An Explainable AI Assessment of Soil Moisture and Atmospheric Circulation as Drivers of Heat Extremes in Seasonal Predictions

Lluís Palma[1], Arnau Garcia[1], Stefano Materia[1] & Markus Donat[1,2]

[1] Barcelona Supercomputing Center, Earth Sciences Department, Spain [2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

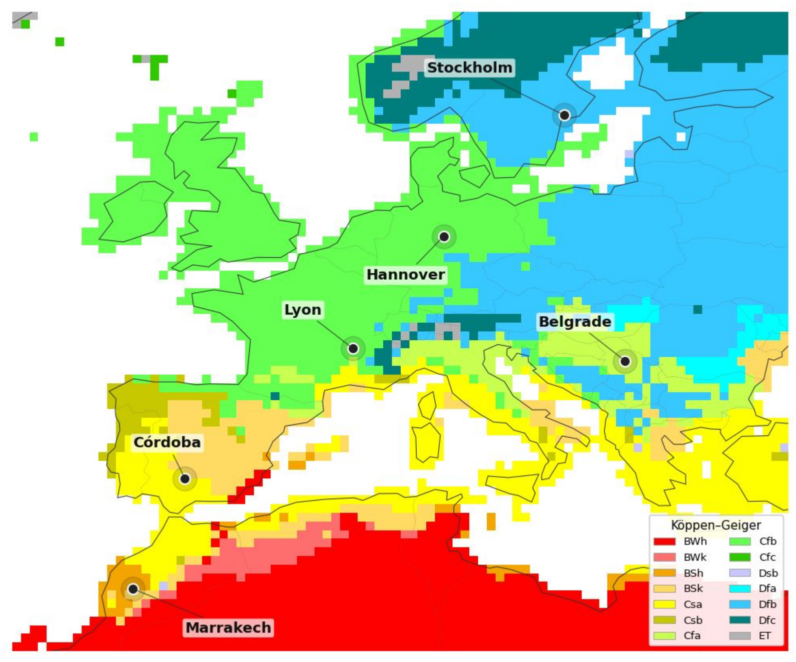
Summary

In this study, we build on [1], which developed an explainable machine-learning framework to disentangle the influences of the atmospheric circulation, soil-moisture, and rising global CO_2 concentrations on boreal-summer temperature extremes. We extend this framework to compare driver contributions between SEAS5 and ERA5, aiming to detect deficiencies in seasonal prediction systems. We investigate six locations across Europe and North Africa with varying land–atmosphere coupling. SEAS5 shows little P90 bias relative to ERA5-Land across lead times, though prediction skill relies heavily on the global warming trend. Driver–extreme relationships closely agree between ERA5(-land) and SEAS5, yet CO_2 importance shows strong biases. Spatial SHAP values reveal location-dependent circulation patterns.

1. Datasets

We leverage ERA5(-Land) reanalysis and SEAS5 seasonal forecasts at six locations with different land–atmosphere coupling regimes.

Soil moisture and sea-level pressure fields serve as local and large-scale predictors, respectively. CO2 concentration serves as proxy for the response trend in extreme temperatures.



Variable	Description	Frequency / Lag	Role
psl	Sea-level pressure anomaly	Daily, 1-day lag	CNN (50×130° grid)
swvl1	Soil moisture 0–7 cm	Daily, lag 1–7 avg	MLP (site point)
swvl2	Soil moisture 7–28 cm	Daily, lag 1–7 avg	MLP (site point)
swvl3	Soil moisture 28–100 cm	Daily, lag 1–7 avg	MLP (site point)
CO_2	Atmospheric CO_2 (JJA mean)	Seasonal	MLP (global scalar)
Tmax	Daily max. temp. > p90	Daily	Binary target

Source ECMWF SEAS5 — 1°×1°, 25 ensemble members pooled

Train / Test 1981-2015/ 2016-2024 · Lead months 4–7 (Apr–Jul → JJA)

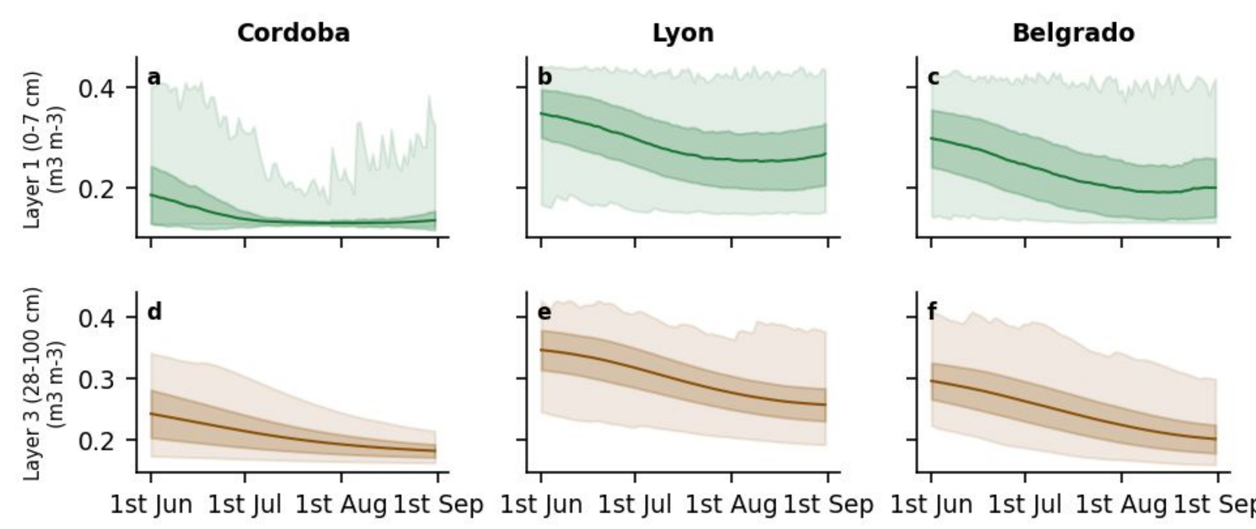
Sites 6 locations across distinct Köppen–Geiger climate zones

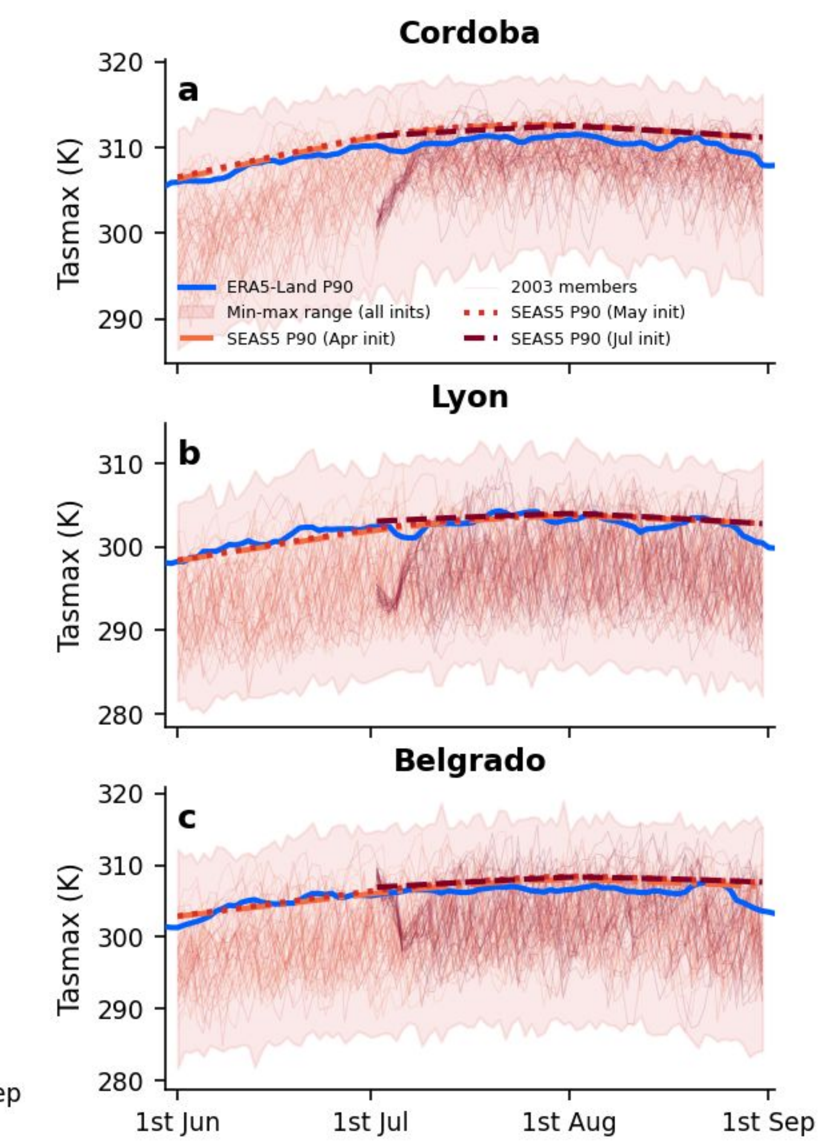
2. Methods

Daily std. anamolaies: μ and σ grouped by day-of-year over the reference period, LOESS-smoothed to remove noise [2]. ERA5 pools across years, SEAS5 pools across years and ensemble members (not init months).

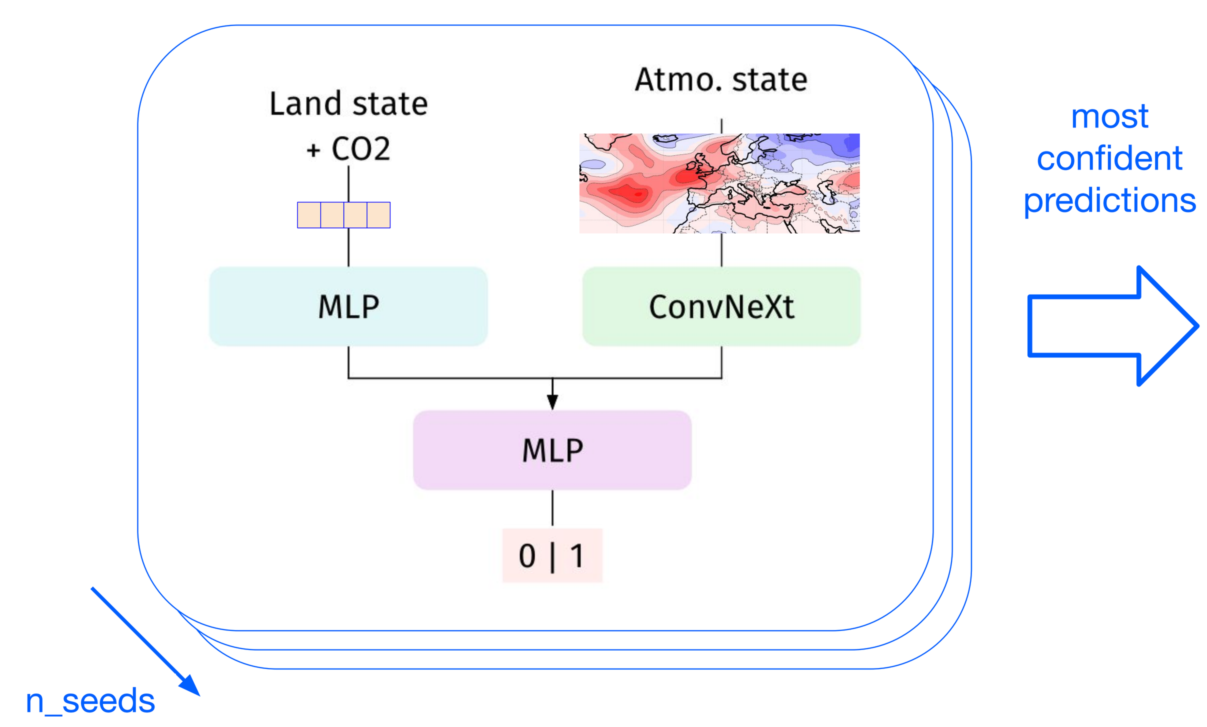
$$(x - \mu\_doy\_loess) / \sigma\_doy\_loess$$

P90 threshold: 90th percentile computed from a 5-day rolling window grouped by day-of-year, LOESS-smoothed. SEAS5 thresholds are computed separately per init month to account for lead-time-dependent biases.





Architecture



Explainability

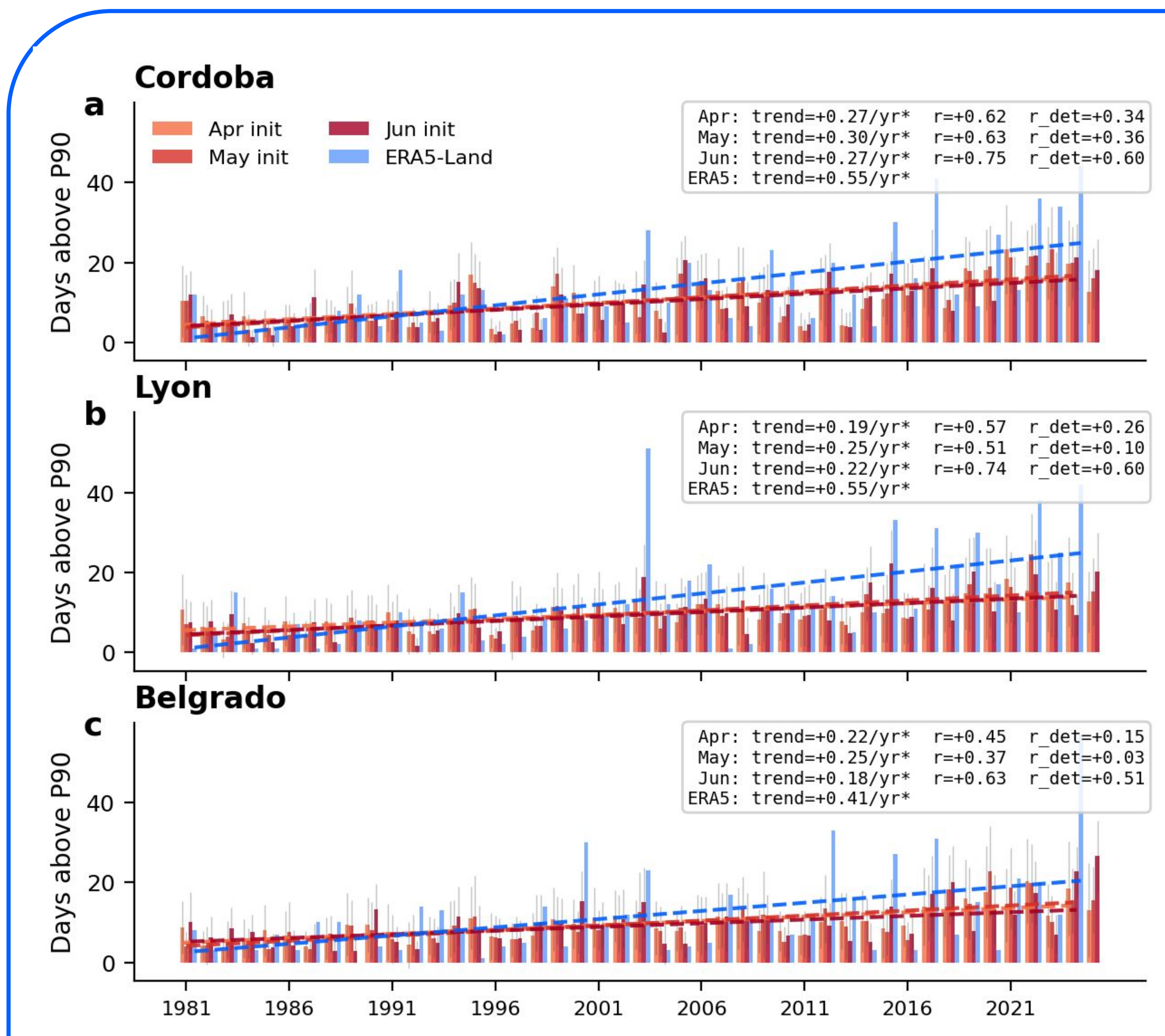
SHAP (GradienExplainer)

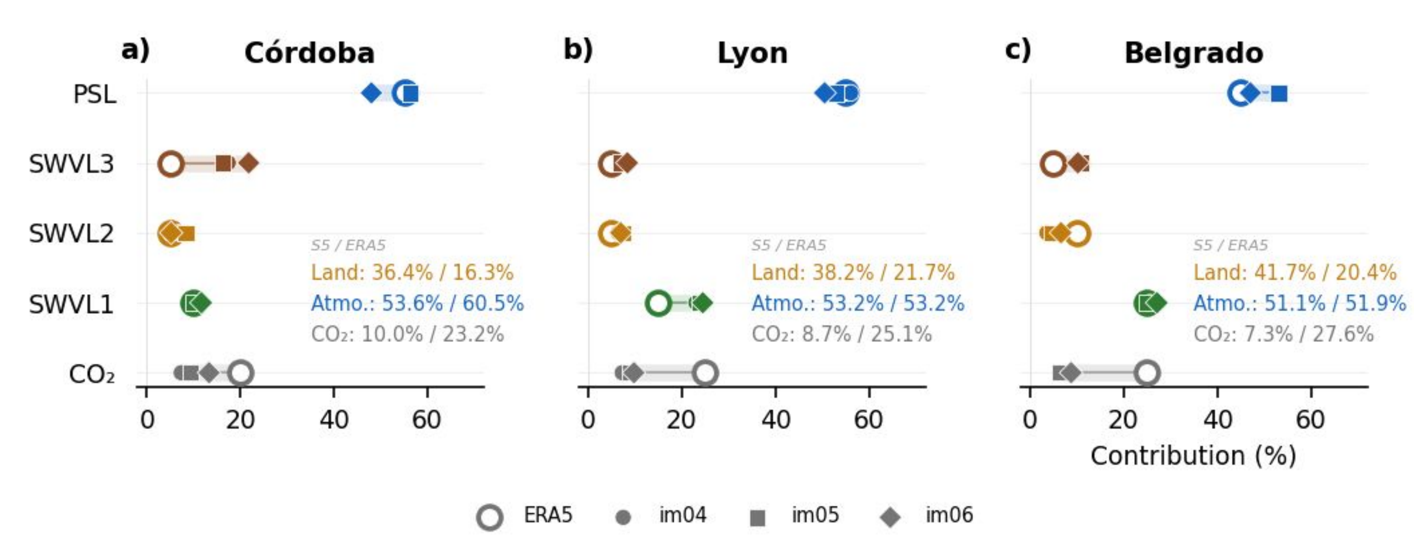
$$f(x) = \phi_0 + \sum_{i}^{N} \phi_i$$

Model Prediction — Baseline — SHAP Value — Number of Features

- SHAP values are additive
- + SHAP value → Increased Likelihood

3. Results



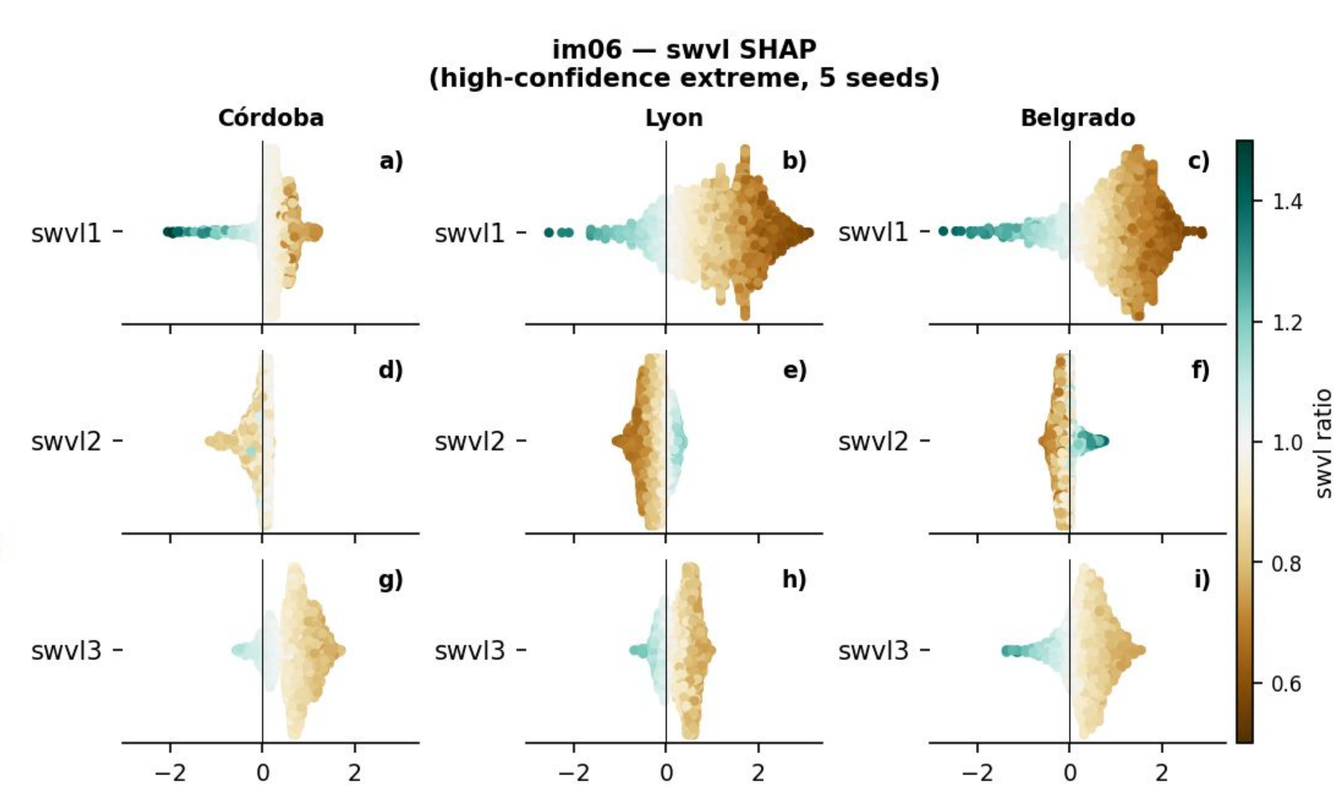


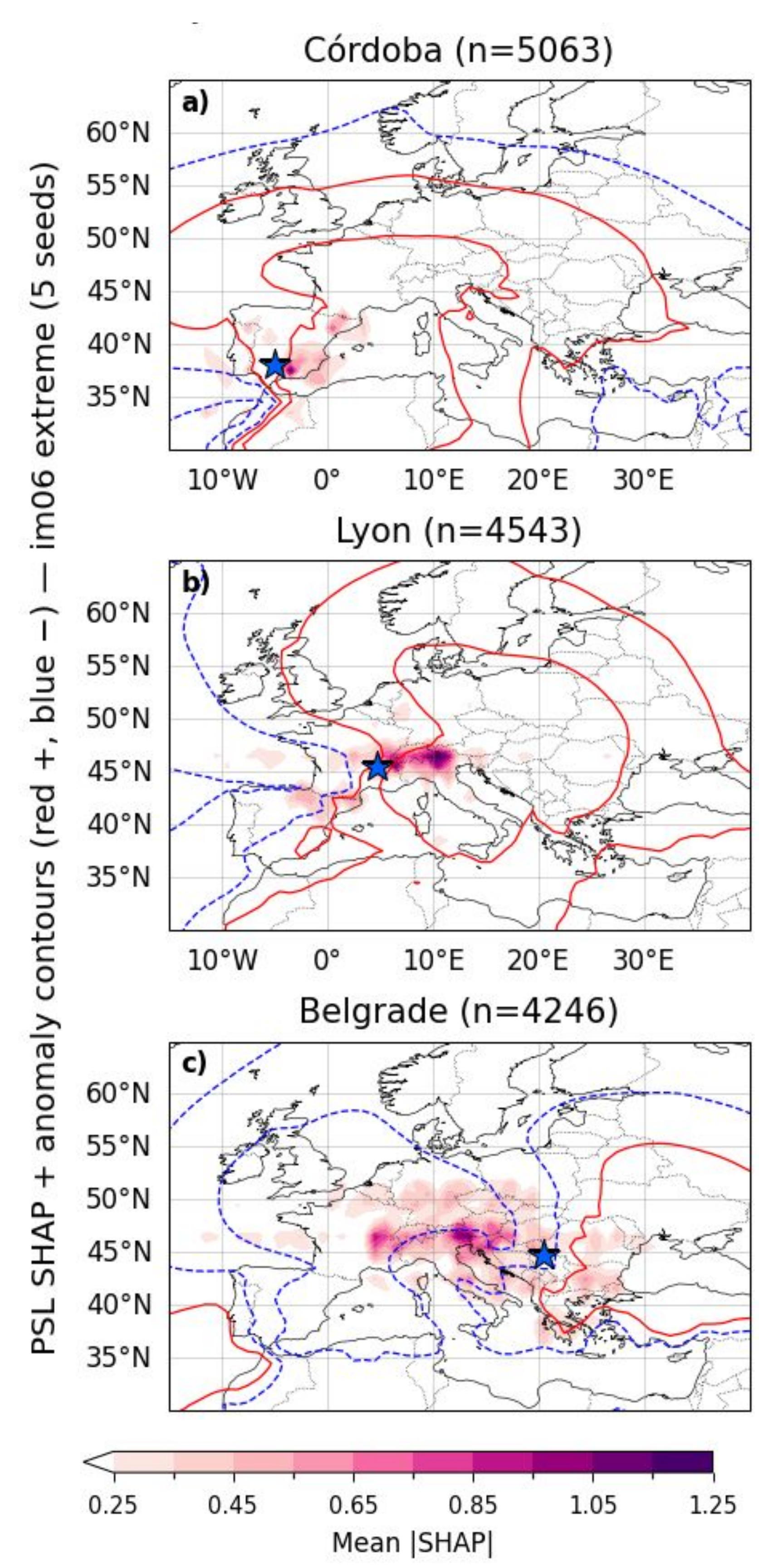
1. The P90 threshold shows little bias (at a coarse resolution) relative to ERA5-Land and is consistent across initialization months, suggesting that SEAS5 retains its ability to model extremes across lead times.

2. At early initialization months the global warming trend sustains skill in predicting JJA days above P90 Tx. Once this trend is removed, only the June initialization retains appreciable skill.

3. Compared to ERA5(-Land), CO_2 shows lower importance as a predictor, consistent with the weaker trend in heat extremes observed. The relationships between drivers and extremes closely agree between ERA5(-Land) and SEAS5.

4. Spatial SHAP values reveal distinct circulation patterns across locations. For example, Córdoba shows localised circulation features, whereas Belgrade shows a more spatially extensive pattern.





[1] Garcia Mesa, A., Palma, L., Donat, M., Materia, S., and Marcos Matamoros, R.: Quantifying atmospheric and land drivers of hot temperature extremes through explainable Artificial Intelligence, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-5392, 2025.

[2] Mahlstein, I., C.Spirig, M. A.Liniger, and C.Appenzeller (2015), Estimating daily climatologies for climate indices derived from climate model data and observations. J. Geophys. Res. Atmos., 120, 2808–2818. doi: 10.1002/2014JD022327.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

lluis.palma@bsc.es