SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year 2026

Project Title: Flow-dependence of the intrinsic predictability limit and its relevance to forecast busts

Computer Project Account: SPDECRAI

Principal Investigator(s): Prof. Dr. George Craig

Affiliation: Meteorologisches Institut
Ludwig-Maximilians-Universität
Theresienstr. 37
80333 München

Name of ECMWF scientist(s) collaborating to the project
(if applicable)

Start date of the project: 2025

Expected end date: 2027

Computer resources allocated/used for the current year and the previous one

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	120M	133.9M	70M	12.7M
Data storage capacity	(Gbytes)	20,000	0	40,000	0

Summary of project objectives (10 lines max)

Forecast busts still occur despite significant advances in weather forecasting, including AI-based approaches. Although rare, they can have severe societal and economic impacts, especially when linked to high-impact weather. Two main causes of busts over Europe have been identified: the extra-tropical transition (ET) of tropical cyclones, and mesoscale convective systems (MCS) over North America. Both couple complex, small-scale diabatic processes to the large-scale dynamics downstream. The project asks whether, and to what extent, forecast busts arise from a genuinely reduced intrinsic predictability limit rather than from model or data-assimilation shortcomings. We address this by estimating uncertainty growth and the remaining improvement potential from ensembles started with tiny, "butterfly-like" initial perturbations.

Summary of problems encountered

No significant problems were encountered.

Summary of plans for the continuation of the project

The plan for the current year is to add more ensemble simulations of the ET Franklin case. The goal is to demonstrate that tiny perturbations can grow very differently when placed around different members of the EDA initial ensemble. In addition, we are considering replicating the results achieved so far by repeating parts of the experiment with a different numerical model (ICON).

List of publications/reports from the project with complete references

None yet.

Summary of results

A central aim of this project is to determine whether forecast busts over Europe – occasional, sharp drops in medium-range skill – are caused merely by transiently low practical predictability, or also by a genuinely reduced intrinsic predictability limit. For average flow conditions the intrinsic limit lies roughly four to five days beyond current forecast skill (Zhang et al., 2019, Selz et al., 2022). This gap, the improvement potential, is expected to vary strongly from case to case. The first year of computing time (2025) was used to address this question for a single, high-impact case: the extra-tropical transition (ET) of tropical cyclone (TC) Franklin over the North Atlantic in early September 2023.

This case is particularly well suited to the project's aims. The transition of Franklin into the midlatitude waveguide was potentially linked to a block onset over Europe, which in turn was associated with several catastrophic weather events, including the floods over northern Africa and Greece. At the same time, the period coincided with a pronounced forecast bust that affected all major forecasting centres and the AI models alike (Figure 1), making it a textbook ET-type bust with clear societal relevance.

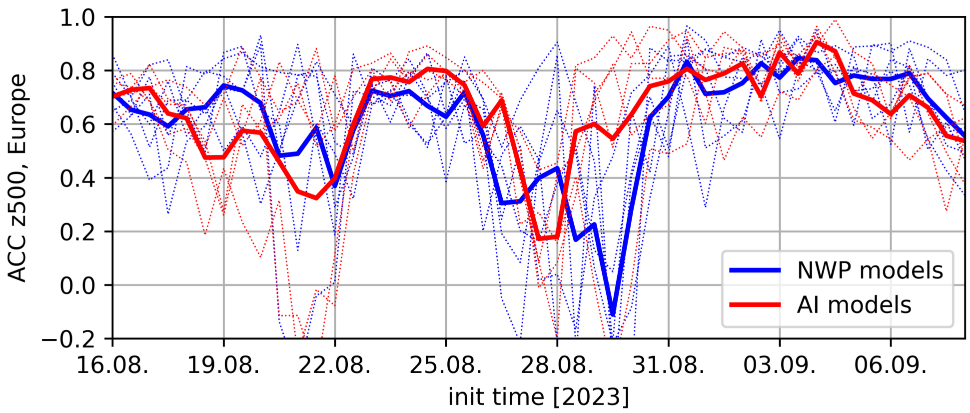

Figure 1: Anomaly correlation of 7-day forecasts over Europe as a function of forecast initialisation time. Dotted blue lines show five NWP models (ECMWF, NCEP, DWD, UKMO, ECCC); dotted red lines show five AI models (Pangu, GraphCast, FourCastNet, NeuralGCM, Aurora), started from the ERA5 reanalysis. The solid blue and red lines show the group means.

Experimental design

To assess the practical and intrinsic predictability of the Franklin ET event, and the associated improvement potential, the following experiment was designed and run with ECMWF's Integrated Forecasting System (IFS). Nine initialisation times were simulated, every 12 hours from 27 to 31 Aug 2023, each with a 50-member ensemble. No stochastic physics (SPP/SPPT) was used, so that the ensemble spread is generated purely by initial-condition perturbations. These perturbations were taken from ECMWF's Ensemble of Data Assimilations (EDA) and rescaled before being added to the analysis, by factors of 100% (unchanged), 50%, 20% and 0.1%. The 100% experiment represents the uncertainty of a current operational analysis and so characterises practical predictability, whereas the smallest factor introduces only tiny, butterfly-like perturbations and probes the intrinsic limit.

The key question of the project is whether the forecast bust (low practical predictability) is also linked to a reduced intrinsic limit – that is, whether we are, in this case, closer to the intrinsic limit than usual. If so, this would imply a reduced potential to improve such forecast busts, even with better models and observations.

Reduced intrinsic predictability and improvement potential

To answer this question for the present case we examine Figure 2, which shows the z500 spread over Europe as a function of forecast lead time for the nine initialisation times and the different rescaling-factor experiments. The improvement potential can be read directly from the separation between these curves: if even the 0.1% butterfly ensemble spread grows to closely match the 100% ensemble, reducing the initial uncertainty no longer helps and the forecast has effectively reached its intrinsic limit. Several initialisation times stand out in exactly this way: the rescaling experiments converge, especially near a lead time of seven days, indicating a much shorter intrinsic limit and reduced improvement potential than usual. This happens only for three initialisation times, namely 28 Aug 00 UTC, 29 Aug 00 UTC and 29 Aug 12 UTC.

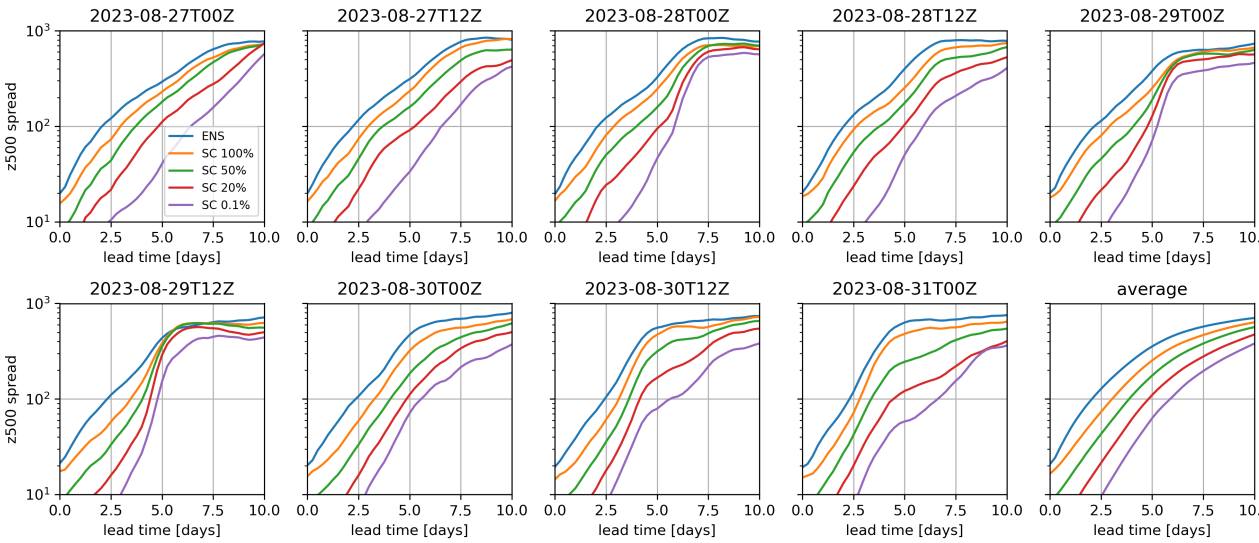

Figure 2: Growth of the 500 hPa geopotential-height spread with lead time over Europe, for the different initialisation times (panels) and rescaling-factor experiments (colour). ENS is the operational ensemble, including physics perturbations and singular vectors. The lower-right panel, "average", shows the spread growth averaged over all initialisation times and over the entire northern midlatitudes, as a reference.

The reason for the rapid spread growth

To understand the underlying dynamics, we examine two contrasting initialisation times in more detail. We begin with why the spread in the 28 Aug 2023 00 UTC butterfly experiment (SC 0.1%) grows so rapidly, in particular between 5.5 and 7.5 days lead time. A k-means cluster analysis (k=2) over Europe, applied to this experiment at 7.5 days lead time (valid 4 Sep 12 UTC), shows that the large spread arises mainly from a two-scenario outcome (Figure 3). In scenario 1 the jet to the north of the block is largely zonal, whereas in scenario 2 it is tilted to the north-east.

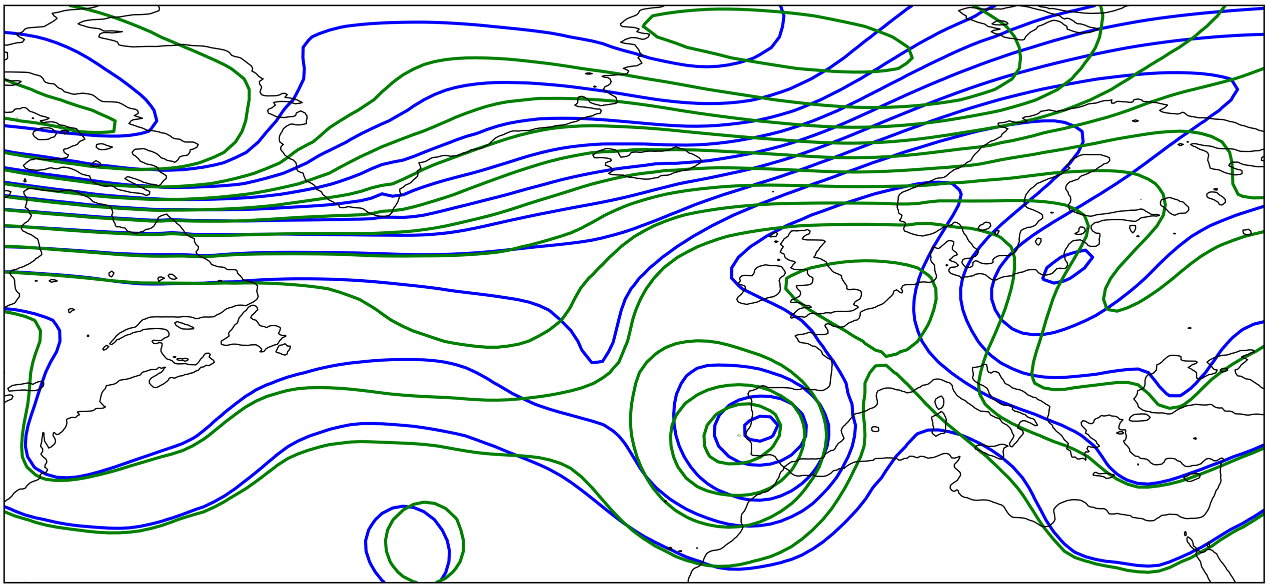

Figure 3: Cluster means from a k-means cluster analysis (k=2) of the 500 hPa geopotential of the 28 Aug 00 UTC, SC 0.1% experiment at 7.5 days lead time (valid 4 Sep 12 UTC). Green lines show cluster mean 1 (scenario 1); blue lines show cluster mean 2 (scenario 2).

The cause of this jet tilt can be traced back to the transition of TC Franklin a few days earlier. Figure 4 shows the TC positions of the 50 ensemble members, colour-coded by the cluster identified above. The outcome over Europe clearly depends on whether the TC moves ahead of the upper-level trough two to three days earlier. If it does not (scenario 1), the jet over Europe evolves into a more zonal state; if it does (scenario 2), the jet is forced to the north-east, producing a stronger block over Eastern Europe.

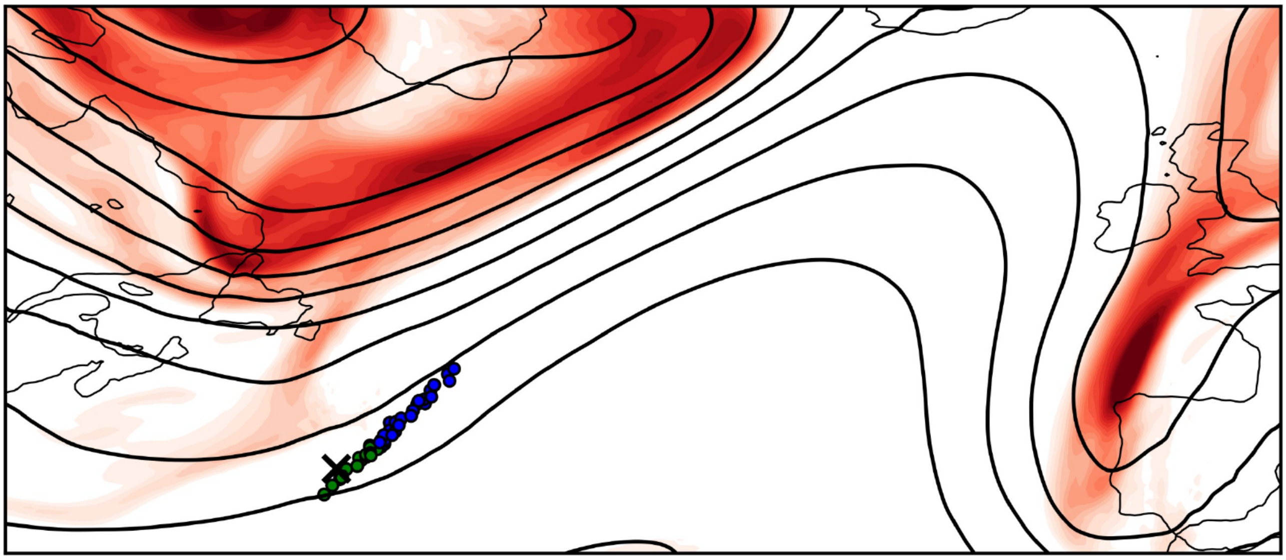

Figure 4: Ensemble-mean 300 hPa geopotential (lines) and 300 hPa potential vorticity (red shading), together with the TC Franklin positions of the 50 ensemble members (dots). Green dots belong to scenario 1 and blue dots to scenario 2. The plot shows the 28 Aug 00 UTC, SC 0.1% experiment at 5 days lead time. The black cross indicates the true TC position, derived from the ERA5 reanalysis.

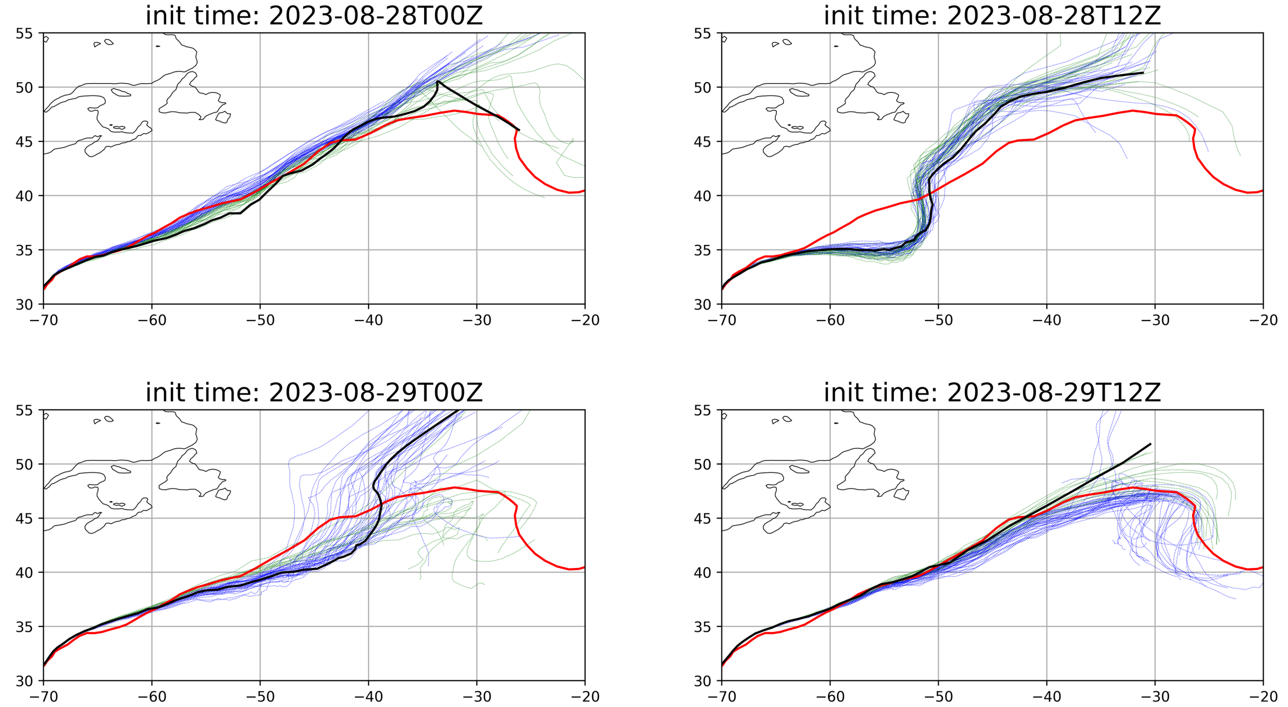

Figure 5: TC Franklin track forecasts of the 50 ensemble members of the SC 0.1% experiment, initialised at 28 Aug 00 UTC (upper left), 28 Aug 12 UTC (upper right), 29 Aug 00 UTC (lower left) and 29 Aug 12 UTC (lower right). The black line shows the track of the unperturbed control experiment (IFS-CTL); the red line shows the observed track derived from ERA5.

The question then arises why the 28 Aug 12 UTC case in particular shows no reduced intrinsic predictability, despite sitting between two cases that do. This can be understood from Figure 5, which shows the SC 0.1% track forecasts for consecutive initialisation times. It can be seen that, in the case that did not lead to a reduced intrinsic limit (28 Aug 12 UTC), the control track clearly deviates both from the observed track and from the control tracks of the other three cases. This strongly reduced the impact of the butterfly perturbations in this case, which are placed around the analysis: all members missed the crucial bifurcation zone and ended up too far south to be drawn into the midlatitudes ahead of the upper-level trough. Consequently the ET occurred later and in a much less sensitive configuration of the midlatitude wave, producing neither enhanced spread nor a tilt of the downstream jet.

Conclusion

In part, the uncertainty growth from the tiny initial perturbations is indeed faster than average, indicating a reduced intrinsic predictability associated with the ET. However, this growth depends strongly on track biases related to the analysis uncertainty. This suggests that the intrinsic predictability downstream of an ET event may be effectively indeterminable: different tracks that are equally consistent with the analysis uncertainty lead to very different sensitivities to the tiny perturbations. In other words, the case-to-case variability of intrinsic predictability in an ET event can be so large that even two different draws from the same day's initial-condition distribution can yield completely different outcomes. As a next step, we plan to demonstrate this explicitly by centring the butterfly perturbations not on the analysis but on different members of the EDA initial ensemble in the 28 Aug 12 UTC case.

References

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.-J. Lin, J.-H. 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.