

THUNDERSTORMS DO NOT GET BUTTERFLIES

BY DALE R. DURRAN AND JONATHAN A. WEYN

Simulated squall lines developing in horizontally homogeneous environments are more sensitive to small initial perturbations on scales of 100 km than to larger perturbations on smaller scales.

The weather will evolve differently from that predicted by a perfect forecast model if the initial state provided to the model differs even slightly from that for the atmosphere. Progressive increases in computing power have made it possible for forecast models to capture atmospheric motions on ever smaller scales, yet such improvements will not necessarily lead to significantly better forecasts. Lorenz (1969) showed that confining errors in the initial state to ever smaller scales, possibly even down to the size of a butterfly, did not significantly extend the time over which his highly idealized model exhibited predictive skill. Nevertheless, as noted by Durran and Gingrich (2014), Lorenz's work also implies that

initial errors at extremely small scales are not likely to impose a practical limit on forecast accuracy, because they would be swamped by minor relative errors in the initial conditions at much larger scales. Here we show that similar conclusions about the importance of larger scales apply to the propagation of initial-condition errors in simulations of thunderstorms and squall lines using a numerical cloud model that provides a far more complete representation of atmospheric motions than that used by Lorenz.

THE LORENZ MODEL. Surprisingly, the practical irrelevance of butterflies and other very small-scale initial errors can be deduced from the original work of Lorenz (1969), who compared the impact on predictability of perturbations introduced at both the smallest and largest scales in an idealized turbulence model. Lorenz found that the impact of errors with equal absolute magnitude was similar whether the errors initially appeared at the smallest scale (his experiment A) or the largest scale (his experiment B). Lorenz devoted only one paragraph to the large-scale case and focused instead on the growth of small-scale errors, including the possible impact of seagull wings.

Lorenz modeled the behavior of ensembles, or families, of turbulent flows in which each family member freely evolves from slightly different initial conditions relative to the ensemble-mean state. Using a slightly updated version of Lorenz's model (Durran

AFFILIATIONS: DURRAN AND WEYN—Department of Atmospheric Sciences, University of Washington, Seattle, Washington

CORRESPONDING AUTHOR: Dale Durran, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195-1640

E-mail: drdee@uw.edu

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-15-00070.1

A supplement to this article is available online (10.1175/BAMS-D-15-00070.2)

In final form 26 May 2015

©2016 American Meteorological Society

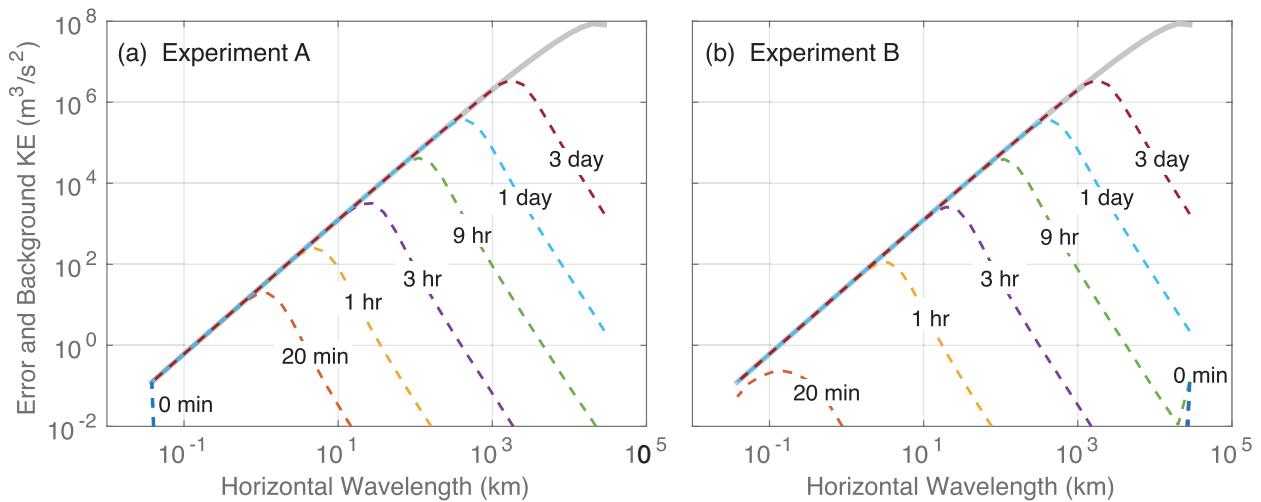


FIG. 1. Error amplitude (perturbation kinetic energy density) plotted as a function of horizontal wavelength for Lorenz's (a) experiment A and (b) experiment B (dashed lines) at simulation times ranging from 20 min to 3 days. The thick gray line shows the background kinetic energy of the ensemble-mean velocities.

and Gingrich 2014), the loss of predictability in his experiments A and B is shown in Fig. 1 by plots of the error amplitude (measured by its kinetic energy KE') as a function of horizontal scale at times ranging from 20 min to 3 days. Also plotted as a thick gray line is the background ensemble-mean atmospheric kinetic energy (KE) spectrum assumed by Lorenz, which increases in proportion to the wavelength λ to the $5/3$ power. When KE' at a given wavelength increases to match KE , the error is saturated and all predictability at that wavelength is lost.

If, as shown in Fig. 1a, the initial error is only at the smallest scale (38 m) and is saturated, then that error modifies somewhat larger scales of motion, which in turn modify even larger scales. After 20 min (red dashed line), the errors have propagated upscale and are saturated between wavelengths of 38 m and 1 km; the errors fall off rapidly at wavelengths longer than 1 km and are too small to plot at wavelengths greater than 11 km. At subsequent times, the error continues to cascade upscale, saturating, for example, at wavelengths up through 100 km in about 9 h.

Initial errors having the same *absolute* magnitude are imposed at the longest wavelength (28,300 km) in experiment B. As shown in Fig. 1b, the large-scale error propagates very rapidly downscale, becoming apparent at the shortest wavelengths within 20 min. These newly generated small-scale errors then start cascading back upscale in a manner similar to that in experiment A. By 3 h into the forecast (dashed purple lines), the spectral distribution of the errors in both experiments has become very similar, with saturated errors at all scales smaller than roughly 20 km. Note that, although the

error in experiment B actually originates in the large scale, it appears to originate in the small scales because that is where the relative error first becomes significant.

To the extent that the Lorenz model correctly describes the atmosphere, the behavior in Fig. 1a implies that inaccurately initialized fields at even smaller wavelengths, the size of butterflies, could indeed ruin a weather forecast. Nevertheless, comparing both panels of Fig. 1, it is evident that butterflies will never be of practical importance because trivial *relative* errors in the large scales will overwhelm 100% relative errors on the very small scales. The initial large-scale KE' in experiment B is smaller by a factor of 10^{-9} than the background kinetic energy at the same scale, corresponding to initial perturbations in the velocity field that are a factor of $10^{-9/2}$ smaller than the background velocities. Even on such large nominally well-observed scales, it is not likely that the atmospheric circulations at a given instant could be determined to within a relative error of $10^{-9/2}$ at any time in the foreseeable future. Within the context of experiments A and B, a complete failure to correctly initialize atmospheric features with a wavelength of 38 m would be of no practical importance unless the largest-scale winds can be initialized with relative errors less than 0.003%. Similarly, the even smaller-scale perturbations generated by butterflies should never have a practical impact on weather forecasts because the initial conditions at larger scales cannot be specified with sufficient accuracy.

SQUALL-LINE SENSITIVITIES. How applicable is the preceding to the real atmosphere? The dynamics underlying Lorenz's turbulence model are

those for a highly idealized two-dimensional flow in which the turbulence is assumed to be spatially homogeneous and statistically stationary. The actual background atmospheric kinetic energy spectrum only follows the $\lambda^{5/3}$ slope assumed by Lorenz for wavelengths smaller than about 400 km (Nastrom and Gage 1985). At scales larger than 400 km, the atmospheric kinetic energy spectrum transitions to a λ^3 slope, and both downscale and upscale error propagation proceed more slowly (Rotunno and Snyder 2008). An important real-world counterexample to the roughly equivalent loss of predictability produced by equal-amplitude large- and small-scale errors might be provided by thunderstorms, which are believed to be particularly effective at rapidly propagating small-scale perturbations to larger scales (Zhang et al. 2006; Hohenegger and Schär 2007). Indeed, many case studies of erroneous weather forecasts have suggested that serious errors can be traced back to regions of deep moist convection (Zhang et al. 2002, 2003, 2007; Rodwell et al. 2013).

We therefore use a numerical cloud model (Durran and Klemp 1983) to generate a pair of 25-member ensemble simulations of collections of thunderstorms, known as squall lines, under atmospheric conditions similar to those specified in Weisman and Klemp (1982) and used in many classic theoretical studies of deep moist convection. In all ensemble members, the initial thunderstorms are triggered by three identical bubbles of warm air near the surface. Variations in the position and strength of the squall line among different members of the ensembles are produced by small initial “errors” in the surface temperature field that decay exponentially over a height of 1 km. The first 25-member ensemble is analogous to Lorenz’s experiment A in that the errors are small scale, having horizontal wavelengths of 8 km in both the east–west and north–south directions. The phase of the 8-km-wavelength perturbation varies randomly among the ensemble members. The second 25-member ensemble, analogous to experiment B, has random-phase initial perturbations at much longer 128-km wavelengths. However, unlike Lorenz’s experiments A and B, where the initial absolute errors had equal magnitudes, the amplitudes of the 128-km temperature perturbations are only one-quarter as large as those at 8 km.¹

After 4 h of simulated time, thunderstorms are present in all ensemble members and are primarily distributed along a roughly north–south line.

¹ The parameters for the numerical simulations are detailed in the supplement (10.1175/BAMS-D-15-00070.2).

The position of the heavy rain cells, as shown by synthetic radar images in Fig. 2, varies considerably between the different members. In members such as those shown in Figs. 2a and 2c, the strongest cells are concentrated into relatively short segments, whereas in other members, such as those in Figs. 2b and 2d, the cells develop into much longer north–south lines. Most significantly, the short- and long-line variations appearing in the ensemble with initial errors at wavelengths of 8 km (e.g., Figs. 2a,b) are comparable to those appearing in the 128-km-wavelength ensemble (e.g., Figs. 2c,d).

KE SPECTRA FOR THE SQUALL-LINE ENSEMBLES.

To obtain a more quantitative comparison of error growth over all the members of the 8- and 128-km ensembles, we compute the same type of spectral decomposition shown for Lorenz’s model in Fig. 1. The kinetic energy of the errors (KE’ computed from the difference between the horizontal velocities of each ensemble member and the ensemble-mean velocities) at a height² of 10 km is plotted as a function of horizontal wavelength by the dashed lines in Fig. 3. In contrast to the situation in Fig. 1, the only initial kinetic energy in these simulations is in a horizontally uniform wind field; the spectra for the ensemble-mean KE therefore develop from zero along with the growing squall line. By 30 min into the simulations, both ensembles have generated very similar mean KE spectra (solid blue lines), but the error spectra (dashed blue lines) are different. In the 8-km ensemble, errors with similar magnitudes have developed at all scales, whereas in the 128-km ensemble, there is a pronounced maximum in KE’ at the scale of the initial perturbation, and much weaker values at other wavelengths.

Yet by 4 h, the KE’ spectra for the 8- and 128-km ensembles (dashed red curves) are virtually identical, implying that all trace of the differences in the scale and magnitude of the initial errors has disappeared. Moreover, at those wavelengths less than 25 km, the values of KE’ and KE are almost the same and all predictability in both ensembles has been lost. The loss of predictability in our simulations is generally consistent with recent assessments of the impact of finescale weather radar data on real-time thunderstorm forecasts from the National Oceanic and Atmospheric Administration’s (NOAA) Hazardous Weather Testbed (HWT) Spring Forecasting

² A height of 10 km is representative of the elevations at which observations of atmospheric spectra were collected by Nastrom and Gage (1985).

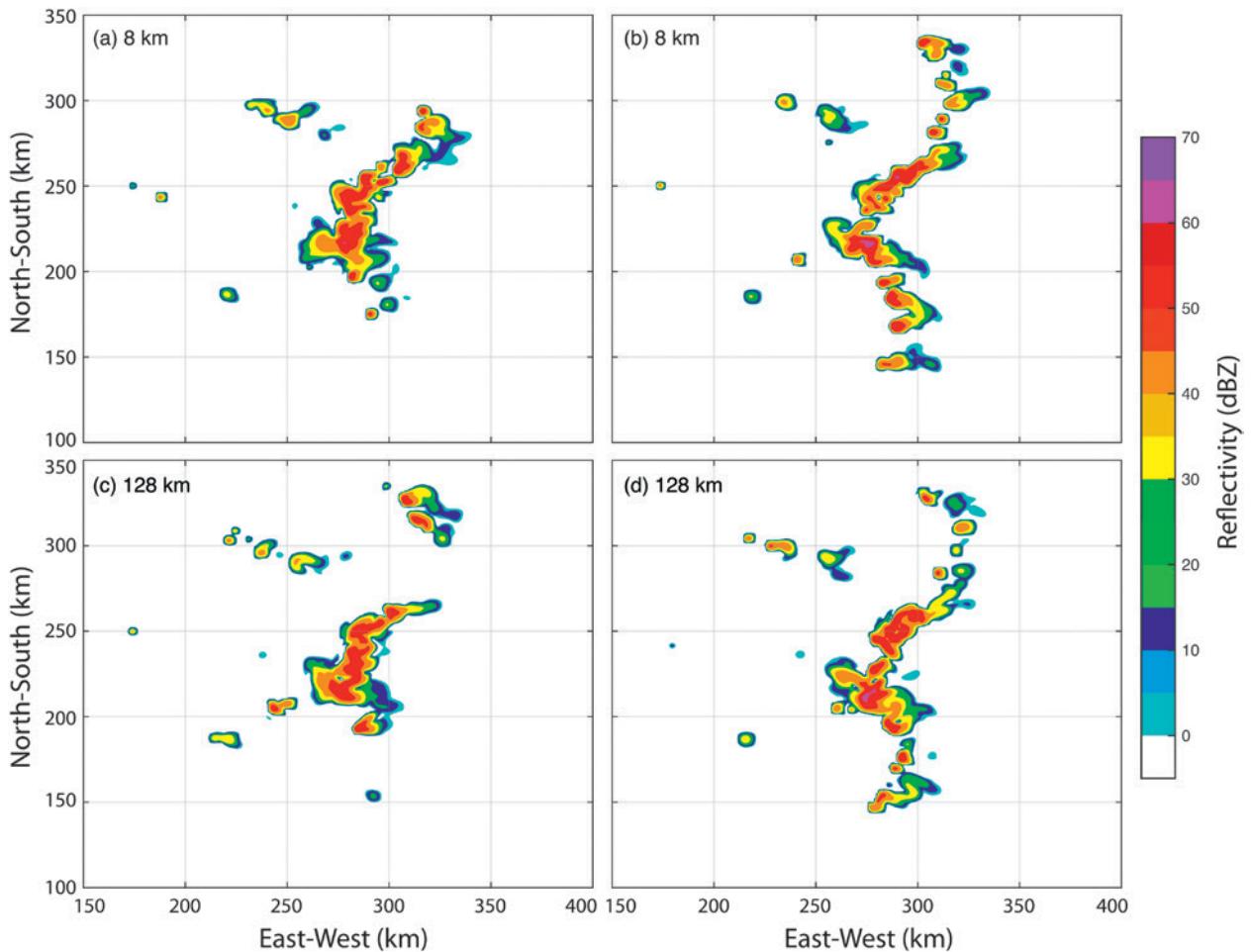


FIG. 2. Synthetic radar reflectivity at hour 4 from the simulated rain fields at a height of 4 km for (a),(b) two members from the 8-km ensemble and (c),(d) two members from the 128-km ensemble. These plots zoom in on the region of active convection.

Experiment (Kain et al. 2010). Stratman et al. (2013) found that assimilating finescale radar data did not add skill at scales less than 40–80 km beyond forecast lead times of a few hours, while the analysis of Surcel et al. (2015) showed a somewhat more rapid loss of predictability.

Also of note is that by 4 h the background KE spectra in Fig. 3 have developed a slope approximating that of the observed $\lambda^{5/3}$ background atmospheric KE spectrum (Nastrom and Gage 1985; thick gray line). Figure 4 illustrates the growth of the background KE spectra at additional times ranging from 30 min to 6 h for the ensemble with 8-km-wavelength initial temperature perturbations.³ At 3 h, the KE spectrum is similar to the observed spectrum for wavelengths between 15 and 100 km. By 6 h, the simulated KE spectrum generally parallels that of the observed

³ The spectra from the 128-km-wavelength ensemble are virtually identical.

background KE at all wavelengths longer than 15 km, but it is stronger than the observations by about a factor of 3, a difference that likely arises from the vigor of the simulated squall-line circulations and the relatively small size of our 512 km × 512 km domain.

Over 35 years ago, Gage (1979) and Lilly (1983) hypothesized that the observed $\lambda^{5/3}$ mesoscale KE spectrum could be generated by deep moist convection through an upscale energy cascade. Their hypothesis has remained controversial, with many researchers suggesting an alternative hypothesis that the spectrum is formed by a downscale energy cascade [see the discussion in Lindborg (2015)]. The growth of the KE spectra shown in Fig. 4 is certainly due to deep moist convection, but its detailed evolution does not particularly resemble a cascade in which energy spreads either upscale or downscale through progressive exchanges with nearby scales. Instead there is simultaneous growth at almost all scales, beginning from an initial distribution that is

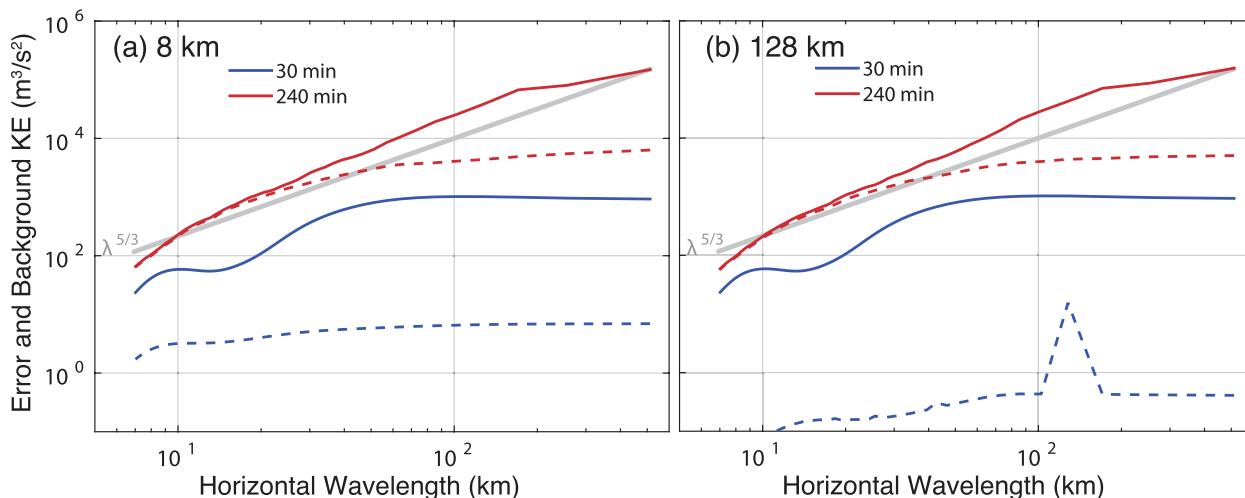


FIG. 3. Error amplitude (perturbation kinetic energy density) at a height of 10 km plotted as a function of horizontal wavelength from the squall-line simulations after 30 min (blue dashed line) and 4 h (red dashed line) for the ensemble with random initial temperature perturbations at scales of (a) 8 and (b) 128 km. Also plotted at the same times are the ensemble-mean kinetic energy density (solid blue and red lines) and a gray line with $\lambda^{5/3}$ slope approximating background kinetic energy spectrum observed in the atmosphere.

relatively flat. The initial horizontal velocity perturbations are dominated by circulations in small-scale updrafts driven by the three warm bubbles, and as such they have spectra roughly similar to the flat (λ independent) spectrum of a delta function.

The KE' spectra at hour 4 in Fig. 3 have far more power at long wavelengths than that found at similar times in Lorenz's model (Fig. 1). It is difficult to make a detailed comparison of the relationship between KE and KE' in the squall-line simulations with that for the Lorenz model because the background KE in the squall-line simulations grows with time. Nevertheless, by hour 3 the KE spectra in the squall-line simulations have developed a slope of approximately $\lambda^{5/3}$, and one can make a rough comparison with Fig. 1 using a normalized perturbation kinetic energy KE'_n [defined mathematically in the supplement (<http://dx.doi.org/10.1175/BAMS-D-15-00070.2>)] representing the value KE' would have if the ratio KE'/KE were preserved while the actual background KE at a given time and wavelength was remapped to the observed mesoscale KE spectrum. Spectra for KE'_n at 3, 4, and 6 h are plotted as a function of λ in Fig. 5. The wavelengths over which the error becomes saturated clearly propagate upscale. Nevertheless, in contrast to the behavior in Fig. 1 and to that obtained with more advanced spectral turbulence models (Leith and Kraichnan 1972; Métais and Lesieur 1986), the overall error growth does not resemble an upscale cascade in which the error at a given scale remains small until those errors at slightly shorter wavelengths saturate. This does not, however,

imply that the dynamics in such turbulence models are irrelevant, since as demonstrated by Durran and Gingrich (2014), the Lorenz model is capable of producing “up amplitude” growth, similar to that apparent for wavelengths between 30 and 60 km in Fig. 5, when the initial KE' distribution is one with constant relative errors at all scales.

CONCLUSIONS. One cannot immediately characterize the absolute initial errors in our squall-line

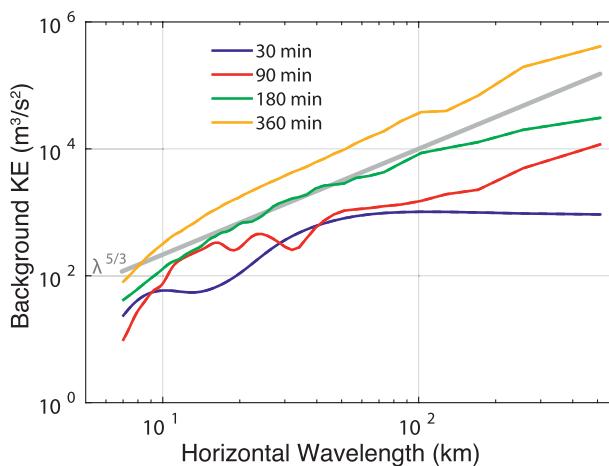


FIG. 4. Evolution of the ensemble-mean kinetic energy density relative to the observed $\lambda^{5/3}$ spectrum (gray line) in simulations with random 8-km-wavelength initial temperature perturbations. Data are plotted at 30 (blue), 90 (red), 180 (green), and 360 (gold) min into the simulations.

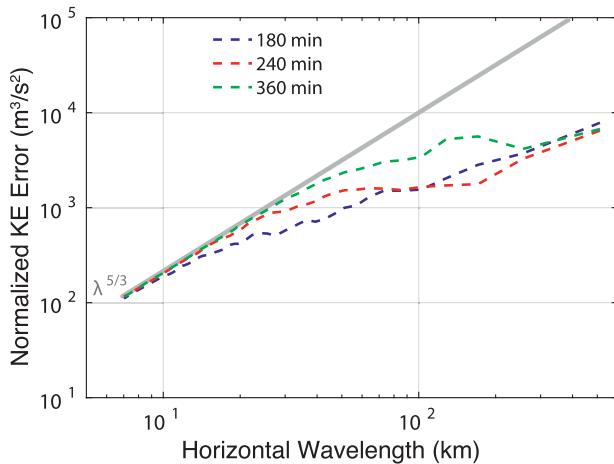


FIG. 5. KE_n' plotted as a function of wavelength for the ensemble of squall-line simulations with random initial temperature perturbations at 8-km wavelength. Spectra are plotted at 3 (blue), 4 (red), and 6 (green) h. The thick gray line shows the background KE spectrum.

simulations as meaningful relative errors because there is no background signal at any finite wavelength in the initial state. Nevertheless, to further compare the relative importance of large- and small-scale errors in these squall-line simulations with that from Lorenz's model, we assume similar error growth would occur if the background temperature spectrum in our simulations matched that given by observations, an assumption supported by the general agreement between the error growth in our results and the HWT forecast experiments. The observed background temperature perturbations (Nastrom and Gage 1985) at wavelengths of 128 km are about 10 times stronger than those at 8 km, which would make the size of the random initial temperature perturbations relative to the background state in the 128-km ensembles one-fortieth as large as those at 8 km. Thus, in qualitative agreement with Lorenz's idealized model, our simulations suggest that after 3 or 4 h, the impact of small relative errors at horizontal scales on the order of 100 km would be similar to that of much larger relative errors at smaller scales. Weather-observing systems and data assimilation strategies for thunderstorm forecasting, therefore, may be most effective at improving forecasts for lead times exceeding 3–4 h if they focus on reducing initial errors at horizontal scales on the order of 100 km, and metrics for evaluating the quality of initial analyses might be best weighted toward larger scales rather than small features, like individual thunderstorm updrafts.

Current operational efforts to forecast weather patterns on small scales, such as over an individual city, are based on the assumption that Lorenz's idealized

model does not apply to the real atmosphere because large-scale motions control the development of many small-scale weather features, including fronts and disturbances produced by mountains or land–sea contrasts (Anthes et al. 1985). Our results together with several recent studies (Nuss and Miller 2001; Bei and Zhang 2007; Durran et al. 2013; Durran and Gingrich 2014; Surcel et al. 2015) suggest this paradigm requires modification. The large scales do indeed appear to exert significant control on small-scale weather, but that control also includes the introduction of the most serious initial errors. Small, but unavoidable, relative errors on the largest scales at which the background kinetic energy spectrum follows a $\lambda^{5/3}$ slope (100–400 km) can rapidly propagate downscale to the smallest resolved features in operational weather forecasts. Errors in those scales subsequently propagate back upscale as if they had simply originated in the small scales. As apparent in Figs. 1–3, there may be no easy way to distinguish such behavior from a situation in which the errors actually originate in the smallest scales.

ACKNOWLEDGMENTS. This research was supported by Grant N000141410287 from the Office of Naval Research.

REFERENCES

- Anthes, R. A., Y.-H. Kuo, D. P. Baumhefner, R. M. Errico, and T. W. Bettge, 1985: Predictability of mesoscale atmospheric motions. *Advances in Geophysics*, Vol. 28, Academic Press, 159–202, doi:10.1016/S0065-2687(08)60188-0.
- Bei, N., and F. Zhang, 2007: Impacts of initial condition errors on mesoscale predictability of heavy precipitation along the Mei-Yu front of China. *Quart. J. Roy. Meteor. Soc.*, **133**, 83–99, doi:10.1002/qj.20.
- Durran, D. R., and J. B. Klemp, 1983: A compressible model for the simulation of moist mountain waves. *Mon. Wea. Rev.*, **111**, 2341–2361, doi:10.1175/1520-0493(1983)111<2341:ACMFTS>2.0.CO;2.
- , and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not important. *J. Atmos. Sci.*, **71**, 2476–2488, doi:10.1175/JAS-D-14-0007.1.
- , P. A. Reinecke, and J. D. Doyle, 2013: Large-scale errors and mesoscale predictability in Pacific Northwest snowstorms. *J. Atmos. Sci.*, **70**, 1470–1487, doi:10.1175/JAS-D-12-0202.1.
- Gage, K. S., 1979: Evidence for a $k^{-5/3}$ law inertial range in mesoscale two-dimensional turbulence. *J. Atmos. Sci.*, **36**, 1950–1954, doi:10.1175/1520-0469(1979)036<1950:EFALIR>2.0.CO;2.

- Hohenegger, C., and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793, doi:10.1175/BAMS-88-11-1783.
- Kain, J. S., and Coauthors, 2010: Assessing advances in the assimilation of radar data and other mesoscale observations within a collaborative forecasting–research environment. *Wea. Forecasting*, **25**, 1510–1521, doi:10.1175/2010WAF2222405.1.
- Leith, C., and R. Kraichnan, 1972: Predictability of turbulent flows. *J. Atmos. Sci.*, **29**, 1041–1058, doi:10.1175/1520-0469(1972)029<1041:POTF>2.0.CO;2.
- Lilly, D., 1983: Stratified turbulence and the mesoscale variability of the atmosphere. *J. Atmos. Sci.*, **40**, 749–761, doi:10.1175/1520-0469(1983)040<0749:STATMV>2.0.CO;2.
- Lindborg, E., 2015: A Helmholtz decomposition of structure functions and spectra calculated from aircraft data. *J. Fluid Mech.*, **762**, R4, doi:10.1017/jfm.2014.685.
- Lorenz, E., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21A**, 289–307, doi:10.1111/j.2153-3490.1969.tb00444.x.
- Métais, O., and M. Lesieur, 1986: Statistical predictability of decaying turbulence. *J. Atmos. Sci.*, **43**, 857–870, doi:10.1175/1520-0469(1986)043<0857:SPODT>2.0.CO;2.
- Nastrom, G., and K. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960, doi:10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2.
- Nuss, W., and D. Miller, 2001: Mesoscale predictability under various synoptic regimes. *Nonlinear Processes Geophys.*, **8**, 429–438, doi:10.5194/npg-8-429-2001.
- Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405, doi:10.1175/BAMS-D-12-00099.1.
- Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz’s model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1076, doi:10.1175/2007JAS2449.1.
- Stratman, D. R., M. C. Coniglio, and S. E. Koch, 2013: Use of multiple verification methods to evaluate forecasts of convection from hot- and cold-start convection-allowing models. *Wea. Forecasting*, **28**, 119–138, doi:10.1175/WAF-D-12-00022.1.
- Surcel, M., I. Zawadzki, and M. K. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *J. Atmos. Sci.*, **72**, 216–235, doi:10.1175/JAS-D-14-0071.1.
- Weisman, M., and J. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504–520, doi:10.1175/1520-0493(1982)110<0504:TDCONSC>2.0.CO;2.
- Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the “surprise” snowstorm of 24–25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632, doi:10.1175/1520-0493(2002)130<1617:MPOTSS>2.0.CO;2.
- , —, and —, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185, doi:10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2.
- , A. M. Odins, and J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166, doi:10.1175/WAF909.1.
- , N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, doi:10.1175/JAS4028.1.