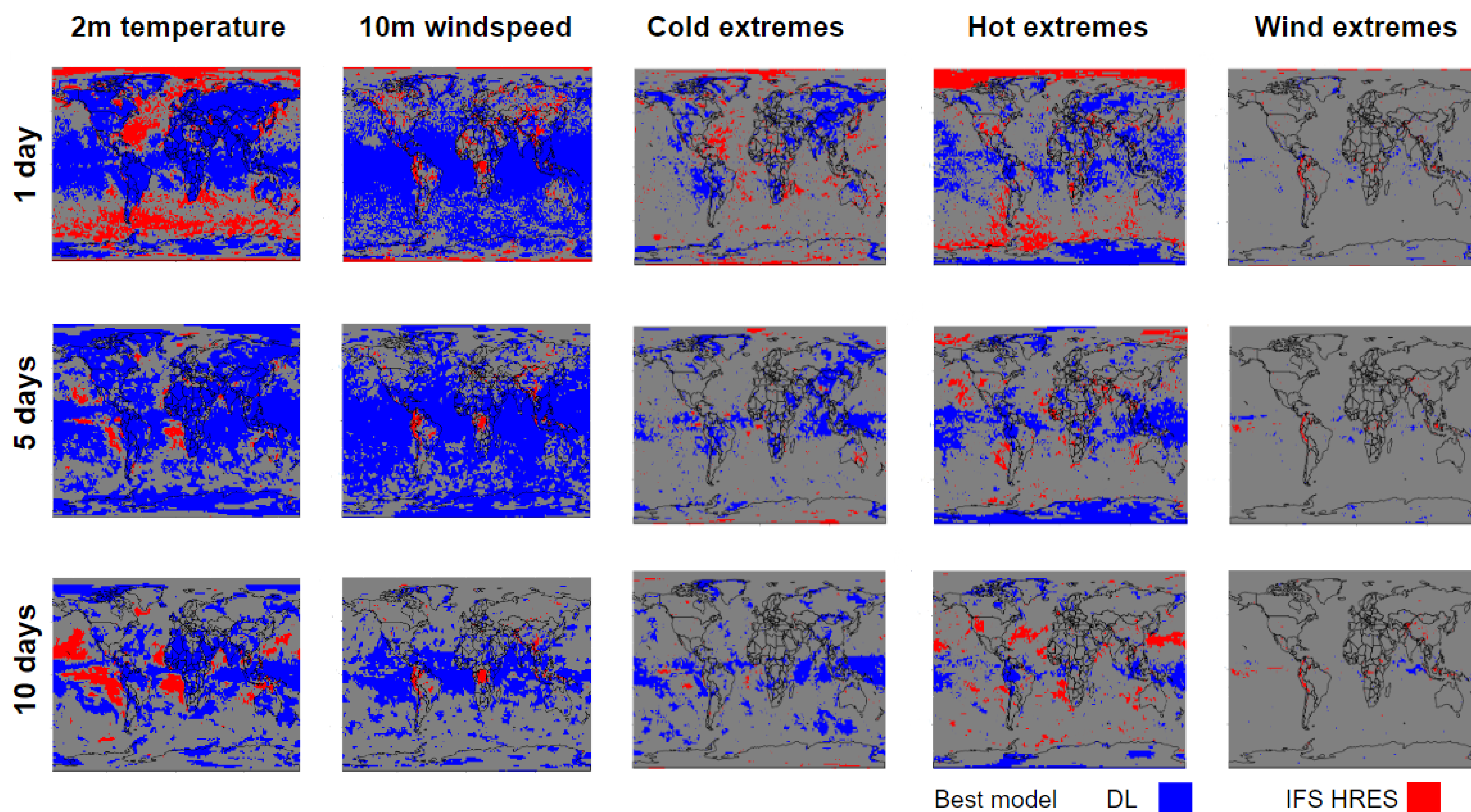


A comparison of data-driven and numerical weather forecasting models on weather extremes

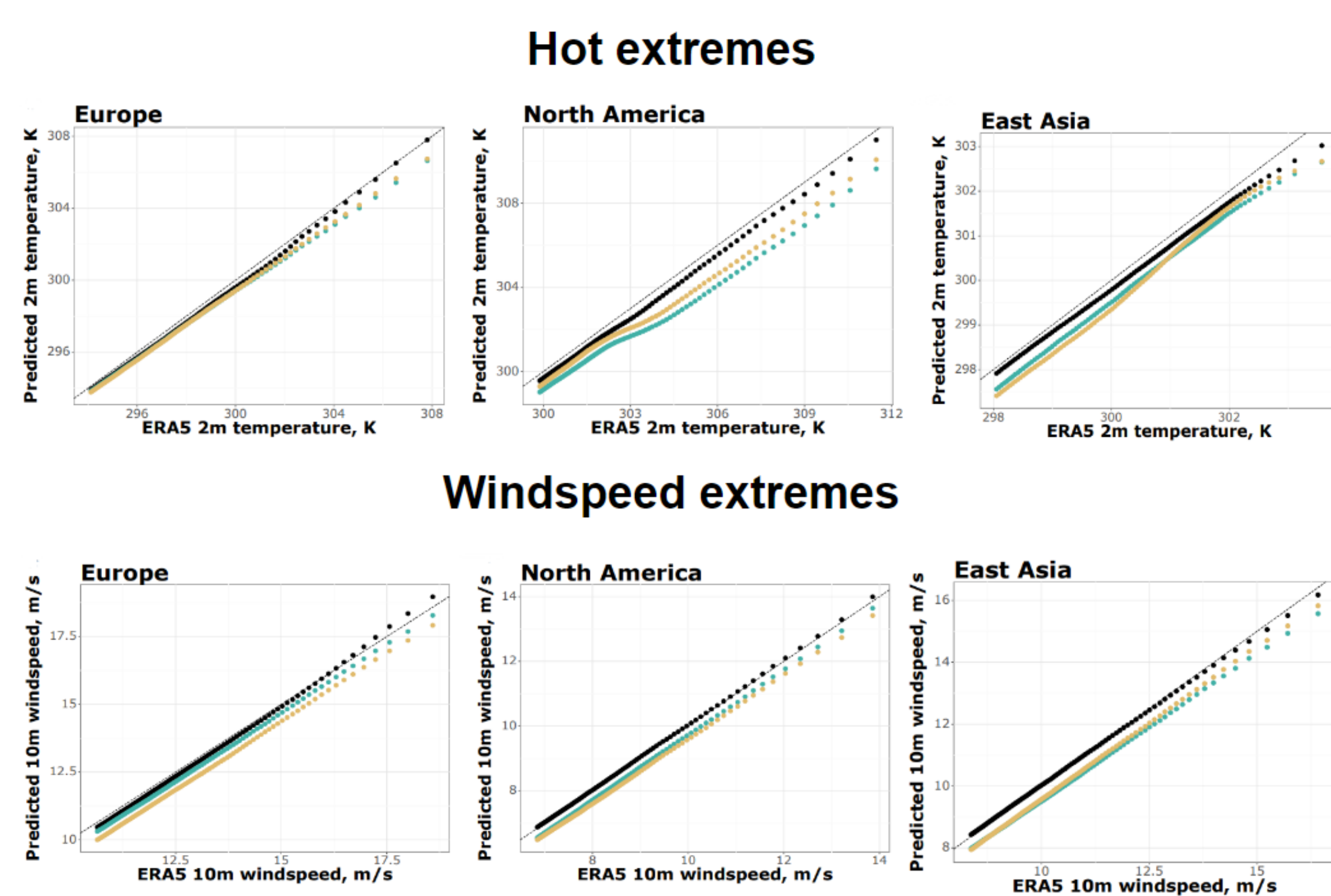
Grid-point level comparisons

- ▶ We evaluate the performance of Pangu-Weather, GraphCast and IFS HRES on near-surface hot, cold and windy extremes globally using the operational forecasts for 2020 provided by the WeatherBench 2 (Rasp et al., 2024).
- ▶ The extremes are defined as the 5% most extreme events for 2020 at each grid-point, based on ERA 5.
- ▶ We compute the grid-point level RMSE based on all data-points and on extreme events only, using ERA5 as ground-truth. We compare IFS HRES to the best DL model.



Tail calibration

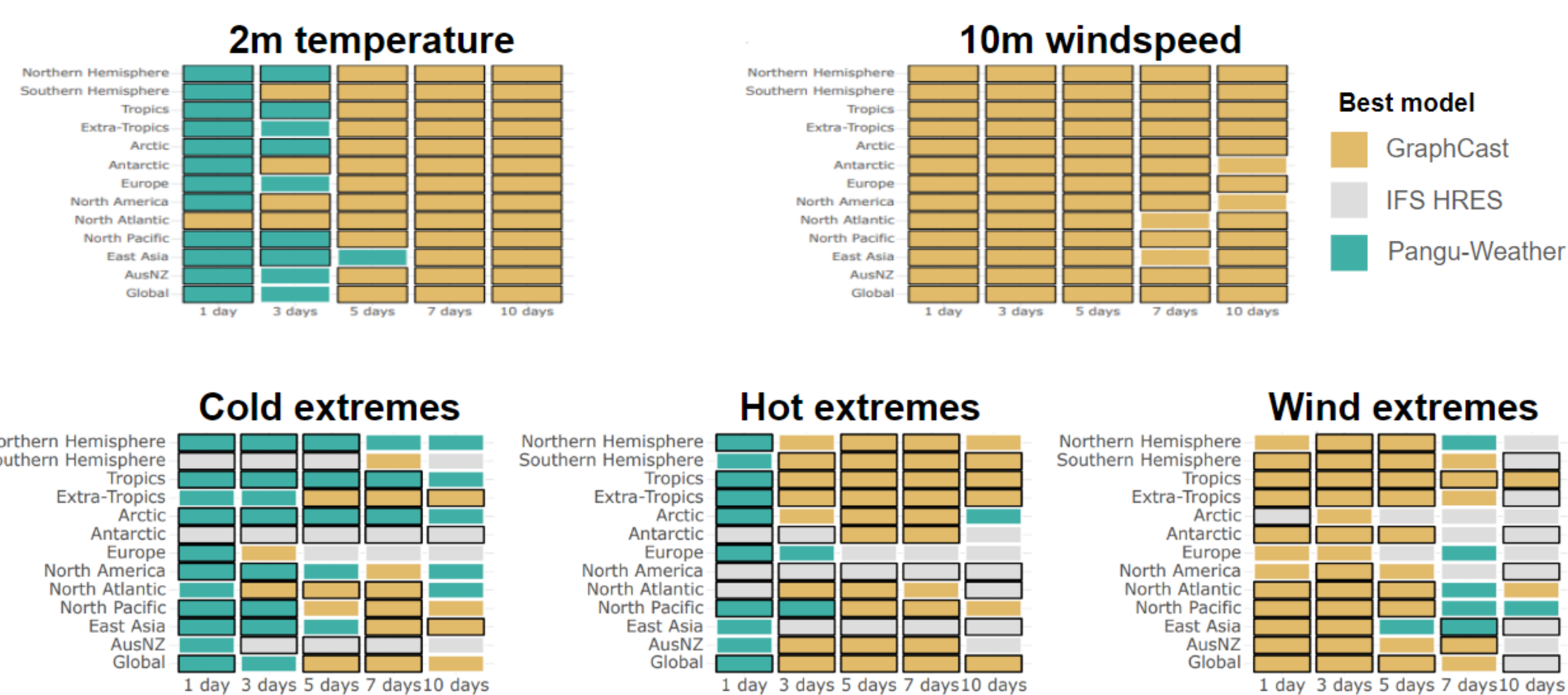
- ▶ We evaluate the tail calibration for regional and global extremes.



QQ-plot 10% most extreme events, 5-days forecast vs ERA5

Regional comparisons

- ▶ We repeat the analysis above at the regional level.



Best model in terms of RMSE. The extremes are defined as the 5% grid-point level most extreme events within the given region.

Main conclusions

Data-driven models perform best:

- ▶ for 1-3 days forecasts
- ▶ in the Tropics
- ▶ for temperature extremes
- ▶ on the west side of ocean basins

Data-driven models perform worse:

- ▶ for 7-10 days forecasts → due to blurring?
- ▶ at higher latitudes → due to use of latitude-based weights?
- ▶ for windspeed extremes → due to separate training of u-and v-wind?
- ▶ on the east side of ocean-basins and in the middle of vast land areas → due to lack of key input variables (e.g. soil moisture and SSTs)?

Leonardo Olivetti^{1,2,3}, Gabriele Messori^{1,2,3,4,5}

¹ Uppsala University; ² CNDS; ³ CLIMES; ⁴ Stockholm University; ⁵ Bolin Centre

Correspondence: Leonardo Olivetti, leonardo.olivetti@geo.uu.se



EGUsphere



GitHub



ERC grant no. 948309