Blending dynamical and empirical information in seasonal forecasts of heat waves

Stefano Materia [1], Martin Jung [2], Markus Donat [1], Lluis Palma [1], Amanda Duarte [1]

[1] Barcelona Supercomputing Center

[2] Max Planck Institute for Biogeochemistry, Jena

02/07/2024

Workshop on Objective Seasonal Forecast

Motivation

- Dynamical seasonal prediction skill of **heat waves** is still rather low in mid-latitudes, and with particular regard in Europe, and often associated to the global warming trend (Prodhomme et al., 2022).
- Statistical techniques, also based on **machine learning** algorithms, have recently demonstrated that **improving dynamical climate prediction** is possible (Cohen et al., 2019; Van Straaten et al., 2022).
- Many local drivers of seasonal predictability, such as **snow cover**, **vegetation** and **land-use** can enhance seasonal prediction, but have often been overlooked.

Motivation

- Dynamical seasonal prediction skill of **heat waves** is still rather low in mid-latitudes, and with particular regard in Europe, and often associated to the global warming trend (Prodhomme et al., 2022).
- Statistical techniques, also based on **machine learning** algorithms, have recently demonstrated that **improving dynamical climate prediction** is possible (Cohen et al., 2019; Van Straaten et al., 2022).
- Many local drivers of seasonal predictability, such as **snow cover**, **vegetation** and **soil moisture** can enhance seasonal prediction, but have often been overlooked.

Using a data-driven approach:

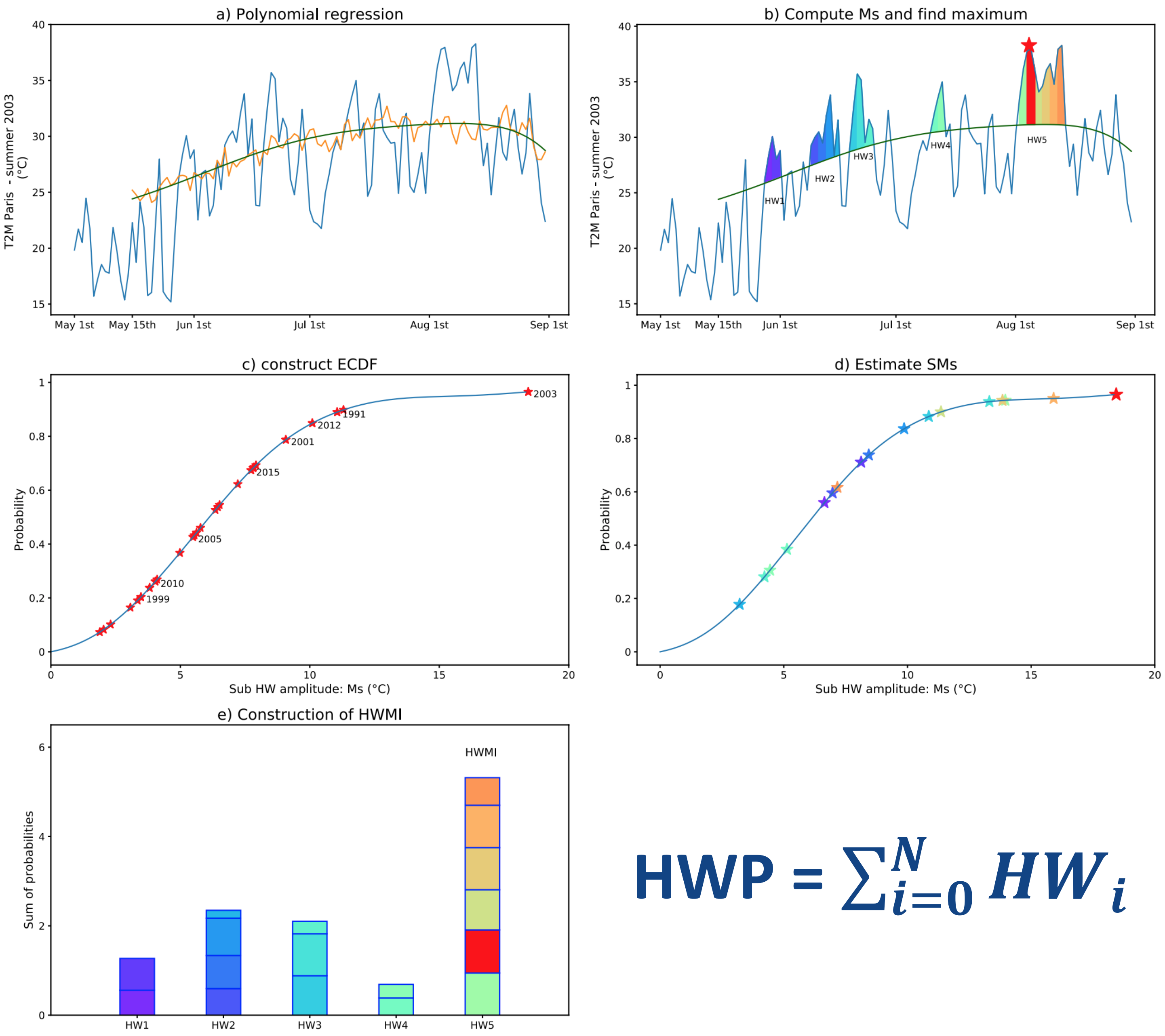
- Can we improve seasonal forecast of heat extremes in key midlatitude regions?
- Can we identify drivers overlooked by dynamical models?

Identification of the target (predictand)

While predicting a specific heat wave at the **seasonal scale** is impossible, seasonal forecasts may have skill for ***heat propensity***, the tendency of a season to be predisposed to the occurrence of an event (Prodhomme et al., 2022).

Heat Wave Propensity (HWP)

integrates in one value **Duration, Intensity, Frequency** of heat waves

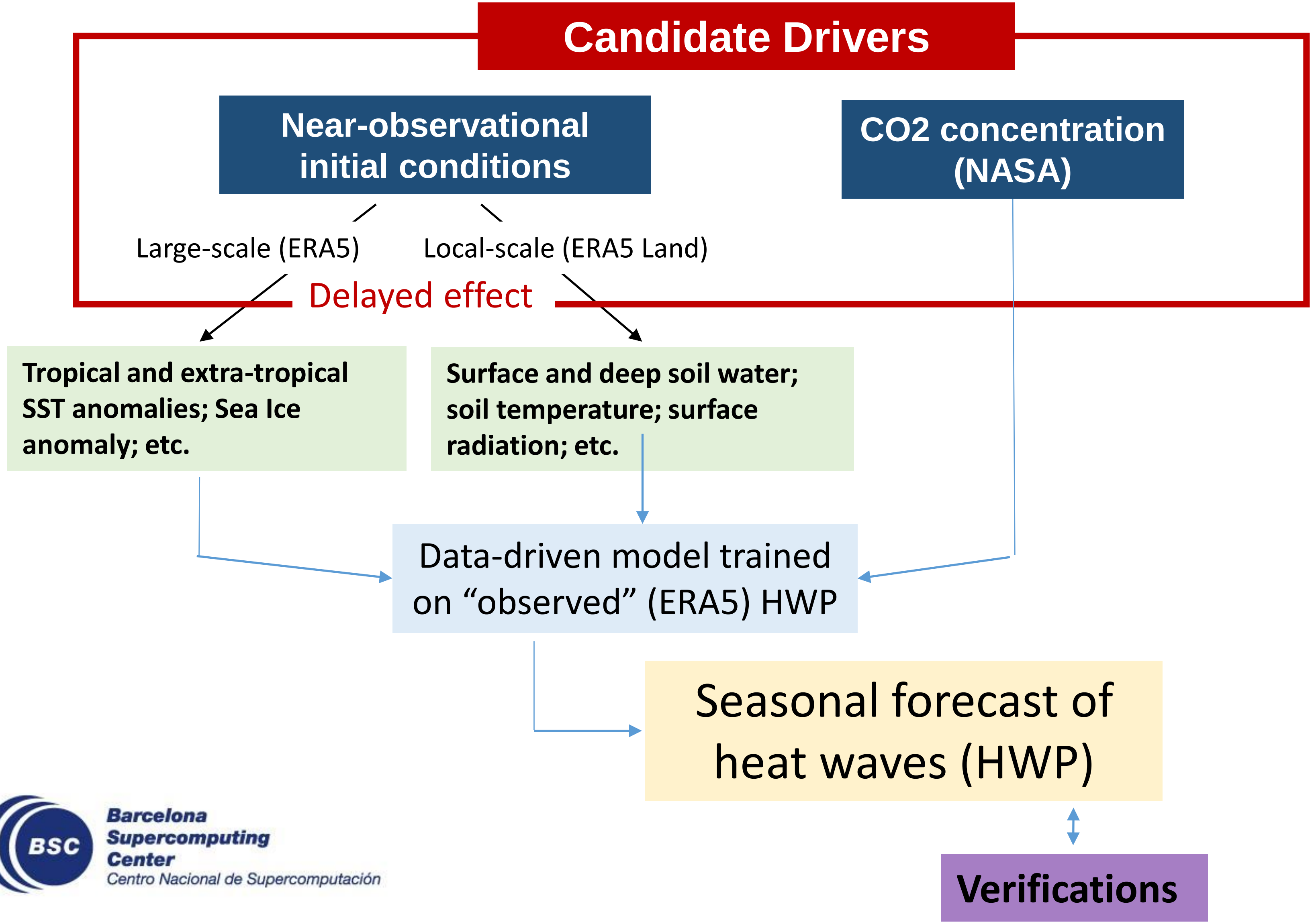


$$\textbf{HWP} = \sum_{i=0}^{N} HW_i$$

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

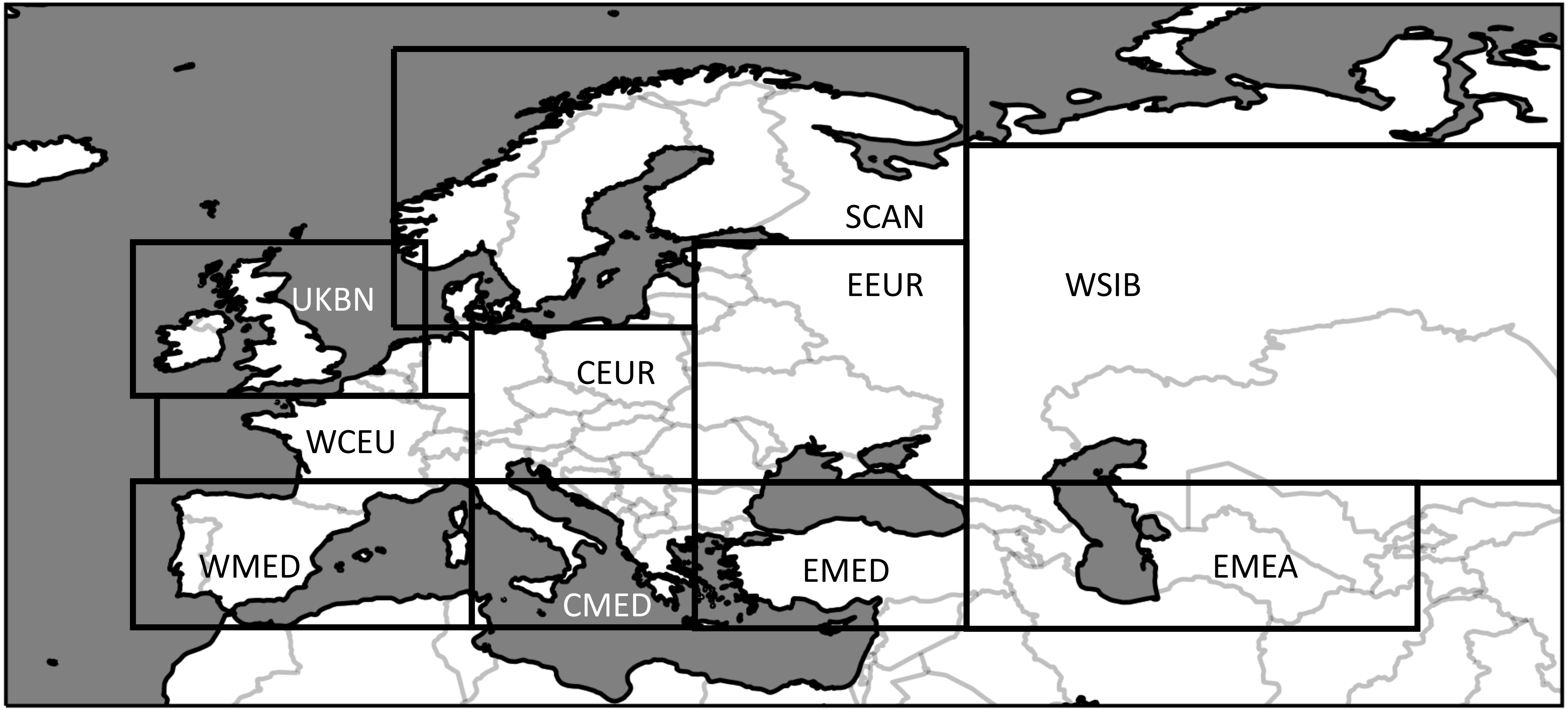
Selection of potential drivers of predictability

Prediction of heat wave propensity (HWP), the tendency of a season to be predisposed to the occurrence of heat waves. The HWP index integrates duration, intensity and frequency within the season



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

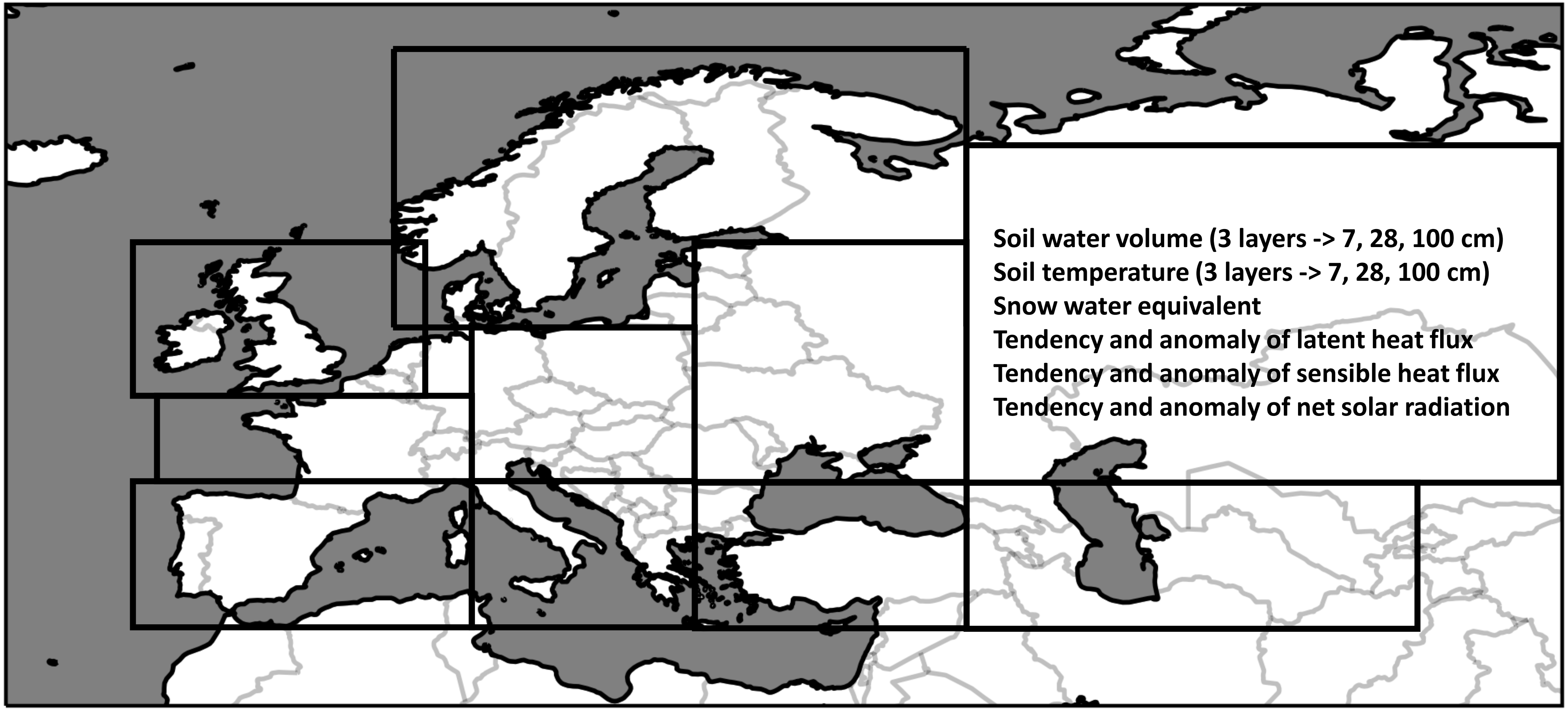
Target: HWP in ten European regions



ERA5

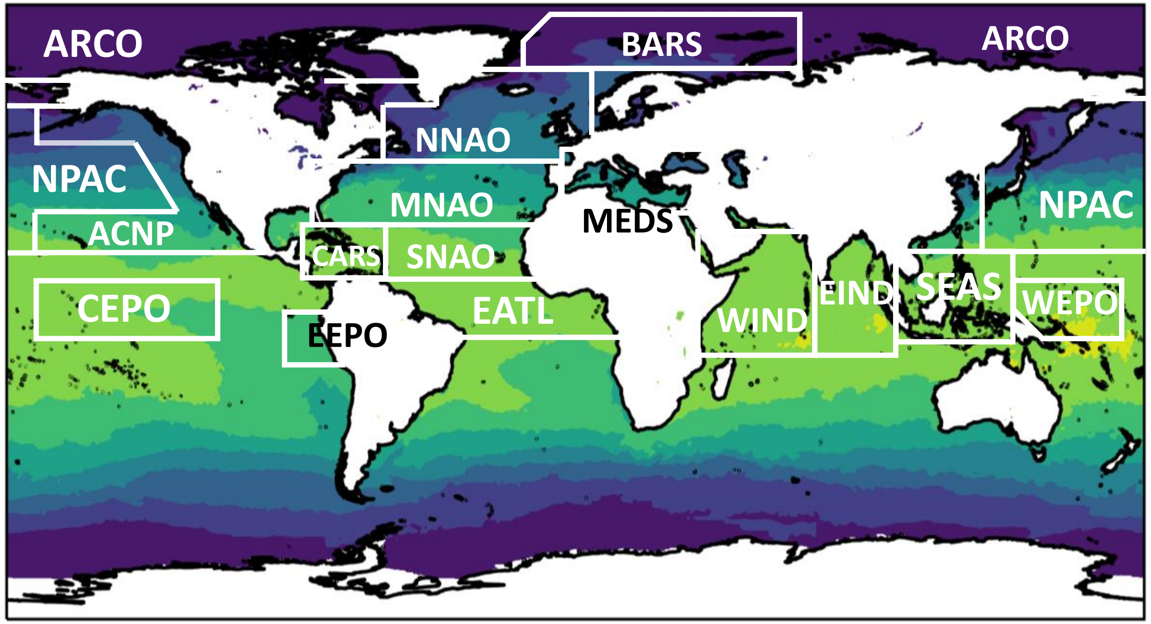
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

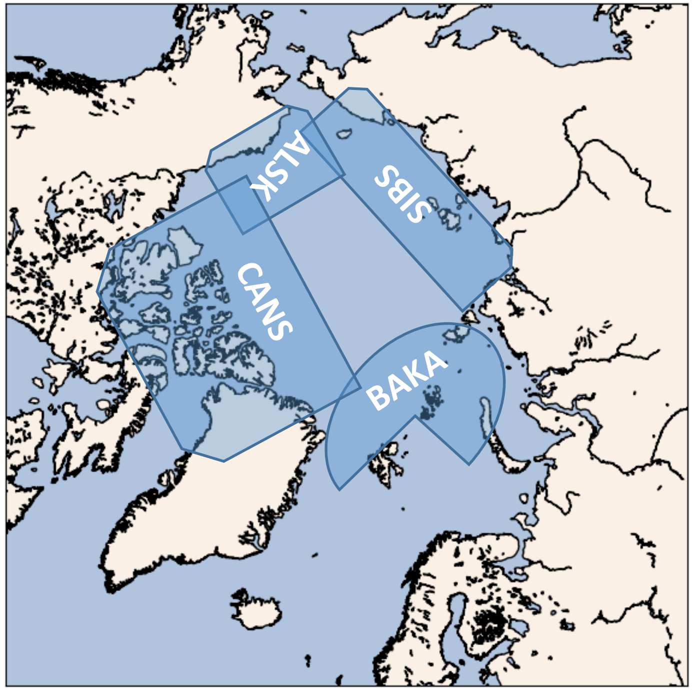
Local predictors: land surface parameters



Each region has its own set of land surface candidate predictors.
Number of candidates: 13. **Time lag is 15 days**

Remote predictors: SST, SI concentration



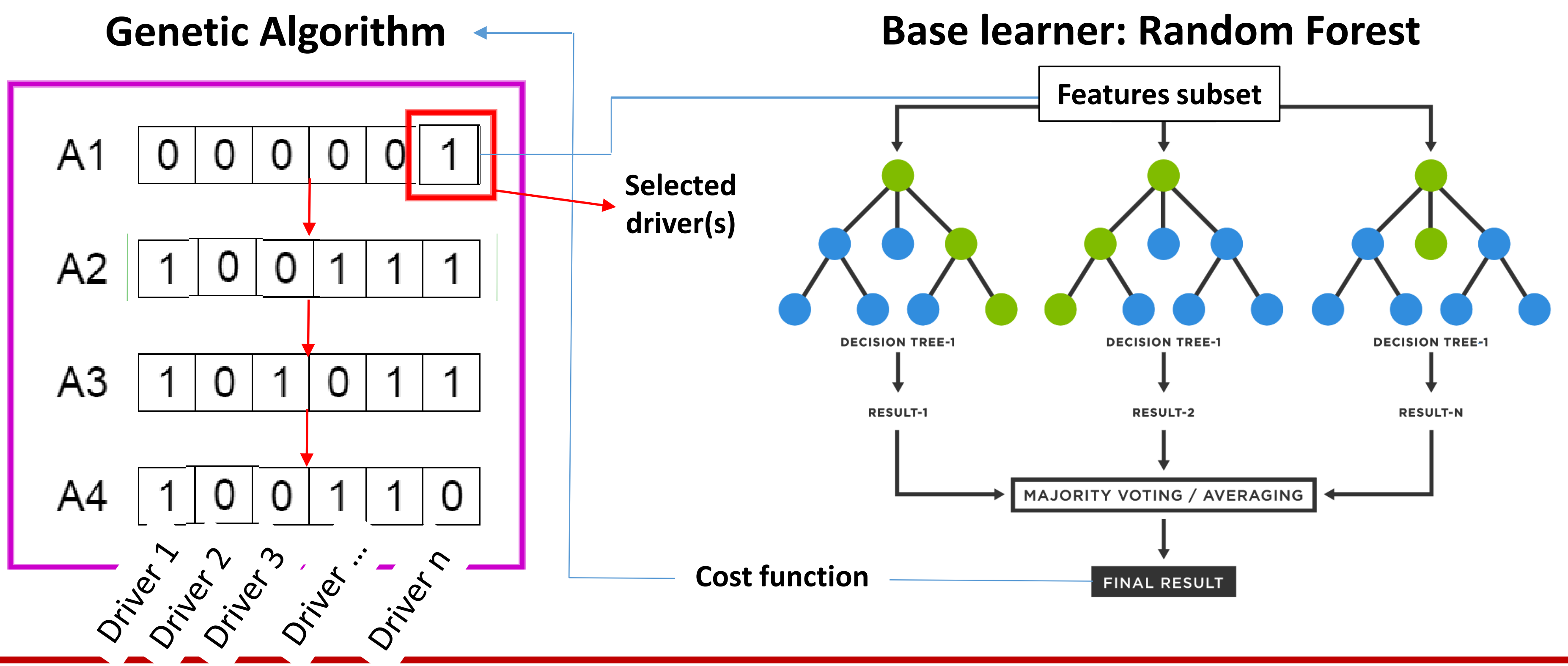


Remote candidate drivers (**time-lag 1 month**)
SST from 18 ocean regions
Sea Ice concentration from 4 Arctic regions

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ML detection of predictability drivers of hot extremes at seasonal scale

A **Feature selection algorithm** that recursively extracts a subset of candidate predictors, wrapped around a **base learner** to predict HWP.



convp = 4

nrestarts = 12

ntrees = 200

All the models with a **cost function <= (MIN cost function + STD cost function)** are considered BEST models and retained

Model skill in terms of r^2

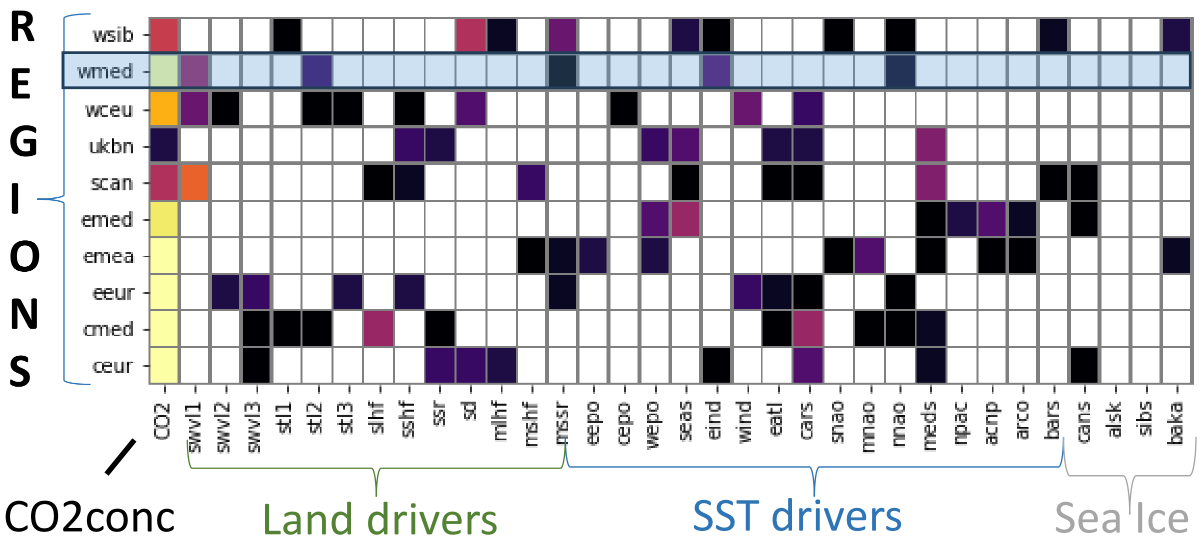
	MAM	JJA	SON	DJF
ceur	0.06	0.22	0.17	-0.19
cmed	0.07	-0.08	0.03	-0.02
eeur	0.20	0.17	0.10	-0.12
emea	0.46	0.38	0.14	0.01
emed	0.04	-0.06	0.01	0.16
scan	0.39	0.18	-0.00	-0.03
ukbn	-0.05	-0.20	0.00	-0.14
wceu	0.12	0.37	0.21	-0.08
wmed	0.05	0.40	-0.09	-0.06
wsib	-0.17	0.19	0.02	-0.26

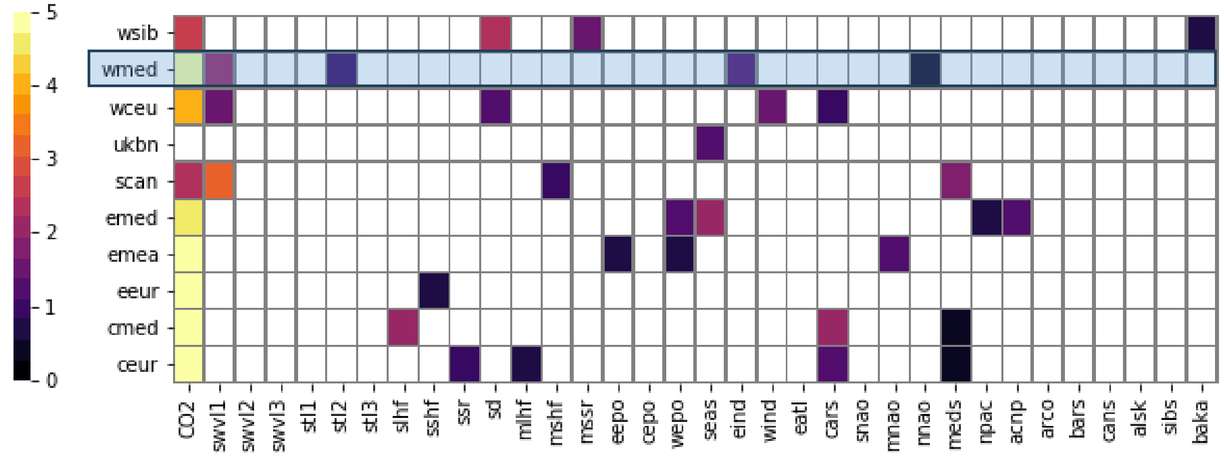
We run the model in cross-validation mode for all the regions in each of the four main seasons

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Summer in Western Mediterranean

- **Each best model has its own subset of features**. Some features are never selected, others are only selected a few times, others are selected by all the best models. The more represented features are generally those with **more importance** for the predictions of European HWP**.**
- **Importance** is a measure given by the **SHAP values (XAI)**, that represent **how much each feature contributes** to the predicted value of the target.

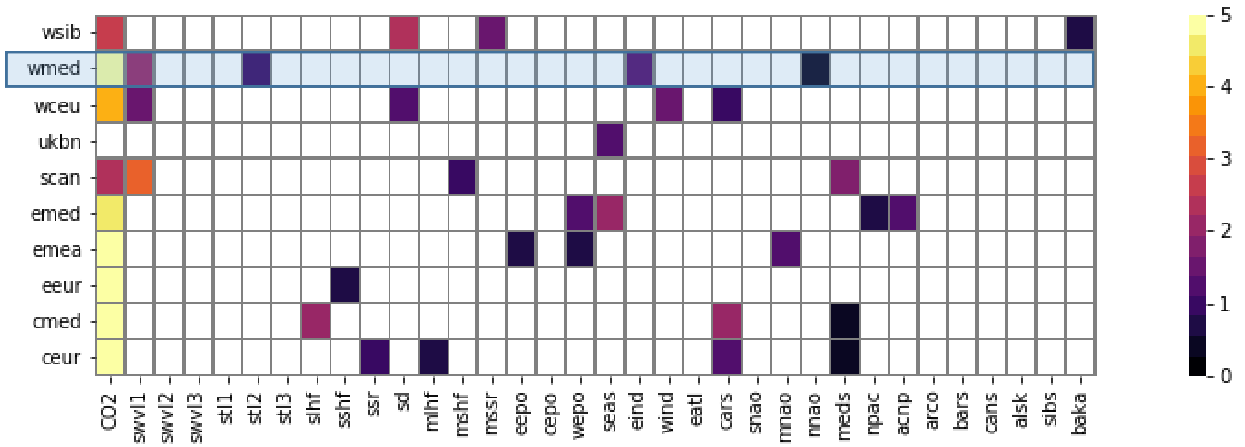




Colors indicate how much each feature contributes to the predicted value of HWP: the higher the number, **the more important the driver is for the HWP prediction**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Western Mediterranean during the JJA season (r2 = 0.4)

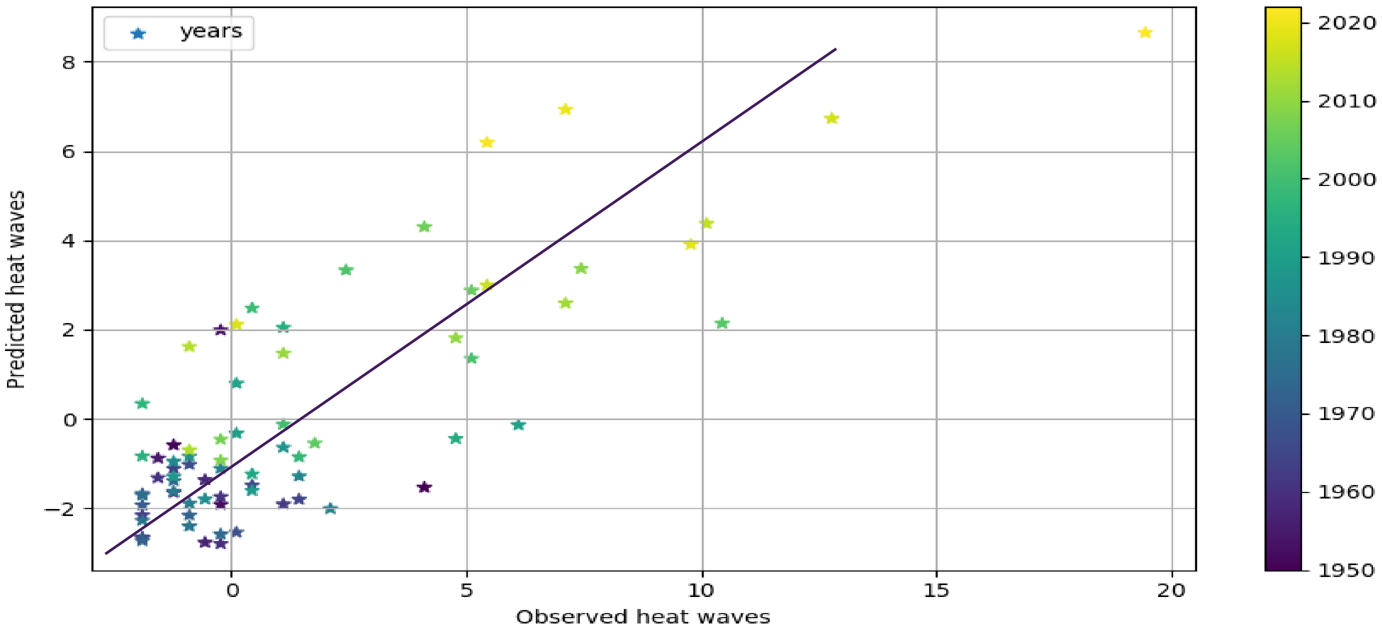


The **global CO2 concentration is the main driver of heat waves**, and this it is also true in almost every European region during summer.

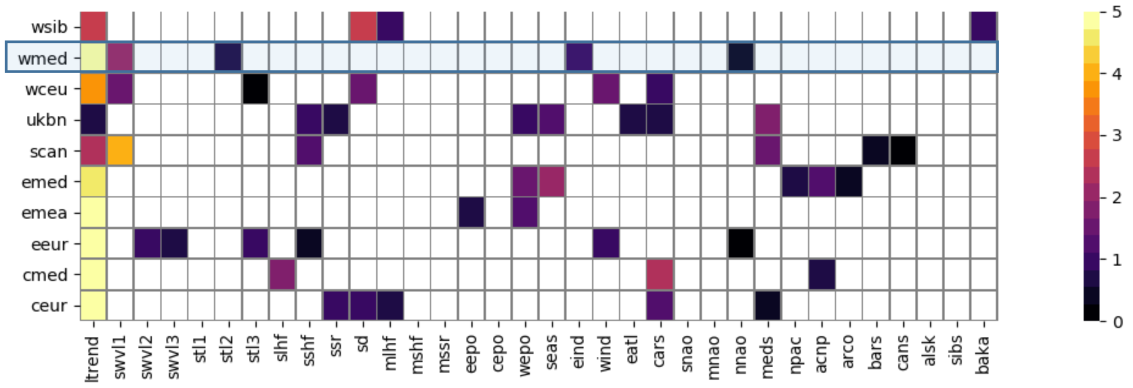
Surface soil moisture and soil temperature also play an important role, with mean absolute shap values higher than 1.5

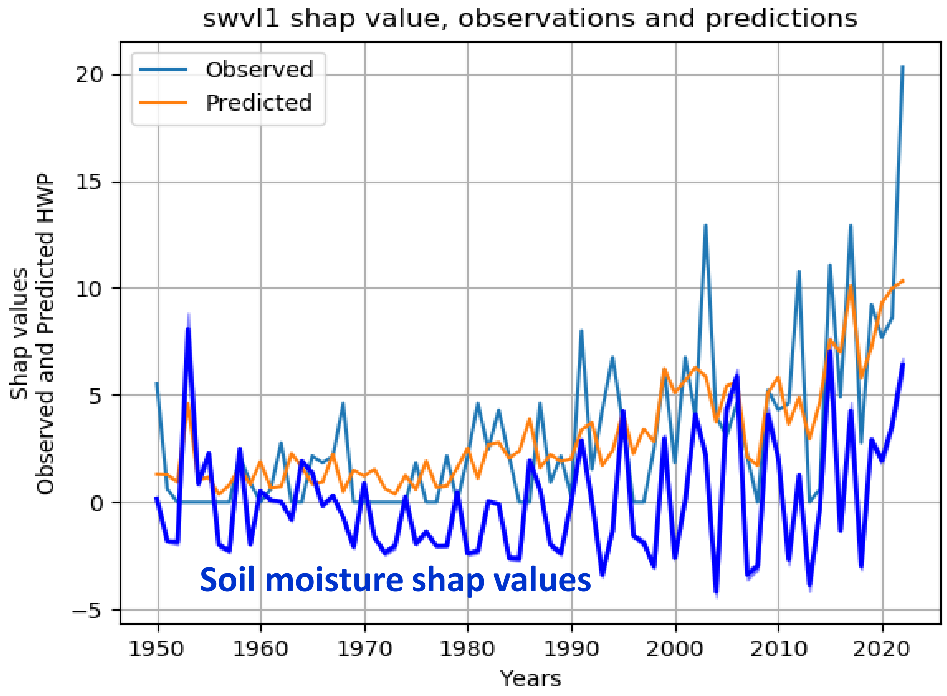
SST in Eastern Indian ocean and Northern North Atlantic also play a role

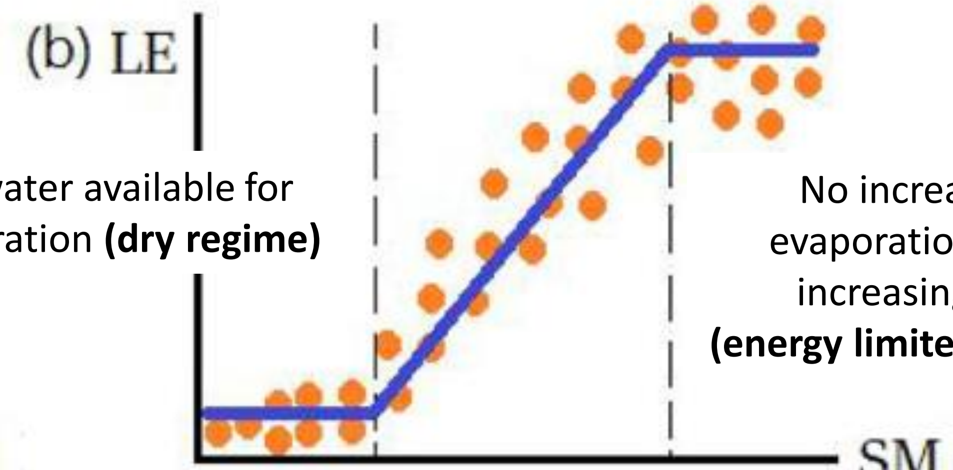
The model explains a good portion of the observed Western Mediterranean HWP variance during summer.



The role of soil moisture







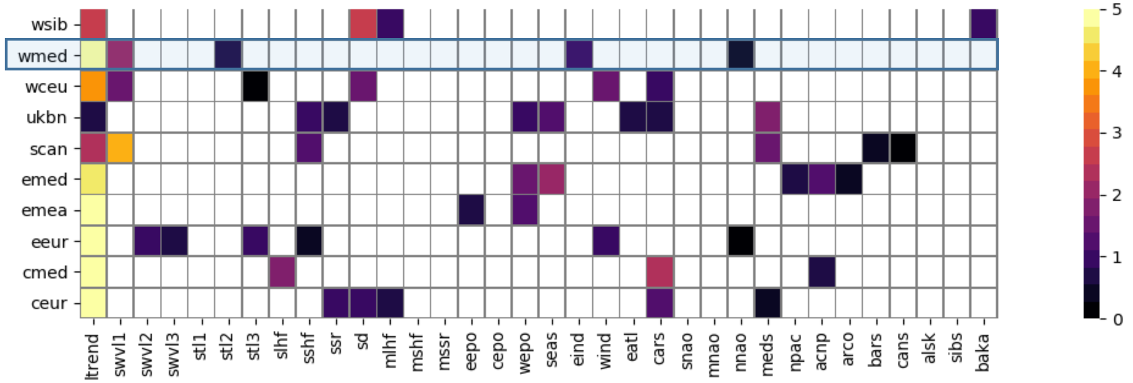
No water available for evaporation **(dry regime)**

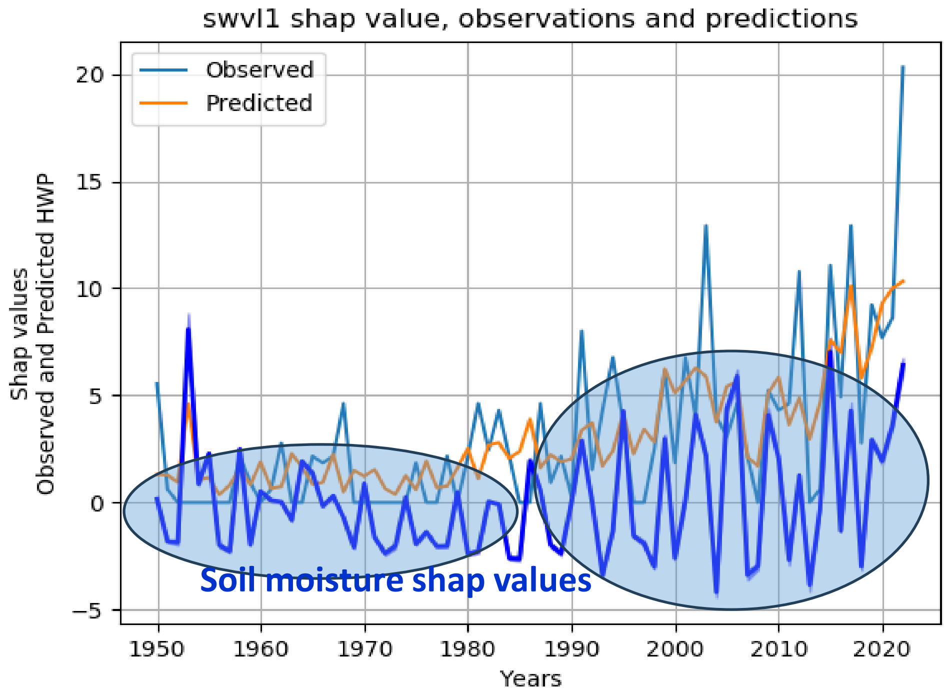
No increasing evaporation with increasing SM **(energy limited regime)**

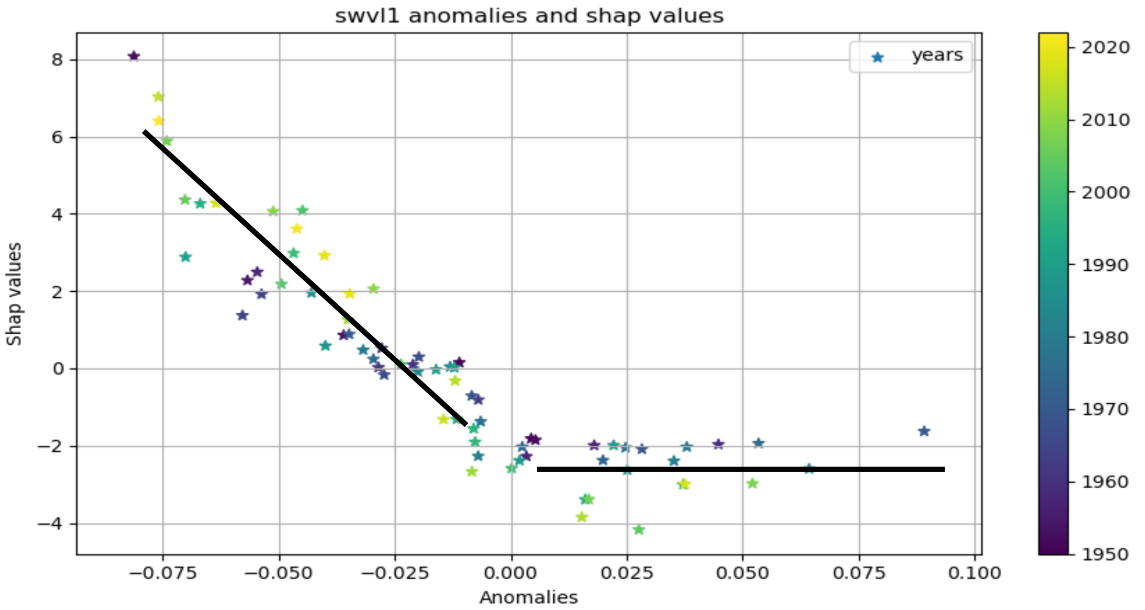
Wilting Point

Critical threshold

The role of soil moisture



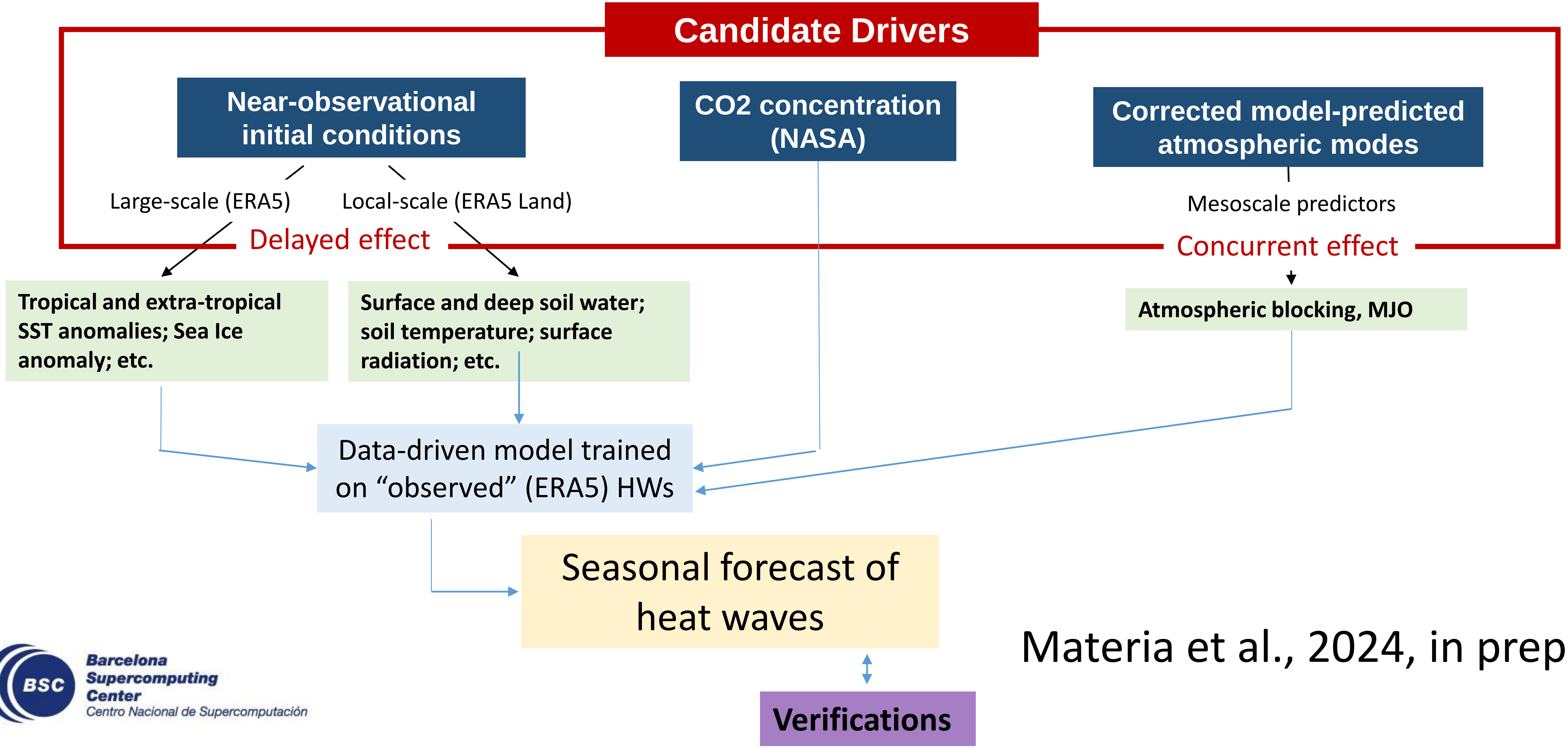




- It seems that soil moisture was not playing a big role in the modulation of HWP until the late eighties, while importance becomes stronger later on (higher variability of SM importance)
- Possible explanation: spring soil moisture is becoming drier due to higher temperatures (and possibly lower precip). An exacerbation of the transitional evaporative regime is taking place in Western Mediterranean during the summer.
- Shap values indicate **greater importance** when antecedent soil conditions are **very dry**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Next step: hybridization using seasonal forecast as a third class of inputs



Materia et al., 2024, in prep.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions

- We built a simplified **machine learning architecture to predict Heat Wave Propensity** in Europe, based on a feature selection algorithm and a decision tree that concur to identify the most important precursors and predict heat waves.
- **Summer** and (at a lesser extent) **spring** are the seasons better predicted. The data-driven system generally performs better in southern Europe, characterized by lower summer variability. The identified drivers are ranked by importance using the SHAP values explainable AI.
- The western Mediterranean has one of the best skill and the identified drivers include the background trend of **increasing GHGs, soil moisture and temperature, and SST in the Indian Ocean and North Atlantic**.
- The role of **soil moisture** appears particularly interesting, since it seems its contribution to summer heat waves has increased in importance in the last 35 years, meaning that drier springs have increased the **soil moisture control over evaporation** during summer.

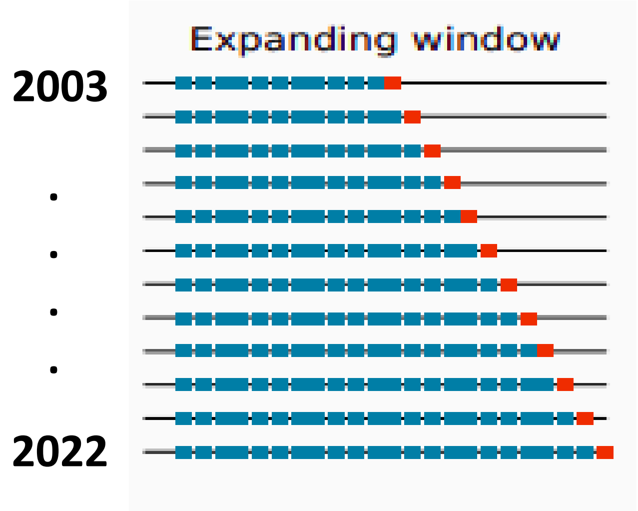
- Materia S., Jung M., Palma L., Duarte A., Weber U., Samso M., Donat M.G. **Data-driven Seasonal Forecasts of European Heat Waves.** *In preparation*
- Materia, S. et al. (2023). **Artificial Intelligence for Prediction of Climate Extremes: State of the art, challenges and future perspectives.** *arXiv preprint arXiv:2310.01944*.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Data-driven seasonal forecast of heat waves in Europe and Western Mediterranean

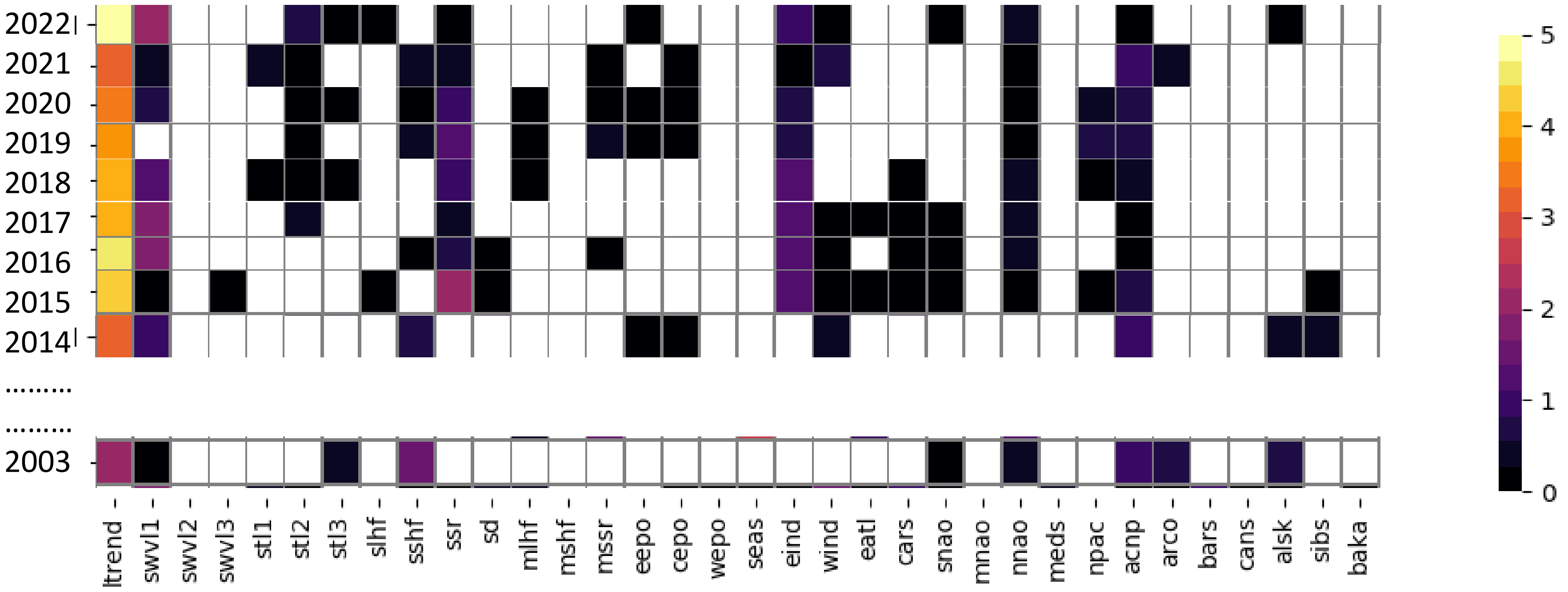
stefano.materia@bsc.es

Forecast evaluation



- The model is trained 19 more times, each time up to a year before the event to test - (1950-2020 to test the 2021 heat waves, etc.).
- The parameters to be used are those from the 2022 training, but the feature are re-selected all the times.
- In this way I will have 20 test years to compare with dynamical seasonal forecasts.

West Med



Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación