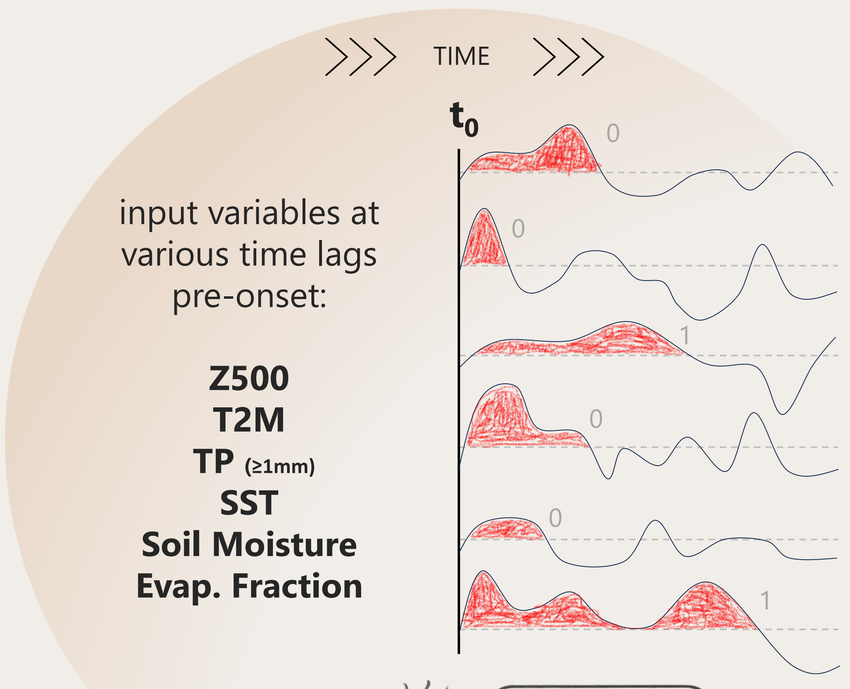
PROBING THE PREDICTABILITY OF **HOT SPELL DURATION** IN EUROPE USING TRADITIONAL ML ALGORITHMS

DUNCAN PAPPERT, MIKEL LEGASA, ALEXANDRE TUEL, PRADEEBANE VAITTINADA AYAR, MATHIEU VRAC, OLIVIA MARTIUS

INTRODUCTION

- Near-surface heat that persists over subseasonal timescales can **severely impact** human and natural systems.
- Existing studies predict heatwave occurrence, but can they capture characteristics like duration?
- Complex **nonlinear dynamics** are involved but remain only partly understood.
- Hypothesis: long hot spells are **preconditioned** (i.e. "set up") for persistence through land, ocean, and atmospheric **signals present before onset**.

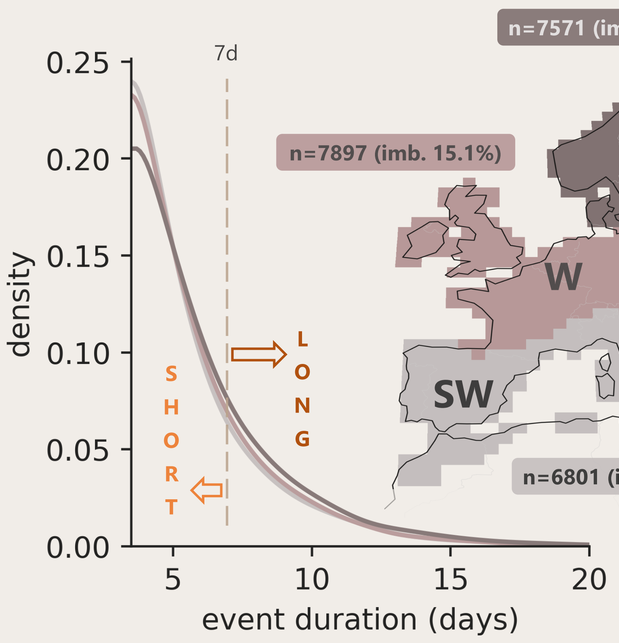


QUESTIONS

WHAT CAN ML ALGORITHMS INFER WHEN ASKED TO PREDICT HOT SPELL DURATION DIRECTLY USING ONLY PRE-ONSET INFORMATION ?

CAN THEY RELIABLY DIFFERENTIATE BETWEEN SHORT AND LONG EVENTS ?

DATA & METHODS



- CESM2-LE historical runs (1951-2000), 50 members
- **probabilistic binary classification** (short=0, long=1) for 3 European regions representing different soil moisture regimes
- input variables cover a large Euro-Atlantic sector, all detrended, deseasonalised and standardised
- splitting procedure across members (10-fold, 90:10)
- **prediction task**: upcoming spell short or long (>7d)?

Tested models:

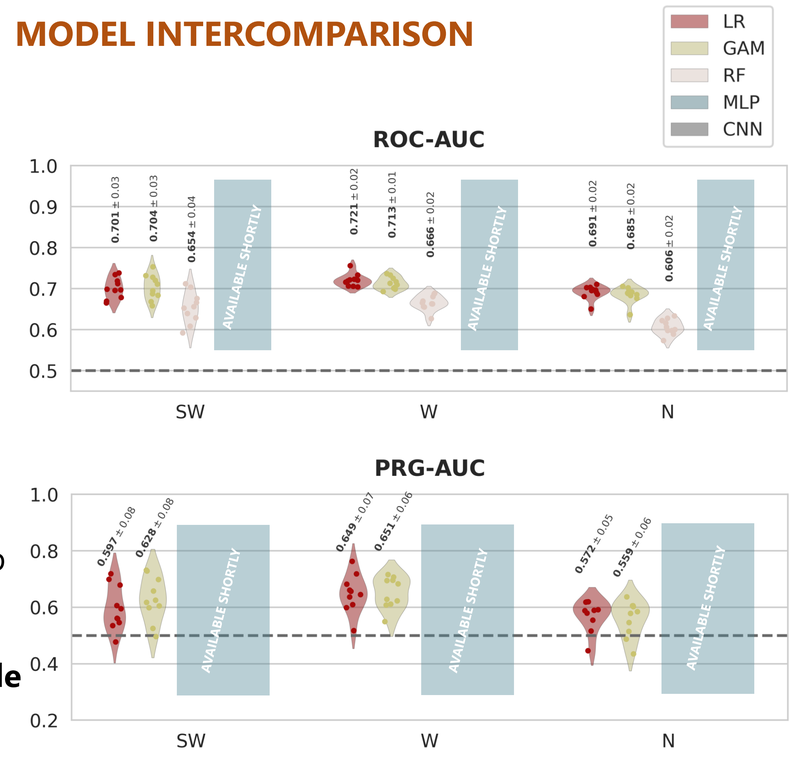
increasing complexity and flexibility

- LOGISTIC REGRESSION (**LR**)
- GEN. ADDITIVE MODEL (**GAM**)
- RANDOM FOREST (**RF**)
- MULTILAYER PERCEPTRON (**MLP**)
- CONVOLUTIONAL NEURAL NETWORK (**CNN**)

RESULTS

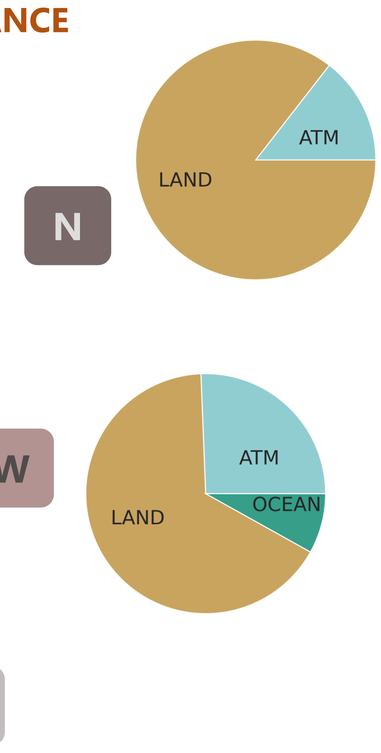
- all regions show above-random skill
- **LR generalises best** overall with a ROC-AUC ~0.7, indicating **modest class discrimination**
- fold variance
- substantial overlap in feature space
- there is **extractable information**, but it is **incomplete**

MODEL INTERCOMPARISON



PERMUTATION FEATURE IMPORTANCE

- Proportion of top-10 features (mean across all test folds) by contribution to model performance (**LR**), grouped by Earth system domain.
- Soil moisture in weeks before spell (both local anomalies and Europe-wide PC patterns) is the dominant driver of the model's predictive signal.



WHAT *IS* GOING ON?

- A **linear model discriminates better** than more complex models, which easily overfit noise and degrade predictive skill. The extractable signal from the provided predictors is **simple and low-dimensional**.
- The best-performing model treats soil moisture more like a decisive predictor than a conditional one, though domain knowledge tells us it represents a facilitating rather than sufficient condition.
- The relatively small sample size (esp. **limited and variable minority class**), the emergent nature of the target, and the constrained training set up pose major challenges. This study could provide a **baseline**.

UNIVERSITÄT BERN
OESCHGER CENTRE CLIMATE CHANGE RESEARCH
LSCE

Contact:
duncan.pappert@unibe.ch