Butterflies and bow echoes: addressing poor forecasts with ensemble simulations

John Lawson
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd
Part of the Meteorology Commons

Recommended Citation
Lawson, John, "Butterflies and bow echoes: addressing poor forecasts with ensemble simulations" (2016). Graduate Theses and Dissertations. 15016.
https://lib.dr.iastate.edu/etd/15016

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Butterflies and bow echoes:
Addressing poor forecasts with ensemble simulations

by

John Lawson

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Meteorology

Program of Study Committee:
William A. Gallus, Jr., Major Professor
Raymond W. Arritt
William J. Gutowski
Andrew VanLoocke
Xiaoqing Wu

Iowa State University
Ames, Iowa
2016

Copyright © John Lawson, 2016. All rights reserved.
DEDICATION

I dedicate all musings herein to my family and friends, for enduring large readings over the years from the book of Who Cares; to fellow researchers and Python code developers, on whose shoulders this work stands; and to the atmosphere, for being rather interesting.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................. vi

ABSTRACT ......................................................... vii

GENERAL INTRODUCTION ........................................... 1
  1 Mesoscale convective systems ................................. 2
  2 Bow echoes and derechos .................................... 3
  3 Predictability horizons ...................................... 5
  4 Numerical simulation ....................................... 7
  5 Key questions .............................................. 10

ON CONTRASTING ENSEMBLE SIMULATIONS OF TWO GREAT PLAINS BOW ECHOES ................................................. 14
Abstract .......................................................... 14
  1 Introduction ............................................... 15
  2 Data and Methods .......................................... 20
  3 Synoptic overviews ........................................ 25
    3.1 NEKS06 ................................................. 25
    3.2 KSOK13 ............................................... 27
  4 ICBC experiments ........................................... 27
    4.1 NEKS06 ................................................. 28
    4.2 KSOK13 ............................................... 28
  5 MXMP experiments ......................................... 30
3 Results .................................................................................. 91
  3.1 Sensitivity of structure to \( \Delta x \) .............................................. 91
  3.2 Sensitivity of spread and skill to \( \Delta x \) ................................ 92
  3.3 Sensitivity of system speed to \( \Delta x \) ...................................... 93
  3.4 Sensitivity of system speed to SKEB ....................................... 96
4 Summary and Conclusions .................................................. 96
Acknowledgments ................................................................. 98

THE RELATIONSHIP BETWEEN SYNOPTIC AND STORM-SCALE PRE-
DICTABILITY OF MESOSCALE CONVECTION SYSTEMS ............ 108
Abstract ................................................................................. 108
1 Introduction ........................................................................... 109
2 Data and Methods .............................................................. 110
  2.1 Datasets .............................................................................. 110
  2.2 Computed fields ................................................................. 112
  2.3 Heatmap methodology ....................................................... 113
3 Results ................................................................................. 114
  3.1 Bow-echo cases ................................................................. 114
  3.2 All cases ............................................................................ 115
4 Discussion ............................................................................. 115
Acknowledgments ................................................................. 117

GENERAL CONCLUSIONS .................................................. 124
ACKNOWLEDGMENTS

I acknowledge my advisor, Prof. William Gallus, Jr., for good humor and constructive guidance over the last three years; my committee, for their patience and effort; Daryl Herzmann and David Flory, for IT support and clearing up my digital mess; and my colleagues for advice and support over the three years.

And like a fractal, this list would continue forever the further you read: thank you everyone.
ABSTRACT

Bow-echo structures, a subset of mesoscale convective systems (MCSs), are associated with damaging wind and hail, and are often poorly forecast within deterministic numerical weather prediction model simulations. Among other things, this may be due to inherent low predictability associated with bow echoes, error in the initial conditions (ICs), and inadequate parameterization schemes (model error). Ensemble simulations account for, and measure, these uncertainties in a forecast.

A study of two bow-echo cases, simulated with multiple ensemble configurations, find that location and timing variation of the simulated systems reduce when certain parameters are fixed (i.e., ICs; microphysical parameterizations). However, variations in convective mode remain substantial. Results suggest the MCS positioning is influenced primarily by ICs, but its mode is most sensitive to the model-error uncertainty.

A modification to the Structure Amplitude Location (SAL) method identifies and compares objects in both forecast and observed composite reflectivity fields. Both the original and modified SAL methods are used to evaluate daily 12-km North American Model (NAM) forecasts during the summer of 2015 for a central United States domain. SAL using reflectivity reveals a diurnal cycle of skill, with minimum skill occurring early-to-late afternoon (local time), and maximum skill occurring just before sunrise.

The modified SAL method is then deployed to evaluate the effect of finer resolution on both ensemble spread and the character of bow-echo development. Due to the increased prominence of noise close to the truncated scale, we expect a larger ensemble spread as horizontal grid spacing decreases. Two ensemble forecasts were generated
using the Weather Research and Forecasting (WRF) model: one used a single domain with 3-km horizontal grid spacing, and another nested a 1-km domain inside the 3-km parent domain with two-way feedback. Ensemble members were then generated from the control with a stochastic kinetic-energy backscatter scheme, with identical initial and lateral-boundary conditions. Results show that the increase in grid resolution reduces both spread and skill, and that the nested ensemble produces a faster bow echo and stronger cold pools. The latter two are most likely due to increased (fractal) cloud surface area within the nested ensemble, which allow more entrainment of dry air and hence increased evaporative cooling.

Finally, we address the poor performance in previous bow-echo studies by evaluating a WRF hindcast dataset designed to capture numerous MCSs in the Great Plains. We may expect the skill of the hindcasts to be dictated by (a) inherent synoptic-scale predictability (i.e., ensemble spread), and (b) the skill of the NAM forecast dataset providing initial and lateral-boundary conditions to the WRF hindcast. However, there is no obvious relationship between the accuracy of MCS convective mode and either factor. When the MCS dataset is confined to cases containing bow echoes, we find that serial bow echoes (i.e., line-echo wave patterns) are better forecast by the WRF hindcasts than progressive bow echoes. Furthermore, stronger rising motion is linked with the propensity for bow echoes to be serial rather than progressive. We therefore speculate that the skill of storm-scale forecasts may inherit only limited characteristics of the large-scale predictability, perhaps due to rapid downscale cascade and growth of initially trivial errors in the initial-condition dataset.

In summary, as model errors appear random (not systematic), and the reduction of IC error yields only diminishing returns, we deem it likely that poor bow-echo forecasts stem from inherent low predictability. This demands the use of well-calibrated ensemble systems, accounting for both model and IC error, to properly gauge the probabilities of bow-echo events and their associated hazards.
GENERAL INTRODUCTION

Does the flap of a butterfly’s wings in Brazil set off a tornado in Texas? This was the title of Edward Lorenz’s iconic presentation in 1972 (Palmer et al. 2014, and refs. therein). The phrase “The Butterfly Effect” has passed into popular culture: the subject of Hollywood films, the name of songs, the excuse of sports players and gamblers. The effect is known more precisely as sensitivity to initial conditions within nonlinear, deterministic, dynamical systems (Williams 1997), or chaos. Lorenz first discovered chaos within his simple computer simulation of convection (Lorenz 1963; Gleick 1987; Palmer et al. 2014), as others did within animal population simulations (May 1976; Feigenbaum 1978), and chaos’ sister phenomenon, fractals (Mandelbrot 1967; Lovejoy and Mandelbrot 1985).

The United States of America experiences some of the most severe weather on the planet. Extreme heat, damaging thunderstorms, extensive flooding, and heavy snowfalls contribute towards hundred of deaths and injuries every year (Greenough et al. 2001, and refs. therein). The National Weather Service is guided by “protection of life and property”, and this mantra drives research into improvements in numerical weather prediction (NWP). Severe summer weather causes significant damage, particularly in the Great Plains, with hazards such as hail, tornadoes, and flooding. The frequency of each hazard varies depending on the morphology, or mode, of the moist convection (Gallus et al. 2008). For instance, thunderstorm lines that contain bowing segments in radar reflectivity data, known as bow echoes, are often associated with damaging winds. Hence, it follows that forecasting the correct mode can greatly assist in communicating the risk to the end user.
However, chaos theory sets an upper bound on the success of NWP, which motivates the investigation of this limit within simulations of thunderstorm complexes in the Great Plains, particularly bow echoes. In presenting evidence that bow echoes are associated with inherently low predictability, the following dissertation comprises a general introduction, four papers, and a final conclusion chapter. The present chapter continues with a review of background literature relating to topics within the dissertation, and an overview of key questions. The first paper describes multiple ensemble configurations that simulate two contrasting bow echoes: one well forecast, another poorly forecast. The second paper describes a modification to the object-based Structure Amplitude Location (Wernli et al. 2008) method, where composite reflectivity is verified (instead of accumulated precipitation) to gauge the skill of the North American Model summertime forecasts in the central United States. This method is then used in the third paper to consider whether decreased horizontal grid spacing ($\Delta x$) increases the spread and/or skill of a bow-echo ensemble forecast. Analysis of the relationship between mesoscale and synoptic-scale predictability of convective systems is shown in the final paper. General conclusions are then presented.

1 Mesoscale convective systems

The mesoscale, often defined in meteorology as the scale between 2 and 2000 km, is the regime in which thunderstorm cells and complexes form (Markowski and Richardson 2010). The ratio of convective available potential energy (CAPE) to vertical wind shear (simply ‘shear’ in this chapter) dictates the evolution of deep moist convection. Single cells comprise one updraft core, and often form in high-CAPE, low-shear environments, unable to organise into a larger system. If shear is moderate (10–20 m s$^{-1}$ between 0 and 6 km), new cells may be triggered repeatedly, and a cluster or line forms. This complex is called a multicell; a menagerie of multicell formations are
named as subsets of mesoscale convective systems (MCSs). MCSs are thunderstorm complexes more than 100 km long in one dimension, and are the broad focus of this dissertation. Finally, if shear is strong (>20 m s$^{-1}$), vertical pressure gradients throughout a large depth of the troposphere induce a stable supercell circulation, in which the cell rotates. The supercell, first documented by Browning and Ludlam (1962) in the United Kingdom, is the thunderstorm system most likely to spawn tornadoes (Moller et al. 1994). Supercells are outside the scope of the dissertation, but some can grow and/or collide with other cells later in its lifecycle to form an MCS (Moller et al. 1994; Klimowski et al. 2003).

MCSs are subclassified into multiple categories based on their appearance in radar reflectivity and satellite data (Markowski and Richardson 2010). Note that a system may match two categories simultaneously. Squall lines or quasi-linear convective systems (QLCSs) are linear in reflectivity data, and are often associated with a stratiform region that may trail or lead the system. The stratiform region, and its orientation, is often associated with serious flooding (Pettet and Johnson 2003). Should a circular, smooth anvil be observed in satellite imagery associated with an MCS, contingent on areal and temperature criteria, the system is also called a mesoscale convective complex (MCC; Maddox 1980). Many areas in the central United States receive up to a fifth of their summertime precipitation from MCCs (Ashley et al. 2003). Finally, linear features that contain bowing segments in reflectivity data are named bow echoes, and are the subject of this dissertation.

## 2 Bow echoes and derechos

Foremost, bow echoes are associated with wind damage in the Great Plains of the United States (Fujita and Caracena 1977; Gallus et al. 2008). The strongest straight-line wind events are named derechos, and are subdivided into two groups (Johns and
This subdivision is based on wind criteria, but in this dissertation, we make a similar distinction using radar reflectivity observations. Hence, when bowing segments occur along a QLCS, they are termed *serial bow echoes* (Fig. 1). These often occur ahead of, and parallel to, cold fronts (serial derechos in Johns and Hirt 1987). Systems with a bowing radius similar to the size of the system are termed *progressive bow echoes* (Fig. 2). These form ahead of, and move perpendicular to, warm fronts, and the prior convection is often elevated (progressive derechos in Johns and Hirt 1987).

The bowing segments in both types form when the rear-inflow jet is brought towards the surface by evaporative cooling (Markowski and Richardson 2010). The negative buoyancy that ensues also creates and accelerates a surface cold pool, which is maintained through dry-air advection from the surround environment relatively low in moisture. Downshear of the cold pool, new cells initiate and develop the bowing signature conspicuous in reflectivity data (Weisman 1993). The relationship between bowing segments and downburst winds was first discussed by Fujita and Caracena (1977), shortly before Fujita coined the term “bow echo”.

When simulating multiple MCS cases and rating their ability to forecast the observed convective mode, Snively and Gallus (2014) found that bow echoes are the mode that is forecast most poorly in terms of their timing and convective mode, and that this was associated with weak vertical wind shear, and excessively high potential temperatures aloft. Poor bow-echo simulations were also seen by Keene and Schumacher (2013). Little relationship between structure and surface wind—contrary to observed systems—was noted by Wandishin et al. (2010). In these studies, no distinction is made between progressive and serial bow echoes. However, in the following discussion on predictability, we may expect progressive bow echoes to be the subtype associated with less skill, due to their smaller length scale, elevated nature, and weaker affiliation with a baroclinic boundary.
3 Predictability horizons

The atmosphere is a fractal, with turbulent circulations embedded on scales from the Hadley cell to viscous dissipation. Chaotic flow has a time limit (or horizon) of predictability (Lorenz 1963; Palmer et al. 2014), after which numerical forecasts of the flow are as unskillful as choosing randomly from the climatological state. This horizon arrives sooner as the scale of motion becomes smaller (Lorenz 1969). The horizon’s dependence on scale can be explained by considering two numerical simulations of the atmosphere, \( A \) and \( B \), each identical apart from a kinetic-energy perturbation \( \Delta KE \) on the smallest scale in \( B \). This \( \Delta KE \) is analogous to the butterfly’s wing flap in Lorenz’s talk. As the simulation is integrated (i.e., time progresses), chaotic motion increases the relative magnitude and length scale of the initial \( \Delta KE \). For instance, this change in kinetic energy may be the difference between an air parcel thermally circulating in the boundary layer in \( A \) versus meeting its level of free convection in \( B \). Tipping points such as thunderstorms grow perturbations very quickly (Zhang et al. 2003). Eventually, chaotic perturbation growth is so large that, on a given scale, \( A \) and \( B \) are as different as two randomly selected atmospheric states. This is error saturation, and this point is reached sooner as a given length scale becomes smaller, giving rise to the predictability horizon.

Lorenz talked of two types of predictability: practical, which is the limit of current NWP systems and datasets, and intrinsic, which is the upper bound possible with a perfect model and dataset (Lorenz 1969). The difference between practical and intrinsic predictability may be somewhat bridged by improvements in ICs and LBCs (as suggested for the case in Melhauser and Zhang 2012). In this dissertation, predictability is a property of the flow, and as such, cannot be verified with a single forecast. Hence, we distinguish between predictability and skill. In meteorology, forecast skillfulness is defined to occur when a forecast is closer to observations than either persistence (e.g.,
repeating yesterday’s observations) or climatology (the average observation over the previous 30 years). Then, the closer the forecast is to observations, the more skillful the forecast. The caveat is that skill, defined in this way, is sensitive to the scoring scheme chosen (for example, as discussed by Mass et al. 2002).

The differences between the observed state and a simulation are known as errors, which stem from deficiencies in initial lateral-boundary conditions (ICs and LBCs, respectively), and approximations or artifacts introduced by the numerical model. Total error comprises a systematic part (the mean error over a sample of cases) and random part (the residual). These small perturbations, or butterflies, can occur at all scales. But what is the practical consequence of Lorenz’s famous quote? Durran and Gingrich (2014) state that butterflies “are not of practical importance”. In their calculations, even a small ($10^{-9}\%$) relative butterfly wingflap at a wavelength of 400 km overwhelms a large (100%) wingflap on a scale of centimeters. This occurs because errors at the large scale grow downscale quickly, becoming larger relative to the scale. The error saturates at the smallest scale and grows back upscale. The up- and downscale cascade of errors in the atmosphere occurs at two distinct rates, depending on the length scale (Nastrom and Gage 1985). Above scales of 400 km, errors cascade at a rate proportional to $k^{-3}$, where $k$ is the horizontal wavelength. Below scales of 400 km, errors cascade at a rate proportional to $k^{-5/3}$. The source of this scale separation around 400 km is much debated, but may be due to the presence of convective eddies on the mesoscale, which through more vigorous mixing destroys predictability faster (Melhauser and Zhang 2012; Durran and Weyn 2016). This means that perturbations on storm scales are increasingly important in limited-area three-dimensional simulations as $\Delta x$ decreases and errors cascade faster both up- and downscale (Rotunno and Snyder 2008), but are negligible on a large quasi-two-dimensional scale (Tennekes 1978). Tennekes articulated this intuitively by pointing out the importance of the fractal (turbulent) nature of cloud edges in the characteristics of cumuli, but
the unimportance of the same detail to a satellite view of smooth-edged extratropical frontal systems. The effect of grid resolution on error growth is discussed later.

Further to this topic, Durran and Weyn (2016) state that “thunderstorms do not get butterflies”, and relate the idealized two-dimensional findings from Durran and Gingrich (2014) to a cloud model ensemble forecast of a QLCS. One ensemble was perturbed using errors at small scales (8 km), another at large scales (128 km). The large-scale perturbations were one-quarter as large as those at the small scale. Durran and Weyn (2016) found substantial variation after 4 h in both ensembles, and a similar kinetic-energy spectrum in both that suggested that the scale of perturbation was irrelevant in small-scale predictability loss. The insignificance of errors that originate on the smallest scales also suggest that storm-scale IC and LBC improvements yield diminishing returns.

4 Numerical simulation

Perhaps some day in the dim future it will be possible to advance the computations faster than the weather advances and at a cost less than the saving to mankind due to the information gained. But that is a dream.

Richardson (1922), quoted above, first proposed that numerical integration of the equations of motion may one day forecast the weather faster than it occurred. However, after numerical instability in his example led to spurious pressure changes, his results were ignored, and the idea was lost for decades (Lynch 2006). Charney (1948) later proposed that changes to the atmospheric state could be predicted through advection of geostrophic vorticity. Charney, von Neumann, and colleagues led progress in NWP as computer power increased (Lewis 2005), and eventually, Richardson’s dream of useful prognoses was reality.
The better the resolution of a forecast (i.e., the smaller the $\Delta x$, or the finer the grid on which the equations of motion are integrated), the more computer power that is required. Furthermore, a twofold increase in three-dimension resolution demands an eightfold increase in power. The horizontal and vertical distances between grid points dictate the smallest processes that a simulation can resolve (usually 4–5 times the horizontal grid spacing $\Delta x$). Over time, as operational centers have dedicated more and more computer resources to reducing $\Delta x$, their forecasts have improved—albeit also for reasons other than resolution (Berner et al. 2012). In fact, there is evidence that the increased computer power yields diminishing returns in terms of skill, e.g., below 12 km (Mass et al. 2002). As $\Delta x$ decreases, more of the faster $k^{-5/3}$ energy-cascade regime is captured, and errors grow faster. This balance may preserve the predictability horizon, although a smaller $\Delta x$ generates more realistic systems by eye (e.g., Mass et al. 2002).

There will always be a scale at which processes are not explicitly resolved, and require an approximation or parameterization. For instance, a $\Delta x$ of 1 km allows moist convection to form more or less explicitly, precluding the need for a cumulus parameterization scheme, but the cloud microphysical processes still require approximation. Parameterizations are simplifications, and introduce error as a result. Some error is systematic—say, a dry bias in the planetary boundary layer—and some is random. This uncertainty in the model (or model error) amplifies within chaotic flow: error growth is rapid for both model and IC errors alike, and is practically indistinguishable (Leutbecher and Palmer 2008).

The existence of chaos implies that a single (deterministic) prognosis of the atmospheric state is insufficient: a forecast must be accompanied by an estimate of uncertainty, i.e., a gauge of the state sensitivity to perturbations. This estimate is produced by running multiple simulations, known as an ensemble forecast, where each member of the ensemble differs slightly within the range of model and IC/LBC uncertainty
(Leutbecher and Palmer 2008). The dispersion of the ensemble, or its spread, can be measured by quantities such as ensemble standard deviation, average differences between all permutations of the members, and so on. Spread and skill are usually inversely proportional (Whitaker and Loughe 1998). For correct estimation of uncertainty, the ensemble must be reliable; that is, all members are equally as likely to be correct, and a $x\%$ chance of an event verifies $x\%$ of the time (Hamill 2001; Gigerenzer et al. 2005). Ensemble output may be calibrated through post-processing (e.g., Gneiting et al. 2005; Hagedorn et al. 2012; Berner et al. 2015a). However, it is most desirable for ensemble output to be as close to reliable as possible, before post-processing (Martin Leutbecher, personal communication), to maximize robustness to model changes in resolution, parameterization setup, etc.

Ensembles can be created by generating perturbations in numerous ways: variation in ICs and LBCs (e.g., Romine et al. 2014); different permutations of parameterization schemes (e.g., Stensrud et al. 2000); forcing introduced by a stochastic kinetic-energy backscatter (SKEB; Shutts 2005) scheme (e.g., Tennant et al. 2011). Both model and IC/LBC error should be accounted for in operational global models, if a probability distribution of atmospheric states is desired. Mixed-parameterization ensembles yield members of different likelihood (related to the bias, quality, and interaction of the parameterization schemes), and as such, output from these ensembles must be calibrated to generate reliable probabilities. Conversely, the SKEB scheme yields members that are dispersed equally across the probability spectrum by randomly injecting kinetic energy into the flow (Shutts 2005). This scheme was initially designed to correct excessive dissipation of kinetic energy between resolved and unresolved scales in large eddy simulations (Mason and Thomson 1992), and was further developed for use in the Weather Research and Forecasting (WRF) model (Berner et al. 2011). Recently, Shutts (2015) adapted the backscatter scheme to target convective processes, the main source of model error. In the case of multiple WRF domains, the instantaneous SKEB forcing
patterns are determined by the largest $\Delta x$, and the perturbations are interpolated to each nest. On a grid of $\Delta x = 15$ km, Romine et al. (2014, their Fig. 1) showed the instantaneous stochastic forcing pattern for the default SKEB settings has dipoles of $O(1000$ km), much larger than the $\Delta x$.

The benefit of stochastic schemes such as SKEB is their potential to span a larger (and well-dispersed) region of phase (outcome) space (Christensen et al. 2014), and the dependence of perturbation growth to the flow regime (Berner et al. 2011). Some stochastic schemes are independent from a given parameterization, such as SKEB, and can be coupled to the model separately; others are written into parameterizations, and may target problem areas such as fog formation (e.g., Random Parameters; Bowler et al. 2008). Berner et al. (2011) found that SKEB alone was outperformed by a mixture of SKEB and parameterization variation: the spread of SKEB members was not sufficient in that case to span the range of outcomes. Hence stochastic parameterizations are an active field of development (Berner et al. 2015b).

5 Key questions

To review, the following papers address these key questions:

- Thunderstorms do not get butterflies: after a short time (e.g., 4 h), the up- and downscale cascade of errors has distributed and grown initial errors regardless of their original wavelength. Hence, as long as systems are allowed time to spin up, can we use SKEB schemes to create ensemble perturbations in our convection-allowing simulations?

- How does the spread of convective mode change with the general ensemble spread, and the method of perturbation generation (IC/LBC perturbations; mixed parameterizations; SKEB)?
• Bow echoes are the worst forecast convective mode. Is this due to low inherent predictability, poor IC/LBCs, poor parameterizations?

• How might grid resolution affect the spread and performance of an ensemble?

• Considering the fast up- and down-scale growth of errors, how linked are the storm and synoptic scales in MCS forecasts?
Figures

Figure 1  Example of a serial bow echo in composite reflectivity observations.
Figure 2  As Fig. 1, but for a progressive bow echo.
ON CONTRASTING ENSEMBLE SIMULATIONS OF TWO GREAT
PLAINS BOW ECHOES

A paper accepted for publication in Weather and Forecasting

John Lawson and William A. Gallus, Jr.

Abstract

Bow-echo structures, a subset of mesoscale convective systems (MCSs), are often poorly forecast within deterministic numerical weather prediction model simulations. Among other things, this may be due to inherent low predictability associated with bow echoes, deficient initial conditions (ICs), and inadequate parameterization schemes. Four different ensemble configurations assessed sensitivity of the MCSs’ simulated reflectivity and radius of curvature to the following: perturbations in initial and lateral-boundary conditions using a global dataset; different microphysical schemes; a stochastic kinetic-energy backscatter (SKEB) scheme; and a mix of the previous two.

One case is poorly simulated no matter which IC dataset or microphysical parameterization is used. In the other case, almost all simulations reproduce a bow echo. When the IC dataset and microphysical parameterization is fixed within a SKEB ensemble, ensemble uncertainty is smaller. However, while differences in the location and timing of the MCS reduce, variations in convective mode remain substantial. Results suggest the MCS’s positioning is influenced primarily by ICs, but its mode is most
sensitive to the model-error uncertainty. Hence, correct estimation of model-error uncertainty on the storm scale is crucial for adequate spread and the probabilistic forecast of convective events.

1 Introduction

Mesoscale convective systems (MCSs) are groups of thunderstorms of length $O(100\,\text{km})$ in at least one direction (American Meteorological Society 2014). These predominantly summertime systems provide the Great Plains of the United States with much of their warm-season rainfall (Fritsch et al. 1986). A subset of these MCSs that contain bowing features, however, bring the risks of damaging winds, 2.5–5 cm (1–2 in) hail, and flash flooding (Gallus et al. 2008). Conspicuous by their convex structure in radar reflectivity (Fig. 1), bow echoes and line-echo wave patterns (LEWPs) are associated with some of the strongest non-tornadic wind events in the Plains, sometimes meeting derecho (damaging straight-line wind) criteria (Johns and Hirt 1987). A bowing structure often develops when stratiform precipitation behind a quasi-linear convective system lowers a rear-inflow jet through evaporative cooling and consequent negative buoyancy (Markowski and Richardson 2010). The cold pool accelerates due to the buoyancy gradient at its leading edge, and is maintained by the jet through advection of drier air. Development of convective cells on the downshear side of the cold pool creates the distinctive bowing shape (Weisman 1993).

Bow echoes and LEWPs, more often than other MCSs, are poorly simulated by numerical model forecasts (Keene and Schumacher 2013; Snively and Gallus 2014). Snively and Gallus (2014) found 0–6 km shear was too weak, and potential temperatures aloft too high, in their deterministic forecasts of bowing segments. The reduced skill of the model was usually related to simulation of the incorrect MCS mode. Snively and Gallus (2014) also surmised that simulations involving elevated convec-
tion may have performed the worst of those in the study. In two studies, Adams-Selin et al. found that performance of numerical simulations, both idealized (2013a) and regarding an observed system (2013b), were acutely sensitive to the chosen micro-physical parameterization. Specifically, when graupel hydrometeors were simulated as lighter and greater in number (i.e., graupel-like rather than hail-like), they resulted in a stronger cold pool and rear-inflow jet, and hence the bowing initiated earlier. The chosen parameterization scheme also strongly affected the magnitude and areal coverage of precipitation, system speed, and wind gusts. But it is unclear whether these findings can be applied generally when considering variations in synoptic regime, the initial-condition (IC) dataset, and model configuration.

While numerical weather prediction (NWP) continues its march towards explicit resolution of smaller and smaller convective features, there are a number of obstacles en route that may inhibit, or even preclude, successful numerical forecasts of bow echoes at a given lead time. Computer models are incomplete and imperfect: while smaller phenomena are resolved explicitly by ever-decreasing grid spacings, there will always be a scale below which wavelengths are truncated, and chaotic, non-linear processes are implicitly resolved, or parameterized. Parameterization is used in operational NWP models, such as the North American Mesoscale (NAM) model and the Global Forecasting System (GFS), to capture the planetary boundary layer (PBL), cloud microphysics, and other sub-grid-scale processes. The ‘spread’ of parameterization schemes, each with their own set of biases and random errors, interact during a simulation without a priori knowledge of the impact on, e.g., simulated radar reflectivity structures. In response to this, Adams-Selin et al. (2013b) called for schemes of opposing biases to be combined in operational mixed-physics ensemble systems. However, we cannot be sure that the biases shown in one study can apply generally to all regions, synoptic regimes, seasons, years, etc. For example, when changing typical hydrometeor characteristics from graupel to hail, Van Weverberg et al. (2011) found increased
surface precipitation amounts; in contrast, Gilmore et al. (2004) did not. To account for these biases \textit{a priori}, Berner et al. (2011) trained their mixed-physics models over a number of months to determine the optimal configuration for spread and skill. This may not be a practical or general approach for operational centers to endorse long-term, when one considers the training sensitivity to many factors, and the frequent updates to NWP systems and parameterizations themselves.

In addition to model uncertainty, the atmosphere as a partly chaotic system is sensitive to IC uncertainty (Lorenz 1969); from this, Lorenz suggested a theoretical predictability horizon (Palmer et al. 2014, and references therein). When assuming purely chaotic (turbulent) flow, Lorenz estimated predictability to be limited to 1–2 h on scales of 10 km (Lorenz 1969). Fortunately from a forecasting standpoint, forecast models show that the atmosphere has inherent predictability at the mesoscale longer than that proposed by Lorenz. This may be due to known forcings that constrain the solution—high terrain, synoptic-scale fronts (e.g., Anthes et al. 1985)—and stable mechanisms that locally limit error growth, such as the helical flow in supercells (Lilly 1990), and in confluent, weak flow (Oortwijn 1998). In addition, limited-area model forecasts are constrained by (and sensitive to) their lateral boundary conditions (LBCs). Palmer et al. (2014) suggest that skillful forecasts beyond a given scale’s Lorenzian horizon may be possible due to the intermittent nature of chaos in the atmosphere (i.e., its regime dependency). In addition, they argue that Lorenz’s pessimistic estimates are due to the overly simplistic nature of the Lorenz-63 system (Lorenz 1963).

Unfortunately for MCS forecasts, moist convection is very destructive to predictability (Zhang et al. 2003). MCSs that form in the Great Plains even influence global-model forecasts of blocking patterns downstream over Europe at the medium-range through diabatic destruction of potential vorticity (Rodwell et al. 2013). In addition, diagnosis of substantially damaging IC error is fraught with difficulty due to both up- and down-scale growth of errors (Durran and Gingrich 2014, and references therein). Notably,
the use of coarse-grid IC/LBC datasets to drive convection-allowing ensemble simula-
tions may result in insufficient variance in convective scales (e.g., Schwartz et al. 2014
and references therein); IC perturbations from a global model do not include variance
below its truncated scale. Errors first propagate downscale and saturate before growing
upscale (Durran and Gingrich 2014). Hence, there is a delay in small-scale variance
growth, which impacts particularly the first 6 h of a numerical simulation (Kühnlein
et al. 2014), and can yield an underdispersive ensemble (Romine et al. 2014, and
references therein).

To address these problems and better sample the spectrum of possible outcomes of
the model atmosphere, many forecast centers use a number of different numerical sim-
ulations (ensemble forecasts; Leutbecher and Palmer 2008). There are different ways
of creating members that differ from their control: through mixed-parameterization
configurations (e.g., Stensrud et al. 2000); through perturbed ICs and LBCs (e.g.,
Romine et al. 2014); through multiple NWP dynamical cores or models (e.g., Hagedorn
et al. 2012); etc. Recently, studies have yielded a method to inject energy (that may be
erroneously dissipated in the model between the resolved and unresolved scales) into
the simulation to better account for model error (Shutts 2005). This so-called stochas-
tic kinetic energy backscatter (SKEB) scheme has been shown to improve ensemble
spread and ultimately provide a more skillful ensemble mean than a mixed-physics
approach (Duda et al. 2016), except at the surface (Berner et al. 2011). Furthermore,
when a SKEB scheme was combined with a mixed-parameterization configuration by
Berner et al. (2011), performance was even better. As of version 3.7, WRF param-
eterizations are deterministic in nature; a stochastic approach is potentially a better
way to account for model error (Palmer 2001). Ensemble forecasts are not only useful
for operational centers, but also can provide a larger corpus of ‘alternative realities’
in which to seek sensitivity of atmospheric phenomena during posterior investigation
(e.g., Hanley et al. 2013).
To address the issue of why bowing structures are often more poorly forecast than other MCS modes, and while not exhaustive or mutually exclusive, we propose four hypotheses:

1. Bow echoes are inherently less predictable features, perhaps due to microscale destruction of predictability within the bowing feature itself,

2. Bow echoes are embedded in less predictable synoptic-scale regimes,

3. There is critical deficiency in ICs and LBCs within simulations and forecasts, and

4. There is critical deficiency in the sub-grid-scale processes of microphysics parameterizations.

Bow echoes are an extreme phenomenon in both rarity and severity, and their specifically local risks (strong wind, flash flooding) do not lend themselves to the smoothing of ensemble means. In this case, choosing the member closest to the ensemble mean (Ancell 2013), perusal of postage-stamp plots, or generating probability of threshold exceedance (Schwartz et al. 2015), is more useful for forecaster interpretation (e.g., Gallus and Lawson 2016). Rather than focus on ensemble means or skill-score statistics, the present study will rather investigate the visual spread of convective mode and radii of curvature in simulated reflectivity, with a secondary focus on surface-wind magnitude, coverage, and exceedance probabilities. Note that bowing structures can occur in two ways: those that appear multiple times along a quasi-linear convective system, typically in parallel with a front (often resulting in serial derechos, Johns and Hirt 1987), and those that are less strongly forced by a large-scale boundary, whose bowing radius of curvature is similar to the size of the system itself (progressive derechos). (There is no differentiation between either type in Snively and Gallus 2014.) Motivated by the wish to concentrate on the more flexible criteria of radar reflectivity signatures, rather than strict (and more arbitrary) surface-wind definitions
of a derecho, the present study refactors this terminology to look at progressive bow echoes.

We will first outline various IC/LBC datasets and model configurations in section 2. The synoptic settings of two progressive bow echoes are presented in section 3. The two cases are contrasted through use of four ensemble configurations. The configuration with perturbed ICs/LBCs (section 4) accounts for uncertainty in the constraining atmospheric-state data. The configuration with mixed microphysics parameterizations (section 5), and two involving SKEB schemes with and without mixed microphysics (section 6), account for model and parameterization uncertainty. The results are synthesized and concluded in sections 7 and 8, respectively, along with discussion of future work, and how the performance of all ensembles is interpreted regarding bowing-structure predictability horizons.

Note, in the present study, we refer to variance between ensemble members as spread or uncertainty interchangeably. This is distinct from error, which hereby is the difference between observations and a dataset, deterministic simulation, or ensemble mean (or ‘mean-like’ interpretation for non-continuous quantities like reflectivity).

2 Data and Methods

The present study focuses on two progressive bow echoes: an eastward-moving system along the Nebraska–Kansas border on 26–27 May 2006, and a southward-moving system that crossed Kansas, Oklahoma, and Texas on 15–16 August 2013. The two cases will be hereby termed NEKS06 and KSOK13, respectively. The former was chosen as one of the poorest simulations in Snively and Gallus (2014); the latter was chosen for contrast as a result of good performance in multiple preliminary simulations. The contrasting synoptic scenarios for both cases (cf. Figs. 2 and 3) also motivated their inclusion. These are described further in section 3.
All numerical simulations were run on the same supercomputer system at Iowa State University to avoid introduction of rounding-error contamination. The simulations were performed with version 3.5 of the Weather Research and Forecasting model (WRF; Skamarock et al. 2008), using the Advanced Research WRF dynamical core. The control parameterization configuration (Table 1) was chosen primarily for demonstrated stability on the Iowa State supercomputers, due to the large number of ensemble runs required with this configuration. The control microphysical parameterization (Thompson) was also selected due to good performance in similar studies (e.g., Snively and Gallus 2014; Romine et al. 2014). The constant domain size was 451 by 451 points with horizontal grid spacing $\Delta x$ set at 3 km. This grid spacing balances the benefits of a finer resolution—better reproduction of convective systems and more skillful forecasts—with the computational demand of multiple ensemble simulations that might yield diminishing returns (Lean et al. 2008). This grid spacing was also used to provide consistency with Snively and Gallus (2014). The domains are shown in Figure 4. Preliminary tests were done with a parent domain to ease the transition from global and regional datasets to the 3-km domain, but use of the parent domain did not substantially change the simulation. The timestep was six seconds (i.e. $2 \Delta x$) after preliminary tests were unstable at longer timesteps. Fifty vertical levels were specified manually as fractions of a terrain-following hydrostatic pressure coordinate. These were stacked more tightly in lower levels (separated on average by $\sim 40$ m in the lowest 20 levels) to better resolve the PBL, as in Adams-Selin et al. (2013b), with the caveat that increased vertical resolution may not always result in a better forecast of the convective system (Aligo et al. 2009).

Depending on the ensemble experiment, ICs and LBCs were provided by one (or all) member(s) of the 11-member Global Ensemble Forecast System Reforecast dataset (GEFS/R2; Hamill et al. 2013), or NAM archived analyses (12 km horizontal grid spacing; 40 vertical levels). We used the limited GEFS/R2 dataset (1° horizontal resolution;
12 vertical levels), readily available online, instead of the full dataset (0.5° horizontal
grid spacing and 42 vertical levels). As the limited GEFS/R2 dataset does not con-
tain sufficient resolution in soil layers for the WRF to run as-is, GFS analyses of soil
temperature and moisture were prescribed for each batch of ICs and LBCs (see Law-
son 2013 for further information on this method). While small changes in variables
such as soil temperature can affect convective initiation (Clark and Arritt 1995), the
absence of perturbations in soil variables was assumed to not preclude useful relative
conclusions. The limited GEFS/R2 dataset performed well in preliminary tests and
provides an interesting contrast to the WRF initialization from the higher resolution
NAM dataset. LBCs were interpolated to, and specified, every 3 h from the same data
set as the ICs. Hence, analyses provided LBCs for GFS- and NAM-based simulations,
and forecasts provided LBCs for GEFS/R2-based simulations.

All runs were initialized on 0000 UTC on the first day of the case study, and ran
for 36 h, to allow (a) mesoscale systems to develop appropriately, (b) perturbations
between ensemble members to grow large enough to observe easily, but not so large
that the timescale of interest was well beyond a predictability horizon for meso-α-scale
motion (Surcel et al. 2014), and (c) use of the once-daily GEFS/R2 data. All MCSs
of interest had at least 18 h between model initialization and convective initiation.
Preliminary tests, using NAM analyses, were started 12 h earlier and later, and did
not improve the simulation performance. Datasets from the Rapid Update Cycle, and
its successor Rapid Refresh (both hereby referred to as RUC), were used for synoptic
overviews, and to supplement observations when initially evaluating model perfor-
ance. However, for the focus of the present study, we verify model performance with
composite NEXRAD Level III radar reflectivity from archives at the Iowa State Uni-
versity (https://mesonet.agron.iastate.edu/docs/nexrad_composites/, accessed
1 September 2015). Base Reflectivity product data are composited through the GEM-
PAK program nex2img, after which suspected false echoes are removed through com-
parison with the Net Echo Top product. We also compared WRF 10-m wind output to National Climatic Data Center (NCDC) storm reports (https://www.ncdc.noaa.gov/stormevents/, accessed 1 September 2015), with the caveat that these reports can occasionally exaggerate or diminish the actual wind strength (Trapp et al. 2006).

Multiple ensemble types (experiments) were created (Fig. 5); those discussed in the following study are listed in Table 2 with their abbreviation and formulation. ICBC ensembles were created by running WRF twelve times, each with a different set of ICs and LBCs from the GEFS/R2 dataset (one control and ten perturbation members) and NAM analyses. Note the GFS-driven simulations provided little variation to NAM and GEFS/R2 datasets, and will not be discussed further in the present study. These ICBC runs used the control configuration (Table 1); hence Thompson is the only microphysical scheme used (ICBC-Thompson). Ensembles were also created by varying the microphysical scheme (MXMP), while holding ICs/LBCs and all else constant. The nine microphysical schemes (including the control scheme, Thompson; Table 3) were chosen to mirror a similar study by Adams-Selin et al. (2013b). In their method, the hydrometeor intercept (their Fig. 2) of graupel was modified in the WRF source code (Adams-Selin, personal communication), so that a parameterization could become ‘hail-like’ or ‘graupel-like’. The smaller intercept used in the ‘hail-like’ modification results in hydrometeors that are larger and more dense, and that fall faster; vice versa for ‘graupel-like’. An identical method has been used in the present study for the WSM6, WDM6, and Morrison schemes to improve the sampling of model-error phase space, resulting in 12 MXMP members. These variations are hereby denoted by “Hail” and “Graupel” (e.g., “WSM6 Hail”). As a caveat to the MXMP method, each member is not of equal likelihood in the same sense as a well-calibrated ensemble. Hence, this ensemble method is more correctly a sensitivity study, and does not rigorously measure predictability. However, it can offer insight into performance of each parameterization scheme.
To further sample the model uncertainty, three more ensemble experiments are used involving a SKEB scheme (e.g., Berner et al. 2011). The SKEB scheme accounts for energy lost between resolved and unresolved scales by randomly \(^1\) injecting kinetic and potential energy back into the model fields. **STCH** prescribes a constant IC/LBC dataset and parameterization. **STMX** couples a SKEB scheme with the same list of microphysical parameterizations as in MXMP \(^2\). Finally, the sensitivity of the STCH method was tested by changing the decorrelation time of temperature and stream-function perturbations from the default 0.5 h to 5.5 h: this variation is called **STCH5** \(^3\). As the kinetic energy spectrum in WRF contains the \(k^{-5/3}\) slope observed by Nastrom and Gage (1985) regardless of the SKEB perturbations (Duda et al. 2016), we tentatively propose that SKEB may instead be used as a ‘variance generator’. In preliminary testing, increasing the decorrelation time effectively turns down variance introduced through SKEB perturbations, but does not obviously degrade the simulation.

To track uncertainty between ensemble members, difference total energy (DTE) is used herein. As the summation of differences in kinetic and thermal energy at every grid point between all members, it serves as a measure of ensemble spread. Its advantage over simple ensemble standard deviation is the integrated impact of wind and temperature differences over three dimensions, including the relevance of diabatic heating to moist convection. Here, DTE is calculated at a given timestep similarly to that in Tan et al. (2004):

\[
DTE = \frac{1}{2} \sum (U_{ijk}^{2} + V_{ijk}^{2} + \kappa T_{ijk}^{2})
\]

\(^1\)The ‘randomness’ is via a seed integer specified in the WRF namelist. Hence unlimited independent ensemble members can be created by changing this value.

\(^2\)Note the seeds used in STCH are different to those specified in STMX.

\(^3\)The seeds used in STCH5 are identical to those in STCH to gauge the effect of increasing decorrelation time.
where \( \kappa \) here \(^4\) is 0.286, and \( U', V', \) and \( T' \) are the differences in \( u- \) and \( v- \) component horizontal wind, and potential temperature, respectively, at every grid point \((i,j,k)\) between two ensemble members. This is summed over all three dimensions to create a time series, or in height to create a latitude–longitude cross-section. For each ensemble, sets of differences were calculated between all permutations of the ensemble members without repetition. Note that, in the following pages, figures that show vertically integrated DTE use a number of contour scales, due to the intermittent rapid growth of DTE with time, and its variation with experiment type.

Also note that simulated composite reflectivity only includes rain and snow hydrometeors in the following figures to enable comparison over all MXMP members. In preliminary testing, this reflectivity was compared to that computed using all hydrometeor species available for each parameterization, and did not substantially affect the conclusion of MCS mode. In fact, reflectivity from all species tended to heavily overestimate reflectivity associated with stratiform precipitation. As a result of using this method (and considering the warm-rain-only nature of the Kessler scheme), we do not analyze inter-member magnitudes of reflectivity in the present study.

3 Synoptic overviews

3.1 NEKS06

The progressive bow echo of 26–27 May 2006 (NEKS06, Fig. 1a) is covered in more detail in Snively and Gallus (2014), where the authors found WRF runs forced by both NAM and GFS forecast datasets incorrectly reproduced the convective mode of the MCS. They also found little sensitivity to microphysical schemes. Regarding this case, May 26 and 27 will be referred to as Day 1 and Day 2, respectively.

\(^4\)DTE can been formulated using this constant value, or as in Tan et al. (2004), via use of a reference temperature.
Figure 2 shows RUC analyses of 500-hPa and 925-hPa geopotential heights, surface frontal positions, and their associated mean sea-level pressure (MSLP) minimum, at 1200 UTC on Day 1. The green star and arrow denotes the location of convective initiation (2200 UTC) and subsequent MCS movement (eastwards), respectively. At 1200 UTC on Day 1, the Nebraska–Kansas border sits underneath the entrance region of a small south-westerly 250-hPa jet maximum of 50 kt, visible in rawinsonde data (not shown), and underneath the axis of a synoptic-scale ridge evident in 500-hPa heights (Fig. 2). Winds become more southerly towards the surface; at 925 hPa, a weak height trough lies along the Nebraska–Kansas border. At the surface, a quasi-stationary warm front, as analysed by the Weather Prediction Center (WPC), stretches through Kansas (Fig. 2). Its associated MSLP minimum in southeast Colorado lies close to the location of convective initiation 10 h later. This synoptic set-up and event evolution, with the MCS of interest moving parallel to a zonal surface front, is similar to Fig. 4 in Bentley et al. (2000), associated with derechos.

Figure 1a presents observed composite radar reflectivity at 2300 (Day 1), 0300 (Day 2), and 0600 UTC (Day 2). Convective initiation of interest occurs at 2200 UTC on Day 1. The cell strengthens in reflectivity intensity, and the mode becomes linear by 0000 UTC on Day 2. While the system continues growing upscale at the beginning of Day 2, the formation of a discrete bowing line is rather sudden between 0200 and 0300 UTC. NCDC storm reports associated with this MCS include hail 2–2.5 cm (0.75–1 in), wind gusts up to 36 m s\(^{-1}\) (70 kt), and a landspout tornado. Between 0400 and 0500 UTC, a second line of moist convection initiates northeast of the first MCS. By 0600 UTC, these two lines of convection form a disconnected arc; a third line of moist convection perpendicular to the arc’s tangent forms in the wake of the primary bowing segment, in a “bow-and-arrow” structure (Keene and Schumacher 2013). The two leading arc segments merge by 0800 UTC as the system moves into western Iowa and
northwestern Missouri. The system weakens in reflectivity as it continues to move east but still produces hail close to 2.5 cm (1 in) in Iowa.

3.2 KSOK13

The progressive bow echo of 15–16 August 2013 (Fig. 1b) brought damaging wind and hail to Kansas, Oklahoma, and Texas. August 15 and 16 are referred to as Day 1 and 2 for this case, respectively. In contrast to midtropospheric west-to-southwest winds in NEKS06, the area of interest at 1200 UTC on Day 1 lies under northwesterly flow at 500 hPa (Fig 3), between an upstream ridge and downstream trough. Winds become weaker and more northerly close to 700 hPa (not shown), and are variable and light at 925 hPa. At the surface, a weak frontal wave (analyzed by the WPC) straddles a MSLP minimum near the Nebraska–Kansas border. This zonally oriented quasi-stationary front slowly migrates south, and initiation (green star in Fig. 3) occurs to the north of this boundary around 2200 UTC. Storm Prediction Center Mesoscale Discussions for this day mention a prior mesoscale convective vortex (MCV) moving southwards, and this is evident in visible satellite data (not shown). The southeastern (downshear; 0–6-km vertical wind shear not shown) edge of this MCV appears to focus moist convection, similar to that seen in idealized simulations by Davis and Weisman (1994). This convection then forms a line by 2200 UTC (Fig. 1b) and begins bowing at 2330 UTC. The line produces a swath of strong wind (up to 34 m s\(^{-1}\) or 67 kt) and large hail (up to 4.4 cm or 1.75 in) primarily in central Kansas and near the Oklahoma–Texas border.

4 ICBC experiments

This section details the results from ensemble simulations that use IC and LBC perturbations from the GEFS/R2 dataset. Note the NAM-driven member for each case
is included in section 5 as the control member of the NAM-MXMP experiment. All ICBC experiments use the control (Thompson) microphysics parameterization.

4.1 NEKS06

No ICBC-Thompson members simulate any substantial reflectivity structures in the region of interest during the first 33 h (not shown); hence the verification (observed convection) falls well outside the envelope of the ICBC-Thompson simulation. There is strong agreement between ICBC-Thompson members regarding frontal location (not shown), but as this consensus position is incorrect in comparison with observations, it suggests inadequate dispersion in the limited GEFS/R2 dataset.

4.2 KSOK13

The first 21 h of this case are simulated poorly by ICBC-Thompson, with moist convection occurring in locations different from that observed; however, performance improves thereafter. At 2100 UTC on Day 1, a line of cells is observed in reflectivity data over north-central Kansas; in ICBC-Thompson members, there is a large spread of solutions in cell evolution (Fig. S1). At 0000 UTC on Day 2 (Fig. S2), eight members have line segments, seven of which have begun bowing; the three remaining members form two regions of cells. Three hours later (0300 UTC, Fig. 6), the observed bow echo has its tightest radius of curvature. In ICBC-Thompson, ten members have a bowing line, but the locations vary from the Nebraska–Kansas border (p04, p08) and central Kansas (p02, p07) to the Kansas–Oklahoma border, the location of the observed bow echo (c00, p01, p03, p05, p09, p10). The last member simulates a straight line in the correct location (p06), but soon after develops bowing.

---

5 All figures with “S” prefix are online-only in the original publication; Figs. S1–S5 are provided in this dissertation as a Supplementary Material appendix after the figures.
In summary, the location of initiation and modes of initial convection do not necessarily predict the resulting simulated system's location and strength. In other words, there is not a 'non-reversible' bifurcation of solution clusters; despite the solution of a bow echo being simulated by all 11 members, some members follow different trajectories en route. The bowing structure is a stable solution despite high sensitivity of location, timing, and intermediate mode to the IC/LBC perturbations. In addition, prior (Day 1) convection was not correctly simulated, but did not preclude the formation of correct mode, timing, and locations of the bow echo in many of the ensemble members.

Integrating DTE vertically shows that, at 0300 UTC on Day 1, uncertainty is larger in two general areas (Fig. 7a): (1) locations with moist convective activity in simulated radar reflectivity, where DTE growth is expected to be larger (Zhang et al. 2003), and (2) along the MSLP trough running west–east in Nebraska. Over the next 6 h, another DTE maximum is associated with the developing MCV (Fig. 7b). At 1800 UTC on Day 1, there is increasing homogeneity in the domain-wide DTE field as moist convection dissipates (Fig. 7c). Yet the local maximum of DTE associated with the MCV stands out from this background field; at 1800 UTC, moist convection initiates on the southeastern (downshear) side of the MCV both in observations and most simulations. In the next 12 h, small variations in location and timing of this convective initiation appear related to differences in the structure of the MCV between ensemble members (Fig. 7d). Eventually these small inter-member differences grow to become large (>5000 m² s⁻²) DTE values, while almost all members generate a bow echo that moves southward through Kansas and Oklahoma, but in a spread of locations with variations in bowing structure (Fig. 6). We see that, in contrast to NEKS06, the GEFS/R2 dataset provides substantial differences in KSOK13 related to the development of convection. However, the mode solution (i.e., a bow echo) appears to be highly predictable, even if the location and specifics of the bowing are more uncertain.
5 MXMP experiments

This section details the results from numerous mixed-microphysics ensemble simulations, forced with either a NAM-analysis dataset or a given ensemble member of the GEFS/R2 dataset.

5.1 NEKS06

Results from c00-MXMP showed poor performance and almost no moist convection during the MCS of interest (not shown), no matter what microphysical scheme was used, in line with ICBC-Thompson results. It is likely this is general regardless of the GEFS/R2 perturbation member used to drive the MXMP experiment, due to insufficient variation between GEFS/R2 members. The c00-MXMP experiment has considerably less spread than ICBC-Thompson (discussed in section 7); DTE calculated between a given parameterization and all others (not shown) shows almost identical DTE growth between most microphysical schemes in this experiment. This reduced uncertainty between microphysical schemes is likely due to the limited amount of moist convection that does not permit spread to grow rapidly through variations in the microphysical parameterizations. However, it is still surprising that c00-MXMP spread is not comparable to that in ICBC-Thompson: Stensrud et al. (2000) found larger variation with varied convective and PBL parameterizations in the first 12 h than variation using perturbed ICs and LBCs. This suggests that, in certain flows with a fixed set of ICs/LBCs, erroneously low ensemble spread cannot be mitigated through parameterization variability alone.

The NAM-MXMP experiment also begins poorly, and does not capture the upscale growth of convection into a line of cells in southern Kansas in the first few hours of the simulation (not shown). As an improvement on c00-MXMP, most members do initiate a northwest–southeast line of convection across Kansas by 0800 UTC on Day 1. The
analogous feature in observations initiates later on Day 1 (1000 UTC) and is orientated NNW–SSE. This suggests that the position of the front may be manipulated by earlier warm-sector convection and subsequent upscale growth of the convective mode, and that accounting for model error is critical to correctly modulate larger-scale baroclinic boundaries. By 0200 UTC on Day 2, cells grow, move northeastward, and grow upscale in both observed and model data, but no ensemble members recreate the bow echo and subsequent turning of the system to the east-southeast as it lengthens in scale. At this point, 26 h into the simulation, all ensemble members appear to critically diverge from verification. The closest member at 0600 UTC on Day 2, by eye, uses the WDM6 Graupel scheme (Fig. 8j), but its simulated bow echo never turns to the east-southeast, and instead continues moving northeast to merge with another linear feature at the Iowa–Nebraska–South Dakota borders. This \( \sim 45^\circ \) error in MCS trajectory is likely related to a comparable error in midtropospheric wind direction (e.g., 500-hPa model winds, not shown) between RUC analyses and both GEFS/R2- and NAM-driven ensemble members, as in Snively and Gallus (2014). This error in storm motion appears critical by taking the developing MCS away from the frontogenesis maximum (which is correctly placed in NAM-MXMP members; not shown), and attendant convergence and positive equivalent-potential-temperature advection originating in the warm sector. The source of such error in large-scale flow is likely to be in ICs and LBCs, which are fixed in MXMP experiments.

5.2 KSOK13

For KSOK13, we first fix ICs and LBCs using the subjectively best ICBC member (p09; see Fig. 6k) to test the sensitivity of a subjectively good simulation to choice of parameterization (p09-MXMP). By 0300 UTC on Day 2 (Fig. 9), all p09-MXMP members create a progressive bow echo with a tight radius of curvature as in the control
(Thompson), with the exception of the Morrison (both Hail and Graupel) members (Fig. 9l,m).

Conversely, while p09-MXMP members resembled observed reflectivity structures, NAM-MXMP members did not. Simulated reflectivity from NAM-MXMP shows much variation between members on Day 1, including swaths of convection in Kansas and Oklahoma at 1200 UTC that is not observed (Fig. S3). At 2100 UTC, after a lull in moist convection, a completely different solution from p09-MXMP unfolds (Fig. S4): a southwest–northeast boundary triggers a line of cells across the Nebraska–Kansas border. By 0000 UTC, NAM-MXMP members display a variety of solutions, some with bowing segments along broken lines. Overall, convection is more scattered and disorganized than in p09-MXMP. By 0300 UTC (Fig. 10), all members have a similar theme: a south-southwest–north-northeast broken or unbroken line, with or without bowing sections embedded within the line (some resembling a serial bow echo). The simulated MCS locations are from the Texas and Oklahoma panhandles towards central Kansas. This is much different from the tightly curved bow echo observed at the Kansas–Oklahoma border. The WDM6 Graupel member maximizes 10-m wind magnitude and areal extent (not shown). This corresponds with the prominent bowing structure in simulated reflectivity, typically associated with the rear-inflow jet and damaging surface winds (Przybylinski 1995; Markowski and Richardson 2010), associated with this member (Fig. 10j).

6 STEX and STHCH experiments

In this section, results from SKEB ensembles (with and without fixed microphysics) are detailed, including the STCH5 variation, using both GEFS/R2 and NAM datasets.
6.1 NEKS06

The addition of a SKEB scheme to a MXMP configuration changes the mode, strength, or location of convection to varying degrees. Figure 11 presents three microphysical schemes without (NAM-MXMP) and with (NAM-STMX) a SKEB scheme, valid at 0600 UTC on Day 2. The three parameterizations (Morrison Graupel, Morrison Hail, and Ferrier) are discussed here for their varying sensitivity to the SKEB scheme. Contrast the Morrison Graupel without and with SKEB (Fig. 11a–b), particularly the split in the latter of the convective line near the South Dakota–Nebraska border. Interestingly, a discrepancy of this magnitude does not occur in the Morrison Hail member, despite the single change in hail–graupel dynamics (Fig. 11c–d). Next, likewise contrast Ferrier without and with SKEB (Fig. 11e–f). In this case, addition of the SKEB scheme changes the orientation of the linear convection.

The NAM-MXMP member closest to the observed bow-echo reflectivity (WDM6 Graupel; Fig. 8j) changes very little with the addition of a SKEB scheme in the NAM-WDM6Graupel-STCH experiment (not shown). The control member (i.e., without SKEB) is fairly representative of NAM-WDM6Graupel-STCH members at 0500 UTC on Day 2 (Fig. S5). In addition, c00-Thompson-STCH does not improve on the poor simulation seen in GEFS/R2-driven ICBC and MXMP experiments. When contrasting GEFS/R2-driven and NAM-driven STCH experiments, we note the dependence of spread on the IC/LBC set chosen (e.g., Alhamed et al. 2002). DTE growth in NAM-WDM6Graupel-STCH follows a similar evolution to NAM-MXMP (Day 1 moist convection is present), whereas DTE growth in c00-Thompson-STCH is closer to ICBC-Thompson and c00-MXMP (without Day 1 moist convection).

This apparently random impact of SKEB perturbations on precipitation structure matches speculation by Romine et al. (2014) that such variation in a 3-km SKEB ensemble simulation “may be a common pattern”. An increase in decorrelation time from 0.5 h to 5.5 h (NAM-WDM6Graupel-STCH and NAM-WDM6Graupel-STCH5, re-
spectively) reduces overall spread, but the DTE field is structurally similar (not shown). Note the SKEB perturbation seeds are identical between the STCH and STCH5 experiments. The reduction in domain-wide spread also does not substantially decrease magnitude of the DTE local maximum embedded within the low-DTE region. This further associates high sensitivity of bow-echo structure with small perturbations, even when large-scale uncertainty is reduced.

6.2 KSOK13

Similarly to Fig. 11, Fig. 12 presents three microphysical schemes without (from p09-MXMP) and with (from p09-STMX) a SKEB scheme, valid at 0300 UTC 16 August 2013 (Day 2). We have chosen WSM6 Graupel, WSM6 Hail, and WDM5 members to highlight the varying sensitivity of the simulated MCS to SKEB perturbations evident in simulated composite reflectivity (a–f), 850-hPa wind (g–l), maximum 10-m wind over 20 min (m–r), and density potential temperature perturbation ($\theta'_{\rho}$, s–x). $\theta'_{\rho}$ is chosen to depict the cold pool strength, as in Markowski and Richardson (2010), and is computed by subtracting density potential temperature $\theta_{\rho}$ from the domain mean at each timestep, where

$$\theta_{\rho} = \theta(1 + 0.61 r_v - r_h)$$

and where $r_v$ and $r_h$ are the mixing ratios of water vapor and the sum of all other hydrometeor species, respectively. Figure 12, and animations of the same fields in Figs. S6–S9 6, raise two points:

- The sensitivity of the microphysical scheme to SKEB may be substantially changed by changing the hail/graupel coefficient. This is also seen in Fig. 11. The introduction of SKEB to the graupel variation of WSM6 (the two top-row panels in

---

6 These animations are available online at the Weather and Forecasting journal website
each six-panel frame) in Fig. 12 straightens the line somewhat (cf. panels a and b), and weakens winds considerably at both 850 hPa (panels g and h) and 10 m (panels m and n). The surface-based cold pool is not noticeably weaker, however (cf. panels s and t). When the hydrometeors are more hail-like (middle rows), there is much less variation in all fields presented here between no-SKEB and with-SKEB simulations. As the SKEB perturbations vary with each member (and simulation initialization time), we cannot make general conclusions about a parameterization’s performance or sensitivity to small perturbations. However, this itself is an important consideration when assessing a parameterization within an ensemble that accounts for model error.

- **An increase in bowing radius—a weaker bow-echo signal—may not be associated with weaker 850-hPa and surface wind magnitudes.** In Fig. 12, the addition of a SKEB scheme to WDM5 weakens the bowing signal (panels e and f) and the rear-inflow jet (panels k and l), but the peak 10-m wind magnitude increases (panels q and r). Conversely, the less impressive bow in panels a and b (WSM6 Graupel), after SKEB is introduced, is associated with weaker winds at both 850 hPa (panels g and h) and the surface (panels m and n). While these figures are a small sample, this lack of consistent relationship between bowing curvature and surface wind in simulations was noticed in different ensemble members, and was noted by Wandishin et al. (2010) in their own simulations.

We now assess the contrasting performance of the subjectively ‘best’ (Thompson) and ‘worst’ (Morrison Hail) p09-MXMP members with STCH experiments. All members contain bowing in the p09-Thompson-STCH experiment at the time of maximum curvature (0300 UTC on Day 2; Fig. 13), though the radius of curvature varies between members. Notably, the control (i.e. no SKEB scheme) is the best member of this experiment. The other members have similar or less bowing in their simulated
systems, suggesting that the initial subjectively best performance of Thompson was partly fortuitous, or that a SKEB scheme degrades the forecast. When we look at the same time for p09-MorrisonHail-STCH (Fig. 14), there is a wider spread in solutions, some of which are as close to verification as p09-Thompson-STCH members. Some members generate two separate bowing segments; others are similar to the control. This shows the Morrison Hail parameterization’s ‘worst’ performance in p09-MXMP was again through insufficient sampling of model phase space. Note as SKEB members in p09-MorrisonHail-STCH outperformed the control, SKEB is unlikely to be systematically degrading forecast skill; however, the limited sample size precludes general statements. The low DTE magnitude in these STCH ensembles compared to the other experiments (discussed in section 7) is related to even more spatial agreement, but only slightly less variation in MCS structure. Maximum 10-m wind over the period of the bow echo (not shown) shows that this variation also affects the locations of surface wind maxima, perhaps associated with downbursts within the bow echo. However, within the simulations, variation in structure is not a reliable predictor of surface-wind magnitude (as seen in Fig. 12). Surface wind is discussed further in section 7.

7 Synthesis

7.1 Ensemble uncertainty

Figure 15 shows time series of DTE integrated over all three spatial dimensions for seven NEKS06 experiments: ICBC-Thompson, NAM-MXMP, NAM-STMX, c00-MXMP, c00-Thompson-STCH, and NAM-WDM6Graupel-STCH and -STCH5. In NAM-driven experiments, DTE decreases to a local minimum around 1800 UTC, likely as the disturbed air advects out of the domain, and as more quiescent flow enters (regression to

\[ \text{We include only GEFS/R2 members here and for KSOK13 to compare spread between experiments. The inclusion of the NAM-driven member would substantially inflate the ensemble spread. Spread of a mixed-model ensemble approach is outside the scope of the present study.} \]
the ensemble mean). Despite large areas of radar reflectivity across the domain (not shown), the precipitation is larger scale and less intense, and less destructive in terms of predictability. DTE rapidly grows after this, the time of maximum solar insolation (∼1800 UTC), on Day 1. This is likely related to the onset of cellular convection at this time and accompanying destruction of predictability (Zhang et al. 2003). There is little difference in spread between NAM-MXMP and NAM-STMX experiments (Fig. 15), showing negligible overall impact of the SKEB scheme with default parameters to uncertainty. Uncertainty growth in the overnight (0300–1200 UTC) periods for both days appears strongly dependent on occurrence of moist convection; the GEFS/R2-based experiments that struggle to initiate moist convection do not have as pronounced bimodality in DTE.

Likewise, Fig. 16 shows time series of DTE for six KSOK13 experiments: ICBC-Thompson, p09-MXMP, NAM-MXMP, p09-STMX, p09-Thompson-STCH, and p09-MorrisonHail-STCH. Ensemble uncertainty is comparable in magnitude between NEKS06 and KSOK13 (cf. Figs. 15 and 16). Similarly to NEOK06 (Fig. 15), KSOK13 displays a twin-peak structure of DTE, with maxima around midnight local time (around 6 and 30 forecast hours). This is again likely related to moist convection during the peaks. Note, in contrast to NEKS06, that ICBC-Thompson has the largest domain-wide DTE, followed by STMX, MXMP, and STCH experiments. The lower diversity in NEKS06 ICBC-Thompson is likely related to the lack of convection associated with GEFS/R2 ICs/LBCs. The better performance of KSOK13 matches previous findings that ensemble skill is largest when IC/LBC uncertainty dominates model uncertainty (Murphy 1988). Also note that NAM-MXMP has larger DTE than p09-MXMP, but a worse forecast (in contrast to NEKS06, where the badly performing experiment had less DTE). A lack of relationship between spread and skill was found in Berner et al. (2011), though a loose relationship was found in Buizza (1997). DTE growth between the two p09-driven STCH experiments, using Morrison Hail and Thompson micro-
physics, is similar up to 2100 UTC on Day 1. After this, spread grows faster in the Morrison Hail member. This corroborates the larger spread, by eye, of modes in simulated reflectivity (cf. Figs. 13 and 14).

Figure 17 shows vertically integrated DTE for a collection of experiments in the KSOK13 case, at 0000 UTC on Day 2, shortly after the MCS of interest has initiated in most ensemble members in all experiments. The panels are in descending order of domain-wide DTE; this is generally seen as a diminishing area of low DTE (blue colors) through the panels. The DTE maximum associated with the simulated MCS is centered in a broad region of low DTE (<2000 m$^2$s$^{-2}$), but still exceeds 6000 m$^2$s$^{-2}$ in all members. As the spread of the MCS’s positioning and timing reduces through the pyramid of experiments, MCS modes still vary between straight and bowing lines (cf. Figs. 6, 9, 10, 13, and 14 at 0300 UTC on Day 2). As DTE is integrated vertically here for each grid point, and as ensembles reach more consensus on the MCS position, DTE generation is concentrated on a smaller area. We do not see a reduction in the local maximum around the MCS, as might be expected with a consensus of position. Hence, the bowing structure is associated with uncertainty (high DTE) on small (~10 km) scales, as expected (Lorenz 1969), with the caveat that no causation is implied between DTE and variance in reflectivity. (It is not apparent whether ensemble spread is creating diversity in MCS mode, or vice versa.)

7.2 Simulated 10-m wind

Simulated wind speeds associated with the bow echo in KSOK13 were too low in general across all experiments (e.g., 10-m wind for KSOK13 shown in Fig. 12). The underestimation may come from the calculation of WRF 10-m wind output, which uses Monin-Obukhov similarity theory. Wind speed output explicitly at the lowest model level (~40 m above ground level) was close to peak observed speeds during the KSOK13 bow-echo event: around double the speed inferred at 10 m (not shown).
Strong winds exist throughout the near-surface levels (some areas perhaps associated with the rear-inflow jet; Fig. 12m–r). It is not clear if the underestimation at 10 m is due to an invalid 10-m computation, model error in the fixed PBL scheme (MYNN Level 2.5), or simply inadequate sampling of model-error phase space. Regarding the latter, while the present study varies only the microphysics parameterizations—likely the largest source of model error—a fixed PBL and surface-layer scheme will limit the spread of surface wind forecasts.

We also note the control (i.e., no-SKEB) member of p09-Thompson-STCH had the weakest winds associated with the KSOK13 bow echo. A similar result is discussed in Lawson and Gallus (submitted to *Monthly Weather Review*), where bow echoes moved faster in SKEB ensemble members versus control members, and may be related to the extra (missing) kinetic energy introduced by the SKEB scheme.

### 7.3 Sensitivity of simulations to hail/graupel variation

Figure 12, and animations of the same fields in Figs. S6–S9, indicate that a change in the hail/graupel coefficient may substantially change the bowing radius of the MCS leading edge. The top- and middle-row panels in the left column of each six-panel frame show graupel and hail variations of WSM6, respectively. Neither the reflectivity bowing structure (cf. panels a and c) nor the 10-m wind (panels g and i) are substantially changed by the change from graupel-like to hail-like fall speeds. However, the rear-inflow jet weakens slightly (panels m and o), while the cold pool is more pronounced (panels s and u). Sensitivity of linear convection to the hail/graupel coefficient is also seen in Fig. 11. Adams-Selin et al. (2013b) found that graupel-like variants of microphysical schemes (i.e. smallest mean size) generated stronger cold pools and rear-inflow jets, and hence MCSs in these simulations bowed earlier, than hail-like variants. This is in contrast to Fig. 12s–v, that show stronger cold pools in the hail variations, and little change in the rear-inflow jet at 850 hPa. However, in
Figs. 12–14, we find that microphysical schemes are sensitive to small SKEB perturbations regarding MCS mode and radius of bow curvature. From this, we suggest that any conclusions about a given microphysical scheme’s performance may be misleading without, e.g., a SKEB ensemble to account further for model error.

8 Conclusions

We have presented two progressive bow echoes, NEKS06 and KSOK13, simulated with multiple ensemble techniques: perturbed ICs and LBCs; mixed microphysical parameterizations; and SKEB perturbations. All ensemble simulations of NEKS06 were poor, with only a few cherry-picked ensemble members simulating an MCS with a bowing structure. On the other hand, simulations of KSOK13 were mostly successful, with a progressive bow echo simulated in almost all cases, timing and location spread notwithstanding. As uncertainty decreases between different ensemble types in KSOK13, so do inter-member differences in location and timing. However, the spread of convective mode remains high, and the locations of strongest surface winds do not substantially lose variation. This suggests relatively high sensitivity to the microscale.

Simulated composite reflectivity fields showed that the spread of convective mode in ensembles using multiple microphysical schemes and those using SKEB perturbations was comparable. Overall uncertainty in the mixed microphysics ensemble, however, was 1.5–2 times the spread in the SKEB ensemble, as measured by ensemble differences in kinetic and thermal energy. Changing the SKEB scheme’s decorrelation time from 5.5 h to 0.5 h, with a prescribed microphysical scheme, increased spread more than adding a 0.5-h SKEB scheme to a mixed-microphysics ensemble. Implementing the SKEB scheme does not noticeably bias the convective mode, but appears to normalize the extreme performers in a mixed-microphysics ensemble. For example, in SKEB ensembles using the ‘best’ microphysics from a previous ensemble, many
members are worse than the no-SKEB control. The SKEB ensemble spread itself is dependant on the flow regime, as expected (Berner et al. 2009), and on the microphysical scheme selected. Moreover, the change in the hail/graupel coefficient within the parameterizations can be critical for bow development, as in Adams-Selin et al. (2013a), and SKEB is itself sensitive to this coefficient. This highlights the complex nature of model error, something that may require stochasticity in the hail/graupel fall-speed coefficient itself, instead of an appended stochastic forcing scheme.

In KSOK13, the uncertainty from ICs and LBCs dominates other sources of uncertainty, while uncertainty from mixed microphysics dominates in NEKS06. That KSOK13 performed better with larger IC/LBC spread is expected from Murphy (1988). These larger differences in ICs/LBCs perturbed the positioning of MCSs but almost all members still formed a bow echo. This suggests in KSOK13 that IC/LBC differences primarily changed the MCS’s position and timing, but spread associated with model error primarily affected the mode of convection. Furthermore, varied mixed microphysics and SKEB perturbations did not improve poor GEFS/R2 ICs/LBCs in NEKS06 and poor NAM ICs/LBCs in KSOK13. This appears to support the idea that small-scale errors (butterflies) are not significant when considering overall model skill (Durran and Gin-grich 2014), but are crucial to spread, and hence determining likelihood of severe weather (correlated with the convective mode).

In light of these findings, we return to address the hypotheses in the first section:

1. **Progressive bow echoes are inherently less predictable than other MCSs.**

   This is most likely, as large convective mode spread is associated with uncertainty on the smallest scales, generated by mixed parameterizations and SKEB. The storm scale is known to have limited variance at short lead times, and has a much shorter predictability horizon than the synoptic scale. Both factors increase the importance of accounting for model uncertainty through perturbation techniques. The poor performance of NEKS06 suggests the ensemble spreads were
insufficient to sample this hypothetical small region of phase space. We specu-
late that the predictability horizon may exist too soon to correctly simulate cell
mergers or growing supercells that precede many bow echoes (Klimowski et al.
2003). The caveat is that KSOK13 shows that MCS mode can be a stable solution
within a perturbed-IC and -LBC ensemble, even if the MCS’s position is displaced
from that observed.

2. **Progressive bow echoes are embedded in less predictable synoptic-scale
   regimes.** If progressive bow echoes are indeed highly sensitive to model uncer-
   tainty, it follows that this sensitivity is compounded in a weakly forced regime,
   where perturbations related to model error dominate over IC/LBC perturba-
   tions. Both cases presented herein occur without particularly strong upstream
   height troughs. The dominance of mixed-microphysics ensemble uncertainty
   over IC/LBC uncertainty in NEOK06 may have contributed to its poor perfor-
   mance.

3. **There is critical deficiency in ICs and LBCs.** The success of KSOK13 but failure
   of NEKS06 leaves an unresolved issue here. Regardless, errors in IC/LBC datasets
   are unavoidable, and hence must be mitigated with well-dispersed ensembles.
   Our results suggest that improvement of ICs and LBCs would yield better timing
   and positioning of MCS systems, but provide diminishing returns on MCS mode.
   Previous studies have raised concern at reduced variance on storm scales within
   global ensemble datasets used to drive limited-area models. While the present
   study does not address suitable spread directly, our results in KSOK13 do show
   that a 24–36 h simulation can successfully capture a progressive bow echo using a
   coarse, global, reforecast dataset; this driving dataset outperforms a limited-area
   analysis dataset.
4. **There is critical deficiency in the microphysics parameterizations.** The contrasting performance by mixed-microphysics ensembles between our two cases suggests that the ICs/LBCs or embedding regime were more important than error from parameterizations. Results showed that parameterizations are substantially sensitive to small perturbations, here introduced through a SKEB scheme, and this sensitivity is not regular. Hence the component of error associated with parameterizations appears complex and strongly non-linear. There is little relationship between bowing radius and simulated wind speed, as in Wandishin et al. (2010), despite strong wind at all low-tropospheric model levels, but this may be due to a calculation unsuitable for bow-echo events to estimate 10-m wind within WRF. Weak winds may also be related to problems within the mechanism of mixing winds in the PBL, but is outside the scope of the present study.

A key question remains outstanding: **is the lack of adequate dispersion in NEOK06 a cause or consequence of convective initiation failure?** Model uncertainty grows to dominate IC/LBC uncertainty in strongly forced cases (Stensrud et al. 2000), where methods like mixed-microphysics and SKEB ensembles are needed to generate small-scale variance in the absence of convective foci. But in the results herein, substantial variance is not generated if convection never initiates. The new stochastic convective backscatter (SCB; Shutts 2015) scheme targets convection as the largest source of model error, but is unable to account for locational error in convection.

Further large-scale conclusions are difficult to make from two cases; future work should address the relationship of storm- and synoptic-scale predictability associated with MCSs. In addition, the impact of grid resolution on bow-echo ensemble simulations is the subject of a recent submission (Lawson and Gallus, submitted to *Monthly Weather Review*).
Acknowledgments

The authors thank: three anonymous reviewers and the editor for their contributions towards improving the manuscript; Rebecca Adams-Selin and her colleagues for supplying the Fortran modifications and advice relating to the microphysical schemes; David John Gagne II, Tim Supinie, Daryl Herzmann, and David Flory for IT assistance; Patrick Marsh, Chris Karstens, Jeff Duda, Adam Clark, and others at the Hazardous Weather Testbed (Norman, Oklahoma) for enlightening discussions; and Xiaoqing Wu, Andy VanLoocke, Ray Arritt, and William Gutowski for constructive criticism. We are grateful for the utility and power of open source Python code and packages (matplotlib, numpy, basemap, netcdf4-python): the modern-day giants’ shoulders. This research was supported by NSF grant AGS-1222384.
Table 1  Control parameterization schemes used in the numerical modeling configuration.

<table>
<thead>
<tr>
<th>Parameterization</th>
<th>Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microphysics</td>
<td>Thompson</td>
</tr>
<tr>
<td>Longwave Radiation</td>
<td>RRTM</td>
</tr>
<tr>
<td>Shortwave Radiation</td>
<td>Dudhia</td>
</tr>
<tr>
<td>Surface Layer</td>
<td>MYNN</td>
</tr>
<tr>
<td>Land Surface</td>
<td>Noah</td>
</tr>
<tr>
<td>Planetary Boundary Layer</td>
<td>MYNN Level 2.5</td>
</tr>
</tbody>
</table>
Table 2  Ensemble configurations addressed in text. Also refer to Fig. 5.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>ICs and LBCs</th>
<th>Microphysics</th>
<th>SKEB scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICBC-Thompson</td>
<td>Varied</td>
<td>Thompson</td>
<td>Off</td>
</tr>
<tr>
<td>NAM-MXMP</td>
<td>NAM</td>
<td>Mixed</td>
<td>Off</td>
</tr>
<tr>
<td>c00-MXMP</td>
<td>GEFS/R2 c00</td>
<td>Mixed</td>
<td>Off</td>
</tr>
<tr>
<td>p09-MXMP</td>
<td>GEFS/R2 p09</td>
<td>Mixed</td>
<td>Off</td>
</tr>
<tr>
<td>p09-STMX</td>
<td>GEFS/R2 p09</td>
<td>Mixed</td>
<td>0.5 h decorrelation</td>
</tr>
<tr>
<td>NAM-STMX</td>
<td>NAM</td>
<td>Mixed</td>
<td>0.5 h decorrelation</td>
</tr>
<tr>
<td>p09-Thompson-STCH</td>
<td>GEFS/R2 p09</td>
<td>Thompson</td>
<td>0.5 h decorrelation</td>
</tr>
<tr>
<td>p09-MorrisonHail-STCH</td>
<td>GEFS/R2 p09</td>
<td>Morrison Hail</td>
<td>0.5 h decorrelation</td>
</tr>
<tr>
<td>NAM-WDM6Graupel-STCH</td>
<td>NAM</td>
<td>WDM6 Graupel</td>
<td>0.5 h decorrelation</td>
</tr>
<tr>
<td>NAM-WDM6Graupel-STCH5</td>
<td>NAM</td>
<td>WDM6 Graupel</td>
<td>5.5 h decorrelation</td>
</tr>
<tr>
<td>c00-Thompson-STCH</td>
<td>GEFS/R2 c00</td>
<td>Thompson</td>
<td>0.5 h decorrelation</td>
</tr>
</tbody>
</table>
Table 3  Microphysical schemes used in the MXMP experiments.

<table>
<thead>
<tr>
<th>Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thompson †</td>
</tr>
<tr>
<td>WSM6 (Hail/Graupel) *</td>
</tr>
<tr>
<td>Kessler</td>
</tr>
<tr>
<td>Ferrier</td>
</tr>
<tr>
<td>WSM5</td>
</tr>
<tr>
<td>WDM5</td>
</tr>
<tr>
<td>Lin</td>
</tr>
<tr>
<td>WDM6 (Hail/Graupel) *</td>
</tr>
<tr>
<td>Morrison (Hail/Graupel)*</td>
</tr>
</tbody>
</table>

† Control parameterisation
* Changed to be either hail- or graupel-like
Figure 1  Observed NEXRAD composite radar reflectivity for the two cases found in the present study, merged over three times each. Panel (a) shows the evolution of a single cell (2300 UTC), to a bow echo (0300 UTC), and finally a bow-and-arrow structure (0600 UTC; note ‘arrow’ feature farther west), on 26–27 May 2006 (NEKS06). Panel (b) shows the development of a linear MCS (2200 UTC) into a bow echo (0200 and 0600 UTC), on 15–16 August 2013 (KSOK13). States are labeled for reference (see Fig. 4 for context).
Figure 2  Geopotential height fields from RUC analysis at 500 hPa (black) and 925 hPa (lavender), contoured every 60 and 30 m, respectively, and valid at 1200 UTC 26 May 2006 (NEKS06 Day 1). Stationary surface front denoted by red/blue line, and low MSLP center marked by red L; both adapted from Weather Prediction Center synoptic analyses. Green star denotes convective initiation of the MCS of interest at 2200 UTC. Green arrow denotes approximate movement of the MCS.
Figure 3  As Fig 2, but valid at 1200 UTC 15 August 2013 (KSOK13 Day 1), with convective initiation at 2200 UTC.
Figure 4  WRF domains for NEKS06 and KSOK13. Labels refer to U.S. states mentioned in text.
Figure 5  Schematic diagram showing the methodology of creating ensembles. The green boxes on the left mark the four experiments; the control and perturbation members are colored blue for IC/LBC perturbations (ICBC), yellow for different microphysical schemes (MXMP), red for SKEB perturbations (STCH uses 0.5 h decorrelation time; STCH5 uses 5.5 h), and orange for a combination of microphysical scheme variation and SKEB perturbations (STMX). Arrows follow example paths down the ‘family tree’ of ensembles.
Figure 6  Observed (panel a) and simulated (b–l) composite reflectivity for GEFS/R2 members of ICBC-Thompson, valid 0300 UTC 16 August 2013 (KSOK13, Day 2).
Figure 7 Evolution of Difference Total Energy in the GEFS/R2 members of ICBC-Thompson, for the KSO13 case, valid at (a) 0300 UTC, (b) 0900 UTC, (c) 1800 UTC on 15 August 2013, and (d) 0000 UTC on 16 August 2013. Scale is in units of m² s⁻².
Figure 8  Observed (panel a) and simulated (b–m) composite reflectivity for NAM-MXMP ensemble members, valid 0600 UTC 27 May 2006 (NEKS06, Day 2).
Figure 9 Observed (panel a) and simulated (b–m) composite reflectivity for p09-MXMP ensemble members, valid 0300 UTC 16 August 2013 (KSOK13, Day 2).
Figure 10  Observed (panel a) and simulated (b–m) composite reflectivity for NAM-MXMP ensemble members, valid 0300 UTC 16 August 2013 (KSOK13, Day 2).
Figure 11 The sensitivity of three microphysical schemes in NEKS06 to the hail/graupel fall speed and addition of a SKEB scheme, taken from NAM-MXMP (a, c, e; also in Fig. 10) and NAM-STMX (b, d, f) members: (a, b) Morrison Graupel, (c, d) Morrison Hail, (e, f) Ferrier parameterizations. Figures valid at 0600 UTC 27 May 2006 (Day 2). Color bar in dBZ.
Figure 12 The sensitivity of three microphysical schemes in KSOK13 to the hail/graupel fall speed and addition of a SKEB scheme, using GEFS/R2 p09 ICs/LBCs. The fields shown are simulated composite reflectivity (a–f), 850-hPa wind (g–l), maximum 10-m wind over 20 min (m–r), and density potential temperature perturbation (s–x), valid 0300 UTC 16 August 2013 (Day 2). Colors and units denoted by legend. Members without SKEB (i.e., p09-MXMP) are on the left of each group of six panels; those with SKEB (i.e., p09-STMX) are on the right. The three microphysical schemes are WSM6 Graupel (top rows of each panel), WSM6 Hail (middle rows), and WDM5 (bottom rows). Note the simulated reflectivity MXMP members (a, c, e) are also shown in Fig. 9.
Figure 13 Observed (panel a) and simulated (b–l) composite reflectivity for p09-Thompson-STCH ensemble members, valid 0300 UTC 16 August 2013 (KSOK13, Day 2).
Figure 14 Observed (panel a) and simulated (b–l) composite reflectivity for p09-MorrisonHail-STCH ensemble members, valid 0300 UTC 16 August 2013 (KSOK13, Day 2).
Figure 15  Domain-integrated Difference Total Energy for multiple experiments, labeled in the top left in descending order of uncertainty at 0600 UTC on Day 2, for the NEKS06 case. Colors roughly follow those used in Fig. 5. Day and hour shown along x-axis in calendar-day/UTC-hour format; y-axis scale is in units of $m^2 \cdot s^{-2} \times 10^8$. 
Figure 16  As Fig. 15, but for the KSOK13 case.
Figure 17  Difference Total Energy, integrated vertically, at 0000 UTC, 16 August 2013 (KSOK13, Day 2), for multiple experiments: (a) GEFS/R2 members of ICBC-Thompson, (b) p09-STMX, (c) NAM-MXMP, (d) p09-MXMP, (e) p09-MorrisonHail-STCH, and (f) p09-Thompson-STCH. Panels in descending order (a–f) of domain-integrated Difference Total Energy at this time (cf. Fig. 16). Scale is in units of m² s⁻².
Supplementary material

Figure S1  Observed (panel a) and simulated (bl) composite reflectivity for GEFS/R2 members of ICBC-Thompson, valid 2100 UTC 15 August 2013 (KSOK13, Day 1).
Figure S2  Observed (panel a) and simulated (bl) composite reflectivity for GEFS/R2 members of ICBC-Thompson, valid 0000 UTC 15 August 2013 (KSOK13, Day 2).
Figure S3 Observed (panel a) and simulated (bm) composite reflectivity for NAM-MXMP members, valid 1200 UTC 15 August 2013 ( KSOK13, Day 1).
Figure S4  Observed (panel a) and simulated (bm) composite reflectivity for NAM-MXMP members, valid 2100 UTC 15 August 2013 (KSOK13, Day 1).
Figure S5 Observed (panel a) and simulated (bl) composite reflectivity for NAM-WDM6Graupel-STCH members, valid 0500 UTC 27 May 2006 (NEKS06, Day 2).
ADAPTING THE SAL METHOD TO EVALUATE MODEL
REFLECTIVITY FORECASTS OF SUMMER PRECIPITATION IN THE
CENTRAL UNITED STATES

A paper submitted to Atmospheric Science Letters

John Lawson and William A. Gallus, Jr.

Abstract

The Structure Amplitude Location (SAL) method was originally developed to evaluate forecast accumulated precipitation fields through identification and comparison of objects in both the forecast and observed fields. The present study describes a small modification for use with instantaneous composite reflectivity forecasts, where object-size and reflectivity-magnitude minima are prescribed. Both SAL methods are used to evaluate daily 0000 UTC 12-km North American Model forecasts during the summer of 2015 for a central United States 4-km domain. The results show considerable differences between the two methods’ results. SAL using reflectivity reveals a diurnal cycle of skill, with minimum skill occurring around 1800–2200 UTC (early to late afternoon local time), and maximum skill occurring around 1000 UTC (just before sunrise). Results show substantial sensitivity to the reflectivity threshold. This is likely related to sampling more signal from convective cell cores, and progressively ignoring stratiform rain areas, as threshold increases. Setting the threshold too high (40 dBZ) yields only 7% of time periods on which error scores can be computed, as opposed to 94% using
a low threshold (5 dBZ). We conclude that both methods yield useful results, but their results may not generalizable to other domains or techniques.

1 Introduction

The Structure Amplitude Location (SAL) method (Wernli et al. 2008; hereby WP08) evaluates accumulated precipitation fields by identifying objects in both a forecast and observed field, and decomposing differences (i.e., error) into three components. This procedure avoids a double penalization for timing and locational errors inherent in verification methods such as root-mean-square error. The errors are normalized by the size of the domain and domain-wide accumulation such that many cases using the same grid can be compared.

The power of SAL also lies in its ability to evaluate the type of error. The structural component $S$ considers the gradient around the object and its size. For instance, a negative $S$ component may indicate, e.g., too high a radial gradient of the forecasted objects, such as a forecast of convective cells when a stratiform area is observed. The amplitude component $A$ considers domain-wide accumulation. Finally, the location component $L$ consists of two parts. One part ($L_1$) measures location differences in centers of mass for the domain-wide observed and forecast fields; the other part ($L_2$) accounts for location differences of all objects weighted by their integrated precipitation. However, as with all so-called objective schemes, there is a subjective element. SAL scores may on occasion be highly sensitive to the choice of minimum threshold (termed the camel effect in W08), which occurs when a bimodal distribution of precipitation may or may not be split into two objects rather than one.

SAL has shown its flexibility in a previous study, where a potential vorticity anomaly component replaced $S$ (Madonna et al. 2015). Given the connection between radar reflectivity and precipitation accumulation, the authors have refactored SAL for use
with composite reflectivity for evaluating ensemble forecasts of mesoscale convective systems (MCSs) in the Great Plains of the United States. The present paper addresses the differences between the original SAL formulation (hereby SALacpc) and its modification for evaluating instantaneous composite reflectivity (hereby SALcref) in section 2. To compare the two SAL methods over an extended period, North American Model (NAM) forecasts of composite reflectivity and 24-h accumulated precipitation were verified over the central United States with radar observations and multisensor estimates of precipitation, respectively. The method and data sources are detailed in section 3, and results are presented and analyzed in section 4. We discuss interpretation of SALcref, along with concluding statements, in section 5.

2 SAL modification

The SALcref method has three modified aspects: (1) instantaneous composite reflectivity is used instead of accumulated precipitation; (2) the minimum area of the object (herein termed its footprint) is specified; and (3) the minimum object threshold is explicitly set in dBZ.

SALacpc deals with a smoothed field (for instance, precipitation accumulated where frontal systems have traversed), whereas SALcref is an instantaneous snapshot of the reflectivity field. The noisier nature of a reflectivity field hence necessitates a minimum footprint. Smaller footprint and threshold parameters yield more, smaller objects in SALcref than in SALacpc. Note the increased likelihood of multiple objects in at least one dataset means the \( L \) component is more likely to be larger (due to a non-zero \( L_2 \) component), and increases the potential frequency of the camel effect. As \( S \) is computed with the average of all objects’ scaled volume, a large structural error will occur if, e.g., observations have a quasi-linear convective system represented by many strong convective cores joined by weak stratiform rain (e.g. in the trailing stratiform mode).
appearing as one object, but if the simulated field has less stratiform precipitation and appears as numerous cell objects.

In addition, it is desirable that the footprint be large enough to ignore radar clutter, but not large enough that growing convective cells suddenly appear as objects between forecast hours and cause a large step increase or decrease in SAL error. Preliminary testing with object identification on a 4-km grid found a footprint of 200 gridpoints (3200 km$^2$) subjectively related best to the field, with little variation when the footprint was varied between 100 and 500 grid points (1600 and 8000 km$^2$). Hence, all $\text{SAL}_{\text{cref}}$ computations in the present study use a footprint of 200 gridpoints (3200 km$^2$). An increase in threshold should not degrade the quality of the component scores. As $S$ is computed using a weighted mean, the splitting of a larger object into its convective cores would not detract from the component's rationale. Note the $A$ component is not sensitive to threshold.

3 Climatology method

We evaluated the NAM model using 0000 UTC initialisations between 1 April 2015 and 31 August 2015 inclusive, verifying 24-h precipitation accumulation between 12 and 36 forecast hours, and composite reflectivity forecasts valid hourly between 12 and 35 forecast hours inclusive.

A 4-km grid was arbitrarily defined inside the continental United States (Fig. 1) as a common grid to which observations and forecast data were interpolated. This is a similar method to W08. We may expect a forecast model in which convection is parameterized, and interpolated to a finer grid, to develop objects too flat ($A > 0$). This may be perceived a limitation of our methodology, but serves as a check for our tests.

Verification of forecasted reflectivity was performed with composite NEXRAD Level III radar reflectivity from archives at the Iowa State University (https://mesonet.
agron.iastate.edu/docs/nexrad_composites/, accessed 1 February 2016). Base Reflectivity product data are composited through the GEMPAK program *nex2img*, after which suspected false echoes are removed through comparison with the Net Echo Top product. Gridded precipitation accumulation datasets (NCEP/EMC 4-km Stage IV), created using rain gauge and radar observations, were obtained from the Earth Observing Laboratory (http://data.eol.ucar.edu/).

Six of the 153 days in our period had missing archived NAM forecast data, and one other day had missing accumulated precipitation data. These days were removed, leaving 146 days for SAL_{acpc} and 147 days for SAL_{cref}. We ran SAL_{cref} for four thresholds: 5, 15, 30, and 40 dBZ. As the SAL methodology requires identification of at least one object in both the observed and forecast fields, times or periods which resulted in spurious SAL-component scores (i.e., exactly 0, −2, or 2) were removed. This did not affect the number of SAL_{acpc} days, but reduced the 3672 instances of composite reflectivity to 3458, 3408, 2560, and 254 times, for 5, 15, 30, and 40 dBZ thresholds, respectively. Note the number of ignored instances of composite reflectivity was sensitive to threshold because higher thresholds eliminated more areas of reflectivity. This is particularly drastic for the 40 dBZ threshold, where only 7% of times contained objects in both forecast and observational datasets.

4 Results

4.1 Accumulated precipitation (SAL_{acpc})

We found that NAM forecasts only weakly overestimate accumulated precipitation on the majority of days in the dataset (Fig. 2). Objects were too flat, which is likely related to the 12-km horizontal resolution of the NAM forecasts. The positive correlation between $S$ and $A$ components is unsurprising, as discussed in W08, due to the physical
relationship between larger objects and larger domain-wide accumulation. There is no obvious relationship between these two components and $L$-component error.

### 4.2 Composite reflectivity ($\text{SAL}_{\text{cref}}$)

Due to the high volume of data, we focus on results from 30 forecast hours (0600 UTC) as an example here. This is around the time of maximum thunder occurrence for summer months in the central United States (Easterling and Robinson 1985). At lower thresholds (5 and 15 dBZ; Fig. 3a and b), the positive correlation between $S$ and $A$ components is similar to that for $\text{SAL}_{\text{acpc}}$ (Fig. 2). However, at 30 dBZ (Fig. 3c), the line of best fit (not shown) is more parallel with the x-axis: the $A$ component error remains positive regardless of $S$ error for almost all points. This suggests the higher signal ratio from convective cells over stratiform precipitation results in positive $A$ error due not simply to the size of objects, but from their radial gradients from the center(s) of mass. As discussed in W08, positive $S$ and negative $A$ (bottom right quadrant) can occur when a forecast misses an observed convective cell. We note the $S$–$A$ relationships shown here, valid at 30 forecast hours for all four thresholds, are consistent throughout the whole 24-h period. We also find the variation in all three components, represented by the ‘spread’ of points and their colors in Fig. 3, is smallest around 24 forecast hours (0000 UTC) and largest between 30 and 36 forecast hours (0600 and 1200 UTC; not shown). In other words, systematic errors in simulated composite reflectivity dominate SAL statistics in the early night period, while random errors dominate towards sunrise.

Median $S$ component is similar (0.6–0.7) in both $\text{SAL}_{\text{cref}}$ and $\text{SAL}_{\text{acpc}}$, but only when the threshold of the former is set at 15 dBZ or higher. However, despite the larger $S$ component using 5 dBZ threshold (Fig. 3a), the $A$ component is around the same as at other thresholds ($\sim$0.5). This suggests that weak stratiform ($<15$ dBZ) precipitation areas are forecast too large in areal coverage and too weak in magnitude.
When median $S$ and $A$ components are plotted for each hour over the whole dataset, we see a diurnal loop tilted positively (Fig. 4). Note 40 dBZ is not discussed here due to its small sample size and noisy nature. For all three thresholds (5, 15, and 30 dBZ), the $S$ and $A$ components increase from 1200 UTC; the trajectories reach their maximum $A$ error at 1800 UTC and maximum $S$ error at 2200 UTC (2100 UTC for 30 dBZ). While diurnal differences in $S$ and $A$ components are $O(0.5)$, $L$ differences are an order of magnitude smaller (not shown). The $L$ component varies similarly to the other two: all thresholds have their largest $L$-component error at 1800 UTC, and this error reduces through the course of the night. As with $S$ and $A$, values of $L$ also retreat towards the origin (i.e., better forecasts) with increasing threshold; minima occur around 1000 UTC. Note the three trajectories terminate slightly further from the origin than their initial points (i.e. 23 h previously), representing the decrease of forecast skill with time.

These times of maximum $S$- and $A$-/$L$-component errors are around 1 pm and 5 pm, respectively, for most of the domain local time (Central Daylight Time; UTC-5). The earlier peak in $A$ may represent forecasted cell initiation that grows too quickly, while the later peak in $S$ may be related to upscale growth (forecast reflectivity objects are too stratiform). The progression of the trajectories towards the origin suggests an increase in forecast skill towards and after sunset, as diurnal convection decays and nocturnal systems develop. This signal that nocturnal systems are better forecast may be due to the propensity of mesoscale convective systems to occur at night (Markowski and Richardson 2010, and refs. therein), whose length scales are larger than (typically daytime) single-cell convection, and whose predictability is therefore theoretically larger (Lorenz 1969; Clark et al. 2007). Yan and Gallus (submitted to Monthly Weather Review) found NAM forecasts of precipitation forecasts to be more skillful between midnight and early morning, and least skillful near noon. While this corroborates results presented here, we note that our location error ($L$) is an order of magnitude smaller.
than structural ($S$) and amplitudinal ($A$) error, whereas displacement error was the main source of low forecast skill in Yan and Gallus. However, as SAL components are normalised, we expect $L$ to be small due to our large domain size.

The total of absolute SAL component values (taSAL) at each time or day allows an estimate of forecast skill. The median taSAL values at each forecast time are shown in Fig. 5 for 5, 15, and 30 dBZ reflectivity thresholds. The decrease of error with increased dBZ threshold may be related to a smaller area of variation in scaled volume to occur within each object, and objects restricted to convective cores, both of which lower the ‘area of freedom’ for potential dBZ values. Surprisingly, given that stratiform precipitation is increasingly ignored with larger thresholds associated with lower error, more power in the SAL signal is given to less predictable convective precipitation.

## 5 Conclusions

The present study presented modifications to the original SAL methodology (SAL_{acpc}) to verify composite reflectivity fields (SAL_{cref}), instead of accumulated precipitation. We evaluated NAM forecasts for a summer (April–August inclusive) season in the central United States with both SAL methods to gauge the impact of our modifications. The two methods draw different conclusions from their respective fields. SAL_{cref} demonstrated a larger positive $S$ component error, likely related to the inability of the convection-parameterizing model to resolve peak maxima associated with convective cells. The positive correlation between $S$ and $A$ components is expected due to the physical relationship between object size and domain-wide composite reflectivity. However, this correlation is not apparent when the minimum threshold of SAL_{cref} is raised to 30 dBZ. SAL_{cref} reveals a diurnal cycle of skill, with forecasts best in the early morning and worst around noon, and with similar patterns in all three SAL components.
These results show the need to set a threshold and footprint small enough to give a sufficient sample size, but large enough to capture the signal of interest—be it convective or stratiform in nature. Our results also reiterate that interpretation of SAL must be isolated to the SAL configuration and field chosen. For instance, use of a large domain reduces the impact of the $L$ component as object displacements are normalized by the diagonal length. A remaining limitation of the SAL method relating to moist convection is its inability to measure error in orientation of objects. The authors are aware of another object-based evaluation systems (the Model Evaluation Tools; http://www.dtcenter.org/met/users/) which considers orientation, but lacks some simplicity and portability of SAL. A fourth component that considers the mode and orientation of convection may improve SAL’s utility for reflectivity fields.

**Acknowledgments**

All statistical processing was performed with the *NumPy* Python package, and figures were produced with the *matplotlib* Python package.
Figure 1  Domain defined for the present study (dark red line), to which radar and NAM forecast data are interpolated.
Figure 2  SAL scores of 24-h accumulated precipitation, valid at 36 h forecast time (i.e., 1200 UTC), for all days in summer 2015. Scatter points involving spurious scores, resulting from insufficient precipitation during the period, are not shown. Each scatter point is colored by its Location component (see inset). The white box spans the middle two quartiles of Structural (x-axis) and Amplitude (y-axis) components. Dotted line denotes the median Structural and Amplitude component score.
Figure 3  As Fig. 2, but for hourly composite reflectivity SAL scores at 30 forecast hours (0600 UTC), for four thresholds: (a) 5 dBZ, (b) 15 dBZ, (c) 30 dBZ, and (d) 40 dBZ. As in Fig. 2, scatter points are omitted if they contain a spurious value.
Figure 4  Trajectories of SAL through Structure–Amplitude component space for the full 24-h period (12–36 forecast hours) for three thresholds: 5 dBZ (orange), 15 dBZ (light blue), and 30 dBZ (green). The colored square denotes the start of the trajectory (i.e., 12 forecast hours; 1200 UTC). The line increases in alpha (decreases in transparency) as time progresses; each line’s terminus denotes 36 forecast hours (i.e., 1200 UTC the next day). Note the 40 dBZ threshold was omitted for clarity due to its noisy nature, and the plot axes are zoomed into the top right quadrant from Figs. 2 and 3 to show detail.
Figure 5  Total absolute SAL error as a function of NAM forecast hour, for three thresholds: 5 dBZ (orange), 15 dBZ (light blue), and 30 dBZ (green). Note that taSAL ranges from 0 to 6; the y-axis has been magnified for detail. The vertical black line marks forecast-hour 24 (i.e., 0000 UTC).
ON THE SENSITIVITY OF A BOW ECHO TO HORIZONTAL GRID SPACING IN A CONVECTION-ALLOWING ENSEMBLE

A paper submitted to Monthly Weather Review

John Lawson and William A. Gallus, Jr.

Abstract

The bow echo, a mesoscale convective system responsible for much hail and wind damage across the United States, is associated with poor skill in forecasts of convective mode. Given the increase in grid resolution within many operational forecasting systems, we investigate the effect of finer resolution on ensemble spread and the character of bow-echo development. Two ensemble forecasts were generated using the Weather Research and Forecasting model: one used a single domain with 3-km horizontal grid spacing, and another nested a 1-km domain inside the 3-km parents domain with two-way feedback. Ensemble members were then generated from the control with a stochastic kinetic-energy backscatter scheme, with identical initial and lateral-boundary conditions. Results presented herein show that the increase in grid resolution reduces both spread and skill, as measured with an adaptation of the Structure Amplitude Location method to identify reflectivity objects. The nested ensemble produces a faster bow echo and stronger cold pools. The latter are most likely due to increased (fractal) cloud surface area within the nested ensemble, which allow more entrainment of dry air and hence increased evaporative cooling.
1 Introduction

Within the group of mesoscale convective systems (MCSs), systems that display bowing structures along the convective line are among the most poorly forecast (Snively and Gallus 2014). Borrowing terminology from Johns and Hirt (1987), these bowing segments may occur along a quasi-linear convective system (a line-echo wave pattern or serial bow echo), or the bowing segment may instead have a radius of curvature similar to the scale of the system (progressive bow echo). In Lawson and Gallus (2016; hereby LG16), it was shown that progressive bow echoes were likely poorly forecast due to inherent low predictability, rather than deficiencies in microphysical parameterizations, and that improvements in synoptic- and mesoscale initial and lateral-boundary conditions (ICs and LBCs, respectively) would yield only minor skill increases at best. The diminishing returns from more accurate large-scale ICs and LBCs were discussed by Durran and Weyn (2016), and stem from the small-scale sensitivity to minuscule error on the large scale via downscale error cascade and growth.

Many operational centers use ensemble forecasting systems (or simply ensembles) to both measure and account for uncertainty in the forecast. Ensembles are created by perturbing numerous simulations from a control by, e.g., varying the ICs and LBCs, or using numerous parameterizations (e.g., Stensrud et al. 2000). Regime uncertainty is measured by the difference between the perturbation members (or spread), which represents heightened sensitivity to earlier perturbations (Leutbecher and Palmer 2008). At a given lead time, higher spread is often associated with lower skill (Whitaker and Loughe 1998), and represents lower inherent predictability, and a shorter predictability time horizon (Lorenz 1969; Palmer et al. 2014).

Increase in computer power year-on-year allows operational centers to decrease horizontal grid spacing ($\Delta x$), which in turn allows smaller and smaller phenomena to be explicitly resolved. However, there is conflicting evidence regarding the bene-
fit of higher resolution. Mass et al. (2002) suggest that forecasts with $\Delta x = 4$ km, while appearing more realistic than those with $\Delta x = 12$ km, are only marginally more skillful. Conversely, Potvin and Flora (2015) found improved grid spacing may be beneficial even with coarsened or inferior ICs. Many studies have found a smaller $\Delta x$ may systematically change MCS characteristics exhibited at a larger $\Delta x$; for example, simulations with a smaller $\Delta x$ in Bryan and Morrison (2012) entrained dry mid-level air faster, developed squall lines more rapidly, than simulations with a larger $\Delta x$. Their higher-resolution simulations contained more evaporation due to better resolved turbulence, which led to stronger cold pools. Furthermore, Lebo and Morrison (2015) found entrainment and detrainment was suppressed in simulations with $\Delta x$ larger than 500 m. Crucially, Johnson et al. (2013) found the biggest impact of increased resolution is on the smallest resolved processes. As LG16 found the largest uncertainty was associated with small length scales near MCSs, this motivates us to ask: how sensitive is ensemble skill and spread, and the structure and speed of bow echoes, to horizontal grid spacing?

Tennekes (1978) showed with two- and three-dimensional vorticity equations (e.g., synoptic and convective scales, respectively) that smaller-scale noise (i.e., random error) dominates in three dimensions, and the finer the grid resolution, the more the noise impacts the simulation. In convection-allowing ensemble simulations of developing MCSs, LG16 found that ensemble spread at 24–36 h was largest geographically near the MCS of interest, and was associated with uncertainty of the MCS convective mode between ensemble members. Hence, considering these two points, we test whether an ensemble forecast of a bow echo has a wider spread of convective modes when resolution is increased. If true, an increase in ensemble size (and computational power) would be required to correctly sample phase space and gauge uncertainty. Furthermore, an increase in spread indicates a shorter predictability time horizon.
Herein, we investigate the sensitivity of a $\Delta x = 3$ km ensemble to the inclusion of a two-way-feedback nested $\Delta x = 1$ km domain for a progressive bow-echo case. The two ensembles use a single set of ICs and LBCs that yield a bow echo in all simulations. From the control in both ensembles, ten equally likely perturbation members are created with a stochastic kinetic-energy backscatter (SKEB) scheme (Mason and Thomson 1992; Shutts 2005; Berner et al. 2011). The SKEB scheme is chosen to (a) allow use of fixed ICs and LBCs that yield a bow echo in each member, and (b) increase the sampling of phase space, considering the unrepresentative nature of deterministic runs due to the sensitivity of bow echoes to small model errors (LG16). We test the hypothesis that the inclusion of a nested 1-km domain increases the spread of ensemble composite reflectivity, and whether the median nested-ensemble forecast is closer to observations than the single-ensemble median. We also detail systematic changes to the ensemble simulation, namely the bow-echo structure and speed. While our focus on a single case and model configuration precludes general conclusions, it allows a deeper analysis of the physical reasons for systematic sensitivity to $\Delta x$.

Section 2 outlines the data and methodology used herein. Results are presented in section 3 and discussed in section 4.

2 Data and Methods

The progressive bow echo of 15–16 August 2013 brought damaging wind and hail to Kansas, Oklahoma, and Texas. This bow echo developed under northwesterly flow at 500 hPa, downstream of a height ridge; winds became weak and variable in direction towards the surface. At the surface, a weak frontal wave associated with a mean sea-level pressure minimum near the Nebraska–Kansas border moved south, and initiation occurred to the north of this boundary around 2200 UTC on Day 1. Initiation and upscale growth appear to be focused by a mesoscale convec-
tive vortex (MCV) embedded within the frontal wave, as analyzed and discussed by Storm Prediction Center forecasters in Mesoscale Discussions (e.g., archived at http://www.spc.noaa.gov/products/md/2013/md1723.html, accessed 1 March 2016). The moist convection formed a line by 2200 UTC and began bowing at 2330 UTC. The MCS produced a swath of strong wind (up to 34 m s$^{-1}$ or 67 kt) and large hail (up to 4.4 cm or 1.75 in) in central Kansas and near the Oklahoma–Texas border (Fig. 1).

Motivated by the desire to simulate a bow echo in most members (to gauge the impact of $\Delta x$ on reflectivity structures), we used a SKEB scheme to generate perturbations, and chose a IC/LBC dataset and parameterization schemes that performed well in preliminary testing. Our choice of methodology followed the results in LG16 that the position and timing of this bow echo was dominated by IC and LBC perturbations, but the convective mode was still modified by the small SKEB perturbations. Additionally, the need for equally likely members precluded the use of mixed parameterization schemes.

SKEB was developed in response to excessive kinetic energy dissipation near the truncation scale of NWP models (Mason and Thomson 1992; Shutts 2005), and adapted for the Weather Research and Forecasting (WRF; Skamarock et al. 2008) model in Berner et al. (2011). In their large eddy simulations, Mason and Thomson (1992) found a SKEB scheme to have little impact on the energy spectra of turbulent planetary boundary layers (PBLs), such as those in the vicinity of strong moist convection. A similar stochastic scheme has since been developed to target convective processes (Stochastic Convective Backscatter; Shutts 2015) but is not deployed herein. We note that the WRF model, tested as version 3.4.1 in Duda et al. (2016), contains a realistic energy spectrum without implementation of the SKEB scheme available in WRF. This may be due to turbulent PBLs near convection (Shutts 2015) within a convection-allowing simulation. Hence the role of SKEB at convective scales, at least within recent versions of WRF, may be primarily to increase ensemble spread. SKEB
is used in this manner within the present study to perturb atmospheric states from the control.

We use WRF version 3.5 to create our ensemble simulations. To drive these WRF simulations, we chose the p09 member of the 11-member Global Ensemble Forecast System Reforecast dataset (GEFS/R2; Hamill et al. 2013), and Thompson microphysics after good performance in LG16. Other parameterizations are described in Table 1. These ICs, LBCs, and parameterizations were also used in LG16 for the same bow echo (KSOK13). Note the original (Δx = 3 km) domain in LG16 was slightly farther west than the parent domain used herein. This change was made to comfortably nest a 1-km domain within the 3-km domain, and ensure this inner domain captured the bow echo.

The single-domain (SINGLE) ensemble has eleven members: the control (c00) and ten SKEB-perturbation members (s01–s10). Its 3-km domain is labeled as such in Fig. 2. Each perturbation member uses a different randomness seed to generate the backscatter pattern. The two-nest (NESTED) ensemble uses the 3-km parent domain from SINGLE with a nested 1-km domain (labeled in Fig. 2). NESTED also has eleven members: the control with no SKEB scheme (c00h) and ten SKEB-perturbation members (s11–s20). The seeds across both SINGLE and NESTED are unique.

Cold pool strength is depicted by the density potential temperature perturbation field (∇θ′), as in Markowski and Richardson (2010). It is computed by subtracting density potential temperature θρ from the domain mean at each timestep, where

\[ θ_ρ = θ(1 + 0.61r_v - r_h) \]  

and where \( r_v \) and \( r_h \) are the mixing ratios of water vapor and all other hydrometeor species, respectively.

We analyze only the 3-km domains of NESTED in the following sections. The simulations in the overlapping sections of the 3-km and 1-km domains of NESTED are
identical—other than the horizontal sampling resolution—due to two-way feedback. The common 3-km grids then allow use of object-based skill scores to evaluate the ensemble spread and performance. For this, we use a modified version of the Structure Amplitude Location (SAL) method devised by Wernli et al. (2008). The original method was formulated for precipitation fields, and identifies objects (i.e., coherent structures that meet a strength threshold) in both simulations and observations. The simulation is then penalized according to normalized differences as follows:

- Structure ($S$), between –2 and +2. A positive value indicates simulated objects are too large and/or too flat.
- Amplitude ($A$), between –2 and +2. A positive value indicates the simulation has overestimated observed domain-averaged precipitation.
- Location ($L$), between 0 and +2. A positive value indicates a displacement of simulated precipitation objects from those observed.

Our modification uses simulated and observed composite reflectivity instead of accumulated precipitation (further detail can be found in the previous chapter). For verification, we obtained composite reflectivity data from composite NEXRAD Level III radar reflectivity archives at the Iowa State University (https://mesonet.agron.iastate.edu/docs/nexrad_composites/, accessed 1 September 2015). Before reaching the archives, Base Reflectivity product data are composited with the GEMPAK program `nex2img`, after which false echoes are removed after comparison with the Net Echo Top product. Total absolute SAL (taSAL), used as a skill score in the present study, is computed as follows:

$$taSAL = |S| + |A| + |L|$$

and varies between 0 (perfect forecast) and 6. To gauge the skill of each ensemble, and as the ensemble mean of composite reflectivity is an unphysical field, we take the
interpolated ensemble median in place of the ensemble mean. This is done by ranking all ten perturbation members by their taSAL score, and taking the value halfway between the 5th and 6th most skillful members.

Objects are identified by ignoring the reflectivity field below a given amplitude threshold, and in our modification, we include objects only when they comprise a given number of gridpoints that exceed a size threshold (known as its footprint). As SAL was originally designed for a smoothed accumulated precipitation field, suitable footprint and threshold values were tested for instantaneous reflectivity fields. Values of 15–30 dBZ and 100–500 gridpoints (900–4500 km² in area) were relatively robust to small changes in these parameters. Above 30 dBZ and 500 gridpoints, there was substantial sensitivity to changes in threshold and footprint, due to a smaller sample size of objects at a given time and across the simulation period. These high values often created rapid increases in SAL from hour to hour as objects ‘appeared’ as they grew critically large. Conversely, much smaller thresholds and footprints captured too much signal from stratiform precipitation, which detracts from the focus on MCS simulation. Ultimately, the 15 dBZ threshold and 200 gridpoint footprint provided a robust compromise, and is used in the following text to estimate spread and skill in the ensemble simulations.

3 Results

3.1 Sensitivity of structure to $\Delta x$

In SINGLE, convective initiation occurs at the same time as in observations (2000 UTC on Day 1; not shown). Between 2100 UTC on Day 1 and 0600 UTC on Day 2, a spectrum of SINGLE solutions—ranging from a collection of cells to a progressive bow echo—are in contrast to the observed line of convection. The simulated MCSs in SINGLE members generally lag $\sim$75 km behind the observed system in its
southward progression. By 0600 UTC on Day 2, most SINGLE members have captured the progressive bow echo in Oklahoma.

The NESTED ensemble has more agreement by 2100 UTC on Day 1 than in SINGLE, with all members creating a progressive bow echo in a similar location to the observed line. At 0000 UTC on Day 2, all members have a bow echo, but it is too large in the zonal direction. In contrast to SINGLE, the bow echoes simulated in NESTED members are collocated with, or in advance (to the south) of, the observed system throughout the lifetime of the bow echo (2100 UTC–0600 UTC). At 0300 UTC on Day 2, almost all members reproduce the observed system, with the correct radius of curvature, and of the correct size. In general, NESTED ensemble members resemble the observed system better than SINGLE members, but accelerate the bow echo too quickly. The following subsection now analyzes whether this translates to higher NESTED skill using SAL scores. Note the sensitivity of system speed to $\Delta x$ is discussed later.

3.2 Sensitivity of spread and skill to $\Delta x$

Following Tennekes (1978), we may expect more spread within the NESTED ensemble due to higher sensitivity of mesoscale processes to small perturbations when $\Delta x$ is decreased. However, the domain-wide standard deviations of composite reflectivity greater than 15 dBZ (to match the SAL threshold), 10-m wind, and 2-m potential temperature (all not shown), show little difference between SINGLE and NESTED.

We investigate the nature of convection in both ensembles further with SAL scores. Figure 3 shows the median (horizontal line), spread of the middle two quartiles (box), and spread of the whole ensemble (whiskers), for the SINGLE and NESTED ensembles. At each forecast time shown (every 60 min), the SINGLE and NESTED ensembles are compared. The smaller median value is colored green (i.e., a better forecast) while the larger is colored red. Likewise, the larger spread of the middle two quartiles is colored yellow (more variation between ensemble members, ignoring outliers), while
the smaller spread is colored gray. Note that no box and whisker is shown at a given forecast time if there are any members in that ensemble that have spurious taSAL values (e.g., lack of convection; reflectivity values not meeting the size and magnitude threshold for object identification). Additionally, if only one ensemble has SAL values, the colors are black and white.

Figure 3 shows that SINGLE has a lower median (better forecast) than NESTED for 75% of the time periods. SINGLE also has more spread (76% of times, neglecting the first 3 h with little convective activity). The exception to this pattern occurs between 0200 UTC and 0500 UTC on Day 2, inclusive (26 h to 29 h forecast hours). At this time, the bow echo is maturing, and there is good agreement between NESTED members, but not in SINGLE. The poor performance of NESTED before 0200 UTC may be related to its overly aggressive development, and excessive west–east length, of the bow echo. After this time, however, the NESTED ensemble has a lower median until 0500 UTC, matching its subjectively better reflectivity fields. Throughout the entire period, there is little correlation between spread and median (skill) at each hour.

3.3 Sensitivity of system speed to $\Delta x$

We now investigate the difference in development and acceleration of the bow echo, with and without the nested 1-km domain. We show representative members within both ensembles by calculating taSAL for each member, and choosing the median member at 0000 UTC (as the bow echo is reaching maturity in observations). The SINGLE (s06) and NESTED (s12) members are hence discussed in the following section. Figure 4 presents observed and simulated composite reflectivity, and simulated $\theta'_\rho$ to depict the near-surface cold pool, at three times: 21 h, 24 h, and 27 h simulation time. At 21 h, the cold pool is $\sim 50$ km farther south in the NESTED member, but the peak magnitude of $\theta'_\rho$ is similar in both members at this time ($\sim 12$ K). Three hours later, the NESTED member cold pool is substantially more developed in areal coverage
and magnitude, and has progressed farther south. Three hours later still, as the bow echo weakens, there is little difference in magnitude between the two simulations, though the NESTED member leaves a more obvious wake of cold air. Values have also decreased, however, as $\theta'_p$ has a strong diurnal dependence. Note, despite the more distinctive bow-echo structure in the NESTED member, that the bow-echo location in the SINGLE member is closer to that observed.

We now use the same median members to investigate cold-pool development. The movement of a cold pool is related to its strength (perturbation of density or 2-m potential temperature) and hence pressure gradient. Prior to the bow echo development at 1800 UTC on Day 1, the gradient of 2-m potential temperature is similar ($\sim 0.75 \times 10^{-3} \text{K m}^{-1}$) in the SINGLE and NESTED members (Fig. 5a,d). Four hours later, the cold pool has moved farther south in the NESTED member (Fig. 5b,e), marked at its leading edge by larger values of potential-temperature gradient ($\sim 2 \times 10^{-3} \text{K m}^{-1}$) than in the SINGLE member. Four hours later still, there is $\sim 125$ km meridional difference between SINGLE and NESTED cold-pool leading edges (Fig. 5c,f), and the NESTED leading edge is associated with a temperature gradient ($2 \times 10^{-3} \text{K m}^{-1}$) double in magnitude of the gradient in SINGLE. An increased gradient along the cold-pool leading edge is also seen in surface pressure (not shown).

The faster movement in the NESTED simulation is similar to behavior documented by Weisman et al. (1997). In their simulations of QLCSs—with $\Delta x$ ranging from 1 km to 12 km—the higher-resolution simulations better developed a feed of low-$\theta_e$ air. They also found a slower evolution with coarser grids, and that MCSs developing near MCVs (such as in the present study) may be more predictable due to associated dynamical balance. To gauge solely the sensitivity of system speed to resolution, we compare the control members from SINGLE and NESTED; the only difference between the two simulations is the addition of a 1-km nest in NESTED (no SKEB scheme is active in the control members). Figure 6 shows perturbation water-vapor mixing ratio ($q'$) at
800 hPa at three times, for the two control members. The 35 dBZ simulated composite reflectivity contour, smoothed with a 9-km Gaussian filter, is overlaid for reference. At 2000 UTC on Day 1, there is little difference between the two simulations (cf. panel a with d). The rear-inflow jet is conspicuous by drier air behind the burgeoning moist convection. As the system intrudes farther into the region covered by the 1-km nest (cf. panel b with e), the drier air in the NESTED member penetrates farther south, and is associated with a more coherent, bowing segment of high reflectivity. This is even more pronounced by 2200 UTC (cf. panel c with f).

Cross-sections perpendicular to the bow-echo apex are shown in Fig. 7 at 2100 UTC on Day 1 (cf. Fig. 6b and e); winds perpendicular to the cross-section transect are contoured and $q'$ is color-filled. Note the cross-sections (Fig. 7a and c) were averaged 6 km (two grid points) in each direction normal to the cross-section transect to improve representivity. The NESTED cross-section (Fig 7c,d) shows winds in excess of 20 m s$^{-1}$ and low $q'$ air descending and feeding into the rear of the bow echo; this is absent in SINGLE (Fig 7a,b), and matches the latitude–longitude cross-section at 800 hPa in Fig. 6. Taking a similar horizontal slice at 800 hPa in the wind field for both SINGLE and NESTED members (Fig. 8), we find a more coherent rear-inflow jet in NESTED. Whereas strong winds do occur along the bow-echo leading edge in SINGLE (southwest of the cross-section transect), associated with cellular development (cf. Fig 7b), they are rather disconnected from the channel of wind farther north. In fact, there is up to 30 m s$^{-1}$ difference in wind vectors associated with the rear-inflow jet (not shown). In summary, the nested 1-km domain appears to have an enhanced and more coherent rear-inflow jet, which in turn increases evaporational cooling. The resultant cold pool is stronger, and accelerates faster due to increased surface pressure gradients.

But why may the rear-inflow jet be stronger with a smaller $\Delta x$? The perimeter of a two-dimensional fractal object (i.e., infinitely complex regardless of zoom level) is sensitive to the measuring interval (Mandelbrot 1967), and similarly for surface area
of three-dimensional objects. Stronger cold pools in higher-resolution simulations are related to evaporation of cloud water content (Bryan and Morrison 2012). Higher horizontal resolution yields a larger surface area of clouds (which are fractal), and hence more interfacing of dry air and cloud water content. This could lead to increased evaporation in the microphysical parameterizations.

3.4 Sensitivity of system speed to SKEB

The region in the wake of an MCS leading edge is turbulent (Droegemeier and Wilhelmson 1987), represented herein by the descending drier air in Fig. 7; hence, entrainment may increase in SKEB members, as the SKEB scheme increases turbulence through the injection of kinetic energy into resolved scales.

We find the control (no-SKEB) member of NESTED has the weakest rear-inflow jet at 2300 UTC on Day 1 out of all members (seen in 800-hPa mixing-ratio perturbation field; not shown), and the least coherent and slowest-moving bow echo until 0330 UTC. This connection between the rear-inflow jet and bow-echo speed is similar to results in the previous subsection. However, in the SINGLE ensemble, the control member does not have the slowest bow echo. As such, further ensemble simulations are needed to address the link between SKEB perturbations and bow-echo speed, outside the scope of the present study.

4 Summary and Conclusions

Two $\Delta x = 3$ km ensemble simulations of a bow echo, one with a nested 1-km domain, have addressed the hypothesis that a smaller $\Delta x$ increases the spread of an ensemble. An increase of spread in the nested simulation does not occur, as measured by standard deviations of various fields. Further analysis of reflectivity objects via the SAL methodology, in fact, suggests:
• Spread is larger in the single-nest ensemble. This disputes the hypothesis that spread increases as $\Delta x$ decreases.

• Skill is better in the single-nest ensemble, despite the ninefold increase in computer power required to run the nested ensemble.

• While both spread and skill are lower in the single-nest ensemble overall, there is a lack of correlation between the two over the hourly forecast times.

In addition, there is systematic bias for systems to move faster with the addition of the second, finer nest. This faster movement in the nested simulation is associated with a stronger rear-inflow jet and surface-based cold pool. We propose this is related to the fractal nature of clouds and turbulence, as follows: in a higher-resolution simulation, a given cloud object will have larger surface area (i.e., its fractal dimension increases). Within the numerical model, this is represented by a greater number of grid points that describe the division between supersaturated and subsaturated air. More resolved turbulence also increases dry-air entrainment. These two factors increase the interfacing of dry air with cloud water content, increasing evaporation. This strengthens the surface-based cold pool and the corresponding pressure gradient behind and ahead of the MCS’s leading edge. In our simulations, this surges the system $\sim$100–200 km farther south at its mature stage (0300 UTC on Day 2) than in coarser simulations. As $\Delta x$ reduces, we do not find convergence towards the observed system as finer features are resolved. However, only two different grid spacings were used.

In the nested simulation, the control (no-SKEB) member simulates the slowest moving bow echo. This may suggest the SKEB scheme may accelerate a bow echo simulation at $\Delta x = 1$ km due to better resolved momentum fluxes and a more developed rear-inflow jet. However, the single-nest simulation does not follow the same pattern.

While we have addressed only one case, we expect the occurrence of faster bow echoes with decreasing $\Delta x$ to be general due its strong signal in humidity and wind
fields. Furthermore, despite the use of a single microphysics scheme, differences resulting from grid-spacing changes dominate over those from varying microphysics schemes (Bryan and Morrison 2012), hence these findings should be generally sound. We note that simulated winds (output at each timestep) are too light throughout the study compared to NCDC storm reports, which may stem from the chosen PBL scheme (MYNN Level 2.5).

Overall, there is little advantage apparent for the ninefold increase in computer power needed to run three times the horizontal resolution. Notably, ensemble spread is not increased by adding a 1-km domain, hence more members should not be required to maintain a good sampling of a higher resolution ensemble. It remains an open question whether the stronger bow echoes and the lack of improved skill in higher-resolution ensembles is general to other cases. This, and the sensitivity of bow-echo speed to SKEB perturbations, should be the subject of further work.

**Acknowledgments**

The authors thank, in alphabetical order: Ray Arritt, Jeffrey Duda, David Flory, Jonathan Flowerdew, Matthew Flourney, Anna Ghelli, Lee Grenci, William Gutowski, Daryl Herzmann, Makenzie Krocak, Anne McCabe, David Schultz, Richard Swinbank, Andy VanLoocke, Xiaoqing Wu. In addition, comments from two anonymous reviewers were appreciated. All computations were performed with Python packages *numpy* and *python-netcdf4*, and figures were generated with Python packages *matplotlib* and *basemap*. Research supported by grant NSF-AGS1222383.
Table 1  Parameterization schemes used in the numerical modeling configuration.

<table>
<thead>
<tr>
<th>Parameterization</th>
<th>Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microphysics</td>
<td>Thompson</td>
</tr>
<tr>
<td>Longwave Radiation</td>
<td>RRTM</td>
</tr>
<tr>
<td>Shortwave Radiation</td>
<td>Dudhia</td>
</tr>
<tr>
<td>Surface Layer</td>
<td>MYNN</td>
</tr>
<tr>
<td>Land Surface</td>
<td>Noah</td>
</tr>
<tr>
<td>Planetary Boundary Layer</td>
<td>MYNN Level 2.5</td>
</tr>
</tbody>
</table>
Figure 1  Overview of the bow echo of interest: (a) composite radar at three times (2200 UTC Day 1, 0200 UTC Day 2, and 0600 UTC Day 2), and (b) National Climatic Data Center storm data for all wind reports exceeding 25 m s\(^{-1}\) (50 kt).
Figure 2  Domains used in the present study. The SINGLE ensemble uses the 3-km domain; the NESTED ensemble nests the 1-km domain inside the 3-km domain.
Figure 3  Box plot of total absolute SAL (taSAL) for the (a) SINGLE and (b) NESTED ensembles. Features shown include median (linearly interpolated; horizontal line), spread of the middle two quartiles (box), and spread of the whole ensemble (whiskers). The lower median (better forecast) at each time is colored green, while the higher median is red. The larger spread of the middle two quartiles at each time is colored yellow.
Figure 4  Observed composite reflectivity (leftmost column), and simulated fields of composite reflectivity (second and third column from left) and density potential temperature perturbation (two rightmost columns) from total absolute SAL median SINGLE and NESTED members, for KSOK13 case. Fields are valid at 2100 UTC on Day 1 (a–e), 0000 UTC on Day 2 (f–j), and 0300 UTC on Day 2 (k–o). Times are listed on the right as hours since initialization.
Figure 5  Potential temperature at 2 m (color-filled) taken from total absolute SAL me-
dian SINGLE (s06; a–c) and NESTED (s12; d–f) members. Fields shown at (a,d) 18 h , (b,e) 22 h, and (c,f) 26 h simulation time. States shown in each
panel, from north to south, are Nebraska, Kansas, Oklahoma, and (the pan-
handle of) Texas
Figure 6  Perturbation water-vapor mixing ratio at 800 hPa (color-filled) taken from control members of SINGLE (c00; a–c) and NESTED (c00h; d–f) ensembles. Black lines contour the 35 dBZ composite reflectivity field, smoothed with a 9-km Gaussian filter, for reference. Fields shown at 20 h (a,d), 21 h (b,e), and 22 h (c,f) simulation time. States shown in each panel are Nebraska (upper) and Kansas (lower), separated by the thin horizontal black line.
Figure 7  Cross-sections at 2100 UTC on Day 1: (a,c) height against horizontal distance in the water mixing ratio perturbation field overlaid with the wind component parallel to the transect (contoured every 5 m s\(^{-1}\); negative values, marked by dotted lines, indicate a component from right to left on the panel); and (b,d) latitude–longitude in the composite reflectivity field from the SINGLE control (c00; a, b) and NESTED control (c00h; c, d) members. The blue transects in the right panels, between points A and B, mark the path of the cross-sections shown in the left panels. Note the height–distance cross-sections (a, c) were averaged 6 km (two grid points) in each direction normal to the cross-section transect.
Figure 8  The 800-hPa wind field for (a) the SINGLE control (c00) and (b) the NESTED control (c00h) at 2100 UTC on Day 1. Colors indicate wind magnitude (see legend) and vectors indicate both wind direction and magnitude. Vectors are shown every third grid point (i.e., every 9 km) for clarity; contour-fill uses all grid points (i.e., $\Delta x = 3$ km). Cross-section transect from Fig. 7 is shown by a black line in both panels.
THE RELATIONSHIP BETWEEN SYNOPTIC AND STORM-SCALE PREDICTABILITY OF MESOSCALE CONVECTION SYSTEMS

A paper in preparation for submission to TBC

John Lawson, William A. Gallus, Jr., and Makenzie Kroca

Abstract

We may expect the skill of Weather Research and Forecasting (WRF) hindcasts of mesoscale convective systems (MCSs) to be dictated by (a) inherent synoptic-scale predictability, as measured by the spread of the Global Ensemble Forecasting System reforecast (GEFS/R2) dataset, and (b) the skill of the North American Model (NAM) forecast dataset providing initial and lateral-boundary conditions to the hindcast. However, there is no obvious relationship between the accuracy of MCS convective mode and either GEFS/R2 ensemble spread or the skill of the NAM archives. When the MCS dataset is confined to bow echoes, we find that serial bow echoes (i.e., line-echo wave patterns) are better forecast by the WRF simulations than progressive bow echoes. Furthermore, stronger rising motion is linked with the propensity for bow echoes to be serial rather than progressive. This corroborates previous findings that stronger mid-level forcing results in better forecasts due to the larger ratio of initial-condition error to model error. However, there is little relationship between bow-echo forecast skill and a number of other fields, including the local instantaneous Lyapunov exponent at numerous pressure levels. We therefore speculate that the skill of storm-scale forecasts may inherit only limited characteristics of the large-scale predictability, perhaps due to
rapid downscale cascade and growth of initially trivial errors in the initial-condition dataset.

1 Introduction

The atmosphere is partly chaotic, and as such, a deterministic forecast of its future state is susceptible to small changes in the initial condition (IC) dataset (Lorenz 1969). Furthermore, this rapid growth of initial perturbations is compounded within moist convection (Zhang et al. 2003) and an imperfect model. Mesoscale convective systems (MCSs), harbingers of severe hail, wind, floods and tornadoes (Gallus et al. 2008, and refs. therein), therefore pose a problem for forecast centers in both their severity and inherent reduced predictability.

To gauge the scale of this problem, Snively and Gallus (2014, hereby SG14) measured the skill of numerous warm-season MCS events in the central United States with convection-allowing hindcasts. Their skill score rated the accuracy of their simulations in reproducing the observed convective mode and its timing. Bow echoes were among the worst modes, and motivated further research into sources of this low skill (Lawson and Gallus 2016).

The present study delves deeper into the relationship between storm-scale predictability, as measured in SG14, and synoptic-scale predictability. We might expect smaller scales to inherit aspects of IC accuracy from larger scales, especially as synoptic-scale features such as fronts may constrain the range of forecast solutions (Anthes et al. 1985). However, Durran and Gingrich (2014, and refs. therein) suggest that the concept of inheritance is flawed due to the extreme sensitivity of the mesoscale to the synoptic scale. If inheritance of predictability exists, we expect limited skill to occur on the smaller scale when uncertainty in the larger scale state is larger. Uncertainty of the state is measured by the differences between multiple forecasts spawned from
slightly different ICs or perturbed by small changes, and known as ensemble spread (e.g., Leutbecher and Palmer 2008).

Stensrud et al. (2000) state that, when mid-level forcing is strong, IC uncertainty dominates over model error, and vice versa for weak forcing. They also note that skill is higher when IC uncertainty dominates. This is borne out in MCS rainfall forecasts, analyzed by Jankov and Gallus (2004), which performed better when forcing was stronger. While strong upper-level forcing is often associated with large vertical wind shear, Weisman et al. (1997) found that convection-allowing simulations were less accurate when vertical wind shear increased. In balance, we hypothesize that the skill of bow-echo forecasts is correlated with synoptic forcing, as defined by 500-hPa rising motion (Stensrud et al. 2000). We also hypothesize that forecast skill of convective mode in general is (1) positively correlated with the driving forecast dataset skill, and (2) negatively correlated with ensemble spread at the synoptic scale. We also analyze potential relationships between multiple fields and bow-echo subtype, as motivated and detailed in section 2. We present results in section 3, with a brief discussion in section 4.

2 Data and Methods

2.1 Datasets

We use the same MCS dataset as SG14, which comprises 37 summer cases in the central United States, chosen from a larger set of cases used by Duda and Gallus (2013). Each event was simulated with the Weather Research and Forecasting (WRF) model with 4 km vertical grid spacing, and used North American Model (NAM) forecast datasets as ICs and lateral boundary conditions (LBCs). All but two hindcasts were initialized at 1200 UTC; the exceptions (29 May 2007 and 12 August 2007) were initialized at 0600 UTC. Each simulation was integrated for 24 h, then given a score on
its accuracy of convective mode and timing (within 3-h periods). This score is referenced herein as \textit{SG14 skill}, and ranges from zero (no skill) to unity (perfect mode and timing). Note that some dates contained two or more events; each event was scored separately in SG14. We point out that the two forecasts initialized 6 h earlier score poorly (0.36 and 0.25), which may be related to the longer forecast time (and hence larger forecast-error-growth potential). However, as these two events do not involve bow echoes and are therefore part of the larger (37-case) dataset, their effect on our general conclusions is likely minimal.

We use ERA-Interim reanalyses from the European Centre for Medium-range Weather Forecasts (ECMWF) as truth on the synoptic scale. These reanalyses were downloaded at 1° by 1° resolution. For mesoscale truth, we used analyses from the Rapid Refresh (RAP) and its predecessor Rapid Update Cycle (RUC). These analyses are at 13 km horizontal grid spacing. To gauge inherent uncertainty (i.e., predictability), we use the spread of the Global Ensemble Forecasting System Reforecasts, version 2 (GEFS/R2). We also downloaded the same NAM forecast dataset used to drive SG14 cases to measure the skill of the hindcast IC/LBC dataset, obtained from the National Operational Model Archive and Distribution System \url{http://www.ncdc.noaa.gov/data-access/model-data/model-datasets}.

To investigate the forecast skill of bow-echo subtypes, we differentiated between serial bow echoes (quasi-linear convective systems with embedded bowing segments, also known as line-echo wave patterns) and progressive bow echoes (bowing segments with a radius similar to the system size), similar to the derecho subclassification in Johns and Hirt (1987). To diagnose the subtypes, we accessed observed composite reflectivity plots at the Mesoscale and Microscale Meteorology Laboratory website \url{http://www2.mmm.ucar.edu/imagearchive/}, accessed 1 March 2016.

We divided the 20 cases in SG14 that contained bow echoes into six serial and eight progressive bow echoes. The remaining six cases were ignored: two had missing radar
data, and four had ambiguous signatures. We then sampled numerous RUC/RAP fields every hour at the location of (a) convective initiation and (b) where the MCS began bowing, both diagnosed from observed composite reflectivity. These fields are detailed in the following subsection.

2.2 Computed fields

Horizontal temperature advection ($T_{adv}$, at 700 and 850 hPa), was chosen for two reasons. First, Johns and Hirt (1987) reported a strong relationship between $T_{adv}$ at these levels and the occurrence of derechos (windstorms associated with a strong subset of bow echoes). Second, while simulating a rapidly deepening extratropical cyclone, Doyle et al. (2014) found high sensitivity of strong winds to perturbations in the moisture and temperature fields within warm conveyor belts. Herein, $T_{adv}$ is computed as:

$$T_{adv} = -\nabla \vec{V}_H \cdot T$$

where $\vec{V}_H$ is the horizontal wind in x- and y-components, and $T$ is drybulb temperature. With similar motivation from Doyle et al. (2014), we also include relative humidity ($RH$) at 700 and 850 hPa.

To motivate our test for the relationship between unstable mid-level flow and mesoscale skill, consider that chaotic flow is typified by the ever-increasing distance between orbits of two particles within the flow. This is measured by the Lyapunov exponent (e.g., Lorenz 1965; Williams 1997; Cohen and Schultz 2005). Now, consider chaotic flow that occurs near a cold front. The ‘attraction’ of stable density-driven circulations near the front constrain the orbits of particles, reducing the bounds of uncertainty of these trajectories. Hence, fronts and other constraints to the flow may increase the predictability of certain regimes and geographical regions (Anthes et al. 1985). Conversely, consider a jet-streak exit region, where acceleration and diffluent
streamlines result in rapid growth between orbit positions. This may be associated with lower skill and higher sensitivity to ICs, for instance in blocking scenarios (Oortwijn 1998). Hence, we use local instantaneous Lyapunov exponent $\lambda$ to estimate dynamical error growth, as defined in Cohen and Schultz (2005):

$$\lambda = 0.5 \times (D + (E^2 - \zeta^2) \times 0.5) \tag{2}$$

where $D$ is horizontal divergence, $E$ is total horizontal deformation, and $\zeta$ is relative vertical vorticity. The advantage of this formulation is its straightforward implementation for gridded datasets.

In addition, to correlate forecast skill with synoptic forcing, we included rising motion at 500 hPa ($\omega_{500}$) and vertical wind shear between 0 and 6 km (hereby ‘shear’).

2.3 Heatmap methodology

For each case that involved a bow echo, we defined two 31-gridpoint by 31-gridpoint latitude–longitude boxes within the RUC/RAP 13-km analyses (i.e., 162 409 km$^2$), centered on the location of both convective initiation and first bowing. These dimensions (403 km in x- and y-directions) were chosen to capture meso-$\alpha$ processes (wavelengths greater than 200 km) that link the meso-$\beta$ and synoptic scales, and to account for error in our subjective location of initiation or bowing. We then computed fields within the boxes from RUC or RAP analyses at every forecast hour for the forecast period. The box at each time was averaged, and these box-averages were normalized over all cases for each field to plot heatmaps shown in the following section. These are shown relative to the time of either convective initiation or first bowing, as appropriate. Preliminary testing showed that sensitivity of a given field to box size, within an order of magnitude of our chosen box size, was negligible.

For all MCSs in the SG14 case set, we interpolated GEFS/R2, NAM, and ECMWF ERA-Interim datasets to the same 1° by 1° domain (Fig. 1). For this domain, we then
computed the following, each using 500-hPa geopotential height: (a) GEFS/R2 spread measured as domain-average standard deviation of the height field in perturbation ensemble members every 3 h; (b) NAM error (skill) by computing the absolute difference in height between ECMWF ERA-Interim reanalyses and NAM forecasts every 12 h, and (c) GEFS/R2 mean skill by computing the absolute difference in height between ECMWF Era-Interim reanalyses and the mean of GEFS/R2 perturbation members. Preliminary testing showed that use of multiple pressure levels between 100 and 1000 hPa yielded similar results to those using 500 hPa (shown herein).

3 Results

3.1 Bow-echo cases

When SG14 bow-echo cases were divided into progressive and serial subtypes, SG14 skill scores for serial cases averaged 0.63, while progressive cases averaged 0.38. In addition, the progressive cases tended to have timing issues (SG14). Hence, serial bow echoes appear to be more predictable than progressive bow echoes; this is expected due to their larger length scale.

There was little relationship between average $\lambda$ and SG14 skill, other than a weak tendency for 300-hPa $\lambda$ to be higher in the 12 h before bowing for progressive cases (Fig. 2). When maximum $\lambda$ across the box at was calculated, there was no correlation (not shown). But there is a strong link between negative $\omega_{500}$ (dark orange; Fig. 3) and the serial bow echo subtype in the 6 h after convective initiation. Conversely, progressive bow echoes in our case set were often associated with 500-hPa geopotential-height ridges just upstream of the Plains, and attendant weak rising motion. As serial cases often occur ahead of strong cold fronts (Johns and Hirt 1987), the connection between serial bow echoes, stronger forcing, and synoptic setup is not surprising.
There is no apparent link between SG14 skill and either RH or $T_{adv}$ (not shown). We may expect systems that develop within higher shear to be poorly forecasted (Weisman et al. 1997). However, in Fig. 4, the worst three cases (all progressive) have low shear throughout the period near, and at the location of, first bowing. Moreover, there is a similar, but weaker, signal when considering location and time of convective initiation (not shown).

3.2 All cases

For all cases, we now analyze the relationship of the SG14 skill and synoptic-scale properties. We find no correlation between SG14 skill and either GEFS/R2 spread or NAM skill, at any pressure level (Fig. 5). This is surprising, as we hypothesized that better-forecast cases would occur within low-spread synoptic-regimes, and that more accurate ICs and LBCs would improve the forecast of convective mode. However, there is a positive correlation between the spread of the GEFS/R2 ensemble and the error of the NAM forecasts (Fig. 6a), the errors of the NAM and GEFS/R2 mean (Fig. 6b), and the spread and ensemble-mean-error of the GEFS/R2 (Fig. 6c). This positive correlation between spread and error is well known (Whitaker and Loughe 1998), and confirms for our dataset that large-scale uncertainty tends to yield a worse forecast at synoptic scales. However, as seen in the heatmaps, this does not translate to storm-scale skill.

4 Discussion

This paper has presented relationships between the forecast skill of convective mode, meso-$\alpha$ subsets of upper-level fields, and synoptic-scale predictability. Our investigation is based on the hypothesis that bow-echo skill is correlated to large-scale
forcing, and that skill in forecasting the MCS mode is associated with skill in the IC/LBC dataset and/or low inherent uncertainty in the synoptic flow.

We find that serial bow echoes tend to occur when mid-level forcing at the time and location of convective initiation is strong (i.e., IC uncertainty dominates over model error), and are forecast better. Conversely, progressive echoes tend to occur when forcing is weak (i.e., model error dominates over IC error), and are forecast worse. This relationship is observed in upper-level geopotential-height charts (not shown), where progressive bow echoes typically developed under ridges, whereas serial bow echoes formed ahead of strong vorticity maxima. This is in contrast to (Guastini and Bosart 2016), who found both serial and progressive derechos were associated with upper-level troughs, although the overlap between bow-echo and (the more severe) derecho case sets is unknown. Little or no correlation was observed between the forecast skill of bow-echo development and mid-level humidity, temperature advection, vertical wind shear, and Lyapunov exponent. Hence, while some mid-level parameters may be associated with bow echoes and poor skill scores (SG14), these parameters do not distinguish between bow-echo subtype.

We hypothesized that higher convective mode skill was a result of large-scale predictability and increased skill of the forecast dataset used as ICs and LBCs for hindcasts. However, no relationship between the SG14 skill score and either GEFS/R2 spread or NAM skill was found, suggesting that the storm-scale predictability of bow echoes is not related to synoptic-scale predictability. This follows Durran and Weyn (2016), who suggest that the concept of predictability inheritance from large to small scales requires reconsideration due to rapid growth and downscale cascade of errors that start at the large scales.

The relationships between upper-level fields and bow-echo cases, while weak, are only predictors of bow-echo subtype and not the skill score. For example, the serial subtype is related to the stronger synoptic forcing; serial bow echoes are more likely
to have a higher skill score because the predictability is higher (i.e., the deterministic forecast in SG14 is choosing from a smaller range of outcomes). However, the better forecast progressive bow echoes are not associated with stronger rising motion. In summary, it appears that some aspects of convection, such as convective initiation, are not connected to the synoptic-scale predictability (Duda and Gallus 2013); conversely, other aspects of MCS simulations, such as accumulated precipitation and skill of upscale-growth forecasts, are associated with large-scale forcing (Jankov and Gallus 2004; Duda and Gallus 2013). Our results show that rising motion is not a predictor of bow-echo forecast skill *per se*, but is associated with the propensity of the serial bow-echo subtype.

It is unclear whether the lack of relationship between computed fields, skill scores, and bow-echo subtype is a true reflection of the atmosphere, or simply an artifact of our averaging method and box location. Our method may be improved through use of object-based schemes that identifies $T_{adv}$ and $\lambda$ maxima and structures. Such a scheme was outlined in Madonna et al. (2015), who identified and scored warm conveyor belts with a modification to the Structure Amplitude Location method (Wernli et al. 2008), in addition to schemes within the Model Evaluation Tools (http://www.dtcenter.org/met/users/).

**Acknowledgments**

All computations herein were performed with the Python package *numpy*, and plotted with the Python package *matplotlib*. This work was supported by NSF grant AGS1222383.
Figure 1  Domain to which GEFS/R2, NAM, and ECMWF ERA-Interim data are interpolated (red box).
Figure 2  Heatmap showing average local instantaneous Lyapunov exponent $\lambda$ at 300 hPa, normalized for the entire dataset, (a) relative to initiation time (x-axis) and at the location of convective initiation, and (b) relative to time of first bowing, at the location it occurred. Y-axis is sorted in descending SG14 skill, top-to-bottom. Dark blue indicates high values of $\lambda$. Thick black lines indicate the time of (a) initiation or (b) first bowing, for reference.
Figure 3  Heatmap showing rising motion at 500 hPa ($\omega_{500}$) at the location of convective initiation of each case, and relative to the time of initiation (indicated by vertical black lines). Data are normalized over the entire case set such that dark purple indicates strong rising motion (large negative values of $\omega_{500}$). Y-axis is sorted in descending SG14 skill, top-to-bottom.
Figure 4  As Fig. 3, but for vertical shear relative to time and location of first bowing. Dark grays indicates high shear.
Figure 5  Heatmaps showing the relationship between SG14 skill and: (a) NAM forecast skill every 12 h, and (b) GEFS/R2 spread every 3 h. Values on the heatmap are normalized over the entire map, where white is zero and dark purple is unity.
Figure 6  Correlations at 12, 24, and 36 h, as colored in the legend, between (a) NAM skill and GEFS/R2 spread, (b) GEFS/R2 mean skill and NAM skill, and (c) GEFS/R2 mean skill and GEFS/R2 spread. Each scatter point represents a SG14 case.
GENERAL CONCLUSIONS

This dissertation has addressed the poor forecasts of bow echoes with ensemble simulations. Key questions outlined in the first chapter have been addressed as follows:

- **Can SKEB schemes, whose perturbations act on wavelengths much longer than \( \Delta x \), be used to create ensemble perturbations in our convection-allowing simulations?** We find that the mode of the MCS is highly sensitive to SKEB perturbations. Even as domain-wide uncertainty is limited with fixed ICs, LBCs, and parameterizations, uncertainty associated with the MCS mode is still substantial. Moreover, SKEB decorrelation time can be tuned to increase variance of the ensemble without obviously degrading the forecast quality.

- **How does the spread of convective mode change with the general ensemble spread, and the method of perturbation generation?** In the first paper, we found that the location and timing of convection varied substantially with IC/LBC dataset, but the convective mode (progressive bow echo) was less uncertain. However, when ensemble uncertainty of timing and location was smaller (with fixed ICs and LBCs), the mode still varied considerably. Traditional skill scores may not capture error stemming from incorrect mode, which is important for forecasting the hazard associated with the MCS.

- **Bow echoes are the worst forecast convective mode. Is this due to low inherent predictability, poor IC/LBCs, poor parameterizations?** As discussed in the first paper, it is unlikely that ICs and LBCs are responsible for poor forecasts, as we observed diminishing returns on convective mode as IC and LBC error
was reduced. Furthermore, all microphysical schemes performed well, when an optimum IC/LBC dataset was chosen. In addition, subjectively poor ensemble members using a given microphysics scheme could be improved within a SKEB ensemble (and vice versa for subjectively good members). This all suggests that predictability is inherently poor for the bow echoes chosen.

- **How might grid resolution affect the spread and performance of an ensemble?** Using an updated object-based scoring scheme outlined in the third paper, we found a smaller grid spacing ($\Delta x = 1$ km) was associated with less spread, less skill, and faster acceleration of the progressive bow echo than the ensemble with $\Delta x = 3$ km.

- **Considering the fast up- and down-scale growth of errors, how linked are the storm and synoptic scales in MCS forecasts?** We found in the fourth paper that more skillful and predictable synoptic regimes were not correlated with the skill of MCS mode forecasts.

To summarize, the poor forecasts of bow echoes in SG14 are likely due to an upper bound on bow echo forecasts: their predictability horizon. Results from the first and fourth paper corroborate findings in Durran and Weyn (2016) that the error cascade and growth means that (1) improvements in ICs and LBCs yield diminishing returns, and (2) synoptic skill is not necessarily inherited by the mesoscale.

This implies that those calibrating ensembles should consider not only large-scale skill, but also the spread and skill of convective mode. The orientation and structure of convection is important to determine the associated hazards, and as shown herein, these characteristics may be poorly forecast despite good large-scale skill and predictability. Furthermore, our evidence that a decrease in $\Delta x$ is not a guarantee of better skill, and may result in a bias in rear-inflow jet strength, demands additional caution when increasing resolution of a convection-allowing ensemble system.
The smallest butterflies, mentioned by Lorenz in his famous talk, may determine tipping points such as whether a line of thunderstorms develops bowing segments or not. However, it is the largest butterflies, whose wingflaps are initially minuscule in context, that are likely responsible for poor forecasts of bow echoes.

Future work

Our focus on reflectivity and convective mode is subjective in nature. Future work may involve analysis of kinetic-energy spectra and development of further objective schemes. Moreover, these findings mined from a small number of cases should be tested for generality across multiple cases.
REFERENCES


Richardson, L. F., 1922: *Weather prediction by numerical process*. Cambridge Univ. Press.


