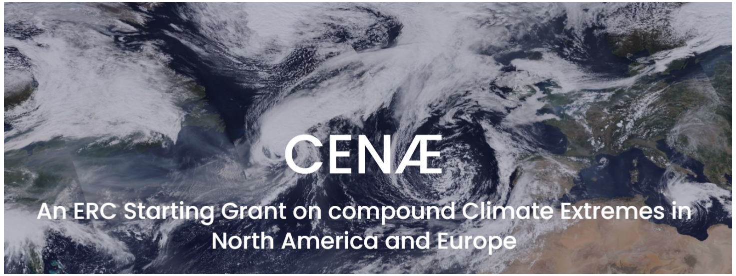
On data-driven forecasting of near-surface extreme weather events

Leonardo Olivetti[1], Gabriele Messori[1,2], ELLIIT seminar, 30/09



https://cenae-erc.eu/

ERC grant no. 948309

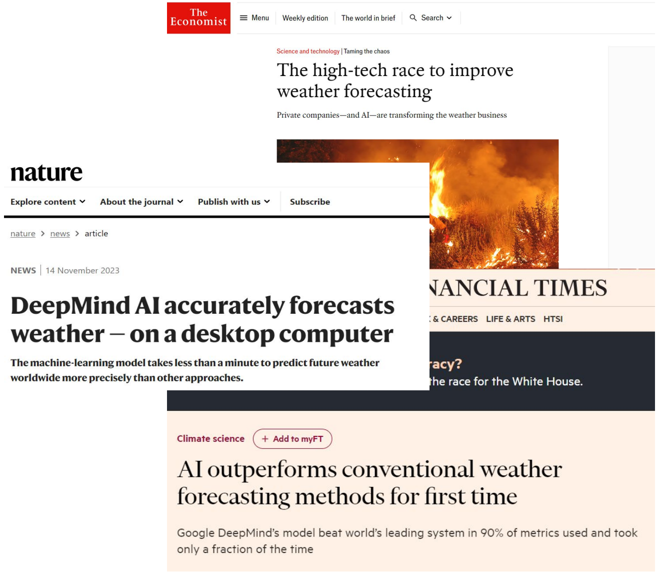
1 Department of Earth Sciences, CLIMES and CNDS, Uppsala University, Sweden
2 Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Sweden

Background

- Deep learning (DL) weather models can produce deterministic forecasts with skill comparable to SOTA physics-based, numerical models. (e.g. Bi et al., 2023, Lam et al., 2023, Chen et al., 2023, Lang et al., 2024)

- Many promising probabilistic approaches are on the horizon (e.g. Kochkov et al., 2024, Price et al., 2024, Oskarsson et al., 2024)

- ECMWF and NOAA have recently announced their intention to operationalise data-driven models.

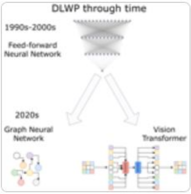


Review and perspective paper | Highlight paper |

21 Mar 2024

Advances and prospects of deep learning for medium-range extreme weather forecasting

Leonardo Olivetti and Gabriele Messori



Abstract

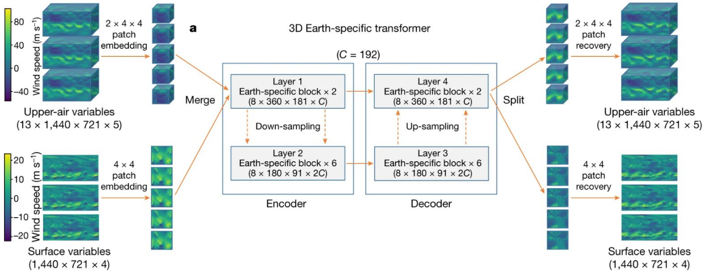
In recent years, deep learning models have rapidly emerged as a stand-alone alternative to physics-based numerical models for medium-range weather forecasting. Several independent research groups claim to have developed deep learning weather forecasts that outperform those from state-of-the-art physics-based models, and operational implementation of data-driven forecasts appears to be drawing near. However, questions remain about the capabilities of deep learning models with respect to providing robust forecasts of extreme weather. This paper provides an overview of recent developments in the field of deep learning weather forecasts and scrutinises the challenges that extreme weather events pose to leading deep learning models. Lastly, it argues for the need to tailor data-driven models to forecast extreme events and proposes a foundational workflow to develop such models.

How to cite. Olivetti, L. and Messori, G.: Advances and prospects of deep learning for medium-range extreme weather forecasting, Geosci. Model Dev., 17, 2347–2358, https://doi.org/10.5194/gmd-17-2347-2024, 2024.

Article in GMD

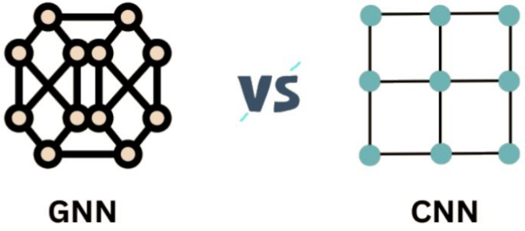
What sparked the recent developments?

Vision Transformer



Credits: Bi et al. (2023)

Graph Neural Network



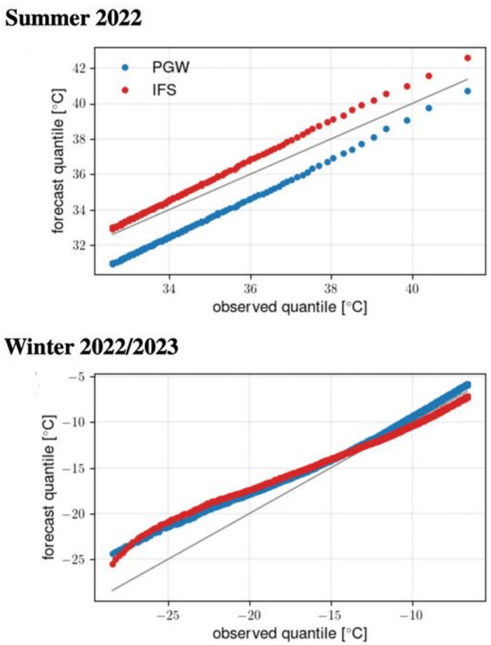
Credits: Reworking of an image by Pathak (2023), "Graph Neural Networks Explained in 5 Minutes".

- High-quality, high-resolution reanalysis data, such as ERA5 (Hersbach et al., 2020)
- Benchmark datasets, such as WeatherBench (Rasp et al., 2020)
- New DL architectures
- Big tech's increased interest for weather forecasting + open source approach

Do global DL models struggle with weather extremes?

Possible issues:

- Limited sample size and challenges related to extrapolation (Watson, 2022)
- Choice of loss function - e.g. global MAE/MSE (Xu et al., 2024)
- Global and temporal averaging across variables and time scales (Bonavita, 2024)



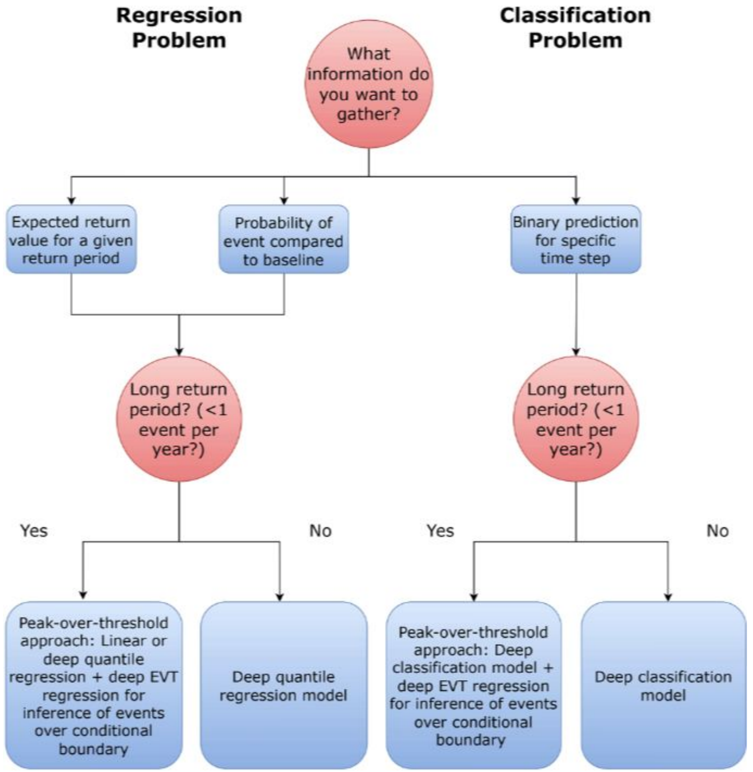
QQ-plot t2m 2020, 6-day forecast vs ERA 5 (Ben Bouallègue et al., 2024)

Possible solutions

- Ensemble approach for DL weather forecasting (e.g. Hu et al., 2023, Price et al., 2024, Kockhov et al., 2024)

- Regional/limited area modelling (e.g. Oskarsson et al., 2023, Nipen et al., 2024)

- Two-stage extreme value theory model



Credits: Olivetti and Messori, 2024

Do (deterministic) **data-driven models struggle** with weather **extremes in practice?**

10 Apr 2024

Do data-driven models beat numerical models in forecasting weather extremes? A comparison of IFS HRES, Pangu-Weather and GraphCast

Leonardo Olivetti ✉ and Gabriele Messori

Abstract. The last few years have witnessed the emergence of data-driven weather forecast models able to compete and in some respects outperform physics-based numerical models. However, recent studies question the capability of data-driven models to provide reliable forecasts of extreme events. Here, we aim to evaluate this claim by comparing the performance of leading data-driven models in a semi-operational setting, focusing on the prediction of near-surface temperature and windspeed extremes globally. We find that data-driven models outperform ECMWF's physics-based deterministic model in the average prediction of 10 m windspeed and 2 m temperature, and can also compete with the physics-based model in terms of extremes in most regions. However, the choice of best model depends strongly on region, type of extreme and sometimes even lead time. Thus, we conclude that data-driven models may already now be a useful complement to physics-based forecasts in those regions where they display superior tail performance, but that some challenges still need to be overcome before widespread operational implementation can take place.

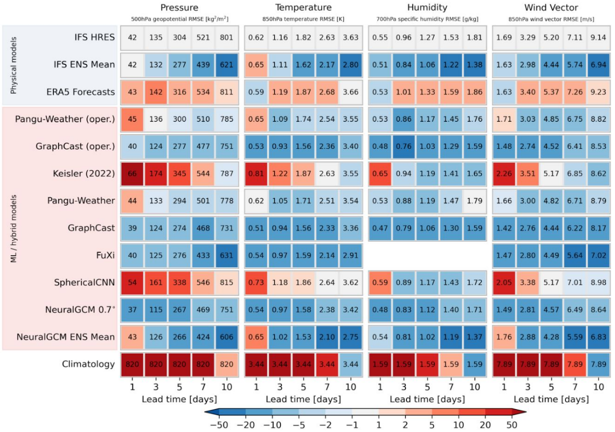
How to cite. Olivetti, L. and Messori, G.: Do data-driven models beat numerical models in forecasting weather extremes? A comparison of IFS HRES, Pangu-Weather and GraphCast, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-1042, 2024.

Initial EGUSphere preprint

Updated preprint accepted by GMD (Paper II in the folder)

Setup

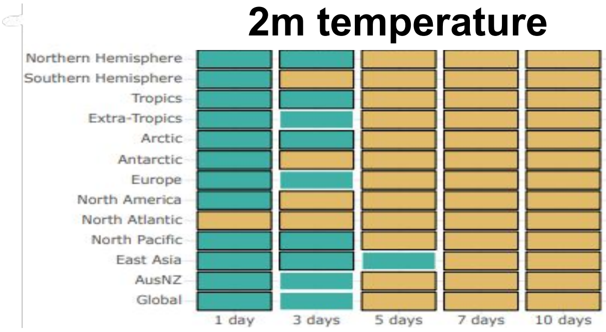
- Deterministic models only, taking IFS HRES t=0 as input (analysis, not reanalysis).
- We compare models in terms of overall and tail forecast RMSE at the global, regional and grid-point level.
- We also assess global and regional tail forecast calibration.



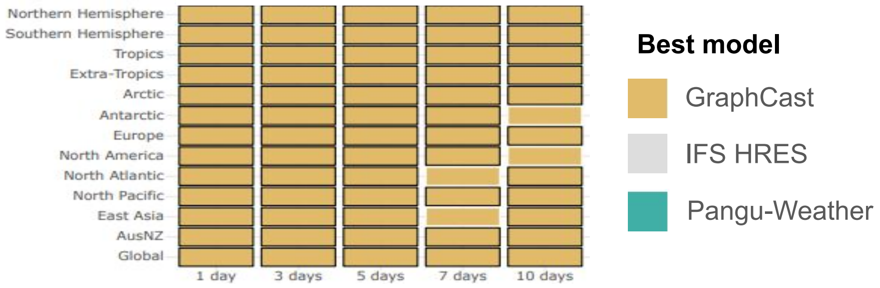
Weatherbench 2, Rasp et al., 2024

Can data-driven models compete with physics-based models in terms of global and regional RMSE?

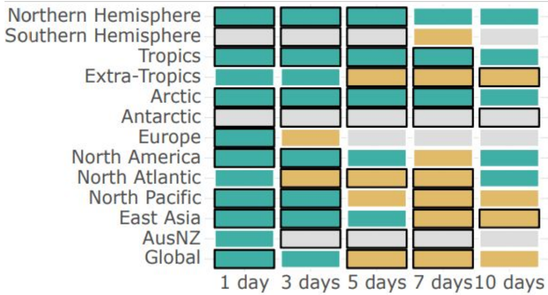
2m temperature



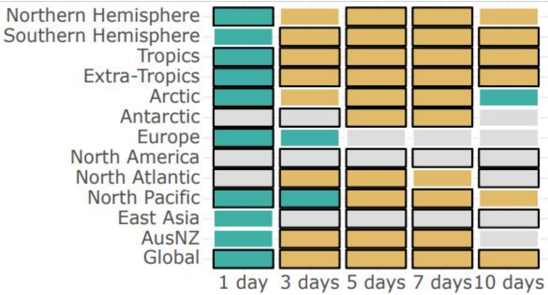
10m windspeed



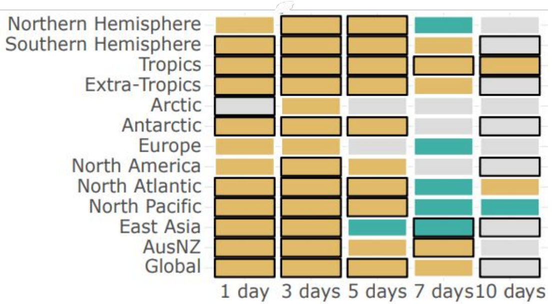
Cold extremes (5% coldest)

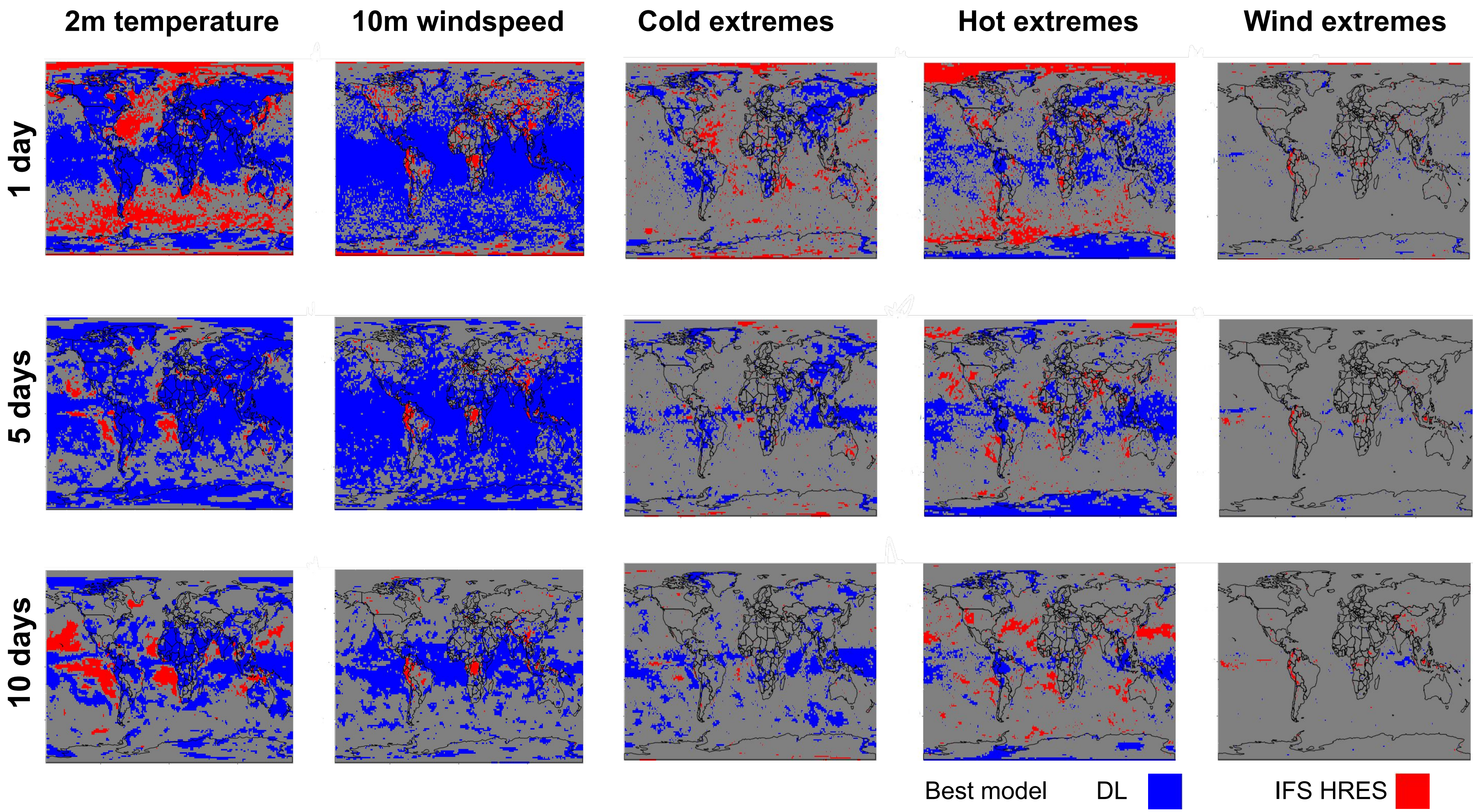


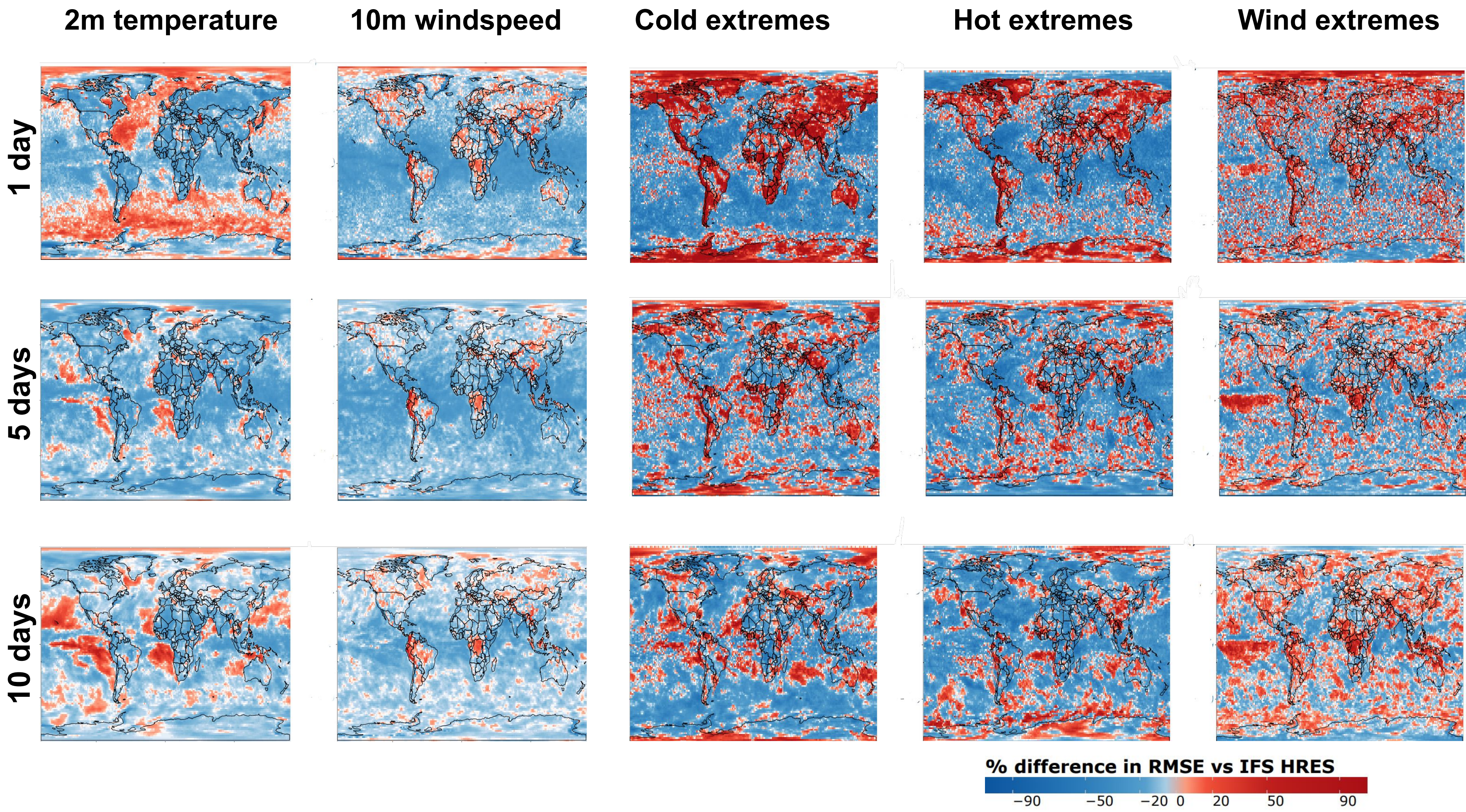
Hot extremes (5% hottest)



Wind extremes (5% windiest)



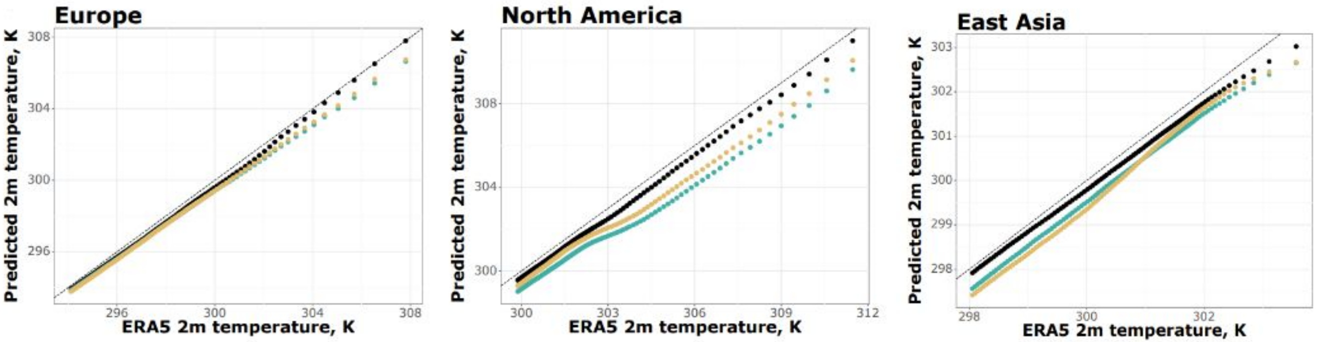

2m temperature
10m windspeed
Cold extremes
Hot extremes
Wind extremes
1 day
5 days
10 days
Best model
DL
IFS HRES



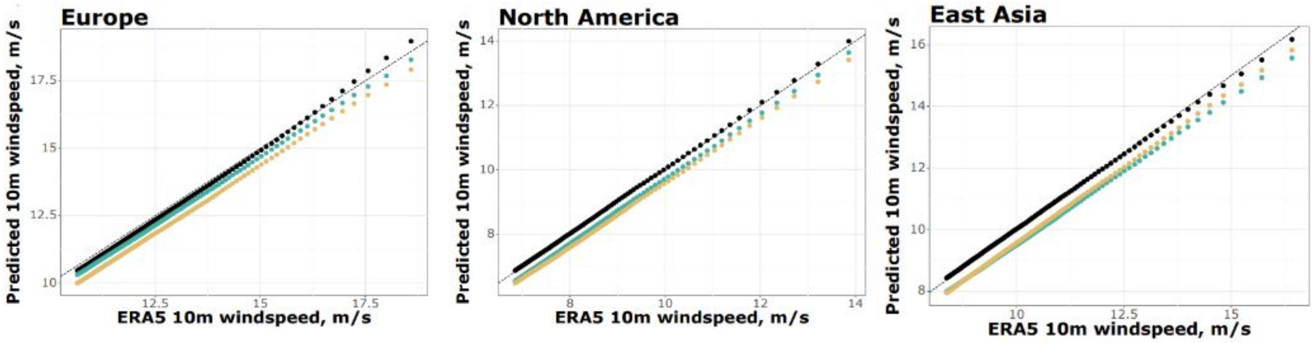
2m temperature
10m windspeed
Cold extremes
Hot extremes
Wind extremes
1 day
5 days
10 days
% difference in RMSE vs IFS HRES
−90
−50
−20
0
20
50
90

Tail calibration

Hot extremes



Windspeed extremes



Main findings

Data-driven models perform best:

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

Data-driven models perform worse:

- for 7-10 day forecasts
- at higher latitudes
- for windspeed extremes
- on the east side of ocean-basins and in the middle of vast land areas

Some more general takeaways

- Global data-driven models produce **deterministic forecasts with comparable or superior skill to IFS HRES** in terms of standard performance metrics.

- However, they show **inconsistencies in forecast quality** across regions and lead times, especially for extremes. This raises questions about fairness and equity.

- They also display evident **blurring at longer lead times and physical inconsistencies** (see also Bonavita, 2024).

Bonus: Not only deep learning…

JAMES | Journal of Advances in Modeling Earth Systems®

Research Article | Open Access

A Quantile Generalized Additive Approach for Compound Climate Extremes: Pan-Atlantic Extremes as a Case Study

Leonardo Olivetti, Gabriele Messori, Shaobo Jin

First published: 24 January 2024 | https://doi.org/10.1029/2023MS003753

SECTIONS PDF TOOLS SHARE

Abstract

We present an application of quantile generalized additive models (QGAMs) to study spatially compounding climate extremes, namely extremes that occur (near-) simultaneously in geographically remote regions. We take as an example wintertime cold spells in North America and co-occurring wet or windy extremes in Western Europe, which we collectively term Pan-Atlantic compound extremes. QGAMS are largely novel in climate science applications and present a number of key advantages over conventional statistical models of weather extremes. Specifically, they remove the need for a direct identification and parametrization of the extremes themselves, since they model all quantiles of the distributions of interest. They thus make use of all information available, and not only of a small number of extreme values. Moreover, they do not require any a priori knowledge of the functional relationship between the predictors and the dependent variable. Here, we use QGAMs to both characterize the co-occurrence statistics and investigate the role of possible dynamical drivers of the Pan-Atlantic compound extremes. We find that cold spells in North America are a useful predictor of subsequent wet or windy extremes in Western Europe, and that QGAMs can predict those extremes more accurately than conventional peak-over-threshold models.

Article in JAMES

Motivation

- Sometimes we may want to **understand and quantify** the effect of **a specific driver on the values in the tail of the distribution of the outcome,** i.e. extreme values

- Extreme value theory regression requires a number of assumptions, enough sample size and a clear cut definition of the extremes.

- **Quantile regression** is a natural approach to this question, but **assumes linear relations** or requires prior knowledge of the non-linear effects.

Generalised additive models

- Generalised additive models (Hastie and Tibshirani, 1990) are a flexible class of statistical models requiring very limited knowledge of the relationship between the input and the output.

$$g(E(Y)) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_i(X_i)$$

where E(Y) is the expected value of the outcome, g() is the link function, describing the relationship between the linear predictor and the expected value of the outcome, β0 is an intercept, and f1(X1) + f2(X2) + ⋯ + fi(Xi) are smooth functions of the predictors.

Quantile generalised additive models

- Quantile generalised additive models (QGAMs, Fasiolo et al., 2021) extend generalised additive models to flexibly model a conditional quantile of interest.

$$Q_{Y|X}(\tau) = f_{1\tau}(X_1) + f_{2\tau}(X_2) + \cdots + f_{i\tau}(X_i),$$

where QY|X(τ) is a conditional quantile of choice of the dependent variable.

Minimisation problem

- Quantile generalised additive models (Fasiolo et al., 2021) minimise a loss function similar to the pinball loss,

$$\min E\left\{\rho_\tau^*(Y - X^*\beta)\right\},$$

where ρτ∗ is defined as:

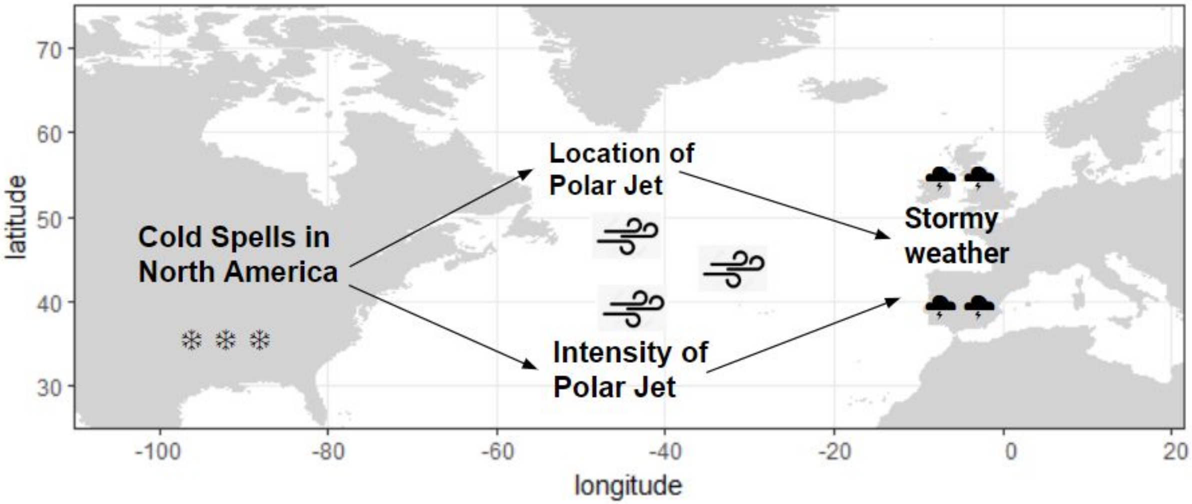
.

$$\rho_\tau^* = (\tau - 1)\frac{z}{\sigma} I(z < 0) + \lambda \log\left(1 + e^{\frac{z}{\lambda\sigma}}\right)$$

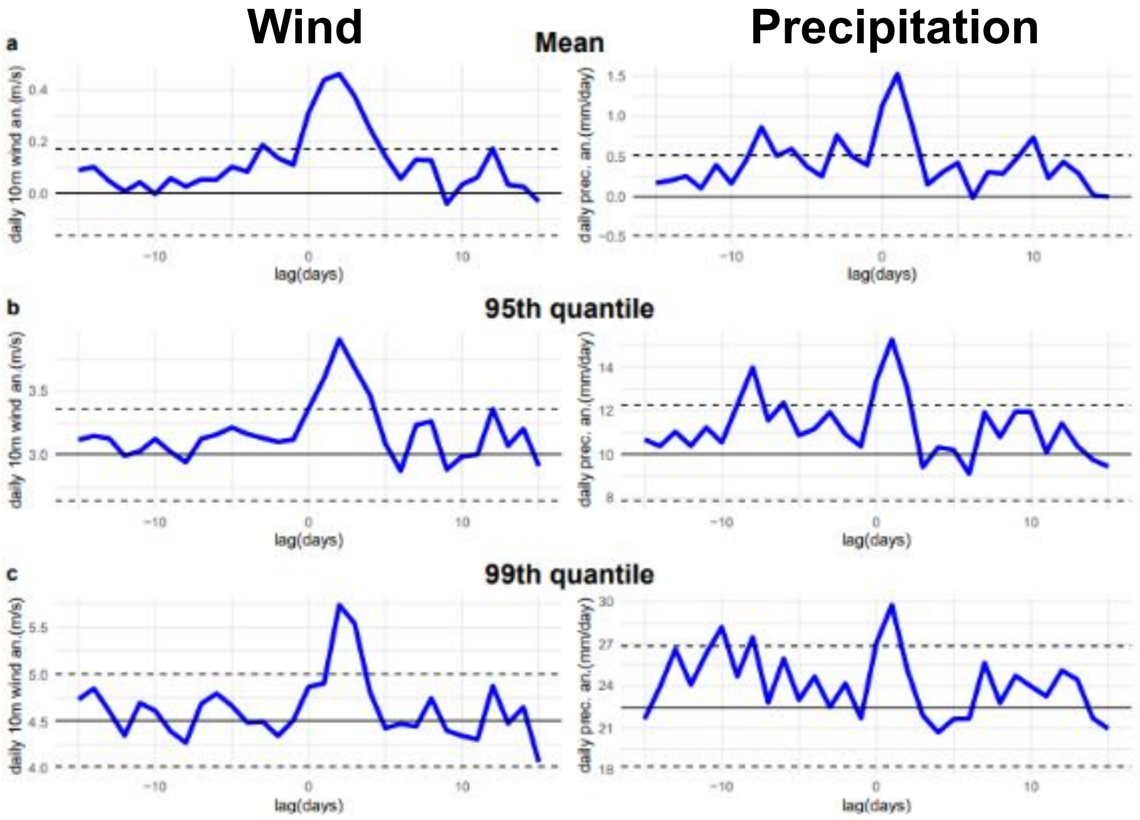
Pτ∗ is the extended log-f loss, which, similarly to the pinball loss, punishes predictions which are further away from the quantile of interest. $\sigma > 0$ is a scale parameter and $\lambda > 0$ is a penalty term, meant to prevent excessive functional complexity.
As λ approaches 0, ρτ∗becomes equivalent to ρ_τ, the pinball loss used in quantile regression (Fasiolo et al.,2021)

Case study: Pan-Atlantic compound extremes

- Several studies e.g. Riboldi et al., 2023, Leeding et al., 2023, Messori et al., 2016 have investigated **the relationship between wintertime cold spells in North America and wet-windy extremes in Europe**.
- The relationship between the two is likely to be strong, but **highly non-linear and hard to quantify**.


70
60
50
40
30
latitude
-100
-80
-60
-40
-20
0
20
longitude
Cold Spells in North America
Location of Polar Jet
Intensity of Polar Jet
Stormy weather

Daily 10m windspeed and precipitation anomalies in Iberia in connection with North American Cold Spells



Aim

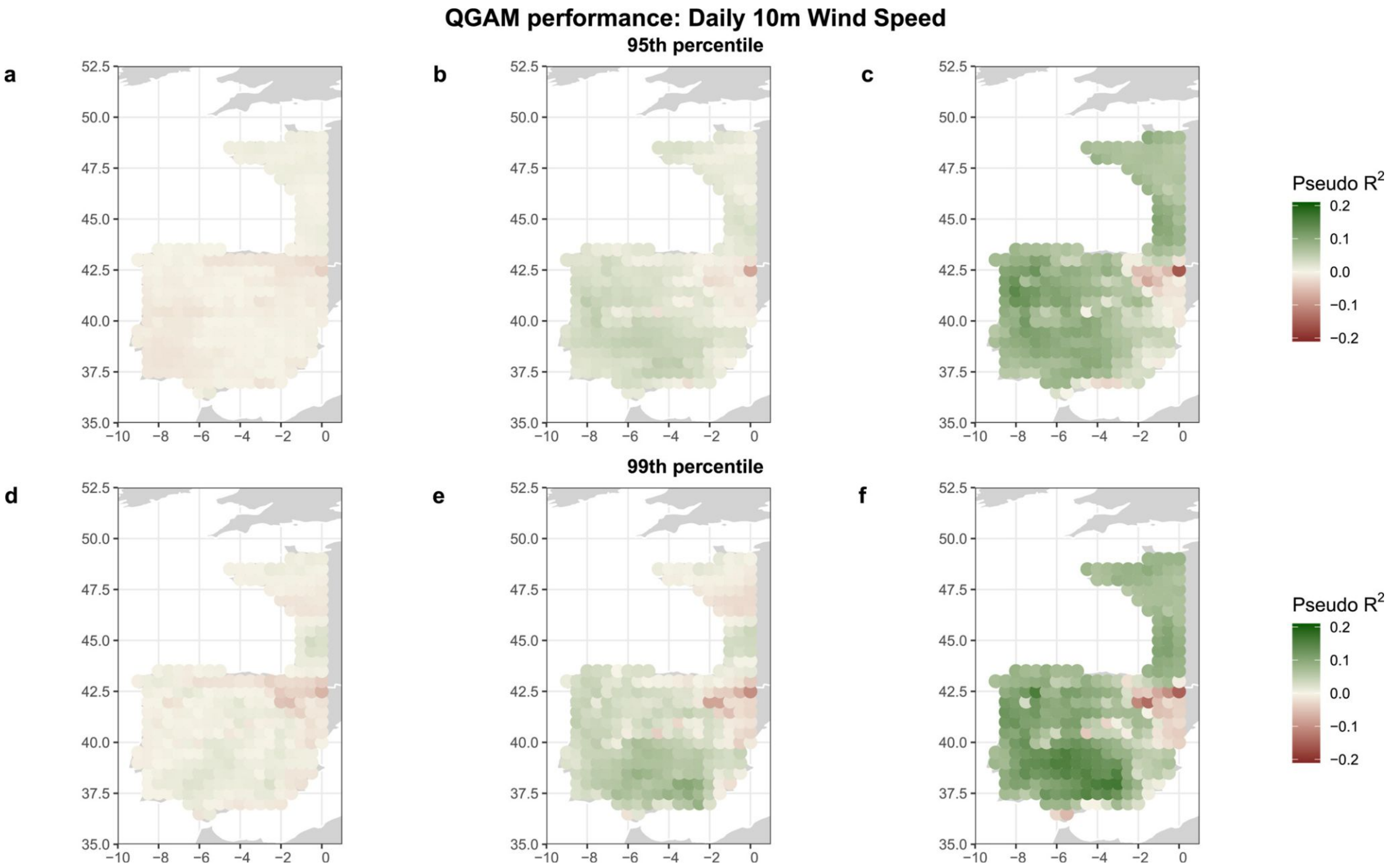
We aim to:

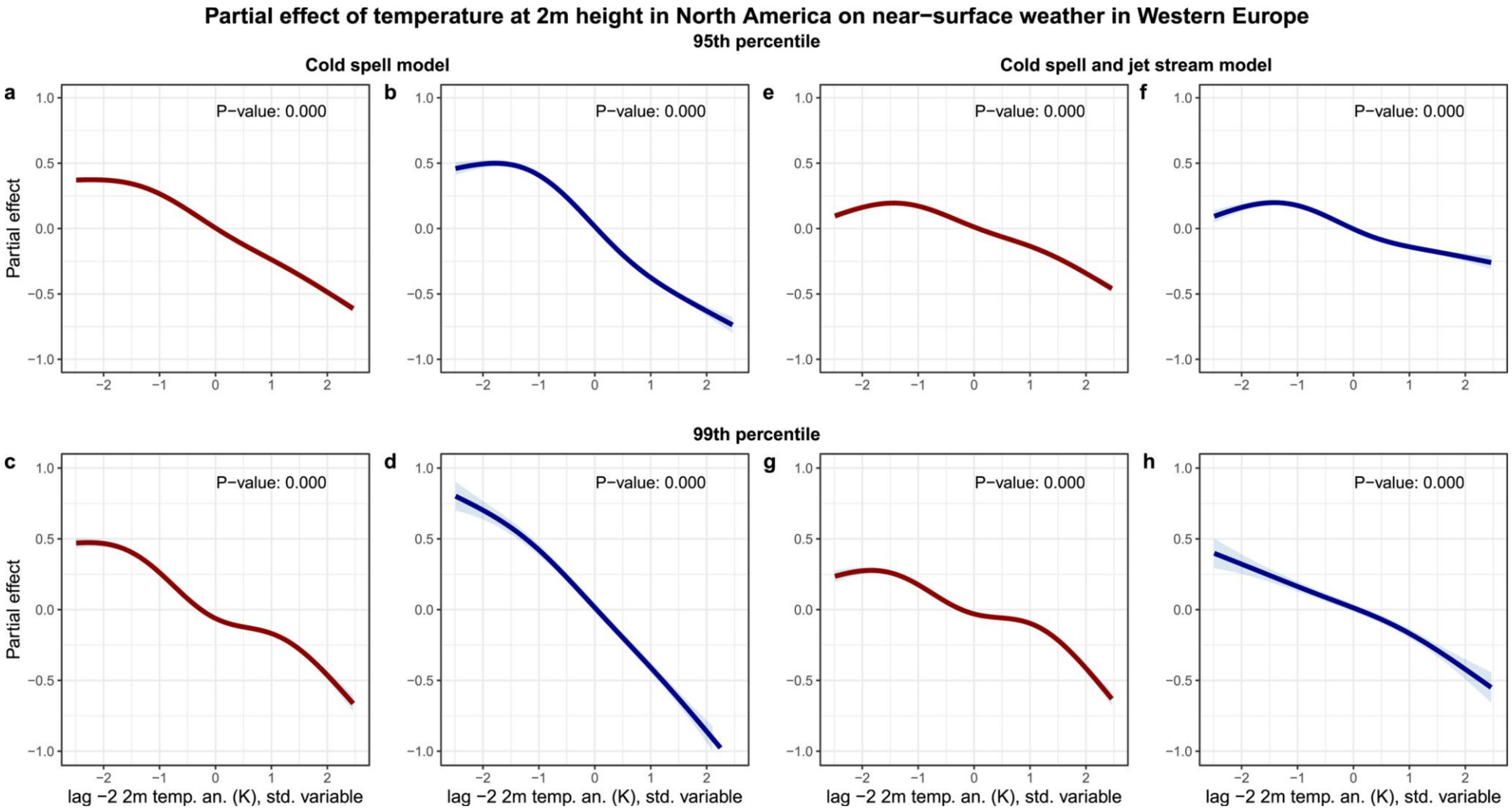
- quantify the association between t2m in North America and wet-windy extremes in Western Europe
- evaluate its significance
- show that QGAMs as a whole behave “reasonably”, i.e. can estimate extreme quantiles with similar skill to conventional models and become progressively better as more information is added to the models.

Setup

We build three models of increasing complexity to estimate extreme quantiles of daily 10m windspeed and accumulated precipitation in South-Western Europe:

- A basic model, making predictions as a function of time, latitude, longitude, and a training-set based seasonal climatology, only.
- A cold spell model, where we include lagged t2m in North America
- A cold spell and jet stream model, where we also include proxies for the location and strength of the Polar jet stream


QGAM performance: Daily 10m Wind Speed
95th percentile
99th percentile
a
b
c
d
e
f
52.5
50.0
47.5
45.0
42.5
40.0
37.5
35.0
−10
−8
−6
−4
−2
0
Pseudo R2
0.2
0.1
0.0
−0.1
−0.2


Partial effect of temperature at 2m height in North America on near−surface weather in Western Europe
95th percentile
Cold spell model
Cold spell and jet stream model
a
b
e
f
P−value: 0.000
99th percentile
c
d
g
h
Partial effect
lag −2 2m temp. an. (K), std. variable

Conclusions

- Quantile general additive models (QGAMs) can model the relationship between compound climate extremes flexibly and robustly

- North American cold spells are significantly associated with wet and windy extremes in Western Europe

- North American cold spells hold some predictive skill for wet or windy extremes in Western Europe, even when accounting for confounders