

Can Artificial Intelligence Global Weather Forecasting Models Capture Extreme Events? A Case Study of the 2022 Pakistan Floods

Rodrigo Almeida¹ Noelia Otero¹ Miguel-Ángel Fernández-Torres² Jackie Ma¹



¹Applied Machine Learning Group, Fraunhofer Heinrich-Hertz Institute, Berlin, Germany, rodrigo.almeida@hhi.fraunhofer.de

²Department of Signal Theory and Communications, Universidad Carlos III de Madrid (UC3M), Madrid, Spain

Key Points

- Deterministic weather prediction on the 2022 Pakistan floods.
- Global AI models: **FuXi** [1], **GraphCast** [2], **SFNO** [3] perturbed with **Gaussian**, Hemi-Centric Bred Vectors (**HCBV**), Huge Ensembles (**HENS**) [4].
- **HENS**: Most skillful ensembles, improved precipitation forecasts.
- **ENS** outperforms all AI models, especially at long lead times.

Methodology

- **Data**: **ERA5** reanalysis initial conditions to initialize forecasts.
- **Perturbation Methods**: 50-member ensembles.
- **Benchmarks**: **ERA5** (ground truth), **ENS** (NWP reference).
- **Evaluation**: **Root Mean Squared Error (RMSE)** for deterministic skill, **Continuous Ranked Probability Score (CRPS)** for probabilistic accuracy, and **Receiver Operating Characteristic Skill Score (ROCSS)** for extreme-event discrimination.
- **Tools**: **earth2studio** (forecasting), **weatherbenchx** (evaluation).

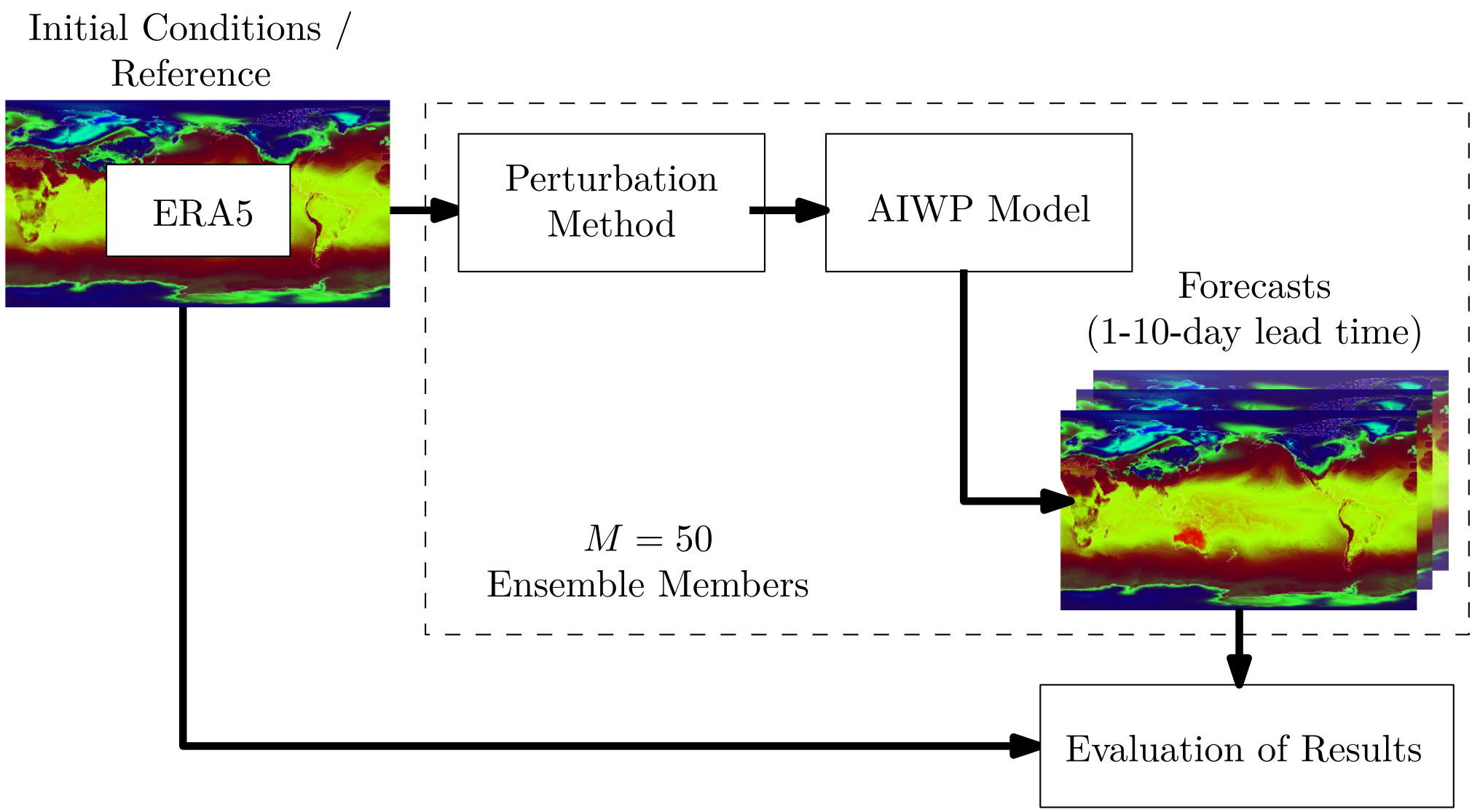


Figure 1. Overview of the methodology in this study.

August 2022 Extreme Precipitation in Pakistan

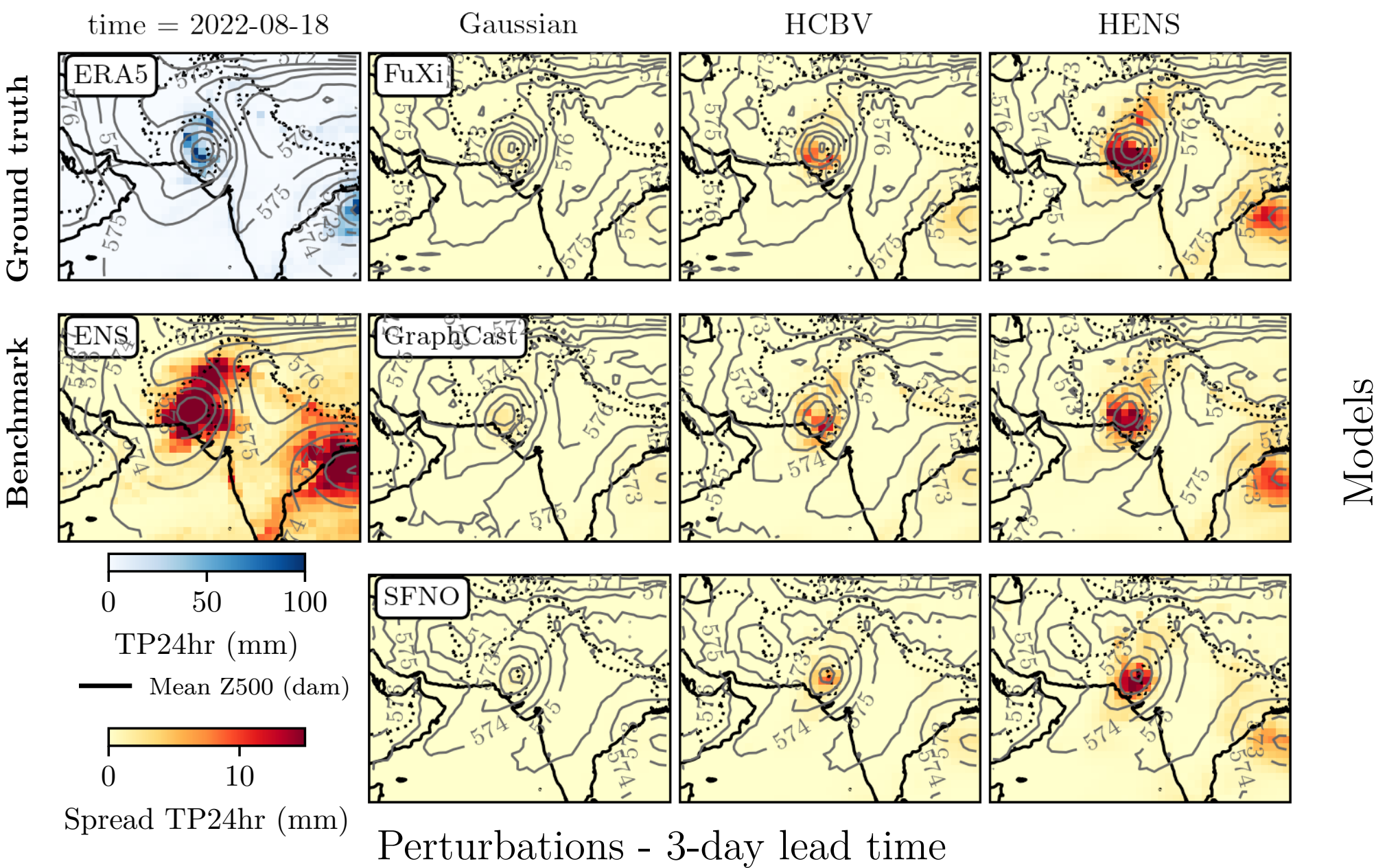


Figure 4. Daily accumulated precipitation spread for the different ensemble models (ENS, AIWPs) over Pakistan. Daily average geopotential height at 500 hPa is displayed as contour lines.

References

[1] Lei Chen et al. *FuXi: A Cascade Machine Learning Forecasting System for 15-Day Global Weather Forecast*. Oct. 2023.
[2] Remi Lam et al. "GraphCast: Learning Skillful Medium-Range Global Weather Forecasting". In: *Science* (Dec. 2023).
[3] Boris Bonev et al. *Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere*. June 2023.
[4] Ankur Mahesh et al. *Huge Ensembles Part I: Design of Ensemble Weather Forecasts Using Spherical Fourier Neural Operators*. Feb. 2025.

Motivation and Objectives

- Extreme weather more frequent and severe due to climate change.
- AI Weather Prediction (AIWP) models show strong deterministic skill but lack uncertainty quantification.
- Deterministic nature limits extreme-event prediction skill.
- Objective: Input perturbations to generate probabilistic AI forecasts.
- Focus: August 2022 Pakistan floods driven by extreme precipitation.

Quantitative Results

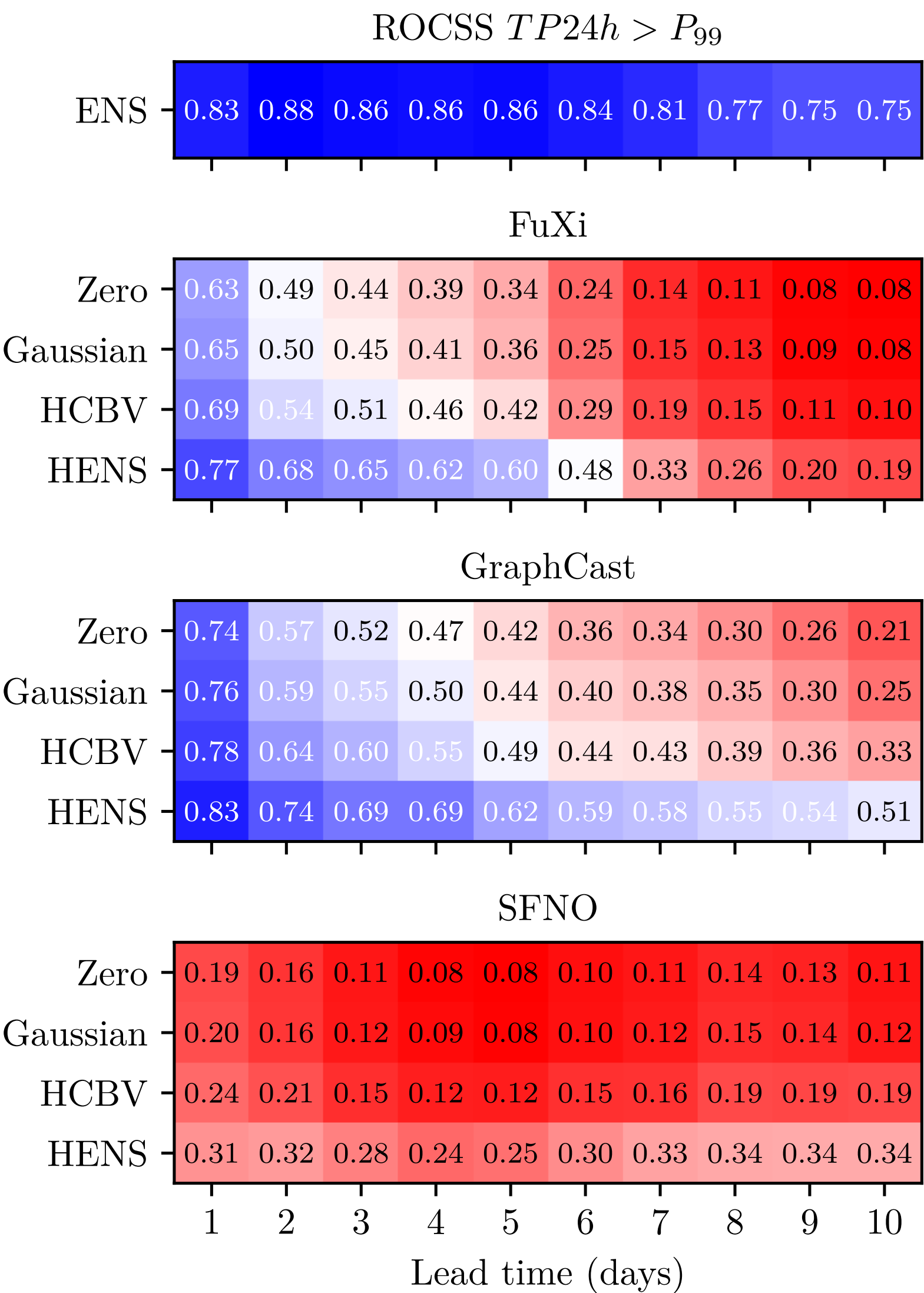


Figure 2. ROCSS for the 99th percentile of the ERA5 1990-2020 climatology for daily precipitation, in the Pakistan region in August 2022. Higher ROCSS values mean better performance. HENS perturbations improve the detection of extreme precipitation, extending skill to longer lead times across all AI models.

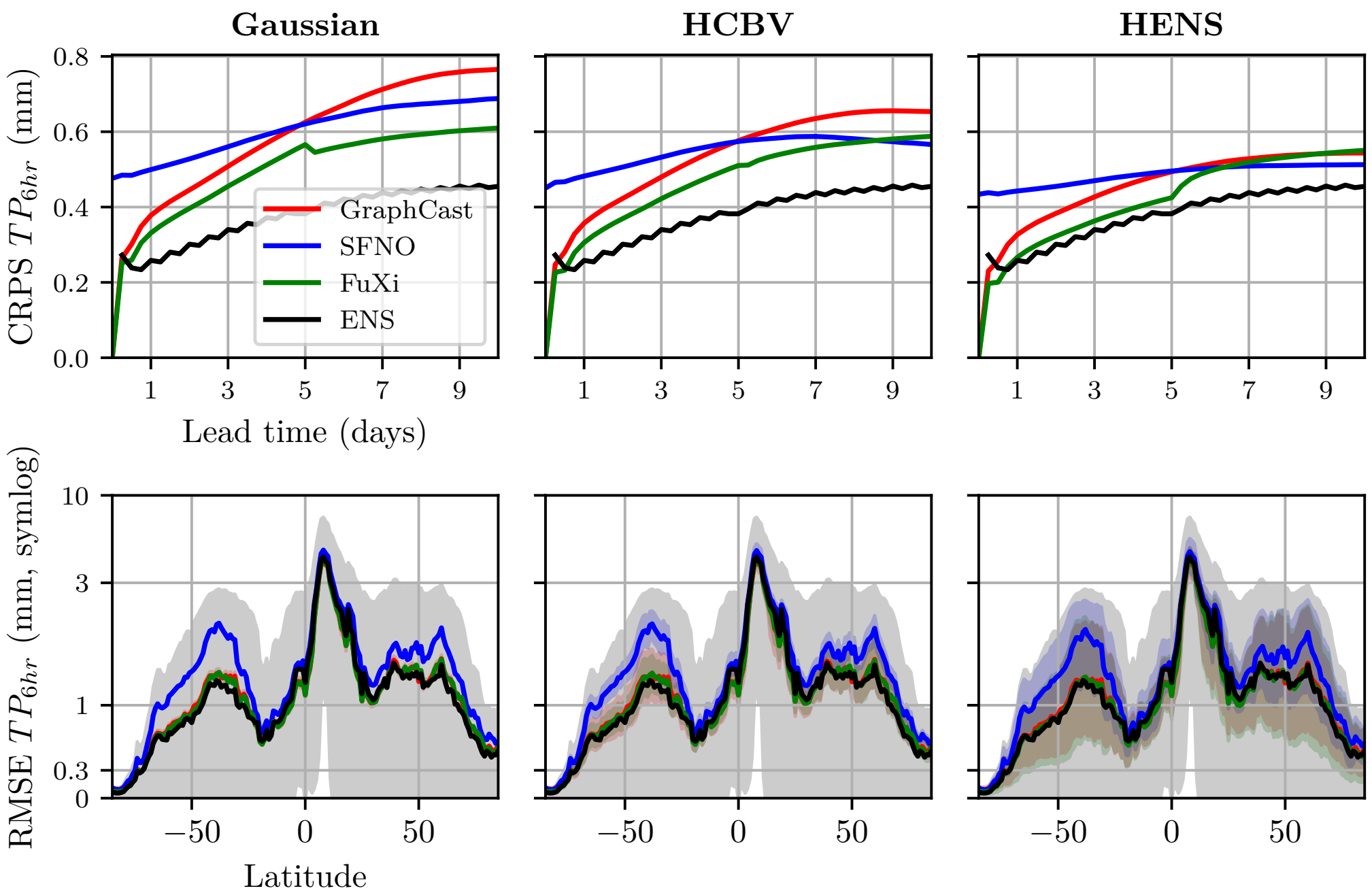


Figure 3. Global CRPS and RMSE across lead times and latitudes for August 2022 for 6 hourly precipitation. HENS ensembles for GraphCast achieve the lowest CRPS and most realistic RMSE spread. After 5 days lead time all models converge in performance with HENS.