FEATURE SELECTION FOR SEASONAL FORECASTING OF EUROPEAN SUMMER HEATWAVES

Ronan McAdam

Junior Scientist, Foundation CMCC – Euro-Mediterranean Centre on Climate Change, Italy

Workshop on "Resilience to Natural Hazards through AI solutions"
Session 3: AI for last-mile communication

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici 20 Years

CLINT CLIMATE INTELLIGENCE www.climateintelligence.eu Universidad de Alcalá Helmholtz-Zentrum hereon the climate data factory POLITECNICO MILANO 1863

DATA-DRIVEN SEASONAL FORECASTING OF HEATWAVES

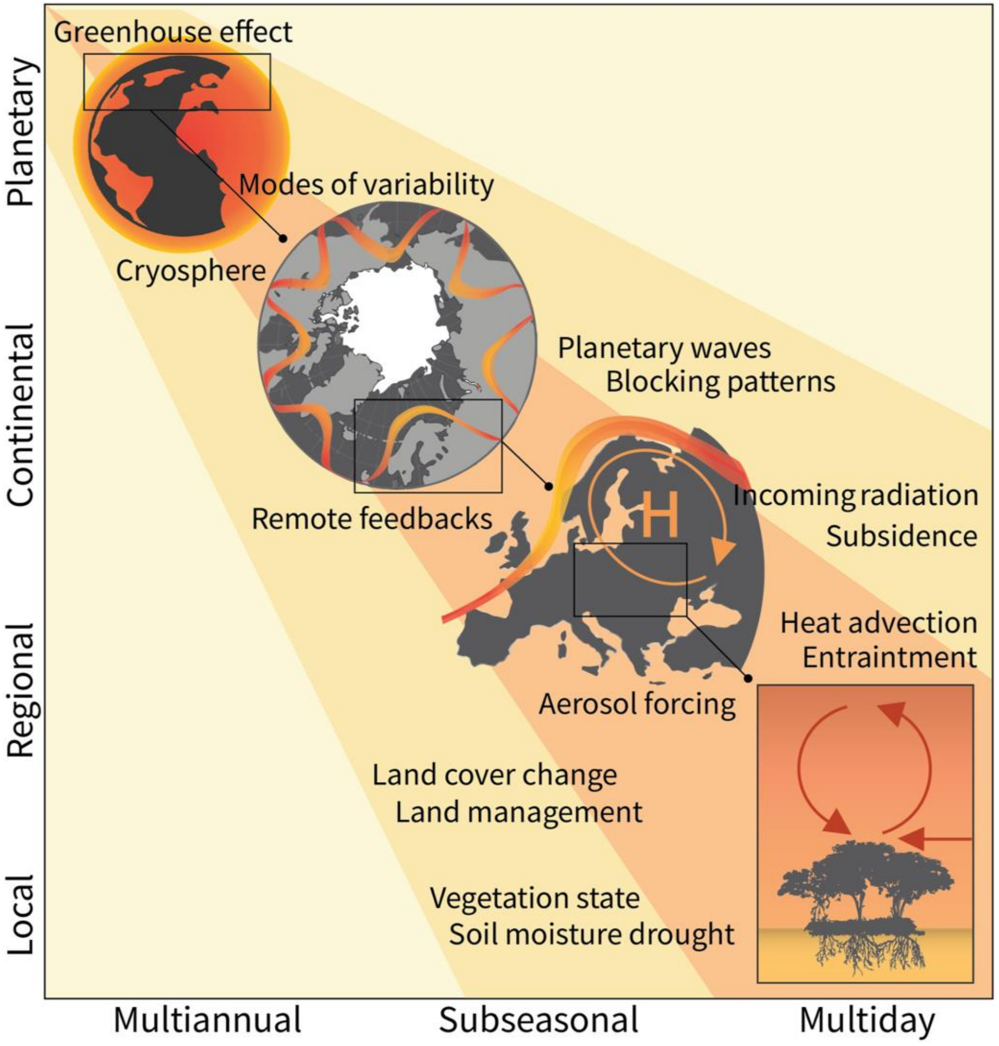
1) Background: State-of-the-art in forecasting extremes
2) The Prototype: methods and key results
3) Future Steps
4) Potential Impact

DATA-DRIVEN SEASONAL FORECASTING OF HEATWAVES

1) **Background: State-of-the-art in forecasting extremes**
2) The Prototype: methods and key results
3) Future Steps
4) Potential Impact

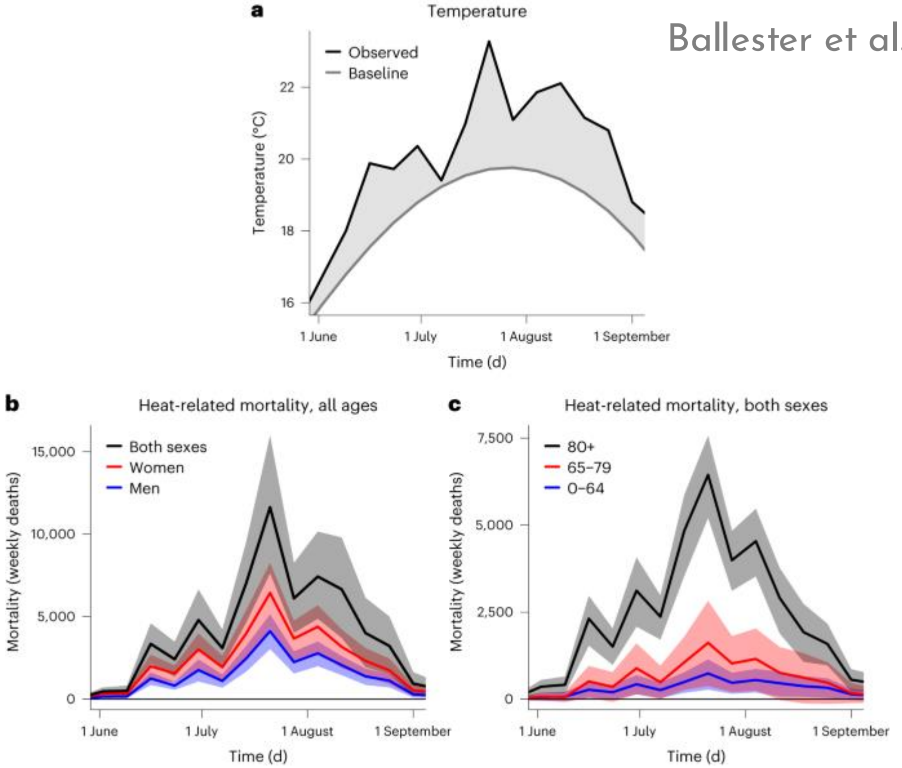
CLINT
CLIMATE INTELLIGENCE

HEATWAVE DRIVERS AND IMPACTS: COMPLEX AND OVERLAPPING



Barriopedro et al., 2023

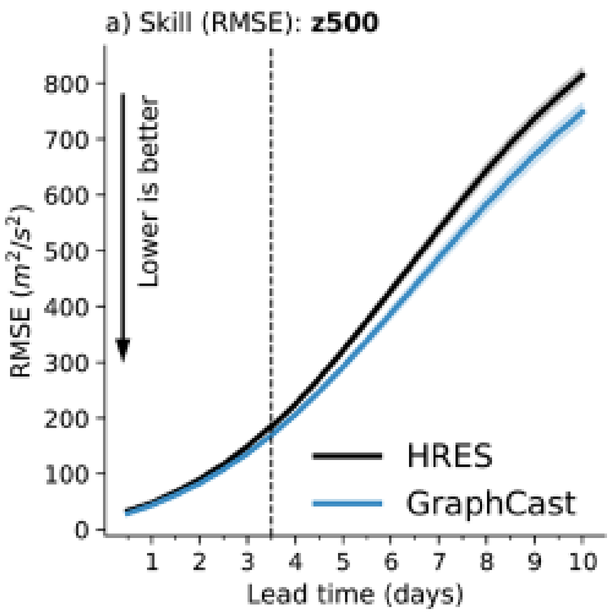
Ballester et al., 2023



CLINT
CLIMATE INTELLIGENCE

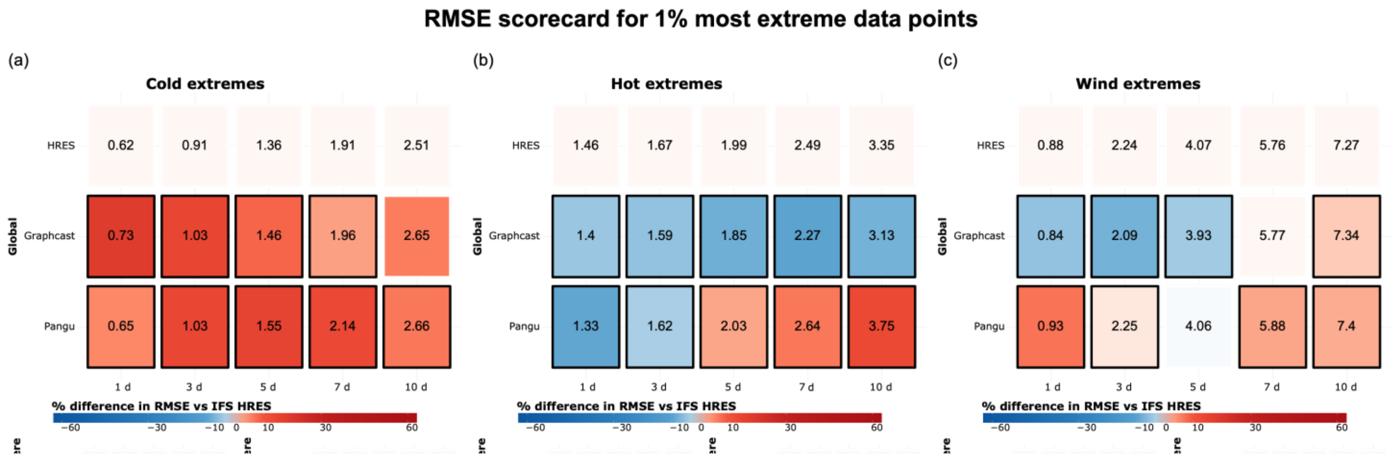
THE STATE OF THE ART OF FORECASTING

The dawn (and fast rise) of data-driven weather forecasting.



Learning skillful medium-range global weather forecasting. Lam et al., 2021

A computationally-efficient alternative which can outperform dynamical systems…but not always!

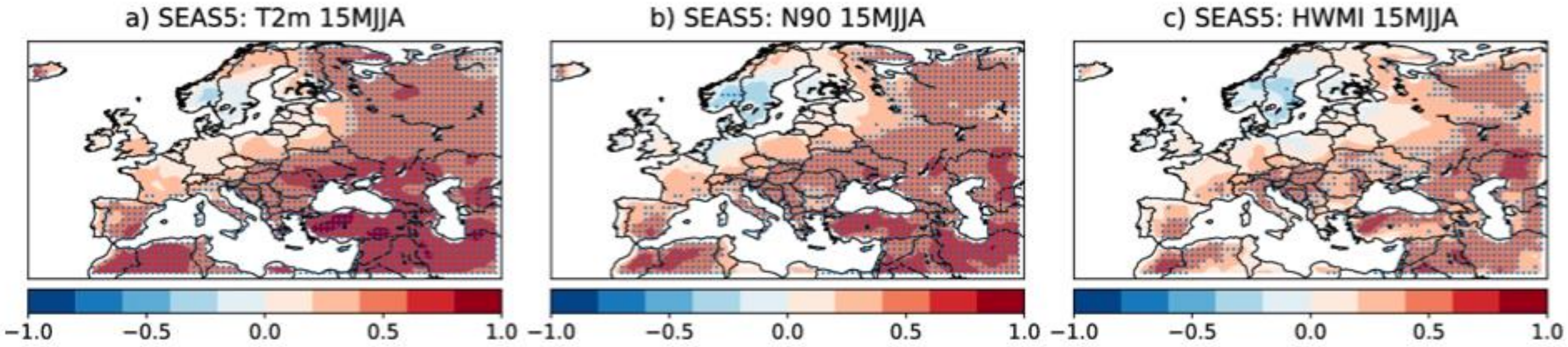


Do data-driven models beat numerical models in forecasting weather extremes? A comparison of IFS HRES, Pangu-Weather, and GraphCast. Olivetti & Messori, 2024

CLINT CLIMATE INTELLIGENCE

THE STATE OF THE ART OF FORECASTING

The current state-of-the-art of dynamical seasonal forecasting



Seasonal prediction of European summer heatwaves. Prodhomme et al., 2021.

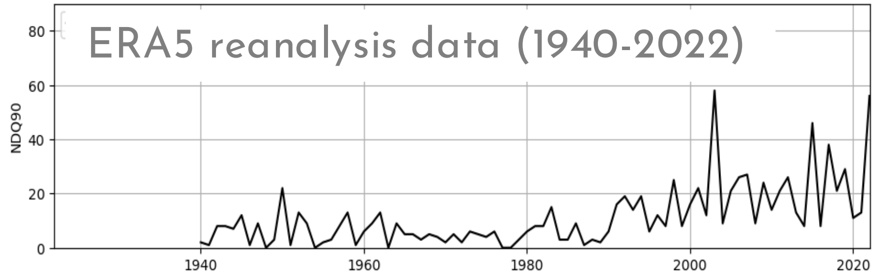
Extension of data-driven approaches to the subseasonal-to-seasonal timescales (and beyond).

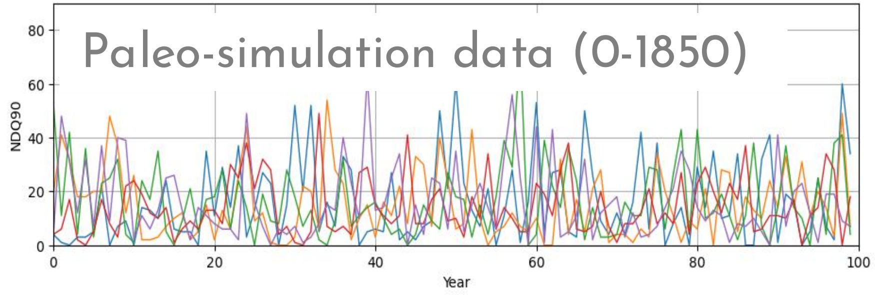
1) **Background: State-of-the-art in forecasting extremes**
2) **The Prototype: methods and key results**
3) Future Steps
4) Potential Impact

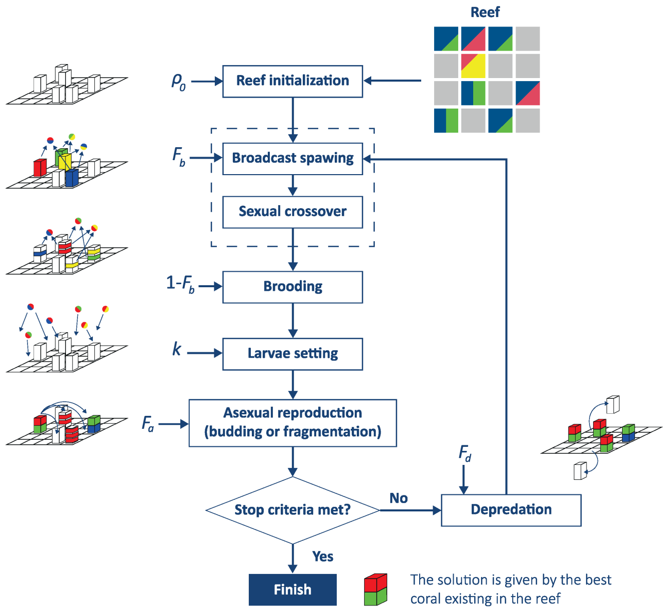
METHOD: FEATURE SELECTION

Key Components

- Multi-Method Ensemble **Optimisation Algorithm**
- Use of long-term **paleo-simulation** to boost training

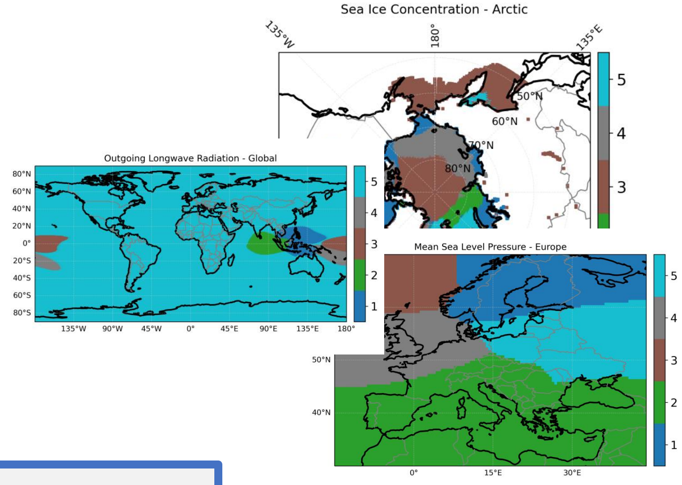






Salcedo-Sanz et al, 2014; 2017; Perez Aracil et al., 2023.

- **Target:** number of summer heatwave days
- **Potential predictors**: atmospheric, ocean, land and sea ice variables
- **Dimensionality-reduction**: k-means clustering



AIM

Identifying optimal combination of predictors for forecasting heatwaves months in advance

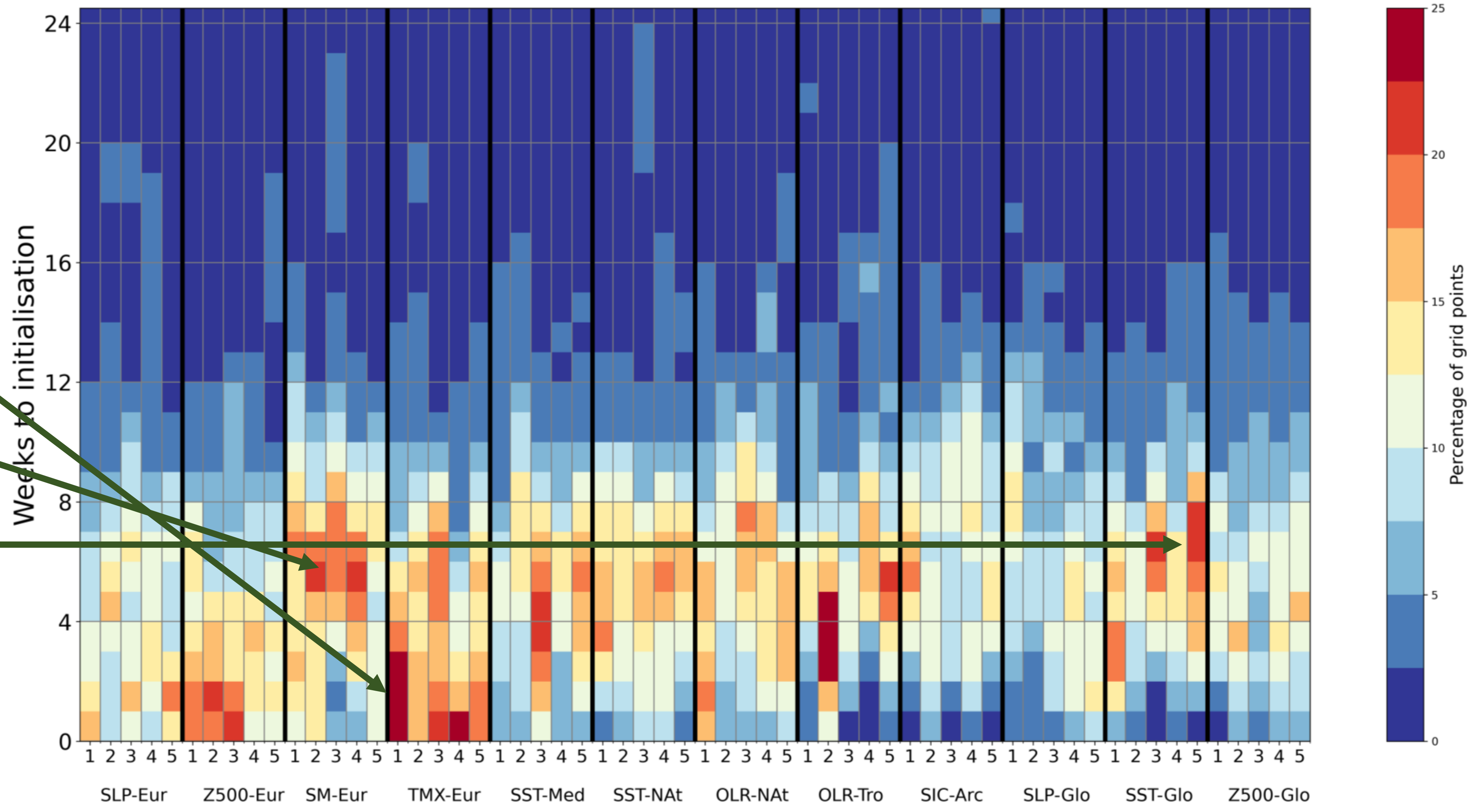
KEY OUTPUT: PREDICTORS OF SUMMERTIME HEATWAVE INDICES

What are the selected features/predictors?

European-wide view of “best” solutions and lag-times

European “VIPs” variables:

- Local temperature and atm. circulation
- March soil moisture
- March-April ENSO

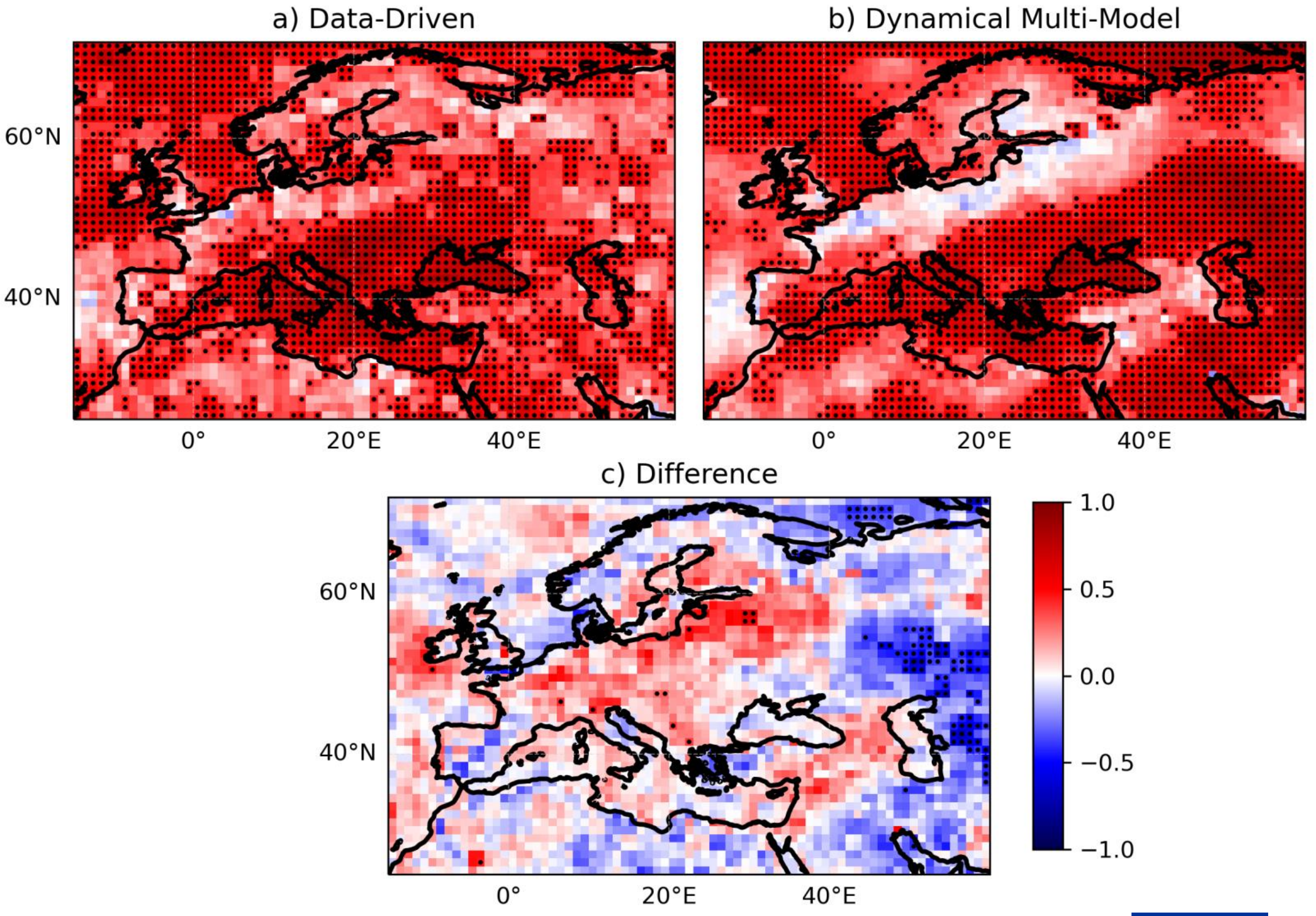


CLINT
CLIMATE INTELLIGENCE

KEY OUTPUT: SKILLFUL DATA-DRIVEN FORECASTS

From simulation to "real" (ERA5) predictors ➡️

- Correlation score – number of HW days over summer - 1993-2016 – against ERA5
- Number of HW days May-July
- Initialized in May
- Data-driven, trained on 1850 years of past2k
- Dynamical multi-model ensemble mean from C3S



DATA-DRIVEN SEASONAL FORECASTING OF HEATWAVES

1) **Background: State-of-the-art in forecasting extremes**
2) **The Prototype: methods and key results**
3) **Future Steps**
4) Potential Impact

FUTURE STEPS

Expansion to **global** domain

Scientific follow-up: first-step in **understanding mechanisms** for newly-identified predictors

Further **refining** of first prototype: different training sets, more predictors

Use in ML for Climate Science **training** (e.g. SCEWERO project with Babes-Bolyai University, Cluj, Romania)

A sub-seasonal version eligible for **AI WeatherQuest competition** (by adjusting lag times of predictors and target variables)

CLINT
CLIMATE INTELLIGENCE

DATA-DRIVEN SEASONAL FORECASTING OF HEATWAVES

1) Background: State-of-the-art in forecasting extremes
2) The Prototype: methods and key results
3) Future Steps
4) Potential Impact

POTENTIAL IMPACT

Index-specific forecasting: why forecast the whole atmosphere when certain case-studies/stakeholders needs very specific information?

Operationalisation: multi-month, cloud-based forecasting of heatwaves

Training: Open-access code for use in ML for Climate Science training (e.g. SCEWERO project with university of Cluj, Romania)

Scientific relevance: first-step in understanding mechanisms for newly-identified predictors

With thanks to...

Ronan McAdam[1], Jorge Perez Aracil[2], Cesar Pelaez-Rodriguez[2], Antonello Squintu[1], Felicitas Hansen[3], Harilaos Loukos[4], Veronica Torralba[5], Leone Cavicchia[1], Matteo Giuliani[6], Eduardo Zorita[3], Sancho Salcedo-Sanz[2], Enrico Scoccimarro[1].

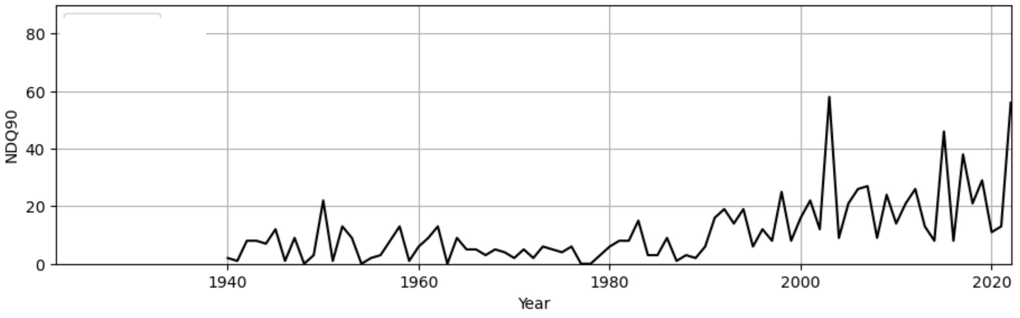
[1]Foundation CMCC – Euro-Mediterranean Centre on Climate Change, Italy
[2]Universidad de Alcalá, Spain
[3]Hereon, Germany
[4]The Climate Data Factory (TCDF), France
[5]Barcelona Supercomputing Centre, Spain
[6]Politecnico di Milano (POLIMI), Italy

www.climateintelligence.eu

1. Set-up

Target: summer HW days (NDQ90)

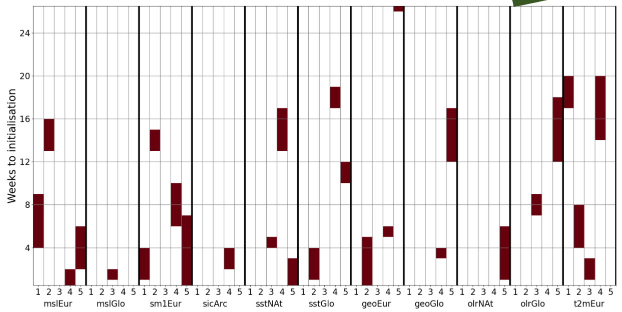
Dimensionality-reduction of predictors



Paleo-simulation

2. Feature Selection

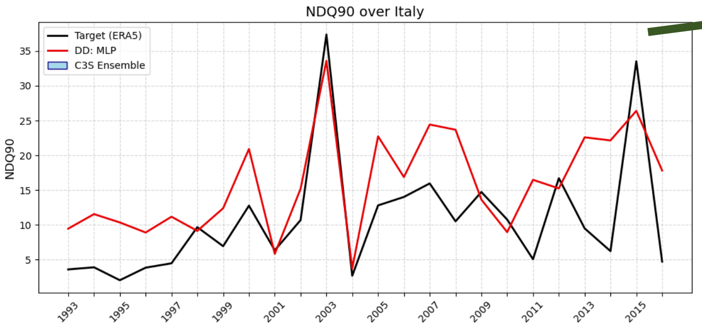
Identifying predictors of summer HW indices



ERA5

3. Prediction

Input predictors into range of ML models



CLINT

CLIMATE INTELLIGENCE

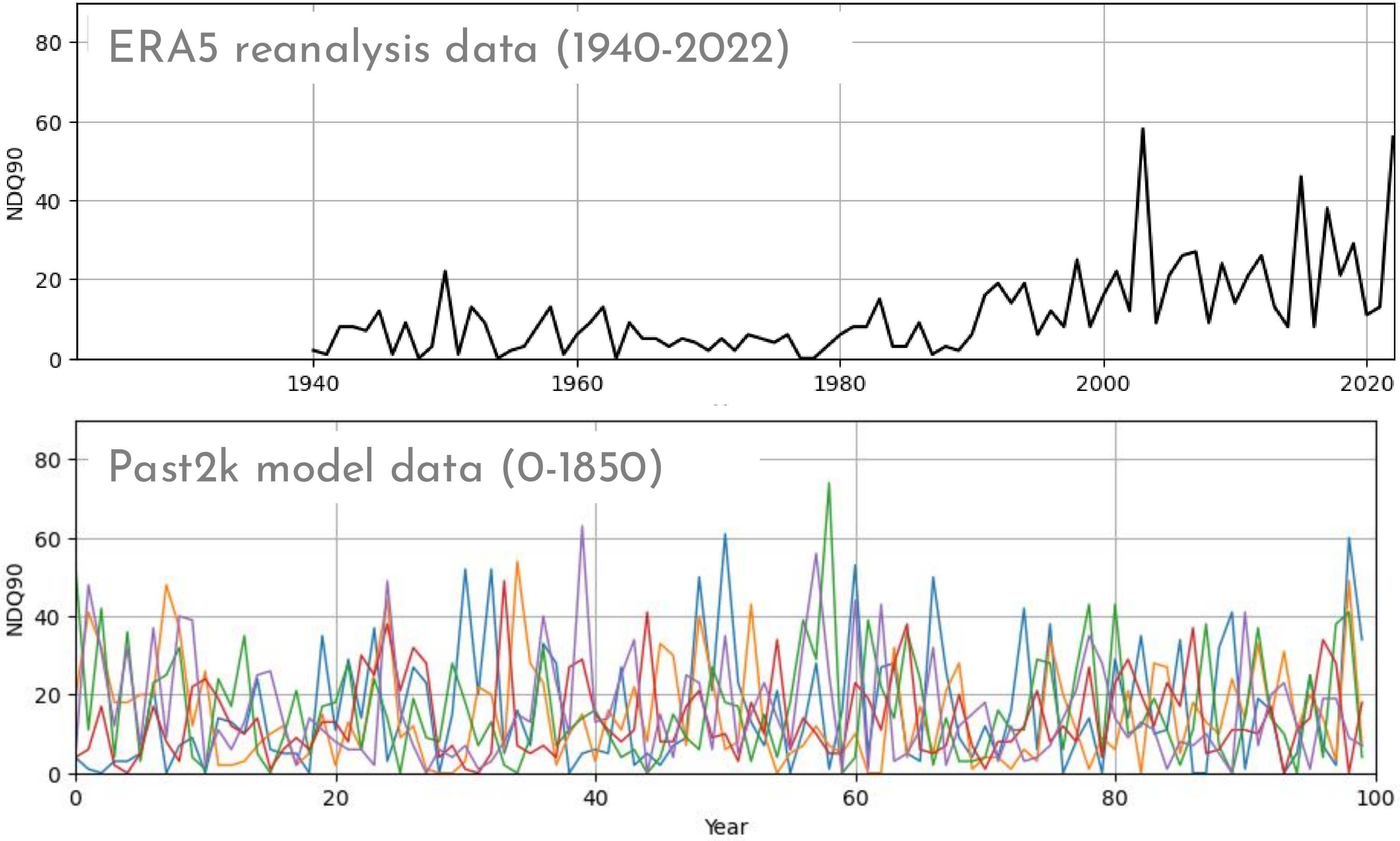
METHOD: Long-term training data

Leveraging a 2000-year climate simulation

MPI-ESM "Past2k" - Climate reconstruction of 0-1850 (Jungclaus et al., 2017)

<u>Atmosphere:</u> ECHAM6
(1.875 deg, 47 vertical levels)
<u>Ocean/Sea Ice:</u> MPIOM
(1.5 deg, 40 vertical levels)

Reconstructions of land-cover, volcanic aerosols, solar forcings (with artificial 11 year cycle) and monthly average ozone concentrations.



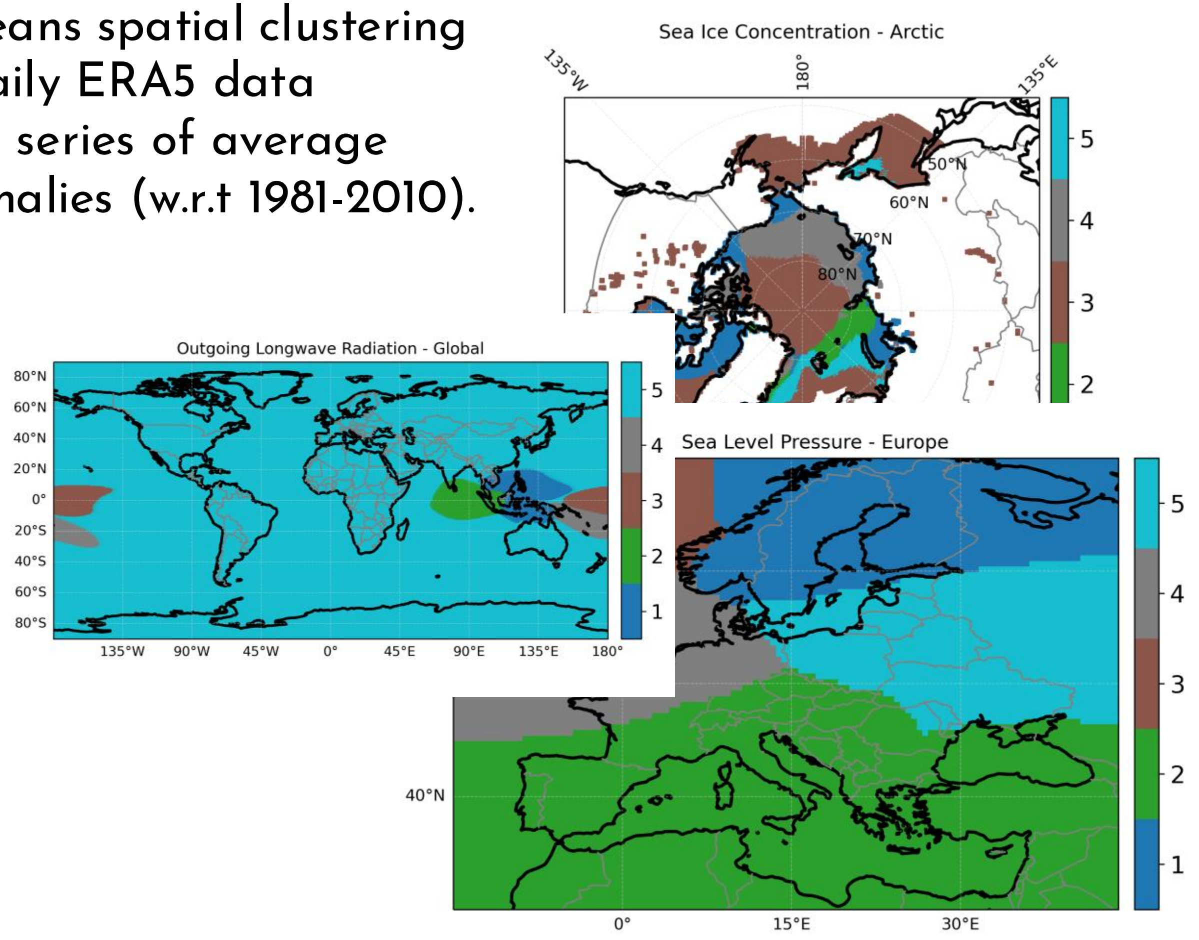
CLINT
CLIMATE INTELLIGENCE

METHOD: Potential predictors

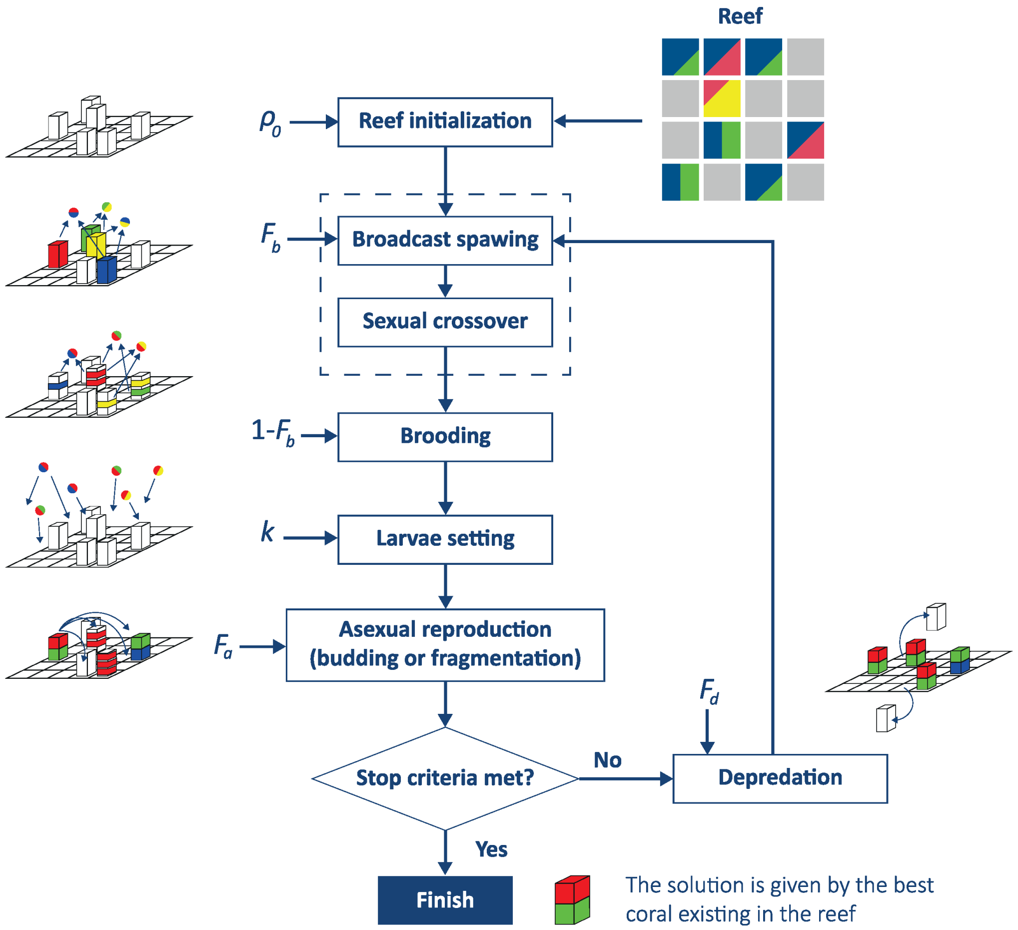
Statistical and ML-based forecasts reduce the dimensionality of problem by using pre-defined area-averages/PCs/clusters as inputs…**can we select the most useful ones?**

Variable	Region
2m temperature (T2M)	Europe
Mean Sea Level Pressure (MSLP)	Europe
Soil Moisture (SM)	Europe
Geopotential Height 500hPa (Z500)	Europe
Total Precipitation (TP)	Europe
Outgoing Longwave Radiation (OLR)	North Atlantic
Sea Surface Temperature (SST)	North Atlantic
Sea Ice Content	Arctic
MSLP	Global
Z500	Global
SST	Global
OLR	Global

K-means spatial clustering of daily ERA5 data
Time series of average anomalies (w.r.t 1981-2010).



METHOD: Coral reef optimisation algorithm



Pool of potential predictors:
Cluster selection → [0,1]
Time lag of each cluster → [0,26] weeks
Sequence length of each cluster → [1,4] weeks

Each evaluation: Linear Regression trained on selected solutions

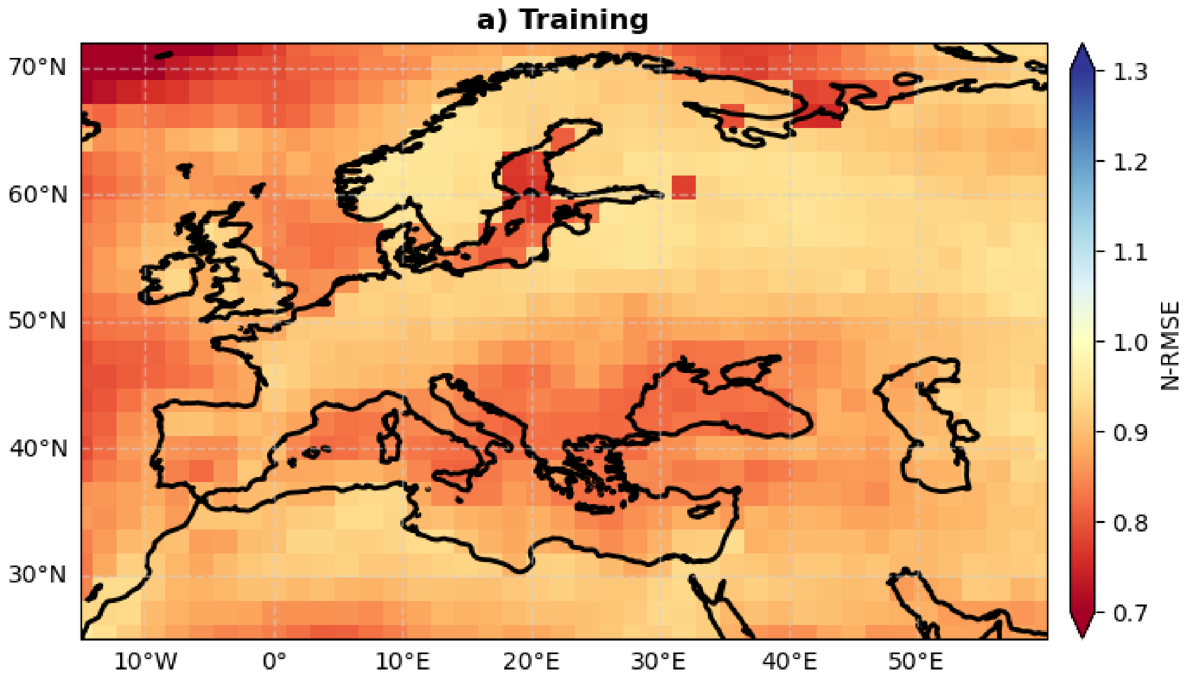
Optimization of 3 solution values based on skill (e.g. normaised root mean squared error - NRMSE)

Target: number of HW days (May-July)
Potential predictors: up to 6 months prior to May 1st

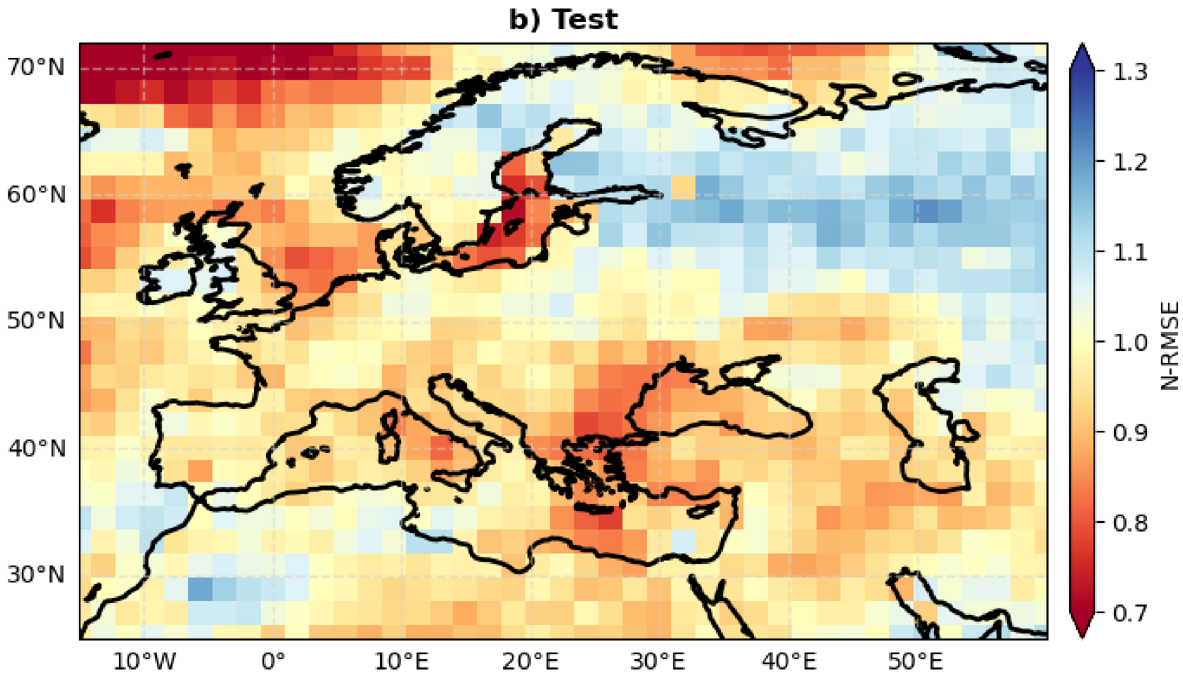
Salcedo-Sanz et al, 2014; 2017;
Aracil Perez et al., 2023.

RESULTS: Optimal skill in model-world

Training period cross-validation
0-1600



Test period
1601-1850



In other words, how well can we recreate model world summer extremes using early information?

The next question is…how transferable is this training/learning to the real world?

CLINT CLIMATE INTELLIGENCE

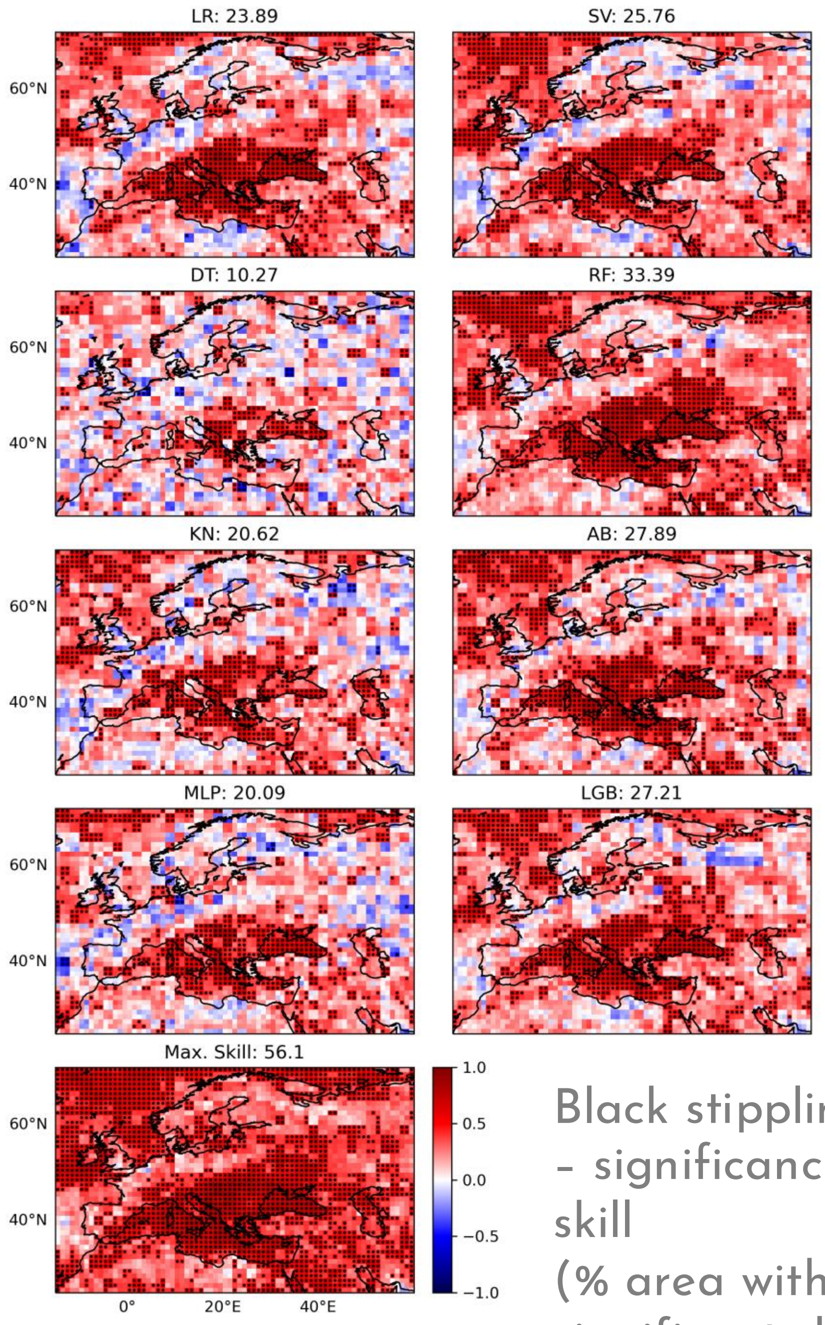
RESULTS: MACHINE LEARNING MODELS

From simulation to "real" (ERA5) predictors ➡️

- Correlation score – number of HW days over summer - 1993-2016 – against ERA5
- Number of HW days May-July
- Initialized in May
- Data-driven, trained on 1850 years of past2k.

Comparing different ML models - which is the best?

Linear Regression - LR
Support Vector - SV
Decision Tree – DT
Random Forest – RF
K Nearest Neighbour – KN
Ada Boost – AB
Multi-Layer Perceptron
Light Gradient Boost - LGB



Black stippling – significance skill (% area with significant skill)

CLINT CLIMATE INTELLIGENCE

SUMMARY

1. Development of (Coral Reef) **optimization-based feature selection** of HW drivers; adapted from *(Perez-Aracil et al., 2025; in review)*

2. Training with millennial-scale paleo-climate – transferable to “real-world” -> **identification of predictors of extreme summers.**

3. Significant skill of summer HW propensity over large parts of Europe – **competing with/matching dynamical state-of-the-art.**

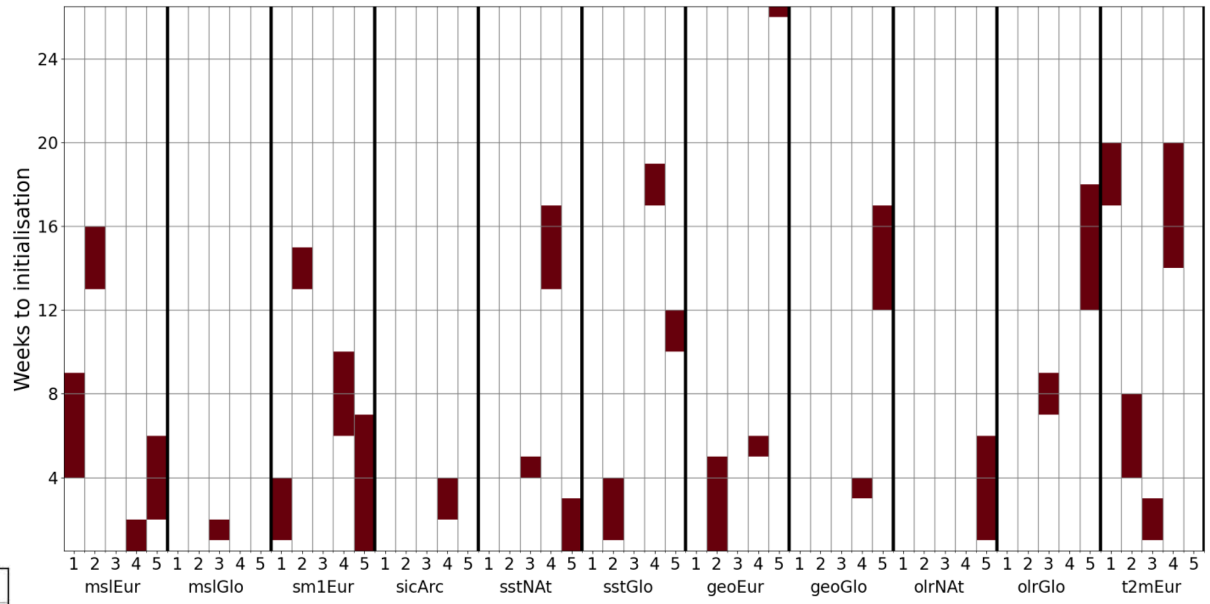
4. Potential improvement to corridor of low skill over northern Europe to Scandinavia.

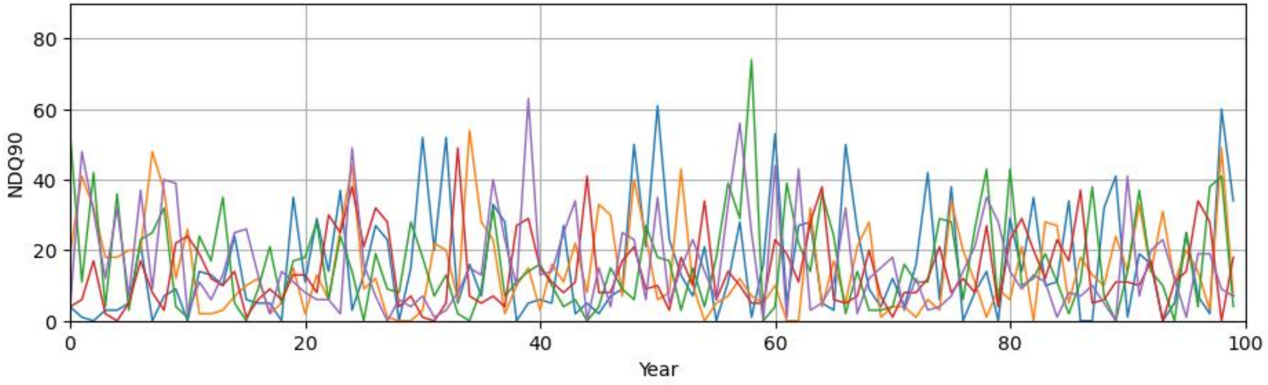
5. Ongoing work: data-driven seasonal forecast system currently undergoing **benchmarking/sensitivity analysis** *(McAdam et al.; in review)* : including deep learning models, SHAP analysis…

CLINT
CLIMATE INTELLIGENCE

METHOD: Validation set-up

1. Selection of potential predictors.



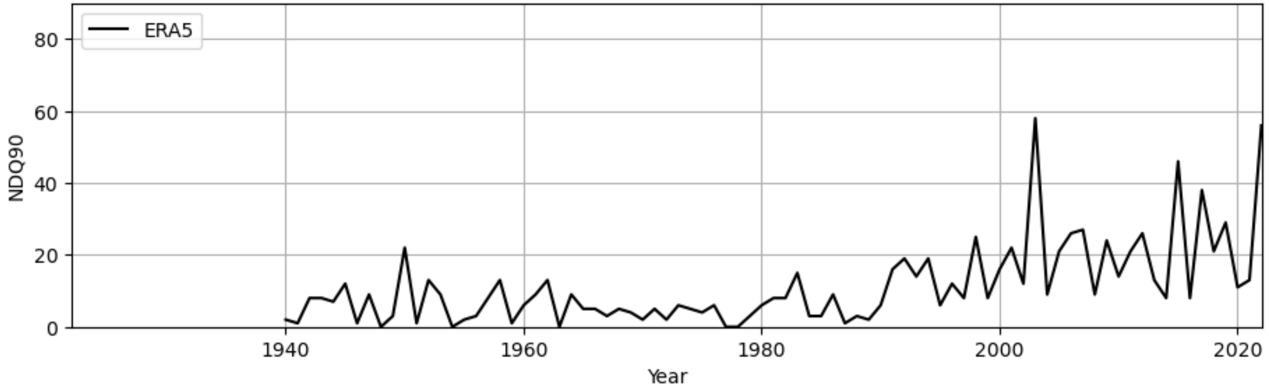


2. Training in model world (0-1600) – Testing in model world (1600-1850)

Can reduced dimensionality approach recreate HW index?

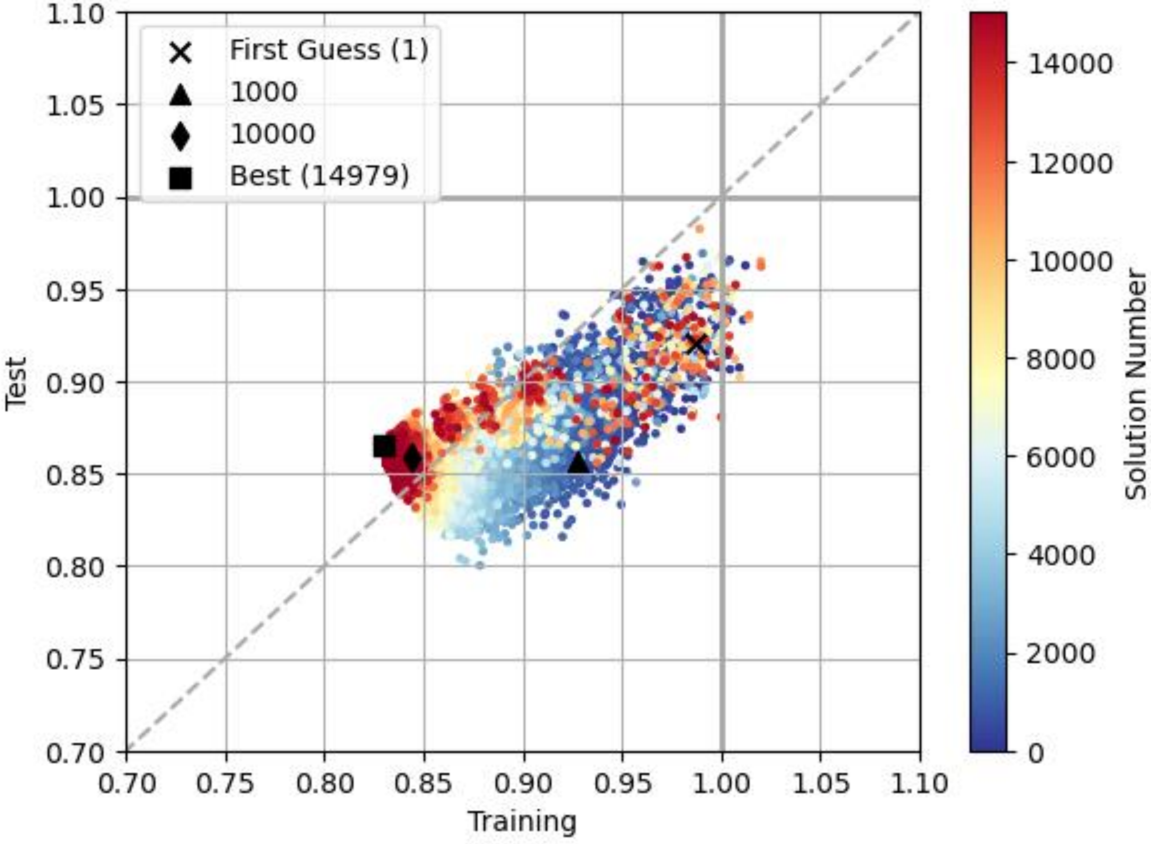
3. Training in model world (0-1850) – Testing in real world (1993-2016, ERA5)

Can we transfer learning to the real world?

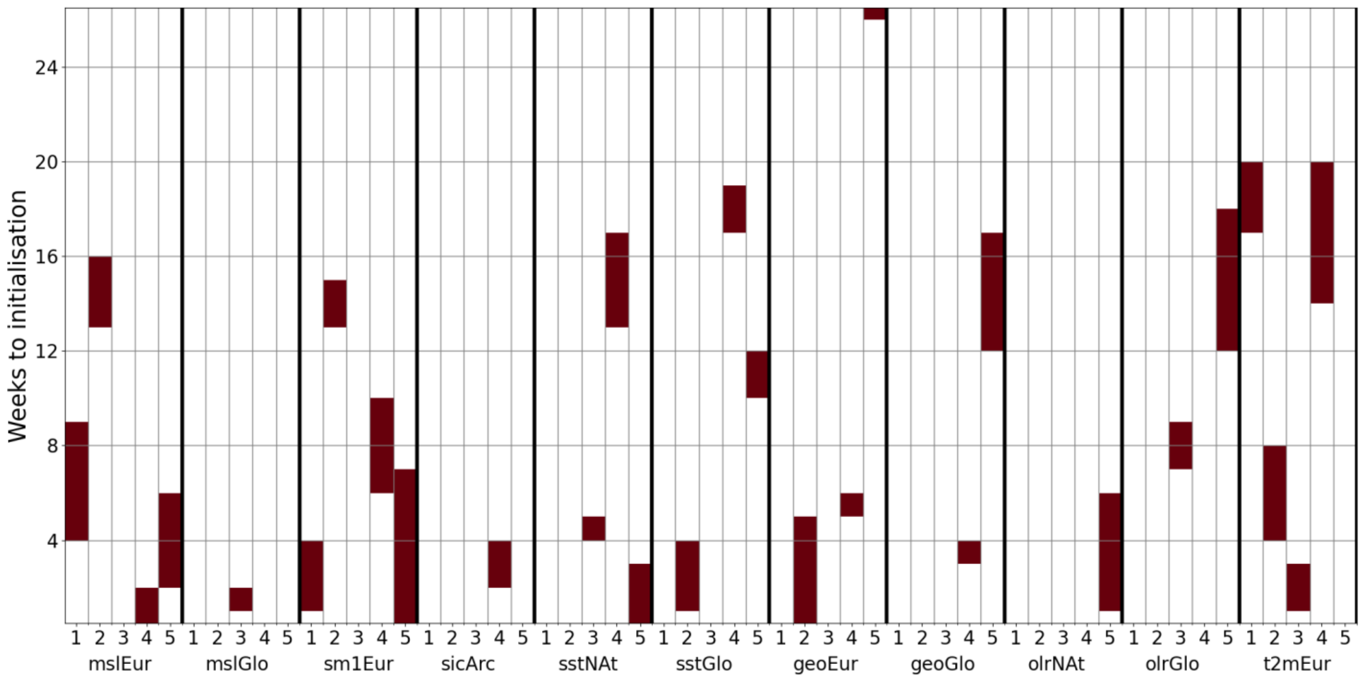


RESULTS: Optimising data-driven forecasts (example over greece)

Evolution of optimal combination of predictors (minimum N-RMSE)



"Heatmap" of optimal solutions ("best")



CLINT
CLIMATE INTELLIGENCE

RESULTS: Forecasts

Evolution of optimal combination of predictors (minimum N-RMSE)

"Heatmap" of optimal solutions ("best")

References / Reading List

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R.F., Pegenaute, F., Herrmann, F.R., Robine, J.M., Basagaña, X., Tonne, C., Antó, J.M. and Achebak, H., 2023. Heat-related mortality in Europe during the summer of 2022. *Nature medicine*, *29*(7), pp.1857-1866. https://doi.org/10.1038/s41591-023-02419-z

Barriopedro, D., García-Herrera, R., Ordóñez, C., Miralles, D.G. and Salcedo-Sanz, S., 2023. Heat waves: Physical understanding and scientific challenges. Reviews of Geophysics, 61(2), p.e2022RG000780. https://doi.org/10.1029/2022RG000780

European Environmental Agency, Climate change, impacts and vulnerability in Europe 2012 – Summary, https://www.eea.europa.eu/publications/climate-impacts-and-vulnerability-2012/climate-change-impacts-and-vulnerability/view

García-Herrera, R., Díaz, J., Trigo, R.M., Luterbacher, J. and Fischer, E.M., 2010. A review of the European summer heat wave of 2003. *Critical Reviews in Environmental Science and Technology*, *40*(4), pp.267-306. https://doi.org/10.1080/10643380802238137

Hobday, A.J., Alexander, L.V., Perkins, S.E., Smale, D.A., Straub, S.C., Oliver, E.C., Benthuysen, J.A., Burrows, M.T., Donat, M.G., Feng, M. and Holbrook, N.J., 2016. A hierarchical approach to defining marine heatwaves. *Progress in oceanography*, *141*, pp.227-238. https://doi.org/10.1016/j.pocean.2015.12.014

Kornhuber, K., Coumou, D., Vogel, E., Lesk, C., Donges, J.F., Lehmann, J. and Horton, R.M., 2020. Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions. *Nature Climate Change*, *10*(1), pp.48-53. https://doi.org/10.1038/s41558-019-0637-z

Perkins, S.E., 2015. A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, *164*, pp.242-267. https://doi.org/10.1016/j.atmosres.2015.05.014

Prodhomme, C., Materia, S., Ardilouze, C., White, R.H., Batté, L., Guemas, V., Fragkoulidis, G. and García-Serrano, J., 2021. Seasonal prediction of European summer heatwaves. *Climate Dynamics*, pp.1-18. https://doi.org/10.1007/s00382-021-05828-3

Russo, S., Sillmann, J. and Fischer, E.M., 2015. Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters*, *10*(12), p.124003. **DOI** 10.1088/1748-9326/10/12/124003

Russo, S., Sillmann, J. and Sterl, A., 2017. Humid heat waves at different warming levels. *Scientific reports*, *7*(1), p.7477. https://doi.org/10.1038/s41598-017-07536-7

Schaller, N., Sillmann, J., Anstey, J., Fischer, E.M., Grams, C.M. and Russo, S., 2018. Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles. *Environmental Research Letters*, *13*(5), p.054015. **DOI** 10.1088/1748-9326/aaba55

Torralba, V., Materia, S., Cavicchia, L., Álvarez-Castro, M.C., Prodhomme, C., McAdam, R., Scoccimarro, E. and Gualdi, S., 2024. Nighttime heat waves in the Euro-Mediterranean region: definition, characterisation, and seasonal prediction. *Environmental Research Letters*, *19*(3), p.034001. 10.1088/1748-9326/ad24cf