SPECIAL PROJECT FINAL REPORT

Project Title:	Flow-dependence of the intrinsic predictability limit and its relevance to forecast busts
Computer Project Account:	spdecrai
Start Year - End Year :	2022 - 2024
Principal Investigator(s)	Prof. Dr. George Craig
Affiliation/Address:	Meteorologisches Institut Ludwig-Maximilians-Universität München Theresienstr. 37 80333 München Germany
Other Researchers (Name/Affiliation):	Dr. Tobias Selz Deutsches Zentrum für Luft- und Raumfahrt Münchner Straße 20 82234 Weßling Germany

Summary of project objectives

Recent numerical experiments suggest that, on average, the accuracy of medium-range weather forecasts in the midlatitudes has not yet reached the intrinsic predictability limit proposed by the "butterfly effect", and further improvements in skill and lead time are possible. However, estimates of the intrinsic limit so far are averages over large regions and/or based on a very small number of samples. In this project, we aimed to investigate the spatial-temporal variability of the intrinsic limit and the improvement potential by simulating a large number of error-growth experiments to keep sampling uncertainty low. In addition, we investigated the ability of recent AI-based weather models to simulate the butterfly effect.

Summary of problems encountered

None

Experience with the Special Project framework

We are pleased with the special project application and reporting process and don't have any specific feedback.

Summary of results

The first goal of this special project was to improve the accuracy of low-resolution models simulation with respect to upscale error growth. Earlier studies already showed, that a stochastic convection scheme is able to amplify and spread small-amplitude errors much more realistically compared to a deterministic convection scheme (Selz and Craig, 2015). A new version of such a convection scheme based on the idea of Machulskaya and Seifert, 2019 has been implemented into the ICON model, using a better plume model and simulating vertical momentum transport. To test this new stochastic scheme, with first years computational resources a high-resolution (2.5km) global reference ensemble was computed, started from small-amplitude initial condition uncertainty. This simulation was then compared to lower-resolution simulations with and without the stochastic convection scheme. The result is given in Figure 1.

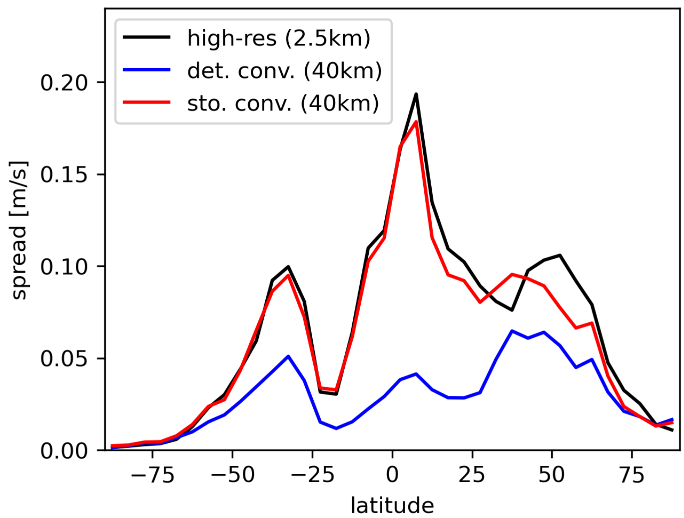

Figure 1:
Ensemble spread of the 300hPa meridional wind after 12h as a function of latitude.

It can be seen from the figure, that indeed the simulation with the stochastic convection scheme produces a much more realistic spread at early lead times compared to the simulation without the stochastic convection scheme, which largely underestimated the fast initial upscale error growth. The high-resolution reference experiment was later also used to estimate the ability of recent AI models to simulate the butterfly effect (see below).

Encouraged by this result, we spent the computational resources of years 2 and 3 to conduct a large set of ensemble simulations using the ICON model at lower resolution (40km) complemented with the stochastic convection scheme. The goal was to estimate the spatial-temporal variability of the predictability time from different levels of initial condition uncertainty. The low resolution made it possible to simulate a relatively large number of cases (i.e. every 4th day of a 2-year period, in total 183) and ensemble members (20) as well as three different levels of initial condition uncertainty (100%, 10%, 1%), which was taken from the ECMWF Ensemble data-assimilation system. From the predictability time difference between those experiments, the improvement potential that is left in weather forecasting can be estimated. Because of the large number of cases and the number of ensemble members, this analysis goes far beyond previous estimates of the intrinsic limit and the improvement potential, which only used case studies and/or twin experiments or very small ensembles (e.g. Judt, 2018, Zhang et al., 2019, Selz et al., 2022).

Thanks to the now larger sample size and the associated smaller sampling uncertainty we are now able to provide an estimate of the spatial-temporal variability of the intrinsic limit and the improvement potential. The results are given in Figure 2-4, which show the current limit estimate, the intrinsic limit estimate and the improvement potential estimate, respectively. Though these results are still preliminary and a careful evaluation of their statistical significance has not yet been performed, several features seem to emerge. First, the intrinsic limit and improvement potential is longer in the tropics compared to the midlatitudes. This result is consistent with Judt 2020, however in addition a clear longitudinal variability of the tropical improvement potential is present in our results. It is largest over the Indian ocean and lowest over Africa and the Atlantic, mostly due to variations of the current limit. In the midlatitudes, the longitudinal variations are smaller, but a larger seasonal variability appears. In the northern hemisphere the improvement potential is much shorter in summer. However, for the southern hemisphere, it is mostly independent of season.

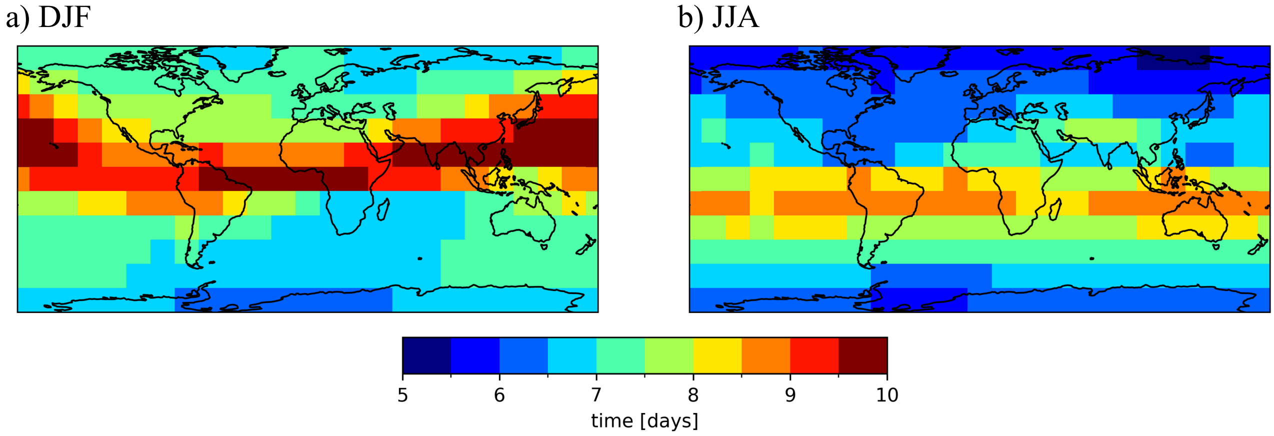

Figure 2: Current predictability limit, estimated with 300hPa DKE from the 100%-EDA experiments and a threshold of 25% with respect to of the climatological variability. Panel a) shows the norther winter season, panel b) the northern summer season. The estimates were computed over 60°x40° boxes, shifted all over the globe. The color pixels in the plot show the box center.

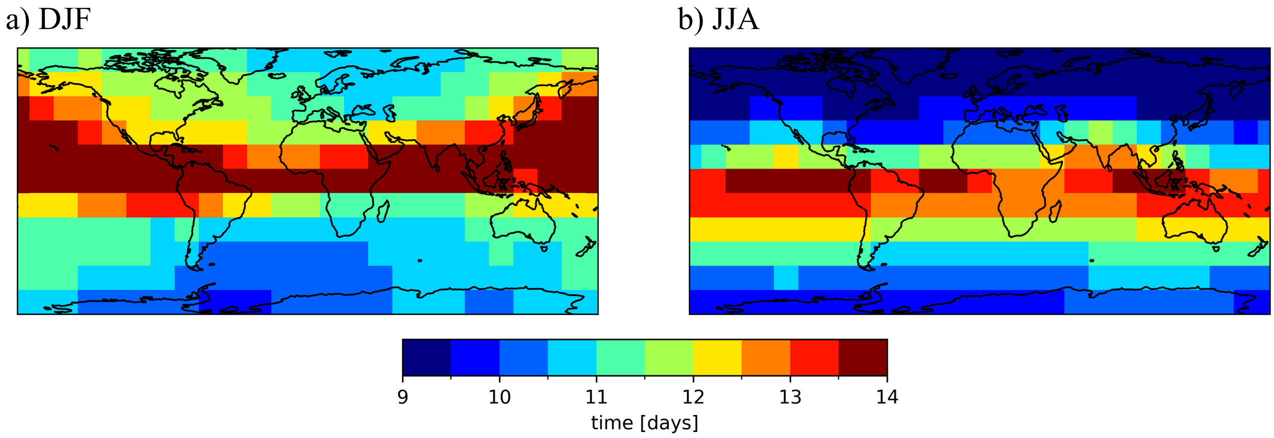

Figure 3: Intrinsic predictability limit, estimated from the 1%-EDA experiments. Otherwise same as Fig. 2.

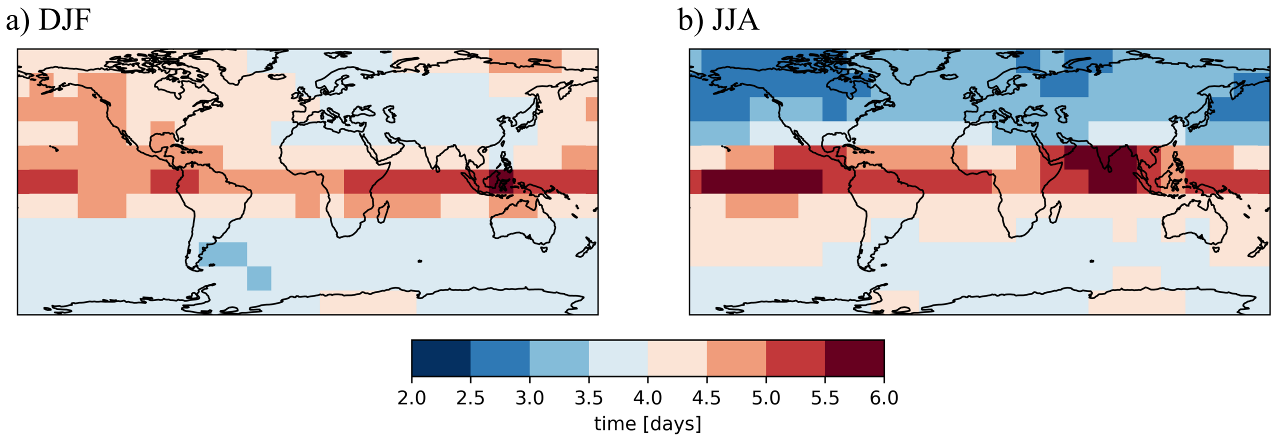

Figure 4: Estimated improvement potential as the difference between the intrinsic predictability limit (Fig. 3) and the current predicability limit (Fig. 2).

Another key finding is the similarity of the spatial patterns of the current and intrinsic limit (Figs. 2 and 3). This result was unexpected, since the initial error growth processes strongly depend on the amplitude of the initial uncertainty and should hence show different regional and seasonal characteristics: The uncertainty of the 100% experiment is amplified by large-scale processes, related to geostrophic balance, while the 1% experiment is initially amplified by convective-scale uncertainty, which quickly grows in amplitude and spreads out in scale (Selz et al., 2022). Nevertheless, the predictability time pattern of the 1% experiments mostly resembles the pattern of the 100% experiment.

The likely explanation for this phenomenon is the fact, that the fast growing and upscale propagating stage in the 1% experiments only lasts for a rather short time, which is of the order of one day. After that, the uncertainty growth continues, but now driven by large-scale balanced dynamics. Hence the 1% experiment "inherits" the characteristic growth rates of the 100% experiments for the specific region and season. Since both experiments spent most of their time in this stage of error growth, it is the growth rate of the large-scale balanced dynamics that to a large extent determines the predictability time. Because the 1% experiment spends even more time in this stage, the same pattern emerges also in the improvement potential (the difference, Figure 4). The consequence of this finding for future weather forecasting based on smaller initial condition errors is that spatial-temporal differences in predictability get more pronounced: There is more improvement potential in regions and seasons, where the forecast is already good and less improvement potential, where it is already worse. A paper about these results is currently in preparation. To what extent this result can be extrapolated to individual cases like forecast busts is however still unclear and the subject of our follow-up special project (see below).

In a second line of work, we have investigated the ability of current artificial intelligence-based weather prediction models to simulate the butterfly effect, that is the very fast initial growth of uncertainty from tiny amplitudes of initial condition uncertainty, which eventually leads to the existence of an intrinsic limit. Although computing AI simulations is cheap and has not been done at ECMWF, the ICON simulations used as reference have been calculated with the Special Project's resources, in particular the global convection-permitting ensemble. An evaluation of the AI model PANGU has already been published (Selz and Craig, 2023), but has been complemented by several other AI models since then. Figure 5 shows a comparison of four AI-models and ICON with respect to uncertainty growth. While for current level of initial condition uncertainty (100%) all AI models basically reproduce the growth rate of ICON (disregarding an initial drop due to low effective resolution), they fail to reproduce the butterfly effect (here with a 0.1% initial perturbation) and incorrectly suggest an infinite predictability of the atmosphere. Only GraphCast seem to simulate a somewhat accelerated growth rate from the tiny initial condition uncertainty. However, it slows down too quickly and a closer inspection revealed that the increase in growth rate is only due to unphysical gridscale noise.

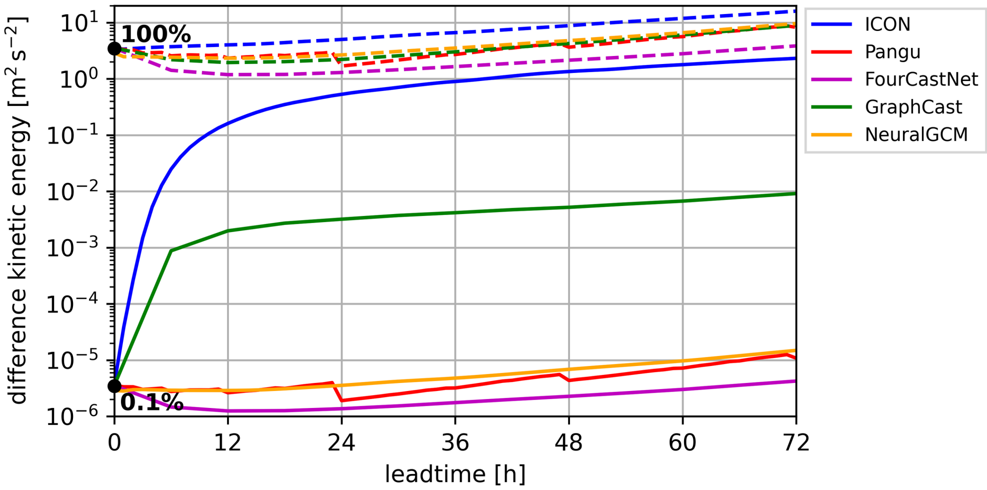

Figure 5: Time series of globally-integrated 300hPa difference kinetic energy for two levels of initial condition uncertainty (100%, 0.1%) and for four different AI-models (Pangu, FourCastNet, GraphCast, NeuralGCM). ICON as a "classic" PDE-based weather prediction model serves as a reference, with the ICON-0.1% experiment computed at global convection-permitting resolution (2.5km gridsize).

References

Selz, T. and G.C. Craig, 2015: Simulation of upscale error growth with a stochastic convection scheme. *Geo. Res. Lett.*, 42, 8, https://doi.org/10.1002/2015GL063525

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.

Judt, F., 2020: Atmospheric Predictability of the Tropics, Middle Latitudes, and Polar Regions Explored through Global Storm-Resolving Simulations. J. Atmos. Sci., 77, 257–276, https://doi.org/10.1175/JAS-D-19-0116.1.

Machulskaya, E., and A. Seifert, 2019: Stochastic Differential Equations for the Variability of Atmospheric Convection Fluctuating Around the Equilibrium, *J. Adv. Model. Earth Syst.*, 11, 2708–2727. doi:https://doi.org/10.1029/2019MS001638

Selz, T., M. Riemer, and G. C. Craig, 2022: The Transition from Practical to Intrinsic Predictability of Midlatitude Weather. *J. Atmos. Sci.*, 79, 2013–2030, https://doi.org/10.1175/JAS-D-21-0271.1

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S. Lin, J. Chen, and K. Emanuel, 2019: What Is the Predictability Limit of Midlatitude Weather? J. Atmos. Sci., 76, 1077–1091, https://doi.org/10.1175/JAS-D-18-0269.1.

List of publications/reports from the project with complete references

Selz, T. and M. Ahlgrimm, 2023a: A new stochastic scheme for deep convection based on Tiedtke-Bechtold. Conference presentation, 18. General Assembly of the IUGG, https://doi.org/10.57757/IUGG23-4393.

Selz, T. and G. Craig, 2023: Can artificial intelligence-based weather prediction models simulate the butterfly effect? *Geophysical Research Letters*, 50, e2023GL105747, https://doi.org/10.1029/2023GL105747.

Selz, T. and G. Craig, 2024: Can artificial intelligence-based weather prediction models simulate the butterfly effect? EMS Conference, Barcelona. https://vimeo.com/showcase/11449061

Selz, T. and G. Craig, 2025: Spatiotemporal variability of predictability and forecast improvement potential. ICCARUS conference, German Weather Service (DWD). https://go.dwd-nextcloud.de/index.php/s/HT67ZNE9gn5gLqf?dir=undefined&path=%2FPlenary-Wednesday&openfile=1189002

Selz, T. and G. Craig, 2025: AI weather models cannot simulate the butterfly effect. ECMWF annual seminar, poster presentation. https://events.ecmwf.int/event/418/contributions/4915/attachments/2906/4974/AS2025_Selz.pdf

Future plans

We will continue this line of work with a follow-up special project, which has already started in 2025. The emphasis now will be in the investigation of selected forecast-bust cases and if and to what extent these busts are a consequence of a transiently short intrinsic predicability limit. Depending on the process likely responsible for the bust (e.g. extra-tropical transition, continental convection) different resolutions, perturbations and model setups will be used to estimate the intrinsic limit and the improvement potential in those cases. The dataset created in this special project will serve as a reference.