Estimating the Intrinsic Limit of Predictability Using a Stochastic Convection Scheme

TOBIAS SELZ

Meteorologisches Institut, Ludwig-Maximilians-Universität, Munich, Germany

(Manuscript received 1 December 2017, in final form 31 October 2018)

ABSTRACT

Global model simulations together with a stochastic convection scheme are used to assess the intrinsic limit of predictability that originates from convection up to planetary scales. The stochastic convection scheme has been shown to introduce an appropriate amount of variability onto the model grid without the need to resolve the convection explicitly. This largely reduces computational costs and enables a set of 12 cases equally distributed over 1 year with five ensemble members for each case, generated by the stochastic convection scheme. As a metric, difference kinetic energy at 300 hPa over the midlatitudes, both north and south, is used. With this metric the intrinsic limit is estimated to be about 17 days when a threshold of 80% of the saturation level is applied. The error level at 3.5 days roughly compares to the initial-condition uncertainty of the current ECMWF data assimilation system, which suggests a potential improvement of 3.5 forecast days through perfecting the initial conditions. Error-growth experiments that use a deterministic convection scheme show smaller errors of about half the size at early forecast times and an estimate of intrinsic predictability that is about 10% longer, confirming the overconfidence of deterministic convection schemes.

1. Introduction

Forecast skill has improved continuously over the last 40 years. The rate of improvement reached about one forecast day per decade, which means that a 6-day forecast today is as good as a 5-day forecast was 10 years ago. This considerable improvement together with its high socioeconomic impact has been recognized as a "quiet revolution" by Bauer et al. (2015). However, such studies of past successes immediately raise the question of how far this progress will go on in the future, especially since the rate of improvement considerably slowed down in the past 5 years.

There are two "types" of predictability that need to be distinguished (Lorenz 1969; Melhauser and Zhang 2012): practical predictability, which is the ability to predict based on the procedures currently available, and intrinsic predictability, which is the extent to which prediction is possible if perfect procedures are used. Early investigations already showed that predictability may be intrinsically limited to about 2 weeks by upscale error growth from small scales (Lorenz 1969). In agreement with Lorenz, Durran and Weyn (2016) pointed out that the scale of the initial error is of minor importance to the subsequent upscale error growth. Even a very accurate knowledge of the small-scale initial conditions can be destroyed by a tiny relative error on the largest scales. Already Hohenegger and Schär (2007) recognized that vastly different initial-condition errors have the same effect on the convective scale after about 10 h.

These results suggest that no matter how accurate our observations and data assimilation procedures will become in the future there is no hope of forecasting the position and strength of individual convective cells for longer than a few hours. Zhang et al. (2007) developed a three-stage conceptual model for upscale error growth in which stage 1 describes this decorrelation of the cloud field that then in stage 2 starts to spin up geostrophically balanced motions. This spinup is related to the divergent outflow at the cloud tops, which undergoes geostrophic adjustment (Bierdel et al. 2017, 2018). In stage 3 of the

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JAS-D-17-0373.s1.

Corresponding author: Tobias Selz, tobias.selz@lmu.de

DOI: 10.1175/JAS-D-17-0373.1

conceptual model the perturbation projects onto the baroclinic instability, although recent work suggests that barotropic near-tropopause interactions may be more important (Baumgart et al. 2018).

If the error of the initial condition becomes small enough in amplitude the upscale process described above dominates direct growth on synoptic scales and basically determines the intrinsic limit of weather prediction up to planetary scales. Already Lorenz (1969) suggested this with his experiment C. It has been further confirmed by the sensitivity of the magnitude of a large-scale error after a certain lead time on moisture (Zhang et al. 2007) or on the convection scheme Selz and Craig (2015a). Indeed, error-growth experiments with global numerical models suffer from the low grid resolution and the need of a parameterization scheme for deep convection. Selz and Craig (2015a) showed by explicitly comparing simulations with convection-permitting resolution to simulations with parameterized convection on coarser grids that error growth is significantly slowed down when convection is parameterized.

Global convection-permitting studies are usually computationally too expensive. However, very recently such a study has been done (Judt 2018), although only one pair of experiments for one case has been simulated up to global saturation. The predictability limit of the upper troposphere was estimated to be 17 days. Another attempt to determine the intrinsic limit of predictability was also very recently done by Zhang et al. (2019) using the European Centre for Medium-Range Weather Forecasts (ECMWF) model at 9-km resolution and reducing the initial-condition uncertainty from the ensemble data assimilation by 90%. Six cases have been simulated, and it was concluded that predictability can be increased by 3–5 days compared to current forecasting capabilities.

Here we will take a different approach to the problem of intrinsic predictability by using coarser, computationally cheaper model resolution and replacing the convection scheme with a stochastic version. The lower resolution will allow for the simulation of more cases. A comparison to error-growth simulations with a deterministic convection scheme will also be performed.

2. Methods

a. Experimental design

The stochastic convection scheme of Plant and Craig (PC; Plant and Craig 2008) will be used in this study to assess intrinsic predictability limits without the need to resolve the convection explicitly. This approach is justified by the results of Selz and Craig (2015a), who showed for a regional model that the stochastic convection scheme is able to generate similar synoptic-scale errors from upscale error growth at 60-h lead time as a convection-permitting simulation. The scheme explicitly models individual updrafts within a grid box, where the strength of each updraft (the cloud-base upward mass flux) is randomly drawn from a distribution. The total upward mass flux M is given by the sum of the individual updrafts (clouds) in the grid box. An ensemble of convective realizations (microstates) consistent with the larger-scale environment (macrostate) can be created by using different random seeds. A similar closure assumption as for a conventional convection scheme is used to determine the ensemble average of the total upward mass flux $\langle M \rangle$. We refer to Craig and Cohen (2006) for more details on the theory behind the stochastic convection scheme.

The stochastic perturbations introduced by the PC scheme are illustrated in Fig. 1, where the average upward mass flux $\langle M \rangle$ is shown in the top-left panel and the realized mass flux M of three ensemble members is shown in the remaining panels. Figure 1 demonstrates the kind of perturbations that are initially introduced into the simulations while everything else is unchanged.

This reshuffling of the clouds is consistent with the end of stage 1 (complete decorrelation of the cloud field) of the conceptual error-growth model of Zhang et al. (2007). It is important to note that any initial-condition uncertainty on any scale is able to decorrelate the clouds within $O(10)$ h because of the low intrinsic predictability of the convective scale (Hohenegger and Schär 2007; Durran and Weyn 2016). The exact duration of this stage 1 will however depend on the magnitude of the initial-condition uncertainty, but it is short compared to predictability estimates for the planetary scale and will be neglected in the rest of the paper.

As already pointed out in the introduction, if the initial-condition uncertainty is small enough in amplitude the upscale error growth will be faster than the direct growth of synoptic-scale errors within large-scale instabilities. Therefore simulations that initially only use decorrelated cloud fields generated by the PC scheme with different random seeds are capable of assessing the intrinsic predictability limit on synoptic and planetary scales. The predictability limit will be assessed by the decorrelation of the largest Rossby modes, and this study will thus focus on the upper troposphere and the midlatitudes.

b. Numerical model

For the simulations in this paper the icosahedral nonhydrostatic (ICON) model is used (Zängl et al. 2015), which has been recently developed at the German Weather Service (DWD) and the Max Planck Institute for Meteorology. ICON uses an icosahedral triangular grid and a nonhydrostatic dynamical core. For ICON the PC convection scheme is available, and basic testing has

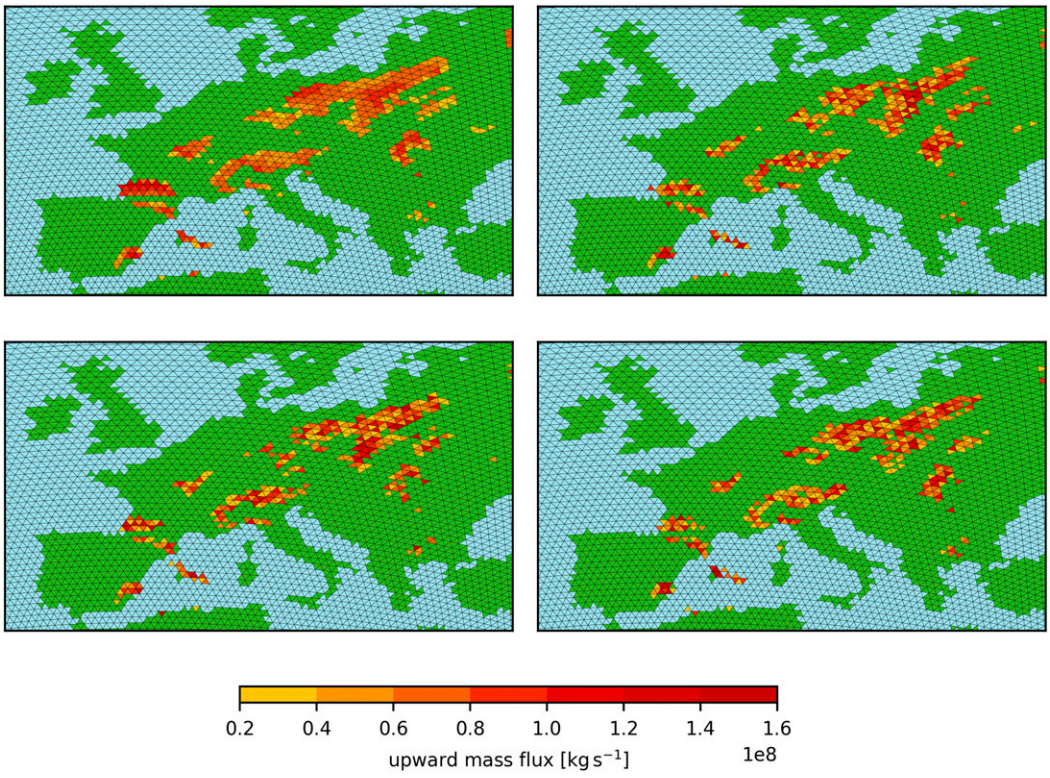

FIG. 1. (top left) Ensemble-mean mass flux of the July 2017 run after 12 h on a subdomain. (top right),(bottom) Three ensemble members (random draws) from the mean mass flux. The triangles indicate the ICON model grid.

been performed (Keane et al. 2014). The model is used on a R2B6 grid for this study which refers to a grid spacing of about $\Delta x \approx 40$ km. Twelve cases were simulated, starting from the first day of each month from October 2016 to September 2017 and run for 31 days. Initial conditions are provided by the ECMWF analysis.

Each case is simulated five times with a different random seed of the convection scheme to create a five-member ensemble with different convective realizations within an unchanged macroscopic state. Technically all simulations are started 12 h earlier (i.e., 12 h before the first day of each month). Within these 12 h the seed of the scheme is the same for all members and the simulations are identical. This 12-h period is introduced because a certain adjustment of the model state and a spinup of the convection scheme is needed when the model is started from interpolated ECMWF analysis fields. After the 12 h, the seed is changed, and the ensemble members start to differ. For the rest of this paper we will ignore the spinup period for simplicity and pretend that each simulation begins at the first day of each month.

Because of the good agreement of growing errors with respect to convection-permitting simulations when the stochastic convection scheme of Plant and Craig is used (Selz and Craig (2015a), the ICON model in this setup is considered perfect. In fact, it is sufficient to assume that differences between two simulations are perfect since no single simulation will be evaluated in this paper.

A second set of ICON ensemble forecasts has been created by using the deterministic Tiedtke–Bechtold (TB) convection scheme (Bechtold et al. 2001) and random noise as perturbation. The Tiedtke–Bechtold scheme is the scheme that is operationally used in ICON. These experiments will be discussed later in section 3d.

Output of the 300-hPa horizontal wind and height fields was generated on a regular 1° longitude–latitude grid every hour for the ICON-PC and ICON-TB simulations.

c. *Forecast metrics*

Since the focus of this study lies on the development of upper-level and midlatitude dynamics we will use the difference kinetic energy (DKE) at 300 hPa as error metric. The DKE between two ensemble members i and j of a case c is defined as

$$\mathrm{DKE}_{ij}^{c}(x,y,t) = 0.5[(u_i - u_j)^2 + (v_i - v_j)^2], \qquad (1)$$

where u is the zonal wind and v is the meridional wind. More generally one can define the DKE of an ensemble member i to the mean of the ensemble for case c by

$$\mathrm{DKE}_i^c(x,y,t) = (u_i - \overline{u})^2 + (v_i - \overline{v})^2, \tag{2}$$

where the overbar denotes the ensemble mean. One can also define the DKE of the whole ensemble of a case c by averaging (1) over all possible combinations of i and j, which can be expressed as simply taking the variance across the ensemble members:

$$\mathrm{DKE}^c(x,y,t) = \mathrm{var}(u) + \mathrm{var}(v). \tag{3}$$

Thus these three definitions of DKE are related through

$$\mathrm{DKE}^c = \frac{1}{N-1}\sum_i \mathrm{DKE}_i^c = \frac{1}{N(N-1)}\sum_{ij}\mathrm{DKE}_{ij}^c. \tag{4}$$

The DKE will be analyzed in midlatitude bands, which we define in this paper as 40°–60°N/S. It is integrated over these bands resulting in a time series that will be referred to as domain-integrated DKE (diDKE).

In addition to these time series, spectral representations of the difference kinetic energy are computed. The spectra are calculated only in longitudinal direction using a Fourier transform. A mean dx of 71 km and thus a mean k_x is used for the entire band.

For two ensemble members their difference kinetic energy matches twice the background kinetic energy when their phase relations are completely decorrelated (Selz and Craig 2015b). The distance of the difference kinetic energy to the background kinetic energy can thus be used as a measure of how much predictability is left in a certain mode. For an N-member ensemble this can be generalized by averaging over one-half of the difference kinetic energy of all possible ensemble member pairs, which can be rewritten as

$$0.5\mathrm{DKE}^c(k_x,y,t) = k_x\frac{N}{N-1}\left[\frac{1}{N}\sum(|\tilde{u}_j|^2 + |\tilde{v}_j|^2) - \left|\frac{1}{N}\sum(\tilde{u}_j + \tilde{v}_j)\right|^2\right], \tag{5}$$

where $\tilde{u}$ and $\tilde{v}$ are the Fourier transforms in zonal direction of the zonal and meridional velocity components of individual ensemble members. In short, (5) states that the spectral DKE can be computed as the mean of the transformed ensemble members minus the transformation of the ensemble mean. The spectral DKE in (5) is also averaged over the northern and southern latitudinal band resulting in $\mathrm{diDKE}^c(k_x, t)$, which is then only a function of wavenumber and time for each case. When half of the diDKE is divided by the background spectrum a relative spectral DKE can be computed that equals one where the modes have decorrelated and is below one for modes where correlation and thus predictability is left. This metric is similar to the decorrelation scale used in Surcel et al. (2015). Note that (5) has a factor k_x in front because we define the spectral DKE to fulfill

$$\mathrm{DKE}^c(y,t) = \int_{k_{\min}}^{k_{\max}} d\ln k_x\, \mathrm{DKE}^c(k_x,y,t). \tag{6}$$

In our opinion this convention is more appropriate when spectra are investigated on a logarithmic x axis.

3. Results

a. Time series of DKE

As an example Fig. 2 shows the error growth of the geopotential of the July 2017 ensemble run over the Northern Hemisphere at four different forecast times. At time zero the ensemble members are identical and thus the five lines lie exactly on top of each other. After 7 days into the forecast one can see slight perturbation to the jet, which then grow further, and after about 14 days of simulation the ensemble members seem to be completely decorrelated. An animated version of Fig. 2 can be found in the online supplemental material.

For a more quantitative analysis of error growth, time series of the diDKE can be used. Figures 3a and 3b display the distribution of the diDKE of the ensemble member time series given by (2). The distribution and the average error show the typical form of an error-growth function: fast initial growth that slows down and eventually saturates at a certain level. The distribution reveals that the saturation level is spread out, partly because it is slightly higher for the Southern Hemisphere and slightly higher in the cold season likely due to the stronger jet.

Taking a closer look at the distribution and its mean it appears that it cannot easily be modeled by a simple function in its entire range. In fact for the first circa 100 h the diDKE function seems to follow a power law that then flattens out. Therefore we only try to model the time development of diDKE from $t_0 = 7$ days on, with a function adopted from Selz and Craig (2015b):

$$\mathrm{diDKE}(t > t_0) = \mathrm{diDKE}(t_0)\exp(a\{1 - \exp[-b(t - t_0)]\}). \tag{7}$$

This function models an initial exponential increase of the diDKE with an exponentially decaying growth rate. It is thus bounded and approaches

$$\mathrm{diDKE}(\infty) = \mathrm{diDKE}(t_0)\exp(a) \tag{8}$$

as t goes to infinity. Figure 3 shows that this is a reasonable fit to the mean of the distribution and also to the selected individual cases, which quantifies the continuous loss of predictability.

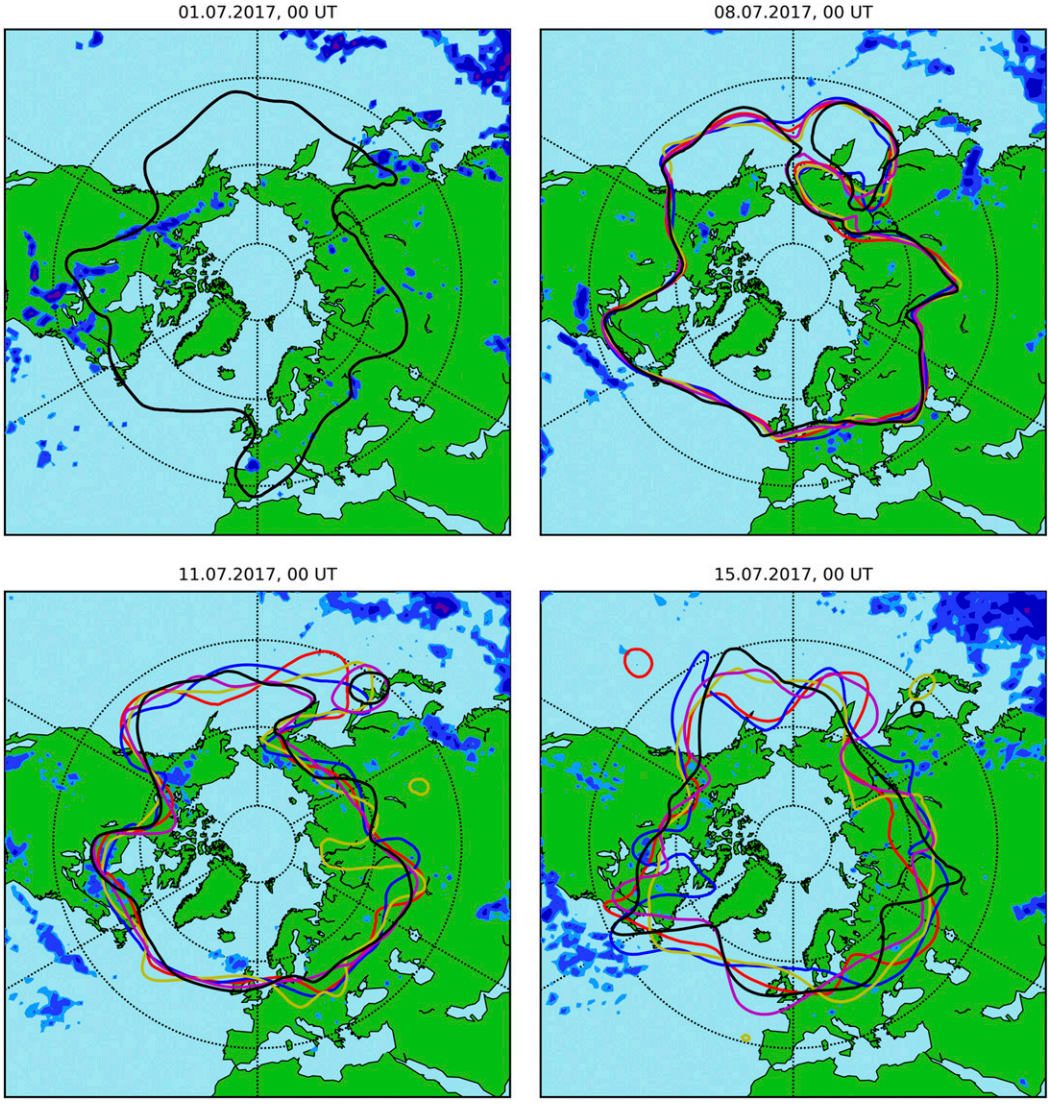

FIG. 2. Error-growth example of the July 2017 run. (top left) At forecast start, (top right) after 7 days, (bottom left) after 10 days, and (bottom right) after 14 days. Shown is the 92 000 $m^2 s^{-2}$ geopotential contour line at 300 hPa for each ensemble member (color). The bluish shading shows the ensemble-mean convective precipitation.

With the aid of the fit function in (7) one can get a predictability time estimate τ by defining a threshold at which (e.g., for a certain user of the forecast) predictability is reduced to the point that it is no longer valuable. Here we define this threshold to be 0.8diDKE(∞), and thus, the predictability time τ can be derived for the mean and all cases and both latitudinal bands: the least predictable case (τ = 13.3 days) is the July 2017 case in the Northern Hemisphere which was already introduced in Fig. 2, while the best predictable case is the January 2017 case in the Southern Hemisphere (τ = 24.6 days). Both cases are plotted in Fig. 3b, together with the fit to (7).

On average the predictability time with the chosen threshold is 17.4 days, a bit more than 2 weeks. It is somewhat lower in the Northern Hemisphere (τ = 17.1 days) and higher (τ = 17.7 days) in the Southern Hemisphere. It is also somewhat lower in summer (τ = 17.4 days) than in winter (τ = 18.0 days). The slightly higher predictability in winter might reflect the weaker convective activity and less disturbance of the large-scale flow by the convection, despite the stronger and more unstable jet. The estimated average predictability time of 17 days agrees exactly with the result from the global convection-permitting error-growth study of Judt (2018). This might however be a coincidence since in

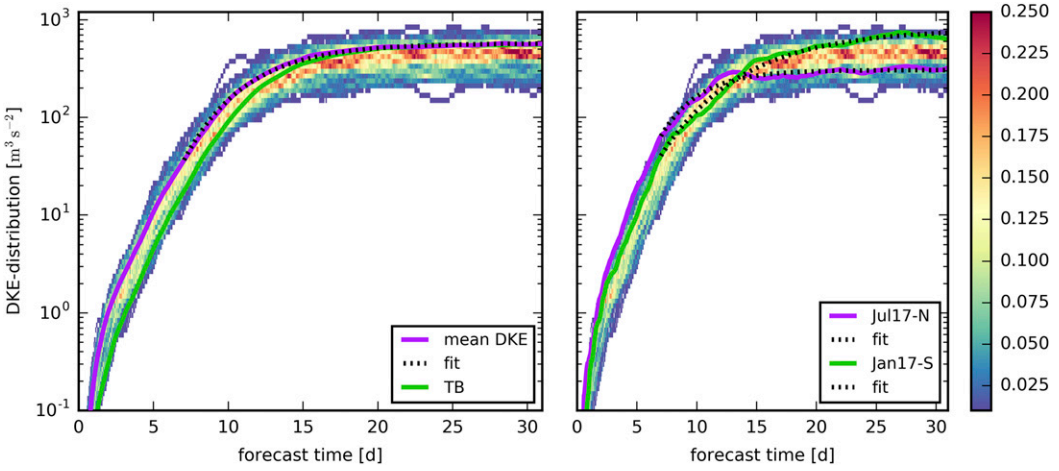

FIG. 3. Distribution of diDKE of all ensemble members using (2) (shading). The plot has been divided into bins, and the relative population of each bin is indicated by the color. Every vertical sum of bins equals one. (left) The diDKE for the ICON-PC experiments. The purple line indicates the average over all cases and latitudinal bands. The black dashed line indicates its fit using (7). The green line shows the mean over all ICON-TB experiments. (right) As in the left panel, but plotted together with (3) for the cases of lowest (July 2017, Northern Hemisphere) and highest (January 2017, Southern Hemisphere) predictability (see text) and their fits to (7) as dotted black lines.

that study only one case was simulated and there is significant variability of the intrinsic predictability estimate in our dataset, which means that some flow patterns are intrinsically more predictable than others. In addition the saturation threshold of 80% is arbitrary.

b. *Spectral analysis of DKE*

A second and more detailed method to measure error growth and saturation can be achieved with the spectral analysis described in the second part of section 2c. Figure 4 shows the error growth as spectral DKE according to (5) averaged over all cases and both latitudinal bands for four different forecast lead times together with the background spectrum. One can see that at smaller scales the error spectra match the background spectrum first, which indicates that at those scales the phase of the modes of the ensemble members are completely decorrelated and thus all predictability is lost. With the chosen normalization the spectra are not flat at the large-scale end (like in Selz and Craig 2015b) but show a pronounced maximum. This maximum moves to larger scales for longer forecast times.

To put these results in context the ECMWF ensemble prediction system's initial-condition uncertainty is considered, which is handled in a similar way using (5) but now with $N = 50$ ensemble members (dotted line in Fig. 4). The initial-condition uncertainty is very close to the output of the ensemble data assimilation system since the amplitude of the singular vectors is initially very small. It thus represents the uncertainties resulting from the current observation and assimilation system. By comparison one can see that the ECMWF initial-condition error is approximately similar to 3.5 days of ICON-PC simulation for a wavelength larger than about 500 km. This suggests that by perfecting the initial conditions one can gain at most 3.5 days of forecast time. This result agrees with the estimate of 3–5 days from Zhang et al. (2019).

c. *Scale-dependent predictability times*

Similar to the diDKE time series, there is no single point in time after which all predictability is suddenly

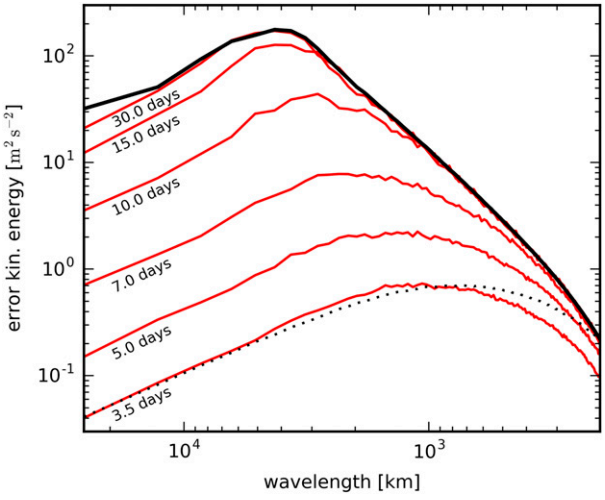

FIG. 4. Spectra of 0.5DKE according to (5), averaged over all cases and both latitudinal bands, at several forecast times shown in red for the ICON-PC experiments. The background spectrum is drawn with a solid black line. The dotted black line shows the initial-condition uncertainty of the ECMWF ensemble forecasting system.

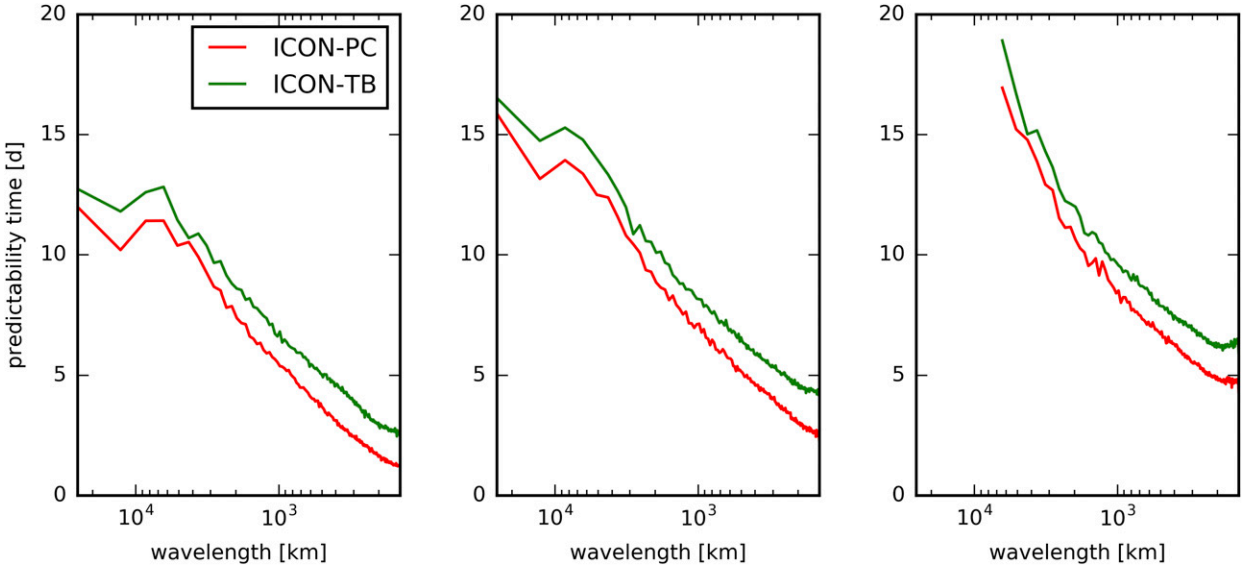

FIG. 5. Predictability times over spatial scale for several thresholds: (left) 0.25, (center) 0.5, and (right) 0.8.

lost, but the loss of predictability is a continuous process. Again, a predictability time can be assigned together with the definition of a threshold that indicates how much uncertainty a certain users of the forecast is willing to tolerate. In the case of the spectral representation of the DKE a threshold results in a scale-dependent predictability time, which is defined as the time when at a certain wavelength the relative DKE first reaches the threshold. Figure 5 shows these predictability times for three different thresholds (0.25, 0.5, and 0.8), averaged over all cases and both latitudinal bands. Note that the threshold here is applied relative to the background spectrum and not the asymptote of the errors.

In the left and center panels (0.25 and 0.5 thresholds) all modes reach the threshold within the simulation time, and thus, the predictability time is well defined for all scales. In contrast the 0.8 threshold is not hit within the simulated time at the largest scales, leaving the predictability time undetermined and >31 days. This means that the largest modes in the ensemble do not completely decorrelate with respect to the background spectrum and thus likely involve a climatological component. For example, orography and the land–sea distribution may hinder the largest planetary waves from freely evolving. In addition the ICON simulations have fixed sea surface temperatures.

d. Comparison to simulations with a deterministic convection scheme

A second set of simulations has been performed using the ICON model but this time in its standard setup with the deterministic TB convection scheme (Bechtold et al. 2001). With this setup, an ensemble was created by introducing random, gridscale noise to the temperature field with zero mean and a standard deviation of 0.01 K at forecast start (i.e., 12 h after spinup). It is important to consider these runs since previous studies have shown that noise (or errors) are not sufficiently amplified when a deterministic convection scheme is used (Selz and Craig 2015a), which might lead to an overconfidence of the ensemble whenever such processes become important. Furthermore, such experiments have been performed in the past (e.g., Zhang et al. 2003; Straus and Paolino 2009).

Indeed the ICON-TB simulations show a slower error growth, which is evident from the green line in Fig. 3. The relative difference reaches at shorter forecast times about a factor of 2. At longer forecast times, the Tiedtke–Bechtold simulations saturate at the same level but later. On average the 80% threshold is reached at 19.2 days and thus 1.8 days later than for the ICON-PC experiments. Similarly Fig. 5 shows a higher predictability time for the ICON-TB simulations and thus an overconfidence of this ensemble configuration over the whole range of scales. Again, this is most pronounced at smaller scales and equals about a factor of 2, which reduces to about a 10% difference at larger scales and longer forecast lead times. This decrease in relative difference can be explained by the fact that growth rates slow down significantly at some point (see Fig. 3) and an initial difference becomes relatively less important at later times. This result also suggests that previous studies that used a deterministic scheme together with noise

perturbations may still be valid approximations when longer forecast times are considered. Note however, the error might get larger with lower resolutions and with different convection schemes.

4. Summary and discussion

Twelve cases of five-member global ICON ensembles are generated with different random realization of the stochastic convection scheme by Plant and Craig (ICON-PC). These ensembles estimate the impact of the convective uncertainty on synoptic and planetary scales without the need to resolve the convection explicitly (Selz and Craig 2015a). Being set up with exact same initial conditions and no further perturbations, the ICON-PC ensemble is suitable to assess the intrinsic limit that is imposed by the convection on synoptic and planetary scales. It is found that the intrinsic limit of forecasting planetary scales is about 17 days (80% threshold), while at 3.5 days the initial-condition error of the current ECMWF data assimilation system is reached. Comparing experiments using the deterministic Tiedtke–Bechtold convection scheme showed an estimate of the intrinsic limit of about 19 days, which is, based on the results of Selz and Craig (2015a), likely too long. This again confirms that deterministic convection schemes do not amplify errors fast enough and thus may lead to an overconfidence of ensemble systems.

The results presented here only hold on average over many cases. We used 12 cases distributed over 1 year, which result in 24 midlatitude flow realizations if both hemispheres are considered. However, the intrinsic predictability seems to be flow dependent and varies depending on the metric that has been used up to 10 forecast days, which indicates that there are flow conditions with higher and lower intrinsic predictability. Seasonal and hemispheric subsamples show however only slight differences with higher intrinsic predictability in winter and in the Southern Hemisphere. If these differences are significant and which factors and processes basically control the intrinsic predictability are interesting questions for future investigations.

The forecast limit of the ECMWF forecasting system was recently estimated by Buizza and Leutbecher (2015). For instantaneous gridpoint fields they estimated the forecast skill horizon to lie between 16 and 23 days. These time scales are longer than those estimated in this study and seem to contradict our results since this forecast skill horizon already exceeds the intrinsic limit estimated here. It is important to note, however, that the forecast horizon (or in our formulation the predictability time) strongly depends on the metric that is used. With their much larger dataset that in turn enables a more sophisticated metric for the forecast horizon, Buizza and Leutbecher (2015) were able to detect even very small remains of predictability. However, larger datasets and thus the use of a metric like continuous ranked probability score and tests against a calibrated model climate distribution are beyond our computing capacities.

We further note that this study refers to instantaneous gridpoint fields and to atmospheric turbulence only. Other Earth system components like the ocean, the stratosphere, the ground, or sea ice can enhance predictability and may lead to skillful predictions beyond the limit estimated here. Such enhancements of predictability can however usually only be detected by using temporal and spatial averages, as is done with seasonal forecasts (Buizza and Leutbecher 2015).

The 3.5-day gap between the ICON-PC simulations and the ECMWF initial conditions suggests that current forecasts are in general far away from intrinsic limits and thus not bounded by them. This does not exclude the fact, however, that locally and occasionally intrinsic limits can become important. It is an interesting topic for future work if some of the forecast bust cases (like, e.g., Rodwell et al. 2013) are results of intrinsic limits. Also, error growth due to baroclinic and shear instabilities as well as error growth in regions of stratiform precipitation might be underestimated by the ICON-PC simulations due to the coarse resolution and the inability to resolve sharp gradients. Thus our estimate of 17 days of intrinsic predictability might be too optimistic. Irrespective of that, it is an open question if technologies and procedures can ever be perfected to get close to the intrinsic limit. It seems more likely that the progress will slow down at some point (or already has) and saturates.

In this study we gave a general overview on the generated dataset of ICON simulations using the stochastic convection scheme of Plant and Craig, evaluated it with simple metrics, and justified its implication. A follow-up study is in preparation, which will investigate the different physical processes behind the error-growth stages using more sophisticated metrics.

Acknowledgments. The author would like to thank George Craig for his helpful comments and support. The use of the computing and archive facilities of the ECMWF is gratefully acknowledged. This research was carried out as part of Project ''A1—Upscale impact of diabatic processes from convective to near-hemispheric scale'' of the Transregional Collaborative Research Center SFB/TRR 165 ''Waves to Weather'' funded by the German Research Foundation (DFG).

REFERENCES

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.

Baumgart, M., M. Riemer, V. Wirth, T. Franziska, and S. Lang, 2018: Potential vorticity dynamics of forecast errors: A quantitative case study. *Mon. Wea. Rev.*, **146**, 1405–1425, https://doi.org/10.1175/MWR-D-17-0196.1.

Bechtold, P., E. Bazile, F. Guichard, P. Mascart, and E. Richard, 2001: A mass-flux convection scheme for regional and global models. *Quart. J. Roy. Meteor. Soc.*, **127**, 869–886, https://doi.org/10.1002/qj.49712757309.

Bierdel, L., T. Selz, and G. Craig, 2017: Theoretical aspects of upscale error growth through the mesoscales: An analytical model. *Quart. J. Roy. Meteor. Soc.*, **143**, 3048–3059, https://doi.org/10.1002/qj.3160.

——, ——, and ——, 2018: Theoretical aspects of upscale error growth on the mesoscales: Idealized numerical simulations. *Quart. J. Roy. Meteor. Soc.*, **144**, 682–694, https://doi.org/10.1002/qj.3236.

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.

Craig, G. C., and B. G. Cohen, 2006: Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.*, **63**, 1996–2004, https://doi.org/10.1175/JAS3709.1.

Durran, D. R., and J. A. Weyn, 2016: Thunderstorms do not get butterflies. *Bull. Amer. Meteor. Soc.*, **97**, 237–243, https://doi.org/10.1175/BAMS-D-15-00070.1.

Hohenegger, C., and C. Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478, https://doi.org/10.1175/2007JAS2143.1.

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.

Keane, R. J., G. C. Craig, C. Keil, and G. Zängl, 2014: The Plant–Craig stochastic convection scheme in ICON and its scale adaptivity. *J. Atmos. Sci.*, **71**, 3404–3415, https://doi.org/10.1175/JAS-D-13-0331.1.

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

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.

Plant, R., and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, https://doi.org/10.1175/2007JAS2263.1.

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.

Selz, T., and G. C. Craig, 2015a: Simulation of upscale error growth with a stochastic convection scheme. *Geophys. Res. Lett.*, **42**, 3056–3062, https://doi.org/10.1002/2015GL063525.

——, and ——, 2015b: 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.

Straus, D. M., and D. Paolino, 2009: Intermediate time error growth and predictability: Tropics versus mid-latitudes. *Tellus*, **61A**, 579–586, https://doi.org/10.1111/j.1600-0870.2009.00411.x.

Surcel, M., I. Zawadzki, and M. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *J. Atmos. Sci.*, **72**, 216–235, https://doi.org/10.1175/JAS-D-14-0071.1.

Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (icosahedral non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, https://doi.org/10.1002/qj.2378.

Zhang, F., C. Snyder, and R. Rotunno, 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. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of mid-latitude weather? *J. Atmos. Sci.*, https://doi.org/10.1175/JAS-D-18-0269.1, in press.