What Is the Predictability Limit of Midlatitude Weather?

FUQING ZHANG AND Y. QIANG SUN

Department of Meteorology and Atmospheric Science, and Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania

LINUS MAGNUSSON AND ROBERTO BUIZZA

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

SHIAN-JIANN LIN AND JAN-HUEY CHEN

NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

KERRY EMANUEL

Lorenz Center, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

Understanding the predictability limit of day-to-day weather phenomena such as midlatitude winter storms and summer monsoonal rainstorms is crucial to numerical weather prediction (NWP). This predictability limit is studied using unprecedented high-resolution global models with ensemble experiments of the European Centre for Medium-Range Weather Forecasts (ECMWF; 9-km operational model) and identical-twin experiments of the U.S. Next-Generation Global Prediction System (NGGPS; 3 km). Results suggest that the predictability limit for midlatitude weather may indeed exist and is intrinsic to the underlying dynamical system and instabilities even if the forecast model and the initial conditions are nearly perfect. Currently, a skillful forecast lead time of midlatitude instantaneous weather is around 10 days, which serves as the practical predictability limit. Reducing the current-day initial-condition uncertainty by an order of magnitude extends the deterministic forecast lead times of day-to-day weather by up to 5 days, with much less scope for improving prediction of small-scale phenomena like thunderstorms. Achieving this additional predictability limit can have enormous socioeconomic benefits but requires coordinated efforts by the entire community to design better numerical weather models, to improve observations, and to make better use of observations with advanced data assimilation and computing techniques.

1. Introduction

Weather forecasting has improved dramatically since the introduction of numerical weather prediction (NWP) nearly six decades ago (Bauer et al. 2015). This has been accomplished through ever-increasing computing power, improved models running at ever-increasing resolution with more accurate representation of atmospheric physical processes, and more sophisticated four-dimensional data assimilating algorithms that can better ingest ever-increasing volumes and quality of in situ and remotely acquired observations (WMO 2015). A widely used measure of global NWP forecast quality is the anomaly correlation coefficient (ACC) of 500-hPa geopotential height between the forecasts and observations. In practice, 60% is usually used as a threshold for measure of skillful synoptic-scale weather forecast. Examining the evolution of ACC (Fig. 1), useful deterministic forecasts by arguably the most advanced NWP model at the European Centre for Medium-Range Weather Forecasts (ECMWF) could at best be made up to around 10 days: this number was 7 days 30 years ago (Simmons and Hollingsworth 2002; Bauer et al. 2015).
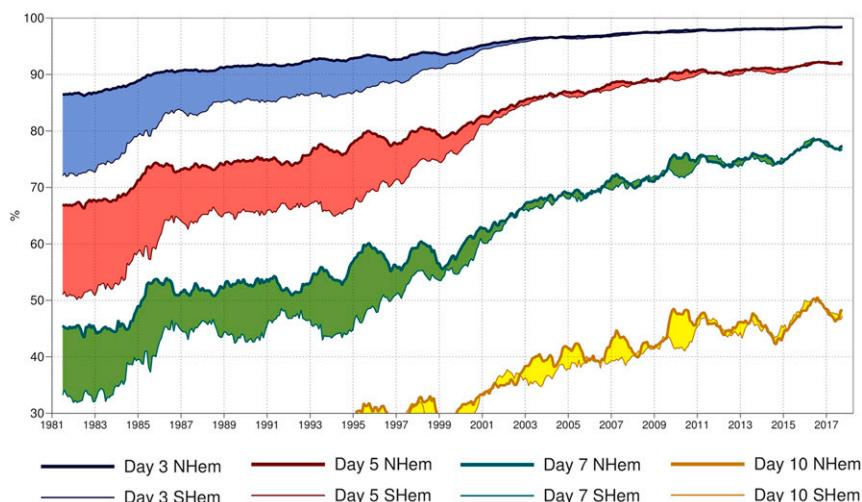
FIG. 1. Annual evolution the ECMWF NWP deterministic control forecast performance in terms of anomaly correlation of 500-hPa-height predictions. Shading indicates the different forecast skill between NH and SH, which has almost disappeared in recent years. This plot is directly adapted from ECMWF official website (https://www.ecmwf.int/en/forecasts/charts/catalogue/plwww_m_hr_ccaf_adrian_ts).

More improvements can be seen in the Southern Hemisphere where the traditional observing network is sparser but which has now been densely covered by satellite observations.

Improved NWP can have significant socioeconomic benefits by better predicting the occurrence of natural disasters, saving lives, and protecting property. For example, improved NWP is largely credited with the dramatic improvement in tropical cyclone prediction worldwide. The present tropical cyclone track forecast accuracy at the U.S. National Hurricane Center on average has gained almost a day lead time per decade (e.g., Zhang and Weng 2015): the yearly averaged 5-day lead-time track forecast error for the Atlantic basin in 2016 is smaller than the 2-day lead-time forecast error in 1990, which may have saved billions of dollars (Katz and Murphy 2015).

Yet improvement of NWP has limits. From the perspective of predictability, this concept of "atmospheric predictability limit" can be grossly categorized into intrinsic versus practical predictability (Lorenz 1996; Melhauser and Zhang 2012). As discussed in Ying and Zhang (2017), intrinsic predictability refers to "the ability to predict given nearly perfect representation of the dynamical system (by a forecast model) and nearly perfect initial/boundary conditions, an inherent limit due to the chaotic nature of the atmosphere and cannot be extended by any means" (Lorenz 1963, 1969; Zhang et al. 2003, 2007; Sun and Zhang 2016). Practical predictability, also commonly referred to as our weather prediction skill, is "the ability to predict given realistic uncertainties in both the forecast model and initial and

boundary conditions" (Lorenz 1982, 1996; Zhang et al. 2002, 2007). This practical predictability can be extended through reduction in key limiting factors of the forecast errors, including initial-condition errors, boundary condition errors, and model errors. All these factors, especially the initial-condition errors, have been greatly and could be further reduced with better NWP models ingesting high-accuracy observations using advanced data assimilation approaches along with advanced computing power (e.g., Zhang et al. 2009; Zhang and Weng 2015; Emanuel and Zhang 2016). Nevertheless, given our desire for better weather forecasting at all temporal and spatial scales, it is natural to ask whether an *intrinsic predictability* of the midlatitude weather exists. If yes, what is this inherent limit given nearly perfect NWP models with nearly perfect initial conditions? This is a crucial question that meteorologists have sought to answer ever since the beginning of NWP (e.g., Thompson 1957; Lorenz 1969; Leith 1971). Answering this question could provide guidance to society in decisions to enhance observing networks, improve models, and better assimilate observations into the forecast models.

Excellent work on this subject area has been pioneered by Lorenz, who first introduced the concept of "butterfly effect," which described the existence of the intrinsic predictability limit using a spectral turbulence model (Lorenz 1969). Lorenz showed that, for flow whose spectral slope is shallower than −3, error-doubling time decreases with decreasing scales, which led to an upscale error spreading and could provide an effective intrinsic

limit to the predictability of the flow. For flow with a slope steeper than −3, unlimited predictability might be achieved. This butterfly effect concept also inspires many subsequent studies using a hierarchy of turbulence models, which further confirmed Lorenz's theory (e.g., Leith and Kraichnan 1972; Rotunno and Snyder 2008; Durran and Gingrich 2014). While it remains unclear how these turbulence model results relate to our real atmosphere, it is widely accepted that the real atmosphere very likely also has an intrinsic limit of predictability (Palmer et al. 2014).

Estimates of this intrinsic predictability limit for a deterministic forecast can be made based on numerical integrations of model equations from two (identical-twin experiments) or more rather similar or even identical initial states (Lorenz 1963, 1969). The limit will occur at a time when the spread between these nearly identical runs starts to saturate and becomes as much as the spread among some randomly selected, but dynamically and statistically possible states. The accuracy of this kind of estimate is dependent on the accuracy of the forecast model used (Lorenz 1996). Earlier studies have used models of increasing complexity to investigate this intrinsic predictability and the error growth behavior of our atmosphere (e.g., Leith 1971; Daley 1981; Zhang et al. 2003, 2007; Mapes et al. 2008; Morss et al. 2009; Ngan et al. 2009). While these studies all agree on the existence of an intrinsic predictability limit for the respective weather systems, detailed error growth behavior differs among different models and different weather systems being studied. For example, in addition to an error cascade from smaller to larger scales (upscale growth; e.g., Lorenz 1969; Morss et al. 2009), some recent studies also show errors could grow spontaneously at all scales (up magnitude) without saturating at smaller scales (e.g., Mapes et al. 2008; Durran and Gingrich 2014).

Given that there is a degree of model dependency, many studies now tend to explore atmospheric predictability under more realistic frameworks with either regional (e.g., Zhang et al. 2003, 2007; Selz and Craig 2015; Ying and Zhang 2017) or global (e.g., Simmons and Hollingsworth 2002; Tribbia and Baumhefner 2004; Froude et al. 2013) NWP models. Regional models, which require boundary conditions, generally constrain longer-term error growth and propagation within the domain boundaries. Previous global predictability studies, on the other hand, usually do not have sufficient model resolutions to explicitly resolve mesoscale processes and moist convections, which have been shown to be critical for the initial error growth (Zhang et al. 2003, 2007; Selz and Craig 2015; Sun and Zhang 2016). Indeed, there has been increasing evidence that mesoscale error

growth shows similarity with the turbulence case under the shallower −5/3 kinetic energy spectrum, which is not well simulated in most coarse-resolution global NWP models (Augier and Lindborg 2013; Sun and Zhang 2016; Weyn and Durran 2017).

With recent advancement in computing capability, we now have entered a new era of global convection-permitting NWP models (Putman and Suarez 2011; Skamarock et al. 2014). Mapes et al. (2008) examined the predictability behavior of the atmosphere using global 7-km aquaplanet identical-twin simulations, with a focus in the tropics. Judt (2018) studied the atmospheric predictability through a pair of convection-permitting identical-twin simulations with the newly developed global Model for Prediction Across Scales (MPAS; Skamarock et al. 2014). Building on the findings of previous theoretical and modeling studies, our work here seeks to estimate the intrinsic limit of day-to-day weather predictability using ensemble simulations with the most advanced global NWP models at both ECMWF and U.S. NOAA. Our particular emphasis will be synoptic-scale weather systems dominated by baroclinic instability in the midlatitudes, where most of the world population resides. In particular, we showcase the practical versus intrinsic predictability limits of the global midlatitude weather during two periods in boreal winter and summer, respectively. These periods also endured two recent hazardous regional weather events: a wintertime cold-surge event affecting northern Europe in early January 2016, and a summertime rainfall-flooding event in China during July 2016. The choice of these two events are rather subjective and somewhat random with the intent to represent the typical midlatitude predictability while in the meantime covering some notable weather events in recent years. Nevertheless, neither of these two cases fall into the "forecast bust" cases using the criteria identified by Rodwell et al. (2013). Moreover, to the best of our knowledge, there were no severe weather outbreaks during these two periods in the midlatitude atmosphere of the Southern Hemisphere whose predictability will be simultaneously examined and compared with the Northern Hemispheric midlatitudes that have notable weather events.

Section 2 of this paper introduces the experiment design of our work, including the model we used and the perturbations added for each ensemble. Analysis for the ensemble spread from different perspectives are given in section 3, physical interpretation based on the results and hence the estimated predictability limit is also provided. Discussions on the limitations of current work and concluding remarks are presented in section 4.

2. Experimental design

This study adopted established methodologies introduced in the introduction for studying atmospheric predictability using perfect-model, identical-twin experiments where the ensemble members with minute initial-condition differences are explored. A series of ensemble simulations with the state-of-the-science global NWP model at ECMWF [viz., the Integrated Forecast System (IFS)], and U.S. Next-Generation Global Prediction System (NGGPS) with finite-volume cubed-sphere dynamical core (FV3) are designed to address the following two key questions: 1) What is the intrinsic predictability limit of multiscale midlatitude weather assuming a perfect model with nearly perfect initial conditions? 2) How much longer can the practical predictability be increased by reducing initial-condition uncertainties to different degree of accuracy?

a. Model details

1) ECMWF/IFS MODEL

The IFS control and ensemble forecasts presented herein uses the latest upgrade (cycle 41r2) of ECMWF, the highest-resolution-ever (~9 km) global operational NWP model. More details of this model upgrade can be found on the official website of ECMWF (http://www.ecmwf.int/). Different from previous versions, this new ECMWF IFS model implements a cubic octahedral reduced Gaussian grid (with spectral truncation denoted by $T_{CO}1279$) instead of the linear reduced Gaussian grid. With this cubic reduced Gaussian grid, the shortest resolved wave is represented by four rather than two grid points. The octahedral grid is also globally more uniform than the linear reduced Gaussian grid. In the vertical, the ECMWF model has 137 levels and a model top at 0.01 hPa. This corresponds to over 900 million grid points in total after this resolution upgrade.

In addition to resolution increase, the realism of the kinetic energy spectrum is also significantly improved with more energy in the smaller scales due to a reduction of the diffusion and removal of the dealiasing filter, enabled by the change to using a cubic truncation for the spectral dynamics. The semi-Lagrangian departure point iterations used to solve the primitive equations are also increased in the new model to remove numerical instabilities. The integration time step upgraded accordingly to 450 s. As intrinsic predictability implies the upper limit for our weather prediction given a nearly perfect model, no perturbation is applied to any model parameter and no stochastic physics scheme is adopted.

2) U.S. FVGFS SYSTEM

The newly developed Geophysical Fluid Dynamics Laboratory (GFDL) FV3 with Global Forecast System (GFS) physics (fvGFS) modeling system (Zhou et al. 2019; Hazelton et al. 2018; Chen et al. 2018) is used to further cross-examine the sensitivity of multiscale predictability to different model parameterizations and resolutions under future global convection-permitting NWP. This system was built during the NGGPS phase II, using the nonhydrostatic FV3 coupled to physical parameterizations from the National Centers for Environmental Prediction's GFS (NCEP/GFS). The GFDL FV3 was recently chosen as the dynamical core for the U.S. NGGPS as detailed in an online report (https://www.weather.gov/sti/stimodeling_nggps_implementation_atmdynamics); a report on this NGGPS development can also be found in Voosen (2017). In this study, we used the global uniform 3-km fvGFS configuration without ocean coupling. This model has 63 vertical layers and the model top is set at 0.6 hPa. The physical parameterizations include the Rapid Radiative Transfer Model for GCMs (RRTMG; Iacono et al. 2008) and the GFDL 6-class single-moment microphysics scheme (Chen and Lin 2011, 2013; Zhou et al. 2019). No cumulus scheme is adopted.

b. Ensemble experiments: EDA and EDA0.1

We first perform two types of ensemble experiments (denoted as EDA and EDA0.1, with explanations later in this subsection) with the current operational 9-km IFS model, running 10-member ensembles for 20 days beginning at 6 different times (3 consecutive days of 24–26 December 2015 and 3 consecutive days of 24–26 June 2016). All simulations are initialized at 0000 UTC. The initial condition and perturbations for the EDA ensembles are derived directly from the first 10 of 21 available operational ensemble four-dimensional variational data assimilation (4DVar) analyses (Bonavita et al. 2012) that represent the current realistic initial-condition uncertainties by the best-performing global NWP model (i.e., IFS at ECMWF). The design of the EDA ensemble using realistic initial-condition uncertainties is to explore more on the practical predictability side of the atmosphere as an assurance that the model used for this study could capture synoptic-scale dynamics and has typical predictive skills during the event periods selected for this study. Note that the EDA system uses the covariances derived from a coarser-resolution ($T_{CO}639$; ~16 km) ensemble forecast and thus small scales are not strongly constrained by observations. When initializing the model at higher resolutions, there would be a transient adjustment process (within hours; see Skamarock et al. 2014) to the small-scale energy

spectrum. This adjustment process will potentially excite a spurious cascade, which might bring faster initial error growth at these smaller scales. However, the impact of this process is expected to be small at synoptic scales and will be neglected in particular for the current study with a perfect-model assumption.

In comparison, the initial conditions for the EDA0.1 ensembles are perturbed with only 10% of the initial perturbations in the corresponding EDA ensembles centered at the control operational analysis of the IFS. With perturbation kinetic energy error only 1% of the current-day state-of-the-science analysis uncertainties, the EDA0.1 ensembles can be regarded as using nearly perfect initial conditions. The use of nearly perfect initial conditions, along with the use of the same model without physics perturbations, is in the spirit of perfect-model identical-twin experiments, which are designed to understand the intrinsic predictability limit of the atmosphere. Although the number of ensemble runs is still limited, to the best of our knowledge, this is the first time such a high-resolution global-model ensemble performing at the convection-permitting resolution is used for exploring the intrinsic limit of atmospheric predictability.

For the NGGPS FV3 model experiment, at 3-km grid spacing, computational costs permit us to run only one pair of identical-twin simulations starting at 0000 UTC 24 December 2015 for the Northern Hemispheric winter event: one initialized with the control member from the IFS model and the other initialized with the same initial perturbations as in member 1 of EDA0.1.

Although only one pair of identical-twin 3-km FV3 simulations can be afforded computationally for this study, it does offer a direct comparison of the error growth to the same pair of identical-twin simulations using the operational IFS model that has a different dynamical core and different resolution. In the meantime, for typical midlatitude synoptic systems of 5000 km in horizontal wavelength, there are about 5–10 such concurrent synoptic weather events in either hemisphere. In essence, this single pair of identical-twin experiments could represent a predictability estimate of multiple events under more general global statistics.

3. Predictability limit

To exemplify the limit of intrinsic predictability of day-to-day weather, we first select the January 2016 cold-surge event during which most areas of northern Europe experienced temperature anomalies below −5°C, as shown in the observational analysis (Fig. 2a). Near-normal temperature is observed over most of the contiguous United States and Canada except for a moderate warm anomaly over the Great Lakes region. The corresponding 15-day control forecast (Fig. 2b) by the ECMWF 9-km operational model IFS initialized at 0000 UTC 24 December 2015 failed to predict the northern Europe cold anomaly while it underpredicted the surface temperature over most of contiguous United States and overpredicted temperature over most of Canada.

A 10-member ensemble (EDA0.1), constructed by perturbing the control forecast with minute initial perturbations that are *an order of magnitude smaller* than the current analysis uncertainty, produced drastically different 15-day forecasts, each of which is nearly indistinguishable from a random sample of the climatology of this day. For example, member 1 of this reduced-perturbation ensemble (Fig. 2c) initialized also at 0000 UTC 24 December 2015 predicted a slightly above normal temperature (instead of the observed cold surge) over northern Europe while forecasting extremely cold conditions over most of the contiguous United States (instead of the observed normal to slightly warmer anomalies). The differences in predicted synoptic flow patterns between EDA0.1 member 1 and the unperturbed control forecast are comparable to the differences between the control run and the observational analyses represented by the sea level pressure maps in Fig. 2, except for the quasi-stationary planetary low pressure centers over the northern Atlantic and Pacific Oceans typical of climatological mean patterns. Failure of the control forecast (compared to observational analysis) and drastic forecast divergence between the control forecast and EDA0.1 ensemble member 1 that is perturbed with hypothetical minute initial perturbations (likely beyond the reach of future analysis accuracy) suggests a complete loss of predictability at the 15-day lead time (i.e., the intrinsic limit of day-to-day midlatitude weather predictability may not be extended beyond 2 weeks, at least in this case).

a. Evolution of ensemble spread

As mentioned in the introduction, the forecast uncertainty and the limit of predictability can be more systematically quantified by the evolution of the spread between the ensemble members and the time when it starts to saturate. Figure 3 shows midlatitude mean ensemble variance of the 500-hPa winds (a measure of ensemble kinetic energy spread) from two ensemble hindcasts initialized on three consecutive days (24–26 December). The choice of 500-hPa winds is because that it is directly linked to the kinetic energy spectrum, which will be discussed later. Nonetheless, metrics using geopotential height give very consistent results (not shown here). The EDA ensemble sets in Fig. 3 are initialized with the current realistic analysis uncertainties represented by the ECMWF ensemble of 4DVar analyses,
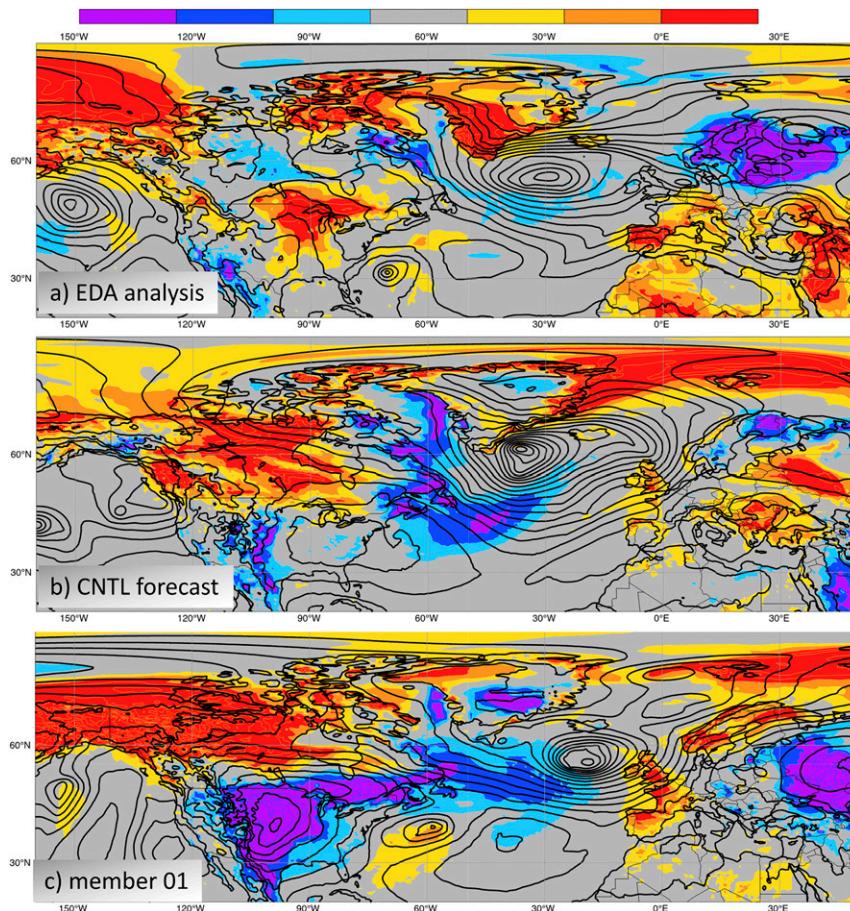
FIG. 2. Sample 15-day numerical weather predictions vs observations. The 2-m surface temperature anomalies (color shading) and the sea level pressure (contoured every 4 hPa) and valid at 0000 UTC 8 Jan 2016 for (a) the ECMWF observational analysis, (b) the 9-km IFS 15-day-lead-time control forecast initialized at 0000 UTC 24 Dec 2015 with the unperturbed EDA analysis, and (c) 15-day forecast from a member of the ensemble (EDA0.1) perturbed with 10% of the EDA ensemble perturbations that the current level of uncertainty in the EDA analysis. The temperature anomalies are calculated from a climatology based on a 20-yr, 10-member ensemble reforecast dataset using the same model with a 15-day lead time.

while EDA0.1 ensemble sets are initialized with nearly perfect initial conditions (initial kinetic energy error is 1% of that in EDA). As shown in Fig. 3a (normalized results shown in Fig. 3c), the spread of the EDA ensembles with realistic initial-condition uncertainties grows nearly two orders of magnitude larger before saturating at approximately 10–12 days, while the spread of the EDA0.1 ensembles, with minute initial perturbations (i.e., nearly perfect initial conditions), grows nearly four orders of magnitude larger before saturating at the same level as the EDA ensemble around 14–15 days (as a strong indication of the intrinsic predictability limit).

Similar quantitative statistics, representing intrinsic versus practical predictability limits assuming perfect model, can also be inferred from the same pairs of ensembles for the Southern Hemisphere (Fig. 4), as well as from pairs of Northern Hemisphere midlatitude 20-day 10-member global ensemble (Figs. 3b and 3d) initialized from three consecutive summer days (24–26 June) in 2016. During the 20-day simulation period in June, vast areas of the Yangtze River basin of China observed historical flooding (NASA; https://earthobservatory.nasa.gov/NaturalHazards/view.php?id=88467). Moreover, calculation of the ACC between the ensemble forecast and the observations over the global midlatitudes also gives quantitatively similar estimates for both the practical and intrinsic predictability limits (Fig. 5), with correlation dropping to 60% at around 10 days for the EDA ensemble and 13–15 days for EDA0.1 ensemble.

The growth of the ensemble variance, representative of the forecast error, fits surprisingly well with the simple
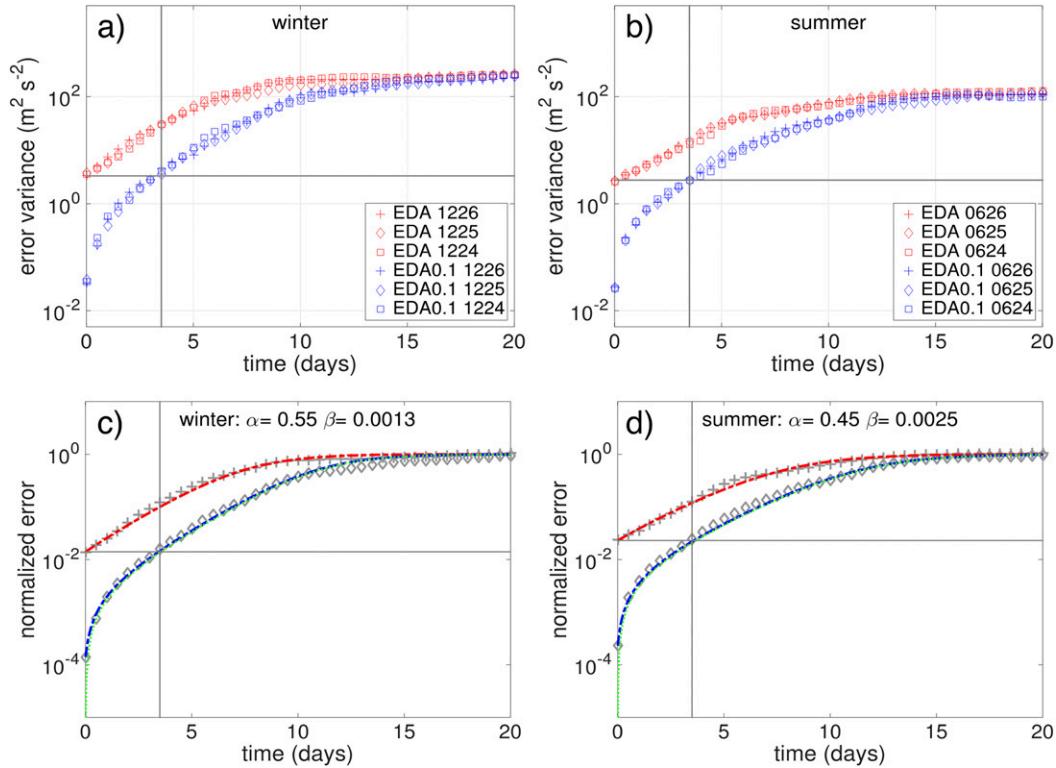
FIG. 3. (top) Evolution of forecast uncertainty growth in terms of ensemble variance of 500-hPa wind energy averaged over the NH midlatitudes (40°–60°N) for three ensemble simulations with current-day realistic initial-condition uncertainties (red symbols) and the corresponding ensembles with minute (1%) initial-condition errors (blue symbols) initialized on (a) 3 consecutive days for the winter (24–26 Dec 2015) and (b) 3 consecutive days in the summer (24–26 Jun 2016). (bottom) Evolution of normalized error variances (gray symbols) averaged for the (c) winter and (d) summer cases, as well as the fitted error growth curves (red for EDA, blue for EDA0.1) from Eq. (1). The units for α and β are day^{-1}. Green lines in (c) and (d) show the error growth curve according to the fitted Eq. (1) if the initial-condition error is further reduced to 1×10^{-10}. Note the log scale of the perturbation energy.

error growth model that was originally proposed in Lorenz (1982), modified later (Dalcher and Kalnay 1987; Reynolds et al. 1994) and here as

$$\frac{d\varepsilon(t)}{dt} = [\alpha\varepsilon(t) + \beta][1 - \varepsilon(t)]. \qquad (1)$$

Here $\varepsilon(t)$ is the normalized error where $\varepsilon \sim 1$ means it reaches the maximum or becomes saturated, and α is the synoptic-scale error growth rate. Previous studies (e.g., Magnusson and Källén 2013) usually use β as a measure for model error. Given that we are comparing between different ensemble members using the same forecast model, β here represents the error growth rate induced by the intrinsic upscale error propagation such as from small-scale moist processes (e.g., convection) even when we have nearly perfect initial condition (Sun and Zhang 2016). Figures 3c, 3d, 4c, and 4d show the evolution of normalized error averaged for both the winter and

summer cases, respectively, as well as the fitted error growth curves from Eq. (1).

Figures 3 and 4 show that exponential error growth (quasi-linear line in the logarithmic plot) dominates the first few days of the EDA ensembles, with a growth rate determined by α. The β term has little impact on the error growth curve for the EDA experiment due to relatively large initial-condition error. However, compared with EDA, much faster initial error growth is observed for the EDA0.1 ensembles. We can also deduce that the error growth rate (slope of the error growth curve in Figs. 3 and 4) in EDA0.1 will increase with decreasing ε, implying that there will eventually be diminishing returns from further reducing the initial-condition errors. This "superexponential" initial error growth in EDA0.1 is caused by the presence of the β term (representing the intrinsic upscale error growth and propagation from small scales) in Eq. (1). For example, the green line in Figs. 3c and 3d shows the
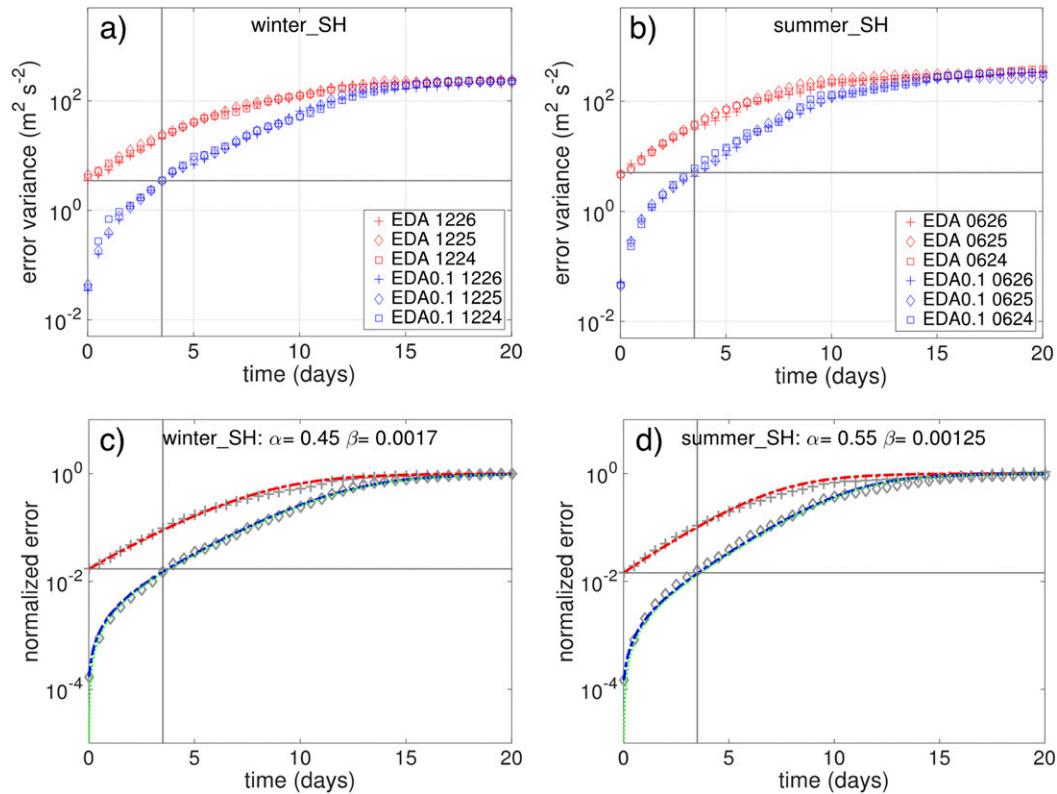
FIG. 4. As in Fig. 3, but for SH (40°–60°S).

predicted error growth curve derived from Eq. (1) when the initial-condition error is reduced to 1.0×10^{-10}. It is nearly identical to the blue line, which means there is not much more room for improvement. In other words, if Eq. (1) holds, further reduction in the initial-condition or model error would not help extend our forecast lead time much longer (maybe only in hours or even minutes).

The errors in EDA0.1 grow to an amplitude similar to the EDA initial ensemble spread in 3–4 days. Subsequent error growth and saturation in the EDA0.1 ensembles mimic those of the EDA ensembles except for a 3–4-day delay in forecast lead times. The overall reference error kinetic energy saturates ($\varepsilon \sim 1$) at around 10–12 days for all the EDA ensembles and 14–15 days for all the EDA0.1 ensembles. This remains true for
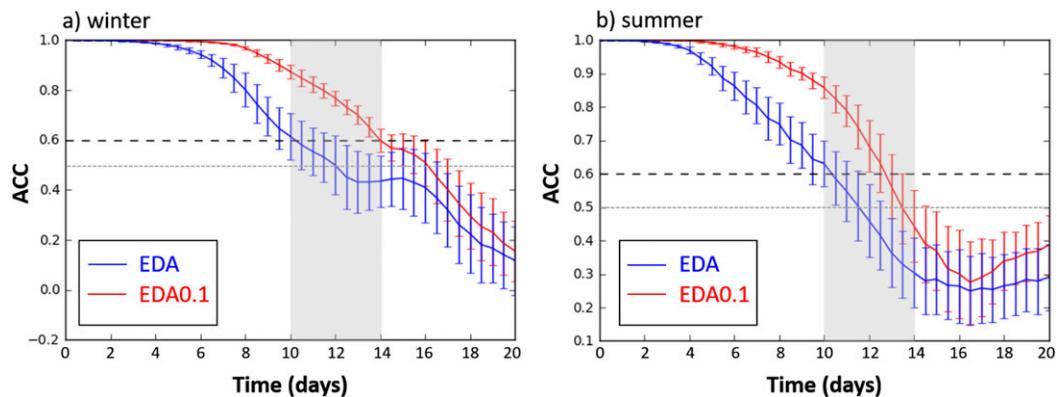
FIG. 5. Anomaly correlation coefficient of 500-hPa heights averaged over the Northern Hemisphere midlatitudes (30°–60°N) for (a) the January 2016 case and (b) the July 2016 summer case. The blue line shows the ACC for all of the EDA members, and the red line shows the results for EDA0.1 members.
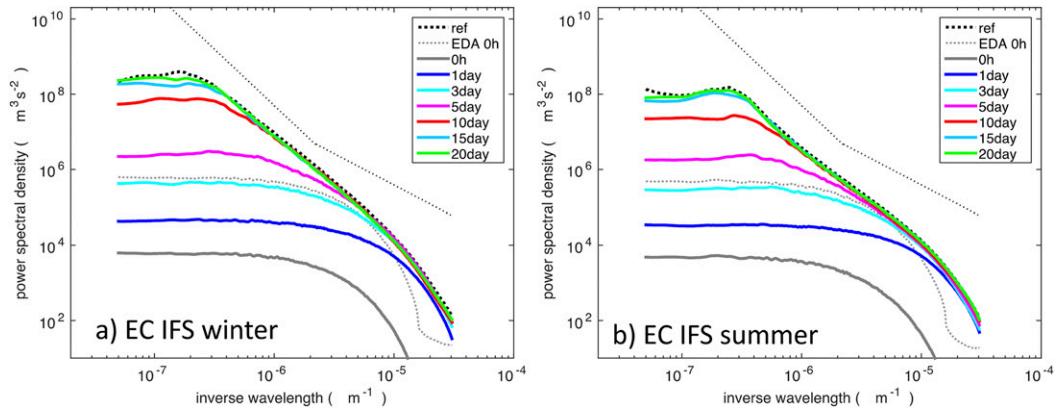
FIG. 6. Forecast error growth and saturation for different horizontal scales. Evolution of ensemble-mean error spectral kinetic energy (colored lines) averaged over NH midlatitudes (40°–60°N) for three 9-km, 10-member ECMWF IFS global ensemble simulations with minute (1%) initial-condition errors (solid gray lines) initialized on 3 consecutive days for (a) the winter period and (b) the summer period. The reference background kinetic energy spectra are derived from the respective 20-day mean spectra of the control forecast (dark dotted line). The corresponding initial spread of the EDA ensemble with realistic analysis uncertainties are shown for comparison (light-gray dotted lines). The two straight dotted line segments denote the spectral slopes of −3 and −5/3.

different initialization times and for both the winter and summer days of the Northern Hemisphere and the Southern Hemisphere.

These unprecedented high-resolution 9-km global ensembles of a state-of-the-science NWP model, initialized with both realistic and nearly perfect initial-condition uncertainties, suggest that the ultimate limit of midlatitude day-to-day weather predictability is about 2 weeks, but there is still a potential of 3–5 more days of additional forecast lead time to be gained through improving the current practical predictability, which is about 9–10 days. Such improvements may be gained from reducing initial-condition and model uncertainties through better observations, better data assimilation, and better forecast models running at higher resolution with ever-increasing computing capability.

b. Spectral analysis

While 3–5 days serves as the estimated potential for extended weather forecast lead time, the atmospheric predictability limit is also scale dependent. For example, small-scale thunderstorms are much less predictable than the synoptic system in which they are embedded. Therefore, it is important to examine the scale dependence of predictability limit. Spectral decomposition of perturbation kinetic energy across all zonal wavenumbers averaged over the midlatitudes (40°–60°N) for both winter and summer periods are displayed in Fig. 6. The corresponding spectra for the Southern Hemisphere midlatitudes (40°–60°S) are shown in Fig. 7.

The kinetic energy spectrum here is calculated as in (Skamarock 2004). We have chosen to compute the one-dimensional (1D) spectrum of the velocity fields along zonal direction. The advantage of this 1D spectrum is that we could fully utilize the periodicity of the global model in the zonal direction while focusing on the midlatitude only. Let $u_{i,j,n}$ and $v_{i,j,n}$ denote the zonal and meridional velocity components for the nth ensemble member, subtracting the ensemble mean fields first if we are calculating the kinetic energy spectra for the perturbations. For the spectra in Fig. 8, the differences between the perturbed run and the unperturbed run are used. The Fourier transforms of the velocity components $\hat{u}_{j,n}(k)$ and $\hat{v}_{j,n}(k)$ are then computed along the zonal direction for each ensemble member and all the meridional j indices. Then the kinetic energy spectra density can be written as

$$E_{j,n}(k) = \frac{\Delta x}{2N_x}[\hat{u}_{j,n}(k)\hat{u}^*_{j,n}(k) + \hat{v}_{j,n}(k)\hat{v}^*_{j,n}(k)], \quad (2)$$

where N_x is the number of grid points along the zonal direction of the model. The asterisk denotes the complex conjugate. We can then average $E_{j,n}(k)$ over j and n to get the kinetic energy spectrum for the full ensemble and the latitude band of interest (40°–60°N for the midlatitudes; the results are not very sensitive to this choice; the 30°–60°N average give very similar plots). When the spectrum of the perturbation kinetic energy (amplitude of "noise") at a given wavelength reaches the reference background spectral kinetic energy (signal to be predicted), it is saturated, after which no single deterministic forecast will have any predictive skill.

Consistent with Fig. 3, Fig. 6 also shows that it takes slightly more than 3 days for the perturbation kinetic energy in the reduced-perturbation ensemble (EDA0.1)
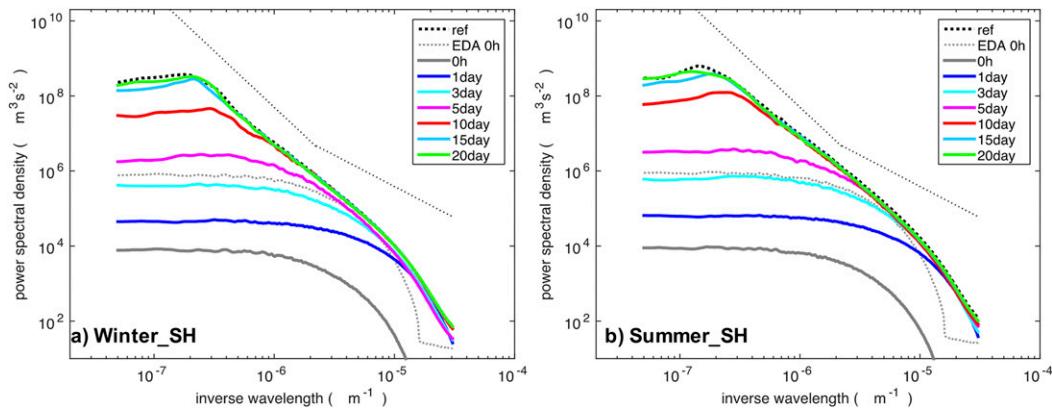
FIG. 7. As in Fig. 6, but for SH (40°–60°S).

to grow two orders of magnitude across all resolvable wavelengths to a level comparable with the realistic analysis uncertainty represented by the EDA for the periods of the winter and summer events, respectively. Also, Fig. 6 shows that the perturbation spectral kinetic energy from the EDA0.1 ensemble saturates at the amplitude of the reference kinetic energy across all synoptic scales by 15 days, again consistent with the overall intrinsic predictability limit estimated from Fig. 3.

Moreover, saturation time for different scales is different. With reduced initial-condition uncertainties, as in EDA0.1, forecast error first saturates at smaller scales, then subsequently grows rapidly in magnitude and in scale, consistent with past regional modeling studies (Zhang et al. 2007; Selz and Craig 2015; Sun and Zhang 2016). A simple estimation from Fig. 6 shows that

the forecast error saturation time (and thus intrinsic limit of predictability) is less than 3 days for horizontal scales less than 200 km, less than 5 days for horizontal scales less than 400 km, and less than 10 days for horizontal scales less than 1000 km. Also worth noting, Fig. 6 illustrates the synoptic-scale predictability in terms of 500-hPa horizontal winds. Much more limited predictability is expected for vertical velocity and instantaneous precipitation rate forecasts (Bei and Zhang 2007, 2014), which possess very different reference energy spectra.

c. Sensitivity study using U.S. NGGPS model

Although we could only afford to perform one pair of 10-day forecasts using the U.S. NGGPS model based on the FV3 dynamical core with 3-km convection-permitting
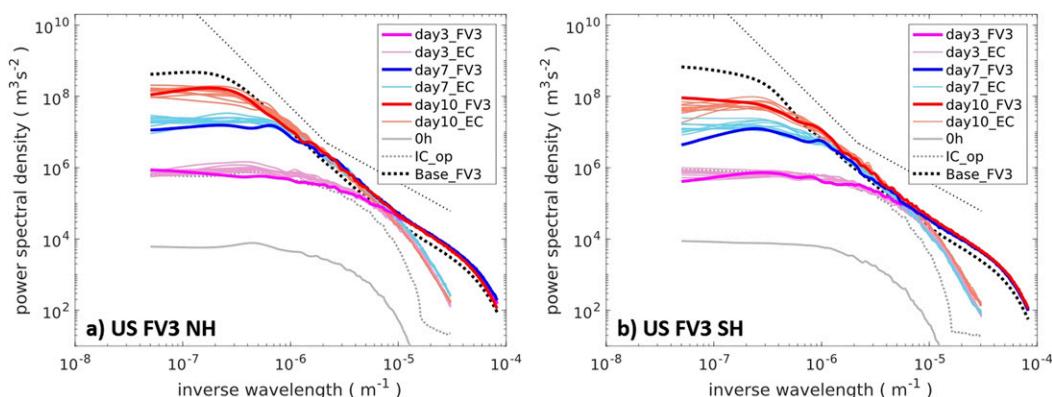
FIG. 8. Sensitivity of error growth and saturation with the 3-km FV3 model. Evolution of perturbation spectral kinetic energy between two forecasts (one initialized with the control EDA analysis and the other perturbed with the same initial perturbations as in EDA0.1 ensemble member 1 performed with the 3-km U.S. FV3 model for (a) NH and (b) SH. Computational costs permit us to run only one pair of 10-day simulations starting at 0000 UTC 24 Dec 2015 for the winter event. Thin lines show results for EC model calculated based on the differences between 10 perturbed EDA0.1 members and the control run. The FV3 results lie well within the 10 ensembles produced by the EC model. Note at small scales, different with the perturbation kinetic energy spectra in Fig. 6, the difference spectral kinetic energy has a saturation level that is twice the background spectrum.

horizontal grid spacing, the results (Fig. 8) show that for both winter- and summer-hemispheric midlatitudes, such limits are rather insensitive to the forecast model or resolution, and likely arise from the intrinsic dynamics of the atmosphere (Zhang et al. 2007; Rotunno and Snyder 2008; Sun and Zhang 2016). Due to the use of a higher horizontal resolution, the NGGPS FV3 model better resolves the small-scale atmospheric motions as can be seen by the extended background energy spectrum at smaller scales (Fig. 8). Yet the evolution of the forecast error in the NGGPS FV3 model, as reflected by the perturbation kinetic energy spectrum at 500 hPa, does not show significant differences with that in the ECMWF model. Once again, we find that after 3–5 days the differences between two initially nearly identical runs is comparable with our current operational analysis uncertainty (Fig. 8).

This consistency between two completely different models (with different dynamical cores, different physics, and different resolutions) also strengthens our confidence that these two state-of-the-science NWP models are ''appropriate'' to assess the intrinsic predictability limit, at least for the periods examined in this study. It is safe to say that minute uncontrollable initial-condition uncertainties originating from convective and mesoscale instabilities can grow upscale and will eventually limit the predictability of various weather systems at increasingly larger scales. The impacts of the background governing dynamics and instabilities on the limits of intrinsic predictability may also be inferred from the differences in Eq. (1) fitted values of α (synoptic-scale error growth rate likely controlled by synoptic instabilities such as baroclinicity) and β (upscale error growth rate likely controlled by small-scale instabilities and moist physics including convention). As denoted in Figs. 3c and 3d, a larger value of α and a smaller value of β are derived from the winter cold-surge event than those derived from the summer flooding event. This is consistent with stronger baroclinicity and weaker convective instability in the winter than in the summer, although more research is needed to further quantify such relationships (Reynolds et al. 1994; Magnusson and Källén 2013).

d. Possible dynamic processes

Given both models agree on the forecast lead time we could gain from reducing initial-condition uncertainty, the question then arises, What dynamic processes control the error growth and eventually limit the predictability of midlatitude weather? While this is surely an important question and needs future research (Rosinski and Williamson 1997; Magnusson 2017), some details on the evolution of the ensemble spread during specific cases may give us some insights and guidance into this question.
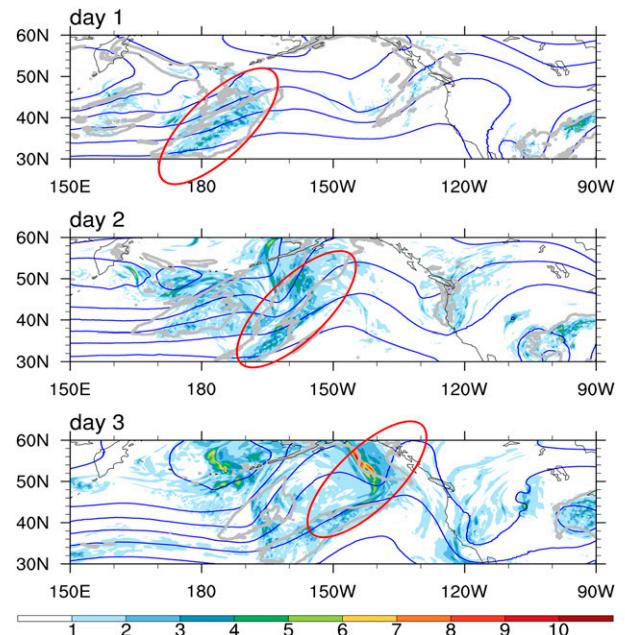
FIG. 9. Evolution of the ensemble spread of 500-hPa meridional winds (color-shaded) for the first 3 days in EDA0.1 integrated from 0000 UTC 26 Dec 2015. Blue contours are the 500-hPa geopotential heights, and gray contours indicate the regions with 12-h precipitation greater than 0.1 mm. Red circles imply the movement of the developing low pressure synoptic storms.

Figure 9 shows the evolution of the ensemble spread of 500-hpa meridional winds for the first 3 days in EDA0.1 integrated from 0000 UTC 26 December 2015, especially focused on a developing extratropical cyclone over the west coast of the United States. The blue contour is the 500-hPa geopotential height and the gray contour is the region with 12-h precipitation greater than 0.1 mm. We can see that, at day 1 the ensemble spread first shows up in the precipitating region. Then the spread increases and moves with the synoptic system, propagating both upstream and downstream in the meanwhile. After 3 days, the ensemble spread could be found anywhere in the midlatitude bands, although with a maximum in the synoptic storms. The key idea that minute perturbations will first generate errors in small-scale moist convective systems and the errors then grow upscale is consistent with previous studies (Zhang et al. 2007; Sun and Zhang 2016). We could also take a first look at the propagation of the ensemble spread using the Hovmöller diagram, which is plotted in Fig. 10 for the EDA0.1 ensembles initialized at 0000 UTC 24 December 2015. As marked subjectively by different type of arrows, there exist at least three characteristic pathways for the error to evolve and grow through time globally. The dotted ones, mainly shown up in the first few days, have an approximate eastward speed of
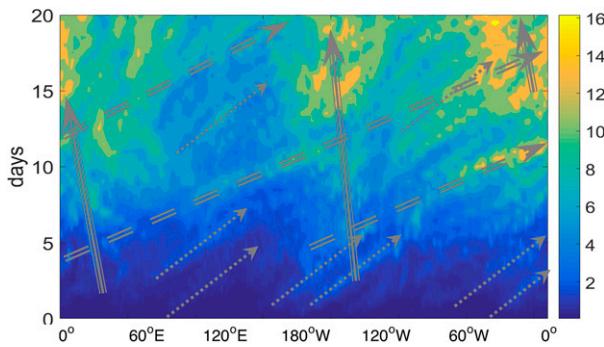
FIG. 10. Longitude–time Hovmöller diagram for the mean ensemble spread in terms of root-mean-square difference of kinetic energy $(m\,s^{-1})$ derived from the 9-km, 20-day, 10-member ECMWF IFS global ensemble initialized at 0000 UTC 24 Dec 2015 as a demonstration of different modes of error propagation that lead to the error saturation and intrinsic predictability limits (see text for details).

$10\text{–}15\,m\,s^{-1}$ and are consistent with the phase speed of individual synoptic weather systems. The double-dashed arrows that have an approximate eastward speed of $25\text{–}30\,m\,s^{-1}$ likely follow the downstream energy propagation of different baroclinic wave packets, and the thick solid arrows with a slow westward progression signal the error enhancement near the quasi-stationary planetary-scale low pressure centers. These error growth pathways expand beyond the multistage error growth mechanisms identified in a previous regional-scale predictability study (Zhang et al. 2007) and will be examined in more detail in our future study.

4. Concluding remarks

The promising finding of the current study is that, assuming the current-generation state-of-the-science NWP models could capture the most essential physical processes in the real world, we can further improve the forecast accuracy of day-to-day weather events such as the ones we discussed, by up to 5 days, if we reduce the initial-condition uncertainties by a factor of 10. In particular, we examined the predictability of weather forecasts in two showcase studies, and we have looked at the multiscale midlatitude error evolution across different spatial scales. Our study suggests that we are currently still quite far from the ultimate limit of predictability, and it is apparent that we have ample room for further improvement in the day-to-day weather predictability likely for decades ahead.

More quantitatively, it can be inferred from Figs. 3 and 6 that reducing the current initial-condition error represented by EDA by about 20% (~40% smaller in error kinetic energy) can potentially lead to a gain of one more day, and reducing by about 50% for a gain of two more days of additional predictable forecast lead time. Achieving this additional predictability limit can greatly benefit society by saving lives and property but requires continued coordinated efforts by the entire meteorology community and beyond to design more accurate NWP models performing at refined resolutions, improve and enhance the observing techniques and networks, and make better use of observations with advanced data assimilation and computing techniques.

It is possible those two individual local weather events (cold surge and floods) have slightly different predictability than more typical weather patterns but our findings here are based on mean error growth statistics of the global midlatitudes averaged over many wavelengths and multiple initialization times, not just the regions of the localized hazardous events. Additional calculations that exclude these two local events show consistent results (Fig. 11), as do the calculations in the opposing hemispheric midlatitudes that have no
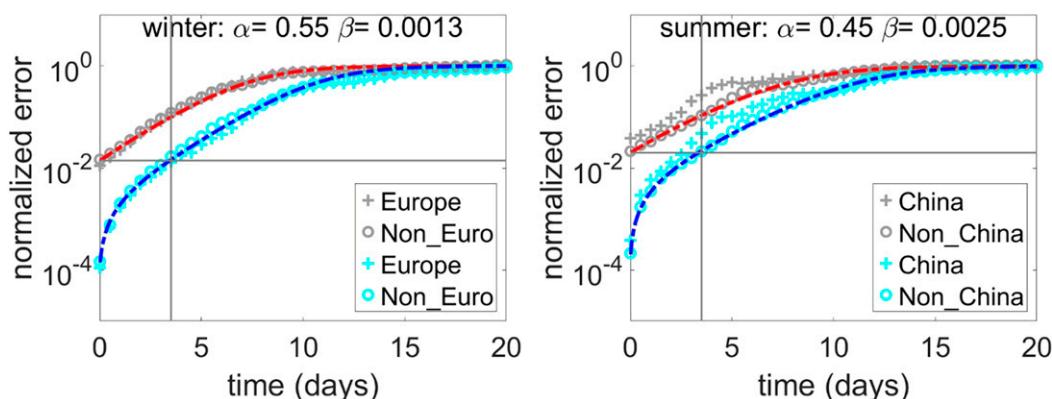
FIG. 11. As in Figs. 3c and 3d, but separating the error statistics in the hazardous-event region vs the rest of the midlatitudes. Red and blue curves are the same fitted curve as in Fig. 3, implying the global averaged results. The plus signs represent the hazardous-event region (60°-longitude width), and the circles represent the nonhazardous-event region (300°-longitude width).

remarkable severe weather events during the same periods examined. Nonetheless, further research is needed to extend the findings to more case studies during all seasons, and preferably with further refined, convection-permitting model resolutions to capture more realistic rapid initial error growth from small-scale moist physics including convection. Although limited sensitivity experiments suggest that the predictability horizon of the day-to-day midlatitude weather is controlled by dynamics and instabilities of the atmosphere and are not particularly dependent on the specific numerical models, it remains possible the estimates may change if future improved models have different perturbation growth's characteristics. While it remains possible that the incorporation of additional unresolved scales and phenomena could actually lead to an increase in the upper bound of predictability (Lorenz 1982), it is generally acknowledged that improved models will resolve more smaller-scale instabilities and thus the error growth is likely to further increase, at least in smaller scales, in which case the current estimate of intrinsic predictability limit may be on the optimistic side.

Also, despite the use of ensembles, the current study focuses on whether a limit of deterministic forecast longer than 2 weeks can be reached if we concentrate on the instantaneous weather that we experience every day. If we define the "forecast skill horizon" as the lead time when ensemble forecasts cease to be, statistically, more skillful than a climatological distribution, the predictability horizon can be longer than 2 weeks for some variables at large synoptic and planetary scales and with longer periods and lower frequencies (Buizza and Leutbecher 2015; Shukla 1998; Palmer 2017). It is beyond the scope of this study to determine the intrinsic limit of the probabilistic prediction. It is also beyond the limit of the current study to determine the predictability horizon for lower-frequency oscillations such as the Madden–Julian oscillation (Madden and Julian 1971; Zhang 2005; Zhang et al. 2017) or the background-mean weather regimes that potentially have predictive skills at the seasonal-to-intraseasonal time scales and beyond. Moreover, even at the convection-permitting resolution, the current sets of ensemble experiments might be insufficient to fully reproduce multiscale tropical waves coupled with moist convection, and thus future studies on the error growth dynamics and predictability limits for tropical systems and their interactions with midlatitude systems (e.g., Ying and Zhang 2017) are warranted.

REFERENCES

Augier, P., and E. Lindborg, 2013: A new formulation of the spectral energy budget of the atmosphere, with application to two high-resolution general circulation models. *J. Atmos. Sci.*, **70**, 2293–2308, https://doi.org/10.1175/JAS-D-12-0281.1.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Bei, N., and F. Zhang, 2007: Mesoscale predictability of the torrential rainfall along the mei-yu front of China. *Quart. J. Roy. Meteor. Soc.*, **133**, 83–99, https://doi.org/10.1002/qj.20.

——, and ——, 2014: Mesoscale predictability of moist baroclinic waves: Variable and scale-dependent error growth. *Adv. Atmos. Sci.*, **31**, 995–1008, https://doi.org/10.1007/s00376-014-3191-7.

Bonavita, M., L. Isaksen, and E. Holm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **138**, 1540–1559, https://doi.org/10.1002/qj.1899.

Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, https://doi.org/10.1002/qj.2619.

Chen, J.-H., and S.-J. Lin, 2011: The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. *Geophys. Res. Lett.*, **38**, L11804, https://doi.org/10.1029/2011GL047629.

——, and ——, 2013: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model. *J. Climate*, **26**, 380–398, https://doi.org/10.1175/JCLI-D-12-00061.1.

——, X. Chen, S.-J. Lin, L. Magnusson, M. Bender, L. Zhou, and S. Rees, 2018: Tropical cyclones in the global 8.5km GFDL fvGFS. *33rd Conf. on Hurricanes and Tropical Meteorology*, Ponte Vedra, FL, Amer. Meteor. Soc., 9B.4, https://ams.confex.com/ams/33HURRICANE/webprogram/Paper339827.html.

Dalcher, A., and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39**, 474–491, https://doi.org/10.3402/tellusa.v39i5.11774.

Daley, R., 1981: Predictability experiments with a baroclinic model. *Atmos.–Ocean*, **19**, 77–89, https://doi.org/10.1080/07055900.1981.9649102.

Durran, D., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.*, **71**, 2476–2488, https://doi.org/10.1175/JAS-D-14-0007.1.

Emanuel, K., and F. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **73**, 3739–3747, https://doi.org/10.1175/JAS-D-16-0100.1.

Froude, L. S. R., L. Bengtsson, and K. I. Hodges, 2013: Atmospheric predictability revisited. *Tellus*, **65**, 19022, https://doi.org/10.3402/tellusa.v65i0.19022.

Hazelton, A. T., L. M. Harris, and S.-J. Lin, 2018: Evaluation of tropical cyclone structure forecasts in a high-resolution version of the multiscale GFDL fvGFS model. *Wea. Forecasting*, **33**, 419–442, https://doi.org/10.1175/WAF-D-17-0140.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived

greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

Judt, F., 2018: Insights into atmospheric predictability through global convection-permitting model simulations. *J. Atmos. Sci.*, **75**, 1477–1497, https://doi.org/10.1175/JAS-D-17-0343.1.

Katz, R. W., and A. H. Murphy, 2015: *Economic Value of Weather and Climate Forecasts*. Cambridge University Press, 240 pp.

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.*, **28**, 145–161, https://doi.org/10.1175/1520-0469(1971)028<0145:APATDT>2.0.CO;2.

——, and R. H. Kraichnan, 1972: Predictability of turbulent flows. *J. Atmos. Sci.*, **29**, 1041–1058, https://doi.org/10.1175/1520-0469(1972)029<1041:POTF>2.0.CO;2.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

——, 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

——, 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513, https://doi.org/10.3402/tellusa.v34i6.10836.

——, 1996: Predictability—A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–18.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50-day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.

Magnusson, L., 2017: Diagnostic methods for understanding the origin of forecast errors. *Quart. J. Roy. Meteor. Soc.*, **143**, 2129–2142, https://doi.org/10.1002/qj.3072.

——, and E. Källén, 2013: Factors influencing skill improvements in the ECMWF forecasting system. *Mon. Wea. Rev.*, **141**, 3142–3153, https://doi.org/10.1175/MWR-D-12-00318.1.

Mapes, B., S. Tulich, T. Nasuno, and M. Satoh, 2008: Predictability aspects of global aqua-planet simulations with explicit convection. *J. Meteor. Soc. Japan*, **86A**, 175–185, https://doi.org/10.2151/jmsj.86A.175.

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, https://doi.org/10.1175/JAS-D-11-0315.1.

Morss, R. E., C. Snyder, and R. Rotunno, 2009: Spectra, spatial scales, and predictability in a quasigeostrophic model. *J. Atmos. Sci.*, **66**, 3115–3130, https://doi.org/10.1175/2009JAS3057.1.

Ngan, K., P. Bartello, and D. N. Straub, 2009: Predictability of rotating stratified turbulence. *J. Atmos. Sci.*, **66**, 1384–1400, https://doi.org/10.1175/2008JAS2799.1.

Palmer, T., 2017: The primacy of doubt: Evolution of numerical weather prediction from determinism to probability. *J. Adv. Model. Earth Syst.*, **9**, 730–734, https://doi.org/10.1002/2017MS000999.

——, A. Döring, and G. Seregin, 2014: The real butterfly effect. *Nonlinearity*, **27**, R123–R141, https://doi.org/10.1088/0951-7715/27/9/R123.

Putman, W. M., and M. Suarez, 2011: Cloud-system resolving simulations with the NASA Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophys. Res. Lett.*, **38**, L16809, https://doi.org/10.1029/2011GL048438.

Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305, https://doi.org/10.1175/1520-0493(1994)122<1281:REGING>2.0.CO;2.

Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405, https://doi.org/10.1175/BAMS-D-12-00099.1.

Rosinski, J., and D. L. Williamson, 1997: The accumulation of rounding errors and port validation for global atmospheric models. *SIAM J. Sci. Comput.*, **18**, 552–564, https://doi.org/10.1137/S1064827594275534.

Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1076, https://doi.org/10.1175/2007JAS2449.1.

Selz, T., and G. C. Craig, 2015: Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Mon. Wea. Rev.*, **143**, 813–827, https://doi.org/10.1175/MWR-D-14-00140.1.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, **282**, 728–731, https://doi.org/10.1126/science.282.5389.728.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647–677, https://doi.org/10.1256/003590002321042135.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, https://doi.org/10.1175/MWR2830.1.

——, S.-H. Park, J. B. Klemp, and C. Snyder, 2014: Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. *J. Atmos. Sci.*, **71**, 4369–4381, https://doi.org/10.1175/JAS-D-14-0114.1.

Sun, Y., and F. Zhang, 2016: Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *J. Atmos. Sci.*, **73**, 1419–1438, https://doi.org/10.1175/JAS-D-15-0142.1.

Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus*, **9**, 275–295, https://doi.org/10.1111/j.2153-3490.1957.tb01885.x.

Tribbia, J. J., and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: An updated perspective. *Mon. Wea. Rev.*, **132**, 703–713, https://doi.org/10.1175/1520-0493(2004)132<0703:SIAAPA>2.0.CO;2.

Voosen, P., 2017: The weather master. *Science*, **356**, 128–131, https://doi.org/10.1126/science.356.6334.128.

Weyn, J. A., and D. R. Durran, 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci.*, **74**, 2191–2210, https://doi.org/10.1175/JAS-D-17-0006.1.

World Meteorological Organization, 2015: Seamless prediction of the Earth system: From minutes to months. WMO Rep. 1156, 483 pp.

Ying, Y., and F. Zhang, 2017: Practical and intrinsic predictability of multi-scale weather and convectively coupled equatorial waves during the active phase of an MJO. *J. Atmos. Sci.*, **74**, 3771–3785, https://doi.org/10.1175/JAS-D-17-0157.1.

Zhang, C., 2005: Madden–Julian oscillation. *Rev. Geophys.*, **43**, RG2003, https://doi.org/10.1029/2004RG000158.

Zhang, F., and Y. Weng, 2015: Predicting hurricane intensity and associated hazards: A five-year real-time forecast experiment with assimilation of airborne Doppler radar observations. *Bull. Amer. Meteor. Soc.*, **96**, 25–32, https://doi.org/10.1175/BAMS-D-13-00231.1.

——, C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the "surprise" snowstorm of 24-25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632, https://doi.org/10.1175/1520-0493(2002) 130<1617:MPOTSS>2.0.CO;2.

——, ———, and ——, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185, https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM> 2.0.CO;2.

——, N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.*, **64**, 3579–3594, https://doi.org/10.1175/ JAS4028.1.

——, Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, https://doi.org/ 10.1175/2009MWR2645.1.

——, S. Talaphdar, and S. Wang, 2017: The role of global circumnavigating mode in the MJO initiation and propagation. *J. Geophys. Res. Atmos.*, **122**, 5837–5856, https://doi.org/ 10.1002/2016JD025665.

Zhou, L., S.-J. Lin, J.-H. Chen, L. M. Harris, X. Chen, and S. Rees, 2019: Toward convective-scale prediction within the Next-Generation Global Prediction System. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-17-0246.1, in press.