Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: Dependence on background flow and method of perturbation

Aaron Johnson[1,2], Xuguang Wang[1,2]

Ming Xue[1,2], Fanyou Kong[2], Gang Zhao[1,2], Yunheng Wang[2], Kevin W. Thomas[2], Keith A. Brewster[2], and Jidong Gao[3]

[1]School of Meteorology, University of Oklahoma, Norman, Oklahoma
[2]Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma
[3]NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

Submitted to
Monthly Weather Review
June 17, 2013
Revised
Oct 12, 2013

Corresponding author address:
Dr. Xuguang Wang
School of Meteorology
University of Oklahoma
120 David L. Boren Blvd.
Norman, OK, 73072
xuguang.wang@ou.edu

Abstract

Multiscale convection-allowing precipitation forecast perturbations are examined for two forecasts and systematically over 34 forecasts out to 30-h lead time using Haar Wavelet decomposition. Two small scale initial condition (IC) perturbation methods are compared to the larger scale IC and physics perturbations in an experimental convection-allowing ensemble.

For a precipitation forecast driven primarily by a synoptic scale baroclinic disturbance, small scale IC perturbations resulted in little precipitation forecast perturbation energy on medium and large scales, compared to larger scale IC and physics (LGPH) perturbations after the first few forecast hours. However, for a case where forecast convection at the initial time grew upscale into a Mesoscale Convective System (MCS), small scale IC and LGPH perturbations resulted in similar forecast perturbation energy on all scales after about 12h. Small scale IC perturbations added to LGPH increased total forecast perturbation energy for this case. Averaged over 34 forecasts, the small scale IC perturbations had little impact on large forecast scales while LGPH accounted for about half of the error energy on such scales. The impact of small scale IC perturbations was also less than, but comparable to, the impact of LGPH perturbations on medium scales. On small scales, the impact of small scale IC perturbations was at least as large as the LGPH perturbations. The spatial structure of small scale IC perturbations affected the evolution of forecast perturbations, especially at medium scales. There was little systematic impact of the small scale IC perturbations when added to LGPH. These results motivate further studies on properly sampling multi-scale IC errors.

1. Introduction

Limited predictability of warm season precipitation forecasts has been demonstrated by low deterministic forecast skill (Fritsch and Carbone 2004), theoretical arguments (Thompson 1957; Lorenz 1963), sensitivity to small perturbations (e.g., Hohenegger et al. 2006, Hohenegger and Schär 2007a, 2007b, Zhang et al. 2003, 2006), and sensitivity to model and physics differences (e.g., Zhang and Fritsch 1988; Zhang et al. 2006; Johnson et al. 2011ab; Johnson and Wang 2012, 2013). The ability to resolve small scale features associated with rapid non-linear error growth limits the predictability of convection-scale forecasts even more than that of coarser resolution forecasts (Elmore et al. 2002; Walser et al. 2004; Hohenegger et al. 2006; Hohenegger and Schär 2007a, 2007b; Zhang et al. 2003, 2006). Predictability studies at convection-allowing[1] resolution have been limited to a small number of forecasts, rather than systematic evaluation over a period of many forecasts.

Understanding perturbation growth is important for ensemble design because ensemble perturbations are intended to sample the error growth leading to forecast uncertainty (Leith 1974; Toth and Kalnay 1997). The optimal design of Storm Scale (≤24 h forecasts with 1-4 km grid spacing) Ensemble Forecast (SSEF) systems remains largely unknown, although coarser resolution ensembles have been relatively well studied. For example, medium range (~1 week) synoptic scale (~100 km grid spacing) ensembles have been studied for almost two decades (Buizza and Palmer 1995; Toth and Kalnay 1997; Houtekamer et al. 1996; Wang and Bishop 2003; Wang et al. 2004). Short-range (~1-3 days) mesoscale (~10-20 km) ensembles have also been the focus of many past studies (Du et al. 1997; Stensrud et al. 1999; Marsigli et al. 2001; Xu et al. 2001; Grimit and Mass 2002; Eckel and Mass 2005; Lu et al 2007; Li et al. 2008;

[1] Equivalently called convection-permitting or cloud-system resolving in other published studies

Berner et al. 2011). However, the optimal design of SSEFs may be quite different than that of coarser resolution ensembles (Hohenegger and Schär 2007b).

Hohenegger and Schär (2007a) found similar convection-allowing precipitation forecast sensitivity to different perturbation methods after about 11 hours for a case study. However, it is not known if these results are characteristic of other cases with different background flow and/or a different role of topography. Other studies have demonstrated large differences in predictability for different events. For example, Zhang et al. (2006) showed reduced sensitivity to small scale IC perturbations for a warm season heavy precipitation event compared to a cold season large scale cyclone event. Walser et al. (2004) and Hohenegger et al. (2006) further found that some warm season cases in the Alpine region characterized by stratiform precipitation exhibited greater predictability than some cases characterized by deep moist convection. However, it was also found that deep convective cases can exhibit higher predictability, depending on other factors such as the presence of topography and the residence time of the perturbations in convectively unstable regions. Done et al. (2012) have also related different aspects of predictability on two case studies to whether convection is in statistical equilibrium with large scale forcings.

The evolution of different types of perturbations has yet to be systematically studied over a period of many convection-allowing forecasts. The present study systematically evaluates the characteristics of evolution of different perturbations for 34 forecasts. Two case studies are also evaluated in detail to expand on the types of flow regimes considered in past case studies. In contrast to the Mesoscale Alpine Program cases studied by Walser et al. (2004) and Hohenegger et al. (2006), this study focuses on the Great Plains of the United States where topography plays a less dominant direct role, severe convective weather is more frequent and intense (Brooks et al. 2003), and the latitude is farther south from the main belt of the westerlies.

Given the range of resolvable scales at convection-allowing resolution, the growth and interaction of perturbations on different scales is of particular interest. Multiscale evolution of convection-allowing forecast perturbations have been studied on even fewer cases than predictability in general (Zhang et al. 2003, 2006; Walser et al. 2004; Luo and Zhang 2011). The present study focuses on evaluating the characteristics and evolution of forecast perturbations by decomposing them into multiple scales using a Harr wavelets analysis method.

A few additional deterministic forecasts were generated by the Center for Analysis and Prediction of Storms (CAPS) during the 2010 National Oceanographic and Atmospheric Administration Hazardous Weather Testbed (NOAA HWT) Spring Experiment (Kong et al. 2010, Xue et al 2010, Clark et al 2012) to complement the CAPS Spring Experiment real-time SSEF. The general design of CAPS SSEF did not include small scale IC perturbations. These additional forecasts were therefore designed to study the sensitivity to small scale IC perturbations. The present study has three main goals. The first goal is to determine the forecast sensitivity to small scale IC perturbations, relative to the larger scale IC and physics perturbations already included in the SSEF design. The second goal is to compare the sensitivity to two methods of generating such small scale IC perturbations. The third goal is to explore the impact of adding small scale IC perturbations on top of the existing large scale IC and physics perturbations. These goals are addressed using two case studies with different background flows and systematic evaluation of all 34 available cases. Since the existing method of perturbation actually includes multiple perturbation sources (IC and physics), additional forecasts were later generated for the two case studies, with the physics perturbations excluded, to aid interpretation of the results and better understand the impact of IC perturbations at various scales.

The paper is organized as follows. In section 2 the model configuration, scale decomposition, and perturbation methods are described. A brief overview of the two cases that are selected for detailed study is given in section 3 and results are presented in section 4. Section 5 contains conclusions and a discussion.

2. Model configurations and methods

a. Control forecast configuration

Forecasts were generated with 4 km grid spacing at 0000 UTC on 34 weekdays from 3 May to 18 June during the 2010 NOAA HWT Spring Experiment (Xue et al. 2010; Kong et al. 2010). The control forecast used the Weather Research and Forecasting (WRF) model Advanced Research WRF (ARW; Skamarock et al. 2005). The control forecast ICs were obtained from the operational National Centers for Environmental Prediction's North American Model (NCEP NAM) 0000 UTC NAM Data Assimilation System (NDAS; Rogers et al. 2009) analysis at 12 km grid spacing, interpolated to the 4 km WRF grid. Additional radar and mesoscale observations were then assimilated using ARPS 3DVAR and cloud analysis package (Xue et al. 2003; Gao et al. 2004; Hu et al. 2006). Radial velocity from over 120 radars in the Weather Surveillance Radar (WSR)-88D network, as well as surface pressure, horizontal wind, potential temperature, and specific humidity from the Oklahoma Mesonet, METAR (Meteorological Aviation Report), and Wind Profiler networks were assimilated with ARPS 3DVAR. The ARPS cloud analysis package uses radar reflectivity along with surface data, Geostationary Operational Environmental Satellite (GOES) visible and 10.5 micron infrared data to estimate hydrometeor species and adjust in-cloud temperature and moisture (Hu et al., 2006). The control forecast was configured with the Thompson et al. (2008) microphysics scheme, the Mellor-Yamada-Janic (Janjic´ 1994) boundary layer scheme, the Rapid Radiative Transfer Model longwave radiation scheme (Mlawer et al. 1997), the Goddard shortwave radiation (Tao et al. 2003) scheme and the NCEP-Oregon State University-Air Force-NWS Office of Hydrology (NOAH; Ek et al. 2003) land surface model. The vertical turbulent mixing was represented in the boundary layer scheme and sub-grid scale horizontal turbulence mixing was represented by Smagorinsky parameterization. No additional numerical diffusion was applied.

b. Forecast perturbation methods

In the general design of the SSEF during the 2010 HWT Spring Experiment, perturbations that sample model and physics uncertainty as well as IC and Lateral Boundary Condition (LBC) perturbations derived from the Short Range Ensemble Forecast system (SREF; Du et al. 2009) are included. Since the SREF was run at grid spacing of 32-45 km (corresponding to a wavelength of 64-90 km, Du et al. 2009), SREF perturbations are on scales much larger than the SSEF model resolution. Thus the perturbations from SREF do not include small scales (i.e., order of tens of kilometers). Methods to generate perturbations on multiple scales, ranging from the synoptic to the convective scales, have yet to be systematically studied. As a first step to help guide development of practical methods of sampling errors across multiple scales in a SSEF system, during the 2010 Spring Experiment additional forecasts were generated with small scale IC perturbations. For each perturbation method described below, one perturbed deterministic forecast was generated and compared to the control member.

Six methods of perturbation are investigated in this study. Perturbations RAND (random) and RECRS (recursive filter) are designed to simulate random small scale errors in the initial state. Perturbation LGPH (large scale and physics) is designed to simulate the medium and large scale (i.e., order of hundreds and thousands of kilometers, respectively) IC errors and model physics errors. Perturbation LGPH is what is currently adopted in the standard CAPS SSEF system. Perturbation LGPH_RECRS (large scale and physics with recursive filter) is a combination of the LGPH and RECRS perturbation methods. This method is designed to explore the impact of adding small scale IC perturbations on top of the existing large scale IC and physics perturbations. For the two case studies, two additional perturbations are evaluated. Perturbations LG and LG_RECRS are identical to LGPH and LGPH_RECRS, respectively, except without any physics differences from the control member. The latter two methods are designed to better understand the impact of IC perturbations at various scales and to infer the impacts of using different physics parameterization schemes on the results. Since the primary goal of the study is a comparison of small scale IC perturbations to the LGPH perturbations in the current CAPS SSEF design, LG and LG_RECRS are not generated for all forecasts and are only limited to the two case studies to facilitate understanding of the results.

The RAND perturbation is obtained by adding spatially uncorrelated, Gaussian random numbers to the IC temperature and relative humidity (standard deviation of 0.5 K and 5%, respectively). The RECRS perturbation is obtained similarly, except with a recursive filter applied to the random perturbations to create spatially correlated perturbations with a 12 (3) km horizontal (vertical) de-correlation scale. The RAND perturbation is conceptually similar to the random perturbations of Hohenegger and Schär (2007a). The RECRS perturbation is conceptually similar to the Gaussian perturbation of Hohenegger and Schär (2007a), except RECRS is applied homogenously across the domain instead of only at a single location.

The LGPH IC perturbation is obtained from the difference between a 3 hour forecast of a SREF WRF-ARW member (labeled P1 in Du et al. 2009) and the corresponding SREF control member forecast. The SREF perturbations of u and v wind components, potential temperature, and specific humidity are rescaled to have a root mean square value of 1 m s^{-1}, 0.5 K, and 0.02 g/kg, respectively. In addition to the IC perturbation, the LGPH forecast uses a different physics configuration than the control forecast to approximate physics errors. Unlike the control forecast, the LGPH perturbation uses Morrison et al. (2008) microphysics scheme, RUC land surface model (Benjamin et al. 2004) and Yonsei University (Noh et al. 2003) boundary layer scheme. The LGPH_RECRS perturbation is identical to LGPH except with additional recursive filtered random perturbations added in the same way as for the RECRS perturbation.

Although only temperature and humidity (and wind in the case of LGPH and LGPH_RECRS) are directly perturbed, results are evaluated in terms of precipitation differences. Thus, the focus is on the net effect, rather than the processes, of perturbation growth and evolution.

c. Scale decomposition method

Following Casati et al. (2004), precipitation fields are decomposed into components of different spatial scale using 2D Haar Wavelets with the Model Evaluation Tools package from the Developmental Testbed Center, available at http://www.dtcenter.org/met/users. The decomposition is defined over a 2^n by 2^n grid point domain for $n>1$. The original field is decomposed into its component on each of $n+1$ scales, and is equal to the sum of its components. The i^{th} component can be calculated as the difference between the original field averaged in boxes of 2^{i-1} by 2^{i-1} grid points and the original field averaged in boxes of 2^i by 2^i grid points for $1 \leq i \leq n$. The $(n+1)^{th}$ component is the domain average value. Each component therefore represents the variation over a spatial scale of $4 * 2^{i-1}$ km from a larger scale average. Analogous to the more familiar Fourier decomposition, in the rest of the paper the wavelet-decomposed spatial scales are referred to in terms of a corresponding wavelength. Thus, for example, the smallest resolvable scale of 4 km (e.g., Fig. 1b) corresponds to the smallest resolvable wavelength of 8 km. A verification domain (shown in Fig. 3) of 512 by 512 grid points (2048 by 2048 km) within the larger forecast domain (shown in Fig. 2) of 1163 by 723 grid points (4652 by 2892 km) is used in this study. Further details of the wavelet decomposition are described in Casati et al. (2004). Precipitation forecast energy is defined as the square of the one-hour accumulated precipitation field, averaged over the verification domain. The energy on a particular scale is defined similarly, using only the component of the precipitation field on that scale. The error (or perturbation) energy is the square of the precipitation field difference between a forecast and the observations (or control forecast). The evolution of a perturbation, or difference, energy metric is a common method of quantifying sensitivity to forecast perturbations (e.g., Zhang et al. 2006, Hohenegger et al. 2006).

Figure 1 illustrates the 2D Haar wavelet decomposition of the difference between the 6h control forecast and corresponding observation of hourly accumulated precipitation on the 20 May case. The distribution of difference energy across scales is also found in Fig. 9 (dashed cyan line). Objectively, there is a maximum of difference energy at 32-64 km wavelength scales and a smaller secondary maximum at the 256 km scale (Fig. 9). The total difference field (Fig. 1a) subjectively looks most similar to the difference fields on 32-64 km scales (Fig. 1d,e), suggesting that the high amplitude, small-scale features on these scales account for most of the total difference. The subjectively apparent displacement of the MCS in Oklahoma and Arkansas (Fig. 1a) also corresponds to increased energy on the 256 km scale (Figs. 1g and 9).

For presentation of results we define the large scale as the sum of scales with wavelengths of 4096, 2048, and 1024 km, the medium scale as the sum of scales with wavelengths of 512, 256, 128 km and 64 km and the small scale as the sum of scales with wavelengths of 32, 16, and 8 km. The small scales are those that are too small to be represented with the current SREF-derived perturbations.

3. Case study overview

Forecasts initialized at 0000 UTC on 10 and 20 May 2010 (hereafter, 10 May case and 20 May case, respectively) are selected for comparison of the differences in precipitation forecasts resulting from different sources and scales of perturbations during different flow regimes. The following sub-sections describe the reasons for selecting these cases, a synoptic scale overview of environmental conditions, and the evolution of the unperturbed control forecast.

a. 10 May 2010

The 10 May case is selected because a synoptic scale baroclinic disturbance generated widespread precipitation in the control forecast. During much of the forecast period (i.e., the first ~24h), the forecast evolution was determined primarily by large scale influences (e.g., fronts, jets and temperature advection). This event is also of interest because of a significant tornado outbreak that occurred in the southern Plains on the afternoon of 10 May (e.g., Palmer et al. 2011). The control forecast of this event contains substantial errors in comparison to observations. It is therefore of interest to examine the sensitivity of the control forecast to different types of perturbations.

At the time of forecast initialization (0000 UTC 10 May) there was an embedded shortwave trough over the western US and a broad ridge over the central US aloft (Fig. 2a). There was southerly flow and a warm front in central Texas at the surface (Fig. 2b). By 0000 UTC the negatively-tilted shortwave had propagated to the central US, inducing surface cyclogenesis and an intersecting dryline, cold front and warm front in the southern Plains (Fig. 2c,d,e,f). An initial wave of observed scattered showers associated with the low-level warm advection developed in Arkansas and Missouri by 0600 UTC and moved eastward into Tennessee and northern Alabama by 1800 UTC (Fig. 3b,d,f). Convection also developed near the Kansas/Nebraska border by 1200 UTC, moving eastward into northern Missouri by 1800 UTC (Fig. 3d,f). At 0000 UTC more intense convection was occurring in the southern Plains.

The control forecast predicted the initial wave of scattered showers, although with a southwestward displacement and with greater intensity than observed (Fig. 3a,b,c), as well as the development of convection along the Kansas-Nebraska border, although with more linear organization, weaker intensity and a slight northward displacement (Fig. 3c,d). The most prominent difference between the forecast and observation was the absence of the intense convection over the southern Plains at 0000 UTC (Fig. 3g). Storms eventually developed in the control forecast but they were several hours slower to develop than observed and did not extend as far south as observed (not shown).

b. 20 May 2010

In contrast to the 10 May case, the 20 May case is selected because early in the control forecast (i.e., first ~12h) an MCS grew upscale from initially smaller scale convection. The MCS evolution then influenced the regional-scale characteristics of subsequent convection, for example through the strength and location of its surface cold pool outflow boundary.

At the time of forecast initialization (0000 UTC 20 May 2010) there was a slow-moving, broad trough aloft with an embedded shortwave rounding its base (Fig. 4a,c,e). At the surface, a weaker surface low than in the 10 May case propagated from central Oklahoma into western Missouri between 0000 UTC 20 May and 0000 UTC 21 May without substantial intensification (Fig. 4b,d,f). By 0600 UTC 20 May, cellular convection from the previous evening (Fig. 5b) was organizing into an MCS in eastern Oklahoma, Arkansas and Missouri that was dissipating by 1200 UTC (Fig. 5d,f). The remnant outflow boundary was the focus for additional convection that developed the following afternoon (Fig. 5h,j). Stratiform precipitation also developed by 1200 UTC, from southeastern Nebraska to southeastern Missouri (Fig. 5f,h), weakening later in the day (Fig. 5j).

The control forecast reflects the upscale organization and intensification of convection, subsequent dissipation of the MCS, development of stratiform precipitation and regeneration of convection the following afternoon (Fig 5a,c,e,g,i). However, the forecast MCS evolved a different structure than the observed MCS (Fig. 5c,e). The coverage, timing and location of subsequent convection along the remnant outflow boundary was also qualitatively different than observed (Fig. 5g,i).

4. Characteristics of perturbation growth

The characteristics of the precipitation forecast perturbation evolution are evaluated using the change in perturbation energy with time in total and on the small, medium and large scales as well as the change in perturbation energy with spatial scale for selected fixed times. When perturbations are related to the background flow, the background flow refers to the control forecast upon which the perturbations were added, which may be different than the observations. Precipitation observations are from the National Severe Storm Laboratory Q2 product (Zhang et al. 2011).

An optimal ensemble design should contain members that are equally plausible, and therefore equally skillful (Leith 1974). Although lower skilled members can add value to an ensemble (Eckel and Mass 2005) and this study focuses on forecast sensitivity rather than skill, the impact of the perturbations on forecast skill should also be considered when designing an ensemble system. Among the forecasts evaluated systematically in this study, only the physics perturbations at some lead times (~2-5h and ~22-27h) and the RECRS perturbations during the first hour resulted in significant decreases in skill compared to the control forecast (not shown). The differences in skill resulting from physics perturbations are in large part related to differences in forecast bias resulting from the use of different physics schemes. How to optimally sample model and physics error is still an open research question for SSEF design. The inclusion of LG and LG_RECRS perturbations in the case studies helps to understand the impacts of forecast biases resulting from different physics schemes and the sensitivity to IC perturbations of various scales. The early loss of skill resulting from recursive filter perturbations is a result of spurious precipitation that formed over large areas on many cases (not shown). This is clearly not desirable in an ensemble and it is suggested that the spatial scales and amplitude of such perturbations should be more carefully studied before this perturbation method is used for ensemble forecasting.

The following case studies and season-average results address the three research goals stated in section 1 by a comparison of LGPH (and LG) with RAND and RECRS, a comparison of RAND with RECRS, and a comparison of LGPH_RECRS (and LG_RECRS) with LGPH (and LG).

a. 10 May 2010

For the 10 May case, the control forecast error energy shows maxima in forecast error energy at lead times of about 10-15h and 24-27h (Fig. 6d). The general trend of two error energy maxima superimposed on an overall increasing trend is found on all scales (Fig. 6a-c). The magnitude of error energy is an order of magnitude greater on the medium and small scales than on the large scales. Compared to the control forecast error energy, the perturbation energy for most lead times and methods is too small in magnitude (Fig. 6). In general, those perturbations involving LG (i.e., LGPH, LGPH_RECRS, LG, LG_RECRS) capture about half of the total error energy while small scale IC perturbations (i.e., RAND and RECRS) capture about one quarter of the total error energy. A particularly pronounced absence of medium scale perturbation energy with a scale of about 64-256 km at 24h for all perturbation methods shown in Fig. 7, compared to forecast error, is consistent with Fig. 3. The medium scale storms in the southern Plains at this time (Fig. 3h) are absent in the corresponding forecast (Fig. 3g), contributing to the medium scale forecast error energy. However, all perturbation methods also missed these storms (not shown) so the perturbation energy does not reflect that particular forecast error. In addition, as shown in Fig. 7, compared to the control forecast error energy, the perturbation energy for most lead times and methods is also too small in the spatial scale of maximum energy except for LGPH and LGPH_RECRS at 24h.

On the 10 May case the evolution of perturbation energy on different scales depends strongly on the method of perturbation. Compared to LGPH, which is currently used in the standard CAPS SSEF, RAND and RECRS show less pronounced growth for large and medium scales, but comparable growth for small scales (Fig. 6). Without physics perturbations, LG perturbation energy is less than LGPH on medium and large scales at later lead times (Fig. 6). However, the qualitative comparison of RAND and RECRS to LGPH is consistent with the comparison to LG. Between the two small scale perturbation methods, RECRS shows an increase of perturbation energy over RAND on the medium scales and on the small scales after ~20h (Fig. 6b,c). When small scale IC perturbations are combined with LGPH and LG, LGPH_RECRS and LG_RECRS are similar to LGPH and LG, respectively (Fig. 6).

The characteristics of perturbation growth are also seen in the perturbation energy spectra at selected lead times (Fig. 7). None of the perturbation methods generates much energy during the first 6h. The perturbation method affects both the spectral width and the wavelength of maximum energy of the resulting precipitation forecast perturbation. For example, at 12h the wavelength of maximum energy of 32 km for RAND (Fig. 7a) is smaller than the 64 km for LGPH (Fig. 7c) and RECRS (Fig. 7b). The LGPH spectrum after 6h is broader than the spectra for RAND and RECRS (Fig. 7a,b,c), indicating perturbations across a wider range of scales in LGPH. The RECRS spectrum is also broader than the RAND spectrum (Fig. 7a,b). When combining the small scale IC perturbation with LGPH, the wavelength of maximum energy for LGPH_RECRS after 6h (Fig. 7d) tends to be larger than LGPH (Fig. 7c). However, such difference was not observed for the combined small and large scale IC-only perturbations (LG and LG_RECRS; Fig. 7e,f).

The perturbations involving RECRS (i.e., RECRS, LGPH_RECRS and LG_RECRS) show perturbation energy maxima at 16-32 km wavelength at 1h (Fig. 7). Such maxima correspond to the spurious small scale precipitation mentioned above. This spurious precipitation may be a result of adding unrealistically large perturbations on such scales, a lack of realistic coupling between the temperature and moisture perturbations, or some other imbalance resulting from the temperature and humidity perturbations in RECRS. The lack of spurious precipitation in the RAND perturbations may be a result of diffusion quickly reducing the amplitudes of the small scale perturbations when the perturbations are of grid scale.

In summary, for the 10 May case the perturbation methods considered, especially small scale IC perturbations, do not reflect the forecast error magnitude or temporal variability. The shape of the perturbation energy spectrum also does not reflect the shape of the forecast error energy spectrum for many lead times and perturbation methods. Compared to the standard LGPH perturbation in CAPS SSEF, RAND and RECRS show less perturbation growth at medium and large scales, resulting in narrower perturbation energy spectra with a smaller wavelength of maximum energy at some lead times. When the physics perturbation is eliminated from LGPH, LG generally has less perturbation energy than LGPH at later lead times on medium and large scales. However, the smaller perturbation growth by RAND and RECRS at medium and large scales is also seen compared to LG. The difference between RAND and RECRS is mainly on medium scales where perturbation energy is increased for RECRS, resulting in a broader spectrum at some lead times. The impact of adding small scale IC perturbations to LGPH and LG is generally small. The relative lack of medium and large scale forecast perturbations in RAND and RECRS compared to LG and LGPH, and the minimal impact of combining small and large scale perturbations, suggests a relative insensitivity of this forecast at such scales to random small scale IC perturbations compared to larger scale perturbations such as LGPH and LG. As shown below, this result is case-dependent.

b. 20 May 2010

As in the 10 May case, the 20 May case shows forecast error energy with a maximum at early lead times followed by a larger maximum at ~24-27h (Fig. 8d). The error energy on 20 May does not show an increasing trend as clearly as on the 10 May case. This may be due to the already much larger error energy on the 20 May case than on the 10 May case at early lead times, especially on small and medium scales (Fig. 8b,c). Although the error energy during the first maximum is again under-represented by the forecast perturbations, the perturbation energy follows the error energy more closely on this case during the second maximum than on the 10 May case. Compared with the 10 May case where all perturbation methods generate maximum error energy on smaller scales than the forecast error energy during the first 12h, only a few perturbation methods (RAND, LG, RECRS) fail to capture the error energy maximum wavelengths at some lead times (Fig.9). By 24h, all perturbation methods reflect the maximum error energy on the 64 km wavelength scale on the 20 May case.

The evolution of perturbation energy on the 20 May case is generally less dependent on the method of perturbation than on the 10 May case. There is not a consistent separation between LGPH and RAND/RECRS on medium and large scales during most of the forecast period (Fig. 8a,b). RAND and RECRS have even more perturbation energy than LGPH on small scales at ~20-27h. Eliminating the impact of physics perturbations, small scale (RAND/RECRS) and large scale (LG) IC perturbations have similar perturbation energy (Fig. 8). During the early forecast hours RECRS has more perturbation energy than RAND on small and medium scales (Fig., 8b,c). In contrast to the 10 May case, this difference diminishes and RAND and RECRS become similar by ~10-12h. Combining the small scale IC perturbation with LGPH also shows a larger impact compared to the 10 May case. In particular, LGPH_RECRS shows greater perturbation energy than LGPH at early lead times on small scales (Fig., 8c), most lead times on medium scales (Fig., 8b), and at the 1h lead time, corresponding to regional variation in the spurious precipitation response to RECRS, on large scales (Fig., 8a). These differences are even more pronounced when only the IC perturbations are considered (i.e., LG_RECRS vs. LG).

The impact of physics perturbation is also evaluated by comparing LG and LGPH. The differences between LG and LGPH are most pronounced on medium scales at early lead times and small scales at later lead times for this case (Fig. 8b,c). On the medium scales LG and LGPH become similar after ~15h, suggesting that medium scale forecast sensitivity is dominated by the IC, rather than physics, perturbations at later lead times. LGPH_RECRS energy is also less than RECRS alone at 1h for large and small scales (Fig. 8). Since LG_RECRS is more similar to RECRS at 1h, this seemingly counter-intuitive result is due to a damping effect of the LGPH physics configuration which is different from that used in RECRS. The physics configuration of LGPH showed less systematic bias in the LGPH forecast than RAND and RECRS forecasts at these lead times (not shown). It is not clear whether this damping effect is related to the differences in microphysics or boundary layer parameterization. This also explains why LG is more similar to RECRS/RAND than LGPH for small scales at later lead times (Fig. 8c).

The perturbation energy spectra are also generally less sensitive to the perturbation method on 20 May than on 10 May (Fig. 9). A prominent difference from 10 May is that in the

20 May case the small scale IC perturbations energy grow substantially, creating total energy that is similar or greater to larger scale IC (with or without physics) perturbations (Fig. 9). At 6 and 12h the wavelength of maximum energy for LGPH and LGPH_RECRS is again larger than for RAND and RECRS (Fig., 9a,b,c,d). This difference is largely due to the physics perturbations since the LG and LG_RECRS spectra (Fig. 9e,f) at these times are more similar to the RAND and RECRS spectra. The differences between RAND and RECRS spectra are more pronounced in the first 6h due to the spurious precipitation. Combining the small scale RECRS perturbation with the large scale IC and physics perturbations slightly broadens the spectra (i.e., LGPH vs. LGPH_RECRS and LG vs. LG_RECRS).

The different sensitivities of the 10 and 20 May cases to different perturbations are illustrated subjectively with representative RAND, LGPH and LG forecast perturbations at the 24h lead time (Fig. 10). On 10 May it is primarily the convective scale details of an incipient MCS over southeast Kansas, and the small scale features within the stratiform precipitation farther north that are substantially affected by the RAND perturbation (Fig. 10a). However, the LGPH perturbation alters the mesoscale structure of the stratiform precipitation region, and more dramatically changes the structure and location of the incipient MCS which is displaced ~100 km to the northwest (Fig. 10b). The LG perturbation shows a similar displacement, although the amount of displacement in the forecast perturbation by LG is less than that by LGPH and in the opposite direction (Fig., 10c). In contrast, even the mesoscale characteristics and location of the MCS forecast over the southern part of the domain on 20 May are substantially changed by the RAND perturbation (Fig. 10d) at least as much as the LG and LGPH perturbations (Fig. 10e,f).

In summary, the perturbation energy is again smaller than the error energy at early lead times but, unlike the 10 May case, is similar to the error energy after ~15h. Unlike the 10 May case, RAND/RECRS show similar or greater energy compared to LGPH. The distribution of perturbation energy across spatial scales is generally more similar among the different perturbation methods on this case than on the 10 May case. Also in contrast to the 10 May case, combining the small scale IC perturbations with larger scale IC and physics perturbations shows a clear impact in terms of the magnitude of perturbation energy growth. These results suggest that small scale IC errors on this case contribute to the forecast uncertainty at least as much as the larger scale IC and physics errors. Therefore, adding small scale IC perturbations to the larger scale IC and physics perturbations may be advantageous to the SSEF design in certain situations.

c. Season average results

On average, the forecast error energy grows approximately linearly on the large scale with much less magnitude than on smaller scales (Fig. 11). On medium and small scales, the forecast error energy follows the diurnal cycle of convection, with maxima during the early forecast hours and during the following afternoon (Fig. 11b,c). The medium scale afternoon maximum of the second day persists into the evening while the small scale maximum decreases after ~23h (i.e., ~2300 UTC; Fig. 11b,c). All perturbation methods result in less total energy than the forecast errors (Fig., 11d). The under-estimation of forecast errors is most pronounced for medium and large scales and for the RAND and RECRS perturbations (Fig., 11a,b,c).

Differences among the average perturbation energies in Fig. 11 are tested for statistical significance using one-sided permutation resampling (Hamill 1999) at the 95% confidence level. On medium and large forecast scales, LGPH has significantly more perturbation energy than RAND and RECRS, except at early lead times due to the spurious precipitation of RECRS and except at 19-24h on the medium scale where the difference between LGPH and RECRS is not significant (Fig. 11a,b). Only LGPH and LGPH_RECRS account for a substantial fraction of the error energy on large scales (Fig. 11a). On small scales LGPH is slightly, but significantly, greater than RAND at 3-9h and is markedly less than RAND and RECRS at 16-30h (Fig. 11c). The reduced LGPH perturbation energy compared to RAND and RECRS on small scales at 16-30h is a systematic result of the physics-related bias difference discussed for the 20 May case. Besides the first few hours, dominated by spurious precipitation for RECRS, significantly greater energy for RECRS than RAND is found at most lead times for large and medium scales and at several lead times for small scales (Fig. 11a,b,c). This difference is qualitatively most pronounced on the medium scales (Fig., 11b). On average, the medium scale differences between LGPH and RAND/RECRS are less pronounced than on the 10 May case. The RAND/RECRS medium scale perturbation energy is 50% or more of the LGPH perturbation energy on average at most lead times. This suggests systematic upscale growth of the small scale IC errors throughout the 30h forecast period. However, the differences between LGPH and LGPH_RECRS on average are generally small and/or not significant, again excluding early lead times dominated by spurious precipitation (Fig. 11).

The total average perturbation energy from all perturbation methods becomes similar after ~16h (Fig. 11d), 4h later than the 11h time scale of insensitivity to the small scale IC perturbation method suggested by Hohenegger and Schär (2007a). The differences between RAND and RECRS perturbation energy, especially on the medium scales, throughout the forecast period suggests that the impact of the structure of small scale IC perturbations may persist longer into the forecast than expected.

The RAND and RECRS perturbations do not reflect the spectral evolution of error energy as well as LGPH (Fig. 12a,b,c). LGPH already approximately reflects the error energy maximum of ~32-128 km wavelength by 6h (Fig. 12c). However, RAND and RECRS still do not even reflect the error energy maximum of 64 km wavelength at 12h (Fig. 12a,b). By 24h, all methods reflect the error energy maximum of 32 km wavelength (Fig. 12). At later lead times, LGPH generally has a broader spectrum, with more energy on the larger scales, than RAND and RECRS (Fig. 12a,b,c). Except for the very early lead times where RECRS and LGPH_RECRS are dominated by the spurious precipitation, there are not substantial differences in perturbation energy spectra between RAND and RECRS or between LGPH and LGPH_RECRS.

5. Summary and discussion

The purpose of this study is to understand the multiscale characteristics of the evolution of different sources of perturbations on convection-allowing precipitation forecasts for two case studies and for 34 forecasts on average, for the purpose of guiding the optimal SSEF design. In particular, three main goals are addressed. First, the impact of small scale IC perturbations (RAND and RECRS) is compared to the impact of larger scale IC and physics perturbations (LGPH and LG) that are currently used in the CAPS Spring Experiment SSEF. Second, two methods of generating small scale IC perturbations (RAND and RECRS) are compared to each other. Third, LGPH is compared to a method of combining the small and large scale IC perturbations (LG_RECRS) and combining multiscale IC and physics perturbations (LGPH_RECRS).

It is found that the relative impacts of the different types of perturbation are case-dependent. On the 10 May case the evolution of the precipitation systems in the background forecast are driven primarily by a synoptic scale disturbance. After the first few hours, the 10 May forecasts containing large scale IC perturbations, with or without physics perturbations, have more perturbation energy than the small scale IC-only perturbations, RAND and RECRS, on medium and large scales while the small scale forecast perturbation energy is similar for all methods. As a result, the perturbation energy spectra are generally broader for LG and LGPH than RAND and RECRS. On this case the RECRS method creates more forecast perturbation energy than RAND at most lead times for the medium scales and for many lead times after ~20h for the small scales. LGPH_RECRS and LG_RECRS do not increase the perturbation energy relative to LGPH and LG, respectively, on this case. In contrast, the 20 May case has ongoing convection in the background forecast at the initial time that grows upscale into an MCS. The 20 May forecasts are generally less sensitive to the scale of IC perturbations, with LG and LGPH not showing a clear increase of perturbation energy, relative to RAND and RECRS, on any scale. The perturbation energy spectra are also less sensitive to the perturbation method on 20 May than on 10 May. There is less forecast energy for LGPH than for RAND and RECRS on small scales at ~20-27h due to the physics scheme differences. On 20 May, RECRS shows increased perturbation energy, relative to RAND, for only the first ~12-15h on small and medium scales. Unlike the 10 May case, the 20 May case shows a greater impact of combining small scale IC perturbations with larger scale IC and physics perturbations, with perturbation energy at ~20-26h for LGPH_RECRS and LG_RECRS being larger than LGPH and LG, respectively.

One of the main differences in perturbation evolution between the two cases is the greater sensitivity to the small scale IC perturbations, relative to the larger scale IC and physics perturbations, on the 20 May case. This is consistent with past case studies suggesting that lower predictability generally results from the release of deep moist convective instability (e.g., Hohenegger et al. 2006). However, Zhang et al. (2006) found *less* sensitivity of the mesoscales to small scale random IC perturbations for a warm season heavy precipitation event than a large scale winter cyclone event. This contrasts with the results in the present study. Reasons for this difference may include the direct consideration of precipitation forecasts, instead of wind and temperature differences as in Zhang et al. (2006), as well as differences in the forcing mechanisms of the precipitation systems. For example, our 20 May case is characterized by upscale growth of convection due to internal storm dynamics rather than the large scale moisture transport interacting with topography in Zhang et al. (2006).

The perturbations are evaluated over a large number of forecasts to better understand their systematic behavior, independent of the many factors of individual cases that can affect the predictability. Averaged over 34 forecasts, there is a diurnal cycle of forecast error and perturbation energy on the small and medium scales. Compared to RAND and RECRS, the forecast sensitivity is dominated by LGPH and LGPH_RECRS perturbations on large and medium scales. However, on medium scales RAND and RECRS alone can generate at least half as much forecast perturbation energy as LGPH throughout the forecast period. This sensitivity of the medium forecast scales to small scale IC perturbations is more similar to the 20 May case than the 10 May case. This similarity is consistent with the expectation that during the late spring and early summer season convective episodes are often dominated by localized and/or diurnal forcings, such as those on the 20 May case, rather than large scale forcing like the 10 May case (Stensrud and Fritsch 1993). Also similar to the 20 May case, perturbation energy for LGPH and LGPH_RECRS is systematically reduced on small scales during the diurnal convective maximum due to the different biases of the physics schemes. The most prominent difference between RAND and RECRS is an increase of medium scale perturbation energy at all times for RECRS. RECRS also shows greatly increased energy at 1-2h due to spurious precipitation. Refinement of the RECRS method or investigation of better methods to generate small scale IC perturbations would therefore be necessary before inclusion in an ensemble forecast system. On average, LGPH_RECRS does not create significantly more perturbation energy than LGPH on any scale after the first few hours which are dominated by the spurious precipitation.

The dominant impact of large scale IC and physics perturbations suggests that the current CAPS ensemble configuration, sampling only large scale IC and physics errors, already samples the primary forecast sensitivity. The comparable, although lesser, impact of small scale IC-only perturbations on medium scales also implies a process of upscale growth of the initially small errors that can substantially contribute to the medium scale forecast sensitivity. However, the method of generating multi-scale IC perturbations represented by LGPH_RECRS does not show a systematic increase in medium scale perturbation energy, relative to LGPH. The three most likely reasons for this lack of impact are that (1) better methods of combining multiple scales of IC perturbation need to be developed, (2) there is only an advantage of including small scales in the IC perturbations under certain conditions such as rapid upscale error propagation (e.g., the 20 May case), or (3) the downscale energy cascade of the large scale IC perturbations implicitly accounts for small scale errors that are not explicitly sampled.

More work is needed to understand how to realistically and efficiently sample, and optimally combine, all scales of uncertainty, from synoptic to convective, into IC/LBC perturbations, along with physics perturbations, for SSEFs. The methods of defining the small scale IC perturbations in this study are not flow-dependent, may not reflect the actual analysis errors, and can result in unbalanced initial fields that are detrimental to short term forecasts. For example, the RAND perturbations exhibit no initial spatial structure and result in less growth than the RECRS perturbations. The RECRS perturbations are defined to have a fixed, uniform spatial structure and amplitude but create spurious precipitation at early lead times. The differences between RAND and RECRS, especially on the medium forecast scales, show the importance of the spatial structure of small scale IC perturbations. Flow-dependent methods should be developed to better sample the small scale error structure in the ICs. Future work will investigate the use of ensemble based data assimilation and its variants (e.g., Wang et al. 2008) to provide flow-dependent multi-scale IC perturbations for SSEFs. In addition to IC/LBC perturbation methods, different physics perturbations may also yield different results. Investigation of physics perturbation methods such as using different physics schemes and different parameters within a fixed scheme (Duda et al. 2013) is left for future study. While this study focuses primarily on the spatial scales of forecast perturbation, the questions of which variables should be perturbed and what the covariance should be among the perturbed variables for SSEF design remains an open question. Ensemble-based data assimilation may also be useful to address such questions.

Acknowledgements

This research was supported by NSF grant AGS-1046081 award. Computing resources at the National Science Foundation XSEDE National Institute of Computational Science (NICS) and the OU Supercomputing Center for Education & Research (OSCER) were used for this study. The SSEF data analyzed in this study were produced under the support from the NOAA CSTAR program (NA17RJ1227) and NSF AGS-0802888. The manuscript was greatly improved by the comments of three anonymous reviewers. Jun Du is acknowledged for making SREF data available.

References

Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473–494.

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

Brooks, H. E., J. W. Lee, and J. P. Craven, 2003: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmos. Res*., **67 & 68**, 73–94.

Buizza, R., T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456.

Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141-154.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74.

Coniglio, M. C., J. Y. Hwang, D. J. Stensrud, 2010: Environmental factors in the upscale growth and longevity of MCSs derived from Rapid Update Cycle analyses. *Mon. Wea. Rev.*, **138**, 3514-3539.

Done, J. M., Craig, G. C., Gray, S. L. and Clark, P. A. (2012), Case-to-case variability of predictability of deep convection in a mesoscale model. *Quart. J. R. Meteor. Soc*., **138**: 638–648.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125,** 2427-2459.

Du, J., and Coauthors, 2009: NCEP Short-Range Ensemble Forecast (SREF) system upgrade in 2009. Extended Abstracts, *19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting*, Omaha, NE, Amer. Meteor. Soc., 4A.4. [Available online at http://ams.confex.com/ams/23WAF19NWP/techprogram/paper_153264.htm].

Duda, J., X. Wang, F. Kong and M. Xue, 2013: Using Varied Microphysics to Account for Uncertainty in Warm-Season QPF in a Convection-Allowing Ensemble. *Mon. Wea. Rev.* , submitted.

Eckel, F. A., C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.

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

Elmore, K. L., D. J. Stensrud, K. C. Crawford, 2002: Explicit cloud-scale models for operational forecasts: A note of caution. *Wea. Forecasting*, **17**, 873–884.

Fritsch, J. M., R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.,* **85**, 955–965.

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Ocean. Tech.*, **21,** 457-469.

Grimit, E. P., C. F. Mass, 2002: Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Wea. Forecasting*, **17**, 192–205.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167.

Hohenegger, C., D. Lüthi, C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Mon. Wea. Rev.*, **134**, 2095–2107.

Hohenegger, C., C. Schär, 2007a: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478.

Hohenegger, C., C. Schär, 2007b: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793.

Houtekamer, P. L., L. Lefaivre, J. Derome, H. Ritchie, H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.

Hu, M., M. Xue, and K. Brewster, 2006: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.*, **134,** 675-698.

Janjic´, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Johnson, A., X. Wang, F. Kong, and M. Xue, 2011a: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part I: Development of object-oriented cluster analysis method for precipitation fields. *Mon. Wea. Rev.*, **139**,3673-3693.

Johnson, A., X. Wang, M. Xue, and F. Kong, 2011b: Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring

Experiment. Part II: Season-long ensemble clustering and implication for optimal ensemble design. *Mon. Wea. Rev.*, **139**, 3694-3710.

Johnson, A. and X. Wang, 2012: Verification and calibration of neighborhodd and object-based probablistic precipitation forecasts from a multi-model convection-allowing ensemble. *Mon. Wea. Rev.*, **140**, 3054-3077

Johnson, A. and X. Wang, 2013: Object-based evaluation of a storm scale ensemble during the 2009 NOAA Hazardous Weather Testbed Spring Experiment. *Mon. Wea. Rev.*, **141**, 1079-1098.

Kong, F., M. Xue, K. W. Thomas, Y. Wang, K. Brewster, X. Wang, J. Gao, S. J. Weiss, A. Clark, J. S. Kain, M. C. Coniglio, and J. Du, 2010: Evaluation of CAPS multi-model storm-scale ensemble forecast for the NOAA HWT 2010 Spring Experiment. *24th Conf. Wea. Forecasting/20th Conf. Num. Wea. Pred.*, Amer. Meteor. Soc., P452.

Leith, C. E., 1974: Theoretical skill of monte carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Li, X., M. Charron, L. Spacek, G. Candille, 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443–462.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

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

Lu, C., H. Yuan, B. E. Schwartz, S. G. Benjamin, 2007: Short-range numerical weather prediction using time-lagged ensembles. *Wea. Forecasting*, **22**, 580–595.

Luo, Y., L. Zhang, 2011: A case study of the error growth and predictability of a Meiyu frontal precipitation event. *Acta Meteorological Sinica*, **25**:4, 430-440.

Maddox, R. A., 1983: Large-scale meteorological conditions associated with midlatitude, mesoscale convective complexes. *Mon. Wea. Rev*., **111**, 1475–1493.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. II: Limited area experiments in four Alpine flood events. *Quart. J. Roy. Meteor. Soc.,* **127,** 2095-2115.

McAnelly, R. L., J. E. Nachamkin, W. R. Cotton, M. E. Nicholls, 1997: Upscale evolution of MCSs: Doppler radar analysis and analytical investigation. *Mon. Wea. Rev.*, **125**, 1083-1110.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res*., **102**, 16,663-16,682.

Morrison, H., and W. W. Grabowski, 2008: A novel approach for representing ice microphysics in models: Description and tests using a kinematic framework. *J. Atmos. Sci*., **65**, 1528–1548.

Noh, Y., W. G. Cheon, S. Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.,* **107,** 421–427.

Palmer, R. D., and Coauthors, 2011: Observations of the 10 May 2010 tornado outbreak using OU-PRIME: Potential for new science with high-resolution polarimetric radar. *Bull. Amer. Meteor. Soc.*, **92**, 871–891.

Rogers, E., and Coathors, 2009: NCEP North American mesoscale modeling system: Recent changes and future plans. Extended Abstracts, *19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting*, Omaha, NE, Amer.

Meteor. Soc., 2A.4. [Available online at http://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154114.htm].

Schwartz, Craig S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the advanced research WRF version 2. NCAR Tech Note NCAR/TN-468_STR, 88 pp. [Available from UCAR Communications, P.O. Box 3000, Boulder, CO 80307.].

Stensrud, D. J., H. Brooks, J. Du, M. S. Tracton, E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.*, **127**, 433-446.

Stensrud, D. J., J. M. Fritsch, 1993: Mesoscale convective systems in weakly forced large-scale environments. Part I: Observations. *Mon. Wea. Rev*., **121**, 3326–3344.

Tao, W.-K., and Coauthors, 2003: Microphysics, radiation, and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteor. Atmos. Phys*., **82**, 97–137.

Thompson, P., 1957: Uncertainty in the initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev*., **136**, 5095–5115.

Toth, Z., E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Walser, A., D. Lüthi, C. Schär, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.*, **132**, 560–577.

Wang, X., C. H. Bishop, 2003: A comparison of breeding and Ensemble Transform Kalman Filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.

Wang, X., C. H. Bishop, S. J. Julier, 2004: Which Is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble?. *Mon. Wea. Rev.*, **132**, 1590–1605.

Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116-5131.

Xu, M., D. J. Stensrud, J.-W. Bao, T. T. Warner, 2001: Applications of the adjoint technique to short-range ensemble forecasting of Mesoscale Convective Systems. *Mon. Wea. Rev.*, **129**, 1395–1418.

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, **82,** 139-170.

Xue, M., F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S. Weiss, A. Clark, J. Kain, M. Coniglio, J. Du, T. Jensen, and Y.-H. Kuo, 2010: CAPS realtime storm scale ensemble and high resolution forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. *25th Conf. Severe Local Storms*, Amer. Meteor. Soc., Paper 7B.3.

Zhang, D.-L., J. M. Fritsch, 1988: Numerical sensitivity experiments of varying model physics on the structure, evolution and dynamics of two mesoscale convective systems. *J. Atmos, Sci.*, **45**, 261-293.

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

Zhang, F., A. M. Odins, J. W. Nielsen-Gammon, 2006: Mesoscale predictability of an extreme warm-season precipitation event. *Wea. Forecasting*, **21**, 149–166.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321-1338.

List of Figures

FIG. 1. Difference between control forecast and observed 1-h accumulated precipitation, at 0600 UTC 20 May 2010 using forecast initialized at 0000 UTC 20 May 2010, showing (A) the total precipitation forecast and (B)-(K) the anomalies on each scale identified by the 2D Haar wavelet decomposition.

FIG. 2. Synoptic scale conditions at (A), (B) 0000 UTC 10 May, (C), (D) 1200 UTC 10 May and (E), (F) 0000 UTC 11 May. In (A), (C) and (E), 500 hPa geopotential height of the control member forecast initialized at 0000 UTC 10 May is shown. In (B), (D) and (F) the mean sea level pressure, surface fronts and surface observations from the Hydrometeorological Prediction Center surface analysis archive are shown (http://www.hpc.ncep.noaa.gov/html/sfc_archive.shtml).

FIG. 3. 1-h accumulated precipitation from the control forecast in (A), (C), (E) and (G) and from observations in (B), (D), (F) and (H). Valid times are (A,B) 0600 UTC 10 May, (C,D) 1200 UTC 10 May, (E,F) 1800 UTC 10 May and (G,H) 0000 UTC 11 May. The red outlines show the verification domain.

FIG. 4. As in Fig. 2, except for (A), (B) 0000 UTC 20 May, (C), (D) 1200 UTC 20 May and (E), (F) 0000 UTC 21 May.

FIG. 5. As in Fig. 3, except valid at (A), (B) 0100 UTC 20 May, (C), (D) 0600 UTC 20 May, (E), (F) 1200 UTC 20 May,(G), (H) 1800 UTC 20 May and (I), (J) 0000 UTC 21 May.

FIG. 6. Average squared difference (i.e., energy) between control forecast and observed hourly accumulated precipitation (CNerror), and between each perturbed forecast and the control forecast, during the 10 May case for (A) large scales only, (B) medium scales only, (C) small scales only and (D) without any scale decomposition or filtering.

FIG. 7. Perturbation energy as a function of wavelength for the 10 May case at lead times of 1, 3, 6, 12 and 24 h for (A) RAND, (B) RECRS, (C) LGPH, (D) LGPH_RECRS, (E) LG and (F) LG_RECRS. The CNerror energy is the dashed line in all panels.

FIG 8. As in Fig. 6, except for the 20 May case.

FIG 9. As in Fig. 7, except for the 20 May case.

FIG 10. Forecast perturbations at the 24 h lead time (perturbed forecasts minus the control forecasts shown in Fig. 3g and 3g) for (A) RAND on the 10 May case, (B) LGPH on the 10 May case, (C) LG on the 10 May case, (D) RAND on the 20 May case, (E) LGPH on the 20 May case and (F) LG on the 20 May case.

FIG. 11. As in Fig. 6, except averaged over the entire experiment period. Statistical significance at the 95% confidence level, based on permutation resampling, is indicated as follows. Markers on the RAND, RECRS and LGPH_RECRS lines (circles, triangles and squares, respectively) indicate a significant difference from the LGPH line. Markers (asterisks) above all the lines indicate a significant difference between RAND and RECRS.

FIG. 12. As in Fig. 7, except averaged over the entire experiment period.

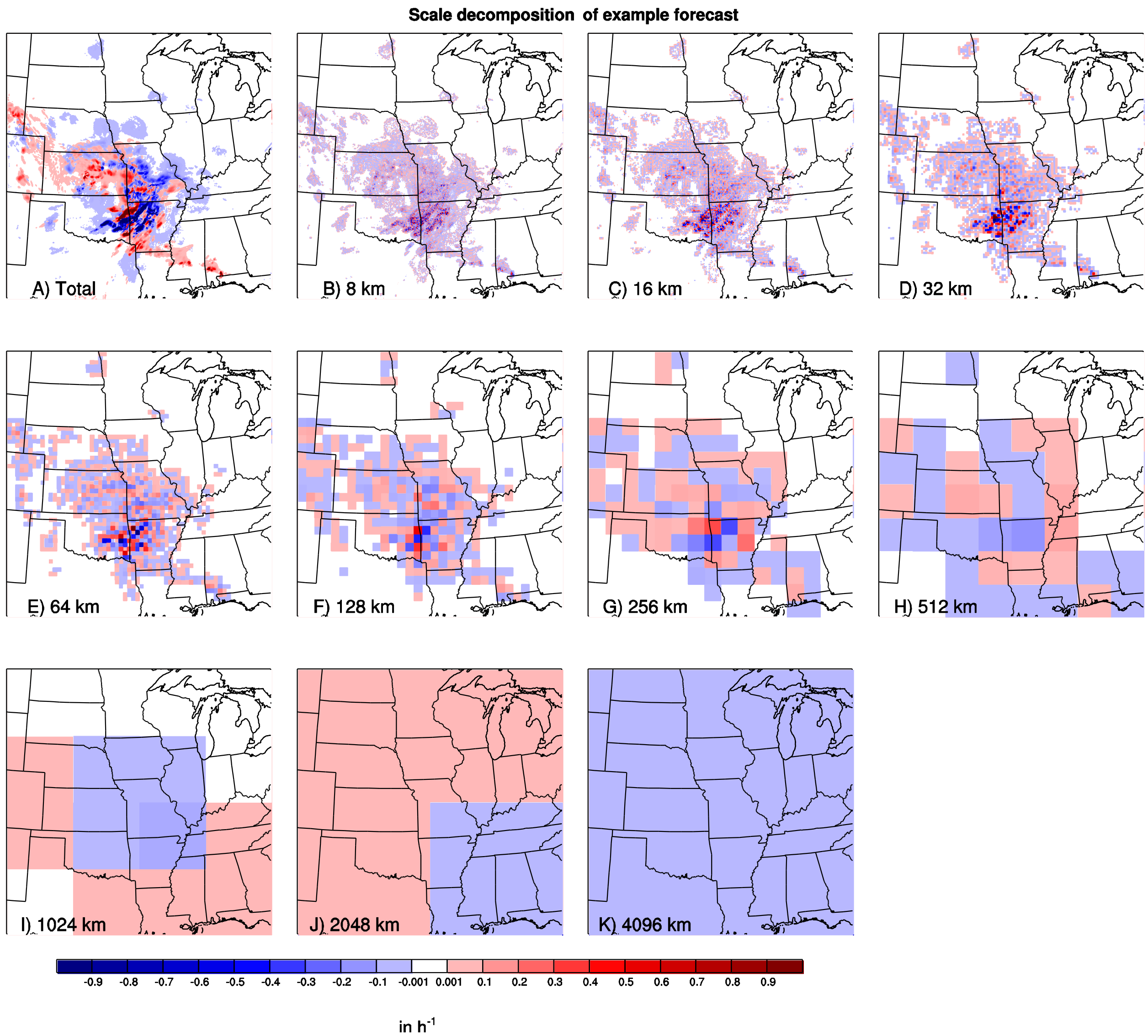

FIG. 1. Difference between control forecast and observed 1-h accumulated precipitation, at 0600 UTC 20 May 2010 using forecast initialized at 0000 UTC 20 May 2010, showing (A) the total precipitation forecast and (B)-(K) the anomalies on each scale identified by the 2D Haar wavelet decomposition.

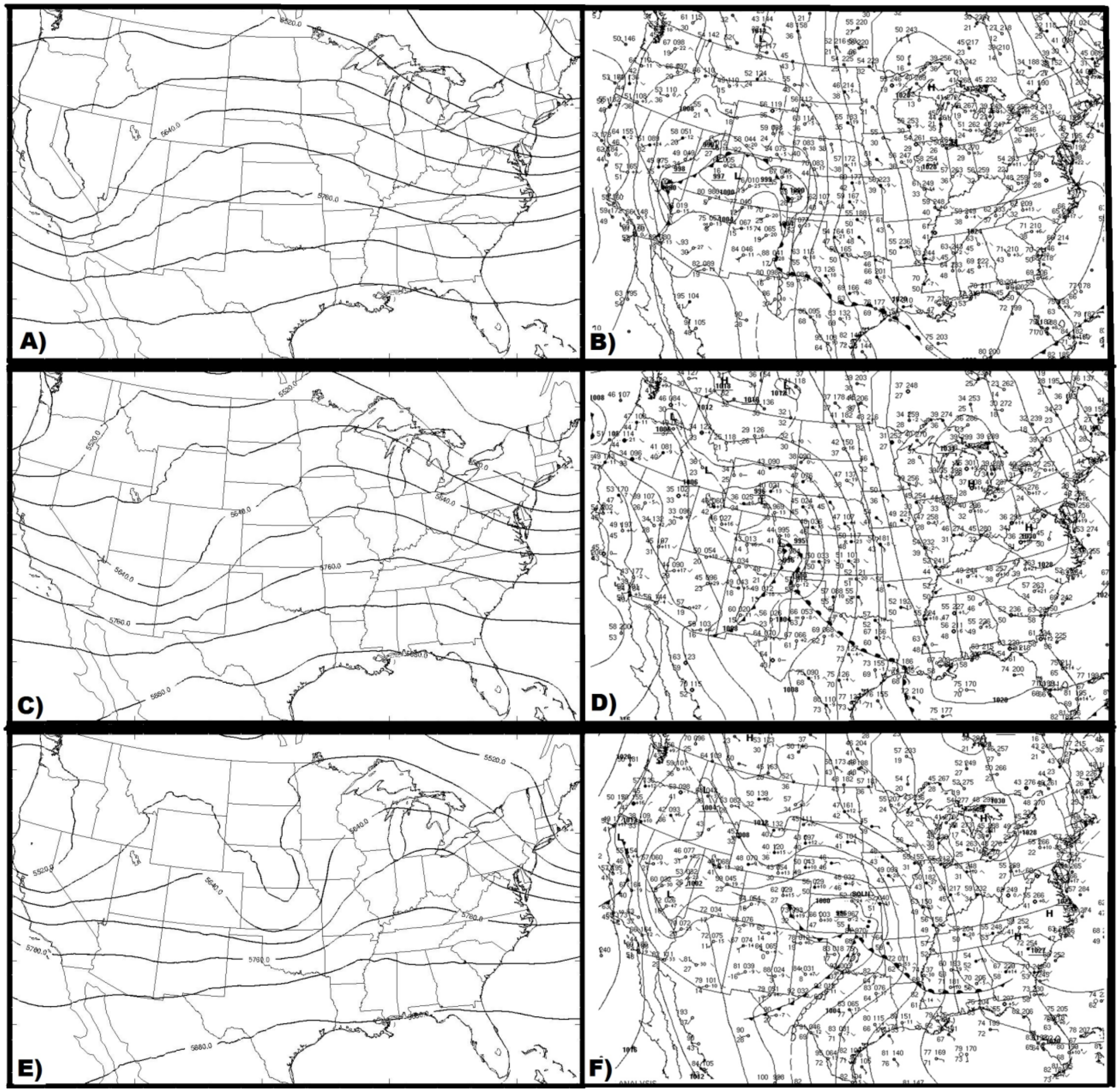

FIG. 2. Synoptic scale conditions at (A), (B) 0000 UTC 10 May, (C), (D) 1200 UTC 10 May and (E), (F) 0000 UTC 11 May. In (A), (C) and (E), 500 hPa geopotential height of the control member forecast initialized at 0000 UTC 10 May is shown. In (B), (D) and (F) the mean sea level pressure, surface fronts and surface observations from the Hydrometeorological Prediction Center surface analysis archive are shown (http://www.hpc.ncep.noaa.gov/html/sfc_archive.shtml).

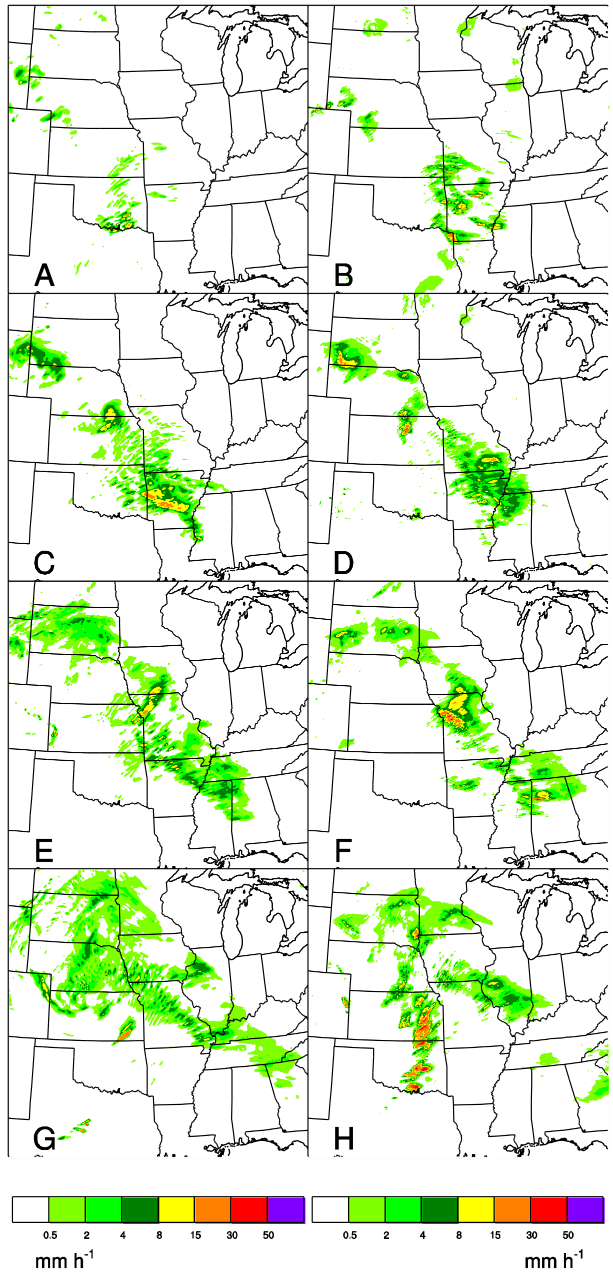

FIG. 3. 1-h accumulated precipitation from the control forecast in (A), (C), (E) and (G) and from observations in (B), (D), (F) and (H). Valid times are (A,B) 0600 UTC 10 May, (C,D) 1200 UTC 10 May, (E,F) 1800 UTC 10 May and (G,H) 0000 UTC 11 May. The red outlines show the verification domain.

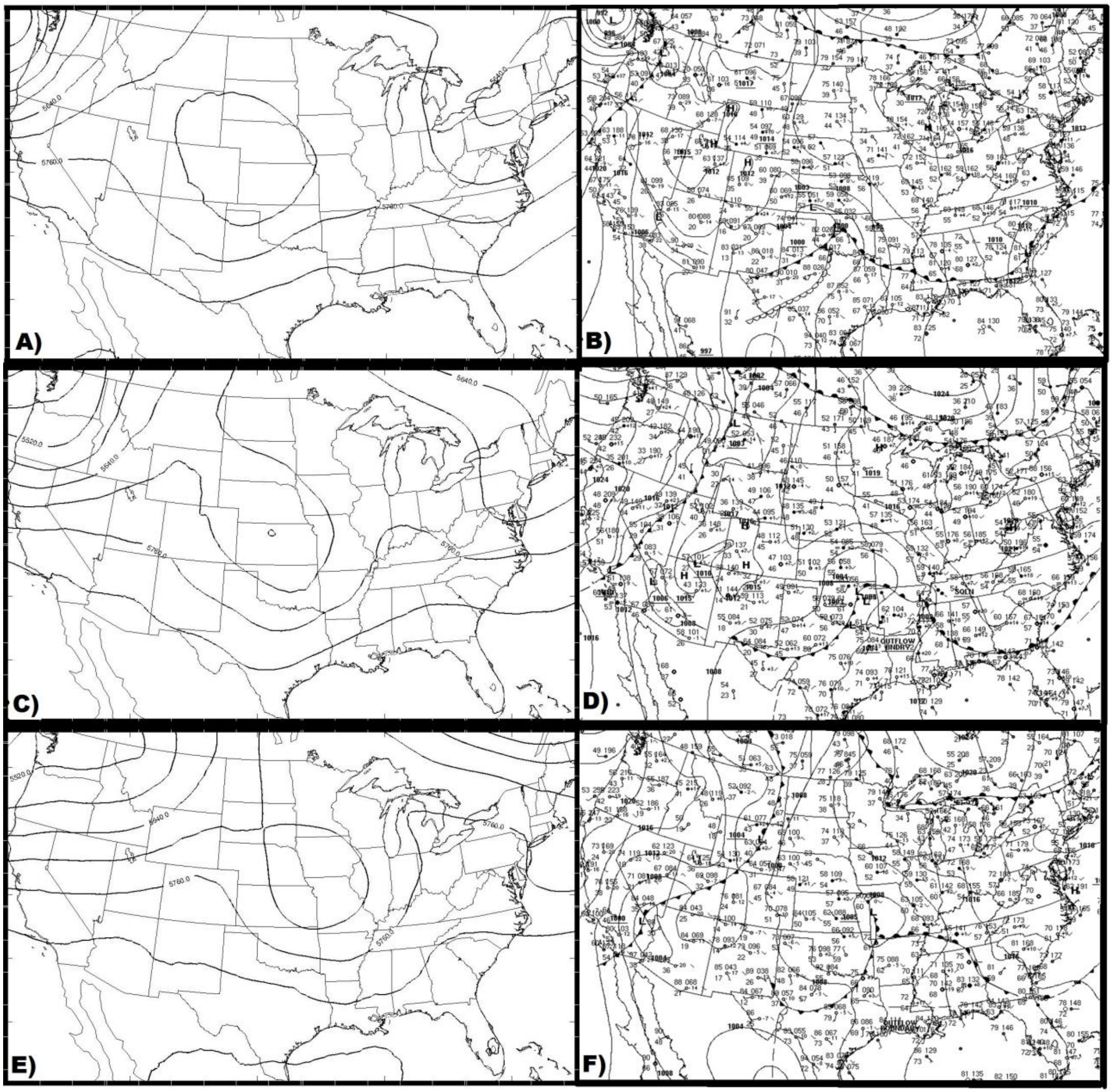

FIG. 4. As in Fig. 2, except for (A), (B) 0000 UTC 20 May, (C), (D) 1200 UTC 20 May and (E), (F) 0000 UTC 21 May.

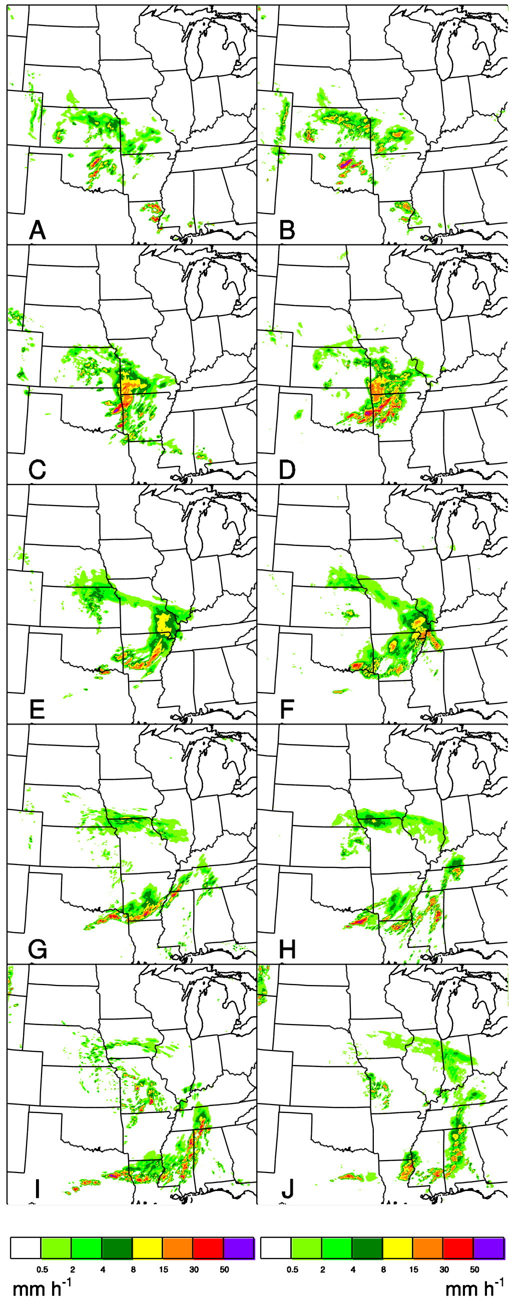

FIG. 5. As in Fig. 3, except valid at (A), (B) 0100 UTC 20 May, (C), (D) 0600 UTC 20 May, (E), (F) 1200 UTC 20 May,(G), (H) 1800 UTC 20 May and (I), (J) 0000 UTC 21 May.

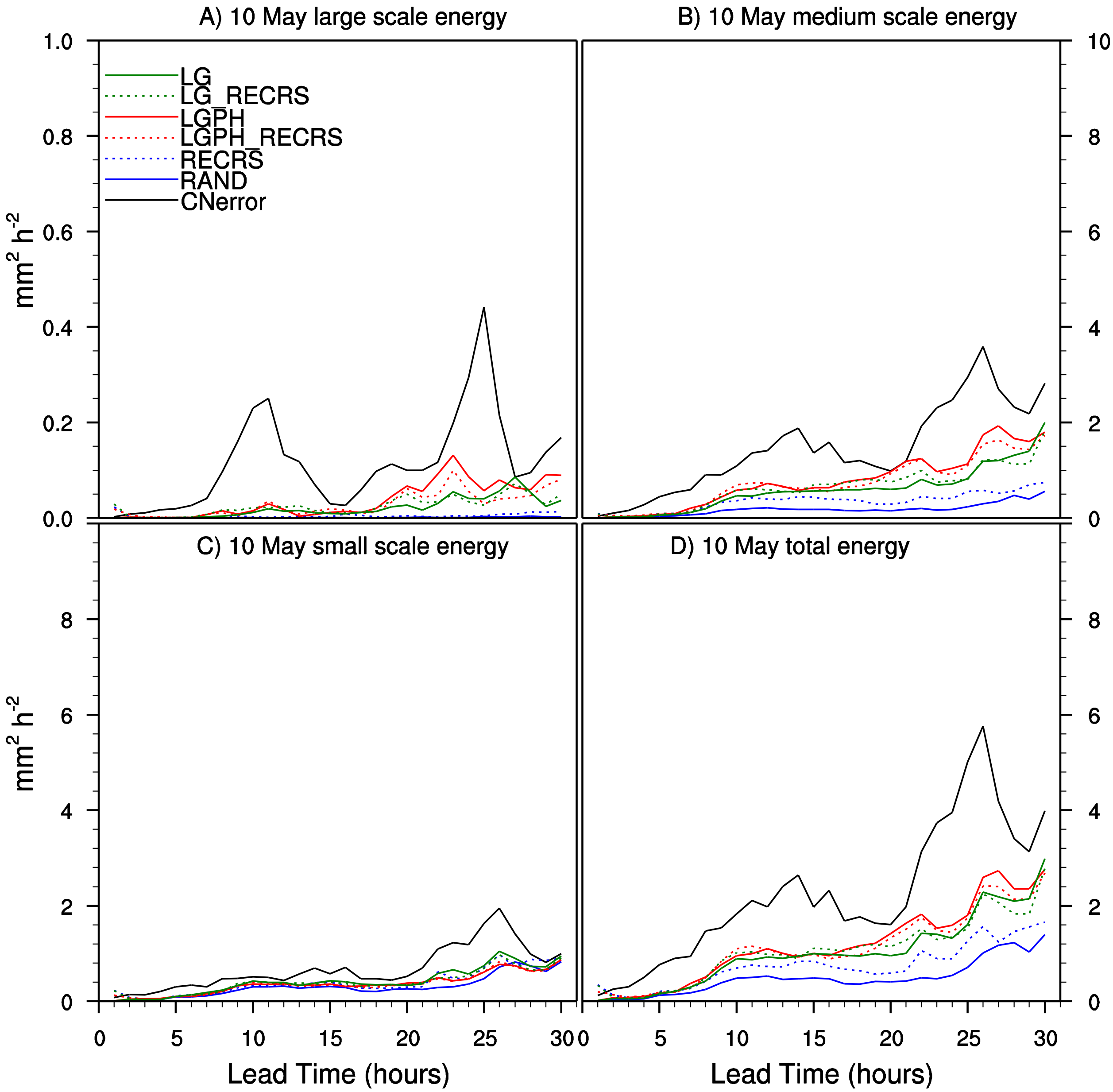

FIG. 6. Average squared difference (i.e., energy) between control forecast and observed hourly accumulated precipitation (CNerror), and between each perturbed forecast and the control forecast, during the 10 May case for (A) large scales only, (B) medium scales only, (C) small scales only and (D) without any scale decomposition or filtering.

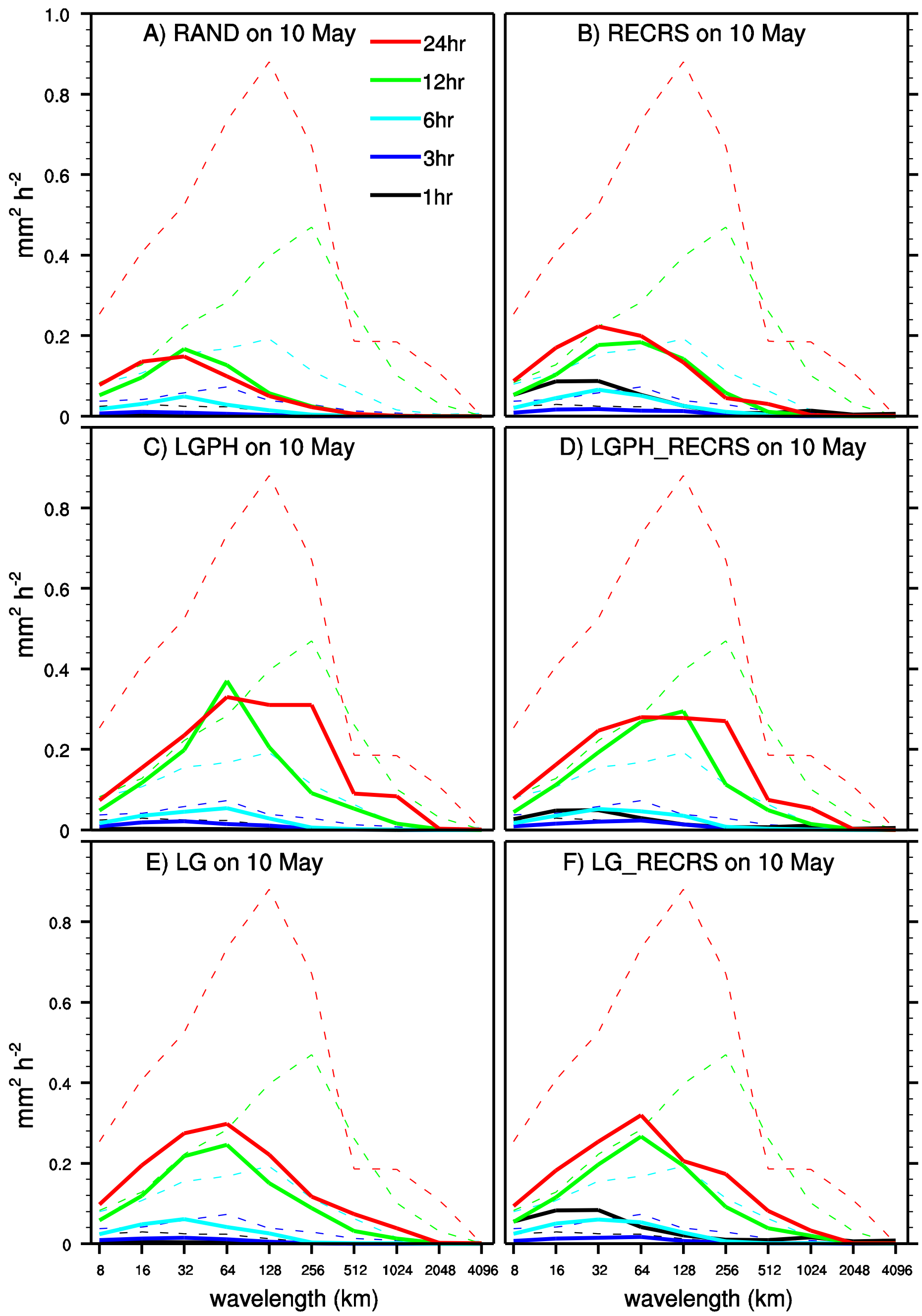

FIG. 7. Perturbation energy as a function of wavelength for the 10 May case at lead times of 1, 3, 6, 12 and 24 h for (A) RAND, (B) RECRS, (C) LGPH, (D) LGPH_RECRS, (E) LG and (F) LG_RECRS. The CNerror energy is the dashed line in all panels.

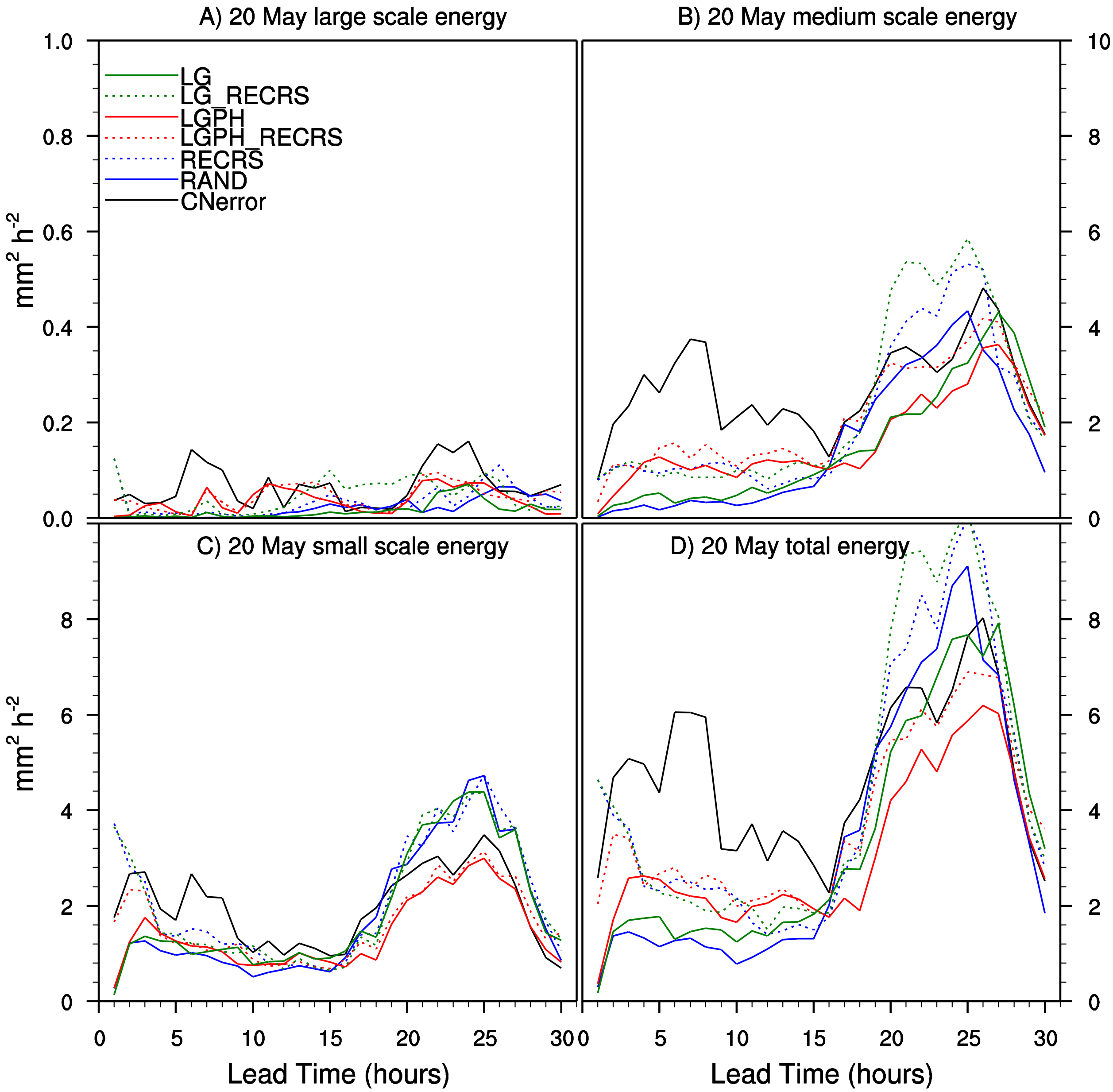

FIG 8. As in Fig. 6, except for the 20 May case.

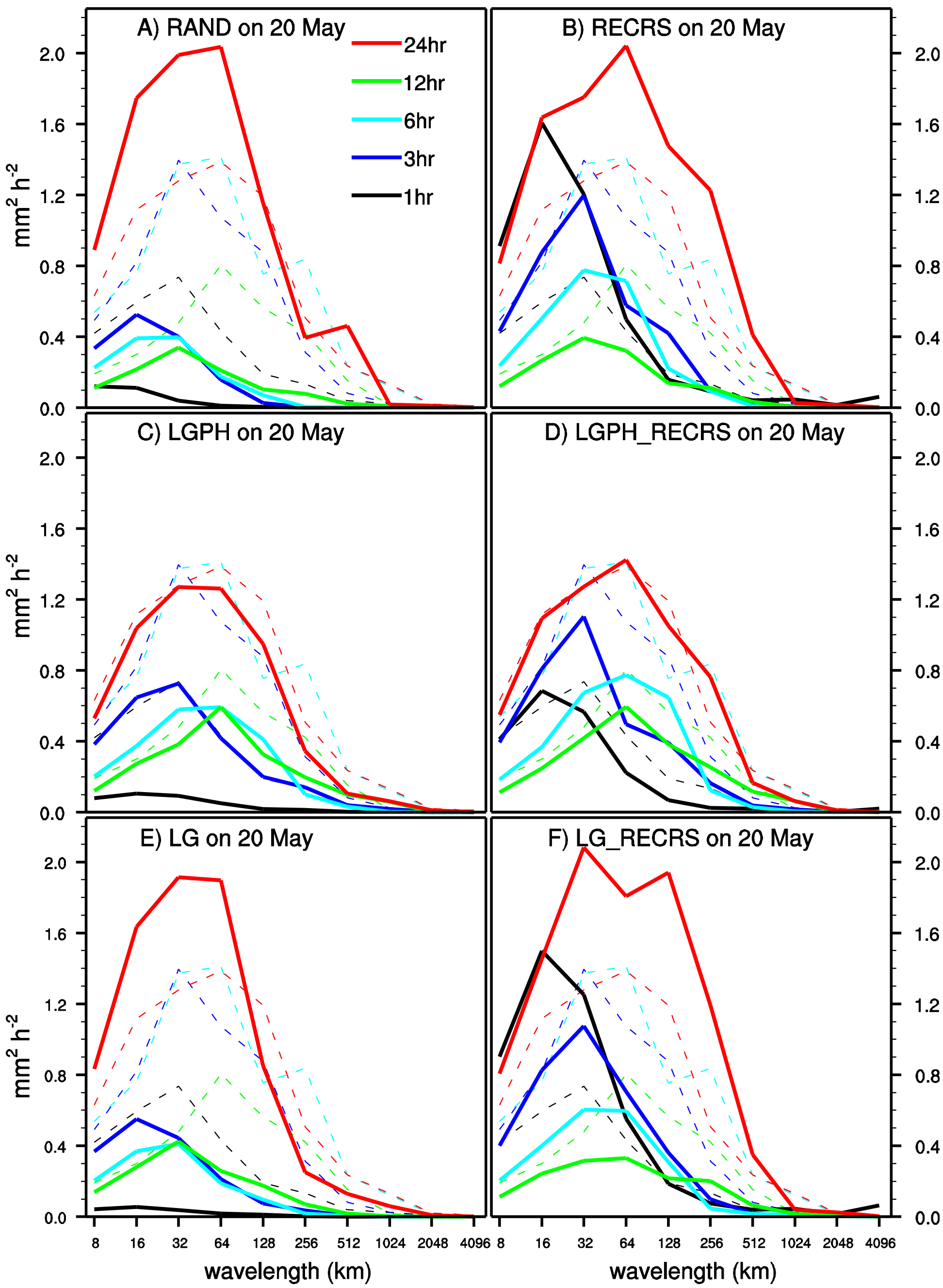

FIG 9. As in Fig. 7, except for the 20 May case.

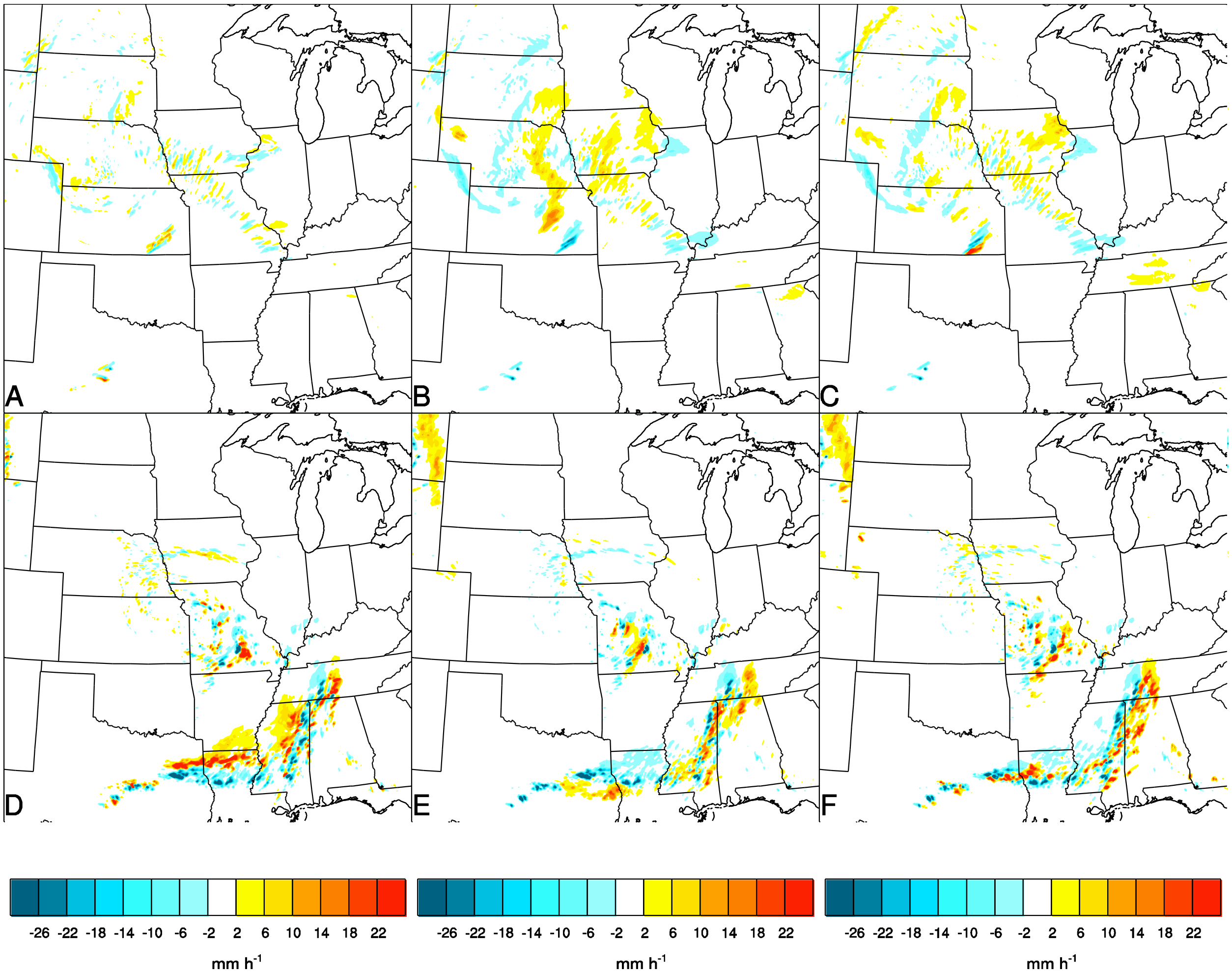

FIG 10. Forecast perturbations at the 24 h lead time (perturbed forecasts minus the control forecasts shown in Fig. 3g and 3g) for (A) RAND on the 10 May case, (B) LGPH on the 10 May case, (C) LG on the 10 May case, (D) RAND on the 20 May case, (E) LGPH on the 20 May case and (F) LG on the 20 May case.

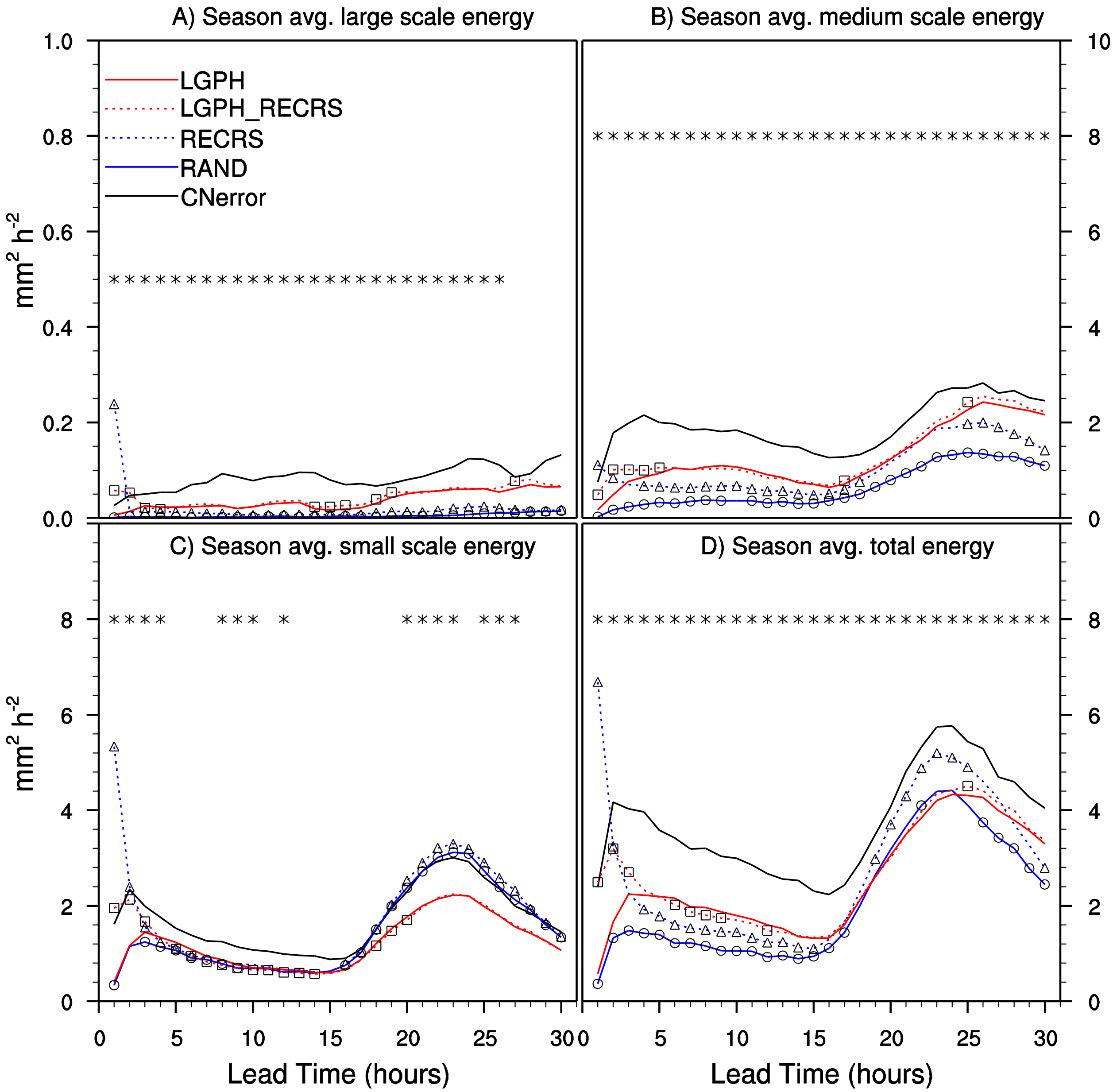

FIG. 11. As in Fig. 6, except averaged over the entire experiment period. Statistical significance at the 95% confidence level, based on permutation resampling, is indicated as follows. Markers on the RAND, RECRS and LGPH_RECRS lines (circles, triangles and squares, respectively) indicate a significant difference from the LGPH line. Markers (asterisks) above all the lines indicate a significant difference between RAND and RECRS.

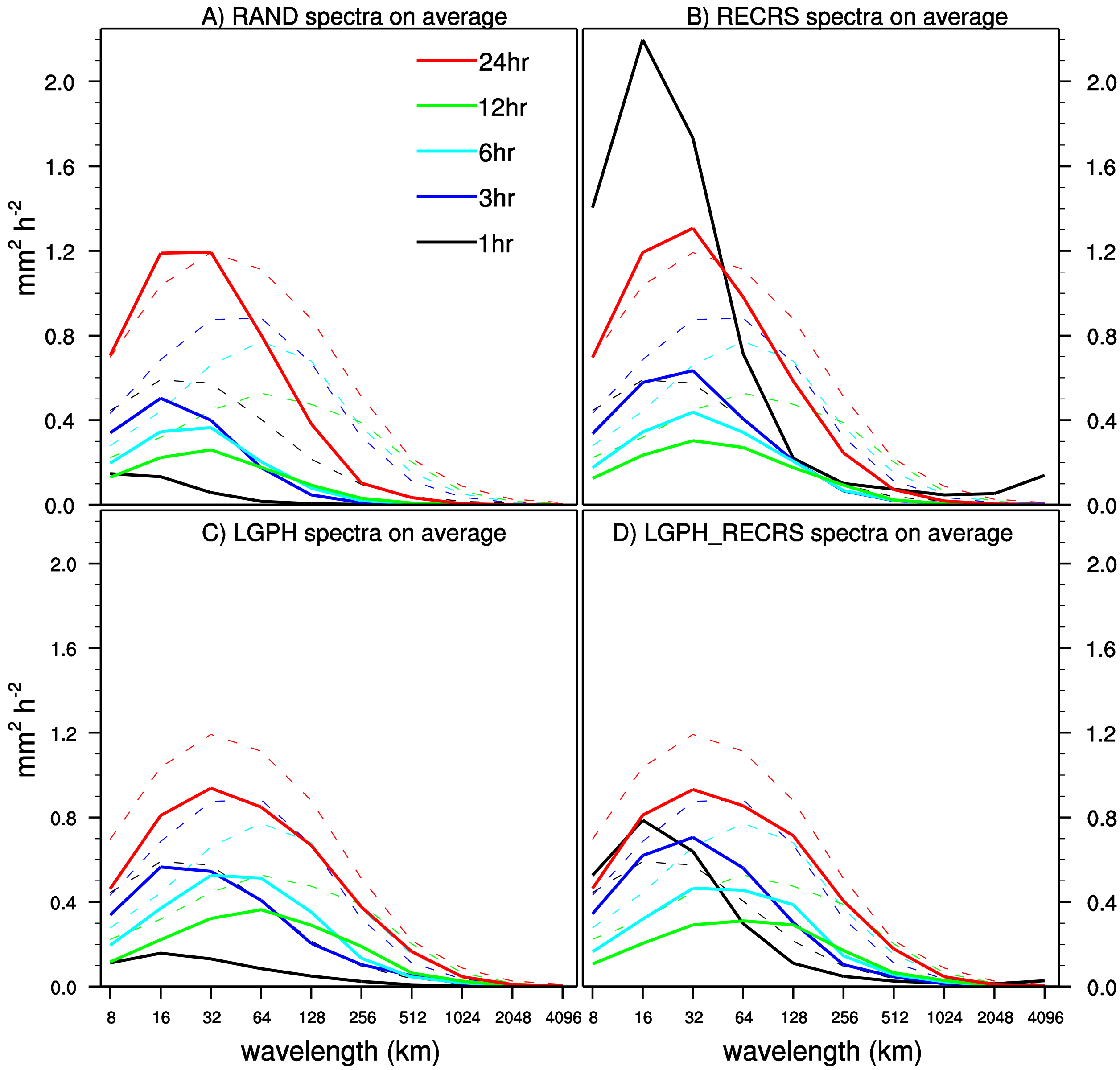

FIG. 12. As in Fig. 7, except averaged over the entire experiment period.