Numerical Models Still Outperform AI models in Forecasting Record-breaking Events

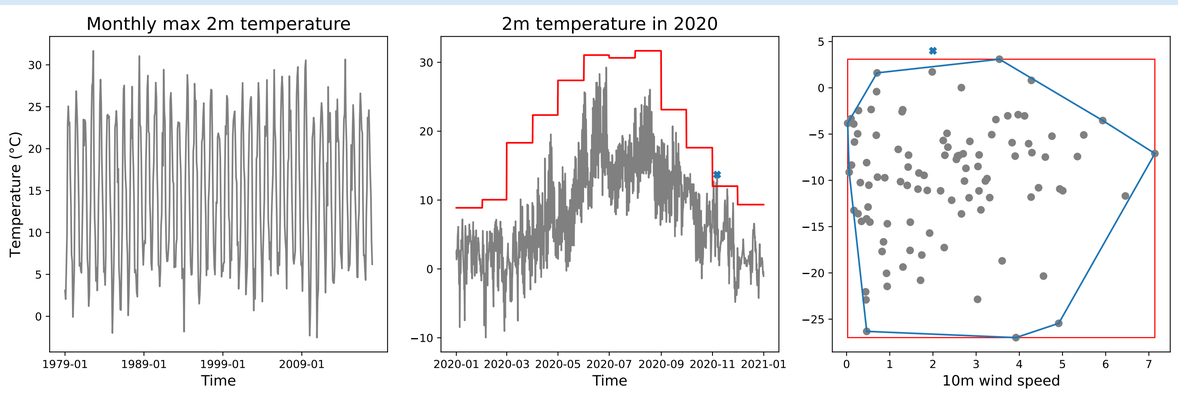
Zhongwei Zhang[1,2], Erich Fischer[3], Jakob Zscheischler[4,5], Sebastian Engelke[1]
[1]UNIGE (GSEM-RISIS), [2]KIT (SCC), [3]ETH Zurich (IAS), [4]UFZ (CER), [5]TUD (Hydro Sciences)
Contact: zhongwei.zhang@kit.edu

Abstract

- Critical question: can AI weather forecasting models improve the prediction of unprecedented extreme events that are unseen in their training data?
- For now, only case studies address this question (Pasche et al., 2025)
- We evaluate models on records to systematically assess their extrapolation capabilities
- For key variables (2m temperature and 10m wind speed over land), GraphCast, Pangu, and Fuxi consistently under-perform HRES on record-breaking events, across nearly all lead times
- This remains true for the operational version of these models

Record and extrapolation

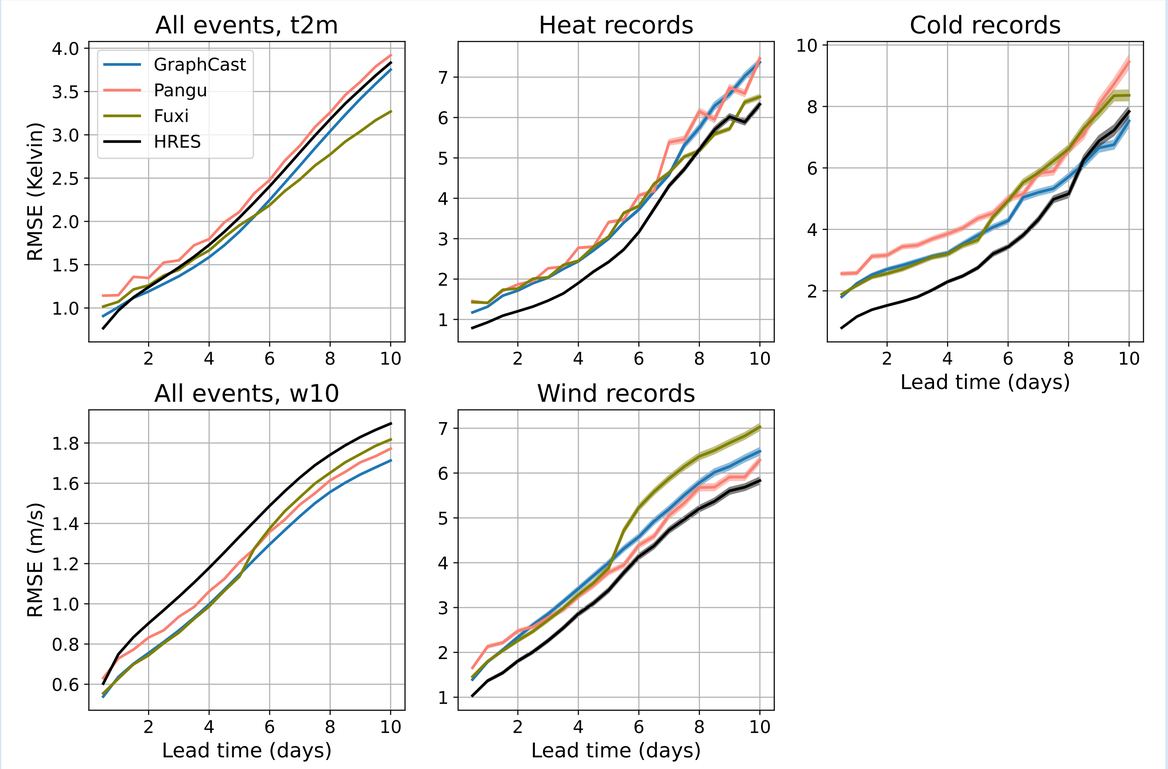
Record: defined on training data of AI models (ERA5 1979–2017); per variable and grid cell at $0.25°$ resolution; per month (red line)

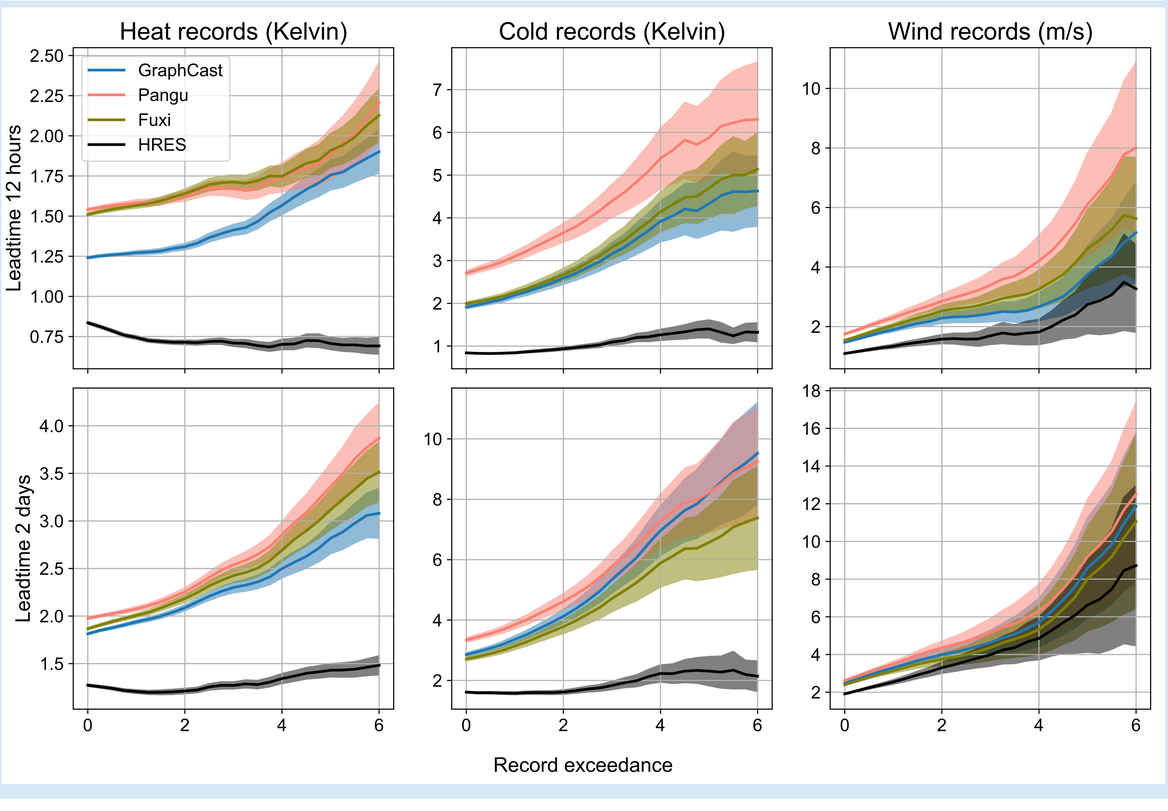


Extrapolation: when the observation, input, or forecast data of test year 2020 is beyond the training record (blue cross)

Non-operational forecast

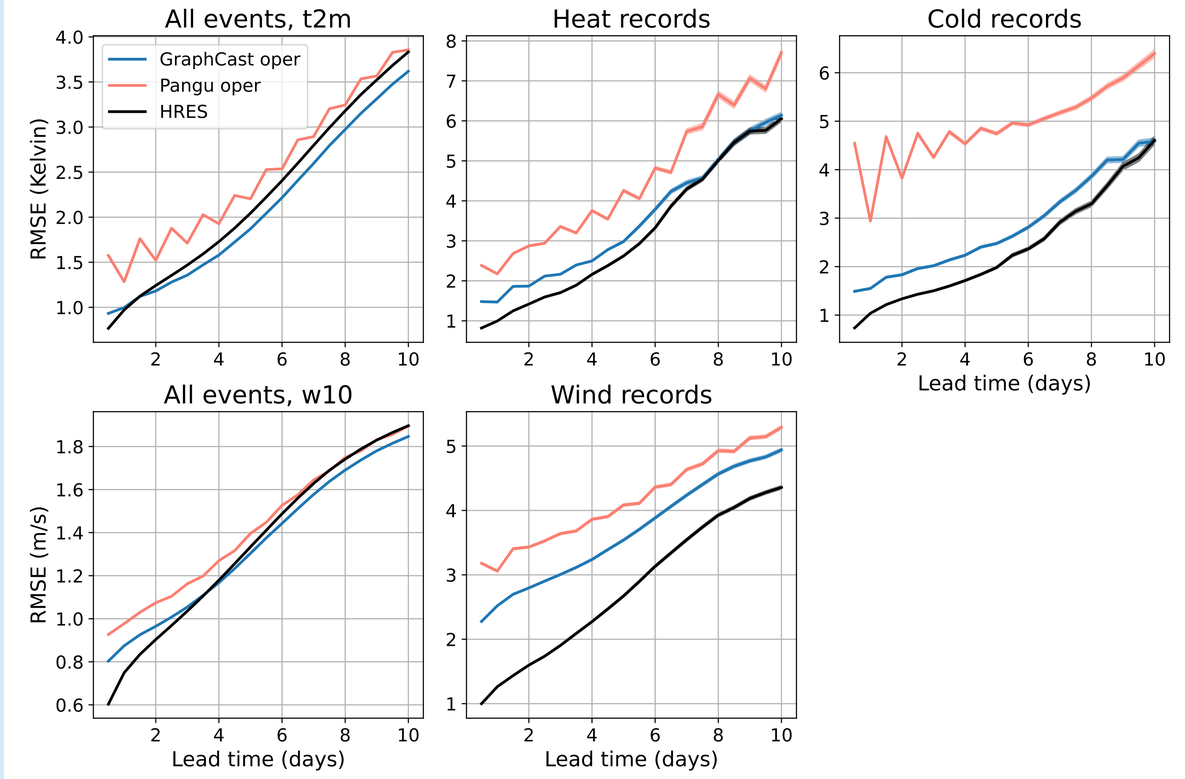
RMSE on record data set over land (Antarctic excluded) in 2020
latitude-weighted; record-breaking events selected from ERA5; AI forecasts compared against ERA5, HRES against HRES-fc0

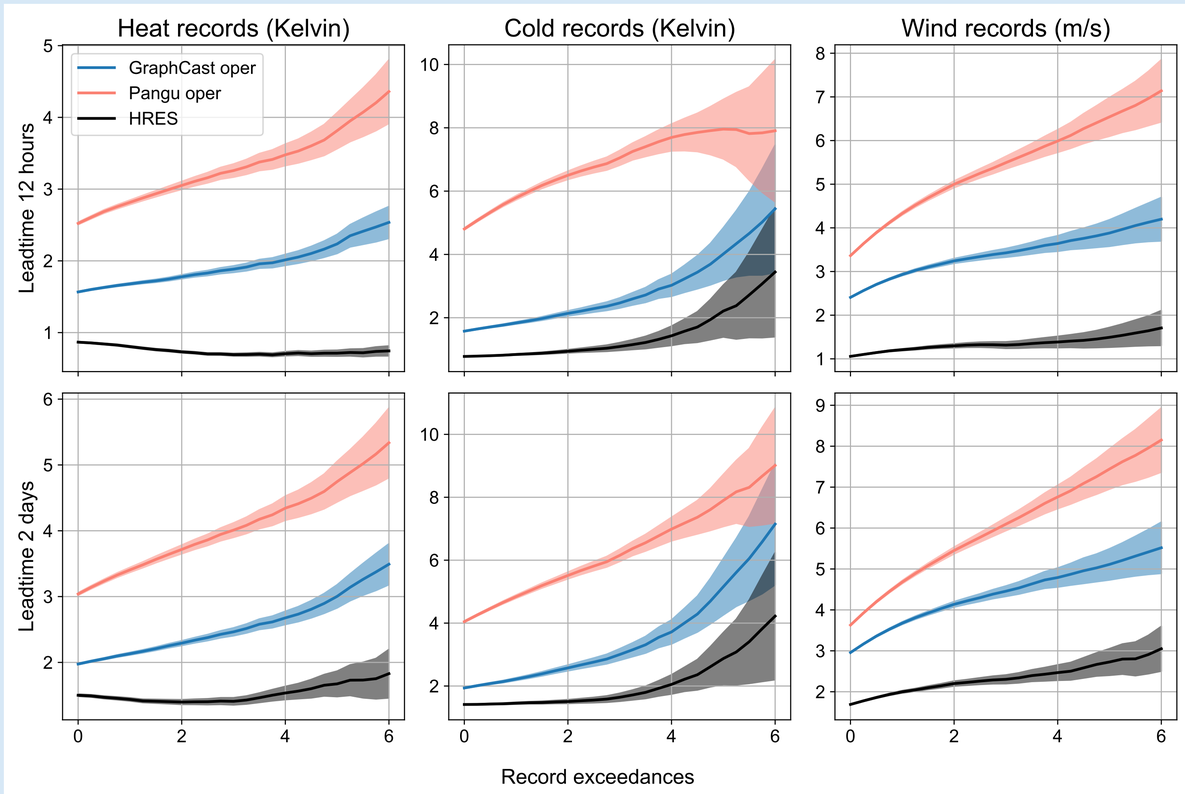




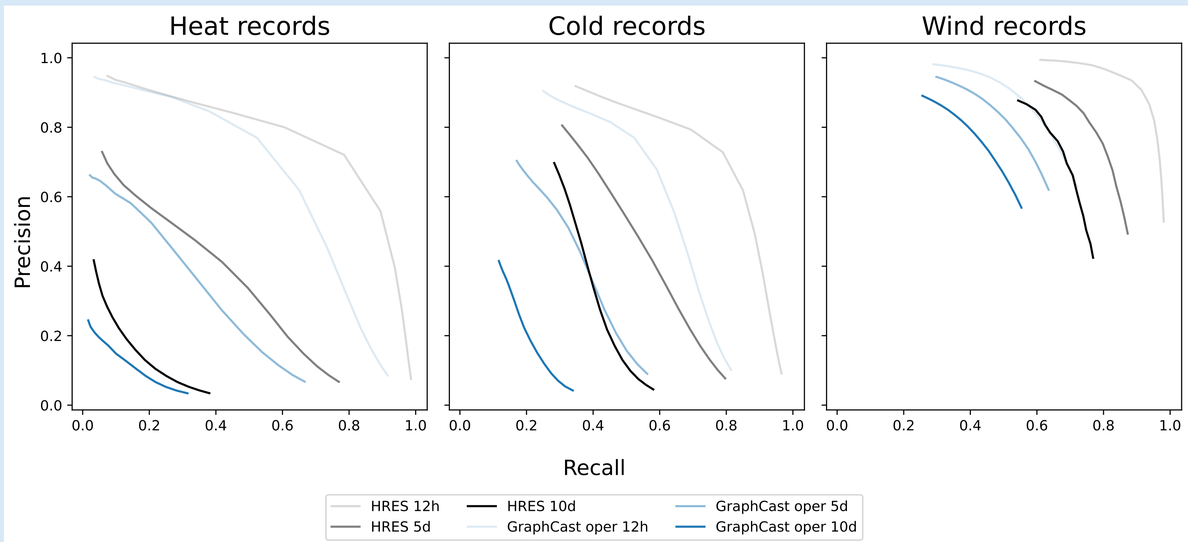
Operational forecast

RMSE on record data set over land (Antarctic excluded) in 2020
latitude-weighted; record-breaking events selected from HRES-fc0; all forecasts compared against HRES-fc0





Precision-recall curve over land in 2020



Discussion

- Extrapolation is known to be challenging for neural networks that underlie all AI weather models
- Our results show that state-of-the-art AI models still have limitations for forecasting unprecedented events
- The numerical weather model HRES systematically outperforms AI models on record-breaking events
- Hybrid models that incorporate physical knowledge into AI models might extrapolate better

Reference

- Pasche, O. C., Wider, J., Zhang, Z., Zscheischler, J. & Engelke, S. Validating deep learning weather forecast models on recent high-impact extreme events. Artif. Intell. for Earth Syst. 4, e240033 (2025).