Unravelling European Heatwaves Dynamics

Climatic Drivers, Quantile Regression, and Causal Approaches

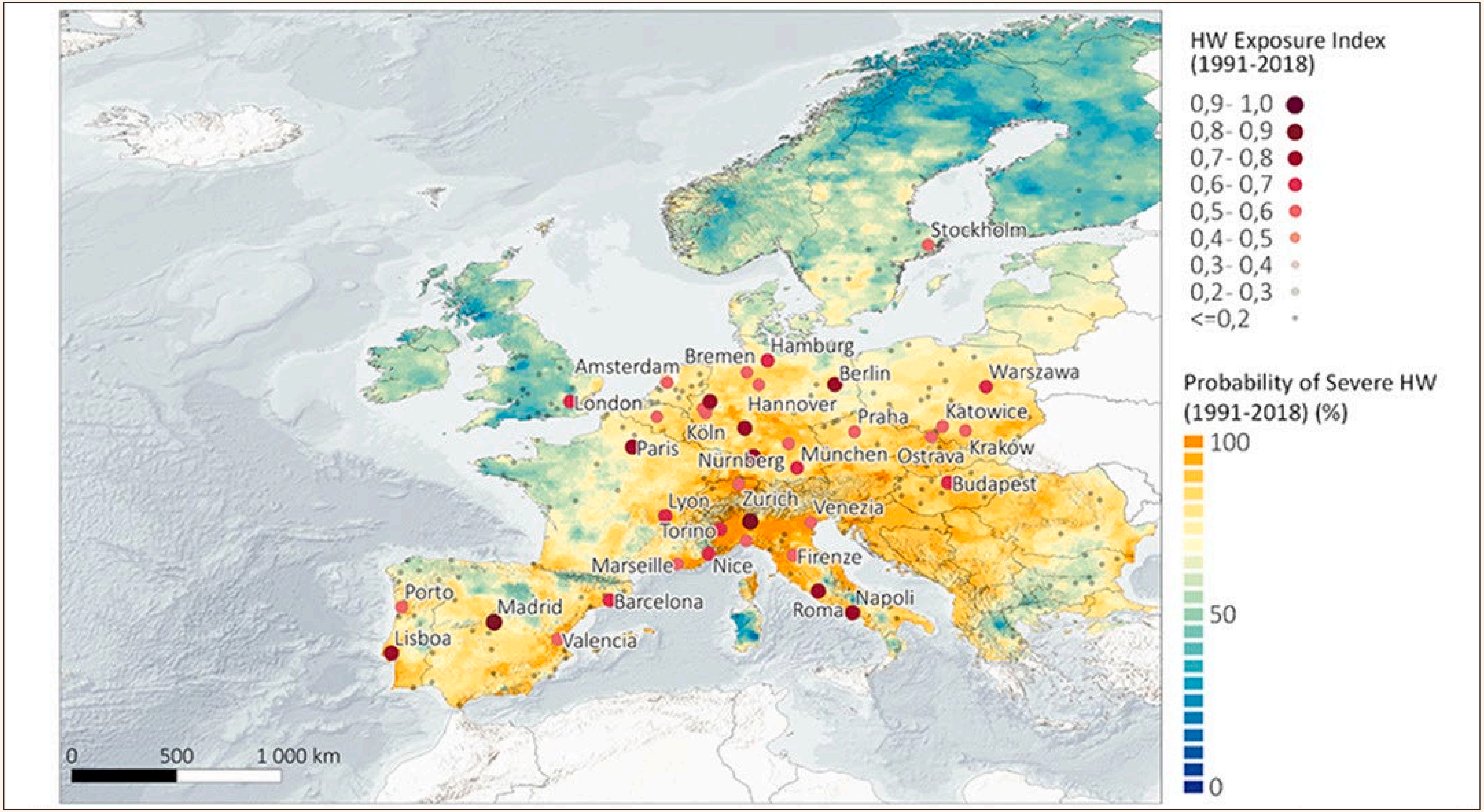
Deborah Bassotto

ELLIIT Focus Period – September 2024

Content

1. A short introduction I: Heatwaves
2. Comparing change of drivers between present and future CMIP6 climates
3. Quantile regression (pinball) loss for extreme events
4. A short introduction II: Causal ML for Climate Sciences

A short introduction I
Heatwaves



From Oliveira, A., Lopes, A., & Soares, A. (2022). Excess heat factor climatology, trends, and exposure across European functional urban areas. *Weather and Climate Extremes*, *36*, 100455.

Processes

Physical Attribution

- Usually, the results of the succession of events
- Example: cold anomalies in the North Atlantic -> strong sea surface temperature gradient -> Rossby wave train -> stationary Jet Stream -> high pressure + extreme temperatures all over Europe (Duchez et al., 2016)

Impacts (non-exhaustive list)

- Health risks and increase in mortality
- Infrastructure strain: pressure on healthcare, energy, water, and transportation systems
- Economic impacts: reduced productivity and agricultural disruptions
- Social impacts: disrupted daily activities and potential mental health issues, “climatic migration”
- Environmental stress: potential damage to ecosystems, increased wildfire risk

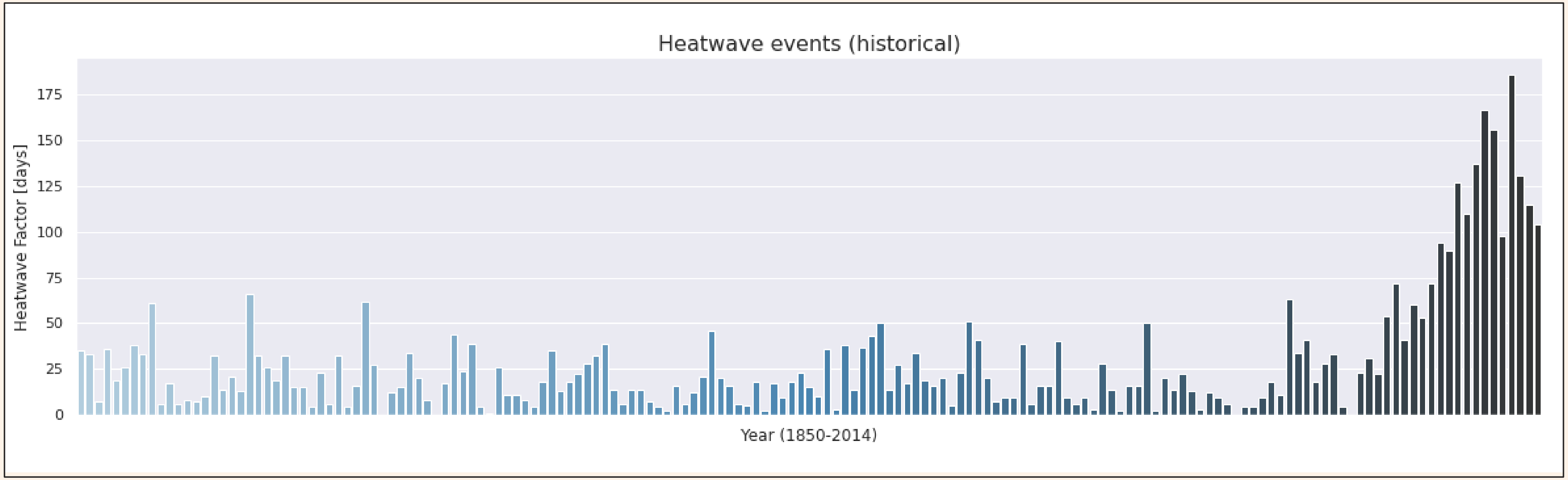
DEFINITIONS

Heatwave Factor

Excess Heat Factor (EHF)	• Daily average of max and min 2m air-temperature
Anomaly	• EHF>90th percentile threshold
Heatwave Event	• Anomaly>0 (each calendar-day, using a 15-days moving window) • For at least 3 consecutive days

Heatwave Factor

CMIP6 Dataset (GFDL-CM4 - Historical)



Comparing Heatwaves Drivers Between Present and Future CMIP6 Climates

Project Framework

Present (i.e., historical)

1) 3D, 2D global fields potential drivers

PLS

2) Extract relevant global fields patterns for UE heatwaves

[redacted]

XGBoost

4) Prediction of European heatwaves based on reduced feature's time-series

Test directly in future climate

Future (e.g., $4xCO_2$)

5) Does it fail?

(If not: drivers quasi–steady)

Yes

5-b) Repeat the same process as for present climate and show which patterns changed

Robustness

6) Test across different CMIP6 models

Partial Least Square Regression (PLS)

Objective

- Maximize covariance between the predictor and the target

Key Elements

- Supervised method
- Suited for datasets where there are more predictor variables than observations **and** when multicollinearity exists among predictors
- Achieves dimensionality reduction while preserving information most relevant to the prediction task

Datasets – CMIP6

Heatwave drivers

1. Air temperature
2. Upper soil moisture (Perkins, 2015; Liu et al.,2020)
3. Meridional and zonal winds (Perkins, 2015)
4. Snow fraction (Zhang et al., 2022)
5. Geopotential (Kornhuber et al. 2019)
6. Precipitation (Liu et al. 2020)
7. Humidity (Liu et al.2020)
8. Sea surface temperature (Duchez et al., 2016; Mecking et al. 2019)

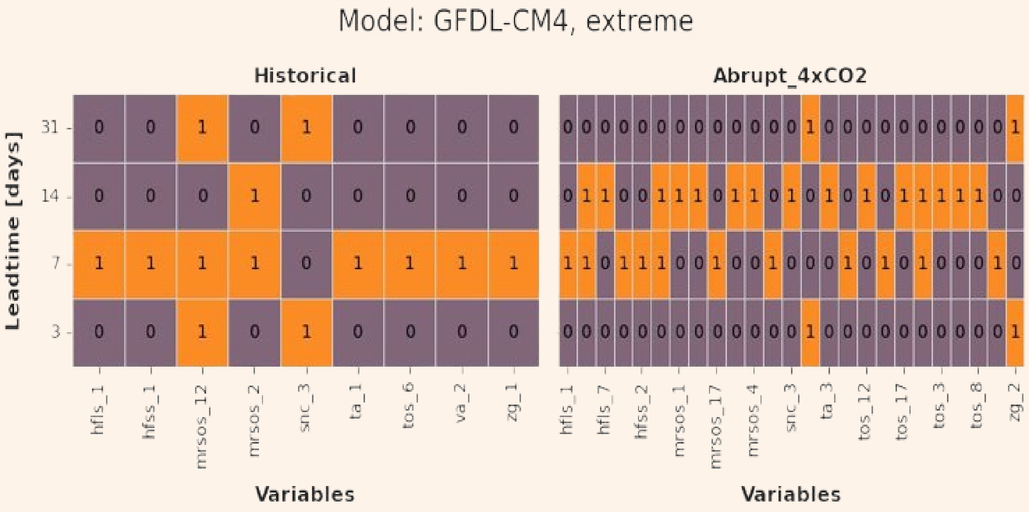
CMIP Experiments

1. Historical (1850-2014)
2. ssp585 (future)

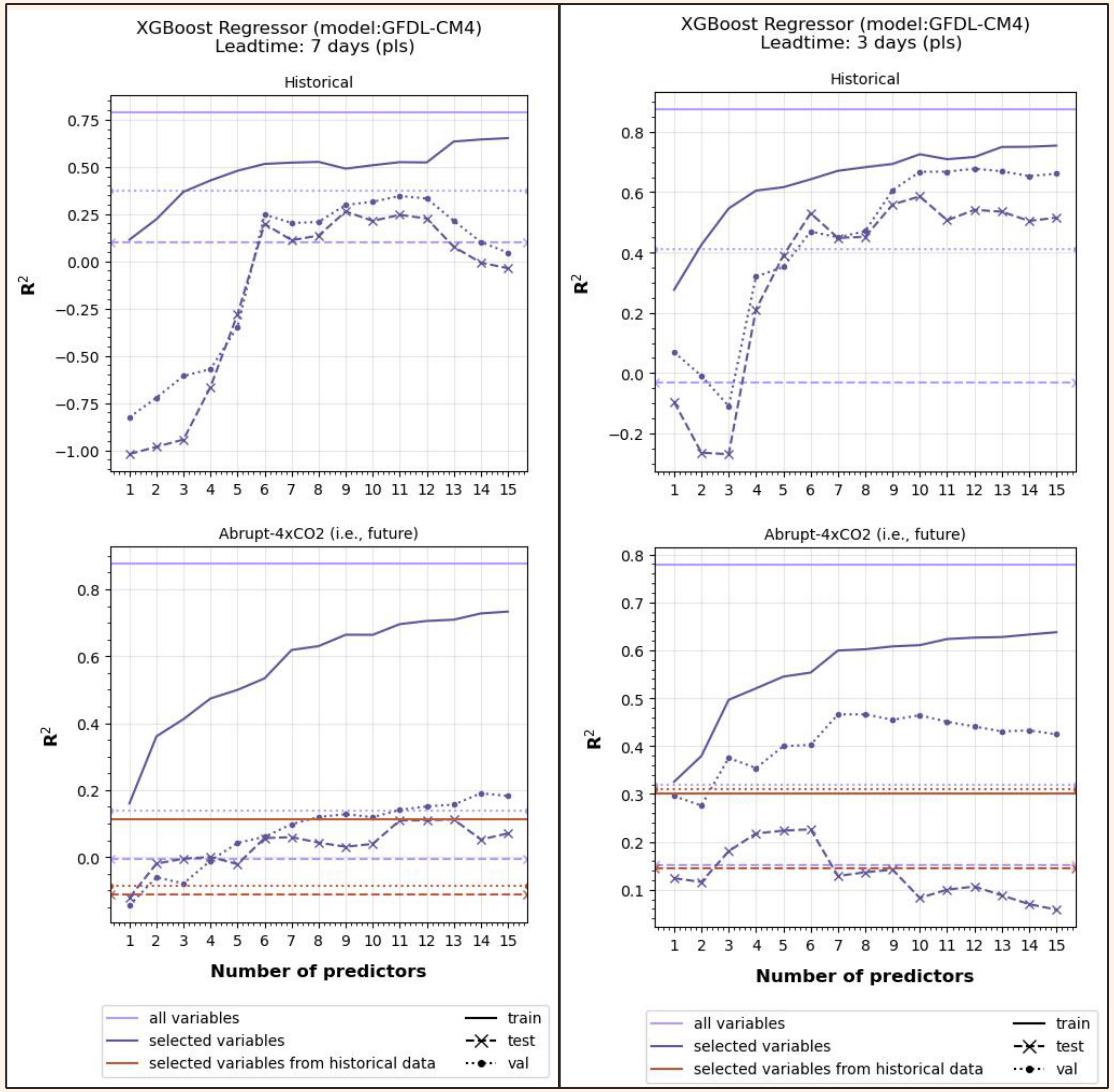
CMIP Models

1. GFDL-CM4
2. ACCESS-CM2
3. ???

(Pre-)Results



Model: GFDL-CM4, extreme
Historical
Abrupt_4xCO2
Leadtime [days]
31
14
7
3
hfls_1
hfss_1
mrsos_12
mrsos_2
snc_3
ta_1
tos_6
va_2
zg_1
hfls_1
hfls_7
hfss_2
mrsos_1
mrsos_17
mrsos_4
snc_3
ta_3
tos_12
tos_17
tos_3
tos_8
zg_2
Variables
Variables



XGBoost Regressor (model:GFDL-CM4)
Leadtime: 7 days (pls)
Historical
Abrupt-4xCO2 (i.e., future)
R2
Number of predictors
all variables
selected variables
selected variables from historical data
train
test
val
XGBoost Regressor (model:GFDL-CM4)
Leadtime: 3 days (pls)
Historical
Abrupt-4xCO2 (i.e., future)
R2
Number of predictors
all variables
selected variables
selected variables from historical data
train
test
val

Spatio-temporal Nonlinear Quantile Regression for Extreme Events

Motivation

Heatwaves trends: more frequent and intense

Challenging to detect and forecast with traditional, parametric indices: few observations, spatio-temporal gaps in dataset, variable distribution, non-linear dynamics, etc.

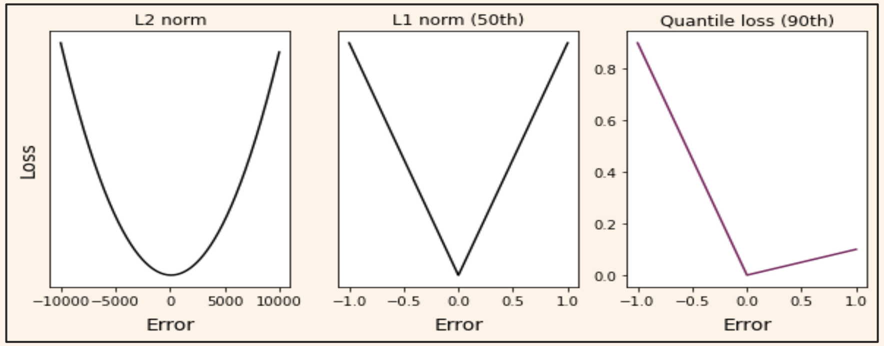
- **Use of a quantile-based method for measuring feature importance for extremes**

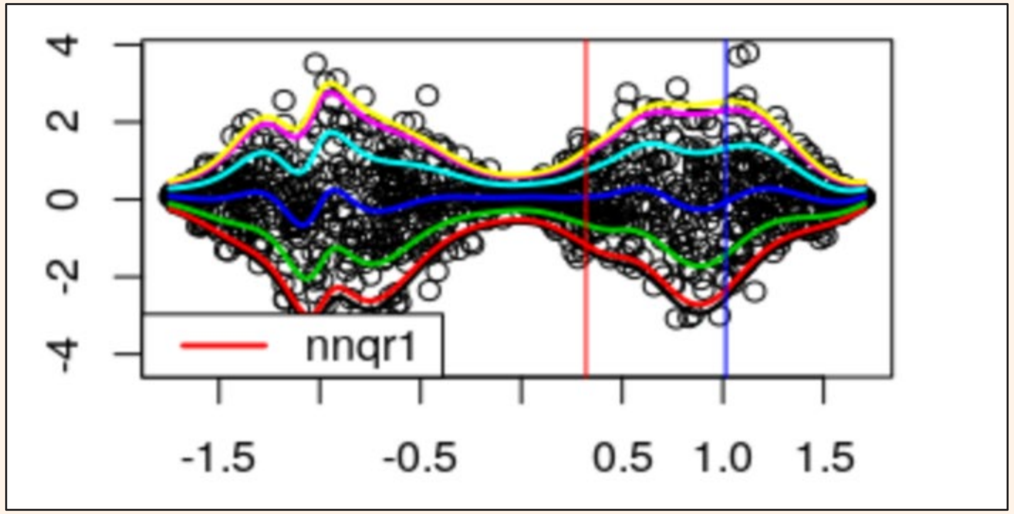
  Maximum temperature detection using atmospheric and surface variables

Quantile Regression - Pinball Loss

- **L2-norm (squared error)** prone to outliers, big errors are penalized a lot
- **L1-norm (MAE)** robust to outliers but symmetric too
- **Pinball (quantile) loss** asymmetric, quantile specific

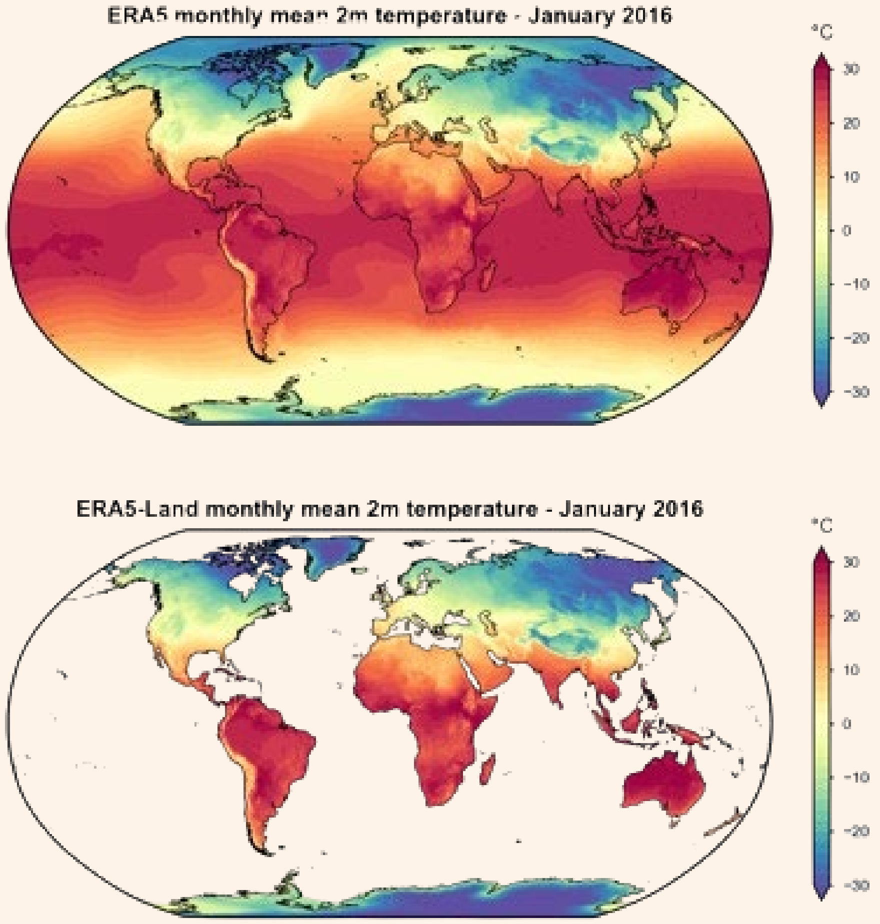
$$
\begin{aligned}
L_\tau(y,z) &= (y-z)\tau & \quad if\ y \geq z \\
&= (z-y)(1-\tau) & \quad if\ y \leq z
\end{aligned}
$$





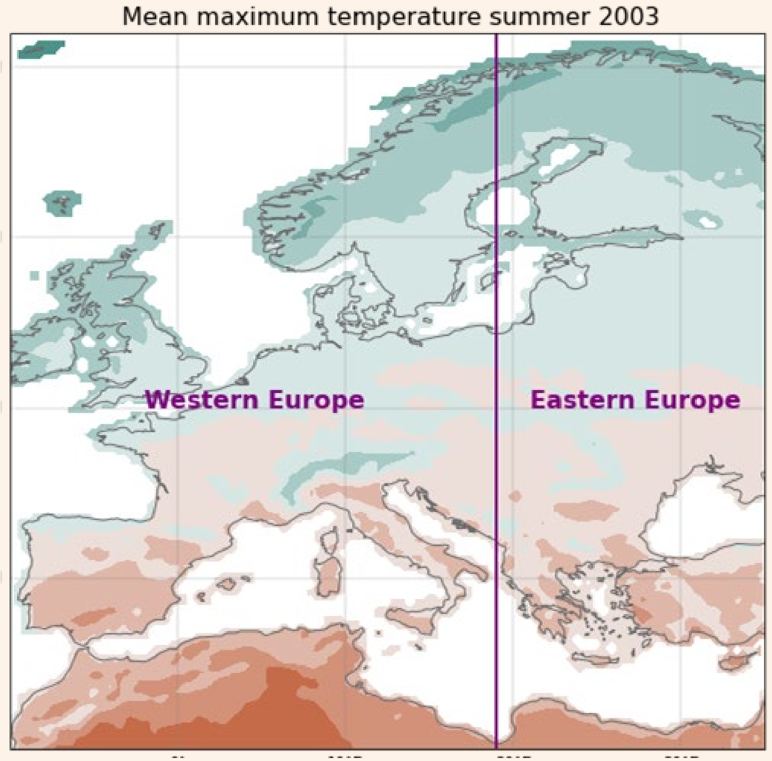
ERA5 Dataset

- **ERA5 Land 6-hourly --> daily mean**
- **Variables : 7**
  - Geopotential (surface, 200hPa, 850hPa)
  - Eastward wind (surface, 200hPa, 850hPa)
  - Soil water content (1st layer)
- **Studied areas : 2**
  - Western Europe, summer 2002-2004
  - Easter Europe, summer 2009-2011
- **Spatial resolution: ~9km**



Experimental Setup

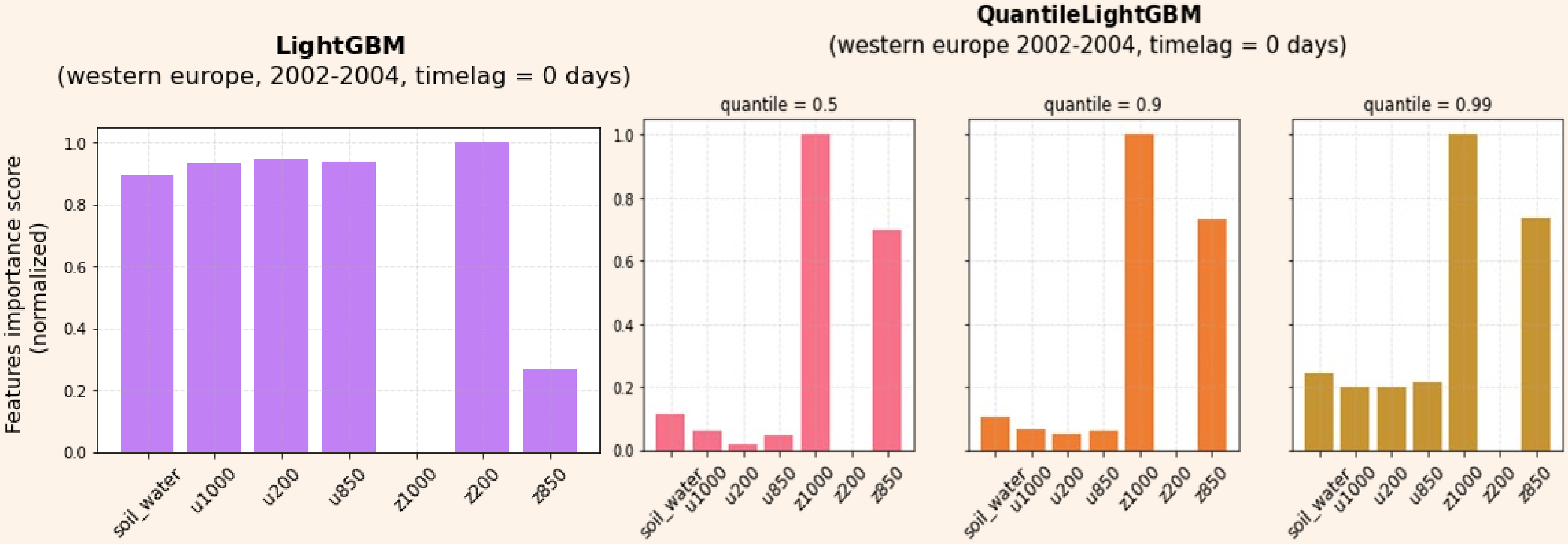
- **Task:** nowcasting daily max temperature prediction
- **ML methods:**
  - LightGBM (quantile vs normal regression)
  - With and without the pinball loss
- **3 models' setup:**
  - Random pixel-wise spatio-temporal selection (1960-2022)
  - Region 1: western Europe (2002-2004)
  - Region 2: eastern Europe (2009-2011)
- **Evaluation metrics**
  - R^2 (modified), RMSE
  - Explainability - feature ranking (sparsity)
  - Using different time-lag for predictors ("forecasting")



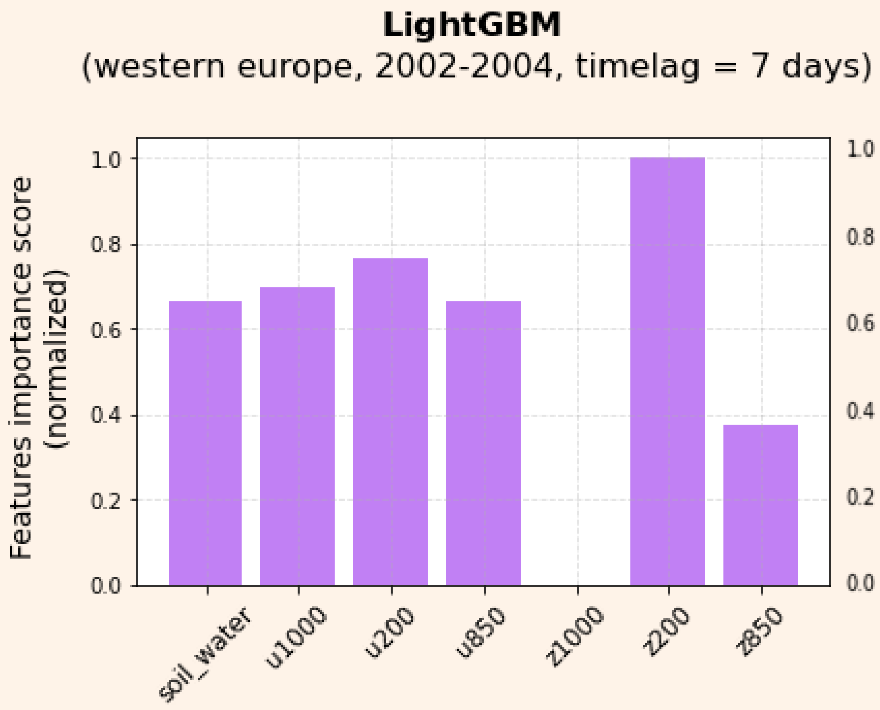
Qualitative Results

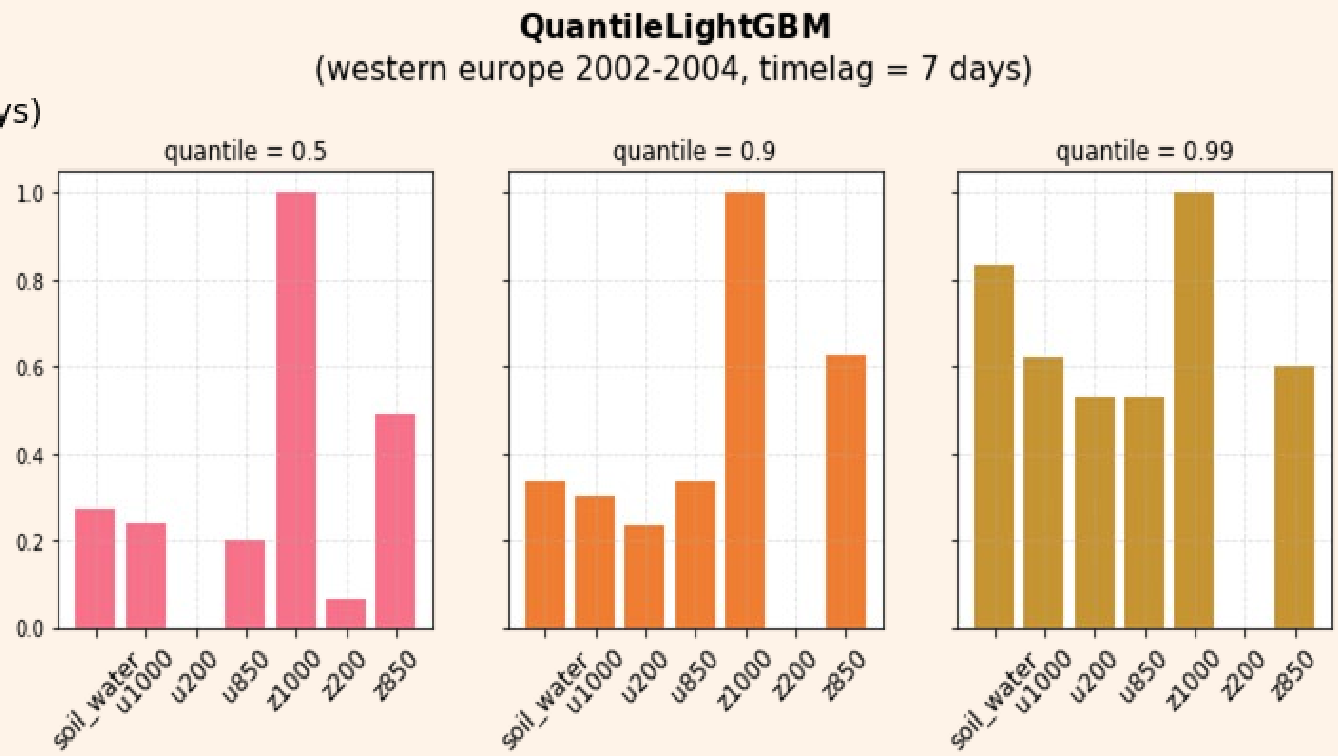
		Quantile LGBM			LGBM
		quantile			
		0.5	**0.9**	**0.99**	
western europe lag = none	R^2	0.935	0.941	0.938	0.839
	RMSE	2.433	3.819	6.134	3.819
western europe lag = 7 days	R^2	0.783	0.78	0.769	0.392
	RMSE	4.452	7.427	11.841	7.427
eastern europe lag = none	R^2	0.938	0.94	0.89	0.844
	RMSE	2.427	3.848	6.885	3.848
eastern europe lag = 7 days	R^2	0.764	0.729	0.636	0.3
	RMSE	4.76	8.171	12.626	8.171
random pixels europe lag = none	R^2	0.937	0.95	0.944	0.858
	RMSE	2.46	3.686	6.152	3.686

Explainability – Features Importance



Time Improves and Impacts Results





A short introduction II

Causal ML for Climate Sciences

Definitions

- **Causal Discovery**

  “Causal discovery algorithms aim to (infer) find **causal structures** from observational data, typically represented as directed acyclic graphs (DAGs).” (Spirtes et al., 2001)

- **Causal Inference**

  “The process of determining the independent, actual **effect** of a particular phenomenon that is a component of a larger system." (Pearl, 2009)

Key Concepts

- **Directed Acyclic Graphs (DAGs):** Graphical representations of causal relationships where nodes represent variables and edges represent causal influences.
- **Structural Causal Models (SCMs):** Mathematical models that represent causal relationships as a set of structural equations.
- **Cofounders**: variable that influences both the treatment (driver) and the outcome (event), potentially leading to a spurious association between them.
- **Topological ordering:** representation of causal structures as a system of structural equations (i.e., each variable is a function of its parents in the graph)

Graphical representation of SCMs

Acyclicity

- **Temporal ordering:** causal relationships inherently imply a temporal order (i.e., causes precede effects -> cyclicity lead to ambiguous causal relationships)
- **Identifiability:** methods that rely on the acyclic nature of the graph (e.g., backdoor/front-door criterion) can identify sets of variables that estimate causal effects
- **Infinite regress:** prevented (i.e., cycles could lead to situations where A causes B, B causes C, and C causes A -> logical paradoxes and impossible to determine the true causal structure).
- **Feedback loops:** can be represented by "unrolling" the cycle over time steps (i.e., acyclicity maintained)

Cyclicity

- **Causal Markov condition:** Broken (i.e., states that a variable is independent of its non-descendants given its parents).
- **Difficulties in intervention analysis:** in causal inference, we often want to reason about the effects of interventions. Cycles make it challenging to predict the outcomes of interventions because changes propagate indefinitely through the cycle.

We want DAGS

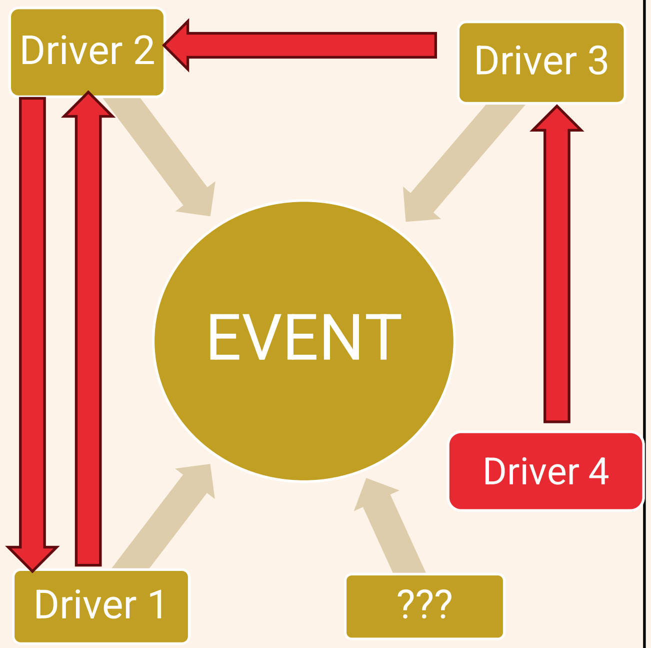
Causality for Climate Science - Why?

Conventional climate models sometimes fail to give good prediction of extremes:

- Spatio-temporal gaps in datasets
- Non-linear dynamics and non-Gaussian distribution of variables
- Model parameterization
- High-dimensional and synergistic effects but small size of observed data

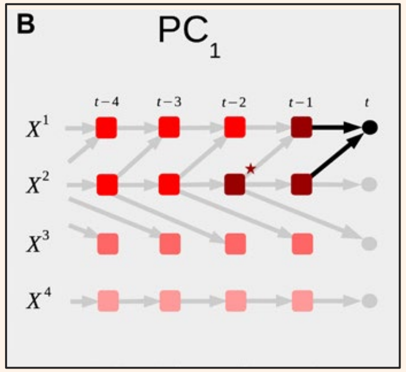
Causality for Climate Science - Challenges

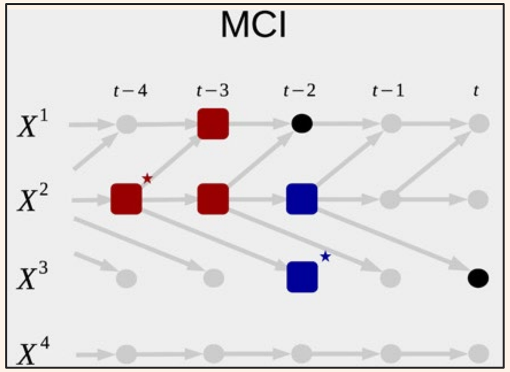
- **Feedback loops**: mechanisms that create cyclic causal structures (i.e., not DAGs)
- **Time-lagged effects**: climate processes often involve time delays between causes and effects.
- **Non-stationarity**: systems can change over time due to unobserved drivers.



Causality for Climate Science - Example: PCMCI (J. Runge, et al. 2019)

- **Condition selection:** PC algorithm identifies relevant conditions (potential causal parents) for each variable in the time series (i.e., tests for independence between variables)
  - Removes irrelevant conditions
  - Reduces the dimensionality of the problem
- **Causal test:** MCI test assesses causal relationships between variables (i.e., finds autocorrelation in data)
  - Removes indirect links and common drivers





References

1. Oliveira, A., Lopes, A., & Soares, A. (2022). Excess heat factor climatology, trends, and exposure across European functional urban areas. *Weather and Climate Extremes*, *36*, 100455.

2. J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M. D. Mahecha, J. Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. Nature communications, 10(1):1–13, 2019a.

3. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, *5*(11), eaau4996.v

4. S. E. Perkins. A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. Atmospheric Research, 164:242–267, 2015.

5. X. Liu, B. He, L. Guo, L. Huang, and D. Chen. Similarities and differences in the mechanisms causing the european summer heatwaves in 2003, 2010, and 2018. Earth's Future, 8(4): e2019EF001386, 2020.

6. K. Kornhuber, S. Osprey, D. Coumou, S. Petri, V. Petoukhov, S. Rahmstorf, and L. Gray. Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. Environmental Research Letters, 14(5):054002, 2019.

8. R. Z. Zhang, X. J. Jia, and Q. F. Qian. Seasonal forecasts of eurasian summer heat wave frequency. Environmental Research Communications, 4(2):025007, 2022.

9. Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms* (p. 288). The MIT Press.

10. Pearl, J. (2009). *Causality*. Cambridge university press.

11. Bassotto, D., Diaz, E., and Camps-Valls, G.: Spatio-temporal Nonlinear Quantile Regression for Heatwave Prediction and Understanding, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-17694, https://doi.org/10.5194/egusphere-egu24-17694, 2024.

12. J. Mecking, S. Drijfhout, J. J. Hirschi, and A. Blaker. Ocean and atmosphere influence on the 2015 european heatwave. Environmental Research Letters, 14(11):114035, 2019.

13. Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search*. MIT press.

14. A. Duchez, E. Frajka-Williams, S. A. Josey, D. G. Evans, J. P. Grist, R. Marsh, G. D. McCarthy, B. Sinha, D. I. Berry, and J. J. Hirschi. Drivers of exceptionally cold north atlantic ocean temperatures and their link to the 2015 european heat wave. Environmental Research Letters, 11(7): 074004, 2016.