Received: 29 March 2018 | Revised: 07 June 2018 | Accepted: 09 June 2018 | Published on: 25 October 2018

DOI: 10.1002/qj.3367

SUPPLEMENT ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

The scale dependence of initial-condition sensitivities in simulations of convective systems over the southeastern United States

Jonathan A. Weyn | Dale R. Durran

Department of Atmospheric Sciences, University of Washington, Seattle, WA

Correspondence
Jonathan A. Weyn, Department of Atmospheric Sciences, University of Washington, Seattle, WA 351640.
Email: jweyn@uw.edu

The sensitivity of ensemble simulations of deep convective events in the southeastern United States to initial-condition (IC) errors is examined by imposing idealized moisture perturbations at small and large scales. Four severe weather events are considered, ranging from a springtime frontal system to convection driven almost exclusively by daytime heating. Events with strong synoptic-scale forcing were insensitive to the scale of IC errors, but weakly forced events exhibited greater sensitivity to small-scale than large-scale IC errors. Additional ensemble simulations of idealized convective systems suggest that the greater sensitivity to small-scale IC errors of the weakly forced cases arises from their higher sensitivity to the strength and location of the first convective elements. Ensemble spread and predictability are characterized by two measures: the ratio of the perturbation kinetic energy (KE) about the ensemble mean to the background KE and the neighborhood-based fractions skill score (FSS) of hourly precipitation with respect to that in an unperturbed reference simulation. For simulations of both observed events and idealized convective systems, the FSS appears to be a more discriminating indicator of differences in predictability between different convective events.

KEYWORDS

convective systems, error growth, initial conditions, numerical weather prediction, predictability, severe weather

1 | INTRODUCTION

The accurate prediction and advance warning of high-impact severe weather events is vital to protecting lives and property. As more powerful computational resources and better meteorological data have enabled numerical weather prediction at increasingly finer resolution, recent attention has focused on the predictability of mesoscale (5–400 km) features, including isolated and organized convection. The seminal work of Lorenz (1969) established the so-called "butterfly" effect, whereby the intrinsic limits on weather forecasts arise from the upscale growth of errors beginning at very small scales, possibly as small as the size of a butterfly. Many studies have shown that errors that first become detectable at small convective scales are eventually responsible for the loss of predictability on much larger scales (for example, Zhang *et al.*, 2003; 2006; 2007; Selz and Craig, 2014; Sun and Zhang, 2016). Recent work has demonstrated, however, that relatively small initial-condition (IC) errors on much larger scales of $\mathcal{O}(100)$ km can propagate rapidly downscale and subsequently reduce predictability in a similar manner to that produced by relatively large errors in the conditions on small scales, thereby making the identification of the actual initial scale of errors responsible for forecast degradation ambiguous (Durran *et al.*, 2013; Durran and Gingrich, 2014; Durran and Weyn, 2016; Weyn and Durran, 2017). In particular, Weyn and Durran (2017, hereafter WD17) showed that intrinsic predictability in simulations of idealized mesoscale convective systems (MCSs) was independent of the horizontal scale of equal-absolute-amplitude IC errors across a range of different MCS types. This study seeks to answer the question of whether this scale independence is also present in

simulations of observed high-impact convective events using a full-physics numerical weather prediction model.

Evaluations of the performance of numerical simulations of severe convection have shown significant variations in the predictability of various convective events, making universal characterizations of forecast uncertainty difficult. Done *et al.* (2012) and Flack *et al.* (2018) showed that convection in "non-equilibrium" convective conditions (Keil *et al.*, 2014), that is, under the influence of fronts and other localized synoptic features, exhibited better predictability in a case study than convection in "quasi-equilibrium" conditions with a fairly uniform large-scale environment. Similarly, Surcel *et al.* (2016) showed that diurnally forced convection under weak synoptic forcing exhibited lower predictability and higher sensitivity to initial conditions than widespread convection under strong synoptic forcing. On the other hand, Duda and Gallus (2013) examined a wide range of over 30 convective-initiation cases and classified each by the magnitude of synoptic-scale forcing diagnostics, including 700 hPa omega, 200 hPa divergence, and quasi-geostrophic forcing terms. While they found variability in forecast skill among the cases, they did not find a statistically significant relationship between the synoptic-scale forcing strength and the rate of error growth.

Understanding how different atmospheric states respond to perturbations on different scales is vital for improving data assimilation. Johnson *et al.* (2014) showed that synoptically forced precipitation events show little response to small-scale IC perturbations, while precipitation events controlled by the upscale evolution of convective elements into an MCS show as much sensitivity to small-scale IC perturbations as to large-scale IC perturbations and variations in model physics. The importance of small-scale perturbations in upscale MCS development was also noted by Nielsen and Schumacher (2016). Potvin *et al.* (2017) showed that simulations of supercells exhibit little sensitivity to the resolution of initial conditions, and therefore little sensitivity to fine-scale initial errors; Johnson and Wang (2016) found that assimilating small-scale IC data produced modest improvements to 5 hr ensemble forecasts of mesoscale convection. These examples demonstrate that the impact of small-scale data assimilation on the predictability of mesoscale convective events, while generally positive, is not consistently beneficial.

The rate of error growth may also vary from case to case, depending on how predictability is quantified. Many different methods have recently been proposed to analyze forecast skill and predictability (for example, Roberts and Lean, 2008; Fang and Kuo, 2015; Surcel *et al.*, 2015; Dey *et al.*, 2016). Spectral decomposition of kinetic energy (described further in section 4) has long been used to study the structure of background atmospheric flows in both observations and models (for example, Nastrom and Gage, 1985; Skamarock, 2004; Waite and Snyder, 2013; Skamarock *et al.*, 2014), and spectra of error kinetic energy have been used to evaluate the scale dependence of errors (for example, Lorenz, 1969; Leith, 1971; Zhang *et al.*, 2003; Mapes *et al.*, 2008; Durran *et al.*, 2013; Durran and Gingrich, 2014; Durran and Weyn, 2016; Sun and Zhang, 2016, WD17). This method has the advantage of evaluating perturbation growth characteristics simultaneously at all scales and across the entire model domain, providing a clear picture of the evolution of errors as a function of scale (Surcel *et al.*, 2015; 2016). However, spectral decomposition is not well-suited for evaluating the skill of an ensemble forecast, especially when considering the location or intensity of severe-weather impacts including precipitation. To characterize forecast skill more thoroughly, we also use the neighborhood-based fractions skill score, or FSS (Roberts and Lean, 2008), described further in section 5. The FSS measures how well a model captures the location and frequency of precipitation events exceeding a certain defined threshold, and can be evaluated for neighborhoods of different sizes to measure performance on different spatial scales.

The main goals of this article are twofold. First, we use idealized, monochromatic two-dimensional IC perturbations to examine the growth of small- and large-scale errors in simulations of four widely varying high-impact convective events in the southeastern United States. These perturbation structures are at specific long and short wavelengths, as in Durran and Weyn (2016) and WD17. WD17 showed that equal absolute amplitude small- and large-scale initial errors produced nearly identical spectra of error kinetic energy at lead times of 4–5 hr in three idealized simulations of squall lines formed under widely varying amounts of vertical wind shear. We apply this methodology to real-data cases simulated with a numerical weather prediction model to examine the case-to-case variability in the growth of both small- and large-scale IC errors, although we do not examine the model's ability to make accurate forecasts. Second, we compare the spectral decomposition method and the FSS to examine what each metric contributes to the assessment of predictability in both idealized and observed cases. As will be shown, the FSS complements the spectral methods by evaluating the skill of the ensemble simulations in representing significant weather.

The remainder of this article is organized as follows. The model configuration is presented in section 2, while details of the four simulated real-world events are presented in section 3. Section 4 analyzes the predictability of these events in terms of the degree of saturation of the ensemble mean perturbation kinetic energy, while section 5 presents an FSS analysis of the predictability of these simulations, as well as FSS for the more idealized cases considered in WD17. Lastly, conclusions are provided in section 6.

2 | ENSEMBLE DESIGN AND NUMERICAL PARAMETERS

The model used is the Weather Research and Forecasting (WRF) model version 3.7.1. Individual simulations are performed on a domain covering much of the contiguous United

States at a horizontal resolution of 2.5 km and run for 24 hr using WRF's adaptive time-stepping. There are 60 vertical levels with a model top at 50 hPa. Model physics include Thompson microphysics (Thompson *et al.*, 2008), the Noah land surface model (Tewari *et al.*, 2004), the Yonsei University (YSU) planetary boundary-layer scheme (Hong *et al.*, 2006), Dudhia short-wave radiation (Dudhia, 1989), and Rapid Radiative Transfer Model (RRTM) long-wave radiation (Mlawer *et al.*, 1997). Cumulus parametrization is turned off. The model initialization and lateral boundary conditions are provided by one-degree Global Forecast System (GFS) analysis data.[1] As described further in the next section, four different cases are considered. Although historical data are used to initialize and provide boundary conditions for the simulations, it is not the purpose of this article to evaluate the ability of the model to reproduce observed weather.

Two ensembles are generated for each event by perturbing the water-vapor field in a manner similar to that in WD17. Each ensemble consists of one unperturbed control member and six perturbed members.[2] To allow for model spin-up, the perturbations are added into the control simulation at 6 hr. We will refer to simulation times t_s, defined such that the simulations are initialized at $t_s = -6$ hr and perturbations are introduced at $t_s = 0$. The water-vapor perturbations are of the form

$$q'_v = a(x,y)e^{-z/H}\sin\left[2\pi\left(\frac{x}{L}-\phi_x\right)\right]\sin\left[2\pi\left(\frac{y}{L}-\phi_y\right)\right], \quad (1)$$

where a is the perturbation amplitude, H is the e-folding height scale, and L is the horizontal perturbation scale. The phases ϕ_x and ϕ_y ($0 \le \phi < 2\pi$) are generated randomly for each ensemble member, and hence the differences between ensemble members arise through the difference in phase between the perturbations and the location of convective activity. Each pair of ensembles consists of an "S" ensemble with $L = 20$ km and an "L" ensemble with $L = 200$ km to simulate the effects of small- and large-scale IC errors, respectively. (Note that the true wavelengths of these two-dimensional perturbations are $2^{-1/2}L$ or about 14.1 and 141 km, respectively.) The perturbation vertical scale height is $H = 1$ km. Because there are large spatial variations in background water vapor in the real-data cases, using a constant perturbation amplitude (as in WD17) would introduce much larger relative errors where the water vapor is low. We therefore define $a(x, y)$ to be 1% of the background water vapor. To avoid spurious supersaturation, no perturbation is added in regions where the perturbation would cause the water vapor to exceed the saturation water vapor. Spectral analysis of the IC water-vapor perturbation field confirmed that, despite the spatial dependence of $a(x, y)$, the perturbation remained strongly peaked at wavelengths of 14 or 141 km.[3]

The scale of the short-wavelength perturbations was chosen as $L = 8\Delta x = 20$ km, because this is slightly longer than the nominal longest wavelength ($7\Delta x$) that is significantly dampened by numerical dissipation (Skamarock, 2004).[4] The choice of 200 km for the large-scale perturbations ensures an order of magnitude separation between the "S" and "L" scales. In addition, 200 km is near the large-scale end of the range of mesoscale wavelengths over which the average atmospheric KE spectrum exhibits a $k^{-5/3}$ slope.

When using a limited-area model with prescribed boundary conditions, the dilution of IC perturbations due to forcing from the boundaries is a challenge (Vukicevic and Errico, 1990). Hohenegger and Schär (2007) found that fixed lateral boundaries dilute IC perturbations in convective-permitting limited-area simulations within about 11 hr on a domain roughly 800 km in size. In our simulations, we avoid the influence of the lateral boundary conditions altogether by focusing all of the analysis of perturbation growth on a limited 1500 km × 1500 km subdomain of the complete WRF model domain. To verify that the boundaries did not influence the subdomain, we used the NOAA Hysplit trajectory model[5] to calculate backward trajectories ending within the subdomain at all of the model heights considered in the analysis (not shown). The choice of WRF model domain was made such that no trajectories originated from outside the domain at any time prior to $t_s = 18$ hr, 6 hr after the end of our analysis. Therefore, we are confident that the ensemble error growth is not influenced significantly by the prescribed lateral boundary conditions.

3 | OBSERVED CASES AND THEIR SYNOPTIC-SCALE FORCING

3.1 | Four severe convective events

In the following discussion of the convective events, the reader is referred to Figure 1c for the locations of US states, marked as state abbreviations. For reference, these state abbreviations are included in the text below in parentheses.

3.1.1 | April 29, 2017 (April 2017)

The April 2017 case has the strongest synoptic-scale forcing among the four cases. The synoptic map in Figure 1a shows that there was a deep upper-level trough extending south to Mexico with associated synoptic-scale ascent ahead of the trough and descent behind at the model initialization time of 1800 UTC on April 29, 2017. The jet exit region generated significant localized ascent over a region from Oklahoma (OK) and Arkansas (AR) northeast to northern Illinois (IL).

[1] ftp://nomads.ncdc.noaa.gov/GFS/analysis_only
[2] The number of ensemble members was limited by the computational cost of each simulation.
[3] Perturbations at wavelengths of $L/2$, the next most significant signal in the spectrum, were more than three orders of magnitude weaker than the peak.
[4] WD17 also chose a scale of $8\Delta x$ for their short-wavelength perturbations, which, because $\Delta x = 1$ km in their smaller numerical domain, gave $L = 8$ km.
[5] https://ready.arl.noaa.gov/HYSPLIT.php

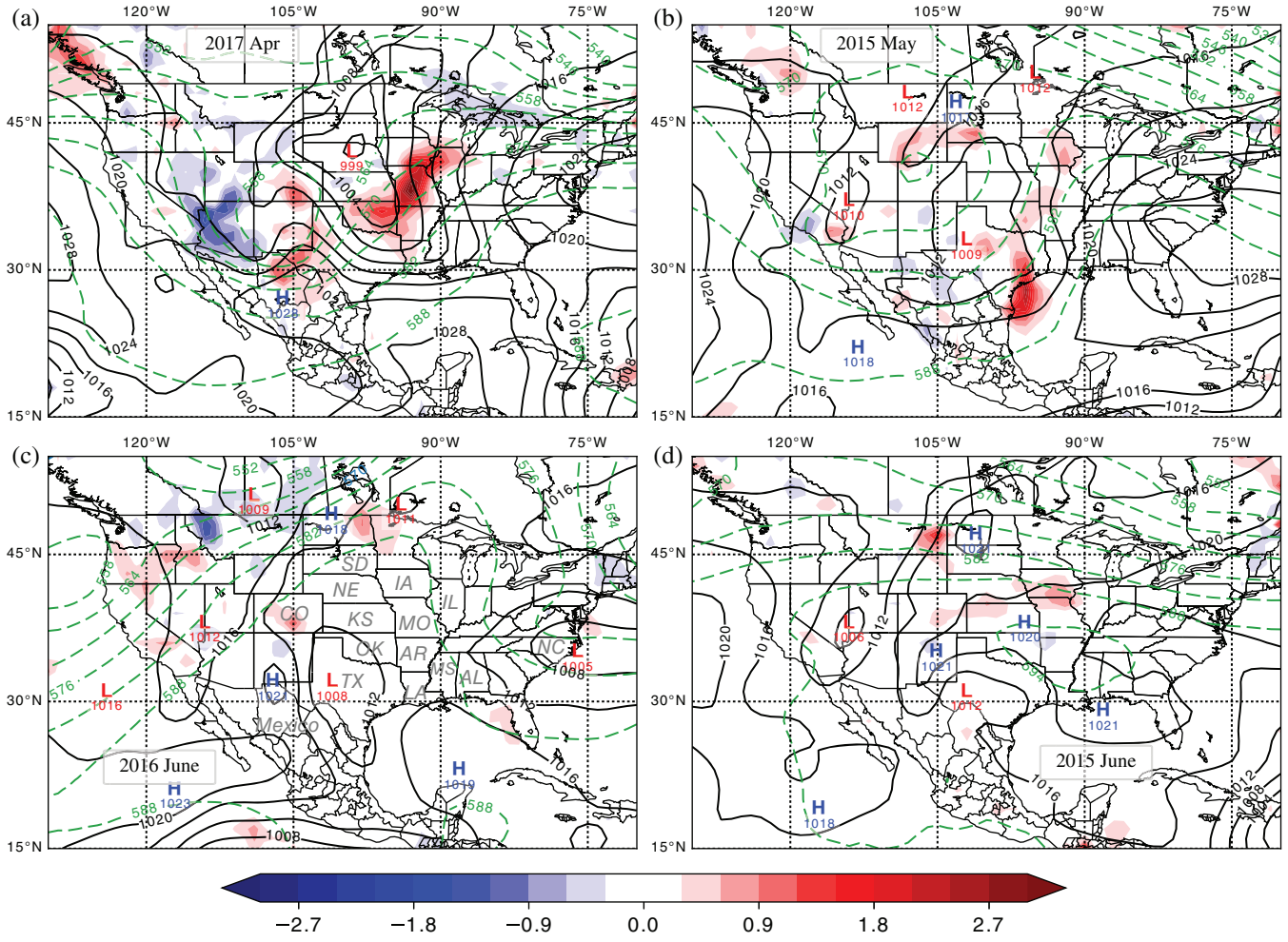

FIGURE 1 Synoptic maps for each case at $t_s = -6$ hr, corresponding to (a) 1800 UTC April 29, 2017, (b) 1200 UTC May 24, 2015, (c) 1200 UTC June 17, 2016, and (d) 1200 UTC June 24, 2015, generated from the Global Forecast System (GFS) analysis data used as model initial conditions. Dashed green lines are 500 hPa heights (dkm), solid black lines are sea-level pressure (hPa), and colored shading is negative 500 hPa omega (Pa/s, red is ascent). Surface pressure highs and lows are also indicated. Abbreviations for US states are included in (c)

An associated 999 hPa surface low was centered ahead of the upper-level trough at the Nebraska (NE)–South Dakota (SD) border and was undergoing rapid development. Vigorous convection was occurring at this time along a cold front spanning from central Texas (TX) to Missouri (MO) and further north, and a subsequent, albeit weaker, region of convective activity developed in association with the synoptic-scale ascent from west Texas (TX) north to Colorado (CO). Twelve hours later ($t_s = 6$ in the simulations), a line of very strong convection ran through eastern Texas (TX) to Arkansas (AR), as shown in the model synthetic reflectivity plot in Figure 2a. The evolution of the model's synthetic reflectivity within the analysis subdomain, presented at higher resolution in Figure 3a,c,e, shows that, although it weakened slightly, the main convective line remained quite distinct at $t_s = 12$ hr. This springtime severe weather system resulted in 364 wind and hail reports and 59 tornado reports from 1200 UTC on April 29 to 1200 UTC on May 1, according to the Storm Prediction Center's (SPC) filtered storm reports.[6]

[6]https://www.ncdc.noaa.gov/stormevents/

3.1.2 | May 24, 2015 (May 2015)

A weak surface frontal system associated with a broad upper-level trough initiated convection in an unstable environment over Oklahoma (OK) and Texas (TX) as much as 24 hr before the start of the next simulation at 1200 UTC on May 24, 2015. A broad line of convection extended from Mexico to Missouri (MO) when the control simulation was initialized, although the model needed to spin those features up. An upper-level trough and weak surface features, along with lift associated with the trough, are pronounced in Figure 1b. The main squall line produced outflow boundaries and surface troughs initiating a sustained line of convection that gradually progressed eastward throughout the model simulation, and was located over a broad area from Missouri (MO) to Louisiana (LA) by $t_s = 6$ hr (Figure 2b). The frontal zone, although weaker than in the previous case, produced several north–south oriented convective lines over the affected area, as shown in Figure 2b. The evolution and eastward propagation of these lines is shown by the model-simulated radar reflectivity in Figure 3b,d,f. Despite weaker frontal forcing, the thunderstorms, along with some supercells triggered by

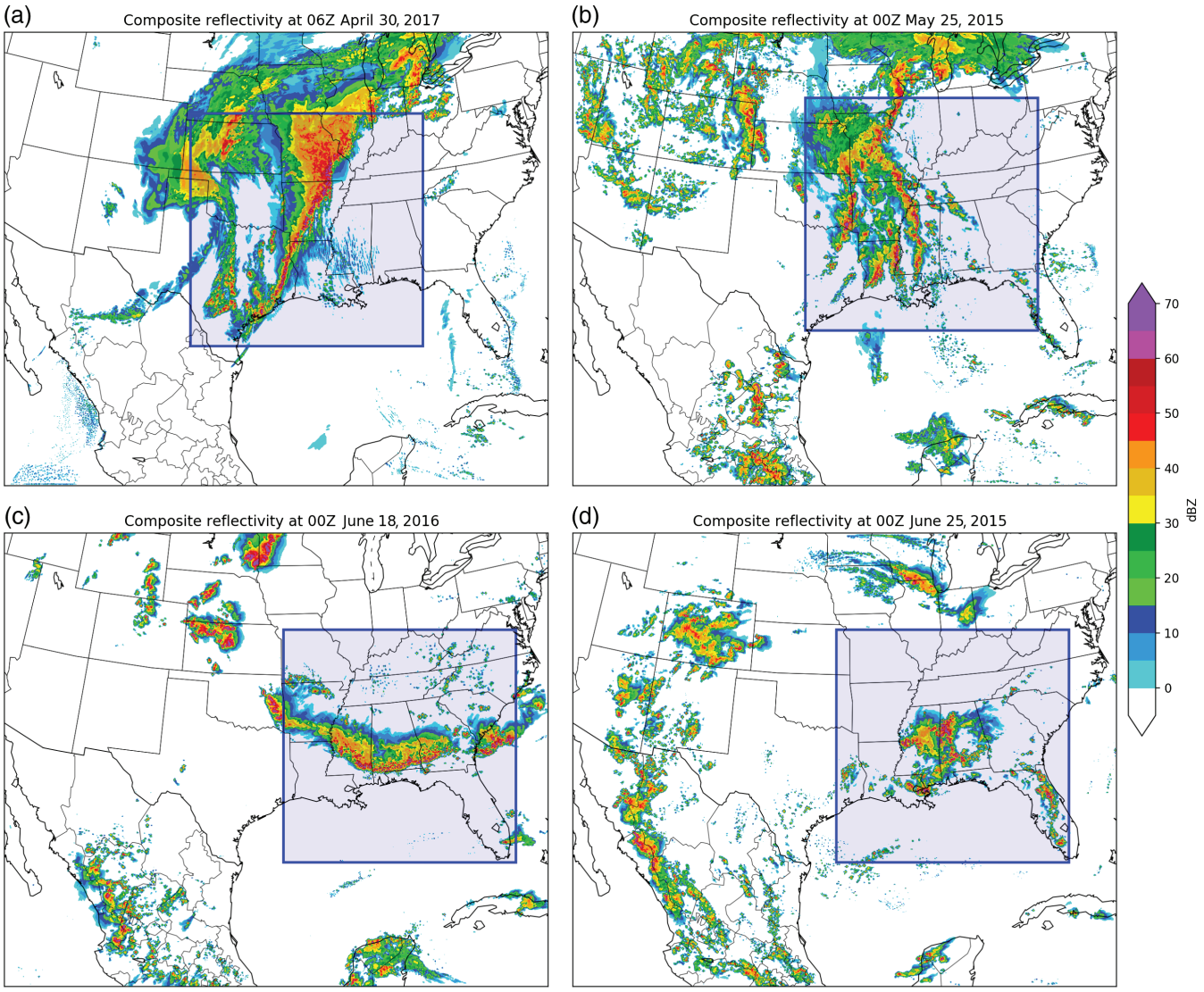

FIGURE 2 Model-generated synthetic reflectivity for the control members at $t_s = 6$ hr, for (a) April 2017, (b) May 2015, (c) June 2016, and (d) June 2015 cases. The plotted domain is the entire model domain, and the blue boxes denote the subdomain where further analysis is performed (see text)

the surface features in eastern Colorado (CO), produced significant severe weather, with the SPC recording 92 filtered wind and hail reports and 32 tornado reports.

3.1.3 | June 17, 2016 (June 2016)

The initial upper-level flow at 1200 UTC on June 17, 2016 consisted of a pronounced ridge over the central United States and a deep, almost west-east-oriented trough over the eastern US, as shown in Figure 1c. Associated with a weak surface low located offshore of North Carolina (NC) was a very weak front extending westward through Arkansas (AR). Diurnal forcing triggered strong thunderstorms in the unstable, warm air mass south of the weak frontal boundary beginning at around $t_s = -2$ hr. Despite relatively weak forcing along the air-mass boundary, these storms organized into a remarkably continuous and strong squall line, as shown in Figure 2c. Outflow uplift sustained the squall line as it gradually progressed southward (see Figure 4a,c,e), but it eventually weakened after sunset and collapsed into a smaller MCS. The squall line produced 328 filtered wind and hail reports according to the SPC, the vast majority being strong straight-line winds.

3.1.4 | June 24, 2015 (June 2015)

At the time of model initialization at 1200 UTC on June 24, 2015, the upper-level features were characterized primarily by nearly zonal 500 hPa heights with little synoptic-scale forcing (Figure 1d). Around $t_s = 0$ hr, the convection of primary interest was triggered by diurnal forcing along a weak surface trough in the southeast states, primarily in Mississippi (MS) and Alabama (AL). As shown in Figure 4b,d,f, these storms aggregated into one large but disorganized MCS, showing significant development by $t_s = 6$ (Figure 2d). By the early morning hours of 25 June ($t_s = 12$ hr), the system had weakened greatly. Despite being the smallest and weakest convective system among the four cases considered, the SPC logged 138 filtered wind and hail reports in the southeast United States for 24 June.

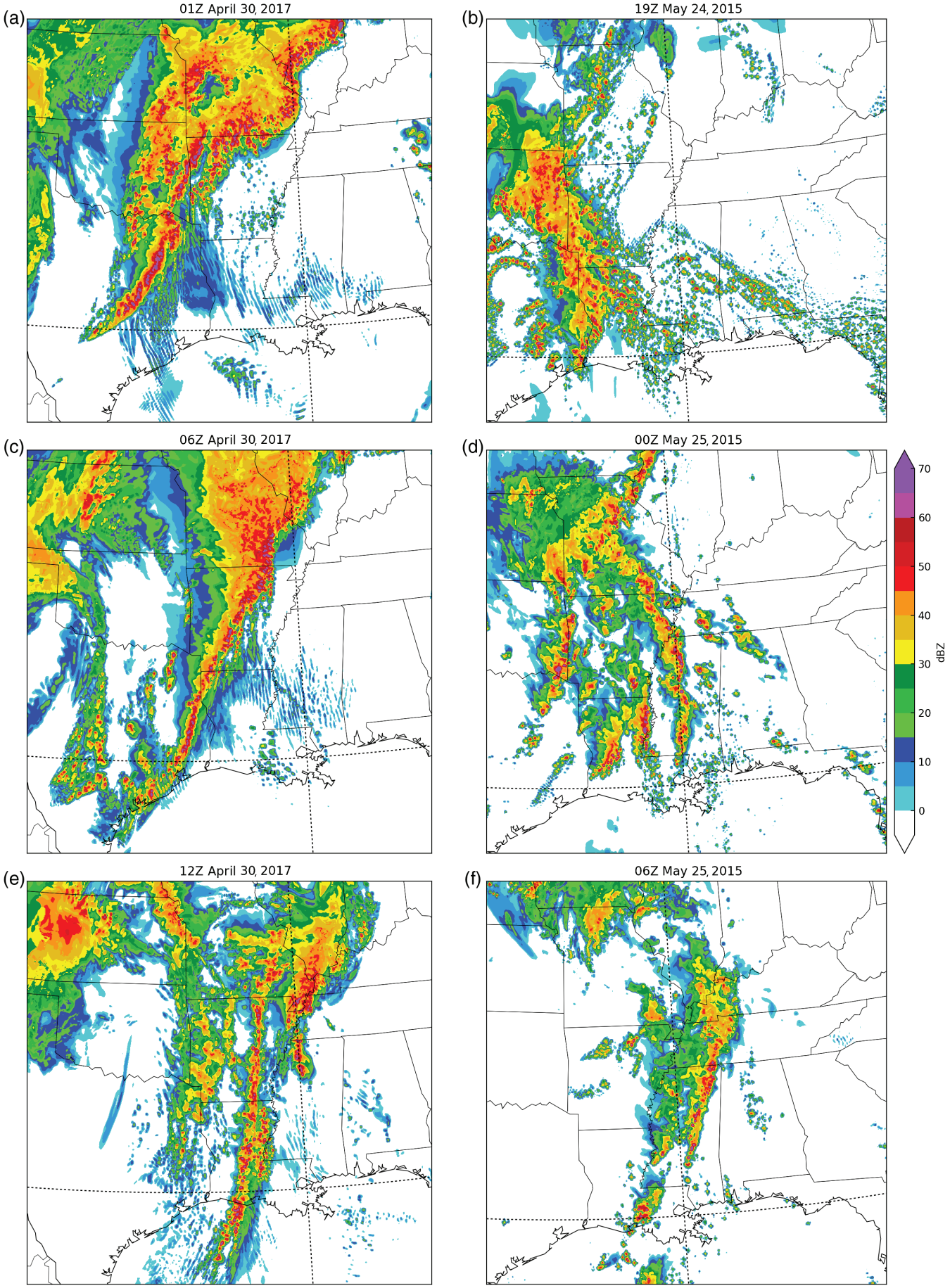

FIGURE 3 Model-generated synthetic reflectivity for the control members at t_s equal to (a,b) 1 hr, (c,d) 6 hr, and (e,f) 12 hr, for (a,c,e) the April 2017 case and (b,d,f) the May 2015 case. Only the subdomains shown in Figure 2 are plotted

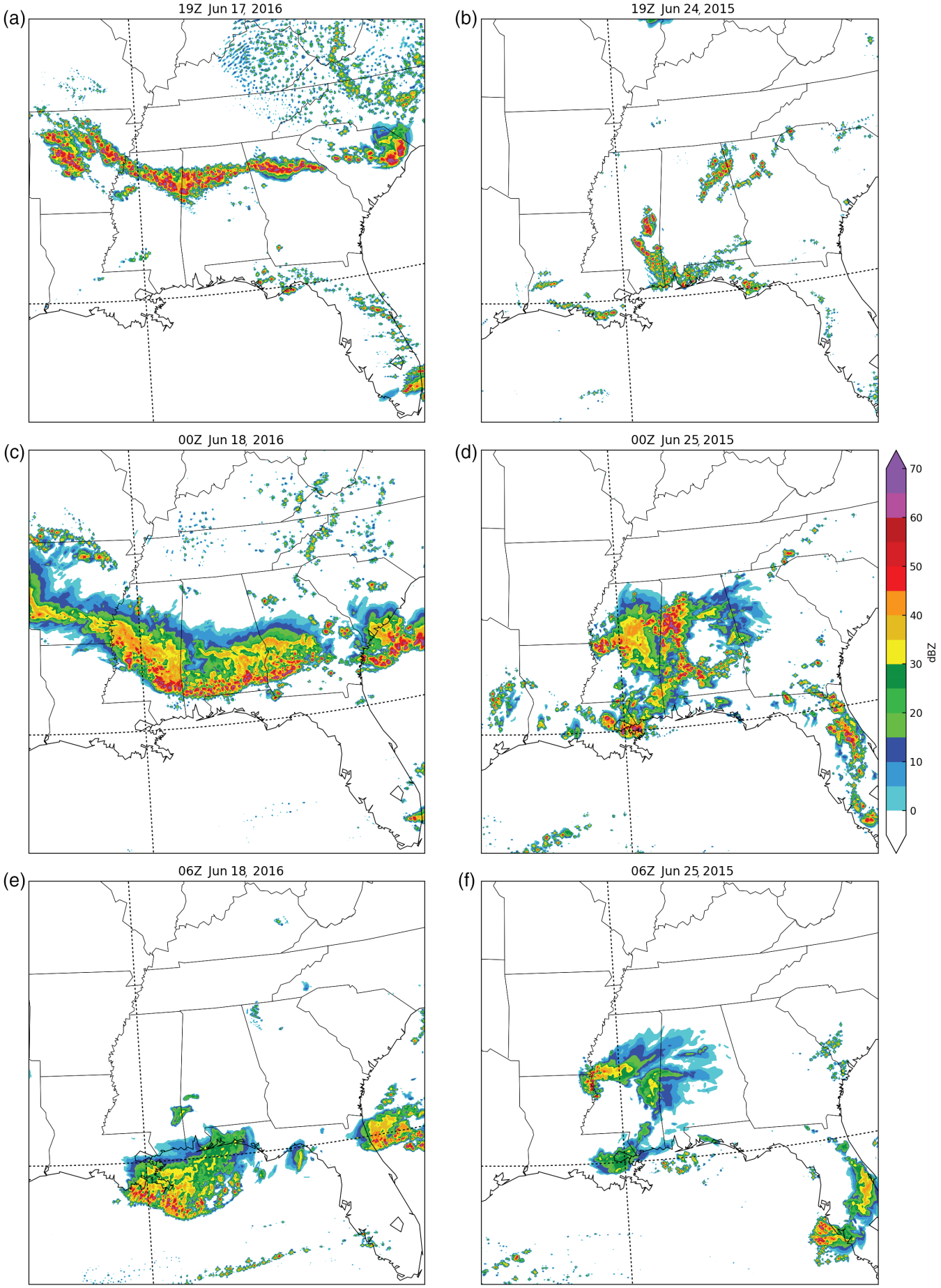

FIGURE 4 As Figure 3, but for (a,c,e) the June 2016 case and (b,d,f) the June 2015 case

TABLE 1 Synoptic-scale forcing diagnostics for the four events computed at the times shown in Figure 1. The numbers in parentheses are normalized values in which each forcing is divided by that for the strongest April 2017 event; A_f is the average of the normalized forcings for each event

	700 hPa ω	Surface frontogenesis ($\geq$ 25 pct'ile)	200 hPa divergence	500 hPa Q-vector convergence ($\geq$ 75 pct'ile)	
	Pa/s	10^{-9} K m^{-1} s^{-1}	10^{-6} s^{-1}	10^{-9} Pa m^{-2} s^{-1}	A_f
April 2017	−0.36 (1.0)	2.2 (1.0)	38.9 (1.0)	0.47 (1.0)	1.0
May 2015	−0.18 (0.50)	0.45 (0.20)	5.3 (0.14)	0.28 (0.60)	0.36
June 2016	0.009 (−0.02)	0.25 (0.11)	−4.4 (−0.11)	0.35 (0.74)	0.18
June 2015	0.074 (−0.21)	0.33 (0.15)	−1.1 (−0.03)	0.31 (0.66)	0.14

3.2 | Synoptic-scale forcing diagnostics

To quantitatively diagnose the degree of synoptic-scale influence on deep convection in these cases, several dynamically motivated metrics are calculated. As in Duda and Gallus (2013), we calculate the 700 hPa vertical motion (ω), surface frontogenesis, and 200 hPa divergence. The forcing for quasi-geostrophic vertical motion is assessed by computing the 500 hPa Q-vector convergence. The 700 hPa ω and 200 hPa divergence calculations are straightforward. For surface frontogenesis, we neglect the contributions from vertical motions and diabatic heating. The formulation used for the Q vectors is

$$\boldsymbol{Q} = \left(-\frac{R}{\sigma p}\frac{\partial \boldsymbol{V}_g}{\partial x}\cdot\nabla T, -\frac{R}{\sigma p}\frac{\partial \boldsymbol{V}_g}{\partial y}\cdot\nabla T \right), \tag{2}$$

where $\boldsymbol{V}_g$ is the geostrophic wind at 500 hPa, T is the temperature at 500 hPa, R is the universal gas constant, $p = 500$ hPa is the pressure, and $\sigma = (-RT/p)(d\ln\theta/dp)$ is the static stability parameter, with θ the potential temperature. We report the Q-vector convergence as $-2\nabla\cdot\boldsymbol{Q}$, which is roughly proportional to $-\omega$; hence positive values for the Q-vector metric equate to upward motion.

The synoptic-scale forcing metrics are calculated from the IC (at $t_s = -6$ hr) one-degree gridded GFS data. For each event, we use the analysis domain, shown by the blue boxes in Figure 2, which includes the main convective features of interest. Within this analysis domain, we identify those grid points in the GFS data where precipitation is occurring when the ensemble perturbations are introduced at $t_s = 0$, and calculate the diagnostics averaged over those grid points 6 hr earlier, at $t_s = -6$. The coarse resolution of the GFS analysis and spatial averaging dilutes the diagnostics. To capture the localized effects of frontal boundaries, we only average values greater than the 25th percentile for the frontogenesis diagnostic; for the Q-vector diagnostic, we average values greater than the 75th percentile to isolate the effect of the strongest upward vertical forcing. While the values for the diagnostics reported in Table 1 depend on these percentile averages, the relative ranking of the four cases does not.

Table 1 shows the values of the synoptic-scale forcing diagnostics for each of the four events, along with the same values normalized by dividing by the forcing for the strongest event, the April 2017 case. Also listed is the average of all four normalized forcing values A_f. The April 2017 case exhibits the strongest synoptic-scale forcing by far ($A_f = 1.0$). The forcing in the May 2015 case is distinctly weaker ($A_f = 0.36$), but is nevertheless substantially stronger than the forcing for the June 2016 ($A_f = 0.18$) and June 2015 ($A_f = 0.14$) cases. These A_f values quantify the discussion in the previous subsection and facilitate our classification of events by the strength of the synoptic-scale convective organization. Nevertheless, the values in Table 1 are meant to illustrate the range in synoptic-scale forcing among the convective events presented herein, rather than an attempt to classify such events generally. We also do not assume that the characteristics of predictability discussed in the remainder of this article depend invariantly on these metrics.

4 | PERTURBATION AND BACKGROUND KINETIC ENERGY SPECTRA

To measure the scale-dependent structure of perturbation growth, we calculate the total and perturbation kinetic energy (KE and KE′) spectral densities following WD17. At a given height and time, denoting the zonal and meridional velocities of the mth ensemble member, including the unperturbed control member, as u_m and v_m, respectively, and denoting the two-dimensional discrete Fourier transform of a function ϕ as $\hat{\phi}$ and its complex conjugate as $\hat{\phi}^*$, the total (or background) two-dimensional KE spectral density is (Durran *et al.*, 2017, equation 24)

$$\widehat{KE}_m(k_h) = \frac{\Delta x\Delta y\Delta k}{8\pi^2 n_x n_y}\left[\hat{u}_m(k_h)\hat{u}_m^*(k_h) + \hat{v}_m(k_h)\hat{v}_m^*(k_h)\right], \tag{3}$$

where k_h is the magnitude of the horizontal wavenumber, n_x and n_y are the number of grid points in the zonal and meridional directions, respectively, and Δx and Δy are the horizontal grid spacing in the zonal and meridional directions. The spectral density at each individual horizontal wavenumber (k_x, k_y) is added to the bin for which $k_h - \Delta k/2 < \sqrt{k_x^2 + k_y^2} \leq k_h + \Delta k/2$, where $\Delta k = 2\pi/1{,}500$ km^{-1} is the smallest resolved wavenumber corresponding to a full-domain wave. The perturbation kinetic energy spectral density $\widehat{KE'}_m$ is also calculated from Equation 3, except that the velocities are replaced by the difference from the ensemble mean, that is, $u'_m = u_m - \bar{u}$, where the overbar denotes the ensemble mean.[7]

[7]Essentially identical results are obtained if the perturbations are computed as a difference from the unperturbed control member, rather than treating

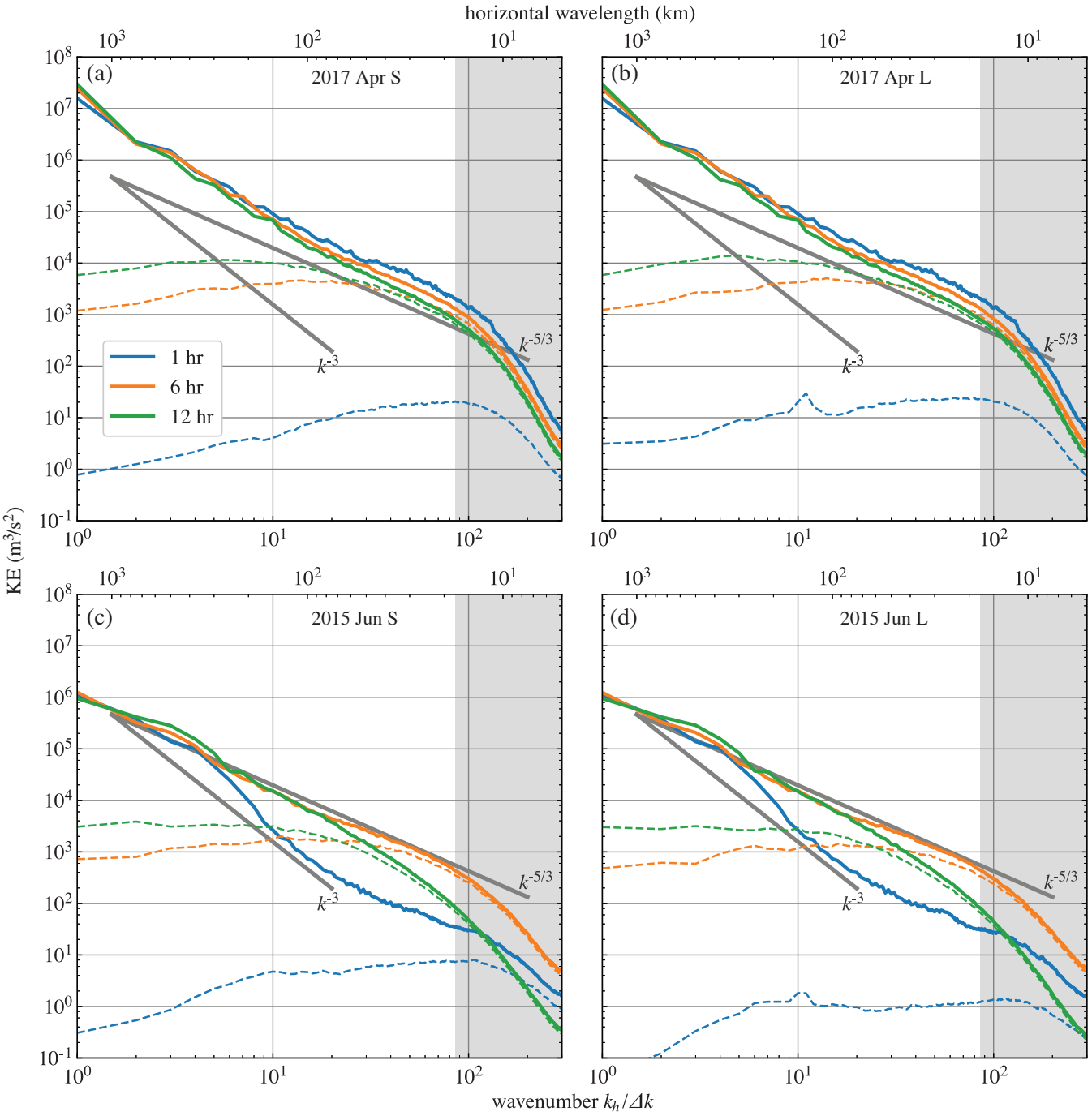

FIGURE 5 Spectra of total (solid lines) and perturbation (dashed lines) kinetic energy averaged over heights of 9–12 km for (a, b) April 2017 and (c, d) June 2015 cases. The S ensembles are shown in (a, c); the L ensembles in (b, d), both at times of $t_s = 1$ hr (blue lines), 6 hr (orange), and 12 hr (green). Solid gray lines are reference lines with slopes of $k^{-5/3}$ and k^{-3}. The gray shaded region denotes wavelengths shorter than $7\Delta x$, which are subject to numerical dissipation

For comparison with the Lorenz model, $\widehat{KE}$ and $\widehat{KE'}$ are computed as the average of $\widehat{KE}_m$ and $\widehat{KE'}_m$ over all ensemble members. Spectra are also scaled by a corrective factor (Durran *et al.*, 2017, equation 29) to reduce systematic noise introduced by binning. As noted in Section 2, spectra are calculated on the $1500 \times 1500\,\text{km}^2$ square analysis subdomains highlighted in blue in Figure 2 after detrending the data following the procedure in Errico (1985).

the control member as part of the ensemble, except that the amplitude of the perturbation spectrum approaches twice that of the background spectrum.

In Figure 5a,b, we show the spectra of KE and KE′ for the strongly forced April 2017 case from the S and L ensembles, respectively. For both ensembles, the KE spectra at $t_s = 1$ and 6 hr for the April 2017 case closely follow a $k^{-5/3}$ slope at scales of 10 km to about 200 km, and transition gradually toward an approximately k^{-3} slope at the largest scales. Wavelengths shorter than 17.5 km (Δx), indicated by shading, show a steep falloff of KE where numerical dissipation produces the strongest damping (Skamarock, 2004). This background spectrum is fairly constant in time, diminishing slightly in amplitude and slightly steepening by 12 hr

as the convection weakens. While the KE spectra are identical between the S and L ensembles, the KE′ spectra in the April 2017 case exhibit a very minor difference at $t_s = 1$ hr: the spectrum for the L case shows a small increase at the wavelength of the water-vapor perturbations (140 km), and slightly larger amplitudes at scales around and larger than that wavelength. By 6 hr, however, the error spectra are essentially identical in both the S and L cases. By comparison, the KE′ spectra in the weakly forced June 2015 case (Figure 5c,d) exhibit a much greater difference between the S and L ensembles in the error at $t_s = 1$ hr than those in the April 2017 cases: the errors in the S ensemble have grown faster across all scales than those in the L ensemble. There is still a small difference in the KE′ spectra between the S and L ensembles at 6 hr, but any difference is essentially negligible by 12 hr.

The background KE spectrum in the June 2015 case is notably different from that for all three other cases. A robust $k^{-5/3}$ slope is present only at the time of most vigorous convection, $t_s = 6$ hr in Figure 5c,d. This is not an artifact of model spin-up (recall the simulation is initialized at $t_s = -6$ hr), but rather is due to the limited duration of the active convection during this event (Figure 4b,d,f). At $t_s = 1$ hr (1 hr after the ensemble perturbations are introduced), the spectral slope around the 141 km wavelength of the large-scale perturbations is much steeper than k^{-3}. According to Lorenz's theoretical model of scale interactions, such a steep background spectral slope would greatly slow downscale error propagation (Lorenz, 1969; Rotunno and Snyder, 2008), potentially explaining why the initial large-scale KE′ perturbations propagate downscale much more slowly in this event than in the April 2017 case (compare with Figure 5b,d). After the diurnal maximum in convection in the June 2015 case, the power at small scales again falls off significantly, and by $t_s = 12$ hr has approached k^{-3}. The tight connection between these variations in convective activity and spectral slope supports the importance of convection in generating the observed $k^{-5/3}$ KE power spectrum in the mesoscales (Lilly, 1983; Durran and Weyn, 2016; Sun *et al.*, 2017, WD17).

To assess the scale-dependent loss of predictability more precisely, we calculate the KE′ saturation ratio, defined as the ratio of the ensemble average of perturbation kinetic energy about the ensemble mean divided by the ensemble average of total kinetic energy, that is, $\widehat{KE'}(k_h)/\widehat{KE}(k_h)$ (WD17). The KE′ saturation ratio is plotted as a function of wavelength in Figure 6 for each pair of S and L ensembles in all four observed cases. In all cases, the KE′ saturation ratio for both ensembles grows consistently over time and exceeds 80% at scales shorter than approximately 20 km by 12 hr. As suggested in Figure 5a,b, there is essentially no difference in error growth between the S and L ensembles in the April 2017 case, indicating that this strongly forced system is insensitive to the scale of the IC errors. In contrast, small-scale errors grow more rapidly in the weakly forced June 2015 case, although the difference between the S and L ensembles becomes small by 12 hr. The other weakly forced event, from June 2016, similarly exhibits a relatively greater sensitivity to small-scale IC errors. The May 2015 case, the second most strongly synoptically forced case, shows slightly more sensitivity to small-scale errors than the April 2017 case, but much less so than in the two June cases. The behavior of the spectra of the KE′ saturation ratio in Figure 6, and the relative rankings of the four cases in terms of synoptic-scale forcing, suggest that strong synoptic-scale forcing greatly diminishes, or eliminates, the sensitivity of convective systems to the scale of IC perturbations. Weakly forced cases, on the other hand, do exhibit more rapid error growth, particularly in the first 6 hr, if the IC errors are small in scale.

5 | FRACTIONS SKILL SCORE

The fractions skill score (FSS) was developed by Roberts and Lean (2008, hereafter RL08) as a means of quantifying the skill of observable parameters in high-resolution forecasts of convective events. It enables comparison of forecast and observed precipitation at any given spatial scale, focusing on localization errors, which are of vital importance in identifying locations threatened by severe weather. In all uses of the FSS in this study, the observed field corresponds to the precipitation field of the control member with no moisture perturbations, while the forecast fields are obtained from the perturbed ensemble members.[8] To calculate the FSS, the *observed* precipitation field is converted to a binary field, I_O, the value of which is one at each point where the observed precipitation exceeds a pre-defined threshold of hourly rainfall and zero otherwise. The precipitation in a local neighborhood of size n surrounding each grid point is then averaged using a spatial kernel $K_{(n)}$ centered at (i,j) to yield the array of observed values

$$O_{(n)i,j} = \sum_{k=1}^{n}\sum_{l=1}^{n} I_O\left[i+k-\frac{n+1}{2}, j+l-\frac{n+1}{2}\right] K_{(n)k,l}, \quad (4)$$

In Equation 4, the summations over k and l span an n-by-n sized square of grid points; the kernel $K_{(n)}$ determines which of these points are actually included in the sum and normalizes by the total number of such points. For a square neighborhood, $K_{(n)} = 1/n^2$ for all k, l. Here, to be consistent with two-dimensional spectral calculations, we define $K_{(n)}$ to be $1/N_{(n)}$ at the $N_{(n)}$ points satisfying $(k-(n+1)/2)^2 + (l-(n+1)/2)^2 \le ((n+1)/2)^2$ (that is, lying within a circle with a radius $(n+1)/2$) and 0 elsewhere. The binary field of *modeled* precipitation, I_M, is similarly defined using the modeled precipitation field and summed over every neighborhood with I_O replaced by I_M in Equation 4 to yield an array of forecast precipitation $M_{(n)}$ (RL08, equation 4).

[8] As discussed earlier, we analyze the internal predictability of the model, rather than its ability to reproduce observed weather.

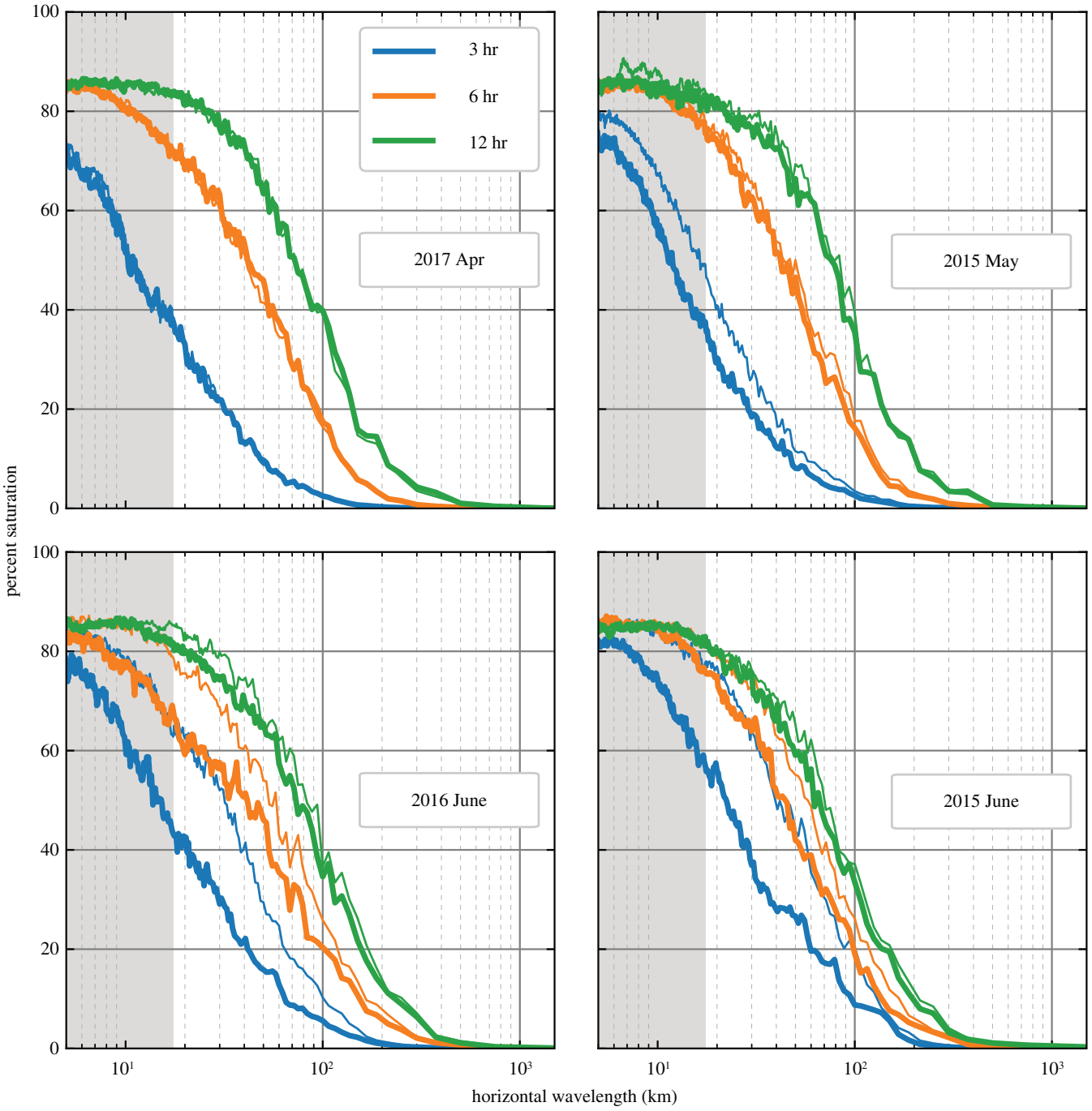

FIGURE 6 Spectra of KE′ saturation ratio (see text) for all four events, as labeled, at times of 3 hr (blue), 6 hr (orange), and 12 hr (green). The spectra for the S (L) ensembles with small-scale (large-scale) IC errors are denoted by the thin (thick) colored lines. The gray shaded region denotes wavelengths shorter than $7\Delta x$, which are subject to numerical dissipation

The mean square error $\mathrm{MSE}_{(n)}$ is computed such that

$$\mathrm{MSE}_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O_{(n)i,j} - M_{(n)i,j} \right]^2, \tag{5}$$

where n again denotes the neighborhood size, and N_x and N_y are the number of grid points in the x and y directions, respectively (RL08, equation 5). The FSS is defined as (RL08, equations 6 and 7)

$$\mathrm{FSS}_{(n)} = 1 - \frac{\mathrm{MSE}_{(n)}}{\mathrm{MSE}_{(n)\mathrm{ref}}}, \tag{6}$$

where

$$\mathrm{MSE}_{(n)\mathrm{ref}} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O^2_{(n)i,j} + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M^2_{(n)i,j} \right]. \tag{7}$$

A perfect forecast has a FSS of 1, while a completely missed forecast has a skill score of 0. RL08 suggest that a skilful forecast should have a skill score exceeding $\mathrm{FSS}_{\mathrm{uniform}} = 0.5 + f_0/2$, where f_0 is the fraction of points within the domain exceeding the precipitation threshold. In subsequent

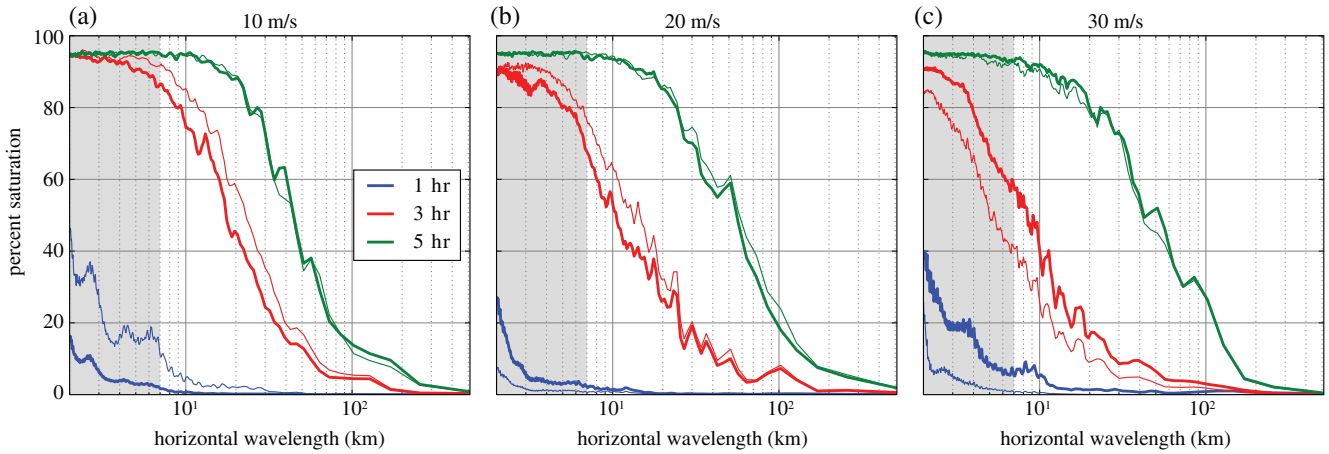

FIGURE 7 Spectra of KE′ saturation ratio for the idealized MCS simulations of Weyn and Durran (2017), adapted from their figure 7, at times of 1 hr (blue), 3 hr (red), and 5 hr (green). The spectra for the S (L) ensembles with small-scale (large-scale) IC errors are denoted by the thin (thick) colored lines. The gray shaded region denotes wavelengths shorter than $7\Delta x$, over which KE′ is most significantly influenced by numerical dissipation. ©American Meteorological Society

figures, we assume that the number of points with precipitation exceeding the threshold is small compared with the total number of grid points and approximate the skilful threshold as 0.5.

5.1 | FSS in idealized MCS simulations

We introduce the application of the FSS by presenting a contextual example from the idealized simulations of MCSs in WD17, where initial warm bubbles triggered convection in environments with horizontally uniform vertical profiles of temperature, humidity and wind with either 10, 20, or 30 m/s of 0–5 km vertical wind shear. Because wind shear organizes squall-line structure strongly, and because all the ensemble members were initialized with identical warm bubbles, the highest-shear case may be considered the most strongly forced idealized convective system in WD17. The ensembles were perturbed with near-surface humidity variations similar to the ones used in this work. The different wind shears produced both different storm structures and different degrees of variability among the ensemble members computed for each profile. For example, the simulations with 20 m/s of wind shear produced ensemble members having squall lines with wide variations in their north–south extent (WD17, figure 3). On the other hand, the simulations with 30 m/s of wind shear produced ensemble members having squall lines with only very minor visual differences (WD17, figure 4). From a practical perspective, a weather forecaster would assign much greater predictability to the 30 m/s shear cases, since all ensemble members showed nearly the same basic squall-line structure and spatial location of convection. Yet, as evident in Figure 7, which is reproduced from figure 7 of WD17 and is analogous to the preceding Figure 6, the spectrum of KE′ saturation ratio shows essentially identical predictability loss by 5 hr regardless of the environmental wind shear. Thus, even though the 30 m/s wind shear case appears much more predictable, the KE′ saturation-ratio metric is unable to distinguish it from less predictable cases.

If, instead, we analyze the FSS for the same idealized cases in WD17, we obtain a different result. Figure 8a–c shows the FSS for an hourly precipitation threshold of 1 mm at neighborhoods of radius 2, 8, and 32 km for all three idealized cases, where the scores for all 20 ensemble members are averaged together.[9] The thin (thick) lines correspond to ensembles initialized with small-scale 5.7 km wavelength (large-scale 90 km wavelength) moisture perturbations. To provide the "observed" values used in the computation of the FSS, the simulations from WD17 were supplemented with one additional simulation in which there were no initial perturbations for each level of environmental wind shear. Scores consistently improve as the neighborhood size increases, as expected, since the spatial location of convective elements is in better agreement when averaged over larger neighborhoods. All of the simulations have skill scores above 0.5 throughout the duration of the simulations, indicating relatively high confidence in the location of precipitation, but there are, nevertheless, substantial differences for different environmental shears. At 6 hr, the FSS for the 30 m/s wind shear case is very high: 0.87 for the 2 km neighborhood and 0.94 for the 8 km neighborhood, whereas the FSS for the 20 m/s wind shear case is much lower: 0.62 and 0.75 for the 2 and 8 km neighborhoods, respectively. These differences in the FSS match the visual difference in the spread of the synthetic radar reflectivity patterns among the ensemble members in the two cases, suggesting that the FSS can distinguish between them in a practical forecasting sense better than the KE′ saturation ratio. Yet the FSS is actually more optimistic about predictability than the KE′ saturation ratio,

[9]The horizontal resolution is 1 km, so these neighborhoods correspond to $n = 5$, 17, and 65.

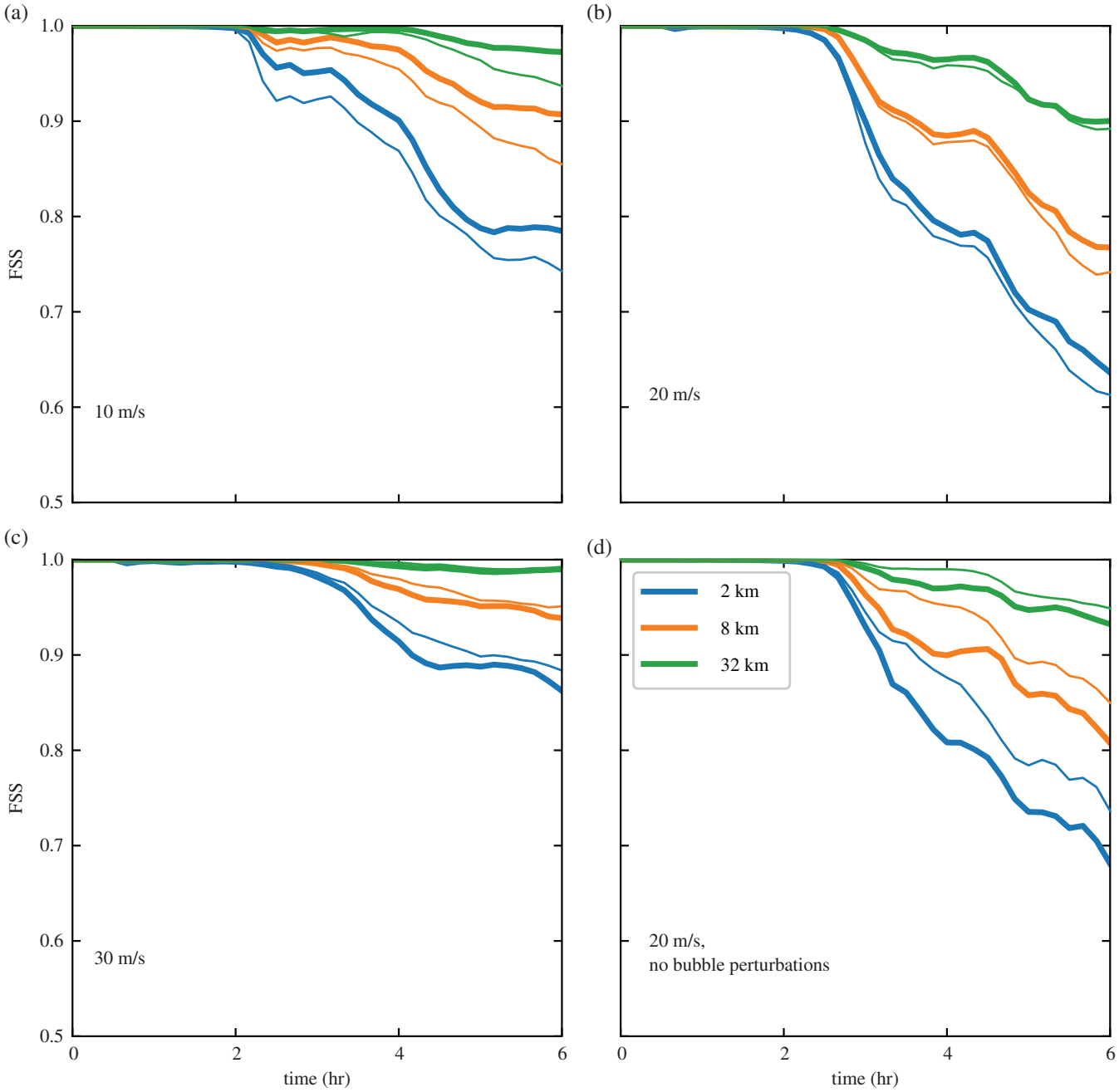

FIGURE 8 Ensemble-averaged fractions skill scores (FSS, see text) for the idealized MCS simulations of Weyn and Durran (2017) with (a) 10 m/s wind shear, (b) 20 m/s wind shear, (c) 30 m/s wind shear, and (d) 20 m/s wind shear, and dampened perturbations near the initial warm bubbles. The FSS is calculated for neighborhoods of 2 (blue), 8 (orange), and 32 km (green). Thin and thick lines correspond to the S and L ensembles, respectively

as even the smallest neighborhood of 2 km remains above the skilful threshold of 0.5 in all simulations at 6 hr.

For a given environmental shear, there are also minor differences between the FSS scores for the S and L ensembles. In the 10 m/s case (Figure 8a), the S ensemble has consistently worse skill scores than the L ensemble; the same is true in the 20 m/s case, although the difference is less pronounced. On the other hand, in the 30 m/s case, the FSS for the S ensemble is slightly higher than that for the L ensemble. This change in the relative rates of error growth between the S and L ensembles as the shear strengthens is also apparent in the spectra of the KE′ saturation ratio in Figure 7, particularly at 1 and 3 hr.

As detailed in WD17, the convection in every ensemble member in these idealized simulations is triggered by three identical warm bubbles, while the perturbations among the ensembles have random phase, and therefore a random spatial location, relative to those bubbles. To investigate whether the difference in skill scores between the S and L ensembles is produced primarily by modifications to the initial warm bubbles or by variations in the environment into which the storms propagate, we ran an additional pair of 20-member

ensembles in which the moisture perturbations in the vicinity of each warm bubble were essentially eliminated through multiplication of the perturbation moisture field by

$$a_{\text{damp}}(x,y) = 1 - \left\{ \sum_{n=1}^{3} \exp\left[-\frac{(x-x_n)^2}{2r^2} - \frac{(y-y_n)^2}{2r^2} \right] \right\}, \tag{8}$$

where n denotes the bubble number, $r = 10$ km is the bubble radius, and (x_n, y_n) is the center of the nth bubble. These ensembles are otherwise identical to the "S" and "L" ensembles with 20 m/s of vertical wind shear in WD17. The FSS for these simulations is shown in Figure 8d. In these simulations, an improvement in the FSS is observed relative to the control case (Figure 8b) for both the S and L ensembles, but there is much more improvement in the S ensemble, which now has higher skill scores than the L ensemble. This suggests that *if the perturbations are small in scale, errors generated on convective scales have greater influence, through their effect on initial storm development, whereas if the perturbations are large in scale, the errors produced in the synoptic-scale environment have greater influence, through their effect on convective organization.* Indeed, the period from about 3–4 hr in the simulations in Figure 8d, where the skill scores in the L ensemble drop off significantly, corresponds to a period of large-scale convective aggregation in the simulations (not shown). It should be noted, however, that this conclusion is drawn from idealized simulations of MCSs with controlled convective initiation, and therefore may not translate exactly to the real world.

5.2 | FSS in real-data cases

Returning to the real-data simulations, Figure 9 shows the FSS for each of the four cases, with values for each individual member of the S and L ensembles plotted along with that for the grand ensemble mean of all S and L simulations. The FSS is calculated over the same analysis domain as the KE spectra, denoted by the blue boxes in Figure 2. As for the idealized cases from WD17, an hourly precipitation threshold of 1 mm is used, but the neighborhood radii are increased to 5, 20, and 80 km, because of the coarser 2.5 km model resolution.[10] At early times around $t_s = 3$ hr, the weakly forced June 2015 and June 2016 cases exhibit the lowest FSS, in agreement with the greater KE′ saturation ratio of those cases at all scales compared with the more strongly forced cases (Figure 6). By $t_s = 6$ hr, the 5 and 20 km neighborhood FSS for the June 2015 case are about 0.8 and 0.95, respectively, while both are very close to 1 in the strongly forced April 2017 case. At $t_s = 12$ hr, by which time the KE′ saturation ratio has become nearly identical in all four cases, the FSSs continue to show significant differences. As shown in Figure 9, the June 2015 case clearly has the lowest predictability as measured by the FSS, followed by the June 2016 case. Nevertheless, the FSS indicates that the ensembles produce skilful forecasts at 12 hr even at the smallest neighborhood size of 5 km, unlike the KE′ saturation ratio, which indicates nearly total predictability loss at small scales.

For the June 2015 and June 2016 cases, and also to a very small degree in the May 2015 case, there is a bifurcation in FSS of individual ensemble members about the grand ensemble mean, particularly for the 5 km neighborhood. As labeled and illustrated most clearly in Figure 9d, the members with FSS below the mean are from the S ensembles, while those above the mean are from the L ensembles. As for the idealized cases shown in Figure 8, the FSS reveals a higher sensitivity to small-scale than to large-scale IC errors in environments with weakly organized convection. The relatively low FSS of the S ensembles in the weakly forced observed cases is also consistent with the relative difference in the KE′ saturation ratios (Figure 6), which indicated greater predictability loss in the June 2015 and June 2016 cases from IC errors at smaller scales.

Lastly, we consider FSS for heavy precipitation by increasing the hourly precipitation threshold from 1 to 10 mm. As expected, there is generally lower predictability for heavier precipitation. As shown in Figure 10, the FSS values are now lower, but the relative skill among the four observed events remains similar for 5 km radius neighborhoods. At radii of 20 and 80 km, however, the scores from the May 2015 case are comparable with those of the June 2015 and June 2016 cases, suggesting that there is less distinction between the three most weakly forced cases when considering high-intensity convection. Nevertheless, Figure 10 continues to demonstrate that the April 2017 case, with the strongest synoptic-scale forcing, still exhibits much better skill scores than the other cases.

6 | CONCLUSIONS

The first goal of this article was to compare the growth rates of small- and large-scale initial-condition errors in numerical simulations of observed severe convective systems. The KE′ saturation ratio, computed for each wavenumber as the ratio of the perturbation KE to the ensemble-mean background KE, revealed that, by 6 hr after the introduction of IC errors, the two cases with the strongest synoptic-scale forcing (April 2017 and May 2015) show essentially no difference in error growth between ensembles with equal absolute amplitude small-scale (14 km wavelength) and large-scale (141 km wavelength) IC errors. This is consistent with the idealized simulations of mesoscale convective systems in WD17, which exhibited similar insensitivities to the scale of IC errors. In contrast, the KE′ saturation ratios for the two cases with the weakest synoptic-scale forcing (June 2016 and June 2015) exhibited greater sensitivity to small-scale IC errors than to

[10]The neighborhood size n remains the same in the idealized and real-data cases.

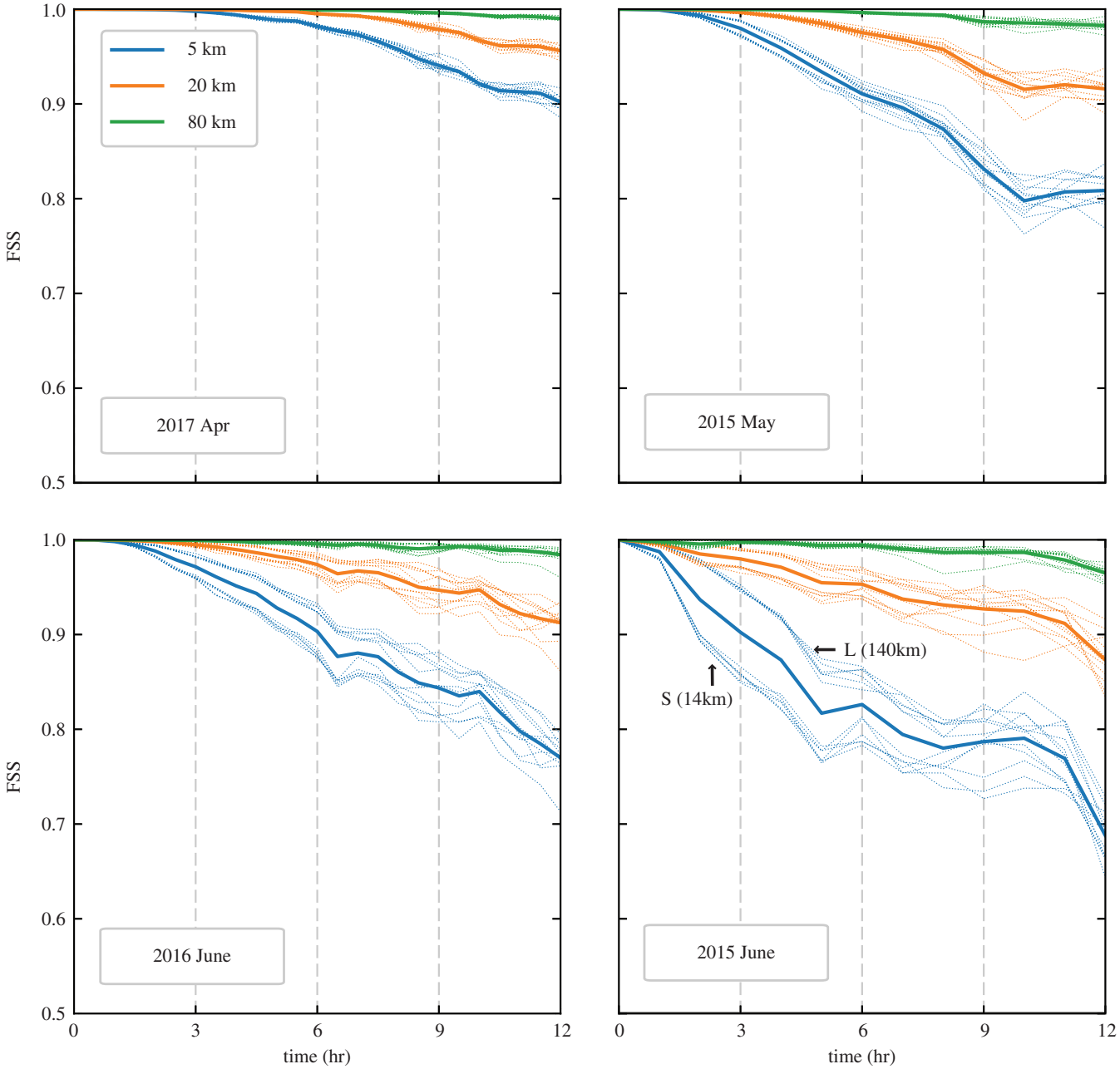

FIGURE 9 Fractions skill scores (FSS, see text) for the WRF simulations of all four events, as labeled. The FSS is calculated for neighborhoods of 5 (blue), 20 (orange), and 80 km (green). Thin dotted lines correspond to individual ensemble members, while thick solid lines correspond to the grand ensemble mean. Where there is a clear split in the FSS of individual ensemble members about the mean, those below the mean correspond to members of the S ensemble, while those above the mean are members of the L ensemble, as labeled in the bottom right panel

large-scale IC errors, although the sensitivity to initial scale was greatly reduced by 12 hr into the simulations.

The second goal of this article was to compare two predictability metrics, the KE′ saturation ratio and the neighborhood-based fractions skill score. In general, the FSS gave a clearer indication of relative predictability among both observed events and idealized convective systems in WD17, which were revisited in this study. The convection in the WD17 cases showed substantial variations in the distribution of synthetic radar reflectivity as a function of environmental wind shear, but by 5 hr into the simulations the KE′ saturation ratio suggested all ensembles achieved roughly the same degree of KE′ saturation, independent of the environmental shear. In contrast, the FSS for these same idealized systems showed substantial variations in skill as a function of the environmental shear, beginning as early as 3 hr and continuing until the end of the simulations. The relative predictability suggested by the FSS for these cases agreed qualitatively with that suggested by the ensemble spread of the synthetic radar reflectivity patterns. In particular, the highest FSS, and nominally greatest predictability, were associated with ensembles simulating the case with the strongest low-level wind shear and, in parallel to the real-world simulations, the

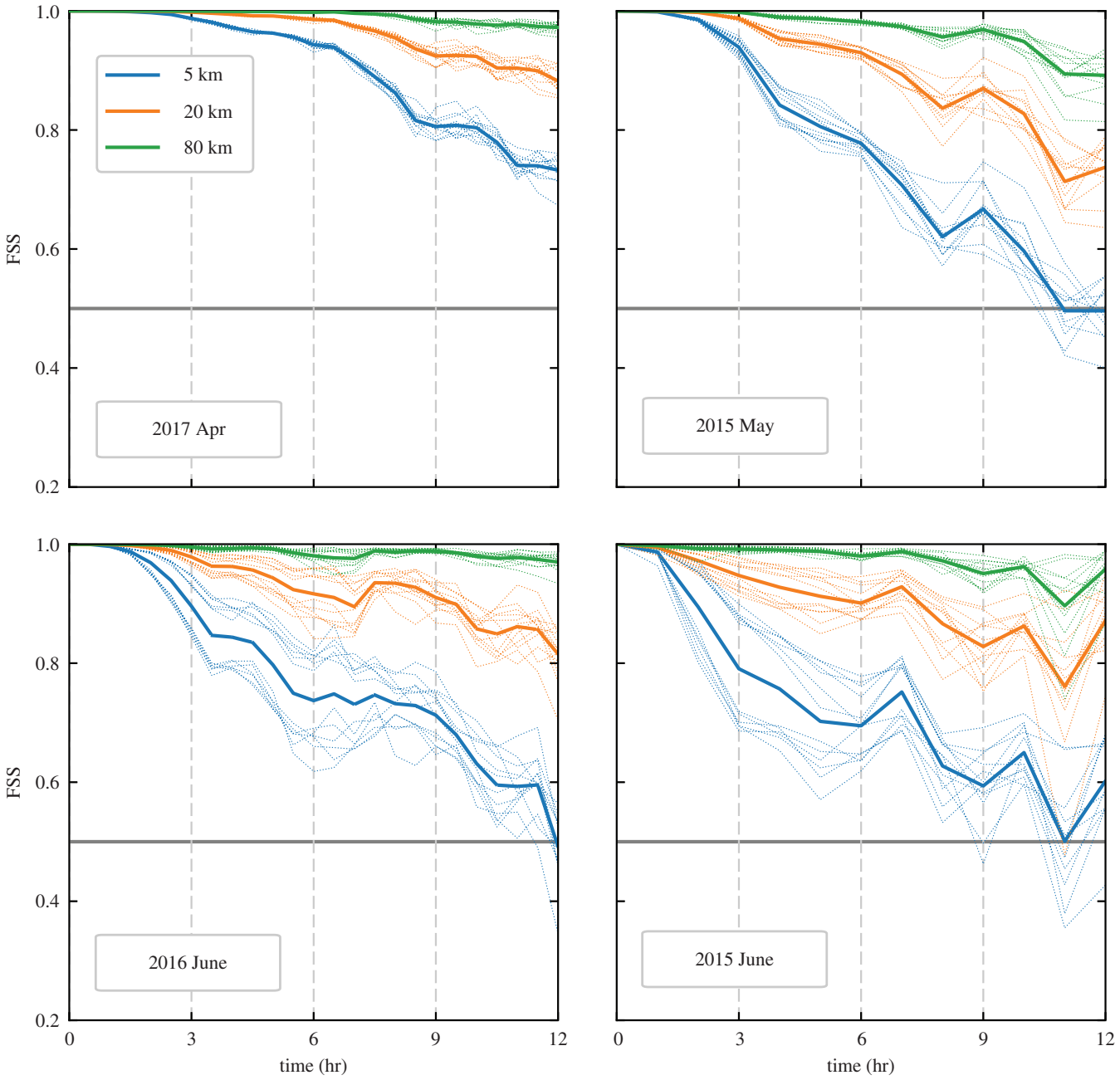

FIGURE 10 As Figure 9, but using the FSS calculated with a precipitation threshold of 10 mm

most strongly forced case. The FSS is likely able to distinguish better between the convective cases primarily because it measures precipitation, a direct and localized impact of convection, instead of horizontal winds, and because it measures in physical space, in contrast to the domain-wide Fourier transforms used to calculate the KE′ saturation ratio.

As with the idealized WD17 simulations, the FSS computed for ensemble simulations of the four observed events showed more variability from case to case than the KE′ saturation ratio. Those differences generally increased with time, up to 12 hr. The cases that were most strongly forced exhibited the highest FSS, while those with the weakest synoptic-scale forcing showed the lowest FSS, consistent with previous results showing greater sensitivity to initial conditions in weakly forced convective events (Done *et al.*, 2012; Surcel *et al.*, 2016). Like the KE′ saturation ratio, the FSS for the weakly forced June 2015 and June 2016 cases showed more sensitivity to the scale of the IC errors than did the cases with stronger forcing. For those weakly forced cases, the FSS for every member of the S ensemble remained lower than the FSS for every member of the L ensemble up to roughly 9 hr.

The more rapid KE′ growth, lower FSS scores, and associated lower predictability of the ensembles with small-scale IC perturbations in weakly forced cases appears to arise from the higher sensitivity of those cases to changes in the position and strength of the first convective elements, which are more significantly modified by short-wavelength perturbations. In contrast, when the IC perturbations are large in scale, they

modify the background environment in which storms organize and propagate more significantly, thereby impacting the strongly forced case more dramatically than those that are weakly forced. This behavior was isolated by comparing the FSS for an idealized ensemble in which the moisture perturbations in the immediate vicinity of the initial bubbles were removed with the FSS for the uniformly perturbed ensemble in the standard 20 m/s wind-shear case. When IC perturbations were imposed everywhere, including the locations of the initial warm bubbles (which were identical for all ensemble members), the FSS for the S ensemble was slightly lower than that for the L ensemble. However, when the perturbations in the immediate vicinity of the warm bubbles were removed, the FSS for the S ensemble became significantly higher than that for the L ensemble (Figure 8b,d).

An additional factor that may contribute to the higher sensitivity of the June 2015 case to small-scale perturbations is the major variation in the slope of the background KE spectrum with time. The background KE spectrum does not develop a robust $k^{-5/3}$ slope until diurnal convection develops later in the simulation. Instead, 1 hr after the perturbations are introduced (and 7 hr after the start of the simulation), the slope of the KE spectrum around the 141 km wavelength of the large-scale perturbations is steeper than k^{-3}, whereas the slope is roughly $k^{-5/3}$ around the 14 km wavelength of the small-scale perturbations. Theoretical models of error growth suggest that this difference in slope could reduce dramatically the rate at which errors spread downscale in the L ensemble compared with the upscale error growth in the S ensemble (Lorenz, 1969; Rotunno and Snyder, 2008).

In conclusion, these results contribute to our understanding of the case-to-case variability of predictability of severe convective systems. Efforts to reduce IC errors at horizontal scales around 10 km, by assimilating fine-scale observations and radar, for example, may be most effective in situations where the convection develops in the absence of significant synoptic-scale forcing. However, even then, at lead times beyond several hours, errors originating at large scales become equally as important as small-scale IC errors. On the other hand, the practical predictability of more strongly forced events may best be improved by reducing IC errors on scales of 100–400 km, as suggested on the basis of the idealized simulations in Durran and Weyn (2016) and WD17.

ACKNOWLEDGEMENTS

This research is funded by Grant N000141410287 from the Office of Naval Research (ONR). J.A.W. was partly supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate (NDSEG) Fellowship Program. We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation, and from Gordon, a supercomputer at the Navy DoD Supercomputing Resource Center (DSRC).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Jonathan A. Weyn http://orcid.org/0000-0002-4789-7594
Dale R. Durran http://orcid.org/0000-0002-6390-2584

REFERENCES

Dey, S.R., Roberts, N.M., Plant, R.S. and Migliorini, S. (2016) A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quarterly Journal of the Royal Meteorological Society*, 142, 1982–1996.

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. *Quarterly Journal of the Royal Meteorological Society*, 138, 638–648.

Duda, J.D. and Gallus, W.A. (2013) The impact of large-scale forcing on skill of simulated convective initiation and upscale evolution with convection-allowing grid spacings in the WRF. *Weather and Forecasting*, 28, 994–1018.

Dudhia, J. (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, 46, 3077–3107.

Durran, D.R. and Gingrich, M. (2014) Atmospheric predictability: why butterflies are not of practical importance. *Journal of the Atmospheric Sciences*, 71, 2476–2488.

Durran, D.R. and Weyn, J.A. (2016) Thunderstorms do not get butterflies. *Bulletin of the American Meteorological Society*, 97, 237–243.

Durran, D.R., Reinecke, P.A. and Doyle, J.D. (2013) Large-scale errors and mesoscale predictability in Pacific northwest snowstorms. *Journal of the Atmospheric Sciences*, 70, 1470–1487.

Durran, D.R., Weyn, J.A. and Menchaca, M.Q. (2017) Practical considerations for computing dimensional spectra from gridded data. *Monthly Weather Review*, 145, 3901–3910.

Errico, R.M. (1985) Spectra computed from a limited area grid. *Monthly Weather Review*, 113, 1554–1562.

Fang, X. and Kuo, Y.-H. (2015) A new generic method for quantifying the scale predictability of the fractal atmosphere: applications to model verification. *Journal of the Atmospheric Sciences*, 72, 1667–1688.

Flack, D.L.A., Gray, S.L., Plant, R.S., Lean, H.W. and Craig, G.C. (2018) Convective-scale perturbation growth across the spectrum of convective regimes. *Monthly Weather Review*, 146, 387–405.

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

Hong, S.-Y., Noh, Y. and Dudhia, J. (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134, 2318–2341.

Johnson, A. and Wang, X. (2016) A study of multi-scale initial condition perturbation methods for convection-permitting ensemble forecasts. *Monthly Weather Review*, 144, 2579–2604.

Johnson, A., Wang, X., Xue, M., Kong, F., Zhao, G., Wang, Y., Thomas, K.W., Brewster, K.A. and Gao, J. (2014) Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: dependence on background flow and method of perturbation. *Monthly Weather Review*, 142, 1053–1073.

Keil, C., Heinlein, F. and Craig, G.C. (2014) The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society*, 140, 480–490.

Leith, C.E. (1971) Atmospheric predictability and two-dimensional turbulence. *Journal of the Atmospheric Sciences*, 28, 145–161.

Lilly, D.K. (1983) Stratified turbulence and the mesoscale variability of the atmosphere. *Journal of the Atmospheric Sciences*, 40, 749–761.

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

Mapes, B., Tulich, S., Nasuno, T. and Satoh, M. (2008) Predictability aspects of global aqua-planet simulations with explicit convection. *Journal of the Meteorological Society of Japan*, 86A, 175–185.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the long-wave. *Journal of Geophysical Research*, 102, 16616–16663.

Nastrom, G.D. and Gage, K.S. (1985) A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *Journal of the Atmospheric Sciences*, 42, 950–960.

Nielsen, E.R. and Schumacher, R.S. (2016) Using convection-allowing ensembles to understand the predictability of an extreme rainfall event. *Monthly Weather Review*, 144, 3651–3676.

Potvin, C.K., Murillo, E.M., Flora, M.L. and Wheatley, D.M. (2017) Sensitivity of supercell simulations to initial-condition resolution. *Journal of the Atmospheric Sciences*, 74, 5–26.

Roberts, N.M. and Lean, H.W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136, 78–97.

Rotunno, R. and Snyder, C. (2008) A generalization of Lorenz's model for the predictability of flows with many scales of motion. *Journal of the Atmospheric Sciences*, 65, 1063–1076.

Selz, T. and Craig, G.C. (2014) Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Monthly Weather Review*, 143, 813–827.

Skamarock, W.C. (2004) Evaluating mesoscale NWP models using kinetic energy spectra. *Monthly Weather Review*, 132, 3019–3032.

Skamarock, W.C., Park, S.-H., Klemp, J.B. and Snyder, C. (2014) Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. *Journal of the Atmospheric Sciences*, 71, 4369–4381.

Sun, Y.Q. and Zhang, F. (2016) Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *Journal of the Atmospheric Sciences*, 73, 1419–1438.

Sun, Y.Q., Rotunno, R. and Zhang, F. (2017) Contributions of moist convection and internal gravity waves to building the atmospheric "−5/3" kinetic energy spectra. *Journal of the Atmospheric Sciences*, 74, 185–201.

Surcel, M., Zawadzki, I. and Yau, M.K. (2015) A study on the scale dependence of the predictability of precipitation patterns. *Journal of the Atmospheric Sciences*, 72, 216–235.

Surcel, M., Zawadzki, I. and Yau, M.K. (2016) The case-to-case variability of the predictability of precipitation by a storm-scale ensemble forecasting system. *Monthly Weather Review*, 144, 193–212.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M.A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J. and Cuenca, R.H. (2004) Implementation and verification of the unified Noah land surface model in the WRF model. In: *20th Conference on Weather Analysis and Forecasting Conference on Numerical Weather Prediction*, Seattle, Washington, USA. Available at: https://ams.confex.com/ams/84Annual/webprogram/Paper69061.html

Thompson, G., Field, P.R., Rasmussen, R.M. and Hall, W.D. (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: implementation of a new snow parametrization. *Monthly Weather Review*, 136, 5095–5115.

Vukicevic, T. and Errico, R.M. (1990) The influence of artificial and physical factors upon predictability estimates using a complex limited-area model. *Weather and Forecasting*, 188, 1460–1482.

Waite, M.L. and Snyder, C. (2013) Mesoscale energy spectra of moist baroclinic waves. *Journal of the Atmospheric Sciences*, 70, 1242–1256.

Weyn, J.A. and Durran, D.R. (2017) The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *Journal of the Atmospheric Sciences*, 74, 2191–2210.

Zhang, F., Snyder, C. and Rotunno, R. (2003) Effects of moist convection on mesoscale predictability. *Journal of the Atmospheric Sciences*, 60, 1173–1185.

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

Zhang, F., Bei, N., Rotunno, R., Snyder, C. and Epifanio, C.C. (2007) Mesoscale predictability of moist baroclinic waves: convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, 64, 3579–3594.

How to cite this article: Weyn JA, Durran DR. The scale dependence of initial-condition sensitivities in simulations of convective systems over the southeastern United States. *Q J R Meteorol Soc.* 2019;145 (Suppl. 1):57–74. https://doi.org/10.1002/qj.3367