Examination of the diurnal cycle of rainfall and ensemble prediction strategies in WRF model simulations

Adam James Clark
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Recommended Citation
Clark, Adam James, "Examination of the diurnal cycle of rainfall and ensemble prediction strategies in WRF model simulations" (2006). Retrospective Theses and Dissertations. 19003.
https://lib.dr.iastate.edu/rtd/19003
Examination of the diurnal cycle of rainfall and ensemble prediction strategies in WRF model simulations

by

Adam Clark

A thesis submitted to the graduate faculty in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE

Major: Meteorology
Program of Study Committee:
Tsing-Chang Chen, Co-Major Professor
William A. Gallus Jr., Co-Major Professor
Igor Beresnev

Iowa State University
Ames, Iowa
2006
Copyright © Adam Clark, 2006. All rights reserved.
Graduate College
Iowa State University

This to certify that the master’s thesis of
Adam J. Clark
has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>CHAPTER 1. GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2. COMPARISON OF THE DIURNAL CYCLES IN A CONVECTION-RESOLVING AND NON-CONVECTION RESOLVING MODEL.</td>
<td>4</td>
</tr>
<tr>
<td>CHAPTER 3. CONTRIBUTION OF MIXED-PHYSICS AND PERTURBED INITIAL CONDITIONS TO THE SKILL AND SPREAD OF PRECIPITATION FORECASTS FROM A 16 MEMBER WRF ENSEMBLE</td>
<td>61</td>
</tr>
<tr>
<td>CHAPTER 4. GENERAL CONCLUSION</td>
<td>105</td>
</tr>
<tr>
<td>APPENDIX A: HISTORICAL PERSPECTIVE ON ENSEMBLE FORECASTING</td>
<td>108</td>
</tr>
<tr>
<td>APPENDIX B: WEBSITE PRESENTATION OF ENSEMBLE FORECASTS OF PRECIPITATION</td>
<td>110</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>112</td>
</tr>
</tbody>
</table>
ABSTRACT

Two studies were conducted addressing issues relevant to warm season precipitation forecasts in the United States. First, the representation of the diurnal cycle of rainfall in a 5 km grid-spacing mesoscale model explicitly representing convection was compared to a 22 km grid-spacing model implicitly representing convection. Improvements were expected in the 5 km grid-spacing model diurnal cycle of rainfall because of its lack of a cumulus parameterization scheme (CPS). Previous works show that problems with CPSs adversely affect the timing of precipitation and the ability of the model to represent the mesoscale dynamics that lead to propagation. Results revealed that the 5 km model did have a significantly better representation of the diurnal cycle of rainfall than the 22 km model. The timing, location, and representation of both propagating and non-propagating rainfall areas were superior in the 5 km model supporting the movement in the research community toward convection-resolving grid-spacing.

Second, a comparison of forecast skill and spread was made between a mixed-physics (MP) and perturbed initial conditions (PI) ensemble. Forecast skill was compared using deterministic forecasts derived from each ensemble using the probability matching technique and using probabilistic forecasts from each ensemble. Results revealed that the MP and PI ensembles had similar skill when the deterministic forecasts were evaluated using equitable threat scores (ETSs). However, when the area under the relative operating characteristic curve (ROC score) was used to evaluate the probabilistic forecasts, the MP ensemble had higher skill at the beginning of the forecast while the PI ensemble had higher skill at the end of the forecast. This behavior was directly related to the spread. In the MP ensemble, because the initial and lateral boundary conditions were the same for each ensemble member, the spread stopped increasing after about 24 hours, and shortly after this time the PI ensemble ROC scores tended to become larger than the MP ensemble ROC scores. However, during the first 24 hours of the forecast, greater spread in the MP ensemble forecasts was accompanied by higher ROC scores than in the PI ensemble. This demonstrates the importance of perturbed lateral boundary conditions in ensembles using limited area models if forecasts are desired beyond 24-36 hours.
GENERAL INTRODUCTION

Problem Statement

The skill for forecasting warm season rainfall in numerical weather prediction (NWP) models continues to be very low compared with forecasts of other meteorological variables (e.g., Fritsch and Carbone 2004; Olson et al. 1995). Fritsch and Carbone (2004) state that, in the foreseeable future, highly skillful deterministic forecasts of moist convection will most likely be limited to less than 3 hours. This presents an important and daunting challenge for researchers, especially in agricultural areas like the Midwestern United States where inaccurate rainfall forecasts can have large negative impacts on the livelihoods of many people. For example, irrigation management decisions based on rainfall forecasts that don’t verify could result in fields not being irrigated leading to the loss of crops.

These difficulties in warm season precipitation forecasts are caused by the nonlinear and chaotic nature of convective storms and the rapid growth of errors associated with them that originate from uncertainties in the initial conditions. Fritsch and Carbone (2004) outline the following areas for the United States Weather Research Program to address in efforts to improve warm season precipitation forecasts: 1) Representation of convection. 2) Methods for statistical post-processing of forecasts. 3) Data assimilation techniques. 4) Generation of probabilistic guidance. 5) Improvement of observing networks. 6) Increased understanding of microphysical processes.

The first study included in this thesis will address issue 1) by comparing and evaluating the diurnal cycle of rainfall in two models to see if a model with lower grid-spacing and the explicit representation of convection is superior to a coarser model implicitly representing convection. The second study will address issue 4) by comparing the skill and spread of rainfall forecasts from a mixed-physics and perturbed initial conditions ensemble.

Thesis Organization

This thesis contains two journal papers. The first paper, *Comparison of the Diurnal*
Cycles in a Convection-Resolving and Non-Convection-Resolving Mesoscale Model, examines diurnal averages from 5 km and 22 km grid-spacing versions of the Weather Research and Forecasting (WRF) model (Skamarock et al. 2001; Michalakes et al. 2001) that explicitly and implicitly represent convection, respectively. The second paper, Contribution of Mixed-physics and Perturbed Initial Conditions to the Spread and Skill of Precipitation Forecasts from a 16 Member WRF Ensemble, examines two 8 member WRF ensembles to identify characteristics in the skill and spread from each ensemble. The time it took for the lateral boundaries to degrade the mixed-physics ensemble forecasts was of particular interest. The thesis is organized into four parts: General Introduction, Comparison of the Diurnal Cycles in a Convection-Resolving and Non-Convection-Resolving Mesoscale Model, Contribution of Mixed-physics and Perturbed Initial Conditions to the Spread and Skill of Precipitation Forecasts from a 16 Member WRF Ensemble, General Conclusion. Both papers will be submitted to Weather and Forecasting. Also, appendices are included that provide background information on the historical evolution of ensemble forecasting (Appendix A: Historical Perspective on Ensemble Forecasting) and the way real-time forecasts were presented on a website (Appendix B: Website Presentation of Ensemble Forecasts of Precipitation).

References


Skamarock, W. C., J. B. Klemp, and J. Dudhia, 2001: Prototypes for the WRF (Weather
COMPARISON OF THE DIURNAL CYCLE IN A CONVECTION-
RESOLVING AND NON-CONVECTION-RESOLVING MESOSCALE
MODEL

A paper to be submitted to *Weather and Forecasting*

Adam J. Clark, William A. Gallus Jr., and Tsing-Chang Chen

Abstract

The diurnal cycle of rainfall in two models is examined by comparing 5 km grid-
spacing, 48 hour forecasts of the Weather Research and Forecasting (WRF) model using the
non-hydrostatic mesoscale model (NMM) dynamic core to 22 km grid-spacing forecasts
from the WRF model using the Advanced Research WRF (ARW) dynamic core, with
explicit and implicit representations of convection, respectively. Forecasts from 1 April to
25 July 2005 are examined for a domain encompassing the eastern two-thirds of the United
States. Traditional skill measures and time-longitude, or Hovmöller, diagrams of daily and
diurnally averaged forecast and observed rainfall are used to compare and evaluate the
simulations.

Forecast skill from both models is found to be significantly affected by the diurnal
cycle. Equitable threat scores (ETSs) were highest around 0000 LST and 0300 LST for both
models. However, the diurnal cycle of rainfall, shown using diurnally averaged Hovmöller
diagrams, was very different between the two models. Bias scores were also indicative of
this difference; the 22 km WRF-ARW had the highest bias scores around 1200 LST and the 5
km WRF-NMM around 1800 LST. Overall, the representation of the diurnal cycle in the 5
km WRF-NMM was superior to the 22 km WRF-ARW, especially during the second diurnal
cycle, which was probably because of the advantages afforded by the explicit representation
of convection in the 5 km WRF-NMM. However, overall, ETSs from the 22 km WRF-ARW
were higher, reinforcing past studies indicating that comparing forecasts from high and low
resolution models can be misleading because of large errors caused by the misplacement of small areas of intense rainfall forecast by high resolution models that are not resolved by coarser models.

1. Introduction

Difficulties in warm season precipitation forecasts are caused by the nonlinear and chaotic nature of convective storms and the rapid growth of errors associated with them that originate from uncertainties in the initial conditions and problems with model physics (e.g., Carbone et al. 2002). Approaches employed to address these issues include improving the model physics and dynamics representations (e.g., Mesinger et al. 1990, 1996; Zhang and Fritsch 1986), improving observing networks (e.g., Bergot et al. 2005; Zou et al. 1998), developing sophisticated data assimilation methods (e.g., Lee and Lee 2003; Xiao 2005; Davolio and Buzzi 2004), and running models at increasingly finer grid-spacing to improve the representation of topographic features and the physical processes related to them (Mass et al. 2002).

Since the beginning of numerical weather prediction (NWP) in the 1950s, a movement towards finer grids has contributed to slow but steady improvements in precipitation forecasts (Olsen 1995). However, when compared to the improvements in the forecasts of other variables, the improvements are small, especially during the warm season. This history of little progress is often blamed on cumulus parameterization schemes (CPSs; Fritsch and Carbone 2004). Because current computational resources in operational models do not allow for grid-spacing that is small enough to adequately resolve convection, these schemes, which account for the collective influence of small scale convective processes in large scale model variables, are a necessity. Some research has even shown that when convective parameterization schemes were active and produced the bulk of the precipitation in a mesoscale model, decreasing grid-spacing had little benefit when below around 50 km (Gallus 1999).

Problems with CPSs include; scale separation issues which arise when CPSs are applied at grid-spacing below 50 km, so that explicitly resolved clouds can form at a grid-point while similar clouds are being parameterized; and a lack of observations regarding the
physical processes that occur at the scales being parameterized (Molinari and Dudek 1992). More recent studies have shown that models using CPSs cannot simulate propagating convection (e.g., Davis et al. 2003; Moncrieff and Liu 2003; and Carbone et al. 2002). These findings imply that the parameterization of convection is not suited for precipitation forecasts in areas like the Midwest where the majority of warm season rainfall comes from propagating MCSs (Fritsch 1986). Because of these CPS deficiencies, many in the modeling community believe that a refinement in resolution to convection-resolving grid-spacing will improve precipitation forecasts in mesoscale models.

In many cases a refinement in resolution to convection-resolving grid-spacing has not led to the improvement of traditional, grid-point based skill measures (e.g., Colle et al. 2001). Thus, there may be a point of diminishing returns for high resolution forecasts when these measures are applied, even when subjective evaluation of the high resolution forecasts show a distinct advantage. This can result from large errors caused by small displacements in small areas of intense rainfall forecast by the high resolution models that are not resolved by the coarser models (Baldwin et al. 2001).

These issues have led researchers to pursue less traditional methods for verifying rainfall forecasts, including “events oriented” approaches that were advocated by Baldwin et al. (2001) and used by Done et al. (2004) and Weisman et al. (2004) to subjectively evaluate the timing, location, and mode of MCSs during the Bow Echo and MCV Experiment (BAMEX; Davis et al. 2004). Another, more systematic method was developed by Ebert and McBride (2000). Their method defined “contiguous rain areas” and calculated the total error in the forecasts of these areas contributed to by shape, amount, and horizontal displacement errors. Also, Skamarock (2004) used kinetic energy spectra and Knievel et al. (2004) used temporal modes of rainfall to evaluate rainfall forecasts. Some of these studies show that convection-resolving forecasts are superior to non-convection-resolving forecasts when the new verification methods are applied, while traditional, grid-point measures may show the opposite.

The goal of this paper will be to evaluate and compare the forecasts of rainfall from two versions of the Weather Research and Forecasting (WRF) model (Skamarock et al. 2001; Michalakes et al. 2001), one explicitly and one implicitly resolving convection, respectively. The results address issues posed by Davis et al. (2003), who state, “Further investigation of
the statistics of cloud-resolving models, especially in three dimensions, is required to address whether the improvement in propagation afforded by using cloud-resolving grids will substantially improve rainfall statistics”. The analyses will be done using diurnally averaged (i.e., averages are calculated for each forecast hour) skill measures and time-longitude, or Hovmöller, diagrams. The objectives are: 1) Identify diurnal oscillations in traditional skill measures and the associated features in the diurnally averaged rainfall forecasts. 2) Identify the model that performs best as shown by each verification measure. 3) Identify rainfall patterns/regimes that correspond to good and bad forecasts in each model as analyzed subjectively using Hovmöller diagrams of daily rainfall.

The paper is organized as follows. Section 2 will include detailed descriptions and specifications of the two models. Section 3 is the data and methodology section, including explanations of Hovmöller diagrams and forecast skill measures. Section 4 examines the results and section 5 offers conclusions and possibilities for future work.

2. Model description

Forecast rainfall data from 3 hour periods from simulations conducted 1 April to 25 July 2005 with 22 km and 5 km grid-spacing versions of the WRF model were examined. The 5 km grid-spacing version used the nonhydrostatic mesoscale model (NMM) dynamic core (Janjic 2003), hereafter, referred to as the 5 km WRF-NMM. The 5 km WRF-NMM simulations were conducted as part of Developmental Testbed Center (DTC) NMM Spring Forecast Experiment, an extension of the DTC Winter Forecast Experiment (DWFE; Bernardet et al. 2005). More information on these experiments can be found at http://www.dtccenter.org/projects/projects.php.

The simulations used 38 vertical levels with a domain encompassing most of the continental United States (Fig. 1). Integrations were performed daily at 00 UTC for 48 hours using initial and boundary condition data from 40 km grid-spacing Eta (now known as NAM; Mesinger 1998, Janjic 1994, Black 1994, and Rogers et al. 1998) model datasets generated from the 12 km grid-spacing operational runs performed at the National Center for Environmental Prediction (NCEP). The physics parameterizations used for the 5 km WRF-NMM included Ferrier microphysics (Ferrier et al. 2002), the Mellor-Yamada-Janjic
planetary boundary layer parameterization scheme (Janjic 1990, 1996b, 2002; Mellor and Yamada 2002), the NOAH land surface model (Ek et al. 2003) and no cumulus parameterization scheme. A smaller sub-region of the original domain was extracted to match the areas in which observed rainfall data were available (Fig. 1).

The 22 km grid-spacing version used the Advanced Research WRF (ARW) dynamic core (Klemp et al. 2000; Wicker et al. 2002; Skamarock et al. 2005), hereafter referred to as the 22 km WRF-ARW. The 22 km WRF-ARW simulations were run at the National Center of Atmospheric Research (NCAR). They were initialized daily at 00 UTC and integrated 72 hours (only the first 48 hours will be used to match the hours available from the 5 km WRF-NMM) using the same Eta model data as the 5 km WRF-NMM for initial and lateral boundary conditions. The physics parameterizations used include NCEP 3-class microphysics (Hong et al. 1998), the YSU boundary layer parameterization scheme (Hong and Pan 1996; Hong and Dudhia 2003), the NOAH land surface model (Ek et al. 2003), and the Kain-Fritsch cumulus parameterization scheme (Kain and Fritsch 1993). 28 vertical levels were used for a domain encompassing the continental United States. A section of this domain was extracted matching the extracted section of the 5 km WRF-NMM (Fig. 1). Additional information on this model can be found at http://wrf-model.org/plots/ncar22km.html.

During the time period used for this study there were dates when data from one or both models were not available. Only the dates when data from both models were available (Table 1) are used in the construction of averaged rainfall plots and calculations of skill measures.

3. Data and methodology

Three hour NCEP stage IV multi-sensor analyses (Baldwin and Mitchell 1997) were used to verify the model forecasts. The stage IV data was interpolated to both the 5 km WRF-NMM and 22 km WRF-ARW grids using a water budget method that conserves the total volume of liquid in the domain. This is the same routine used in verification procedures at NCEP. The stage IV analyses were interpolated to both model grids, rather than interpolating all data to one independent grid, so that the effects of the differing grids would
be retained.

Equitable threat score (ETS; Schaefer 1990) and bias are the traditional skill measures used to verify the forecasts. These are calculated using a contingency table in which the members are hits, false alarms, misses, and correct negatives denoted by $a$, $b$, $c$, and $d$, respectively. In terms of these contingency table elements, ETS and bias are expressed as:

$$ETS = \frac{da - bc}{(b + c) + (da - bc)}$$

$$Bias = \frac{a + b}{a + c}$$

ETS measures the fraction of events that were correctly predicted, accounting for hits that are associated with random chance. The range of values for ETS is $-1/3$ to 1, with scores below 0 having no skill and a score of 1 representing a perfect forecast. The bias score represents the ratio of the frequency of forecast to observed events. Bias indicates whether the model tends to over or under-forecast events. The range of bias values is 0 to infinity; 1 represents a perfect forecast and values above (below) 1 indicate over-forecasting (under-forecasting). When analyzing ETSs it is important to analyze bias, as well, because ETSs are sensitive to high bias scores (Hamill 1999).

Average ETSs and bias scores can be calculated by averaging scores from daily forecasts or summing the contingency table elements from all the forecasts and computing the scores from the summed elements. The first method gives equal weight to each forecast while the second gives more weight to larger precipitation events. Both methods were used, with the only notable differences occurring in the ETSs, which had higher scores using the summed contingency table elements. These higher scores imply that larger, more widespread precipitation events were associated with higher ETSs. Only the summed contingency table element scores are shown (Tables 1 and 2).

Hovmöller diagrams are constructed by computing meridional averages of forecast and observed 3-hr accumulated precipitation between $29^\circ$N and $48^\circ$N. This differs from similar previous works (e.g., Davis et al. 2003) that constructed Hovmöller diagrams using
the frequency of rainfall events above a specified threshold. The methodology in this study was chosen to retain information on the actual rainfall amounts and still be able to infer information on timing and location. The Hovmöller format is useful because it “collapses one spatial dimension of structure in favor of a time/direction of propagation depiction of phenomena” (Carbone et al. 1998). The distance coordinate used is degrees of longitude because rainfall systems in the Midwest typically propagate in the east-west direction. This analysis technique is routinely used in climate diagnostics (e.g., Levey and Jury 1996; Black et al. 1996; Murtugudde et al. 1996). Recently, these diagrams have been used to study the life cycle of precipitation systems using Doppler radar (Carbone et al. 1998; Wilson et al. 2001; Carbone et al. 2002) and to verify mesoscale models (Davis et al. 2003).

4. Results

a. Traditional skill measures

The bias scores from the 5 km WRF-NMM simulations were less than 1 initially and slowly increased to above 1 by the 6-9 hour period (Table 2), reflecting the model “spin up” of microphysical variables from zero. Skamarock (2004) also found, using 4 km and 10 km grid-spacing convection-resolving versions of the WRF model, that “the mesoscale portion of the kinetic energy spectrum, missing at the initial time, develops rapidly and reaches a fully developed state somewhere between 6 to 12 hours into the forecast”. For the rainfall thresholds 0.01 to 0.10, the bias scores approached or exceeded one at forecast hour 9, while for the rainfall thresholds 0.25 and above this occurred at forecast hour 6. Apparently, it takes the model slightly longer to generate areas of light rainfall that are comparable in scale to the observed areas of light rainfall than it does to create the smaller areas of heavy rainfall.

In the 22 km WRF-ARW simulations, at the rainfall thresholds 0.01 to 0.75 inches, the bias scores start relatively high and drop off to lower values at forecast hour 6 before they begin increasing again at forecast hour 9. This behavior is strong evidence of a spurious gap between the convective precipitation generated by the cumulus parameterization scheme and non-convective precipitation that is being resolved on the grid-scale; a phenomenon that has been observed in models using traditional CPSs (Molinari and Dudek 1992). This occurs
because CPS formulations allow the schemes to quickly activate and begin generating precipitation in moist, unstable environments favorable for convection. So, while the CPS can quickly generate areas of rainfall, the grid-resolved component still needs time to "spin up" microphysics variables. At the rainfall threshold of 1.00 inch (also for heavier amounts not shown in Table 2) there is less evidence of this spurious gap and the behavior of the bias scores are similar to the 5 km WRF-NMM, except the rate of increase is much lower in the 22 km WRF-ARW than in the 5 km WRF-NMM, implying that the CPS has trouble generating heavy precipitation.

A distinct diurnal oscillation exists in the bias scores from both models (Table 2). In the 5 km WRF-NMM these oscillations have higher amplitude at higher rainfall thresholds, while the amplitude of the oscillations in the 22 km WRF-ARW forecasts is fairly constant. The 5 km WRF-NMM bias scores generally peak around forecast hours 24 and 48 although the peaks occur in the range of forecast hours 21 to 27 and 45 to 48 (1500 to 2100 LST). Mean bias scores over the entire forecast period (Table 2) reveal that the model tends to overestimate precipitation more at higher rainfall thresholds. This is consistent with past research using the fully explicit approach (e.g., Molinari and Dudek 1992; Weisman et al. 2004).

Molinari and Dudek (1992) attributed the excessive rainfall in fully explicit simulations to the inability of the model to remove convective instability by sub-cloud scale eddies that could not be resolved the model grid. Thus, only grid scale vertical motion acted and advected high low-level equivalent potential temperature (theta-e) upward producing a tropospheric deep layer of convective instability (opposite to the theta-e minimum observed in nature close the mid-tropospheric layer in convection) that was overturned on the scale of the grid producing excess rainfall. Bryan et al. (2003) argues that even at 1 km grid-spacing the resolution is insufficient to resolve these sub-cloud scale eddies. To confirm if this is part of the problem in this study, thermodynamic profiles at grid points that experienced heavy rainfall would need to be examined, but in the present study only rainfall data were available, so this was not possible.

The 22 km WRF-ARW bias scores generally peaked around forecast hours 15 to 18 and 39 to 42 (0900 to 1200 LST), 6 hours earlier than the peaks from the 5 km WRF-NMM. The reasoning for these peaks will be discussed in the following section.
Similar to the bias scores, the ETSs exhibit oscillatory behavior and the effects of "spin up" can be observed during the first few forecast hours (Table 3). In the 5 km WRF-NMM forecasts the ETSs attain their highest values at forecast hours 3-12 with a trend for the heavier rainfall thresholds to have their highest scores earlier in the forecast. This is similar to the bias scores and, once again, implies that the model takes longer to generate areas of light rainfall that match observed areas of light rainfall than it does for heavier rainfall. This behavior is counter-intuitive, and an exploration of the cause of these trends is beyond the scope of this study. However, future work should see if this behavior is caused by differences in dynamical forcing. It makes intuitive sense that when there is weak forcing present the model may generate areas of light rainfall that take a relatively long time to "spin up", while under strong forcing the model quickly generates areas of heavy rainfall that are more likely to be associated with stronger forcing. It is also possible that the model fails to generate stratiform rain areas within organized MCSs, or that organized MCSs may be simulated well, but other areas of light rain are often completely missed.

In the 22 km WRF-ARW simulations the ETSs generally attain their highest values at forecast hours 6 to 12 (Table 3), similar to the 5 km WRF-NMM. This implies that the relatively heavy amounts of rainfall likely generated by the CPS during the first 3 hours of the forecast (indicated by the higher bias scores at these times), are not corresponding to observed areas. Apparently, the highest skill is probably not attained until the grid-resolved component has "spun up".

After initial peaks, the ETSs from both models tend to peak again around forecast hours 30 to 33 (0000 to 0300 LST). There is also a peak in the 22 km WRF-ARW confined to the rainfall threshold of 0.50 inches and above at forecast hour 18. This is the only time from either model that peaks in ETSs match peaks in bias scores. Past studies (e.g., Hamill 1999) have indicated that high ETSs are often associated with high bias scores. Because the highest ETSs in the present study do not occur at the times of the highest bias, it is likely the model truly does have more skill at forecasting rainfall at these times compared to other times. Also, at virtually all forecast hours and rainfall thresholds the ETSs are higher in the 22 km WRF-ARW. Past studies have shown that there may be a point of diminishing returns when applying traditional grid-point verification methods to high resolution forecasts (e.g., Mass et al. 2002; Gallus 2002; Fritsch and Carbone 2004) because coarser grid models,
unlike fine grid models, do not get penalized by having fine-scale details that may be displaced slightly, lowering the ETS.

**b. Relation between diurnally averaged plots of rainfall and skill measures**

Standard x-y plots (Figs. 3-10) and Hovmöller diagrams (Figs. 11-13; Fig. 11 is for the entire domain, while Fig. 12 and Fig. 13 are for the southern and northern halves of the domain, respectively. Figs. 12-13 will be discussed in a later section) of diurnally averaged forecast and observed rainfall were constructed. Using these, it is possible to infer information on what features in the models are causing the observed behaviors in the skill measures. In the 5 km WRF-NMM, bias scores peaked at forecast hours 24 and 48, corresponding to times when peak heating occurs and the non-propagating component of rainfall in the eastern part of the domain is at its maximum amplitude (Fig. 11). This non-propagating rainfall maximum is due mainly to rainfall in the southeast United States (Figs. 4-5). The 5 km WRF-NMM appears to accurately depict the timing of this rainfall maximum occurring at forecast hours 24 and 48 (Fig. 11), especially during the second diurnal cycle simulated by the model. Thus, it can be inferred that the high bias scores are simply a result of the 5 km WRF-NMM over-predicting rainfall at the maximum phases of the non-propagating component of the diurnal cycle. Note that the scale differs between the Hovmöller diagrams from the 5 km WRF-NMM and observations (Figs. 11-13) so that this over-prediction is even greater than what the shading on the plots implies.

The bias scores for the 22 km WRF-ARW were highest around forecast hours 18 and 42, corresponding to times in which the propagating signal in the west and non-propagating signal in the east were at minimum amplitudes (Fig. 11). It is inferred from the diurnally averaged Hovmöller diagrams of forecast rainfall from the 22 km WRF-ARW (Fig. 11) that the relatively high bias scores at these times are the result of the 22 km WRF-ARW simulating the late afternoon non-propagating rainfall maximum too early. Thus, the high bias scores were caused by this phase difference, as opposed to the over-prediction observed in the 5 km WRF-NMM.

As discussed earlier, a possible cause for the over-prediction in the 5 km WRF-NMM is believed to be related to the 5 km grid not being able to resolve sub-cloud scale processes.
that limit instability. Similar behavior to the phase difference in the 22 km WRF-ARW has been identified in past studies in models using traditional CPSs. For example, Baldwin et al. (2001) noted that the Betts-Miller-Janjic (BMJ; Betts 1986; Betts and Miller 1986; Janjic 1994) CPS within the NCEP Eta model had a tendency to remove capping inversions that are typical during the daytime in the Great Plains, because of its shallow mixing parameterization. This was discussed in Davis et al. (2003) who noted similar behavior from simulations of the WRF model that used the BMJ CPS. Also, Dai and Trenberth (2004) noticed that moist convection over land, simulated by version 2 of the Community Climate System Model (CCSM2), was initiated about 4 hours prematurely.

ETSs for both models were generally highest at or around forecast hours 6 to 9 and 30 to 33 (Table 2), corresponding to the times at which the propagating component in the western part of the domain is at its maximum amplitude (Fig. 11). Because the 5 km WRF-NMM accurately depicted the timing and longitude of both rainfall maxima (propagating in the west and non-propagating in the east), it is difficult to ascertain why the ETSs were highest around 0000 to 0300 CDT in this model, corresponding to the maximum amplitude of the propagating component. It is possible that in the southeast United States the non-propagating rainfall maximum is associated with convection that is unorganized, short-lived, and chaotic in nature. In the western high plains the propagating rainfall maximum is associated with long-lived and organized MCSs that are inherently more predictable (Carbone et al. 2002). Thus, small errors in the location of rainfall areas will penalize the ETSs more in the areas with more random and chaotic convection.

In the 22 km WRF-ARW, the peak in ETSs likely occurs at hours similar to the 5 km WRF-NMM because of the enhanced predictability of the rainfall systems that occur at these times. Also, the skill in the 22 km WRF-ARW forecasts are likely degraded more during the afternoon hours than in the 5 km WRF-NMM forecasts because of the timing errors discussed earlier.

Overall, the representation of the diurnal cycle by the 5 km WRF-NMM is superior to the 22 km WRF-ARW (Fig. 11). The 5 km WRF-NMM is able to accurately represent the timing and location of both the propagating rainfall maximum in the western portion of the domain, and the non-propagating rainfall maximum in the eastern portion, especially during forecast hours 24 to 48 (the forecasts during hours 00-24 are not as good because of “spin
up” issues; this will be discussed in the next section). The 22 km WRF-ARW has trouble simulating the coherent axis of propagating rainfall in the western part of the domain and also has trouble with the timing of the late afternoon rainfall maximum in the eastern part of the domain. However, overall, the ETSs from the 22 km WRF-ARW are higher than the 5 km WRF-NMM, likely an artifact of comparing high to low resolution forecasts (Baldwin et al. 2001). Thus, these results reiterate the importance of non-traditional verification techniques for verifying high resolution forecasts.

c. Propagating axes of rainfall emanating during model “spin up”

During the model “spin up” time, approximately 00 UTC to 06 UTC, the peak in the diurnal mode of observed rainfall frequency is occurring from about 105°W to 95°W in the central high plains (Knievel et al. 2004). This is significant because, at this time and location, the strongest propagating signal of rainfall frequency in the United States occurs. MCSs are beginning to organize, usually with the help of a strengthening low-level jet, and propagate to the east across the western high plains. Evidence of this can be seen from the plots of the averaged observed rainfall at forecast hours 03-12 (Fig. 3). A maximum in observed rainfall centered around 103°W at 03 UTC moves east at a speed of roughly 20 m/s and ends up centered around 95°W by 12 UTC. The area of rainfall appears to reach its peak intensity around 06 to 09 UTC. The implications of this strong propagating rainfall signal are that the mesoscale dynamics taking place while the model is “spinning up” starting at 00 UTC are very different from the dynamics that would be occurring while a model that was initialized at 12 UTC was “spinning up”. By initializing the model at 00 UTC, the challenge of resolving developing/ongoing propagating MCSs, systems that are already difficult to predict without having to worry about “spin up” issues, is exacerbated.

For the reasons mentioned above, it is reasonable to think that there should be problems in the averaged simulated precipitation around the times when “spin up” occurs. In the 5 km WRF-NMM, during the first 24 hours of the simulations, it appears that the simulated propagating axis of rainfall in the western part of the domain has been shifted later in time with the maximum corridor of precipitation occurring further east than in the observations. Thus, it appears that model “spin up”, which delays the onset of the
precipitation, has caused the entire propagating axis of rainfall to shift. It is further puzzling that the axis of propagation extends into the hours at which there should be an observed minimum in rainfall (forecast hours 15 to 18) caused by the dissipation of nighttime MCSs. This shift is significant because it means that the “spin up” effects are not limited to the times in which “spin up” occurs; areas downstream from where convection forms during the first 3 to 6 hours of the forecast are also indirectly affected because of timing and placement errors. Also puzzling is the observation that systems initiating too late also dissipate too late. To determine precisely why the dissipation of rainfall is delayed is beyond the scope of this study. Future research should verify that rainfall systems are in fact dissipating too late (as opposed to anomalous generation of distinct new convection at this time) and possible mechanisms for the late dissipation should be investigated. Some possibilities may include the improper simulation of the low-level jet. Past studies (e.g., Maddox 1983) have shown that the weakening of the low-level jet during the morning hours caused by the inertial oscillation (Bonner et al. 1968) is a major factor in the dissipation of MCSs during this time. If the low-level jet is properly simulated, the model may not be properly representing the mesoscale dynamics and circulations within MCSs that lead to dissipation. If systems are dissipating properly and new convection is firing prematurely, future research may want to investigate problems in the planetary boundary layer parameterization. If too much moisture is present leading to over-predicted instability, this could lead to the premature initiation of convection.

In the 5 km WRF-NMM simulations the representation of the propagating axis of rainfall drastically improves during the second diurnal cycle (forecast hours 24-48; Fig. 11). The location and timing match quite well with the observations. This is related to model “spin up” no longer being an issue at these times and is encouraging because it implies that errors from model “spin up” don’t appear to have indirect effects well after initialization. To see if high resolution models can continue to represent the diurnal cycle after hours 24-48, it would be useful to study forecasts going out to at least 72 hours, encompassing one more diurnal cycle.

Generally, the representation of the propagating axis of rainfall in the west in the 22 km WRF-ARW is poor compared to the 5 km WRF-NMM. It is hard to infer that the 22 km WRF-ARW is even simulating a propagating axis because the signal is so weak. However, it
appears that the representation of a propagating rainfall axis is better during forecast hours 00 to 24 than at forecast hours 24 to 48, opposite to the behavior in the 5 km WRF-NMM.

d. Comparison of hovmöller diagrams for the north and south portions of the domain

Because observations showed that different rainfall regimes occurred within the domain in this study, the domain was split into northern and southern halves to identify the features in the forecasts and observations associated with the different regions (Figs. 12 and 13, for the southern and northern halves, respectively). The latitude 38°N was chosen somewhat arbitrarily to separate the domain into two halves. The primary difference between the observed precipitation in the two regions is that the propagating signal, emanating around 104°W in the western part of the domain has the highest amplitude in the northern region, while the non-propagating signal in the eastern part of the domain from around 92°W to 80°W has the highest amplitude in the southern region. Also, from observations in the northern region, two axes of propagation can be identified; one emanating at about 104°W and another weaker axis emanating around 96°W, approximately the same longitude that the strongest part of the western propagation axis ends. This second, weaker axis is not observed in the southern part of the domain. In addition, it appears that there may be an extension of the first, stronger axis into forecast hours 24-48; albeit, it is much weaker at these hours. This is not surprising given that Carbone et al. (2002) found that the longevity of propagating rainfall episodes could be up to 60 hours.

Generally, the 5 km WRF-NMM forecasts correctly delineate between the two different precipitation regimes in the northern and southern regions. As in observations, the 5 km WRF-NMM forecasts show that the propagating signal in the western part of the domain has the highest amplitude in the northern region, while the non-propagating signal has the highest amplitude in the southern region (Figs. 11-12).

In the 22 km WRF-ARW, in the southern part of the domain, the non-propagating component of the diurnal cycle has the highest amplitude. In this respect, the 22 km WRF-ARW performs well, however, there is only a very subtle hint to a propagating component in the western portion which is in stark contrast to the 5 km WRF-NMM simulations that display a very clear propagating component. In the northern part of the domain in the
observations, the propagating component of the diurnal cycle in the west has the highest amplitude. This is not the case in the 22 km WRF-ARW forecasts which hint at very weak propagating signals between 103°W and 92°W at forecast hours 3 to 18 and between 108°W and 100°W at forecast hours 24 to 36. Also, the steep slope of the propagating rainfall axes in the Hovmöller diagrams from the 22 km WRF-ARW implies that the speed of propagation is too slow.

e. Time series of forecast hours 00-24 and 24-48 displayed using Hovmöller diagrams

More information can be gained about the forecasts by constructing time series of Hovmöller diagrams for the forecast periods 00-24 and 24-48 hours (Figs. 14-21 for the 5 km WRF-NMM; Figs. 22-29 for the 22 km WRF-ARW). The forecasts are separated into these two periods because of the differences in the diurnal cycle representations between the two periods discussed in previous sections. Analysis of these time series can show when the errors observed in the averaged plots occurred (such as displaced propagating rainfall axes), and these errors can be matched to the large scale weather regimes that were occurring at the time.

In the forecast and observations eastward propagating areas of rainfall are easily identified by diagonal streaks (Figs. 14-29). Multiple occurrences of precipitation “episodes” are observed, defined by Carbone et al. (2002) as time-space clusters of heavy precipitation that often result from sequences of organized convection. The episodes consist of slow, eastward propagating precipitation areas within which there are faster propagating rainfall areas. Examples during the time period of this study occurred 11-15 May and 3-7 June 2005. Also notice that it is possible to distinguish between periods in which diurnal or synoptic forcing were the dominant forcing mechanisms in generating convection. The periods in which synoptic scale forcing dominates feature areas of rainfall that move across the domain from west to east over the span of 2 or 3 days with little or no daily variation in rainfall amounts, whereas the periods in which diurnal forcing dominates feature areas of rainfall that propagate from west to east over a smaller part of the domain and dissipate within 12-24 hours on a daily basis. There are also periods when both forcing mechanisms help generate rainfall, similar to the aforementioned precipitation “episodes”.
The 00-24 hour forecasts from the 5 km WRF-NMM of the two weather systems that occurred 1-3 and 11-14 April (Fig. 14) did not have timing and location errors observed from the averaged plots (Fig. 11). These were systems that appeared to be mainly associated with synoptic forcing because there was little daily variation in the rainfall. During all of the other active periods that featured rainfall associated with precipitation “episodes” and diurnally forced rainfall, the time and location errors were a regular occurrence in the 00-24 hour forecasts. Periods in which this could clearly be seen were 11-13, 24, and 28-30 May (Fig. 15); and 7-8 and 16-17 June (Fig. 16). During some of the periods in which rainfall was clearly only generated by diurnal forcing mechanisms, the 00-24 hour forecasts badly underestimated or completely missed areas of propagating rainfall; for example 20-29 June (Fig. 16) and 5-10 July (Fig. 17). These were also periods in which non-propagating rainfall areas were overestimated.

The 24-48 hour forecasts from the 5 km WRF-NMM (Figs. 18-21) overestimated rainfall much of the time. However, the timing and location of propagating systems improved. For example, on 24 May the timing and location of a propagating system was much better for the 24-48 hour forecast (though, there were problems with the amount and timing of rainfall when the systems should have been dissipating). Many of the observed rainfall areas that seemed to be purely diurnally forced and were completely missed by the 00-24 hour forecast were simulated well by the 24-48 hour forecasts. The best examples of this were 25-27 June (Fig. 20) and 4-11 July (Fig. 21). Figure 30 is a comparison of the latter period for both the 5 km WRF-NMM and 22 km WRF-ARW. Note that the 5 km WRF-NMM has significantly better forecasts during the 24 to 48 hour forecast period than for the 00-24 hour period. Also, for the 00-24 hour periods, neither model seems to have a significant advantage, however, during the 24-48 hour periods the 22 km WRF-ARW forecasts get much worse and the 5 km WRF-NMM forecasts get much better.

Generally, during many time periods the time series from both models appear very similar. Many apparently propagating rain streaks are observed in the 22 km WRF-ARW and the 5 km WRF-NMM forecasts. This is not surprising; past studies (e.g., Bukovsky and Kain 2004) have shown that models using traditional CPSs can appear to simulate propagating MCSs. However, the propagating signal in the diurnally averaged plots is much more clearly seen in the 5 km WRF-NMM forecasts. Davis et al. (2003) noted a similar
finding in a mesoscale model using a traditional CPS. To see how the propagating signal was eroded from the long term averages they computed a series of time-longitude frequency diagrams for increasing averaging periods. They found that “as the averaging time increased the propagating component of the rainfall de-correlated rapidly such that by the time the averaging interval exceeded approximately 7-11 days the time-longitude frequency diagrams strongly resembled the 2 month climatology.” They concluded that the rain streaks observed in the forecasts were not phase locked to the diurnal cycle like they are in the real atmosphere. In other words, because of timing and placement errors, when the averaging was performed the propagating signal was drowned out. Results in this study indicate that the rainfall forecasts from the 5 km WRF-NMM appear to be phase locked to the diurnal cycle because the long term averaging does not drown out the propagating signal. This does not appear to be the case in the 22 km WRF-ARW forecasts because of the weaker propagating signal, especially during the 24-48 hour period.

5. Summary and future work

This paper examined the representation of the diurnal cycle by a 5 km grid-spacing model that did not use a CPS and compared it with a 22 km grid-spacing version of the WRF-ARW model that used a CPS. The historic lack of progress in improving QPFs is often blamed on CPSs because CPSs are likely unable to simulate the mesoscale dynamics leading to propagating convection (i.e., the unrealistic treatment of downdrafts and resulting poor representation of cold-pools in models with grid-spacing above 10 km; Davis et al. 2003). Also, Bukovsky and Kain (2004) note that, because CPSs act independently in individual model columns (Kain and Fritsch 1998), without a substantial mesoscale circulation, precipitation features will act like a collection of isolated entities not allowing propagation. The major findings from this study are summarized below.

There are distinct oscillations in the diurnally averaged ETS and bias skill measures. In the 5 km WRF-NMM, peaks in the bias scores are caused by the over-prediction of rainfall during the times at which the non-propagating rainfall signal in the eastern part of the domain is at its maximum amplitude. In the 22 km WRF-ARW the peaks in the bias scores are caused by a difference in phase between the times at which the maximum amplitude of
the non-propagating component in the eastern part of the domain is simulated and the times it is observed.

The ETSs from both models peak at the same times, which correspond to the times at which the propagating component of the diurnal cycle in the western part of the domain is at its maximum amplitude. It is speculated that this occurs because the organized MCSs that are responsible for the propagating signal are inherently more predictable than the more random, chaotic convection responsible for the non-propagating signal.

Overall, the ETSs from the 22 km WRF-ARW were higher at virtually all forecast hours and rainfall thresholds. Because the diurnally averaged Hovmöller plots indicate that the 5 km WRF-NMM had a much better representation of the diurnal cycle, this study is further evidence of how traditional verification measures can be misleading when applied to high resolution forecasts.

The 5 km WRF-NMM was generally able to delineate between the two different precipitation regimes in the northern and southern portions of the model domain, separated by 38°N. The 22 km WRF-ARW was not. The amplitude of the propagating rainfall signal was greater (less) than the non-propagating (propagating) component in the northern (southern) portion of the domain.

Because of model “spin up” the axes of propagating rainfall from the 5 km WRF-NMM forecasts in both regimes were displaced later in time and east of what was observed so that indirect effects of “spin up” were probably present up to 24 hours into the forecasts. At forecast hours 24-48 the 5 km WRF-NMM corrected the issues associated with “spin up” and depicted the timing and location of propagating and non-propagating areas of rainfall extremely well. The diurnal cycle representation from the 22 km WRF-ARW was worse than the 5 km WRF-NMM at all forecast hours. However, the 00-24 hour period appeared to be slightly better than the 24-48 hour period, which was opposite to what was observed in the 5 km WRF-NMM forecasts that were more affected by model “spin up” during the 00-24 hour forecast period.

Time series of Hovmöller diagrams covering monthly time periods revealed that the most notable difference between forecast hours 00-24 and 24-48 occurred during time periods in which most of the rainfall appeared to be generated through diurnal forcing mechanisms. During these times, at forecast hours 00-24, the 5 km WRF-NMM had trouble
even generating areas of precipitation; however, at forecast hours 24-48 there was major improvement. This was encouraging because Carbone et al. (2002) noted that the dominance of diurnal forcing corresponds to low skill in the dynamical prediction of convective precipitation. Also, the 5 km WRF-NMM forecasts were far better during the 24-48 hour period than the 22 km WRF-ARW forecasts. Generally, the 5 km WRF-NMM simulated propagating precipitation with the correct timing and location.

The 5 km WRF-NMM has a much better diurnal cycle representation than the 22 km WRF-ARW because the 5 km WRF-NMM is able to simulate propagating rainfall axes while the 22 km WRF-ARW is far less proficient and only displays weak evidence of propagation axes. Our results imply that the rainfall forecasts from the 5 km WRF-NMM are phase locked to the diurnal cycle just like in the real atmosphere while the 22 km WRF-ARW rainfall forecasts are not.

The year that this study was conducted was not a typical year with respect to rainfall. Much of the Midwest, especially portions of Iowa, Illinois, and Missouri experienced a severe drought. It will be necessary to see if the results from this study are consistent with those obtained using simulations conducted during other years. Future work should also focus on the mechanisms that are leading to what appears to be timing errors associated with the dissipation of MCSs during the first 24 hours of the 5 km WRF-NMM forecasts. Also, future studies should investigate what is causing heavier areas of rainfall to “spin up” faster than lighter areas.

**Acknowledgements.** This research was funded by Baker Endowment Fund 497-01-78-3803 and NSF grant ATM-0537043. Data from the 22 km WRF-ARW and 5 km WRF-NMM were obtained from the NCAR Mass Store with the assistance of Wei Wang for the 22 km data and Jim Bresch for the 5 km data.

**REFERENCES**


Chen, F., and J. Dudhia, 2001: Coupling an Advanced Land-Surface/Hydrology Model with the Penn State/ NCAR MM5 Modeling System. Part I: Model Description and


Soc., 85, 955–965.


<table>
<thead>
<tr>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>16</td>
<td>01</td>
<td>16</td>
</tr>
<tr>
<td>02*</td>
<td>17</td>
<td>02</td>
<td>17*</td>
</tr>
<tr>
<td>03*</td>
<td>18</td>
<td>03</td>
<td>18*</td>
</tr>
<tr>
<td>04</td>
<td>19</td>
<td>04</td>
<td>19*</td>
</tr>
<tr>
<td>05*</td>
<td>20</td>
<td>05</td>
<td>20*</td>
</tr>
<tr>
<td>06*</td>
<td>21</td>
<td>06</td>
<td>21*</td>
</tr>
<tr>
<td>07</td>
<td>22</td>
<td>07*</td>
<td>22</td>
</tr>
<tr>
<td>08*</td>
<td>23</td>
<td>08</td>
<td>23*</td>
</tr>
<tr>
<td>09*</td>
<td>24</td>
<td>09</td>
<td>24*</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>10</td>
<td>25*</td>
</tr>
<tr>
<td>11</td>
<td>26*</td>
<td>11</td>
<td>26</td>
</tr>
<tr>
<td>12*</td>
<td>27</td>
<td>12</td>
<td>27*</td>
</tr>
<tr>
<td>13</td>
<td>28</td>
<td>13</td>
<td>28*</td>
</tr>
<tr>
<td>14</td>
<td>29</td>
<td>14</td>
<td>29*</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>15</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 1: Dates in which data from both the 5 km WRF-NMM and 22 km WRF-ARW models were available. The dates denoted by an asterisk specify days in which data from one or both models was not available. There were a total of 91 days in which data was available from both models.
<table>
<thead>
<tr>
<th>Forecast Hour</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.260</td>
<td>1.452</td>
<td>0.192</td>
<td>1.876</td>
<td>0.136</td>
<td>1.675</td>
<td>0.134</td>
<td>0.971</td>
<td>0.157</td>
<td>0.538</td>
<td>0.184</td>
<td>0.370</td>
</tr>
<tr>
<td>6</td>
<td>0.585</td>
<td>1.329</td>
<td>0.760</td>
<td>1.358</td>
<td>0.859</td>
<td>0.960</td>
<td>1.061</td>
<td>0.623</td>
<td>1.348</td>
<td>0.495</td>
<td>1.783</td>
<td>0.458</td>
</tr>
<tr>
<td>9</td>
<td>0.951</td>
<td>1.448</td>
<td>1.103</td>
<td>1.292</td>
<td>1.258</td>
<td>0.969</td>
<td>1.656</td>
<td>0.725</td>
<td>2.237</td>
<td>0.685</td>
<td>2.969</td>
<td>0.716</td>
</tr>
<tr>
<td>12</td>
<td>1.203</td>
<td>1.562</td>
<td>1.395</td>
<td>1.456</td>
<td>1.669</td>
<td>1.203</td>
<td>2.240</td>
<td>1.010</td>
<td>2.970</td>
<td>0.912</td>
<td>3.878</td>
<td>0.866</td>
</tr>
<tr>
<td>15</td>
<td>1.367</td>
<td>1.827</td>
<td>1.686</td>
<td>1.837</td>
<td>2.224</td>
<td>1.570</td>
<td>3.196</td>
<td>1.296</td>
<td>4.304</td>
<td>1.140</td>
<td>5.427</td>
<td>0.996</td>
</tr>
<tr>
<td>18</td>
<td>1.347</td>
<td>2.183</td>
<td>1.741</td>
<td>2.742</td>
<td>2.271</td>
<td>2.401</td>
<td>3.323</td>
<td>1.727</td>
<td>4.638</td>
<td>1.319</td>
<td>6.428</td>
<td>1.233</td>
</tr>
<tr>
<td>21</td>
<td>1.261</td>
<td>1.910</td>
<td>1.710</td>
<td>2.360</td>
<td>2.302</td>
<td>2.051</td>
<td>3.441</td>
<td>1.455</td>
<td>5.162</td>
<td>1.145</td>
<td>7.485</td>
<td>0.952</td>
</tr>
<tr>
<td>24</td>
<td>1.426</td>
<td>1.683</td>
<td>2.039</td>
<td>1.900</td>
<td>2.702</td>
<td>1.468</td>
<td>3.839</td>
<td>1.006</td>
<td>5.312</td>
<td>0.822</td>
<td>7.197</td>
<td>0.852</td>
</tr>
<tr>
<td>27</td>
<td>1.607</td>
<td>1.421</td>
<td>2.212</td>
<td>1.355</td>
<td>2.814</td>
<td>1.113</td>
<td>3.887</td>
<td>0.907</td>
<td>5.328</td>
<td>0.842</td>
<td>6.931</td>
<td>0.864</td>
</tr>
<tr>
<td>30</td>
<td>1.479</td>
<td>1.152</td>
<td>2.049</td>
<td>1.230</td>
<td>2.469</td>
<td>1.058</td>
<td>3.121</td>
<td>0.878</td>
<td>3.975</td>
<td>0.832</td>
<td>5.130</td>
<td>0.893</td>
</tr>
<tr>
<td>33</td>
<td>1.670</td>
<td>1.305</td>
<td>1.835</td>
<td>1.242</td>
<td>2.033</td>
<td>1.054</td>
<td>2.483</td>
<td>0.894</td>
<td>3.198</td>
<td>0.913</td>
<td>4.029</td>
<td>0.963</td>
</tr>
<tr>
<td>36</td>
<td>1.702</td>
<td>1.430</td>
<td>1.730</td>
<td>1.344</td>
<td>1.869</td>
<td>1.149</td>
<td>2.195</td>
<td>0.917</td>
<td>2.761</td>
<td>0.815</td>
<td>3.417</td>
<td>0.818</td>
</tr>
<tr>
<td>39</td>
<td>1.690</td>
<td>1.730</td>
<td>1.779</td>
<td>1.665</td>
<td>2.005</td>
<td>1.413</td>
<td>2.367</td>
<td>1.084</td>
<td>2.989</td>
<td>0.976</td>
<td>3.536</td>
<td>1.033</td>
</tr>
<tr>
<td>42</td>
<td>1.551</td>
<td>2.078</td>
<td>1.735</td>
<td>2.541</td>
<td>1.953</td>
<td>2.242</td>
<td>2.364</td>
<td>1.462</td>
<td>3.091</td>
<td>1.129</td>
<td>4.195</td>
<td>1.050</td>
</tr>
<tr>
<td>45</td>
<td>1.386</td>
<td>1.812</td>
<td>1.664</td>
<td>2.249</td>
<td>1.975</td>
<td>1.997</td>
<td>2.556</td>
<td>1.291</td>
<td>3.696</td>
<td>0.933</td>
<td>5.282</td>
<td>0.771</td>
</tr>
<tr>
<td>48</td>
<td>1.522</td>
<td>1.624</td>
<td>2.016</td>
<td>1.803</td>
<td>2.538</td>
<td>1.362</td>
<td>3.434</td>
<td>0.843</td>
<td>4.553</td>
<td>0.621</td>
<td>5.973</td>
<td>0.582</td>
</tr>
<tr>
<td>mean</td>
<td>1.314</td>
<td>1.622</td>
<td>1.603</td>
<td>1.766</td>
<td>1.944</td>
<td>1.480</td>
<td>2.584</td>
<td>1.067</td>
<td>3.477</td>
<td>0.882</td>
<td>4.615</td>
<td>0.839</td>
</tr>
</tbody>
</table>

Table 2: Bias Scores from the 5 km WRF-NMM and 22 km WRF-ARW calculated using the method of summing all of the contingency table elements from all of the forecasts. Forecast hours at which relative maxima are observed are shaded light gray for the 5 km WRF-NMM and darker gray for the 22 km WRF-ARW.
<table>
<thead>
<tr>
<th>Forecast Hour</th>
<th>Rainfall Threshold (Inches)</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
<th>NMM</th>
<th>ARW</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.100</td>
<td>0.167</td>
<td>0.066</td>
<td>0.118</td>
<td>0.040</td>
<td>0.084</td>
<td>0.025</td>
<td>0.058</td>
<td>0.017</td>
<td>0.040</td>
<td>0.014</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.146</td>
<td>0.164</td>
<td>0.117</td>
<td>0.142</td>
<td>0.084</td>
<td>0.112</td>
<td>0.059</td>
<td>0.072</td>
<td>0.040</td>
<td>0.045</td>
<td>0.025</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.198</td>
<td>0.209</td>
<td>0.141</td>
<td>0.162</td>
<td>0.089</td>
<td>0.110</td>
<td>0.048</td>
<td>0.061</td>
<td>0.026</td>
<td>0.034</td>
<td>0.014</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.211</td>
<td>0.218</td>
<td>0.139</td>
<td>0.165</td>
<td>0.083</td>
<td>0.105</td>
<td>0.045</td>
<td>0.051</td>
<td>0.025</td>
<td>0.017</td>
<td>0.014</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.204</td>
<td>0.202</td>
<td>0.127</td>
<td>0.137</td>
<td>0.075</td>
<td>0.088</td>
<td>0.036</td>
<td>0.047</td>
<td>0.023</td>
<td>0.033</td>
<td>0.018</td>
<td>0.011</td>
</tr>
<tr>
<td>18</td>
<td>1.00</td>
<td>0.181</td>
<td>0.174</td>
<td>0.105</td>
<td>0.101</td>
<td>0.061</td>
<td>0.066</td>
<td>0.031</td>
<td>0.054</td>
<td>0.019</td>
<td>0.043</td>
<td>0.013</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.189</td>
<td>0.181</td>
<td>0.100</td>
<td>0.108</td>
<td>0.060</td>
<td>0.068</td>
<td>0.033</td>
<td>0.045</td>
<td>0.018</td>
<td>0.034</td>
<td>0.010</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.186</td>
<td>0.176</td>
<td>0.099</td>
<td>0.096</td>
<td>0.056</td>
<td>0.056</td>
<td>0.030</td>
<td>0.029</td>
<td>0.016</td>
<td>0.016</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.152</td>
<td>0.160</td>
<td>0.099</td>
<td>0.103</td>
<td>0.066</td>
<td>0.069</td>
<td>0.037</td>
<td>0.041</td>
<td>0.019</td>
<td>0.022</td>
<td>0.010</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.143</td>
<td>0.155</td>
<td>0.112</td>
<td>0.126</td>
<td>0.076</td>
<td>0.087</td>
<td>0.044</td>
<td>0.040</td>
<td>0.026</td>
<td>0.021</td>
<td>0.016</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.164</td>
<td>0.173</td>
<td>0.122</td>
<td>0.126</td>
<td>0.079</td>
<td>0.080</td>
<td>0.043</td>
<td>0.042</td>
<td>0.022</td>
<td>0.017</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>33</td>
<td>0.10</td>
<td>0.160</td>
<td>0.149</td>
<td>0.112</td>
<td>0.110</td>
<td>0.066</td>
<td>0.063</td>
<td>0.031</td>
<td>0.031</td>
<td>0.015</td>
<td>0.012</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.160</td>
<td>0.154</td>
<td>0.107</td>
<td>0.094</td>
<td>0.063</td>
<td>0.056</td>
<td>0.032</td>
<td>0.033</td>
<td>0.020</td>
<td>0.010</td>
<td>0.012</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.142</td>
<td>0.142</td>
<td>0.089</td>
<td>0.074</td>
<td>0.052</td>
<td>0.041</td>
<td>0.033</td>
<td>0.033</td>
<td>0.020</td>
<td>0.033</td>
<td>0.010</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.131</td>
<td>0.154</td>
<td>0.070</td>
<td>0.084</td>
<td>0.040</td>
<td>0.045</td>
<td>0.023</td>
<td>0.032</td>
<td>0.012</td>
<td>0.036</td>
<td>0.005</td>
<td>0.038</td>
</tr>
<tr>
<td>47</td>
<td>1.00</td>
<td>0.134</td>
<td>0.146</td>
<td>0.071</td>
<td>0.075</td>
<td>0.043</td>
<td>0.039</td>
<td>0.023</td>
<td>0.019</td>
<td>0.012</td>
<td>0.013</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>0.161</td>
<td>0.172</td>
<td>0.104</td>
<td>0.114</td>
<td>0.065</td>
<td>0.073</td>
<td>0.036</td>
<td>0.043</td>
<td>0.021</td>
<td>0.027</td>
<td>0.012</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Table 3: Equitable Threat Scores from the 5 km WRF-NMM and 22 km WRF-ARW calculated using the method of summing all of the contingency table elements from all of the forecasts. Forecast hours at which relative maxima are observed are shaded light gray for the 5 km WRF-NMM and darker gray for the 22 km WRF-ARW.
Figure 1: Domains for a) the 22 km WRF-ARW, b) the 5 km WRF-NMM, and c) the analyses.
Day 1 (Hours 03-12): WRF-NMM QPF vs. Stage 4 OBS

Figure 3: Top images display forecast precipitation from the 5 km WRF-NMM averaged over 100 forecasts at forecast hours 3-12. Bottom images display the Stage 4 observed precipitation estimates valid from the same time as the forecasts. The forecast hour is shown in the box in the top right portion of each image.
Figure 4: Same as figure 3 except for the forecast hours 15-24.
Figure 5: Same as figure 3 except for the forecast hours 27-36.
Day 2 (Hours 39-48): WRF-NMM QPF vs. Stage 4 OBS

Figure 6: Same as figure 3 except for forecast hours 39-48.
Day 1 (Hours 03-12): WRF-ARW QPF vs. Stage 4 OBS

Figure 7: Top images display forecast precipitation from the 22 km WRF-ARW averaged over 106 forecasts at forecast hours 3-12. Bottom images display the Stage 4 observed precipitation estimates valid from the same time as the forecasts. The forecast hour is shown in the box in the top right portion of each image.
Day 1 (Hours 15-24): WRF-ARW QPF vs. Stage 4 OBS

Figure 8: Same as figure 7 except for the forecast hours 15-24.
Day 2 (Hours 27-36): WRF-ARW QPF vs. Stage 4 OBS

Figure 9: Same as figure 7 except for the forecast hours 27-36.
Day 2 (Hours 39-48): WRF-ARW QPF vs. Stage 4 OBS

Figure 10: Same as figure 7 except for the forecast hours 39-48.
Figure 11: Hovmöller diagrams of diurnally averaged rainfall from the 22 km WRF-ARW (left), 5 km WRF-NMM (center), and stage IV observations (right) for the area between 29° N and 49° N. Note that because the 22 km WRF-ARW and 5 km WRF-NMM over-predict rainfall so much the scale for the observed rainfall is exactly half of the observed rainfall. This is done to make it easier to compare the trends between the 3 plots.
Figure 12: Same as figure 11 except for the area between 29° N and 38° N.
Figure 13: Same as figure 11 except for the area between 38° N and 49° N.
Figure 14: Hovmöller diagram for (a) 1-15 April 2005, and (b) 16-30 April 2005, for the forecast hours 00-24. The shading represents the model forecasts (5 km WRF-NMM) and the contours represent the observations. The three contour levels match the three levels of shading. Missing days are shaded gray.
Figure 15: Same as figure 14 except for the period of May 2005.
Forecast Hours 00 – 24

Figure 16: Same as figure 14 except for the period of June 2005.
Forecast Hours 00 – 24

Figure 17: Same as figure 14 except for the period of July 2005.
Figure 18: Hovmöller diagram for (a) 1-15 April 2005, and (b) 16-30 April 2005, for the forecast hours 24-48. The shading represents the model forecasts (5 km WRF-NMM) and the contours represent the observations. The three contour levels match the three levels of shading. Missing days are shaded gray.
Forecast Hours 24 – 48

Figure 19: Same as figure 18 except for the period of May 2005.
Figure 20: Same as figure 18 except for the period of June 2005.
Forecast Hours 24 – 48

Figure 21: Same as figure 18 except for the period of July 2005.
Figure 22: Hovmoller diagram for (a) 1-15 April 2005, and (b) 16-30 April 2005, for the forecast hours 00-24. The shading represents the model forecasts (22 km WRF-ARW) and the contours represent the observations. The three contour levels match the three levels of shading. Missing days are shaded gray.
Figure 23: Same as figure 22 except for the period of May 2005.
Figure 24: Same as figure 22 except for the period of June 2005.
Figure 25: Same as figure 22 except for the period of July 2005.
Figure 26: Hovmöller diagram for (a) 1-15 April 2005, and (b) 16-30 April 2005, for the forecast hours 24-48. The shading represents the model forecasts (22 km WRF-ARW) and the contours represent the observations. The three contour levels match the three levels of shading. Missing days are shaded gray.
Figure 27: Same as figure 26 except for the period of May 2005.
Figure 28: Same as figure 26 except for the period of June 2005.
Figure 29: Same as figure 26 except for the period of July 2005.
Figure 30: Hovmöller diagram of forecast (shaded) vs. observed precipitation (contours) for forecast hours 00-24 (left) and 24-48 (right) for the period July 4-11 for the 5 km WRF-NMM (top) and the 22 km WRF-ARW (bottom).
CONTRIBUTIONS OF MIXED PHYSICS AND PERTURBED INITIAL CONDITIONS TO THE SKILL AND SPREAD OF PRECIPITATION FORECASTS FROM A 16 MEMBER WRF ENSEMBLE

A paper submitted to Weather and Forecasting

Adam J. Clark, William A. Gallus Jr., and Tsing-Chang Chen

Abstract

This paper will identify and compare the impacts of perturbed initial and lateral boundary conditions and mixed-physics on the spread and skill of precipitation forecasts from a regional 16 member Weather Research and Forecasting (WRF) model ensemble integrated for 120 hours at 15 km grid-spacing for a domain covering the Midwest United States. Forecast skill is compared using deterministic forecasts derived from each ensemble using the probability matching technique and using probabilistic forecasts from each ensemble.

Results revealed that the mixed-physics (MP) and perturbed initial and lateral boundary conditions (PI) ensembles had similar skill when the deterministic forecasts were evaluated using equitable threat scores (ETSs). However, when the area under the relative operating characteristic curve (ROC score) was used to evaluate the probabilistic forecasts, the MP had higher skill at the beginning of the forecast while the PI ensemble had higher skill at the end of the forecast. This behavior was directly related to the spread. In the MP ensemble, because the initial and lateral boundary conditions were the same for each ensemble member, the spread stopped increasing after about 24 hours, and shortly after this time the PI ensemble ROC scores surpassed the MP ensemble ROC scores. However, during the first 24 hours of the forecast, greater spread in the MP ensemble forecasts led to higher ROC scores than in the PI ensemble. This demonstrates the importance of perturbed initial and lateral boundary conditions in ensembles using limited area models.
1. Introduction

Many different techniques for constructing ensembles have been developed and tested, including the perturbation of an initial state (e.g., Toth and Kalnay 1997; Palmer et al. 1992; Molteni et al. 1996; Houtekamer 1996), using different combinations of physical parameterizations (e.g., Stensrud et al. 2000; Du et al. 2004), using different numerical models (e.g., Du et al. 2004; Wandishin et al. 2001), and using different combinations of these techniques. These methods try to increase the skill of the ensemble by introducing independent information so that all possible states of the atmosphere are simulated (Roebber et al. 2004). Currently, methods for generating ensembles require further research, because the current spread in ensemble members in research systems lacks the breadth of nature (Fritsch and Carbone 2004).

This paper addresses finding the best way to construct an ensemble and finding what temporal scales are best suited for a specific ensemble design. It builds on the recent success of the National Center for Environmental Prediction’s (NCEP) Short Range Ensemble Forecast (SREF) System (Du et al. 2003). Starting in April 2000, NCEP’s SREF system was composed of 10 different versions of regional models run in real-time that covered a domain over the contiguous United States. Because of its initial success, the SREF system has now evolved to include 21 different ensemble members run with grid-spacing ranging from 32-45 km and integrated 3 days (Du et al. 2006). The ensemble members include mixed-physics, mixed initial conditions, and mixed model formulations. Recently six Weather Research and Forecasting (WRF) model (Skamarock et al. 2001; Michalakes et al. 2001) members were added that increased the spread and skill of the forecasts (Du et al. 2006). Eventually, NCEP plans on transitioning the entire SREF system to the WRF model. It is hoped that the results from this study will aid in the construction of NCEP’s all WRF SREF and other similar ensembles.

The objectives of this paper are to compare the spread and skill of precipitation forecasts in two 8 member WRF model ensembles; one constructed using mixed-physics (MP) and one constructed using perturbed initial and lateral boundary conditions (PI). Forecasts from a 16 member ensemble constructed using the members from both mixed-
physics and perturbed initial and lateral boundary conditions ensembles (FULL) will also be examined.

The remainder of the paper is organized as follows. Section 2 includes a description of the data and methodology; section 3 includes the results, and section 4 contains a summary and recommendations for future work.

2. Data and methodology

a. Ensemble member specifications

The domain of the ensemble system covers a large portion of the Midwest (Fig. 1). The 16 WRF members were integrated daily at 00 UTC for a period of 120 hours on a 15 km grid from 12 February to 15 April 2006. The general strategy in constructing the ensemble system was to vary the physical parameterization schemes or dynamic cores in 8 members while using the same initial and lateral boundary conditions. This will be referred to as the MP ensemble, hereafter. Gallus and Bresch (2006) showed that dynamics core changes caused roughly similar rainfall forecast spread in a set of warm season cases as the physics changes. Also, the initial and lateral boundary conditions in 8 members were varied while using the same physical parameterization schemes and dynamic core. This will be referred to as the PI ensemble, hereafter. By splitting the 16 members into two groups strategies can be examined independently and compared to output from the 16 member ensemble system (FULL ensemble, hereafter).

The 8 MP members use initial and lateral boundary conditions from the operational version of NCEP’s Global Forecast System (GFS; Environmental Modeling Center 2003). The Advanced Research WRF (ARW) dynamics core (Klemp et al. 2000; Wicker et al. 2002; Skamarock et al. 2005) was used in 6 of the MP members and the nonhydrostatic mesoscale model (NMM) dynamics core (Janjic 2003) was used in 2 of the MP members. Microphysical and convective parameterization schemes were varied, a decision based on Jankov et al. 2005 who showed that differing convective parameterization schemes was the most efficient way to substantially increase spread. The Kain-Fritsch (KF; Kain and Fritsch 1993) and Betts-Miller-Janjic (BMJ; Betts 1986; Betts and Miller 1986; Janjic 1994)
cumulus parameterization schemes; and Lin et al. (1983), Eta Ferrier (Ferrier et al. 2002), and the WRF single-moment 6-class (Skamarock et al. 2005) microphysical parameterization schemes were used in the MP ensemble.

Six WRF-ARW and two WRF-NMM members were used instead of a more equally balanced ensemble diversified more through mixed model formulations, because the WRF NMM core only had one microphysical parameterization scheme and two convective parameterization schemes available at the time of the simulations. Therefore, the maximum number of members possible using mixed-physics with the NMM core is two, while the ARW core, which has 9 microphysical schemes and 3 convective schemes available, could have 27 mixed physics members. Gallus and Bresch (2006) show that spread from the use of different dynamic cores is a function of the physics schemes used, and can be comparable to that from differing physics, so that as more physics options become available in NMM, further studies should be performed.

The 8 PI members are all run using the NCEP NMM core with the KF cumulus parameterization scheme and the Eta Ferrier microphysical parameterization scheme. All 8 members have unique initial and lateral boundary conditions that consist of 4 positive and 4 negative perturbations taken from the GFS ensemble members. A complete summary of all ensemble member specifications is contained in Table 1.

b. Evaluation of the skill and spread of the precipitation forecasts

This study will focus on forecasts of 3, 6, 12 and 24 hour accumulated rainfall. The observations used for verification are from the stage IV (Baldwin and Mitchell 1997) multi-sensor rainfall estimates. To perform the verification, stage IV data was interpolated to the 15 km WRF-NMM grid and the WRF-ARW grid was also interpolated to the 15 km WRF-NMM grid. This was done using a water budget method that conserved the total volume of liquid in the domain, a procedure routinely used at NCEP. Verification was performed for both deterministic and probabilistic forecasts derived from the ensemble system. The skill measures used for each type of forecast are described in the following two sections.
1) Verification of Deterministic Forecasts

Deterministic forecasts of precipitation are calculated from the ensemble system using a technique known as probability matching. This technique can be used to blend data types with different spatial and temporal properties and is especially useful when one data type has a better spatial representation while the other has greater accuracy (Ebert 2001). A detailed description of the calculation of model forecast rainfall using probability matching is contained in Ebert (2001). Basically, if we assume that the best spatial representation of rainfall is given by the ensemble mean and the best frequency distribution of rain amounts is given by the model quantitative precipitation forecasts (QPFs), we can reassign the rain amounts from the ensemble mean using values randomly selected from the distribution of individual model QPFs. This corrects for the large bias in rain area and underestimation of rain amounts that are caused by the averaging process and results in forecast rain fields that are much more realistic. Using an ensemble consisting of models run at different operational centers, Ebert (2001) concluded that the probability matching method is the most useful deterministic ensemble rainfall forecast for forecasters.

Equitable threat score (ETS; Schaefer 1990) and bias are used to verify the forecasts. These are calculated using a contingency table in which the members are hits, false alarms, misses, and correct negatives denoted by a, b, c, and d, respectively. In terms of these contingency table elements, ETS and bias are expressed as:

\[
ETS = \frac{da - bc}{(b + c) + (da - bc)}
\]  

(1)

\[
Bias = \frac{a + b}{a + c}
\]  

(2)

The ETS measures the fraction of events correctly predicted, accounting for hits associated with random chance. ETSs range from \(-1/3\) to \(1\); scores below 0 have no skill and 1 represents a perfect score. Bias scores represent the ratio of the frequency of forecast events to the frequency of observed events. In other words, it indicates over or under-prediction of events. Bias values range from 0 to infinity, with 1 being a perfect score.
When comparing ETSs, bias scores should also be examined because ETSs are sensitive to high bias scores (Hamill 1999).

Average ETSs and bias scores can be calculated by averaging scores from daily forecasts or summing the contingency table elements from all the forecasts and computing the scores from the summed elements. The first method gives equal weight to each forecast while the second gives more weight to larger precipitation events. The first part of the results section will include a comparison of the two methods; after this, all scores are calculated using the method of summed contingency table elements.

2) VERIFICATION OF PROBABILITY FORECASTS

The area under the relative operating characteristic curve (ROC score hereafter; Mason 1982) will be used to evaluate the forecasts from the 16 member FULL ensemble as well as to compare the 8 member MP and PI ensembles. The ROC score is closely related to the economic value of a forecast system (e.g. Mylne 1999; Richardson 2000b, 2001). Its purpose is to provide information on the characteristics of systems upon which management decisions can be made. The derivation of the ROC score is based on the members of a contingency table. To construct the ROC curve the probability of detection \[\text{POD}=a/(a+c)\] of the forecast system is plotted against its probability of false detection \[\text{POFD}=b/(b+d)\]. In this study, the area under this curve, which begins with the points (0,0) and ends with (1,1), is calculated using the trapezoidal method which can introduce errors when comparing ensembles with a different number of members. An idealized plot of two identical ROC curves; one using one point to compute the ROC score and the other using three points to compute the ROC score is presented in Figure 1 to illustrate how these errors occur. The range of values for the ROC score are 0.5 (no skill; points lie along diagonal line \([0,0],[1,1]\)) to 1.0 (perfect forecast; curve approaches upper-left hand corner). ROC scores will be computed for rainfall accumulated at 3, 6, 12, and 24 hour intervals.

3) EVALUATION OF ENSEMBLE SPREAD

Three methods are used to assess the spread of each 8 member ensemble and the full
16 member ensemble. First, rank histograms (Hamill 2001) are constructed by repeatedly tallying the rank of the rainfall observation relative to forecast values from an ensemble sorted from highest to lowest. Generally, a flat rank histogram is a sign of reliability; a U-shaped rank histogram indicates a lack of spread in the ensemble; and an upside-down U-shaped rank histogram indicates too much spread in the ensemble (Hamill 2001).

Second, root-mean-square differences (RMSD) are calculated by computed the root-mean-square error between all possible pairs of ensemble members and taking the average. The RMSD was first used by Lorenz (1965) to examine the growth rate of the difference between two or more solutions with slightly varied initial states. Lorenz (1965) used this method to gain information about predictability limits in atmospheric models. In our applications, the RMSD is used to give information about the rate at which the ensemble member solutions are diverging; in other words, the rate at which the ensemble spread increases. A steep positive slope indicates that the spread is increasing rapidly, while a steep negative slope indicates that the spread is decreasing rapidly. A negative slope would indicate that the solutions were actually converging; we would not expect to see this. The time at which the RMSD curve becomes flat can be interpreted as the time at which all solutions have completely diverged from one another.

Finally, the spread ratio is the inverse of the correspondence ratio defined by Stensrud and Wandishin (2000) as the ratio of the intersection of two or more fields to the union of the fields. For an ensemble, the spread ratio can be quantified as [adapted from Stensrud and Wandishin 2000]:

\[
SR_p = \frac{\sum_{i=1}^{n} \left\{ \frac{1}{m} \sum_{j=1}^{m} d(f_{i,j}) \right\} \geq P}{\sum_{i=1}^{n} f_{1,i} \cup f_{2,i} \cup f_{3,i} \cup \ldots \cup f_{m,i}}
\]

where,

\[
d(f_{i,j}) = \begin{cases} 
1 & \text{For } f_{i,j} \geq \text{threshold} \\
0 & \text{For } f_{i,j} < \text{threshold}
\end{cases}
\]
Stensrud and Wandishin (2000) note that the correspondence ratio (inverse of spread ratio) can provide more information than the RMSD because it can be calculated for different thresholds. Also, it is especially useful for evaluating discontinuous fields, such as precipitation because it can be used to evaluate the divergence of the forecast fields with time.

3. Results

a. Deterministic forecasts

Figures 3-4 show ETSs for a range of thresholds at forecast hour 24 calculated for precipitation accumulated during 3, 6, 12, and 24 hour intervals using the two methods discussed earlier. The deterministic forecasts are calculated from all 16 ensemble members using the probability matching method. These figures all show higher ETSs using method 1 (summed contingency table elements). In fact, for rainfall accumulated at 24 hour intervals, the ETS using method 1 are nearly 3 times larger than those of method 2. As the time interval in which rainfall accumulates is decreased the difference between the ETSs calculated using the two methods also decreases but is still higher for method 1, even at 3 hour accumulation periods. These differences occur because more weight is given to heavier and more widespread precipitation events using method 1. Because heavier, more widespread events are more likely to occur over an increased time interval, the difference in scores between the two methods increases with the increasing time interval.

Time series ETSs and bias scores from each 8 member ensemble and the full 16 member ensemble were constructed for 3, 6, 12, and 24 hour intervals at the 0.01, 0.25, and 1.00 inch rainfall thresholds (Figs. 5-8). First, some general trends consistent with the three sets of ETSs should be noted. The ETSs increase for the longer accumulation periods. Wandishin et al. (2001) explained that this is because the scores for the longer accumulation periods are affected less by timing errors. For example, if a model forecast predicted rainfall 4 hours too early, the 3 hour accumulation period may miss the entire event while the 6 hour period may capture most of the event. Also note the slopes of the ETS time series.
Intuitively we’d expect to see a general decreasing trend caused by the model and initial condition errors growing with time. However, when the time series reaches a constant value we can assume that the model has reached its limit of predictability and that any observed skill is equal to that of climatology. This would explain why the 0.01 inch threshold ETS time series flattens out at higher scores than the 1.00 inch ETS time series; rain events above 0.01 inches occur more frequently than rain events above 1.00 inch. Also, the time at which the ETS time series flattens occurs later in the forecast for the longer accumulation periods for the reasons discussed above.

The diurnal cycle of rainfall is having an impact on the scores as is evident by relative maxima occurring at 0300 LST and minima occurring at 1200 LST seen in the scores every 24 hours. The maxima (minima) correspond to the time at which the propagating component of the diurnal cycle in the Midwest is at its maximum (minimum) amplitude (Carbone et al. 2002). This can be most clearly seen at the 0.01 inch rainfall threshold; there is no signal from the diurnal cycle on the ETSs at the 1.00 inch rainfall threshold because errors have already completely saturated the scores at the 1.00 inch threshold before one 24 hour forecast period. At the 0.25 inch threshold, the errors drown out the diurnal cycle signal around the last 36 hours of the forecast period while the signal is clearly seen through the entire 120 hour forecast period at the 0.01 inch threshold. This implies that the models have at least some skill in forecasting areas of rainfall out to at least 120 hours.

There are notable differences between the ETSs of the MP ensemble and the PI ensemble (Figs. 5-8). The bias scores are also provided in these plots because Hamill (1999) noted that comparing ETSs from forecasts with differing bias scores can be misleading. Considering this, it is noted that for all accumulation periods and rainfall thresholds, the PI ensemble bias scores are higher. This may be a result of the differences in the convective parameterization used within the ensembles. All of the PI ensemble members use the Kain-Fritsch scheme while half of the MP ensemble members use the Kain-Fritsch scheme and the other half use the Betts-Miller-Janjic scheme. Past research (e.g., Gallus 1999) has shown that rainfall forecasts can be very sensitive to the choice of convective parameterization. It is quite possible that the Kain-Fritsch scheme may be causing the high bias scores in the PI ensemble.

Generally, Figures 5-8 show that the PI ensemble has the highest ETSs for the shorter
time intervals while the MP ensemble has the highest ETSs for longer time intervals. The largest difference between the ensembles is observed for the 0.01 inch rainfall threshold at 24 hour time intervals. At most times and rainfall thresholds the difference between the sets of scores is not very large.

b. Probabilistic forecasts

Figures 9-10 display average ROC scores for a range of thresholds at forecast hour 24 for precipitation accumulated at 3, 6, 12, and 24 hour intervals. Similar to the ETSs, the methods of summing the contingency table elements and averaging the daily scores are compared. Trends in these plots are similar to those observed from ETSs. At virtually every rainfall threshold the scores calculated using the sum of the contingency table elements are higher than the average daily scores. Also, as the time interval for which scores are calculated increases, the difference between the scores from the two methods increases. The reasoning is the same as for the ETSs.

Figures 11-14 display time series of average ROC scores at 0.01, 0.10, 0.25, 0.50, 0.75, and 1.00 inch rainfall thresholds calculated for 3, 6, 12, and 24 hour intervals. Note that because the trapezoidal method was used to calculate the area under the ROC curve, a fair comparison cannot be made between the FULL ensemble and the PI and MP ensembles. This is because of the difference in the number of members. The trapezoidal method will tend to give higher scores to the ensemble with the greater number of members even if the skill is equal (Fig. 2). The FULL ensemble scores are still included in the plots for reference.

Generally, at all of the time intervals and rainfall thresholds for which ROC scores are calculated, the MP ensemble has higher scores for a period at the beginning of the forecast and the PI ensemble has higher scores for the rest of the forecast (Figs. 11-14). This relates to the spread of each ensemble (discussed in next section). Basically, the MP ensemble members quickly generate variance but experience “spread saturation” near the beginning of the forecast. This is caused by the propagation of lateral boundary influences through the domain. The PI ensemble members generate variance more gradually until the end of the forecast period. The PI ensemble doesn’t quickly experience “spread saturation” because its lateral boundary conditions are different for each member. Problems associated
with lateral boundary conditions in limited area models are thoroughly discussed by Warner et al. (1997) and the importance of perturbing lateral boundaries are discussed in Nutter et al. (2004).

In general, at the 0.01 inch rainfall threshold, the times at which the ROC scores from the PI ensemble overtake the MP ensemble, are much later in the forecast than at the higher rainfall thresholds. For example, for the ROC scores calculated at 6 hour intervals at the 0.01 inch rainfall threshold, the PI ensemble ROC scores become larger than the MP ensemble ROC scores at forecast hour 72, while at the 0.10 inch rainfall threshold and all heavier ones this occurs at forecast hour 36.

It is also interesting that the FULL ensemble ROC scores and the PI ensemble ROC scores only decrease very slowly within the 120 hour forecast period. While starting off with higher ROC scores, the MP ensemble scores decrease much more rapidly because of lateral boundary condition influences (Nutter et al. 2004; Warner et al. 1997).

c. Ensemble spread

1) ROOT-MEAN-SQUARE DIFFERENCE (RMSD)

Figures 15-16 display time series of RMSD for the PI, MP, and FULL ensembles at 3, 6, 12, and 24 hour intervals. A very coherent diurnal oscillation is evident in the 3 ensembles at 3, 6, and 12 hour intervals with peak amplitudes corresponding to 2100 LST. RMSD is sensitive to rainfall amounts so that the peak amplitudes most likely occur at the times when the models generate the most rainfall, assuming displacement errors don’t vary substantially.

The RMSD curve for the MP ensemble starts off higher than the PI ensemble and then becomes lower than the PI ensemble RMSD at some point for the rest of the forecast. The forecast hours that this occurs at are 27, 30, 36, and 48 for the RMSD calculated at 3, 6, 12, and 24 hour intervals, respectively. These are also the times at which the RMSD curve for the MP ensemble flattens out. This means that the spread generated by the MP ensemble members becomes constant at these times (there are still diurnal oscillations but the oscillations occur around a constant value).

As evident from Figures 15-16, the spread generated by the PI ensemble is, on
average, increasing during the entire 120 hour forecast period. Thus, it makes sense to observe from these figures that the spread generated from the FULL ensemble is also increasing, on average, for the entire period. Also, there is a time at which the RMSD from the PI ensemble becomes greater than the FULL ensemble. This occurs close to forecast hour 60 for the RMSD calculated at 3, 6, and 12 hour intervals, and forecast hour 72 for 24 hour intervals. This means that there is little, if any, benefit to having mixed-physics members in the ensemble past these forecast hours without lateral boundary perturbations.

The time period between the time at which the RMSD curve for the MP ensemble flattens out and when the PI ensemble RMSD becomes greater than the FULL ensemble RMSD can be interpreted as the time at which the MP ensemble has stopped generating spread among its own members, but is still contributing spread to the FULL ensemble.

2) SPREAD RATIO

The spread ratio can build on the information provided by the RMSD because values of spread ratio are computed at different rainfall thresholds and RMSD values are not. Figures 17-20 display spread ratios from the PI, MP, and FULL ensembles at 0.01, 0.10, and 0.25 inch rainfall thresholds. Generally, these time series reveal that the spread ratio increases more at higher rainfall thresholds. This is simply because heavier areas of rainfall are smaller than lighter areas of rainfall so that small displacements in rainfall areas result in larger spread ratio increases at higher rainfall thresholds. Also, the effects of the lateral boundary condition influences in the MP ensemble forecasts are very apparent. For all time intervals and rainfall thresholds, spread ratios level out after about forecast hour 24, while the spread ratios in the PI ensemble become significantly larger. Surprisingly, although the RMSD time series showed that the MP ensemble tended to have greater spread during the first 24 hours of the forecast, the spread ratios did not. At virtually all forecast hours and rainfall thresholds the PI ensemble had higher spread ratios.

3) RANK HISTOGRAMS

Figures 21-24 display rank histograms for the PI, MP, and FULL ensembles.
calculated for 3, 6, 12, and 24 hour accumulation periods. The frequencies displayed in the plots are averages over all of the forecast periods. Rank histograms from individual forecast times are not shown because it was found that they were very similar over all of the accumulation periods. Thus, an average over all periods gives an appropriate representation of the behavior of each ensemble.

In general, the rank histograms for the PI ensemble are the flattest, indicating the ensemble spread is close to accurately representing the forecast uncertainty. As the accumulation periods in the PI ensemble increase from 3 to 24 hours the shape of the rank histogram changes from a slight upside-down U-shape to a slight upright U-shape with a higher tail on the right. This reversal of shape is mostly caused by an increase in the number of events that occur outside of the range of all the ensemble members. This indicates that the ensemble has a little bit too much spread at 3 hour accumulation periods and not quite enough at 24 hour accumulation periods with a relatively large amount of observations verifying at amounts higher than all the ensemble member forecasts.

The MP ensemble is very noticeably right skewed indicating that most of the ensemble members have a tendency to over-predict precipitation. Similar to the PI ensemble the tails of the rank histogram grow with the increasing accumulation periods.

The combination of the PI and MP ensembles into the FULL ensemble leads to a rank histogram that is right skewed. As the accumulation periods increase, the most obvious trend is the growth of the tails on both the left and right. By the time the accumulation period reaches 24 hours, the FULL ensemble resembles a right skewed U-shape indicating there is not enough spread to represent the forecast uncertainty and most of the ensemble members have a tendency to over-predict rainfall.

4. Summary and future work

To construct deterministic forecasts from the ensembles examined, the probability matching technique was used. Using the forecasts derived with this technique, it was found that the method of summing contingency table elements over all of the forecasts, as opposed to averaging the daily ETSs, led to calculations of higher average ETSs. As the accumulation period was increased the difference between the ETSs using the two methods
became larger. These results implied that heavier/widespread rainfall events were associated with more skill.

A comparison of the MP and PI ensembles revealed that ETSs derived using the probability matching technique were very similar at all accumulation periods. The largest difference in the scores occurred at the 0.01 inch rainfall threshold. As the accumulation period at the threshold increased, the difference between the PI and MP ensembles ETSs increased with the MP ensemble having the higher scores.

A comparison of ROC scores for the MP and PI ensembles allowed for evaluation of probabilistic forecasts. For all of the rainfall thresholds and accumulation periods examined, the MP ensemble started with higher ROC scores and the PI ensemble ROC scores surpassed the MP ensemble scores at some point during the forecast, always ending up with higher scores during the last part of the forecast. At the 0.01 inch rainfall threshold this happened later in the forecast than for all of the other rainfall thresholds.

RMSD values revealed that the MP ensemble stopped generating increasing spread around forecast hour 27 while the PI ensemble was generating increasing spread throughout the entire forecast period. However, during the first 24 hours the MP ensemble generated more variance than the PI ensemble. The MP members stopped contributing spread to the FULL ensemble around forecast hour 60. These results reiterate the importance of perturbing the lateral boundary conditions in limited area model ensembles. The trends in the RMSD calculations were analogous to the trends in the ROC scores, indicating that the ability of the PI ensemble to continue to generate increasing variance has a positive effect on the ROC scores.

Spread ratios showed similar results to the RMSD calculations; the MP ensemble stopped generating variance around 24 hours into the forecast. However, unlike the RMSD calculations, spread ratios showed that the PI ensemble generated more spread than the MP ensemble at virtually all time intervals and rainfall thresholds, even during the first 24 hours of the forecast when the MP ensemble might be expected to generate more spread.

Rank histograms revealed that the PI ensemble spread gave a better representation of the forecast uncertainty than the MP ensemble. The FULL ensemble had a tendency to over-predict rainfall and the spread was not quite large enough to represent the uncertainty in the forecasts.
Overall, results indicated that during this time period, 12 February to 15 April 2006, a mixed-physics ensemble is superior to a perturbed initial and lateral boundary conditions ensemble for only a short time (approximately 24-48 hours, at best) at the beginning of a 120 hour forecast because the influence of the lateral boundary conditions begin to degrade the forecast without perturbed boundary conditions. Future work should examine whether or not these results apply to warm season forecasts (May-September), when other studies have suggested that mixed-physics and mixed model formulations are an effective technique to increase spread. Also, future work should investigate the effects of changing the physics and perturbing the boundary conditions over a similar forecast period.

Acknowledgements. This research was funded by Baker Endowment Fund 497-01-78-3803 and NSF grant ATM-0537043. The model simulations were conducted on the 64 processor IBM computing cluster in the meteorology department of Iowa State University.

REFERENCES


Gallus, Jr., W., and J. F. Bresch, 2006: Comparison of Impacts of WRF Dynamic Core, Physics Package, and Initial Conditions on Warm Season Rainfall Forecasts. (in press)


Lorenz, E., 1965b: The Predictability of a Flow which Possesses Many Scales of Motion.


Table 1: Model physics options, dynamics options, and initial and lateral boundary conditions for all 16 ensemble members. The first 8 members are the mixed-physics members and the last eight are the perturbed initial conditions members. In the “Ensemble Member” column, names are given to denote each unique ensemble member. The “WRF core” column specifies the dynamic cores used by the members. In the “Cumulus Scheme” column, BMJ refers to the Betts-Miller-Janjic cumulus parameterization scheme and KF refers to the Kain-Fritsch scheme. The microphysics schemes Lin, Eta, and WSM6 refer to the Lin et al. (1983) scheme, the Eta Ferrier (Ferrier et al. 2002), and WRF single-moment 6-class schemes, respectively. The last column, “ICs and LBCs”, specify the initial and lateral boundary conditions used for each member. GFS denotes the Global Forecast System model run operationally at NCEP and n# GFS and p# GFS denote GFS ensemble members with negative and positive perturbations, respectively.
Figure 1: Domain of the WRF ensemble members.

Idealized ROC curves

(a) Hit Rate vs False Alarm Rate

(b) Hit Rate vs False Alarm Rate

Figure 2: The idealized ROC curves illustrate why it is unfair to compare ROC Scores calculated using the trapezoidal method from ensembles with different numbers of members. The shaded areas represent the difference between the area beneath the "true" ROC curve and the area calculated using the trapezoidal method. When more points are added to the curve (i.e., more ensemble members) this area decreases and the exact same curve will have a higher ROC Score.
Figure 3: ETS at forecast hour 24 calculated by averaging ETSs from individual forecasts (Ave) and by summing contingency table elements (Cont) for rainfall accumulated at 3 hour (a) and 6 hour (b) intervals. The figures in the upper right of the plots denote the averages over all thresholds.
Figure 4: Same as figure 2 except for rainfall accumulated at 12 hour (a) and 24 hour (b) intervals.
Figure 5: Time series of ETSs computed for precipitation accumulated at 3 hour intervals for the 0.01 (a) 0.25 (b) and 1.00 (c) inch rainfall thresholds. Average ETSs are displayed in the upper right part of the plot along with the bias scores.
Figure 6: Same as figure 4 except for precipitation accumulated at 6 hour intervals.
Figure 7: Same as figure 4 except for precipitation accumulated at 12 hour intervals.
Figure 8: Same as figure 4 except for precipitation accumulated at 24 hour intervals.
Figure 9: ROC Scores at forecast hour 24 calculated by averaging ROC Scores from individual forecasts (Ave) and by summing contingency table elements (Cont) for rainfall accumulated at 3 hour (a) and 6 hour (b) intervals. The figures in the upper right of the plots denote the averages over all thresholds.
Figure 10: Same as figure 8 except for rainfall accumulated at 12 hour (a) and 24 hour (b) intervals.
Figure 11: Time series of average ROC Scores at 3 hour intervals for the rainfall thresholds (a) 0.01, (b) 0.10, (c) 0.25, (d) 0.50, (e) 0.75, and (f) 1.00. The scores are calculated using the summed contingency table elements.
Figure 11 continued.
Figure 12: Time series of average ROC Scores at 6 hour intervals for the rainfall thresholds (a) 0.01, (b) 0.10, (c) 0.25, (d) 0.50, (e) 0.75, and (f) 1.00. The scores are calculated using the summed contingency table elements.
Figure 12 continued.
Figure 13: Time series of average ROC Scores at 12 hour intervals for the rainfall thresholds (a) 0.01, (b) 0.10, (c) 0.25, (d) 0.50, (e) 0.75, and (f) 1.00. The scores are calculated using the summed contingency table elements.
Figure 13 continued.
Figure 14: Time series of average ROC Scores at 24 hour intervals for the rainfall thresholds (a) 0.01, (b) 0.10, (c) 0.25, (d) 0.50, (e) 0.75, and (f) 1.00. The scores are calculated using the summed contingency table elements.
Figure 14 continued.
Figure 15: Root-mean-square difference (RMSD) at (a) 3 hour intervals and (b) 6 hour intervals
Figure 16: Same as 18 except at (a) 3 hour intervals and (b) 6 hour intervals
Figure 17: 3 hourly spread ratios from the MP and PI ensembles at the (a) 0.01, (b) 0.10, and (c) 0.25 inch rainfall thresholds.
Figure 18: Same as Figure 17 except for 6 hourly spread ratios.
Figure 19: Same as Figure 17 except for 12 hourly spread ratios
Figure 20: Same as Figure 17 except for 24 hourly spread ratios
Figure 21: Average Rank Histograms calculated for 3 hour accumulation periods averaged over the 120 hour forecast period for the (a) PI (b) MP, and (c) FULL ensembles.

Figure 22: Same as Figure 24 except for 6 hour accumulation periods.
Figure 23: Same as Figure 24 except for 12 hour accumulation periods.

Figure 24: Same as Figure 24 except for 24 hour accumulation periods.
GENERAL CONCLUSION

Summary of Results

Given the chaotic nature and rapid growth of errors associated with warm season precipitation in the Midwest United States, improving forecasts of warm season rainfall has proved especially challenging. The two papers in this thesis used two different approaches to build on knowledge of warm season precipitation forecasts in numerical weather prediction (NWP) models outlined by the research community as potential paths towards improved prediction.

Results in the first paper revealed that the improved representation of convection afforded by a high resolution, convection-resolving model significantly improved the representation of the diurnal cycle, especially during forecast hours 24-48 and during synoptic regimes favoring diurnally forced convection, as compared to a coarser non-convection-resolving model. This most likely resulted from the lack of a cumulus parameterization scheme in the convection-resolving model. The parameterization of cumulus convection has been blamed for the history of little progress in improving warm season precipitation forecasts (Fritsch and Carbone 2004). Analysis of equitable threat scores (ETS; Schaefer 1990), however, revealed that the coarser, non-convection-resolving model had better scores. This most likely resulted from issues discussed by Baldwin et al. (2001) who showed that large errors can be caused by small displacements in small areas of intense rainfall forecast by the high resolution models that are not resolved by the coarser models. These results reiterate the importance of non-traditional verification techniques and also show that significant improvements can be gained by moving to high resolution, explicit forecasts of convection.

The second paper compared limited area model ensembles constructed using mixed-physics and perturbed initial conditions. Major findings were the validation of the importance of perturbing the lateral boundary conditions in limited area models as discussed in past literature (e.g., Warner et al. 1997; Nutter et al. 2004). After approximately 24 hours, advantages to using the mixed-physics were nullified by the propagation of lateral boundary condition influences into the model domain. This was shown by the leveling-off of spread.
ratio (Wandishin et al. 2001) and root-mean-square difference (RMSD; Lorenz 1965) after forecast hour 24 and decreases in the area under the relative operating characteristic curve (ROC score; Mason 1982).

**Recommendations for Future Research**

Future research that could directly build from the two papers in this thesis is as follows. From the first paper, future work should see if results from the present study are consistent with those that would be obtained from simulations conducted during other years. Because the year that this study was conducted was not a typical year with respect to rainfall, this would be important. Also, mechanisms that are leading to the appearance of timing errors associated with the dissipation of mesoscale convective systems (MCSs) during the first 24 hours of the 5 km WRF-NMM forecasts should be diagnosed, as well as what is causing heavier rainfall areas to “spin up” faster than lighter areas.

From the second paper, future work should examine whether or not the results apply to warm season forecasts (May-September). Also, future work should see what effects changing the physics and perturbing the lateral boundary conditions have on the forecasts over a similar forecast period.

In general, there are many avenues through which improvement in warm season rainfall forecasts could potentially be realized that are paving the way for future research. Of particular interest are some questions raised by Roebber et al. (2004): “How can high-resolution forecasts be applied in the context of probabilistic forecasting?” and “Can high resolution ensembles with a small number of members outperform a low resolution ensemble with a large number of members?”

**REFERENCES**


Fritsch, J. M., and R. E. Carbone, 2004: Improving Quantitative Precipitation Forecasts in the


APPENDIX A: HISTORICAL PERSPECTIVE ON ENSEMBLE FORECASTING

Lewis (2005) traces the roots of ensemble forecasting to the first stages of numerical weather prediction (NWP) beginning in the late 1940s when many scientists believed that, in the future, it would be possible to make accurate deterministic predictions of future atmospheric states based on the Newtonian equations of motion. In fact, in the mid-1950s Frederick Sanders who was then a newly appointed professor at MIT recalls a theoretician who was prompted to state "A hierarchy of ever more powerful and sophisticated models will be brought to bear on meteorology until our problems will crumble before the onslaught of mathematical equations".

The belief that computers would eventually be able to provide near perfect weather forecasts was fueled by the early success of researchers studying orbital mechanics that were able to accurately predict the motion of heavenly bodies. However, in the early 1960s, scientists began to realize that meteorological predictions would never share the same success as celestial mechanics because meteorological predictions were less forgiving to errors that would always be present in the initial state (Lewis 2005). This epiphany is credited to Edward Lorenz who made this discovery accidentally when he inadvertently introduced truncation error into a model he was using for part of an experiment (Lorenz 1962). He found that forecasts generated from calculations using numbers rounded to three-digit accuracy were drastically different from forecasts using six-digit accuracy. Because of this, he proposed a procedure using a finite ensemble of initial states that "resembled the observed state closely enough so that the differences might be ascribed to errors or inadequacies in observation". These different initial states would be used to run a model multiple times so that output could be interpreted in terms of probabilities or ensemble means which would provide more information than a single deterministic forecast.

Lorenz’s proposed procedure is the general strategy currently used in global operational ensemble forecasting systems like the National Center of Environmental Prediction (NCEP) ensemble prediction system (Toth and Kalnay 1993; Tracton and Kalnay 1993; Toth and Kalnay 1997).

Incredibly, this proposition was made more than 30 years before these types of
models were implemented operationally.

REFERENCES

Symp. On Numerical Weather Prediction, Tokyo, Japan, Meteorological Society of
Japan, 629-634.
Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of
Tracton, M. S., and E. Kalnay, 1993: Ensemble forecasting at NMC: Operational
APPENDIX B: WEBSITE PRESENTATION OF ENSEMBLE FORECASTS OF PRECIPITATION

A website was constructed to display the forecasts from the mixed-physics and perturbed initial conditions ensemble in near real-time. The website URL is:

http://www.meteor.iastate.edu/~clar0614/wrf/ENSEMBLE/Ensemble_home.html

In order for the general public and meteorological community to benefit from the real-time ensemble forecasts discussed in the second paper, they must be displayed in the most meaningful and convenient form as possible. Also, it is important to realize that different groups of users will prefer to view the forecasts in different ways. For example, an operational forecaster at the NWS may want to compare the forecasts from the mixed physics and perturbed initial conditions ensemble members to see which members are contributing more to the uncertainty of the forecasts. If it is known that one group of ensemble members doesn't have adequate spread in its forecasts than this could have important implications on the products that the forecaster will put out. On the other hand, a farmer may only care about what the chances of rain are and what the average amounts of rain are that the ensemble members are forecasting. These issues were all taken into consideration in the design of the website.

The website consists of a home page that contains a brief introduction with wording geared toward the general public. Links are provided to “meteorological forecasts” and “public forecasts”. The website linked to the “meteorological forecasts” is designed for those with some kind of background in meteorology. It is set up so that probability forecasts and quantitative precipitation forecasts (QPFs) can be viewed for all rainfall thresholds and forecasts times very quickly and easily through 8 panel plots and “rollover images”. It also allows for the comparison of QPF and probability forecasts from the MP and PI ensemble members as well as the FULL ensemble. The website linked to the “public forecasts” is oriented toward the general public. QPFs from the individual ensemble members are not provided on this site, because this information is most likely of little use to the layman; instead, the QPFs from all 16 ensemble members are combined using the probability
matching technique to form one QPF forecast that will most likely be more skillful than any of the individual ensemble member forecasts. Also, to provide the general user with information about forecast uncertainty, probability of rainfall forecasts are provided for 3 rainfall thresholds; "Light (P>0.01 in.)", "Moderate (P<0.25 in.)", and "Heavy (P>1.00 in.)". All of the plots provided are 1 panel to avoid any confusion. Finally, point forecasts are provided for Iowa counties. By clicking on a county on a map of the state of Iowa, the user can view time series of rainfall probabilities and QPF (from probability matching) at 6, 12, and 24 hour intervals from the model grid-point closest to the center of the selected Iowa county.
ACKNOWLEDGEMENTS

Many people supported me in the completion of this work. I would especially like to thank my parents, Jim and Bev Clark; and my fiancé, Beth; without their support this thesis would never have been possible. Also deserving thanks are my co-major advisors Dr. William A. Gallus, Jr. and Dr. Tsing-Chang Chen; without their knowledge and expertise this work would also not have been possible. Thanks also go out to James Correia for technical writing assistance and many thought provoking discussions on various meteorology topics; Isidora Jankov for help with running model codes; and Shih-Yu (Simon) Wang, Jeng-dar (Paul) Tsay, and Wan-Ru (Judy) Huang for help generating graphics.