Sensitivity of bow-echo forecasts to ensemble and model configuration

John Lawson *

Department of Geological and Atmospheric Sciences, Iowa State University, United States of America.

William A. Gallus, Jr.

* *Corresponding author address:* Dept., Institution, Address, City, State/Country.

E-mail: johnroblawson@gmail.com

ABSTRACT

Bow-echo structures, both embedded within quasi-linear convective systems and as singular systems, are often poorly forecast within deterministic numerical weather prediction model simulations. However, the spread of bow-echo structures within an ensemble of model runs is not known. This study assesses the inter-ensemble-member sensitivity of the structures' simulated reflectivity and radius of curvature to the following: perturbations in initial conditions using a global dataset; different microphysical schemes; and in response to the use of a stochastic kinetic-energy backscatter (SKEB) scheme. It is found that the ensemble spread decreases, respectively, through the three aforementioned ensemble types. Interestingly, a poor deterministic forecast using a given microphysical scheme and a given set of initial conditions can be somewhat improved in some SKEB ensemble members, suggesting that model error needs to be better accounted for when forecasting these mesoscale phenomena, and that one-shot evaluations of parameterization schemes are dangerous.

1. Introduction

Mesoscale convective systems (MCSs) are groups of thunderstorms of length $O(100\,\mathrm{km})$ in at least one direction (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, 1–2 in hail, and flash flooding (Gallus et al. 2008). Conspicuous by their convex structure in radar reflectivity (Fig. **??**), bow echoes and line-echo wave patterns (LEWPs) are associated with some of the strongest wind events in the Plains, sometimes meeting derecho (damaging straight-line wind) criteria (Johns and Hirt 1987). A bowing structure develops when stratiform precipitation behind a quasi-linear convective system (QLCS) lowers a rear-inflow jet through evaporative cooling and consequent negative buoyancy (Markowski and Richardson 2010). This jet advects the cold pool faster in its centre, creating the distinctive bowing shape.

Bow echoes in particular can be poorly simulated by numerical model forecasts (**?**). In that study, weaker 0–6 km shear and higher potential temperatures aloft were responsible for deterministic forecast failures in numerical simulations. They also surmised that simulations involving elevated convection may have been amongst the worst. Interestingly, the study found evidence that convective morphology was not substantially sensitive to mesoscale meteorological parameters. In two studies, Adams-Selin and co-authors found that performance of numerical simulations, both idealised (2013a) and real-life (2013b), were acutely sensitive to the chosen parameterisation. Specifically, when graupel hydrometeors were simulated as lighter and greater in number, they resulted in a stronger cold pool and rear-inflow jet, and hence the bowing initiated earlier. The chosen parameterisation scheme also strongly affected pertinent forecast parameters such as precipitation amount, system speed, wind gusts, and areal coverage. The issue with these conclusions is their lack of generality to other events, including variations in synoptic regime, the initial condition dataset, and poor sampling of the model attractor (or model climate) via use of a deterministic approach. While computational and time limitations preclude in-depth analysis repeated for many cases, these methodological shortcomings nonetheless motivate a more general ensemble approach with cases in different synoptic regimes.

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. Our computer models are incomplete and imperfect: while smaller phenomena are resolved explicitly by ever-decreasing grid spacings, there will always be a scale at which chaotic, non-linear processes are implicitly resolved, or parameterised. Parameterisation is currently 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), surface radiation, cloud microphysics, and so on. The spread of parameterisation schemes developed by different institutions and researchers is not controlled, and as such, biases in each scheme can constructively or destructively interfere within a forecast run, without *a priori* knowledge of the impact on the forecast compared to *a posteriori* verification from radar reflectivity observations. For example, Adams-Selin and coauthors (2013b) found bow-echo forecasts to vary substantially when solely changing the microphysical parameterisation. In their study, the authors 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. Likewise, Berner et al. (2011) trained their mixed-physics models over a number of months to determine the optimal configuration for spread and skill. This is, of course, not a practical or general approach for operational centres to endorse long-term, when one considers the training sensitivity to many factors.

In addition to model error, the atmosphere as a partly chaotic system is sensitive to initial condition error (Lorenz 1969). As such, there is a theoretical limit of predictability, estimated when assuming purely chaotic (turbulent) flow. On scales of 10 km, Lorenz estimated predictability to be limited to 1–2 h. Fortunately from a forecasting standpoint, it is evident in forecast models that the atmosphere has inherent predictability at the mesoscale longer than that proposed by Lorenz. This may be due to known forcings — high terrain, synoptic-scale fronts (e.g., Anthes et al. 1985) — and stable mechanisms that locally limit error growth, such as supercells (Lilly 1990), and in confluent, weak flow (Oortwijn 1998). Unfortunately for MCS forecasts, moist convection is very destructive to predictability (Zhang et al. 2003), even affecting global-model forecasts of blocking patterns in Europe at the medium-range (Rodwell et al. 2013). In addition, diagnosis of substantially damaging initial-condition error is fraught with difficulty due to both up- and downscale growth of errors (**?**, in press, and references therein). Notably, Kühnlein et al. (2014) showed that global-model initial-condition perturbations may not capture the variance in convective scales, which impacts particularly the first six hours of a forecast time.

To address these problems and better sample the spectrum of possible outcomes of the model atmosphere, many forecast centres use a number of different numerical simulations (ensemble forecasts). There are different ways of creating members that differ from their control: through mixed-parameterisation configurations; through perturbed initial conditions; etc. Recently, studies have yielded a method to return kinetic energy erroneously dissipated in the model between the resolved and unresolved scales to better account for model error. This so-called stochastic 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, except at the surface (Berner et al. 2011). Furthermore, when a SKEB scheme was combined in that study with a mixed-parameterisation configuration, performance was even better. As parameterisations 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 centres, but can provide a larger corpus of 'alternative realities' in which to seek sensitivity of atmospheric phenomena (Hanley et al. 2013). Hence, it seems prudent to approach ensemble creation through numerous methods, with the caveat that not all methods can rigorously measure predictability in its purest sense.

This paper will provide an initial foray into the following questions: 1) in terms of convective mode and the presence of bowing, is the quality of a (deterministic, ensemble) forecast systematically sensitive to the larger-scale setting? 2) When a NWP model forced by archived (re)forecast data incorrectly resolves the mode of a system, is this poor performance more related to model error or to initial conditions? Of course, these questions will not be answered in sufficient depth at this stage, but the preliminary results may provide a foundation for further investigation.

As a final note on nomenclature, this study will borrow the dichotomous categorisation of derechoes in Johns and Hirt (1987) by dividing bowing features into two groups: those that appear multiple times along a QLCS, typically in parallel with a front (serial bow echoes, Fig. **??**b,c), 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 bow echoes, Fig. **??**a,d). This is motivated by the wish to concentrate on the potentially flexible criteria of radar reflectivity signatures, rather than strict (and more arbitrary) surface-wind definitions of a derecho. Furthermore, not all bow echoes are derechoes, and *vice versa*.

2. Data and Methods

The study focuses on four cases (Fig. **??**): (a) a progressive bow echo, 26–27 May 2006; (b) a serial bow echo, 10–11 September 2009; (c) a serial bow echo, 19–20 May 2011; and (d) a progressive bow echo, 15–16 August. Each case may be referenced in the following sections in short by its letter: Case A, Case B, and so on.

All numerical simulations were performed on the same supercomputer system 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 parameterisation configuration (Table 1) was chosen primarily for stability, due to the large number of ensemble runs required with this (or most of this) control configuration. The control microphysical parameterisation (Thompson) was also selected due to good performance in similar studies (Brian Squitieri, personal communication). The constant domain size is 451 by 451 points with grid spacing Δx set at 3 km. The domain location varied depending on the case study; each location is shown in Figure 1. The timestep was six seconds (i.e. $2\Delta x$). Fifty vertical levels were specified manually to better resolve the PBL, as found in similar studies (David Jahn and Brian Squitieri, personal communication).

Depending on the ensemble experiment, initial and boundary conditions were provided by one member of the 11-member Global Ensemble Forecast System Reforecast dataset (GEFS/R2), or the North American Mesoscale (NAM) archive analyses. As the GEFS/R2 dataset does not contain sufficient resolution in soil layers for the WRF to run as-is, Global Forecast System analyses of soil temperature and moisture were prescribed for each batch of initial and boundary conditions (see Lawson 2013 for further information on this method). While there is no doubt that small changes in variables such as soil temperature can affect convective initiation (Clark and Arritt 1995), the absence of perturbations in soil variables is not likely to affect any relativistic conclusions through this method. Boundary conditions were specified every three hours. All runs were initialised on 0000 UTC on the first day of the case study, and ran for 36 hours (39 hours for Case B) to (a) allow mesoscale systems to develop in good time, (b) to allow perturbations between ensemble members to grow large enough to observe easily, but not so large that the timescale of interest was beyond the theoretical predictability limit for meso-α-scale motion, and (c) to allow use of the once-daily GEFS/R2 data.

Eleven-member WRF–GEFS/R2 ensembles were created by running WRF eleven times, each with a different set of initial and boundary conditions from the GEFS/R2 dataset, with the control configuration (Table 1. The GEFS/R2 control member and ten perturbation members were all used for both inspection of atmospheric variables and computation of error growth. This ensemble setup is hereby termed **ICBC**. Ensembles were also created by varying the microphysical scheme (termed **MXMP**), but holding all else constant (and selecting a constant ICBC input data set). The 11 microphysical schemes (that supplement the control scheme, Thompson) used here are detailed in Table 3. The schemes were chosen to mirror a similar study by Adams-Selin et al. (2013b). In their method, the fall speed of hail was modified in the WRF source code, so that a parameterisation could become 'hail-like' or 'graupel-like'. An identical method has been used in this study for the WSM6, WDM6, and Morrison schemes (Adams-Selin, personal communication), to improve the sampling of 'parameterisation space'. As a caveat to the MXMP method, each member is not *a priori* equally as likely to ultimately simulate the 'correct' solution in the same sense as a controlled ICBC ensemble. Hence, this ensemble method is technically a sensitivity study, and as such does not as rigorously measure predictability. However, it can offer insight into performance of each parameterisation scheme. To more rigorously sample the model error, a final ensemble method is used: a SKEB scheme (Berner et al. 2011) accounts for kinetic energy lost between resolved and unresolved scales by randomly [1] ‘injecting’ kinetic energy back into the wind field. This ensemble method, termed **STCH**, requires a prescribed ICBC dataset and MXMP parameterisation. Overall, notwithstanding the previous caveat about the lack of rigour in use of MXMP members to assess predictability, this methodology is summarised in Figure **??**. While this schematic follows the selection of the best-performing members to assess the stability of a ICBC ensemble member solution, it can also be used to see if a poorly performing member is a ‘one-off’, or is a representative sample of that particular model climate. The assumption that spread decreases in ICBC, MXMP, and STCH ensembles, respectively, is supported by experimental results, as follows in the next section.

To track error growth through the ensemble domains, difference total energy (DTE) at a given timestep was calculated in a similar way to Tan et al. (2004), as follows:

$$\mathrm{DTE} = \frac{1}{2}\sum(U'^2_{ijk} + V'^2_{ijk} + \kappa T'^2_{ijk}) \tag{1}$$

where κ here [2] is 0.286, and U', V', and T' are the differences in zonal wind, meridional 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 solely in height to create a latitude–longitude cross-section. In this study, for each ensemble, n sets of differences were calculated between all permutations of the N ensemble members without repetition, where $n = \frac{N}{N-1} + 1$.

[1] The ‘randomness’ is via a seed integer set in the WRF namelist. Hence unlimited independent ensemble members can be created by changing this value.

[2] DTE has been formulated using this constant value, or as in Tan et al. (2004), via use of a reference temperature. Nonetheless, a constant is used here without ill-effect; the value of DTE in this study is in relative comparison between ensemble methods.

3. Results

Figure **??** shows characteristic radar reflectivity signatures of the four MCS cases. Due to space limitations, model output for Cases A–C will be described briefly rather than shown due to the large (44) amount of ensemble members. A descriptive, subjective summary is found in Table 2.

The progressive bow echo of 26–27 May 2006 (Case A, Fig. **??**a) has been covered in more detail in Snively and Gallus (2014), where the authors found WRF runs forced by both NAM and GFS forecast data to incorrectly reproduce the convective mode of the MCS. They also found little sensitivity to microphysical schemes. An ICBC experiment of this event (not shown) found even poorer performance when forced with GEFS/R2 conditions; a NAM-forced run was a little more successful, and lays the foundation for further study (see Section 4). The serial bow echo of 10–11 September 2009 (Case B) contained LEWP-like features embedded along its (QLCS) length, particularly around 0800 UTC on the 11th (Fig. **??**). This QLCS was associated with a cold front. Case C is a dramatic example of a serial bow echo that spawned multiple tornadoes and other severe weather on 19–20 April 2011. Finally, a progressive bow echo brought damaging wind and hail to Kansas, Oklahoma, and Texas on 15–16 August 2013. This case (D) will be the subject of a more in-depth subsection later in this paper.

a. Error growth in all four cases

In summary: Case A emphatically did not correctly simulate convection; Case B struggled to correctly simulate the QLCS and its orientation, and bowing along the line was intermittent across members, but it was generally successful; Case C did not simulate the dramatic bowing seen in observations, and only some members formed an unbroken line of convection, others simulating more discrete cells; finally, Case D was most successful, with a progressive bow echo simulated in almost all cases, timing and location errors notwithstanding.

Figure **??** shows the time series of DTE summed over all three spatial dimensions at each time, for the ICBC experiment from the four cases in this study. Each plotted line represents the average DTE growth of all eleven (ten perturbation and one control) ensemble-member permutations (without repetition) that involve that member. The black line is then the average of these averages. Absolute DTE values cannot be compared between cases, as the value is highly dependent on the geographical region, mesoscale environment, etc. However, it is useful to look at relative growth rates, which can be generally contrasted between cases. For instance, Case A does not experience rapid (quasi-exponential) DTE growth until around the time of maximum solar insolation on 26 May. This may be related to the onset of cellular convection at this time; before this (when DTE growth is limited), despite large areas of radar reflectivity across the domain (not shown), the mode is larger scale (and potentially less destructive in terms of predictability). Conversely, it could simply be due to the limited small-scale variance seen in the initial six hours of a numerical simulation forced by global data with large-scale perturbations (Kühnlein et al. 2014). Interestingly, despite very poor reflectivity simulation in the ICBC ensemble, error growth is comparable in rate and magnitude to the other, better-simulated progressive bow echo (Fig. **??**d).

Case B, in contrast, sees a slow quasi-exponential growth rate. As errors may saturate and cascade upscale faster at smaller scales (not tested here), the limited DTE growth earlier in the period in Case B may be related to a strong coupling with the synoptic-scale cold front. On the other hand, the orientation of the QLCS varies quite drastically at the time of bowing (around 0800 UTC), perhaps related to the position of the cold front, and we may expect these large-scale errors to propogate downscale (and be reflected in the DTE data), as noted by **?** (in press). To diagnose whether this explains the delay of large DTE growth as seen in the other three cases in Fig. **??**, a spectral analysis amongst other things is required (see Section 4).

The magnitude of DTE in Case C is around 50% larger (note the different y-axis scale in Figure), likely related to the large scale of the QLCS and the strong convection (compare Fig. **??**c with panels a,b,d). The growth rate pattern by eye appears to be a hybrid of Case A (early- then late-period convection) and Case B (potential strong coupling with large-scale boundary). With limited data, we may conjecture this is a reflection of multiple embedded small-scale bowing features within a larger-scale QLCS.

Finally, Case D shows a strong bimodal structure of DTE, with maxima around midnight local time for both days (around 6 and 30 forecast hours). This is likely related to vigorous convection in these periods. More interesting is the intermediate recovery. This may simply be regression to the mean; i.e., the rapid growth of DTE within areas of cellular convection ceases and a more typical environmental balance of variables is assumed. It could also be related to the movement of thunderstorms outside the domain, and advection of more tranquil air into the domain. However, it could be related to the (presumed) nocturnal low-level jet (NLLJ) that is often instrumental in nocturnal convection in the Great Plains. The dependence of the NLLJ on the Coriolis parameter and high terrain (i.e., known forcings) may converge ensemble solutions towards a model attractor. Alternatively, divergent flow within the NLLJ may separate parcel pairs (i.e., increase the Lyapunov exponent) and hence decrease predictability (similarly, convergent flow in the nose of the NLLJ may increase predictability). The relationship of the NLLJ to theoretical predictability is unfortunately discussed little in the literature (see Section 4).

b. Stability of a good ensemble simulation

The focus will now be on the case of 15–16 August 2013, to better evaluate its predictability, to observe the growth of errors, and to assess the performance of the ensembles. Specifically: is the good performance of the ICBC experiment related to fortuitous selection of microphysics and other aspects of the model configuration? (Interestingly, the reforecast-forced ICBC runs outperformed the NAM-analysis-forced run.) To address this, DTE will be assessed, first by integrating vertically to get a spacial sense of its distribution, then via MXMP and STCH experiments based on the best-performing set of initial and boundary conditions (p09, not shown). Figure **??** shows vertically-integrated DTE evolving through Case D at selected time, and for reference, Fig. **??** displays the composite radar reflectivity observed at the same times. At the first time (0300 UTC on the 15th, Fig **??** and Fig **??**), three hours into the simulation, the spread is larger in two general areas: (1) locations with moist convective activity (seen in simulated radar reflectivity), where DTE growth is much larger (Zhang et al. 2003), and (2) along the surface-pressure trough (not shown) running west–east in Nebraska. Over the next six hours, convection dominates the areas of rapid DTE growth (cf. Fig **??**). The arc of locally high DTE in southwest Nebraska to northeast Colorado (0900 UTC on the 15th, Fig **??** and Fig **??**) appears to be related to discrepancies between members in simulating an outflow boundary, which is apparent in individual ensemble members' surface pressure fields (not shown). This area is also associated with a developing mesoscale convective vortex (MCV) as noted by Storm Prediction Center (SPC) mesoscale discussions. This MCV moves east-southeast over the next nine hours; at (1800 UTC on the 15th (Fig **??** and Fig **??**)), note the increasing homogeneity in the DTE field as convection dissipates (and again, cf. Fig **??**). Yet the local maximum of DTE associated with the MCV stands out from this background field; at 1800 UTC, convection initiates in the region of the MCV both in observations and most simulations (not shown). The next twelve hours summarise the essence of chaos theory: these small variations in initial location and timing of convective initiation, related to differences in the structure of the MCV between ensemble members, grow to become large DTE values as almost all members generate a progressive bow echo that moves southward through Kansas and Oklahoma (0600 UTC on the 16th, Fig **??** and Fig **??**), but all in different locations with variations in bowing structure (e.g., Fig. **??** three hours earlier). The mode solution appears to be highly predictable (conceptually, a large basin of attraction), even if the location and specifics of the bowing is not.

Next, the stability of this control simulation is assessed by running a MXMP ensemble with the 'best' initial conditions (subjectively chosen as p09). Figure **??** shows the average DTE growth for each parameterisation permutation pair [3]. Figure **??**, to aid physical interpretation, is the postage-stamp plot of composite reflectivity found in each member. We find both outputs support the assumption in the methodology that spread is reduced (at least, at these forecast ranges) when using MXMP rather than ICBC ensembles. Note this does not necessarily equate to a more skillful ensemble-mean forecast. The composite reflectivity plots show much less variation in location, but still variation in structure. When we delve further, and run a STCH ensemble based on the successful control microphysics, and the same p09 initial and boundary conditions, we find even greater convergence of solutions towards a bow echo that is similar to verification (Fig. **??**). Again in support of the methodology, error growth is lower still (Figs. **??** and **??**). Interestingly, the 'bad' microphysics scheme (deemed to be Morrison (hail-like)) was more successful in some STCH members, showing again the danger of using one deterministic forecast to make conclusions about a model scheme. The decreased DTE magnitude in STCH ensembles is related to even more spatial agreement, and slightly less variation in structure (but still substantial enough to warrant further investigation). Plotting maximum 10-m wind over the period of the bow echo (not shown) finds this variation not merely superficial, but reflected in surface wind potentially associated with downbursts within the bow echo. While this is an initial foray, future work will include investigate the physical background to the variations of the STCH members in this ensemble.

[3] Say that fast three times

4. Summary and Future Work

To simply attribute bow-echo simulation failure on deficiencies with a microphysical parameterisation is dangerous: the concluded sensitivity may itself be sensitive to initial and boundary conditions (the synoptic regime, time of year, etc.) and model configuration (other parameterisations, domain size, etc.). To account for these two sensitivities, ICBC and STCH ensembles, respectively, were run to assess the stability of a 'good' forecast. A MXMP experiment allows comparison with other studies, and can lay the groundwork to better address the physical reasons for model error in future. In particular, Skew-Ts and cross-sections for members that are good or bad within e.g. the STCH ensemble may shed light on physical processes responsible for good and bad simulations.

Preliminary findings include:

- The poor performance of NWP models in Case A seen in Snively and Gallus (2014) is not solved by varying initial conditions. This might be indictive of a trigger so small that available models cannot create perturbations on a scale small enough to recreate the bow echo evolution in at least one member.

- Case D appears to be a highly predictable case. Considering the opposite in Case A, this suggests that poor NWP model performance is not necessarily related to the bow-echo subtype morphology.

- Both Cases B and C contain much variation in LEWP-type bowing structures. Case B forecasts varied more in QLCS orientation, while Case C forecasts varied in the production of slabular versus cellular convection. No members in Case C managed to recreate the dramatically large radii of curvature seem in the LEWP-type bowing features.

- Spread in this study decreases between ensemble members in ICBC, MXMP, STCH, in descending order.

- A poorly performing microphyiscal scheme may not be any worse than others when used with SKEB perturbations (i.e., the poor solution is not stable).

In light of the recategorisation of bow echoes into progressive and serial sub-types, a similar reclassification of cases in Snively and Gallus (2014) may reveal a tendency for one type over the other to be associated with poor skill scores assessed in that study.

To further this study's methodology, MXMP and STCH ensembles will be conducted for the other three cases, and spectral analysis will be performed on DTE data to somewhat normalise error growth by event, and to diagnose the source of errors at an early stage. Finally, additional topics may include the stochastic parameterisation for hail/graupel to better reflect uncertainty in the atmospheric variables; an investigation into the predictability of the NLLJ and potential impact on error growth within mesoscale systems in the Great Plains; and whether the conceptual basin of attraction for convective mode is fractal-like, with very unstable solutions, and if so, whether this basin structure is sensitive to an atmospheric tuning parameter akin to the control parameter in chaotic dynamical systems k (e.g., Williams 1997, p.161).

Acknowledgment. The author thanks Rachael Adams-Selin and her colleagues for supplying the Fortran modifications and advice relating to the microphysical schemes.

Acknowledgments. Start acknowledgments here.

Make your BibTeX bibliography by using these commands:

References

Adams-Selin, R. D., S. C. van den Heever, and R. H. Johnson, 2013a: Impact of graupel parameterization schemes on idealized bow echo simulations. *Mon. Weather Rev.*, **141 (4)**, 1241–1262.

Adams-Selin, R. D., S. C. van den Heever, and R. H. Johnson, 2013b: Sensitivity of Bow-Echo simulation to microphysical parameterizations. *Weather Forecast.*, **28 (5)**, 1188–1209.

Anthes, R. A., Y.-H. Kuo, D. P. Baumhefner, R. M. Errico, and T. W. Bettge, 1985: Predictability of mesoscale atmospheric motions. *Adv. Geophys.*, **28**, 159.

Berner, J., S.-Y. Ha, J. P. Hacker, a. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Weather Rev.*, **139 (6)**, 1972–1995.

Clark, C. A., and P. W. Arritt, 1995: Numerical simulations of the effect of soil moisture and vegetation cover on the development of deep convection. *J. Appl. Meteorol.*, **34 (9)**, 2029–2045.

Fritsch, J. M., R. J. Kane, and C. R. Chelius, 1986: The contribution of mesoscale convective weather systems to the Warm-Season precipitation in the united states. *J. Climate Appl. Meteor.*, **25 (10)**, 1333–1345.

Gallus, W., N. a. Snook, and E. V. Johnson, 2008: Spring and summer severe weather reports over the midwest as a function of convective mode: A preliminary study. *Weather Forecast.*, **23 (1)**, 101–113.

Hanley, K. E., D. J. Kirshbaum, N. M. Roberts, and G. Leoncini, 2013: Sensitivities of a squall line over central europe in a Convective-Scale ensemble. *Mon. Weather Rev.*, **141 (1)**, 112–133.

Johns, R. H., and W. D. Hirt, 1987: Derechos: Widespread convectively induced windstorms. *Weather Forecast.*, **2 (1)**, 32–49.

Kühnlein, C., C. Keil, G. C. Craig, and C. Gebhardt, 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quart. J. Roy. Meteor. Soc.*

Lawson, J., 2013: Analysis and predictability of the 1 december 2011 wasatch downslope windstorm. M.S. thesis, University of Utah, 84 pp.

Lilly, D. K., 1990: Numerical prediction of thunderstorms—has its time come? *Quart. J. Roy. Meteor. Soc.*, **116**, 779–798.

Lorenz, E., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21 (3)**, 289–307.

Markowski, P., and Y. Richardson, 2010: *Mesoscale Meteorology in Mid-latitudes*. Wiley-Blackwell, 407 pp.

Oortwijn, J., 1998: Predictability of the onset of blocking and strong zonal flow regimes. *J. Atmos. Sci.*, **55**, 973–994.

Palmer, T., 2001: A nonlinear dynamical perspective on model error: A proposal for nonlocal stochasticdynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279–304.

Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor Medium-Range weather forecasts for europe. *Bull. Am. Meteorol. Soc.*, **94 (9)**, 1393–1405.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, and others, 2008: A description of the advanced research WRF version 3, 2008. *NCAR Technical Note. NCAR/TN-*, 113.

Snively, D. V., and W. A. Gallus, 2014: Prediction of convective morphology in Near-Cloud-Permitting WRF model simulations. *Weather Forecast.*, **29 (1)**, 130–149.

Society, A. M., 2014: Glossary of meteorology: Mesoscale convective system. Accessed: 2014-4-23, http://glossary.ametsoc.org/wiki/Mesoscale_convective_system.

Tan, Z.-M., F. Zhang, R. Rotunno, and C. Snyder, 2004: Mesoscale predictability of moist baroclinic waves: Experiments with parameterized convection. *J. Atmos. Sci.*, **61 (14)**, 1794–1804.

Williams, G., 1997: *Chaos Theory Tamed*. Joseph Henry Press.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60 (9)**, 1173–1185.

LIST OF TABLES

TABLE 1. Control parameterization schemes used in numerical modelling configuration.

Parameterization	Scheme
Microphysics	Thompson
Longwave Radiation	RRTM
Shortwave Radiation	Dudhia
Surface Layer	MYNN
Land Surface	Noah
Planetary Boundary Layer	MYNN Level 2.5

TABLE 2. Subjective summary of ensemble-member performance in each of the four ICBC experiments.

Date	Convection	Bowing
Case A: 26–27 May 2006	None	N/A
Case B:10–11 September 2009	All	Some
Case C: 19–20 April 2011	All	Too weak
Case D: 15–16 August 2013	All	Most

TABLE 3. Control parameterisation schemes used in numerical modelling configuration.

Thompson †
WSM6 *
Kessler
Ferrier
WSM5
WDM5
Lin
WDM6 *
Morrison *

† Control parameterisation

* Changed to be either hail- or graupel-like

LIST OF FIGURES

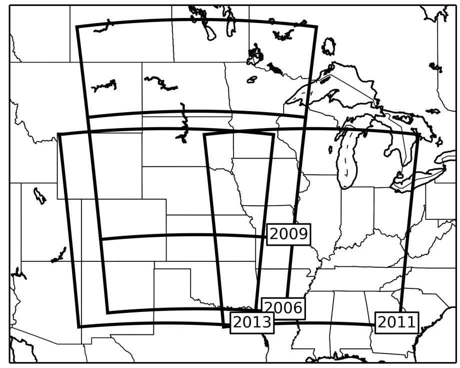

FIG. 1. Domains for each of the four cases. Label denotes the year of each case (c.f. Fig ?.