

# Feature Selection for Data-Driven Seasonal Forecasts of Heatwaves

Ronan McAdam<sup>1</sup>, Jorge Pérez-Aracil<sup>2</sup>, Antonello Squintu<sup>1</sup>, Fabio Merizzi<sup>3</sup>, Cesar Peláez-Rodríguez<sup>2</sup>, Thomas Noel<sup>3</sup>, Guilhem Vignal<sup>3</sup>, Harilaos Loukos<sup>3</sup>, Enrico Scoccimarro<sup>1</sup>

<sup>1</sup>CMCC Foundation - Euro-Mediterranean Centre on Climate Change, Italy

<sup>2</sup>University of Alcalá, Spain

<sup>3</sup>The Climate Data Factory, France



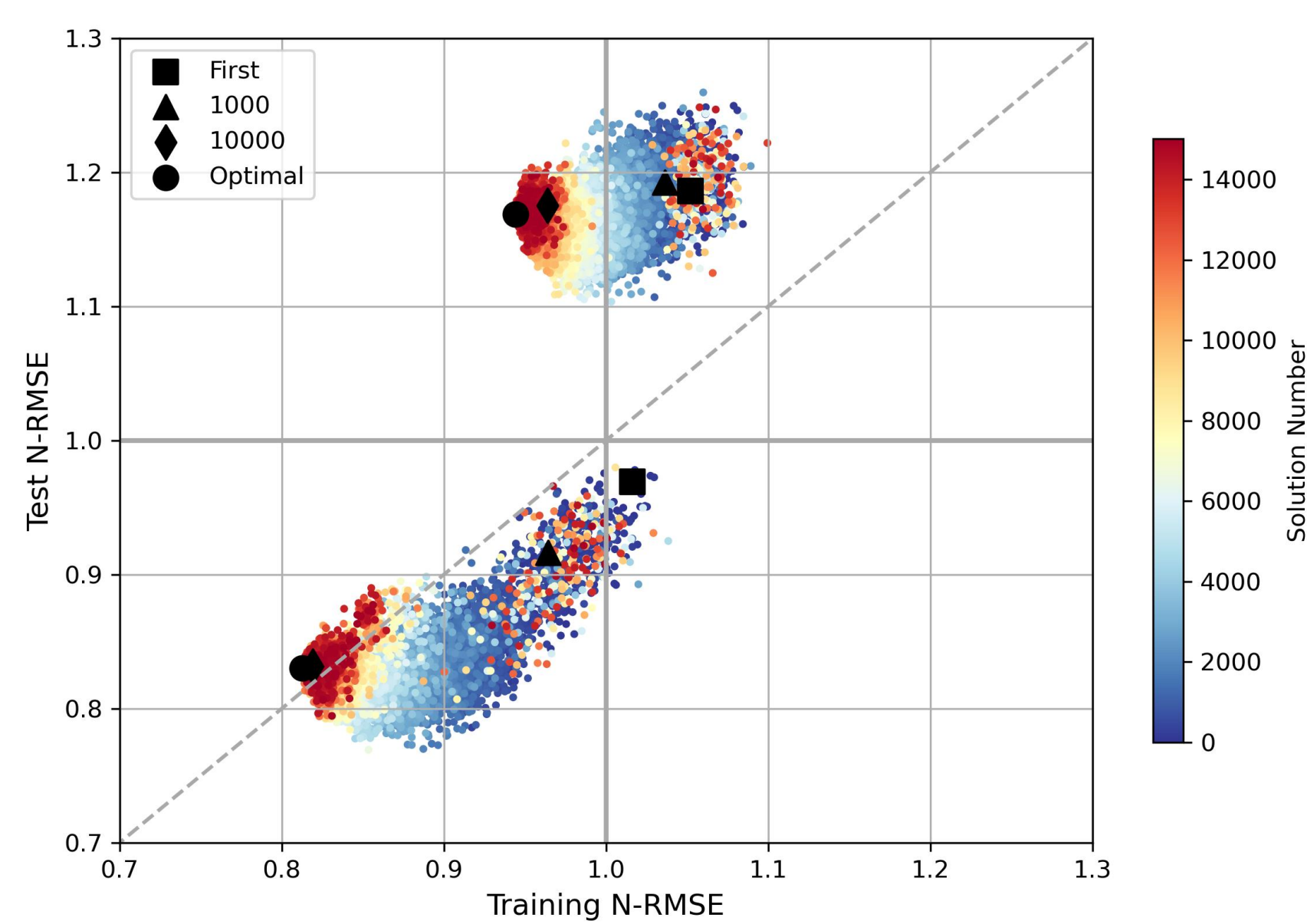
Scan here for the paper

## Key Takeaways

- **Data-driven seasonal forecasts of summer heatwaves** are shown to match, and in places outperform, the skill of the state-of-the-art dynamical systems.
- **Feature Selection** highlights that the greatest contribution to skill comes from predictors at 4-7 weeks time-lag (e.g. mid-March), and identifies predictors which can form the basis for future studies on mechanisms.
- Forecasts are trained purely on a paleo-climate simulation and large dimensionality-reduction of potential predictors.

## 1. Method: Optimisation & Training

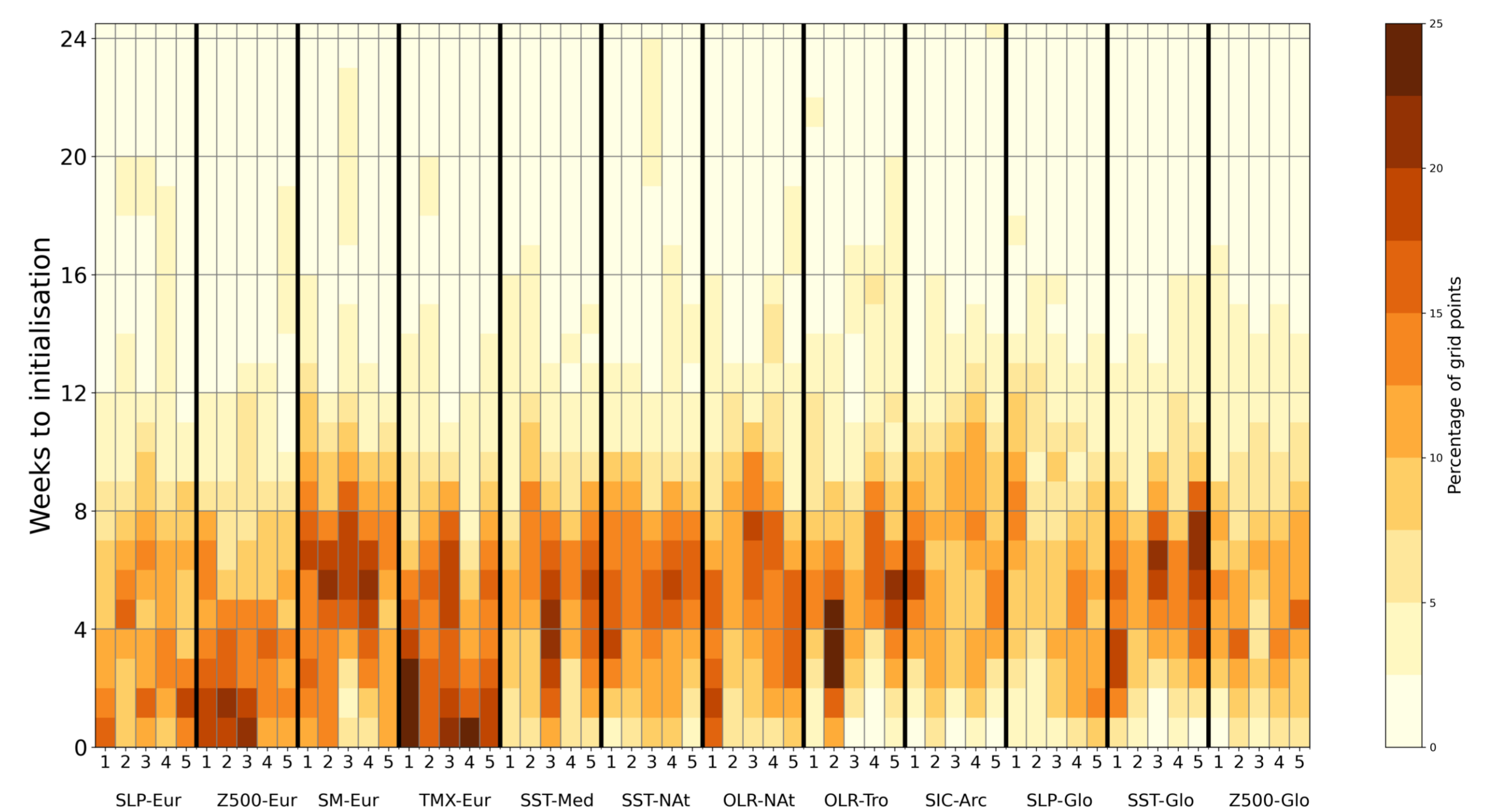
- Target: Number of Heatwave days in May to July.
- Predictors: K-means clustering of soil moisture, 500hPA geopotential height, sea surface temperature, outgoing longwave radiation and more (see paper for details).
- Training: MPI-ESM paleo-climate reconstruction “past2k” covering years 0-1850.
- Feature Selection: Probabilistic Coral Reef Optimisation With Substrate Layers (PCRO-SL).
- PCRO-SL tests thousands of combinations of predictors in a Linear Regression model (training 0-1600, test 1601-1850).



Optimisation of seasonal forecast skill (normalised-RMSE) in paleo-climate simulation for two example points.

## 2. Feature Selection for optimal heatwave predictors

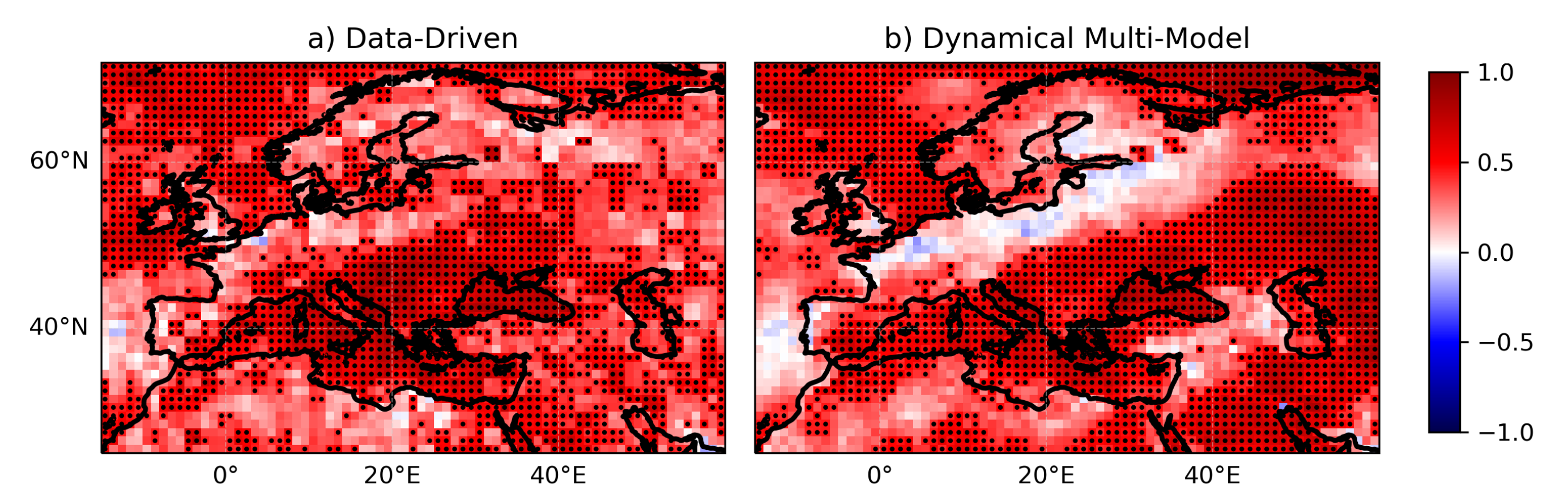
- Identification of optimal heatwave predictors across Europe.
- Peak predictability for summer heatwaves comes from March (e.g. soil moisture and sea surface temperature).



Most selected heatwave predictors across Europe.

## 3. Seasonal Forecasts

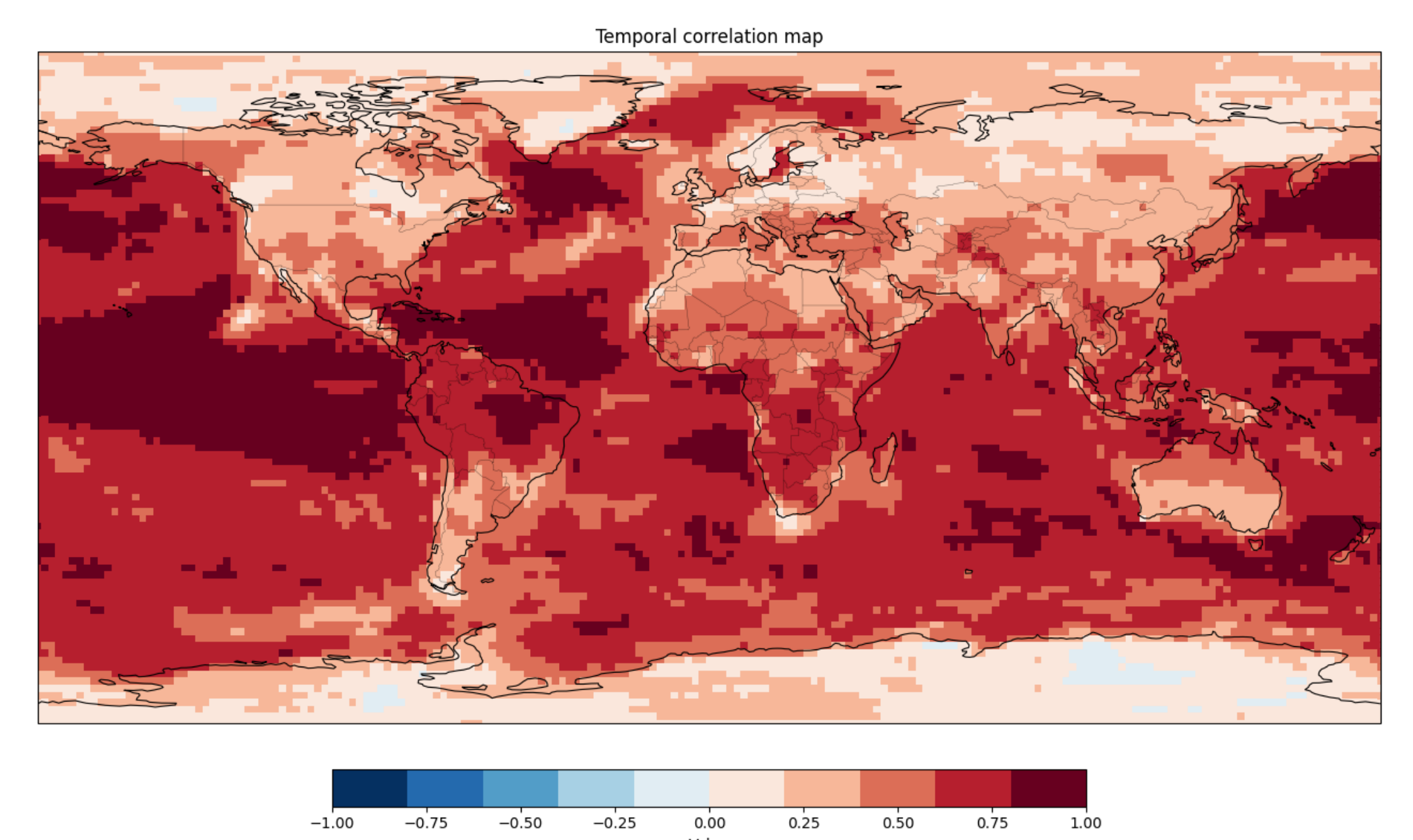
- Application to ERA5 predictors (training 0-1850 past2k, test 1993-2016 ERA5).
- A mix of ML models (Random Forest, Multi-Layer Perceptron) are used and the best performing model for each grid cell is taken.
- Pattern and magnitude of summer heatwave seasonal forecast skill similar between data-driven and dynamical forecasts.



Correlation skill between ERA5 and seasonal forecasts. Dynamical multi-model consists of ensemble mean of Meteo-France9, ECMWF5.1, CMCC3.5 & DWD2.1.

## Next Steps: Global Deep Learning Heatwave Forecasts

- Extend the current framework to a fully deep learning approach, while keeping the same predictors and lead times up to 12 weeks.
- Train on Past2k (0-1700) across all weeks of the year and the full global domain.
- Use a U-Net with separate encoders for each predictor variable.
- Assess performance on May-July heatwave days in the period 1700-1850.
- Subseasonal version being developed for ECMWF AI WeatherQuest competition.



Spatial map of temporal correlation for the deep learning model over the Past2k test period (1700-1850)