*, 136, 5095–5115.

Toth, Z. et al., 2001: Meeting Summary: Ensemble Forecasting in WRF. *Bull. Amer. Meteor. Soc.*, **82**, 695–697.

University Corporation for Atmospheric Research, 2016: *Unidata AWIPS II*. [Available at: http://www.unidata.uncar.edu/software/awips2/. Last accessed 09/06/2016.]

Wang, X. L., V. R. Swail, and F. W. Zwiers, 2006: Climatology and Changes of Extratropical Cyclone Activity: Comparison of ERA-40 with NCEP–NCAR Reanalysis for 1958–2001. *J. Climate*, **19**, 3145–3166.

Weather Prediction Center (WPC), 2015a: *Experimental Day 4-7 Winter Weather Outlook*. [Available at: http://www.wpc.ncep.noaa.gov/wwd/pwpf_d47/pwpf_medr.php. Last accessed 08/25/2016/.]

Weather Prediction Center (WPC), 2015b: *Short-Range Ensemble Forecast (SREF) Derived Winter Weather Impact Graphics*. [Available at: http://www.wpc.ncep.noaa.gov/wwd/impactgraphics/. Last accessed 08/25/2016.]

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko, 2012b: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, **117**, D03109.

Yang, D., B. E. Goodison, J. R. Metcalfe, V. S. Golubev, R. Bates, T. Pangburn, and C. L. Hanson, 1998: Accuracy of NWS 8" Standard Nonrecording Precipitation Gauge: Results and Application of WMO Intercomparison, *J. Atmos. Oceanic Tech.*, **15**, 54-68.

Yuan, X., X. Liang, and E. F. Wood, 2012: WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982-2008, *Clim. Dyn.*, **39**, 2041-2058.

Zhang, D. L. and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer – Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594-1609.