What is the Predictability Limit of Midlatitude Weather?

Fuqing Zhang and Y. Qiang Sun

[1]Department of Meteorology and Atmospheric Science, and [2]Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania.

Linus Magnusson and Roberto Buizza

European Center for Medium Range Weather Forecasting, Reading, United Kingdom.

Shian-Jiann Lin and Jan-Huey Chen

Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, New Jersey.

Kerry Emanuel

Lorenz Center, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Journal of the Atmospheric Science
Accepted, December 2018

*Correspondence to: fzhang@psu.edu; smilesyq@gmail.com.

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 Center for Medium Range Weather Forecasting (ECMWF, 9-km operational model) and identical-twin experiments of the US next-generation global prediction system (NGGPS, 3-km). Results suggest that predictability limit for mid-latitude 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 Center for Medium Range Weather Prediction (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). 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

65 Atlantic Basin in 2016 is smaller than the 2-day lead-time forecast error in 1990, which
66 may have saved billions of dollars (Katz et al. 2015).

67 Yet, improvement of NWP has limits. From the perspective of predictability, this
68 concept of "atmospheric predictability limit" can be grossly categorized into intrinsic
69 versus practical predictability (Lorenz 1996; Melhauser and Zhang 2012). As discussed in
70 Ying and Zhang (2017), intrinsic predictability refers to "the ability to predict given nearly
71 perfect representation of the dynamical system (by a forecast model) and nearly perfect
72 initial/boundary conditions, an inherent limit due to the chaotic nature of the atmosphere
73 and cannot be extended by any means" (Lorenz 1963, 1969; Zhang et al. 2003, 2007; Sun
74 and Zhang 2016). Practical predictability, also commonly referred to as our weather
75 prediction skill, is "the ability to predict given realistic uncertainties in both the forecast
76 model and initial and boundary conditions" (Lorenz 1982, 1996; Zhang et al. 2002, 2006).
77 This practical predictability can be extended through reduction in key limiting factors of
78 the forecast errors, including initial-condition errors, boundary condition errors, and model
79 errors. All these factors, especially the initial condition errors, have been greatly and could
80 be further reduced with better NWP models ingesting high-accuracy observations using
81 advanced data assimilation approaches along with advanced computing power (e.g., Zhang
82 et al. 2009; Zhang and Weng 2015; Emanuel and Zhang 2016). Nevertheless, given our
83 desire for better weather forecasting at all temporal and spatial scales, it is natural to ask
84 whether an *intrinsic predictability* of the mid-latitude weather exists. If yes, what is this
85 inherent limit given nearly perfect NWP models with nearly perfect initial conditions? This
86 is a crucial question that meteorologists have sought to answer ever since the beginning of
87 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 to 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 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 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 shallower -5/3 kinetic energy spectrum, which is not well simulated in most of 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

134 global 7-km aqua-planet identical-twin simulations, with a focus in the tropics. Judt (2018)
135 studied the atmospheric predictability through a pair of convection-permitting identical-
136 twin simulations with the newly developed global model for prediction across scales
137 (MPAS, Skamarock et al. 2014). Building on the findings of previous theoretical and
138 modeling studies, our work here seeks to estimate the intrinsic limit of day-to-day weather
139 predictability using ensemble simulations with the most advanced global NWP models at
140 both ECMWF and U.S. NOAA. Our particular emphasis will be synoptic-scale weather
141 systems dominated by baroclinic instability in the midlatitudes, where most of the world
142 population resides. In particular, we showcase the practical versus intrinsic predictability
143 limits of the global midlatitude weather during two periods in boreal winter and summer,
144 respectively. These periods also endured two recent hazardous regional weather events: a
145 wintertime cold surge event affecting northern Europe in early January 2016, and a
146 summertime rainfall-flooding event in China during July 2016. The choice of these two
147 events are rather subjective and somewhat random with the intent to represent the typical
148 midlatitude predictability while in the meantime covering some notable weather events in
149 recent years. Nevertheless, neither of these two cases fall into the "forecast-bust" cases
150 using the criteria identified by Rodwell et al. (2013). Moreover, to the best of our
151 knowledge, there were no severe weather outbreaks during these two periods in the
152 midlatitude atmosphere of the Southern Hemisphere whose predictability will be
153 simultaneously examined and compared with the Northern Hemispheric midlatitudes that
154 have notable weather events.

155 Section 2 of this paper introduces the experiment design of our work, including the
156 model we used and the perturbations added for each ensemble. Analysis for the ensemble

157 spread from different perspectives are given in section 3, physical interpretation based on
158 the results and hence the estimated predictability limit is also provided. Discussions on the
159 limitations of current work and concluding remarks are presented in section 4.

160

161 **2. Experimental Design**

162 This study adopted established methodologies introduced in the introduction part
163 for studying atmospheric predictability using perfect-model, identical-twin experiments
164 where the ensemble members with minute initial condition differences are explored. A
165 series of ensemble simulations with the state-of-the-science global NWP model at ECMWF
166 (namely the Integrated Forecast System or IFS), and US next-generation global prediction
167 system (NGGPS) with finite volume (FV3) dynamical core are designed to address the
168 following two key questions: (1) what is the intrinsic predictability limit of multiscale
169 midlatitude weather assuming a perfect model with nearly perfect initial conditions? (2)
170 How much longer can the practical predictability be increased by reducing initial condition
171 uncertainties to different degree of accuracy?

172 *2.1 Model details*

173 *2.1.a ECMWF / IFS model*

174 The IFS control and ensemble forecasts presented herein uses the latest upgrade
175 (cycle 41r2) of ECMWF, the highest-resolution (~9 km) ever global operational NWP
176 model. More details of this model upgrade can be found on the official website of ECMWF
177 (http://www.ecmwf.int/). Different from previous versions, this new ECMWF IFS model
178 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, 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 450s. 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.

2.1.b US fvGFS system

The newly developed Geophysical Fluid Dynamics Laboratory (GFDL) fvGFS modeling system (Zhou et al. 2018; 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 next-generation global prediction system (NGGPS) phase II, using the nonhydrostatic Finite-Volume Cubed-Sphere Dynamical core (FV3) coupled to physical parameterizations from the National Center for Environmental Prediction's Global Forecast System (NCEP/GFS). The GFDL FV3 was recently chosen as the dynamical core

for the US NGGPS as detailed in the 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. 2018). No cumulus scheme is adopted.

2.b Ensemble experiments: EDA and EDA0.1

We first perform two types of ensemble experiments (denoted as EDA and EDA0.1 hereafter) with the current operational 9-km IFS model, running 10-member ensembles for 20 days beginning at 6 different times (three consecutive days of 24-26 December 2015, and three consecutive days of 24-26 June 2016), respectively. 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 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 (TCo639, ~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 US 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 under-predicted the surface temperature over most of contiguous US and over-predicted 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 U.S. (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).

3.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, 25 and 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

294 ensemble of 4DVar analyses, while EDA0.1 ensemble sets are initialized with nearly
295 perfect initial conditions (initial kinetic energy error is 1% of that in EDA). As shown in
296 Fig. 3a (normalized results shown in Fig. 3c), the spread of the EDA ensembles with
297 realistic initial condition uncertainties grows nearly two orders of magnitude larger before
298 saturating at approximately 10–12 days, while the spread of the EDA0.1 ensembles, with
299 minute initial perturbations (i.e., nearly perfect initial conditions), grows nearly four orders
300 of magnitude larger before saturating at the same level as the EDA ensemble around 14–
301 15 days (as a strong indication of the intrinsic predictability limit).

302 Similar quantitative statistics, representing intrinsic versus practical predictability
303 limits assuming perfect model, can also be inferred from the same pairs of ensembles for
304 the Southern Hemisphere (Fig. 4), as well as from pairs of Northern Hemisphere
305 midlatitude 20-day 10-member global ensemble (Fig. 3b and Fig. 3d) initialized from three
306 consecutive summer days (24, 25 and 26 June) in 2016. During the 20-day simulation
307 period in June, vast areas of the Yangtze River Basin of China observed historical flooding
308 (NASA:https://earthobservatory.nasa.gov/NaturalHazards/view.php?id=88467). Moreover,
309 calculation of the anomaly correlation coefficients (ACC) between the ensemble forecast
310 and the observations over the global midlatitudes also gives quantitatively similar estimates
311 for both the practical and intrinsic predictability limits (Fig. 5), with correlation dropping
312 to 60% at around 10 days for the EDA ensemble and 13-15 days for EDA0.1 ensemble.

313 The growth of the ensemble variance, representative of the forecast error, fits
314 surprisingly well with the simple error growth model that was originally proposed in
315 Lorenz (1982), modified later (Dalcher and Kalnay 1987; Reynolds et al. 1994) and here
316 as:

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

Here ε(t) is the normalized error where ε~1 means it reaches the maximum or becomes saturated, 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-d and 4c-d show the evolution of normalized error averaged for both the winter and summer cases, respectively, as well as the fitted error growth curves from equation (1).

Fig. 3 and Fig. 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 Fig.3 and Fig. 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 "super-exponential" 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 equation (1). For example, the green line in Figs. 3c-d shows the predicted error growth curve derived from equation (1) when the initial condition error is reduced to $1.0e^{-10}$. It is nearly identical to the blue line

339 which means there is not much more room for improvement. In other words, if equation
340 (1) holds, further reduction in the initial-condition or model error would not help extend
341 our forecast lead time much longer (maybe only in hours or even minutes).

342 The errors in EDA0.1 grow to an amplitude similar to the EDA initial ensemble
343 spread in 3-4 days. Subsequent error growth and saturation in the EDA0.1 ensembles
344 mimic those of the EDA ensembles except for a 3-to-4-day delay in forecast lead times.
345 The overall reference error kinetic energy saturates ($\varepsilon \sim 1$) at around 10-12 days for all the
346 EDA ensembles and 14-15 days for all the EDA0.1 ensembles. This remains true for
347 different initialization times and for both the winter and summer days of the Northern
348 Hemisphere and the Southern Hemisphere.

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

358 *3.b Spectral analysis*

359 While this 3-5 day serve as the estimated potential for extended weather forecast
360 lead time, the atmospheric predictability limit is also scale-dependent. For example, small
361 scale thunderstorms is much less predictable than the synoptic system in which they are

362 embedded. Therefore, it is important to examine the scale dependence of predictability
363 limit. Spectral decomposition of perturbation kinetic energy across all zonal wavenumbers
364 averaged over the midlatitudes (40-60°N) for both winter and summer periods are
365 displayed in Fig. 6. The corresponding spectra for the Southern Hemisphere midlatitudes
366 (40-60°S) are shown in Fig. 7.

367 The kinetic energy spectrum here is calculated as in (Skamarock 2004). We have
368 chosen to compute the one-dimensional (1D) spectrum of the velocity fields along zonal
369 direction . The advantage of this 1D spectrum is that we could fully utilize the periodicity
370 of the global model in the zonal direction while focus on the midlatitude only. Let $u_{i,j,n}$
371 and $v_{i,j,n}$ denote the zonal and meridional velocity components for the nth ensemble
372 member, subtracting the ensemble mean fields first if we are calculating the kinetic energy
373 spectra for the perturbations. For the spectra in Fig. 8, the differences between the perturbed
374 run and the un-perturbed run are used. The Fourier transforms of the velocity components
375 $\hat{u}_{j,n}(k)$ and $\hat{v}_{j,n}(k)$ are then computed along the zonal direction for each ensemble member
376 and all the meridional j indices. Then the kinetic energy spectra density can be written as:

377 $$E_{j,n}(k) = \frac{\Delta x}{2N_x}\left[\hat{u}_{j,n}(k)\hat{u}^*_{j,n}(k) + \hat{v}_{j,n}(k)\hat{v}^*_{j,n}(k)\right] \quad (2)$$

378 where N_x is the number of grid points along the zonal direction of the model. The asterisk
379 denotes the complex conjugate. We can then average $E_{j,n}(k)$ over j and n to get the kinetic
380 energy spectrum for the full ensemble and the latitude band of interest (40-60°N for the
381 mid-latitude; the results are not very sensitive to this choice, 30-60°N average give very
382 similar plots). When the spectrum of the perturbation kinetic energy (amplitude of "noise")
383 at a given wavelength reaches the reference background spectral kinetic energy (signal to

384 be predicted), it is saturated, after which no single deterministic forecast will have any
385 predictive skill.

386 Consistent with Fig. 3, Figure 6 also shows that it takes slightly more than 3 days
387 for the perturbation kinetic energy in the reduced-perturbation ensemble (EDA0.1) to grow
388 two orders of magnitude across all resolvable wavelengths to a level comparable with the
389 realistic analysis uncertainty represented by the EDA for the periods of the winter and
390 summer events, respectively. Also, Fig. 6 shows that the perturbation spectral kinetic
391 energy from the EDA0.1 ensemble saturates at the amplitude of the reference kinetic
392 energy across all synoptic scales by 15 days, again consistent with the overall intrinsic
393 predictability limit estimated from Fig. 3.

394 Moreover, saturation time for different scales is different. With reduced initial-
395 condition uncertainties, as in EDA0.1, forecast error first saturates at smaller scales, then
396 subsequently grows rapidly in magnitude and in scale, consistent with past regional
397 modeling studies (Zhang et al. 2007; Selz and Craig 2015; Sun and Zhang 2016). A simple
398 estimation from Fig. 6 shows that the forecast error saturation time (and thus intrinsic limit
399 of predictability) is less than 3 days for horizontal scales less than 200 km, less than 5 days
400 for horizontal scales less than 400 km, and less than 10 days for horizontal scales less than
401 1000 km. Also worth noting, Fig. 6 illustrates the synoptic-scale predictability in terms of
402 500-hPa horizontal winds. Much more limited predictability is expected for vertical
403 velocity and instantaneous precipitation rate forecasts (Bei and Zhang 2007, 2014), which
404 possesses very different reference energy spectra.

405

406 *3.c. Sensitivity study using US NGGPS model*

Although we could only afford to perform one pair of 10-day forecasts using the US NGGPS model based on the FV3 dynamical core with 3-km convection-permitting 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 equation (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-d, 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).

3.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 as to what dynamic processes control the error growth and eventually limits 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.

Figure 9 shows the evolution of the ensemble spread of 500-hpa meridional winds for the first three days in EDA0.1 integrated from 0000 UTC 26 Dec 2015, especially focus on a developing extratropical cyclone over the west coast of the United States. The blue contour is the 500-hpa geopotential height and grey contour is the regions with 12 hour 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 mid-latitude bands, although with a maximum in the synoptic storms. The key idea that minute perturbations will first generate error in small scale moist convective systems and the errors then grow upscale is consistent

453 with previous studies (Zhang et al. 2007; Sun and Zhang 2016). We could also take a first
454 look at the propagation of the ensemble spread using the Hovmöller diagram, which is
455 plotted in Fig. 10 for the EDA0.1 ensembles initialized at 0000 UTC 24 December 2015.
456 As marked subjectively by different type of arrows, there exist at least three characteristic
457 pathways for the error to evolve and grow through time globally. The dotted ones, mainly
458 shown up in the first few days, have an approximate eastward speed of 10-15 m/s and are
459 consistent with the phase speed of individual synoptic weather systems. The double dashed
460 arrows that have an approximate eastward speed of 25-30 m/s likely following the
461 downstream energy propagation of different baroclinic wave packets, and the bold solid
462 arrows with a slow westward progression signaling the error enhancement near the quasi-
463 stationary planetary-scale low-pressure centers. These error growth pathways expand
464 beyond the multi-stage error growth mechanisms identified in previous regional-scale
465 predictability study (Zhang et. al, 2007) and will be examined in more details in our future
466 study.

467

468 **4. Concluding Remarks**

469 The promising finding of the current study is that, assuming the current-generation
470 state-of-the-science NWP models could capture the most essential physical processes in
471 the real world, we can further improve the forecast accuracy of day-to-day weather events
472 such as the ones we discussed, by up to 5 days, if we reduce the initial condition
473 uncertainties by a factor of 10. In particular, we examined the predictability of weather
474 forecasts in two showcase studies, and we have looked at the multiscale midlatitude error
475 evolution across different spatial scales. Our study suggests that we are currently still quite

476 far from the ultimate limit of predictability, and it is apparent that we have ample room for
477 further improvement in the day-to-day weather predictability likely for decades ahead.

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

487 It is possible those two individual local weather events (cold surge and floods) have
488 slightly different predictability than more typical weather patterns but our findings here are
489 based on mean error growth statistics of the global midlatitudes averaged over many
490 wavelengths and multiple initialization times, not just the regions of the localized
491 hazardous events. Additional calculations that exclude these two local events show
492 consistent results (Fig. 11); so are the calculations in the opposing hemispheric midlatitudes
493 that have no remarkable severe weather events during the same periods examined.
494 Nonetheless, further research is needed to extend the findings to more case studies during
495 all seasons, and preferably with further refined, convection-permitting model resolutions
496 to capture more realistic rapid initial error growth from small-scale moist physics including
497 convection. Although limited sensitivity experiments suggest that the predictability
498 horizon of the day-to-day midlatitude weather is controlled by dynamics and instabilities

499 of the atmosphere and are not particularly dependent on the specific numerical models, it
500 remains possible the estimates may change if future improved models have different
501 perturbation-growth's characteristics. While it remains possible that the incorporation of
502 additional unresolved scales and phenomena could actually lead to an increase in the upper
503 bound of predictability (Lorenz 1982), it is generally acknowledged that improved models
504 will resolve more smaller scale instabilities and thus the error growth is likely to further
505 increase, at least in smaller scales, in which case the current estimate of intrinsic
506 predictability limit may be on the optimistic side.

507 Also, despite the use of ensembles, the current study focuses on whether a limit of
508 deterministic forecast longer than two weeks can be reached if we concentrate on the
509 instantaneous weather that we experience every day. If we define the 'forecast skill
510 horizon' as the lead time when ensemble forecasts cease to be, statistically, more skillful
511 than a climatological distribution, the predictability horizon can be longer than two weeks
512 for some variables at large synoptic and planetary scales and with longer periods and lower
513 frequencies (Buizza and Leutbecher 2015; Shukla 1998; Palmer 2017). It is beyond the
514 scope of this study to determine what is the intrinsic limit of the probabilistic prediction. It
515 is also beyond the limit of the current study what is predictability horizon for lower-
516 frequency oscillations such as the Madden-Julian Oscillations (Madden and Julian 1971;
517 Zhang 2005; Zhang et al. 2017) or the background mean weather regimes that potentially
518 have predictive skills at the season-to-intraseasonal timescales and beyond. Moreover,
519 even at the convection-permitting resolution, the current sets of ensemble experiments
520 might be insufficient to fully reproduce multiscale tropical waves coupled with moist
521 convection, and thus future studies on the error growth dynamics and predictability limits

for tropical systems and their interactions with midlatitudes systems (e.g., Ying and Zhang 2017) are warranted.

Acknowledgments: This study is partially supported by NSF Grants AGS-1305798 and 1712290, and ONR Grant N000140910526. Computing is performed at ECMWF, NOAA ESRL jet cluster, and the Texas Advanced Computer Center. Review comments from the editor and three anonymous reviewers, and discussions with Richard Rotunno, Dale Durran, Robert Nystrom, James Doyle and many other researchers on related subjects are beneficial**.**

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

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

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.

Bei, N. and F. Zhang, 2014: Mesoscale predictability of moist baroclinic waves: Variable and scale-dependent error growth. *Adv. Atmos. Sci.*, **31**, 995–1008.

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

544 Buizza, R. and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy.*
545 *Meteor. Soc.*, **141**, 3366–3382.

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

549 Chen, J.-H., and S.-J. Lin, 2013: Seasonal predictions of tropical cyclones using a 25-km-
550 resolution general circulation model. *J. Climate*, **26**, 380–398.

551 Chen J.-H., X. Chen, S.-J. Lin, L. Magnusson, M. Bender, L. Zhou, and S. Rees, 2018:
552 Tropical Cyclones in GFDL fvGFS – Impacts of Dycore, physics and initial
553 conditions. 33rd AMS Conference on Hurricanes and Tropical Meteorology, Ponte
554 Vedra, FL. Amer. Meteor. Soc., 9B.4,
555 https://ams.confex.com/ams/33HURRICANE/webprogram/Paper339827.html.

556 Daley, R., 1981: Predictability experiments with a baroclinic model. *Atmos.–Ocean*, **19,**
557 77–89.

558 Dalcher, A. and E. Kalnay, 1987: Error growth and predictability in operational ECMWF
559 forecasts. *Tellus*, **39**, 474-491.

560 Durran, D., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of
561 practical importance. *J. Atmos. Sci.*, **71**, 2476–2488.

562 Emanuel, K. and F. Zhang, 2016: On the Predictability and Error Sources of Tropical
563 Cyclone Intensity Forecasts, *J. Atmos. Sci.*, **73**, 3739-3747.

564 Froude, L. S. R., L. Bengtsson, and K. I. Hodges, 2013: Atmospheric predictability
565 revisited. *Tellus*, **65**, 19022.

566 Hazelton, A. T., L. M. Harris, and S.-J. Lin, April 2018: Evaluation of Tropical Cyclone
567 Structure Forecasts in a High-Resolution Version of the Multiscale GFDL fvGFS
568 Model. *Wea. Forecasting*, **33**, 419–442.

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

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

574 Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D.
575 Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with
576 the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103

577 Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos.*
578 *Sci.,* **28**, 145–161.

579 Leith, C. E., and R. H. Kraichnan, 1972: Predictability of turbulent flows. *J. Atmos. Sci.,*
580 **29**, 1041–1058.

581 Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field
582 in a cloud model. *J. Climate Appl. Meteor.*, **22**, 1065–1092

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

585 Lorenz, E. N., 1982: Atmospheric Predictability experiments with a large numerical
586 model. *Tellus,* **34**, 505- 513.

587 Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50-day oscillation in the zonal
588 wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708.

589 Magnusson, L. and E. Källén, 2013: Factors Influencing Skill Improvements in the
590 ECMWF Forecasting System. *Mon. Wea. Rev.*, **141**, 3142-3153.

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

593 Mapes, B., S. Tulich, T. Nasuno, and M. Satoh, 2008: Predictability aspects of global
594 aqua-planet simulations with explicit convection. *J. Meteor. Soc. Japan*, **86**, 175–
595 185.

596 Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of Severe and
597 Convective Weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350-3371.

598 Morss, R. E., C. Snyder, and R. Rotunno, 2009: Spectra, spatial scales, and predictability
599 in a quasigeostrophic model. *J. Atmos. Sci.*, **66**, 3115–3130

600 Ngan, K., P. Bartello, and D. N. Straub, 2009: Predictability of rotating stratified
601 turbulence. *J. Atmos. Sci.*, **66**, 1384–1400

602 Palmer, T. N., A. Döring, and G. Seregin, 2014: The real butterfly effect. *Nonlinearity*, **27**,
603 R123–R141,

604 Palmer, T., 2017: The primacy of doubt: Evolution of numerical weather prediction from
605 determinism to probability. *J. Adv. Model. Earth Syst.*, **9**, 730-734.

606 Putman, W. M., and M. Suarez, 2011: Cloud-system resolving simulations with the NASA
607 Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophys.*
608 *Res. Lett.*, **38**, L16809

609 Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's
610 global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305.

611

612 Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range
613 weather forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405

614 Rosinski, J., and D. L. Williamson, 1997: The accumulation of rounding errors and port
615 validation for global atmospheric models. *Siam Journal on Scientific Computing*, **18**,
616 552-564, doi:10.1137/S1064827594275534.

617 Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the
618 predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1076.

619 Selz, T. and G. C. Craig, 2015: Upscale error growth in a high-resolution simulation of a
620 summertime weather event over Europe. *Mon. Wea. Rev.*, **143**, 813–827.

621 Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate
622 forecasting. *Science,* **282,** 728–731.

623 Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of
624 numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647–677

625 Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy
626 spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.

627 Sun, Y. and F. Zhang, 2016: Intrinsic versus practical limits of atmospheric predictability
628 and the significance of the butterfly effect. *J. Atmos. Sci.*, **73**, 1419–1438.

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

631 Tribbia, J. J., and D. P. Baumhefner, 2004: Scale interactions and atmospheric
632 predictability: An updated perspective. *Mon. Wea. Rev*., **132**, 703–713.

633 Voosen, P., 2017: The weather master. *Science*, **356**, 128-131.

634 Weyn, J. A., and D. R. Durran, 2017: The dependence of the predictability of mesoscale
635 convective systems on the horizontal scale and amplitude of initial errors in idealized
636 simulations. *J. Atmos. Sci.*, **74**, 2191–2210

637 World Meteorological Organization, 2015: *Seamless Prediction of the Earth System: From*
638 *Minutes to Months*. WMO No. 1156, 483pp.

639 Ying, Y., F. Zhang, 2017: Practical and intrinsic predictability of multi-scale weather and
640 convectively-coupled equatorial waves during the active phase of an MJO. *J. Atmos.*
641 *Sci*., **74**, 3771-3785.

642 Zhang, C., 2005: Madden–Julian oscillation. *Rev. Geophys.*, **43**, RG2003.

643 Zhang, F. and Y. Weng, 2015: Predicting Hurricane Intensity and Associated Hazards: A
644 Five-Year Real-Time Forecast Experiment with Assimilation of Airborne Doppler
645 Radar Observations. *B. Amer. Meteor. Soc.*, **96**, 25-32.

646 Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the "surprise"
647 snowstorm of 24-25 January 2000. *Mon. Wea. Rev.*, **130**, 1617-1632.

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

650 Zhang F., N. Bei, R. Rotunno, C. Snyder, C. C. Epifanio, 2007: Mesoscale predictability
651 of moist baroclinic waves: Convection-permitting experiments and multistage error
652 growth dynamics. *J. Atmos. Sci*. **64**, 3579–3594.

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

657 Zhang, F., S. Talaphdar, and S. Wang, 2017: The role of global circumnavigating mode in
658 the MJO initiation and propagation. *J. Geophys. Res.-Atmos.,* **122**, 5837-5856.

659 Zhao, Q., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models.
660 *Mon. Wea. Rev.*, **125**, 1931–1953.

661 Zhou, L., S.-J. Lin, J.-H. Chen, L. M. Harris, X. Chen, and S. Rees, 2018: Enabling and
662 Improving Convective-Scale Predictions using the Variable-Resolution FV3-based
663 Next Generation Global Prediction System, BAMS (in reviewing).

664

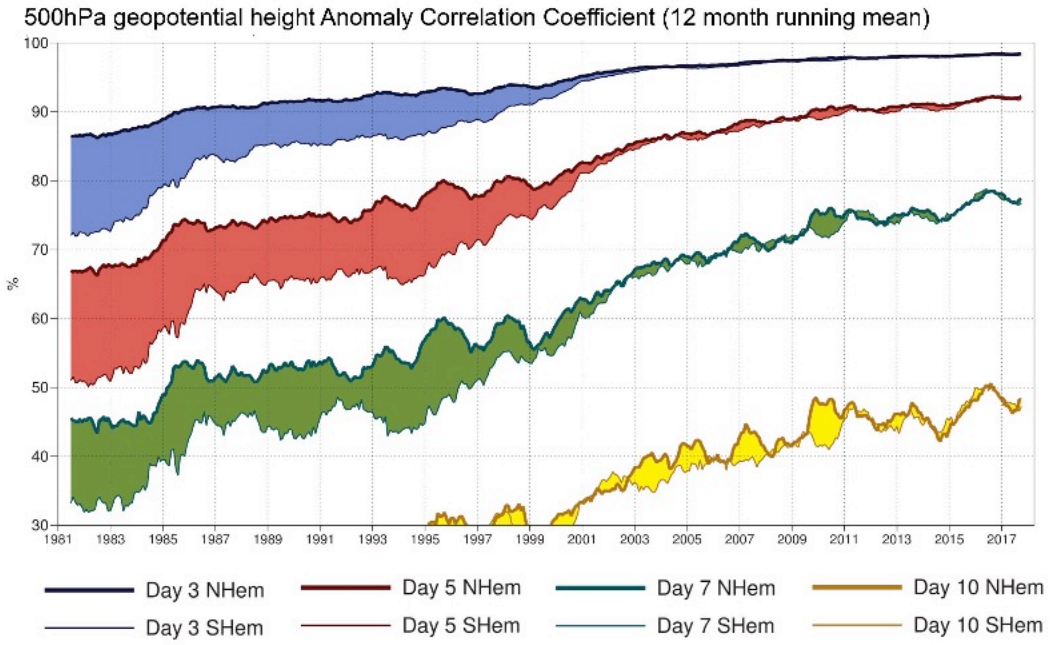

Fig. 1. Annual evolution the ECMWF NWP deterministic control forecast performance in terms of anomaly correlation of 500-hPa height predictions. Shading in this plot indicates the different forecast skill between NH and SH, which has almost gone in recent years. This plot is directly adapted from ECMWF official website

(https://www.ecmwf.int/en/forecasts/charts/catalogue/plwww_m_hr_ccaf_adrian_ts).

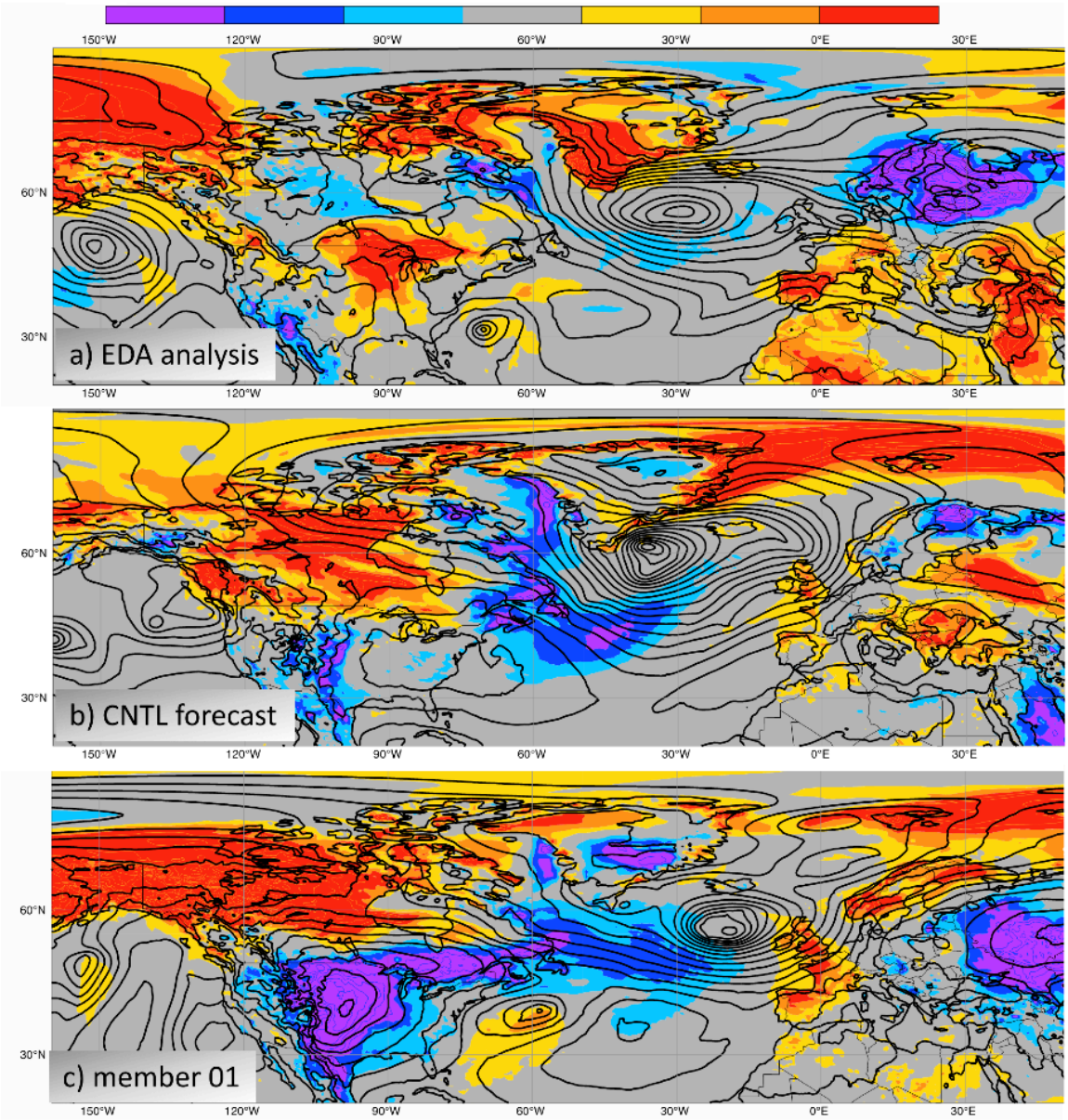

Fig. 2. Sample 15-day numerical weather predictions versus observations. The 2-metre surface temperature anomalies (colored) and the sea-level pressure (contoured every 4 hPa) and valid at 0000 UTC 8 January 2016 for (a) the ECMWF observational analysis, (b) the 9-km IFS 15-day-lead-time control forecast initialized at 0000 UTC 24 December 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-year, 10-member ensemble reforecast dataset using the same model with a 15-day lead time.

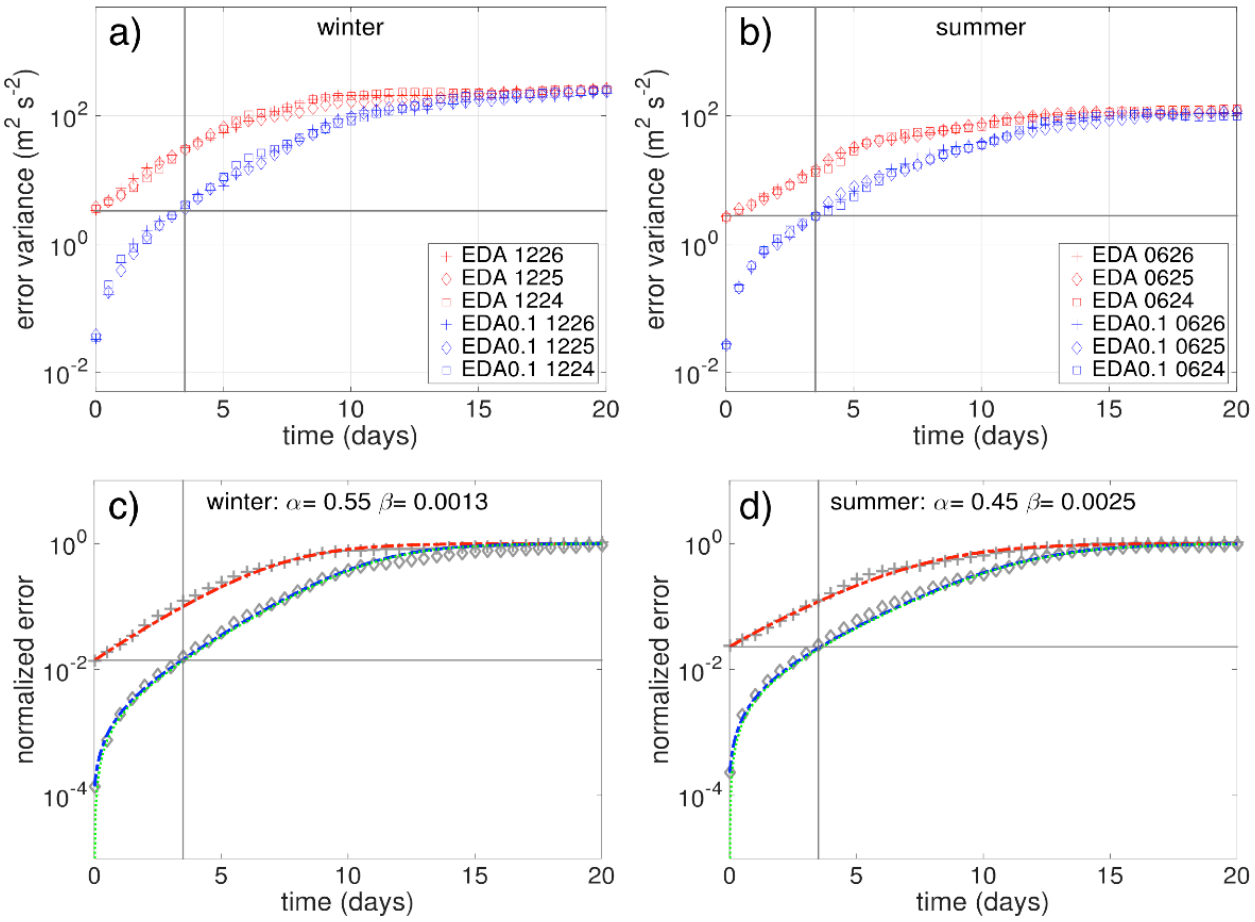

Fig. 3 Evolution of forecast uncertainty growth in terms of ensemble variance of 500-hPa winds energy averaged over NH midlatitude (40-60°N) for 3 ensemble simulations with current-day realistic initial condition uncertainties (red symbols) and the corresponding ensembles with minute (1%) initial condition errors (blue symbols) initialized at: a) three consecutive days for the winter (24-26 December 2015; top) and b) three consecutive days in the summer (24-26 June 2016). c) and d) show the evolution of normalized error variances (grey symbols) averaged for the winter and summer cases, respectively, as well as the fitted error growth curves (red for EDA; blue for EDA0.1) from equation (1). The units for α and β are day^{-1}. Green lines in c) and d) show the error growth curve according to fitted equation 1 if the initial condition error is further reduced to 1e-10. Note the log scale of the perturbation energy.

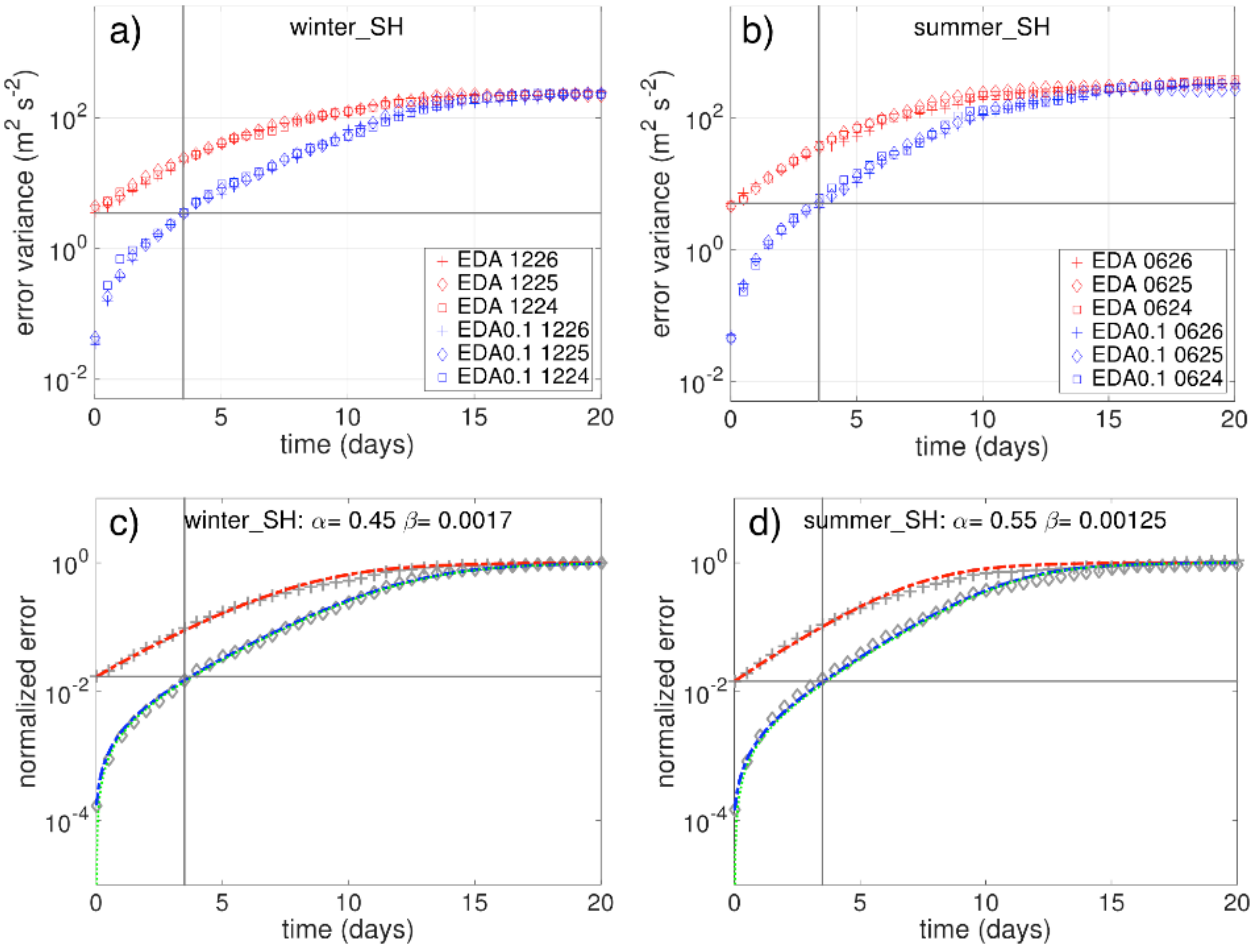

Fig. 4. Similar as in Fig. 3 except for the Southern Hemisphere (SH, 40-60S).

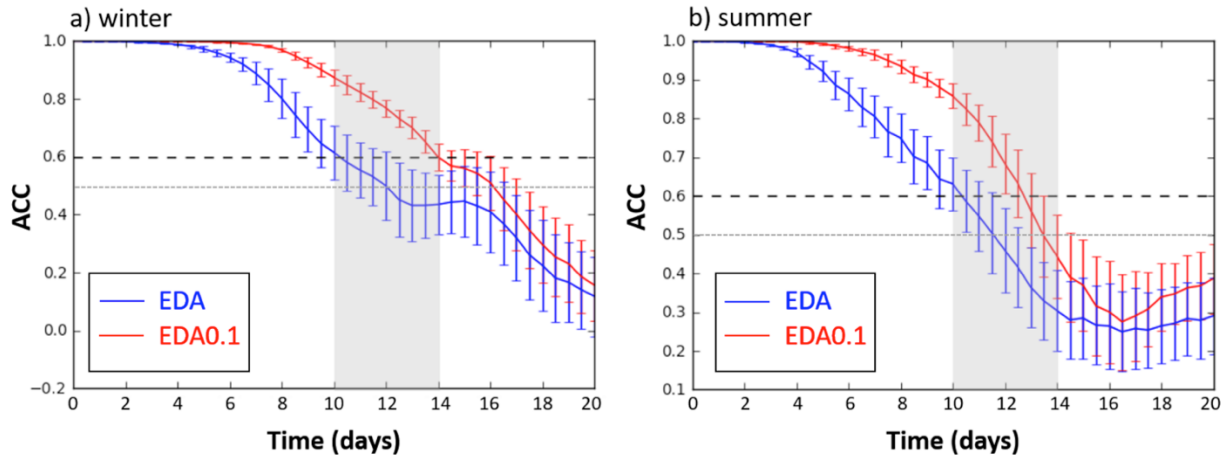

Fig. 5. Anomaly correlation coefficient (ACC) of the 500-hPa heights averaged over Northern Hemisphere mid-latitude (30-60N) for (a) 2016 Jan case and (b) 2016 July summer case. Blue line shows the ACC for all the EDA members, red line shows the results for EDA0.1 members.

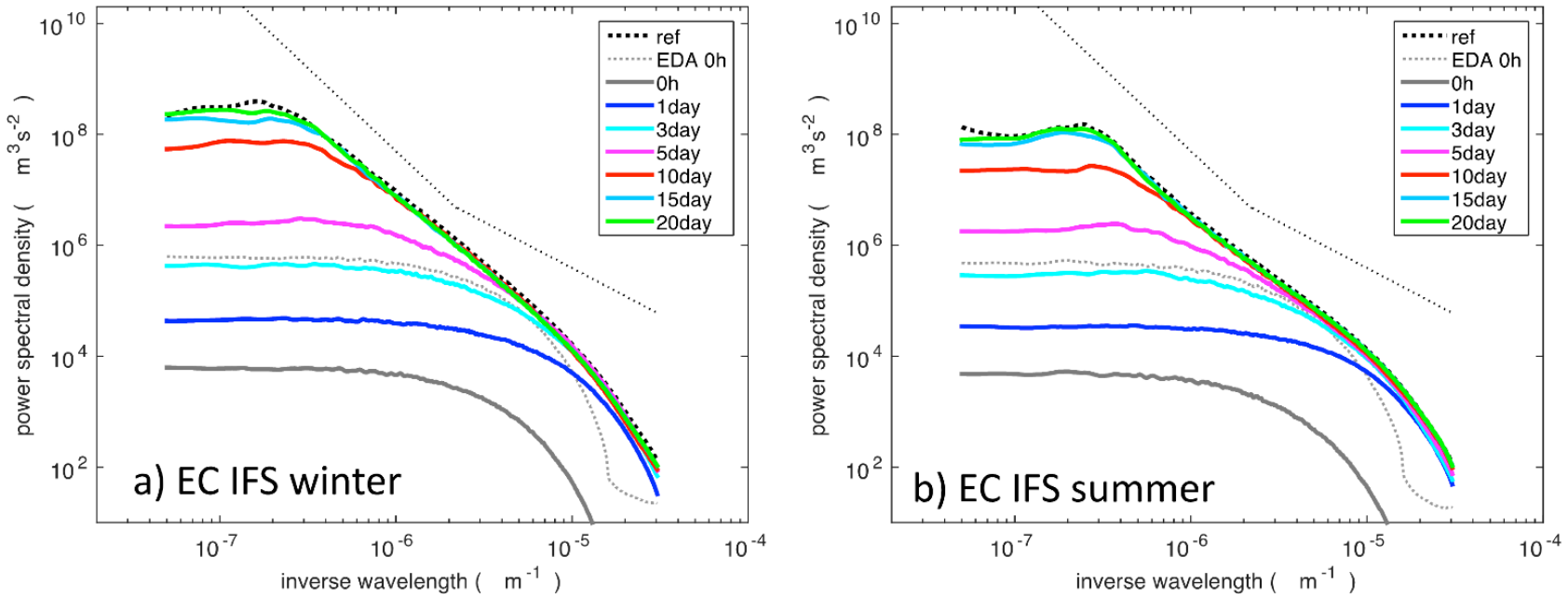

Fig. 6 Forecast error growth and saturation for different horizontal scales. Evolution of ensemble mean error spectral kinetic energy (colored lines) averaged over NH midlatitudes (40-60N) for three 9-km 10-member ECMWF IFS global ensemble simulations with minute (1%) initial condition errors (solid gray lines) initialized on three 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). The corresponding initial spread of the EDA ensemble with realistic analysis uncertainties are shown in light-gray dotted lines for comparison. The two straight dotted line segments denote the spectral slopes of -3 and -5/3, respectively.

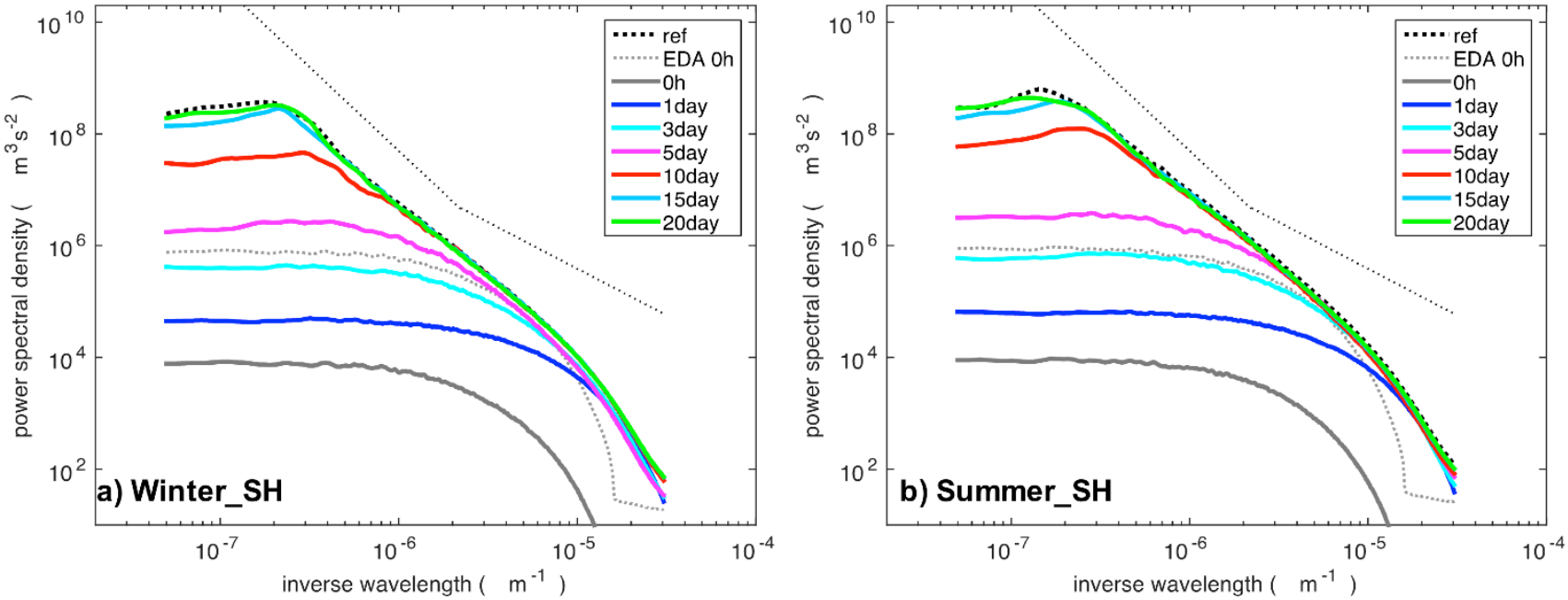

Fig. 7. Similar as Figure 6 except for the Southern Hemisphere (SH, 40-60°S).

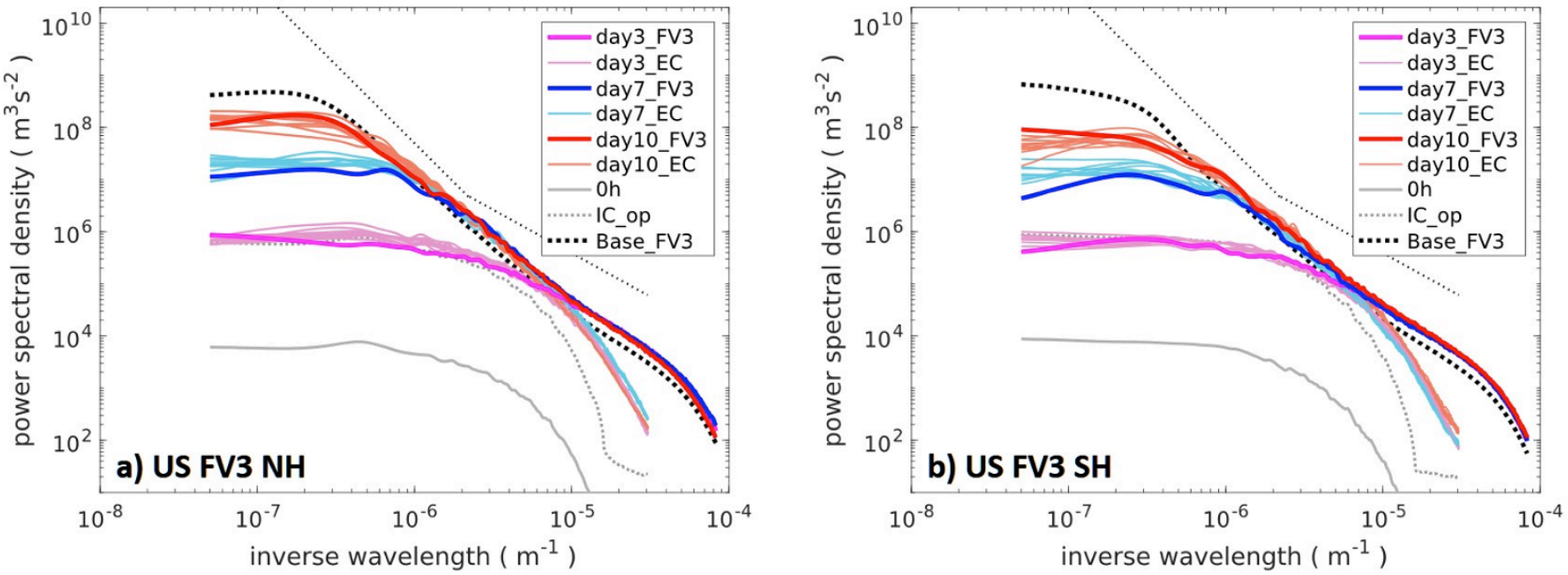

Fig. 8 Sensitivity of error growth and saturation with the 3-km FV3 model. Evolution of perturbation spectral kinetic energy between two forecasts (one initialized with the control EDA analysis and the other perturbed with the same initial perturbations as in EDA0.1 ensemble member 1 performed with the 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 December 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 ctrl run. The FV3 results lies 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.

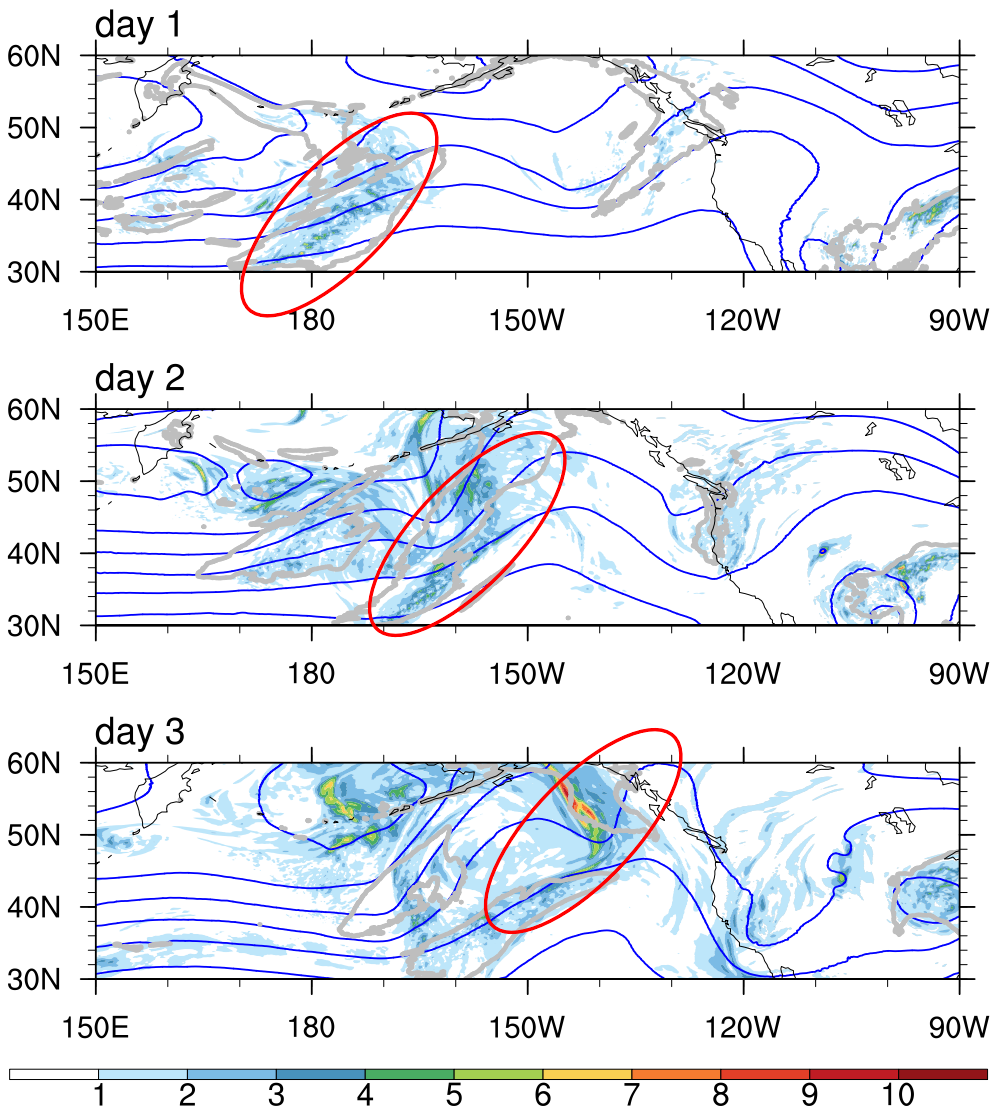

Fig. 9 Evolution of the ensemble spread of 500-hpa meridional winds (color shaded region) for the first three days in EDA0.1 integrated from 0000 UTC 26 Dec 26 2015. Blue contour is the 500-hpa geopotential height and grey contour is the regions with 12 hour precipitation greater than 0.1 mm. Red circles imply the movement of the developing low-pressure synoptic storms.

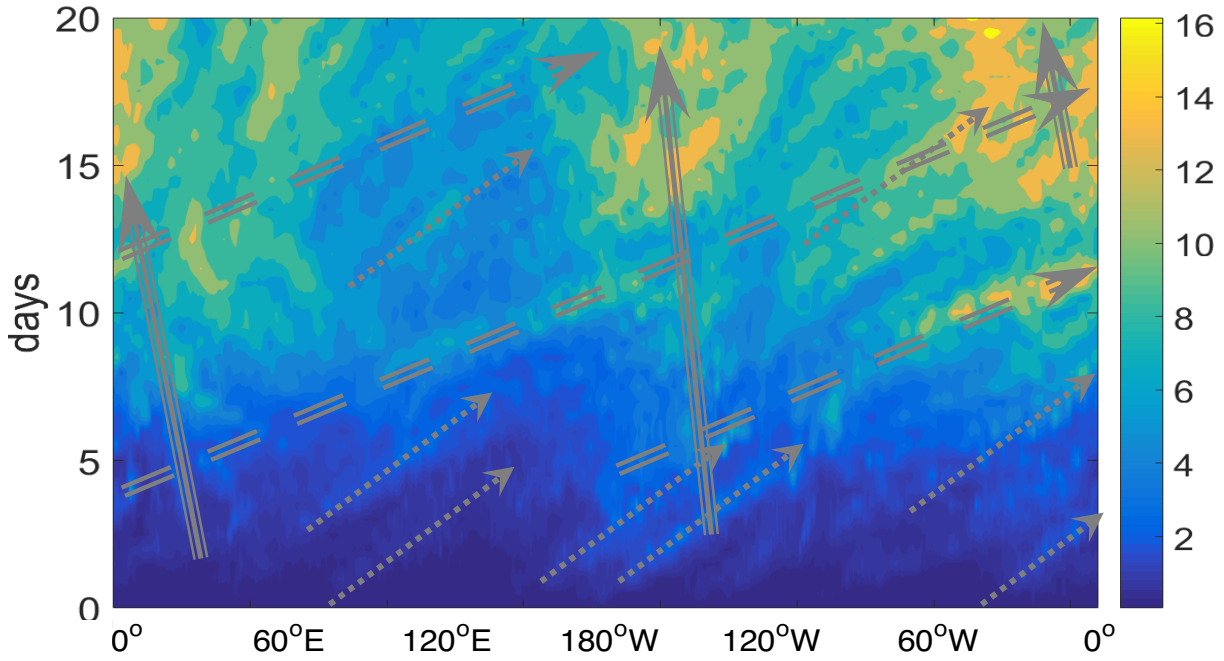

Fig. 10 The longitude-time Hovmöller diagram for the mean ensemble spread in terms of root-mean-square difference of kinetic energy (m/s) derived from the 9-km 20-day 10-member ECMWF IFS global ensemble initialized at 0000 UTC 24 December 2015, as a demonstration of different modes of error propagation that lead to the error saturation and intrinsic predictability limits (see text for details).

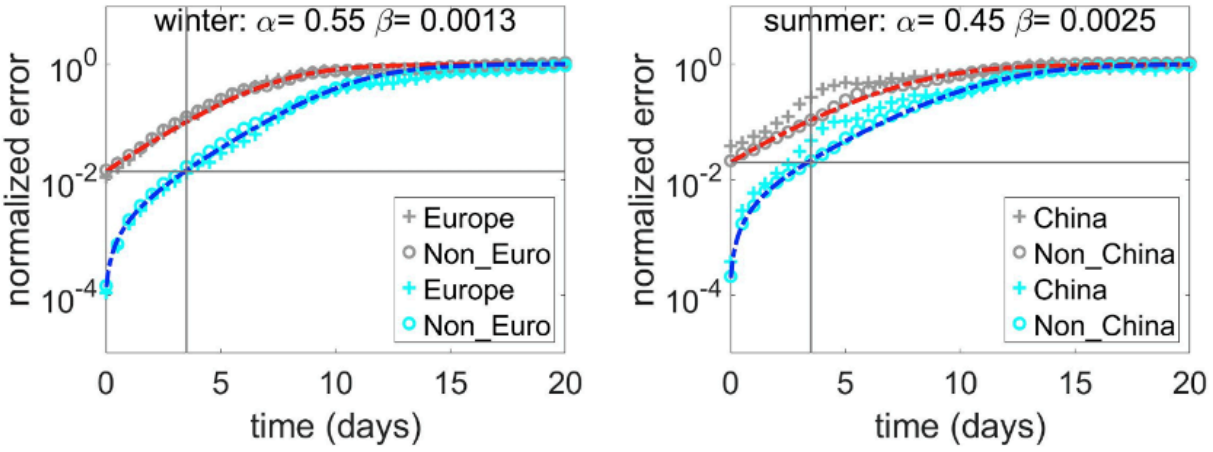

Fig. 11: As in Fig. 3c and Fig. 3d except for separating the error statistics from the hazardous event region versus the rest of the midlatitudes. Red and blue curves are the same fitted curve as in Fig. 3, implying the global averaged results. '+' represents the hazardous event region (60-degree longitude width). 'o' represents the non-hazardous-event region (300-degree longitude width).