Can Artificial Intelligence Global Weather Forecasting Models Capture Extreme Events?

A Case Study of the 2022 Pakistan Floods

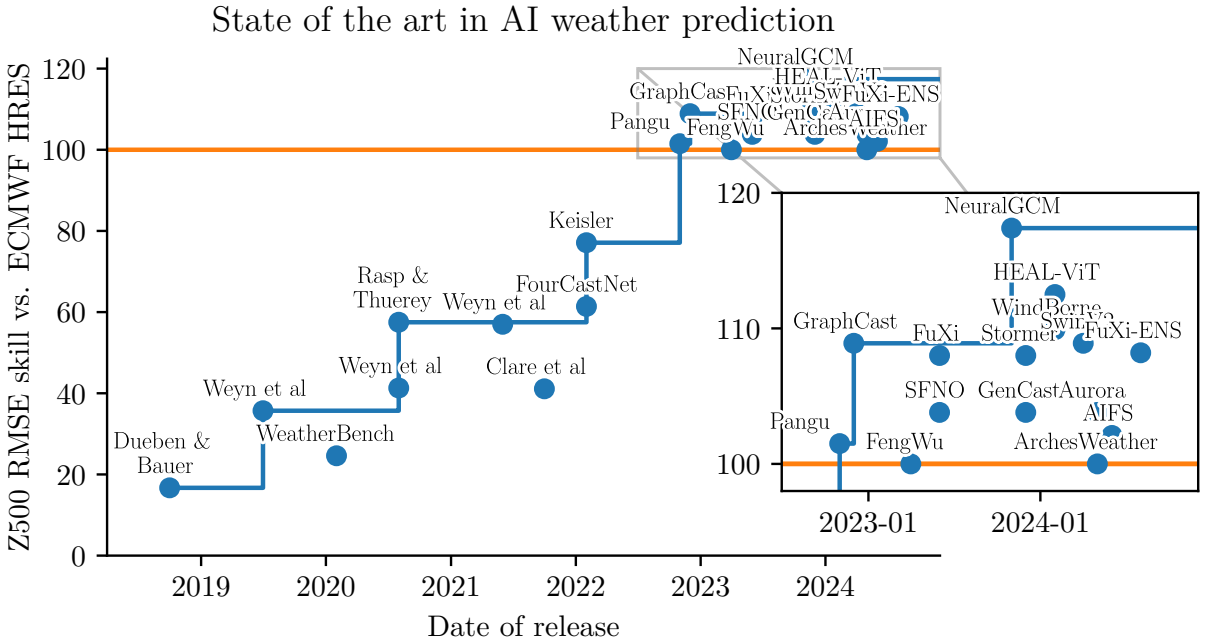
NeurIPS CCAI Workshop 2025

R. Almeida[1] N. Otero[1] M. Fernández-Torres[2] J. Ma[1]

[1]Applied Machine Learning Group
Fraunhofer Heinrich-Hertz Institute
[2]Department of Signal Theory and Communications
Universidad Carlos III de Madrid

AI weather prediction (AIWP) is state of the art*

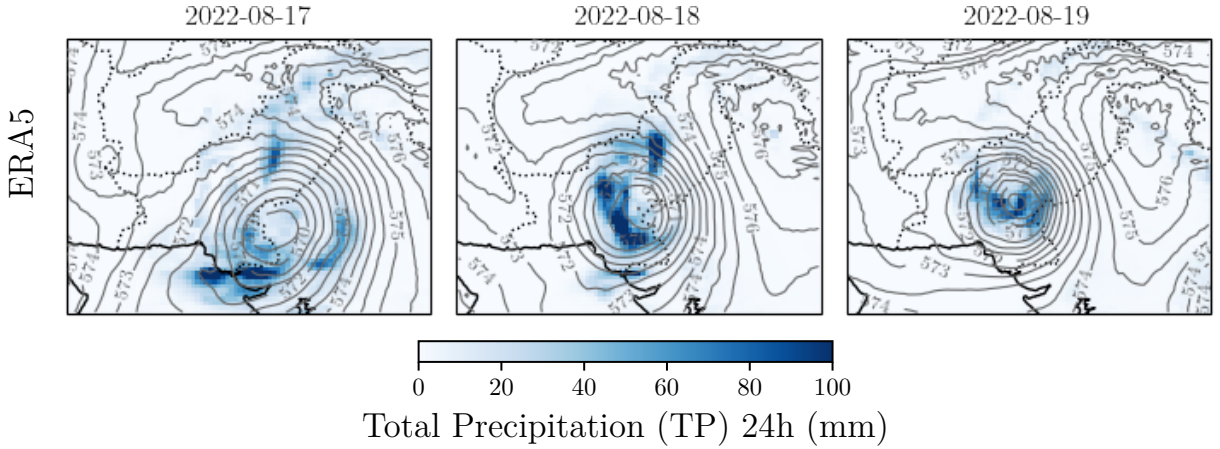
- Most benchmarks are deterministic
- Limits uncertainty quantification (UQ) and extremes performance assessment
- Probabilistic forecasts are critical for decision makers
- **Goal**: Test different perturbations (IC) + deterministic AIWP models on extremes



Adapted from [Hoyer and Rasp, 2025]

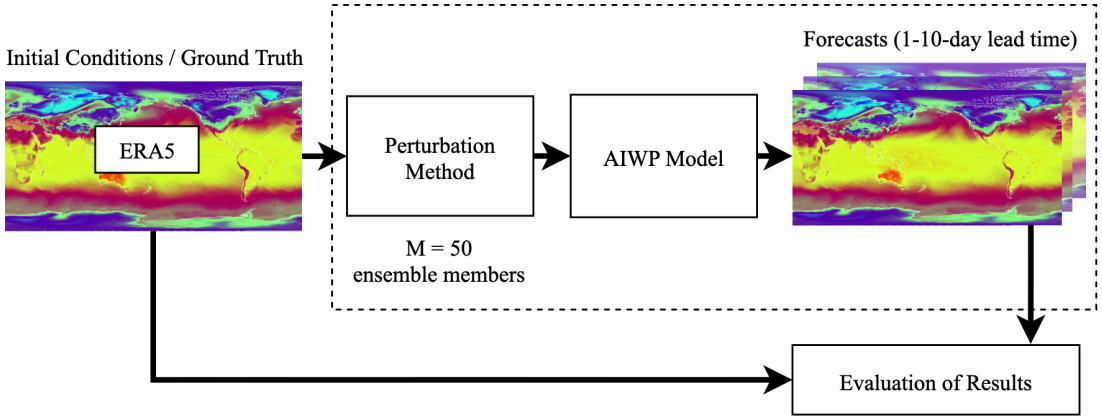
Record breaking floods in Pakistan

- Extreme monsoon rains in August 2022
- Flooding main driver was extreme precipitation [Cui et al., 2025]
- Thousands of deaths, 30M people displaced [Hong et al., 2023]



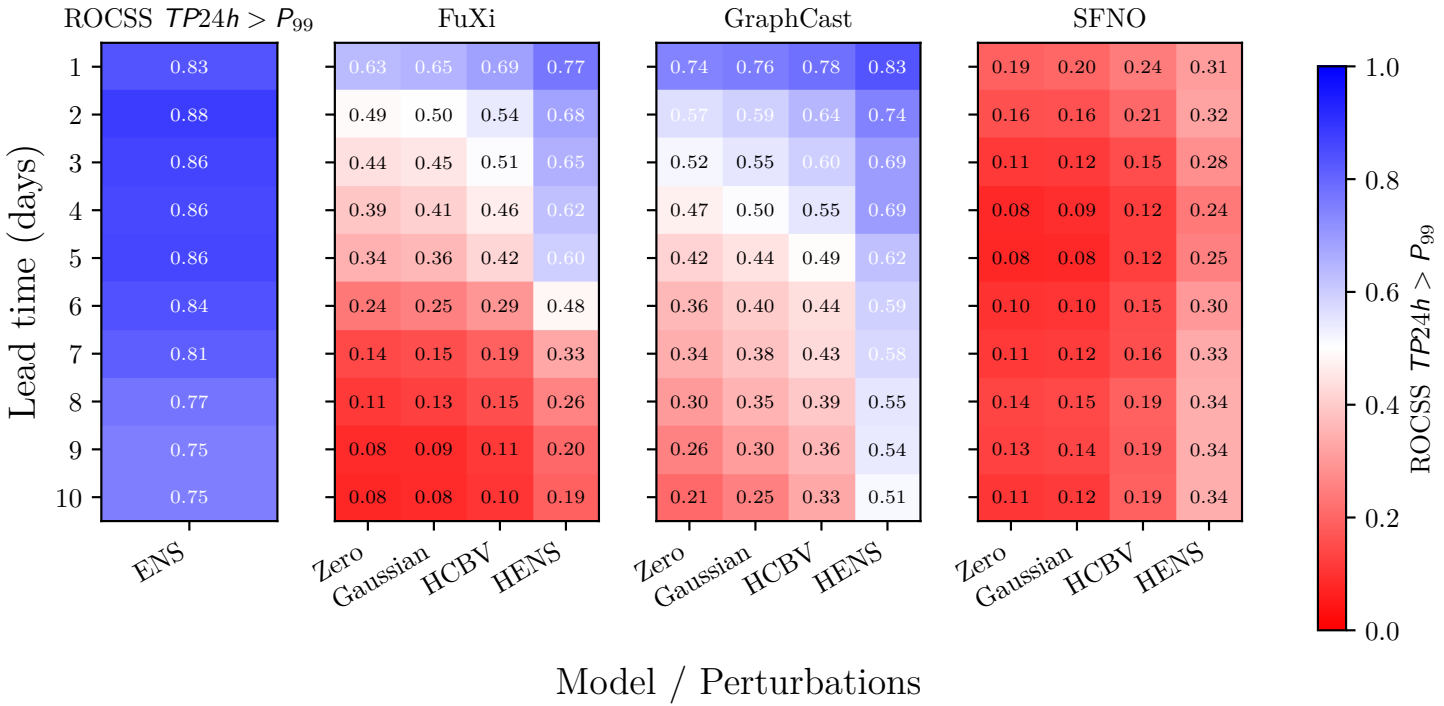
Perturb ⇒ Forecast ⇒ Evaluate

Methodology



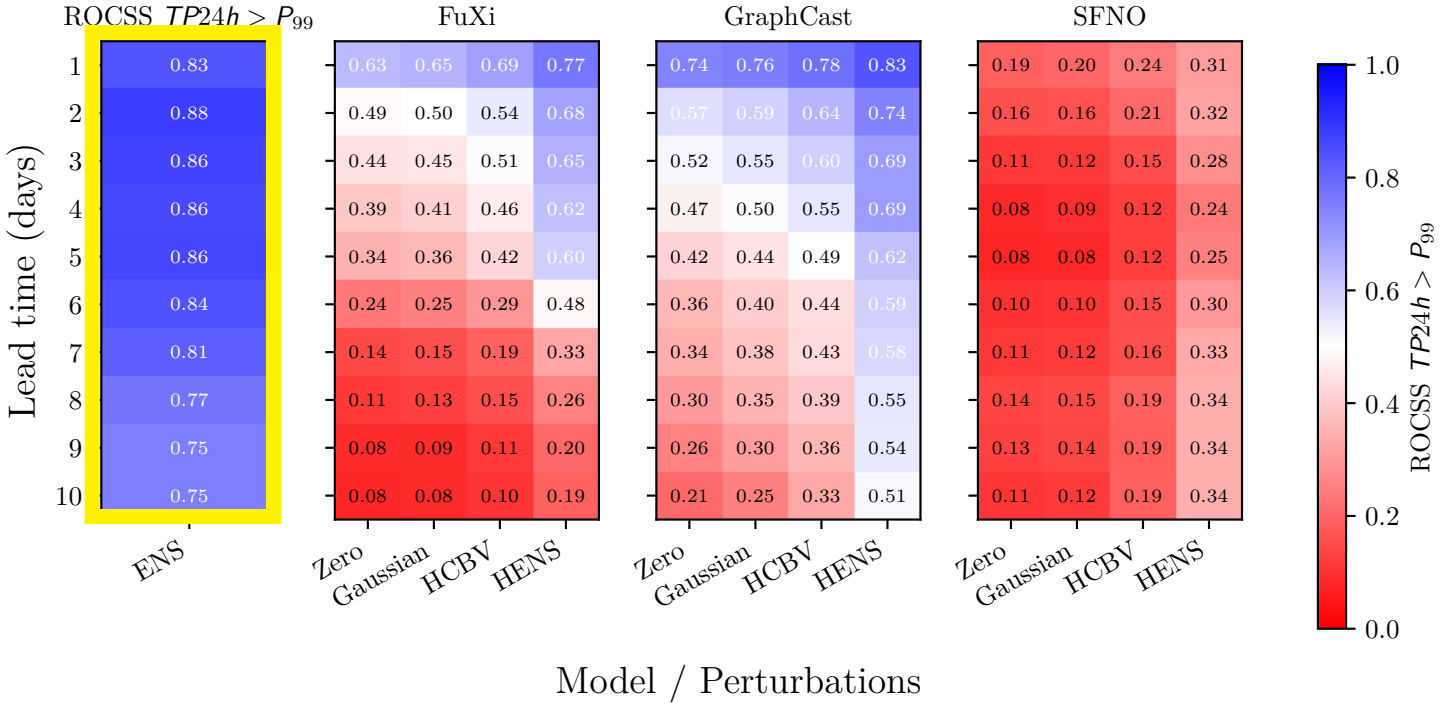
- **Data:** ERA5 [C3S, 2018] and IFS ENS [ECMWF, 2016]
- **Models:** FuXi [Chen et al., 2023], GraphCast [Lam et al., 2023] and SFNO [Bonev et al., 2023]
- **Perturbations:** Gaussian, HCBV, HENS
- **Metrics:** ROCSS, CRPS, RMSE
- **Tools:** `earth2studio` (forecasting, [Geneva and Foster, 2024]), `weatherbenchx` (evaluation, [Goo, 2025, Zhao et al., 2025])

ENS wins, GraphCast + HENS runnerup



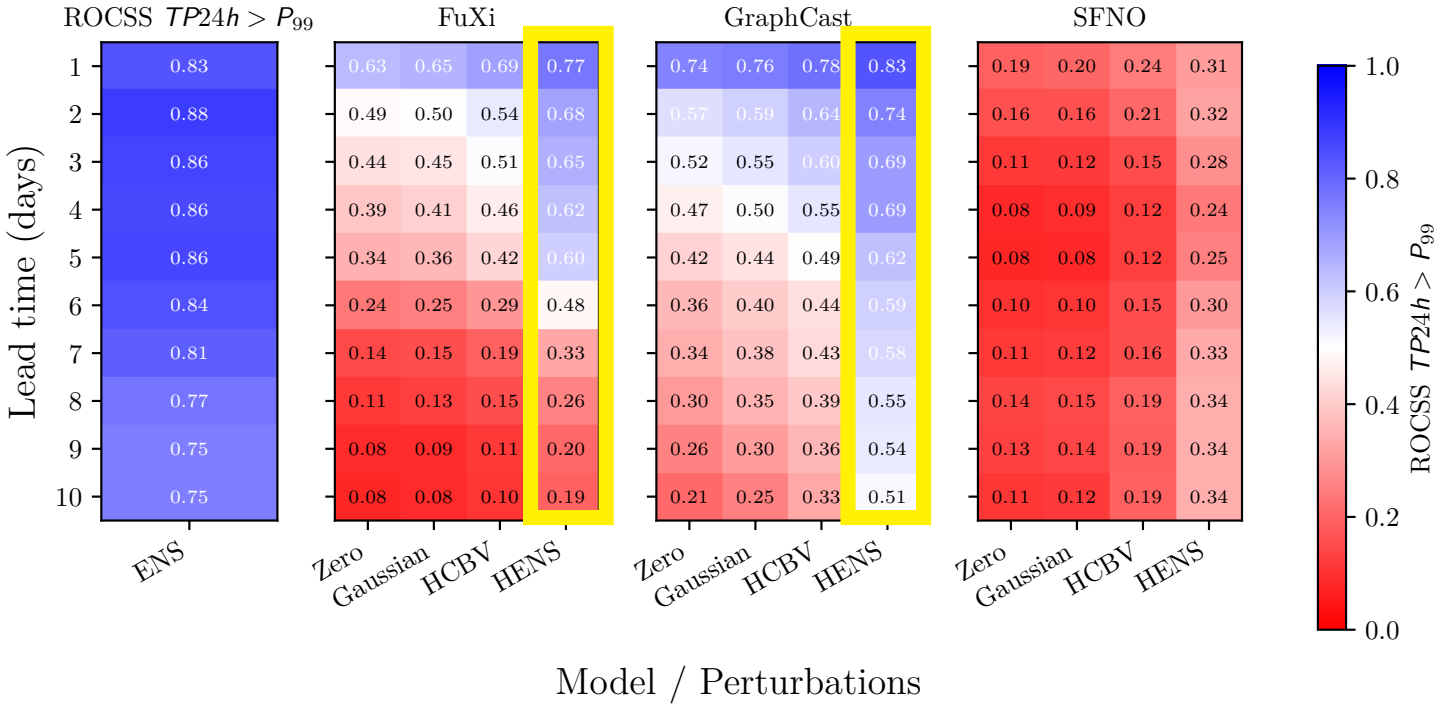
Pakistan region, August 2022.

ENS wins, GraphCast + HENS runnerup



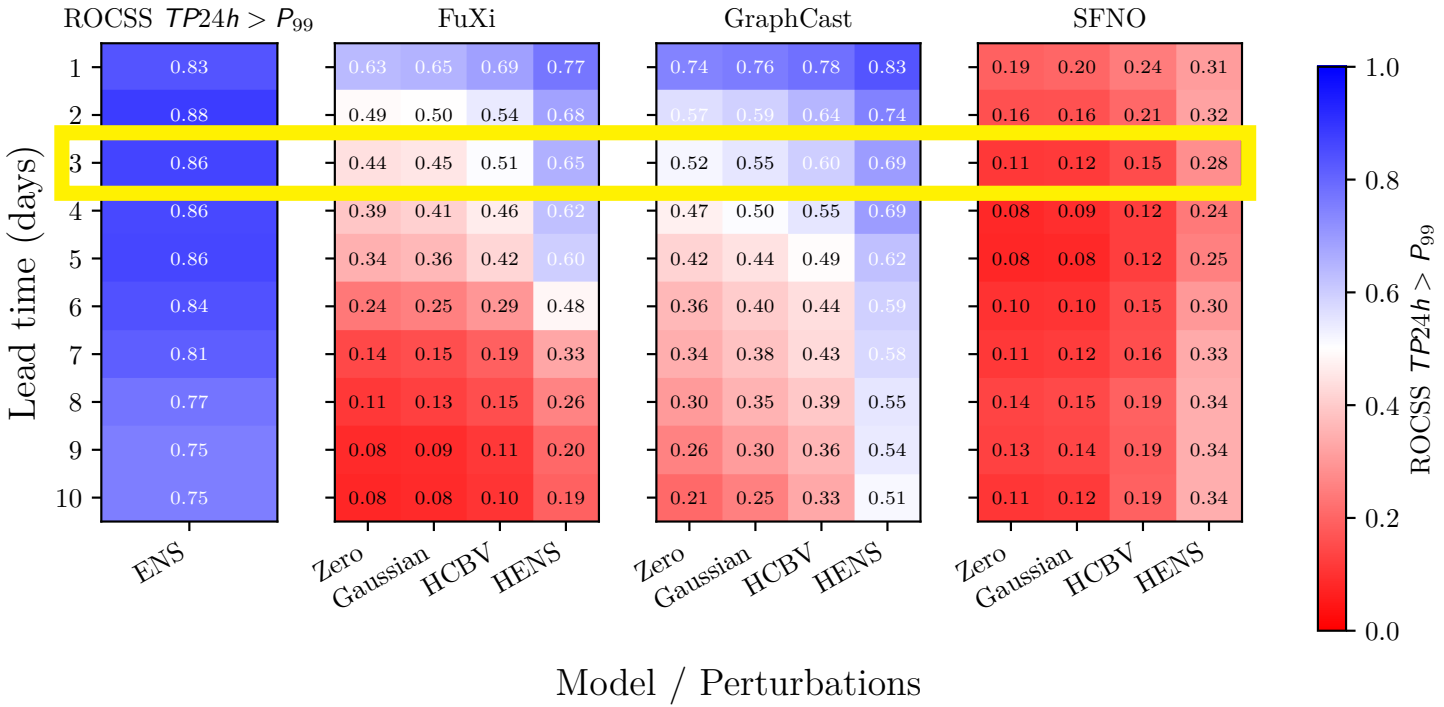
Pakistan region, August 2022.

ENS wins, GraphCast + HENS runnerup



Pakistan region, August 2022.

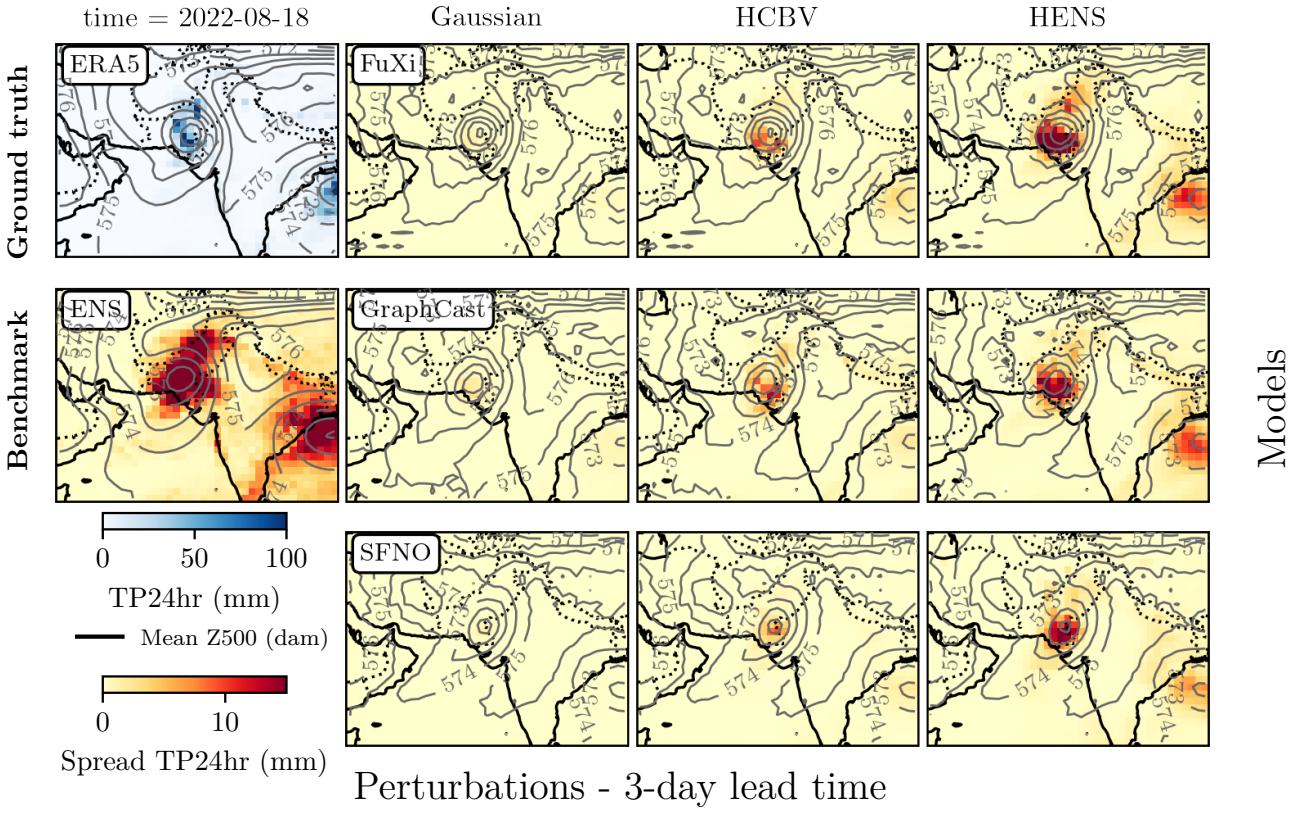
ENS wins, GraphCast + HENS runnerup



Pakistan region, August 2022.

FuXi + HENS is closest to ENS spread

- AIWP does not reproduce ENS spread
- Gaussian is underdispersive

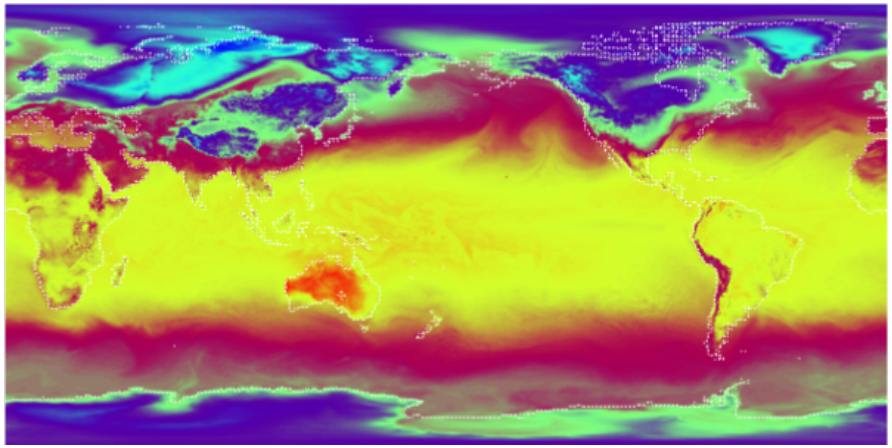


Conclusion & Outlook

- Input perturbations (HENS) improve forecasts but still below ENS for precipitation extremes, in this case
- Deterministic performance is maintained, but to different degrees for the perturbations
- Extremes remain a key challenge
- Future work, compare approaches for extremes (also warm extremes):
  - Multi-model ensembles
  - Checkpoint ensembles
  - Probabilistic AIWP
  - Explainability methods

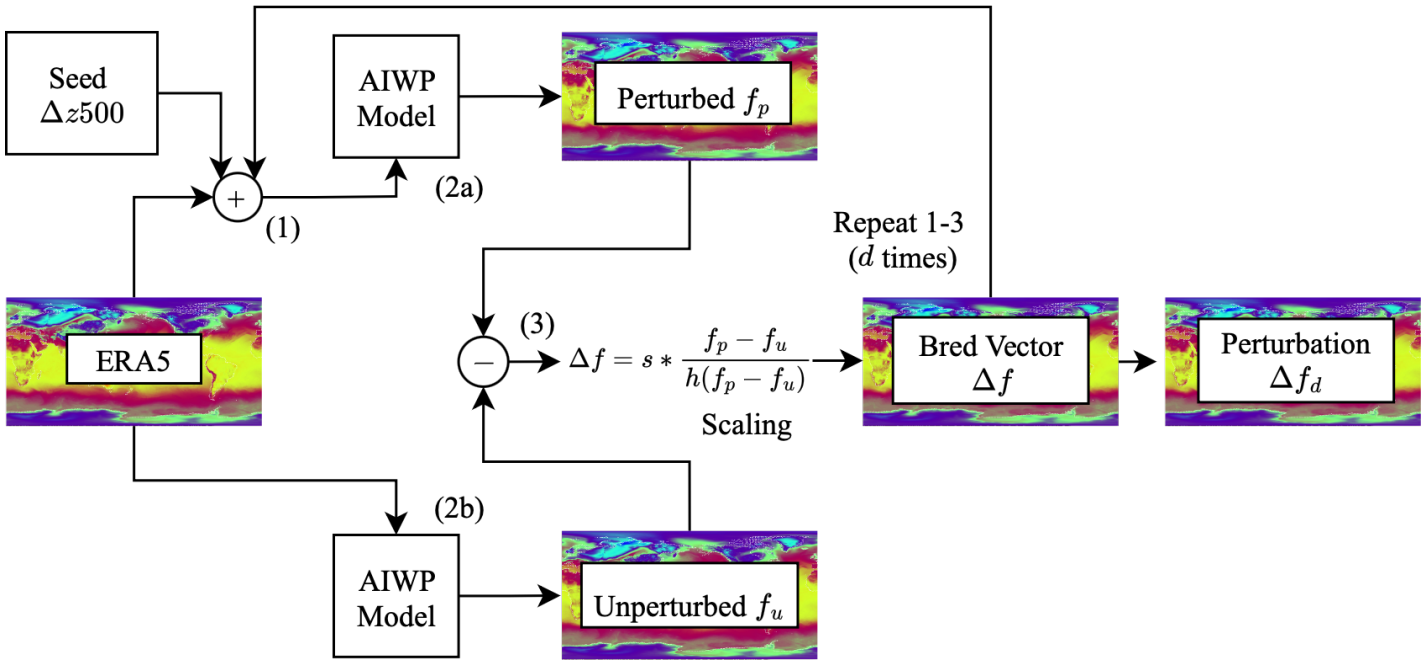
Thank you!

rodrigo.almeida@hhi.fraunhofer.de
rodrigoalmeida94
rodrigo-gtiff.bsky.social

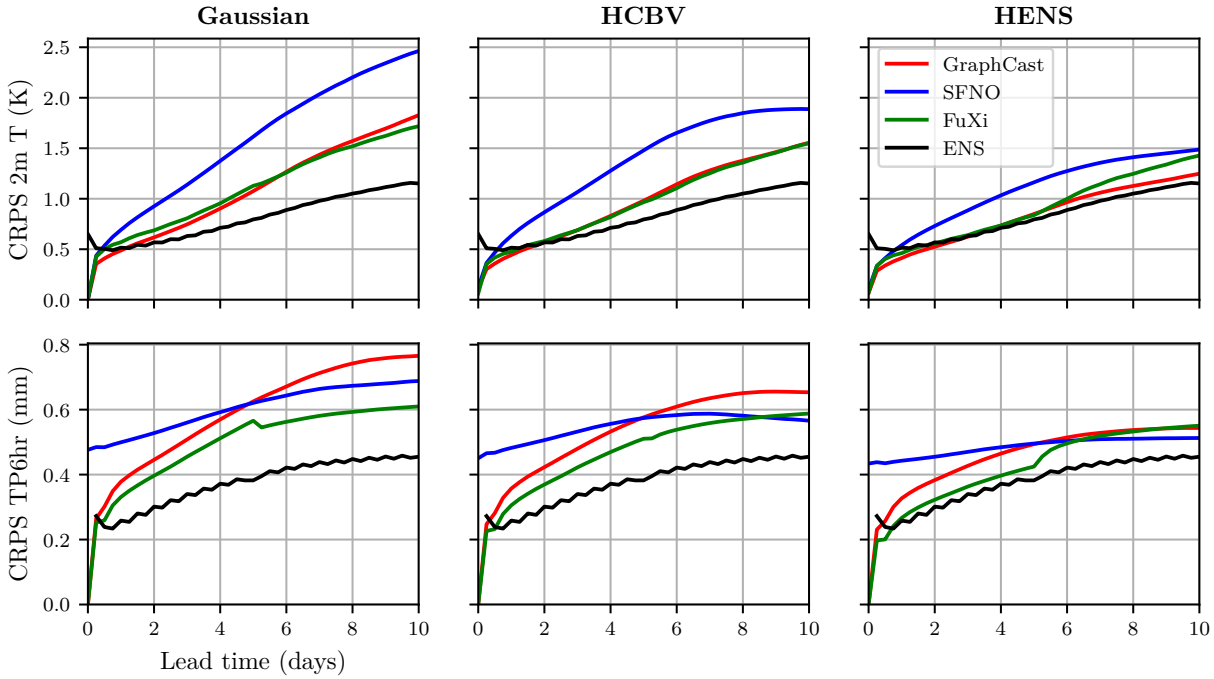


Perturbations

Perturbations: Gaussian, *HCBV, HENS* $\Downarrow$ *s* is different



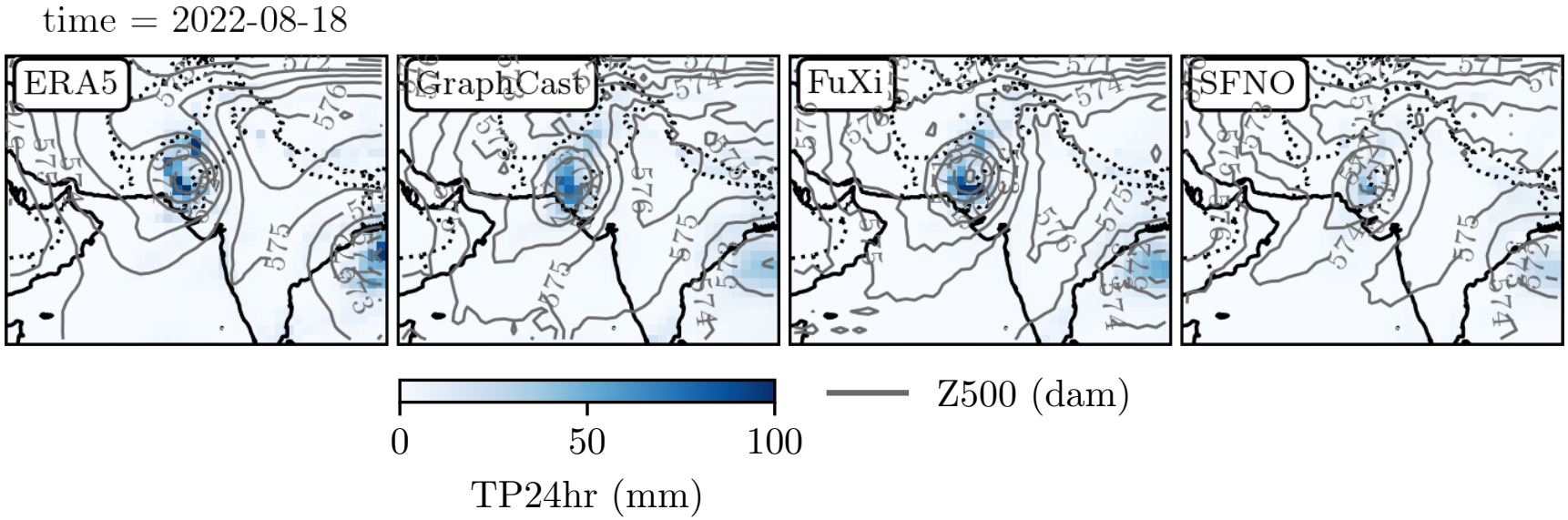
HENS narrows the gap to ENS



Global, August 2022

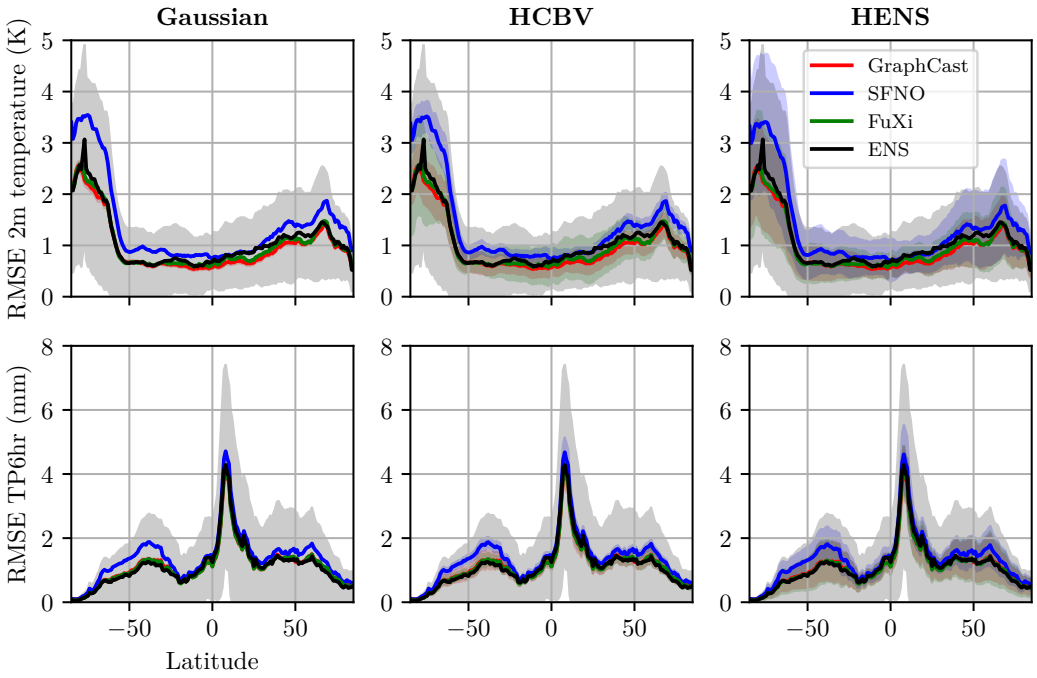
Fraunhofer HHI

Deterministic AIWP forecasts



3 day lead time, August 2022

Global metrics (by latitude)



3 day lead time, August 2022

References I

[Goo, 2025] (2025).
Google-research/weatherbenchX.
Google Research.

[Bonev et al., 2023] Bonev, B., Kurth, T., Hundt, C., Pathak, J., Baust, M., Kashinath, K., and Anandkumar, A. (2023).
Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere.

[C3S, 2018] C3S (2018).
ERA5 hourly data on single levels from 1940 to present.

[Chen et al., 2023] Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., and Li, H. (2023).
FuXi: A cascade machine learning forecasting system for 15-day global weather forecast.

References II

[Cui et al., 2025] Cui, P., Bazai, N. A., Qiang, Z., Jiao, W., Yan, W., Xu, Q., Yu, L., and Bo, Z. (2025).
Flood risk assessment with machine learning: Insights from the 2022 Pakistan mega-flood and climate adaptation strategies.
npj Natural Hazards, 2(1):42.

[ECMWF, 2016] ECMWF (2016).
ECMWF IFS High-Resolution Operational Forecasts.

[Geneva and Foster, 2024] Geneva, N. and Foster, D. (2024).
NVIDIA Earth2Studio.

[Hong et al., 2023] Hong, C.-C., Huang, A.-Y., Hsu, H.-H., Tseng, W.-L., Lu, M.-M., and Chang, C.-C. (2023).
Causes of 2022 Pakistan flooding and its linkage with China and Europe heatwaves.
npj Climate and Atmospheric Science, 6(1):1–10.

References III

[Hoyer and Rasp, 2025] Hoyer, S. and Rasp, S. (2025).
Ai-weather-sota-vs-time.ipynb.
https://gist.github.com/shoyer/f77647f3cc7231ae66c5c63b95bfc4a2.

[Lam et al., 2023] Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P. (2023).
GraphCast: Learning skillful medium-range global weather forecasting.
Science.

[Zhao et al., 2025] Zhao, T., Li, Q., Tu, T., and Chen, X. (2025).
An extension of the WeatherBench 2 to binary hydroclimatic forecasts.